Supplementary material for "Biased-sample empirical likelihood weighting for missing data problems: an alternative to inverse probability weighting"

Yukun Liu[1] and Yan Fan[*2]

[1] KLATASDS-MOE, School of Statistics, East China Normal University Shanghai, China

[2]School of Statistics and Information, Shanghai University of International Business and Economics, Shanghai, China

Abstract

The supplementary material consists of nine sections. Section 1 reviews the conditions and the main results in the main paper: Lemma 2.1, Theorems 2.1 and 2.2, whose proofs are given in Sections 2-4, respectively. In Section 5, we extend the biased-sample empirical likelihood weighting (ELW) method to unequal probability samplings with and without replacement. Theoretically we establish its asymptotic normalities and compare its efficiency with the inverse probability weighting (IPW) and stablized IPW (SIPW) methods in Theorems 5.1 and 5.2, which are proved in Sections 6 and 7, respectively. Section 8 contains additional simulation results. In Section 9, we investigate large-sample properties of the ELW method under over-identified estimating equations.

1 The conditions and the main results

Lemma 2.1. *Suppose $\pi(z_i)$ $(1 \le i \le n)$ take $m \ge 2$ distinct values $\pi_{(1)} < \ldots < \pi_{(m)}$ $(m \ge 2)$. If there exists $\varepsilon \in (0,1)$ such that $\pi_{(m)} - \pi_{(1)} > \varepsilon$ and $n/N < 1-\varepsilon$, then $\kappa \le N/\varepsilon^3$.*

Condition 1. *(i) θ_0 is the unique solution to $\mathbb{E}\{g(Z,\theta)\} = 0$. (ii) The parameter space is a compact set $\Theta \subset \mathbb{R}^r$, $g(Z,\theta)$ is a continuous function of θ for every Z, and there exists a function $\bar{g}(Z)$ such that $\mathbb{E}\{\bar{g}(Z)\} < \infty$ and $\sup_{\theta\in\Theta} \|g(Z,\theta)\| \le \bar{g}(Z)$. (iii) $g(Z,\theta)$ has a continuous partial derivative $g_1(Z,\theta) = \partial g(Z,\theta)/\partial\theta^\top$ in a neighborhood of θ_0 for each Z. There exists a positive function $\bar{g}_1(Z)$ such that $\mathbb{E}\{\bar{g}_1(Z)\} < \infty$ and $\|g_1(Z,\theta)\|_F \le \bar{g}_1(Z)$ for all Z and for β in the neighborhood, where $\|\cdot\|_F$ is the Frobenius norm. (iv)The $r \times r$ matrix $K = \mathbb{E}\{g_1(Z,\theta_0)\}$ is nonsingular.*

[*]Corresponding author: fanyan212@126.com

We denote $A^{\otimes 2} = AA^\top$ for a vector or matrix A, and define

$$B_{gg} = \mathbb{E}\left[\frac{\{g(Z,\theta_0)\}^{\otimes 2}}{\pi(Z)}\right], \quad B_{11} = \mathbb{E}\left\{\frac{1}{\pi(Z)}\right\}, \quad B_{g1} = \mathbb{E}\left\{\frac{g(Z,\theta_0)}{\pi(Z)}\right\}.$$

In the case of $g(Z,\theta) = f(Z) - \theta$, define

$$B_{ff} = \mathbb{E}\left[\frac{\{f(Z)\}^{\otimes 2}}{\pi(Z)}\right], \quad B_{f1} = \mathbb{E}\left\{\frac{f(Z)}{\pi(Z)}\right\}.$$

Theorem 2.1. *Let $\alpha_0 \in (0,1)$ be the truth of α. Suppose that Condition 1 is satisfied, $\mathbb{V}\mathrm{ar}\{\pi(Z)|D=1\} > 0$ and that B_{11} and B_{gg} are both finite. Also suppose that the conditional inclusion probabilities $\pi(Z_i)$ are known. As N goes to infinity,*

(a) $\sqrt{N}(\hat{\theta}_{\mathrm{ELW}} - \theta_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{ELW}})$, where $\Sigma_{\mathrm{ELW}} = K^{-1}\{B_{gg} - B_{g1}^{\otimes 2}/(B_{11}-1)\}(K^{-1})^\top$;

(b) $\sqrt{N}(\hat{\theta}_{\mathrm{SIPW}} - \theta_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{SIPW}})$, where $\Sigma_{\mathrm{SIPW}} = K^{-1}B_{gg}(K^{-1})^\top$;

(c) the ELW estimator $\hat{\theta}_{\mathrm{ELW}}$ is more efficient than the SIPW estimator $\hat{\theta}_{\mathrm{SIPW}}$, i.e. $\Sigma_{\mathrm{ELW}} \le \Sigma_{\mathrm{SIPW}}$, where the equalities hold only if $\pi(Z)$ is degenerate.

(d) If $g(Z,\theta) = f(Z) - \theta$, then $\sqrt{N}(\hat{\theta}_{\mathrm{IPW}} - \theta_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{IPW}})$ with $\Sigma_{\mathrm{IPW}} = B_{ff} - \theta_0^{\otimes 2}$ and $\Sigma_{\mathrm{ELW}} = (B_{ff} - \theta_0^{\otimes 2}) - (B_{f1} - \theta_0)^{\otimes 2}/(B_{11}-1)$; the ELW estimator $\hat{\theta}_{\mathrm{ELW}}$ is also more efficient than the IPW estimator $\hat{\theta}_{\mathrm{IPW}}$.

Condition 2. *(i) There exists β_0 such that $\pi(Z,\beta_0) = \pi(Z)$ for all Z, and the function $\pi(Z,\beta)$ is continuously differentiable in β in a neighborhood of β_0. Let $\pi_1(Z,\beta) = \partial\pi(Z,\beta)/\partial\beta^\top$. (ii) There exist a positive constant ε and positive functions $\bar{\pi}(Z)$ and $\bar{\pi}_1(Z)$ such that $\bar{\pi}(Z) \le \inf_{\beta:\|\beta-\beta_0\|\le\varepsilon} \pi(Z,\beta)$, $\sup_{\beta:\|\beta-\beta_0\|\le\varepsilon} \|\pi_1(Z,\beta)\| \le \bar{\pi}_1(Z)$, $\mathbb{E}\{\pi(Z)/(\bar{\pi}(Z))^2\} < \infty$, $\mathbb{E}\{\pi(Z)\bar{g}(Z)/\bar{\pi}(Z)\} < \infty$, and $\mathbb{E}\left\{\pi(Z)\bar{g}(Z)\bar{\pi}_1(Z)/\{\bar{\pi}(Z)\}^2\right\} < \infty$, where $\bar{g}$ is given in Condition 1.*

Define

$$B_{1\dot{\pi}} = \mathbb{E}\left\{\frac{\pi_1(Z,\beta_0)}{\pi(Z)}\right\}, \quad B_{g\dot{\pi}} = \mathbb{E}\left\{\frac{g(Z,\theta_0)\pi_1(Z,\beta_0)}{\pi(Z)}\right\}.$$

In the case of $g(Z,\theta) = f(Z) - \theta$, define

$$B_{f\dot{\pi}} = \mathbb{E}\left\{\frac{f(Z)\pi_1(Z,\beta_0)}{\pi(Z)}\right\}.$$

Theorem 2.2. *Assume Conditions 1 and 2 and that $\hat{\beta}$ satisfies $\hat{\beta} - \beta_0 = N^{-1}\sum_{i=1}^N h(D_i, Z_i) + o_p(N^{-1/2})$, where the influence function $h(D,Z)$ has zero mean. Suppose that the truth α_0 of α satisfies $0 < \alpha_0 < 1$ and $\mathbb{V}\mathrm{ar}\{\pi(Z,\beta_0)|D=1\} > 0$. As N goes to infinity,*

(a) $\sqrt{N}(\hat{\theta}_{\mathrm{ELW}} - \theta_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{ELW},e})$, where $\Sigma_{\mathrm{ELW},e} = K^{-1}\Omega(K^{-1})^\top$ with

$$\Omega = \mathbb{V}\mathrm{ar}\left\{\frac{Dg(Z,\theta_0)}{\pi(Z,\beta_0)} + \frac{B_{g1}}{B_{11}-1}\left(1 - \frac{D}{\pi(Z,\beta_0)}\right) + \left(\frac{B_{g1}B_{1\dot{\pi}}}{B_{11}-1} - B_{g\dot{\pi}}\right)h(D,Z)\right\}.$$

(b) $\sqrt{N}(\hat{\theta}_{\mathrm{SIPW}} - \theta_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{SIPW},e})$, *where*

$$\Sigma_{\mathrm{SIPW},e} = K^{-1}\mathbb{V}\mathrm{ar}\left\{\frac{Dg(Z,\theta_0)}{\pi(Z,\beta_0)} - B_{g\dot{\pi}}h(D,Z)\right\}(K^{-1})^\top.$$

(c) In the case of $g(Z,\theta) = f(Z) - \theta$, $\sqrt{N}(\hat{\theta}_{\mathrm{IPW}} - \theta_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{IPW},e})$, *where*

$$\Sigma_{\mathrm{IPW},e} = \mathbb{V}\mathrm{ar}\left\{\frac{Df(Z)}{\pi(Z,\beta_0)} - B_{f\dot{\pi}}h(D,Z)\right\};$$

2 Proof of Lemma 2.1

Proof. Let $\xi_i = n/N + (1 - n/N)\pi(z_i)$ for $1 \le i \le n$. The fact $\hat{\alpha} < \min_{1\le i\le n} \xi_i$ guarantees that all $\hat{p}_i$ for $1 \le i \le n$ must be positive. This immediately implies that

$$\kappa = \frac{\max_{1\le i\le n} \hat{p}_i}{\min_{1\le i\le n} \hat{p}_i} = \frac{\max_{1\le i\le n} \xi_i - \hat{\alpha}}{\min_{1\le i\le n} \xi_i - \hat{\alpha}} \le \frac{1}{\min_{1\le i\le n} \xi_i - \hat{\alpha}}, \tag{1}$$

where the inequality follows from $\xi_i \le 1$ for any $1 \le i \le n$.

Let $\xi_{(i)} = n/N + (1 - n/N)\pi_{(i)}$ and k_i be the frequency of $\pi_{(i)}$, $1 \le i \le m$. In the main paper, we have shown that

$$0 = \sum_{i=1}^{n} \frac{\pi(z_i) - \alpha}{\xi_i - \alpha}.$$

This implies

$$k_1 \frac{\hat{\alpha} - \pi_{(1)}}{\xi_{(1)} - \hat{\alpha}} = \sum_{i=2}^{m} k_i \frac{\pi_{(i)} - \hat{\alpha}}{\xi_{(i)} - \hat{\alpha}}.$$

With the facts that $0 \le \pi_{(i)}, \hat{\alpha} < 1$ and $\pi_{(1)} < \hat{\alpha} < \xi_{(1)} < \xi_{(2)} < \ldots < \xi_{(m)}$, we further have

$$k_1 \frac{\hat{\alpha} - \pi_{(1)}}{\xi_{(1)} - \hat{\alpha}} \le \sum_{i=2}^{m} k_i \frac{|\pi_{(i)} - \hat{\alpha}|}{\xi_{(i)} - \hat{\alpha}} \le \sum_{i=2}^{m} \frac{k_i}{\xi_{(2)} - \xi_{(1)}} = \frac{n - k_1}{\xi_{(2)} - \xi_{(1)}},$$

which implies

$$\frac{1}{\xi_{(1)} - \hat{\alpha}} \le \frac{1}{\xi_{(1)} - \pi_{(1)}} \times \left\{\frac{n - k_1}{k_1(\xi_{(2)} - \xi_{(1)})} + 1\right\}.$$

It follows from $\pi_{(2)} - \pi_{(1)} > \varepsilon$ that $(1 - \pi_{(1)}) > \epsilon$, $\xi_{(1)} - \pi_{(1)} = (n/N)(1 - \pi_{(1)}) > n\varepsilon/N$ and $\xi_{(2)} - \xi_{(1)} \ge (1 - n/N)\varepsilon \ge \varepsilon^2$. This together with (1) implies that

$$\kappa = \frac{\max_{1\le i\le n} \hat{p}_i}{\min_{1\le i\le n} \hat{p}_i} \le \frac{1}{\xi_{(1)} - \hat{\alpha}} \le \frac{N}{n\varepsilon} \times \left(\frac{n - k_1}{k_1\varepsilon^2} + 1\right) \le \frac{N}{k_1\varepsilon^3} \le \frac{N}{\varepsilon^3}.$$

□

3 Proof of Theorem 2.1

The proofs of results (b)-(d) are straightforward and omitted. We prove result (a) only. The proof is divided into several steps. We first prove the root-N consistency of $\hat{\alpha}$, and derive linear approximates for $\hat{\alpha}$ and $\lambda(\hat{\alpha})$. Then we prove the consistency of $\hat{\theta}_{\mathrm{ELW}}$, and derive a linear approximate for it. Finally, we prove the asymptotic normality of $\hat{\theta}_{\mathrm{ELW}}$ based on the linear approximate.

We first present a technical lemma, which quantifies the magnitude of $\max_{1\le i\le N}\{D_i/\pi(Z_i)\}$.

Lemma 1. *If* $\mathbb{E}\{\pi(Z)\}^{-1} < \infty$, *then* $\max_{1\le i\le N}\{D_i/\pi(Z_i)\} = o(N^{1/2})$,

Proof. Because $\mathbb{E}\{D/\pi(Z)\}^2 = \mathbb{E}\{\pi(Z)\}^{-1} < \infty$, by Lemma 3 of Owen (1990), we immediately have $\max_{1\le i\le N}\{D_i/\pi(Z_i)\} = o(N^{1/2})$. $\square$

3.1 Root-N consistency of $\hat{\alpha}$

By definition, $\hat{\alpha}$ satisfies

$$\frac{1}{N}\sum_{i=1}^{N}\frac{D_i\{\pi(Z_i)-\hat{\alpha}\}}{(n/N)+(1-n/N)\pi(Z_i)-\hat{\alpha}} = 0 \tag{2}$$

and $\min_{1\le i\le n}\pi(z_i) < \hat{\alpha} < n/N + (1-n/N)\min_{1\le i\le n}\pi(z_i)$.

Let $\hat{\gamma} = \{(n/N)-\hat{\alpha}\}/(1-n/N)$. Equation (2) can be rewritten as

$$\begin{aligned} 0 &= \frac{1}{N}\sum_{i=1}^{N}\frac{D_i}{\pi(Z_i)+\hat{\gamma}}\left\{\pi(Z_i)-\frac{n}{N}+\left(1-\frac{n}{N}\right)\hat{\gamma}\right\} \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{D_i}{1+\hat{\gamma}D_i/\pi(Z_i)}\left\{1-\frac{1}{\pi(Z_i)}\frac{n}{N}+\frac{1}{\pi(Z_i)}\left(1-\frac{n}{N}\right)\hat{\gamma}\right\} \\ &= \frac{1}{N}\sum_{i=1}^{N}D_i\left\{1-\frac{1}{\pi(Z_i)}\frac{n}{N}\right\}+\frac{n}{N}\cdot\frac{1}{N}\sum_{i=1}^{N}\frac{D_i/\pi(Z_i)}{1+\hat{\gamma}D_i/\pi(Z_i)}\left\{\frac{1}{\pi(Z_i)}-1\right\}\hat{\gamma}, \end{aligned} \tag{3}$$

which means

$$\left|\frac{1}{N}\sum_{i=1}^{N}D_i\left\{1-\frac{1}{\pi(Z_i)}\frac{n}{N}\right\}\right| = \left|\frac{n}{N}\cdot\frac{1}{N}\sum_{i=1}^{N}\frac{D_i/\pi(Z_i)}{1+\hat{\gamma}D_i/\pi(Z_i)}\left\{\frac{1}{\pi(Z_i)}-1\right\}\hat{\gamma}\right|. \tag{4}$$

Because the left-hand side of equation (4) is clearly $O_p(n^{-1/2})$, it follows from equation (4) that

$$\begin{aligned} O_p(n^{-1/2}) &\geq \frac{1}{1+|\hat{\gamma}|\max_{1\le i\le N}D_i/\pi(Z_i)}\cdot\frac{1}{N}\sum_{i=1}^{N}\frac{D_i}{\pi(Z_i)}\frac{1-\pi(Z_i)}{\pi(Z_i)}\cdot\frac{n}{N}\cdot|\hat{\gamma}| \\ &= \frac{1}{1+|\hat{\gamma}|\cdot o(N^{1/2})}\cdot O_p(1)\cdot|\hat{\gamma}|, \end{aligned}$$

which implies $\hat{\gamma} = O_p(n^{-1/2})$. Because $n/N - \alpha_0 = O_p(N^{-1/2})$, we have

$$\hat{\alpha}-\alpha_0 = \left(\frac{n}{N}-\alpha_0\right)-\hat{\gamma}\left(1-\frac{n}{N}\right) = O_p(N^{-1/2}).$$

3.2 Approximate of $\hat{\alpha}$

The order $\hat{\gamma} = O_p(N^{-1/2})$ together with $\max_{1\le i\le N} D_i/\pi(Z_i) = o_p(N^{1/2})$ (see Lemma 1) implies $\hat{\gamma}\max_{1\le i\le N} D_i/\pi(Z_i) = o_p(1)$. Therefore equation (3) becomes

$$\begin{aligned}
0 &= \frac{1}{N}\sum_{i=1}^{N} D_i\left\{1-\frac{1}{\pi(Z_i)}\frac{n}{N}\right\} + \frac{n}{N}\cdot\frac{1}{N}\sum_{i=1}^{N}\frac{D_i}{\pi(Z_i)}\left\{\frac{1}{\pi(Z_i)}-1\right\}\hat{\gamma} + o_p(N^{-1/2}) \\
&= \frac{\alpha_0}{N}\sum_{i=1}^{N}\left\{1-\frac{D_i}{\pi(Z_i)}\right\} + \alpha_0(B_{11}-1)\hat{\gamma} + o_p(N^{-1/2}) \\
&= \frac{\alpha_0}{N}\sum_{i=1}^{N}\left\{1-\frac{D_i}{\pi(Z_i)}\right\} + \alpha_0(B_{11}-1)\frac{(n/N)-\hat{\alpha}}{1-(n/N)} + o_p(N^{-1/2}) \\
&= \frac{\alpha_0}{N}\sum_{i=1}^{N}\left\{1-\frac{D_i}{\pi(Z_i)}\right\} + \frac{\alpha_0}{1-\alpha_0}(B_{11}-1)\left(\frac{n}{N}-\hat{\alpha}\right) + o_p(N^{-1/2}).
\end{aligned}$$

This implies

$$\begin{aligned}
\hat{\alpha}-\alpha_0 &= \frac{1}{N}\sum_{i=1}^{N}(D_i-\alpha_0) + \frac{1-\alpha_0}{B_{11}-1}\frac{1}{N}\sum_{i=1}^{N}\left\{1-\frac{D_i}{\pi(Z_i)}\right\} + o_p(N^{-1/2}) \\
&= \frac{1-\alpha_0}{\alpha_0(B_{11}-1)}\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{D_i\{\pi(Z_i)-\alpha_0\}}{\pi(Z_i)} - \frac{1-\alpha_0 B_{11}}{1-\alpha_0}(D_i-\alpha_0)\right\} + o_p(N^{-1/2}).
\end{aligned}$$

3.3 Approximate of $\lambda(\hat{\alpha}) = (N-n)/\{n(1-\hat{\alpha})\}$

Because $\delta_1 = \hat{\alpha}-\alpha_0 = O_p(N^{-1/2})$ and $\delta_2 = n/N-\alpha_0 = O_p(N^{-1/2})$, we have

$$\lambda(\hat{\alpha})-\alpha_0^{-1} = \frac{(1-\alpha_0-\delta_2)}{(\alpha_0+\delta_2)(1-\alpha_0-\delta_1)} - \alpha_0^{-1} = \frac{\delta_1}{\alpha_0(1-\alpha_0)} - \frac{\delta_2}{\alpha_0^2(1-\alpha_0)} + o_p(N^{-1/2}).$$

Let $\delta_3 = \lambda(\hat{\alpha})-\alpha_0^{-1}$. Putting the approximations of δ_1 and δ_2 into the above equation gives

$$\begin{aligned}
\delta_3 &= \frac{1}{\alpha_0(1-\alpha_0)}\times\frac{1}{N}\sum_{i=1}^{N}(D_i-\alpha_0) + \frac{1}{\alpha_0(1-\alpha_0)}\times\frac{1-\alpha_0}{B_{11}-1}\frac{1}{N}\sum_{i=1}^{N}\left\{1-\frac{D_i}{\pi(Z_i)}\right\} \\
&\quad -\frac{1}{\alpha_0^2(1-\alpha_0)}\frac{1}{N}\sum_{i=1}^{N}(D_i-\alpha_0) + o_p(N^{-1/2}) \\
&= -\frac{1}{\alpha_0^2}\times\frac{1}{N}\sum_{i=1}^{N}(D_i-\alpha_0) + \frac{1}{\alpha_0(B_{11}-1)}\frac{1}{N}\sum_{i=1}^{N}\left\{1-\frac{D_i}{\pi(Z_i)}\right\} + o_p(N^{-1/2}) \\
&= \frac{1}{\alpha_0^2(B_{11}-1)}\frac{1}{N}\sum_{i=1}^{N}\left[\frac{D_i\{\pi(Z_i)-\alpha_0\}}{\pi(Z_i)} - B_{11}(D_i-\alpha_0)\right] + o_p(N^{-1/2}).
\end{aligned}$$

3.4 Consistency of $\hat{\theta}_{\text{ELW}}$

The ELW estimator $\hat{\theta}_{\text{ELW}}$ is the solution to

$$\sum_{i=1}^{N} \hat{p}_i g(Z_i, \theta) = \sum_{i=1}^{N} \frac{1}{n} \frac{D_i}{1 + \lambda(\hat{\alpha})\{\pi(Z_i) - \hat{\alpha}\}} g(Z_i, \theta) = 0. \tag{5}$$

Define

$$\Delta_n(\theta) = \sum_{i=1}^{N} \hat{p}_i g(Z_i, \theta) - \mathbb{E}\{g(Z, \theta)\} = \Delta_{n1}(\theta) + \Delta_{n2}(\theta),$$

where

$$\begin{aligned}
\Delta_{n1}(\theta) &= \sum_{i=1}^{N} \hat{p}_i g(Z_i, \theta) - \frac{1}{N} \sum_{i=1}^{N} \frac{D_i}{\pi(Z_i)} g(Z_i, \theta), \\
\Delta_{n2}(\theta) &= \frac{1}{N} \sum_{i=1}^{N} \frac{D_i}{\pi(Z_i)} g(Z_i, \theta) - \mathbb{E}\{g(Z, \theta)\}.
\end{aligned}$$

By Theorem 5.9 of van der Vaart (2000), a sufficient condition for the consistency of $\hat{\theta}_{\text{ELW}}$ is that (1) $\inf_{\theta: \|\theta - \theta_0\| > \varepsilon} \|\mathbb{E} g(Z, \theta)\| > 0$ for every $\varepsilon > 0$, and (2) $\sup_{\theta \in \Theta} \|\Delta_n(\theta)\| = o_p(1)$. As the first point is guaranteed by Condition 1(a), it suffices to prove the second point.

Because $\|g(Z, \theta)\| D/\pi(Z) \le \bar{g}(Z) D/\pi(Z)$ and $\mathbb{E}\{\bar{g}(Z) D/\pi(Z)\} = \mathbb{E}\{\bar{g}(Z)\}$, under Condition 1(b), the result of Example 19.8 of van der Vaart (2000) applies to $\Delta_{n2}(\theta)$, i.e. $\sup_{\theta \in \Theta} \|\Delta_{n2}(\theta)\| = o_p(1)$. As $\|\Delta_n(\theta)\| \le \|\Delta_{n1}(\theta)\| + \|\Delta_{n2}(\theta)\|$, to prove $\sup_{\theta \in \Theta} \|\Delta_n(\theta)\| = o_p(1)$, it suffices to prove

$$\sup_{\theta \in \Theta} \|\Delta_{n1}(\theta)\| = o_p(1).$$

Note that

$$\begin{aligned}
\Delta_{n1}(\theta) &= \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{1}{n/N} \frac{D_i}{1 + \lambda(\hat{\alpha})\{\pi(Z_i) - \hat{\alpha}\}} - \frac{D_i}{\pi(Z_i)} \right\} g(Z_i, \theta) \\
&= \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{1}{\alpha_0 + \delta_2} \frac{D_i}{1 + (\alpha_0^{-1} + \delta_3)\{\pi(Z_i) - (\alpha_0 + \delta_1)\}} - \frac{D_i}{\pi(Z_i)} \right\} g(Z_i, \theta) \\
&= \frac{1}{N} \sum_{i=1}^{N} \xi_{Ni} \frac{D_i}{\pi(Z_i)} g(Z_i, \theta)
\end{aligned}$$

where

$$\xi_{Ni} = \frac{D_i}{\alpha_0 + \delta_2} \cdot \frac{\pi(Z_i)}{1 + (\alpha_0^{-1} + \delta_3)\{\pi(Z_i) - (\alpha_0 + \delta_1)\}} - D_i.$$

It can be verified that

$$\begin{aligned}
\xi_{Ni} &= \frac{D_i}{1 + \delta_2 \alpha_0^{-1}} \cdot \frac{1}{1 + \alpha_0 \delta_3 - (\delta_1 + \alpha_0^2 \delta_3 + \alpha_0 \delta_2 \delta_3) \frac{D_i}{\pi(Z_i)}} - D_i \\
&= D_i \cdot \frac{\alpha_0 \delta_3 + \delta_2 \alpha_0^{-1}(1 + \alpha_0 \delta_3) - (1 + \delta_2 \alpha_0^{-1})(\delta_1 + \alpha_0^2 \delta_3 + \alpha_0 \delta_2 \delta_3) \frac{D_i}{\pi(Z_i)}}{(1 + \delta_2 \alpha_0^{-1}) \left\{ 1 + \alpha_0 \delta_3 - (\delta_1 + \alpha_0^2 \delta_3 + \alpha_0 \delta_2 \delta_3) \frac{D_i}{\pi(Z_i)} \right\}},
\end{aligned}$$

where the first equality holds because D_i takes only two values 0 and 1. Further,

$$\max_{1\le i\le N}|\xi_{Ni}| \le \frac{\alpha_0|\delta_3|+|\delta_2|\alpha_0^{-1}(1+\alpha_0|\delta_3|)+(1+|\delta_2|\alpha_0^{-1})(|\delta_1|+\alpha_0^2|\delta_3|+\alpha_0|\delta_2\delta_3|)\max_{1\le i\le N}\frac{D_i}{\pi(Z_i)}}{(1-|\delta_2|\alpha_0^{-1})\left\{1-\alpha_0|\delta_3|-|\delta_1+\alpha_0^2\delta_3+\alpha_0\delta_2\delta_3|\cdot\max_{1\le i\le N}\frac{D_i}{\pi(Z_i)}\right\}}.$$

We have shown that $\delta_1 = O_p(N^{-1/2})$, $\delta_2 = O_p(N^{-1/2})$, and $\delta_3 = O_p(N^{-1/2})$. By Lemma 1, $\max_{1\le i\le N} D_i/\pi(Z_i) = o(N^{1/2})$. Therefore

$$\max_{1\le i\le N}|\xi_{Ni}| \le \frac{O_p(N^{-1/2})+O_p(N^{-1/2})+O_p(N^{-1/2})\cdot o(N^{1/2})}{\{1-O_p(N^{-1/2})\}\left\{1-O_p(N^{-1/2})-O_p(N^{-1/2})\cdot o(N^{1/2})\right\}} = o_p(1).$$

Because ξ_{Ni} is independent of θ, we have

$$\begin{aligned}\sup_{\theta\in\Theta}\|\Delta_{n1}(\theta)\| &= \sup_{\theta\in\Theta}\left\|\frac{1}{N}\sum_{i=1}^N \xi_{Ni}\frac{D_i}{\pi(Z_i)}g(Z_i,\theta)\right\| \\ &\le \sup_{\theta\in\Theta}\frac{1}{N}\sum_{i=1}^N|\xi_{Ni}|\left\|\frac{D_i}{\pi(Z_i)}g(Z_i,\theta)\right\| \\ &\le \max_{1\le i\le N}|\xi_{Ni}|\cdot\frac{1}{N}\sum_{i=1}^N\frac{D_i}{\pi(Z_i)}\bar{g}(Z_i) \\ &= o_p(1)\times O_p(1) = o_p(1),\end{aligned}$$

where the last equality holds because $\mathbb{E}\{\bar{g}(Z)D/\pi(Z)\} = \mathbb{E}\{\bar{g}(Z)\} < \infty$. This proves the uniformly convergence of $\sum_{i=1}^N \hat{p}_i g(Z_i,\theta)$ to $\mathbb{E}\{g(Z,\theta)\}$ for all $\theta\in\Theta$, and hence the consistency of $\hat{\theta}_{\mathrm{ELW}}$.

3.5 Asymptotic normality of $\hat{\theta}_{\mathrm{ELW}}$

Applying a first-order Taylor expansion to the left-hand side of equation (5), we have

$$\sum_{i=1}^N \hat{p}_i g(Z_i,\theta_0) + \sum_{i=1}^N \hat{p}_i g_1(Z_i,\theta_*)(\hat{\theta}_{\mathrm{ELW}}-\theta_0) = 0, \tag{6}$$

where $\theta_* = \theta_0\rho + \hat{\theta}_{\mathrm{ELW}}(1-\rho)$ with $0\le\rho\le 1$.

In the proof of the consistency of $\hat{\theta}_{\mathrm{ELW}}$, we have proved that

$$\hat{p}_i = \frac{D_i}{\pi(Z_i)} + \frac{D_i}{\pi(Z_i)}\xi_{Ni}$$

and $\max_{1\le i\le N}\xi_{Ni} = o_p(1)$. Under Condition 1 (b) and (c), equation (6) becomes

$$\sum_{i=1}^N \hat{p}_i g(Z_i,\theta_0) + K(\hat{\theta}_{\mathrm{ELW}}-\theta_0) + o_p(N^{-1/2}) = 0,$$

which implies

$$\hat{\theta}_{\mathrm{ELW}}-\theta_0 = -K^{-1}\sum_{i=1}^N \hat{p}_i g(Z_i,\theta_0) + o_p(N^{-1/2}).$$

Using the expression of $\hat{p}_i$, we have

$$\begin{aligned}\sum_{i=1}^{N} \hat{p}_i g(Z_i, \theta_0) &= \sum_{i=1}^{N} \frac{1}{n} \frac{D_i}{1 + \lambda(\hat{\alpha})\{\pi(Z_i) - \hat{\alpha}\}} g(Z_i, \theta_0) \\ &= \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\alpha_0 + \delta_2} \frac{D_i}{1 + (\alpha_0^{-1} + \delta_3)\{\pi(Z_i) - (\alpha_0 + \delta_1)\}} g(Z_i, \theta_0).\end{aligned}$$

By the first-order Taylor expansion, we have

$$\begin{aligned}\sum_{i=1}^{N} \hat{p}_i g(Z_i, \theta_0) &= \frac{1}{N} \sum_{i=1}^{N} \frac{D_i g(Z_i, \theta_0)}{\pi(Z_i)} + \frac{1}{\alpha_0^2} \cdot \frac{1}{N} \sum_{i=1}^{N} \frac{D_i g(Z_i, \theta_0)}{[1 + \alpha_0^{-1}\{\pi(Z_i) - \alpha_0\}]^2} \cdot \delta_1 \\ &\quad - \frac{1}{\alpha_0^2} \cdot \frac{1}{N} \sum_{i=1}^{N} \frac{D_i g(Z_i, \theta_0)}{1 + \alpha_0^{-1}\{\pi(Z_i) - \alpha_0\}} \cdot \delta_2 \\ &\quad - \frac{1}{\alpha_0} \cdot \frac{1}{N} \sum_{i=1}^{N} \frac{D_i g(Z_i, \theta_0)\{\pi(Z_i) - \alpha_0\}}{[1 + \alpha_0^{-1}\{\pi(Z_i) - \alpha_0\}]^2} \cdot \delta_3 + o_p(N^{-1/2}) \\ &= \frac{1}{N} \sum_{i=1}^{N} \frac{D_i g(Z_i, \theta_0)}{\pi(Z_i)} + \alpha_0^2 B_{g1} \cdot \delta_3 + B_{g1}\delta_1 + o_p(N^{-1/2}),\end{aligned}$$

where $B_{g1} = \mathbb{E}\{g(Z, \theta_0)/\pi(Z)\}$.

Using $\delta_2 = (1/N)\sum_{i=1}^{N}(D_i - \alpha_0)$, and the approximations

$$\begin{aligned}\delta_1 &= \frac{1 - \alpha_0}{\alpha_0(B_{11} - 1)} \frac{1}{N} \sum_{i=1}^{N} \left[\frac{D_i\{\pi(Z_i) - \alpha_0\}}{\pi(Z_i)} - \frac{1 - \alpha_0 B_{11}}{1 - \alpha_0}(D_i - \alpha_0)\right] + o_p(N^{-1/2}), \\ \delta_3 &= \frac{1}{\alpha_0^2(B_{11} - 1)} \frac{1}{N} \sum_{i=1}^{N} \left[\frac{D_i\{\pi(Z_i) - \alpha_0\}}{\pi(Z_i)} - B_{11}(D_i - \alpha_0)\right] + o_p(N^{-1/2}).\end{aligned}$$

we have

$$\begin{aligned}\sum_{i=1}^{N} \hat{p}_i g(Z_i, \theta_0) &= \frac{1}{N} \sum_{i=1}^{N} \frac{D_i g(Z_i, \theta_0)}{\pi(Z_i)} + \frac{B_{g1}\alpha_0}{\alpha_0(B_{11} - 1)} \frac{1}{N} \sum_{i=1}^{N} \left[\frac{D_i\{\pi(Z_i) - \alpha_0\}}{\pi(Z_i)} - B_{11}(D_i - \alpha_0)\right] \\ &\quad + \frac{B_{g1}(1 - \alpha_0)}{\alpha_0(B_{11} - 1)} \frac{1}{N} \sum_{i=1}^{N} \left[\frac{D_i\{\pi(Z_i) - \alpha_0\}}{\pi(Z_i)} - \frac{1 - \alpha_0 B_{11}}{1 - \alpha_0}(D_i - \alpha_0)\right] + o_p(N^{-1/2}) \\ &= \frac{1}{N} \sum_{i=1}^{N} \frac{D_i g(Z_i, \theta_0)}{\pi(Z_i)} + \frac{B_{g1}}{\alpha_0(B_{11} - 1)} \frac{1}{N} \sum_{i=1}^{N} \frac{D_i\{\pi(Z_i) - \alpha_0\}}{\pi(Z_i)} \\ &\quad - \frac{B_{g1}}{\alpha_0(B_{11} - 1)} \frac{1}{N} \sum_{i=1}^{N} (D_i - \alpha_0) + o_p(N^{-1/2}) \\ &= \frac{1}{N} \sum_{i=1}^{N} \left\{\frac{D_i g(Z_i, \theta_0)}{\pi(Z_i)} - \frac{B_{g1}}{B_{11} - 1}\left(\frac{D_i}{\pi(Z_i)} - 1\right)\right\} + o_p(N^{-1/2}).\end{aligned}$$

Therefore

$$\hat{\theta}_{\mathrm{ELW}} - \theta_0 = -K^{-1} \frac{1}{N} \sum_{i=1}^{N} \left\{\frac{D_i}{\pi(Z_i)} g(Z_i, \theta_0) - \frac{B_{g1}}{B_{11} - 1}\left(\frac{D_i}{\pi(Z_i)} - 1\right)\right\} + o_p(N^{-1/2}).$$

By Slutsky's theorem and the central limit theorem, we have

$$\sqrt{N}(\hat{\theta}_{\mathrm{ELW}} - \theta_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{ELW}}),$$

where $\Sigma_{\mathrm{ELW}} = K^{-1}\mathbb{V}\mathrm{ar}(u_{1i} - B_{g1}(B_{11}-1)^{-1}u_{2i})(K^{-1})^\top = K^{-1}\{B_{gg} - B_{g1}^{\otimes 2}/(B_{11}-1)\}(K^{-1})^\top$.

4 Proof of Theorem 2.2

We prove only result (a) and omit proofs of results (b) and (c). The proof of result (a) is divided into five steps. We first prove the root-n consistency and derive an approximate of $\hat{\alpha}$ in Sections 4.1 and 4.2. Based on these results, we derive an approximate of $\hat{\lambda}$ in Section 4.3. The consistency and asymptotic normality of $\hat{\theta}_{\mathrm{ELW}}$ are proved in Sections 4.4 and 4.5.

4.1 Consistency of $\hat{\alpha}$

We prove the root-N consistency of $\hat{\alpha}$ by a similar proof to that in Section 3.1. Recall that $\hat{\alpha}$ is the solution to

$$\frac{1}{N}\sum_{i=1}^{N} \frac{D_i\{\pi(Z_i,\hat{\beta}) - \hat{\alpha}\}}{(n/N) + (1-n/N)\pi(Z_i,\hat{\beta}) - \hat{\alpha}} = 0. \tag{7}$$

Let $\hat{\gamma} = \{(n/N) - \hat{\alpha}\}/(1-n/N)$. Similar to the proof in Section 3.1, it follows from equation (7) that

$$J_1 = -J_2\hat{\gamma}, \tag{8}$$

where

$$J_1 = \frac{1}{N}\sum_{i=1}^{N} D_i\left\{1 - \frac{1}{\pi(Z_i,\hat{\beta})}\frac{n}{N}\right\}, \quad J_2 = -\frac{n}{N}\cdot\frac{1}{N}\sum_{i=1}^{N}\frac{D_i/\pi(Z_i,\hat{\beta})}{1+\hat{\gamma}D_i/\pi(Z_i,\hat{\beta})}\left\{\frac{1}{\pi(Z_i,\hat{\beta})} - 1\right\}.$$

Next we identify the magnitudes of J_1 and J_2. It can be seen that

$$J_1 = \frac{1}{N}\sum_{i=1}^{N} D_i\left\{1 - \frac{1}{\pi(Z_i,\beta_0)}\frac{n}{N}\right\} + \frac{n}{N}\cdot\frac{1}{N}\sum_{i=1}^{N} D_i\frac{\pi_1(Z_i,\tilde{\beta})}{\{\pi(Z_i,\tilde{\beta})\}^2}(\hat{\beta}-\beta_0),$$

where $\tilde{\beta} = \rho\beta_0 + (1-\rho)\hat{\beta}$ with $0 \le \rho \le 1$. Condition 2(b) guarantees that

$$\sup_{\beta:\|\beta-\beta_0\|<\varepsilon}\left\|\frac{1}{N}\sum_{i=1}^{N} D_i\frac{\pi_1(Z_i,\beta)}{\{\pi(Z_i,\beta)\}^2} - \mathbb{E}\frac{\pi(Z)\pi_1(Z,\beta)}{(\pi(Z,\beta))^2}\right\| = o_p(1).$$

The consistency of $\hat{\beta}$ implies that $\tilde{\beta} = \beta_0 + o_p(1)$. This together with the continuity of $\pi_1(Z,\beta)$ in Condition 2(i) implies that

$$\frac{1}{N}\sum_{i=1}^{N} D_i\frac{\pi_1(Z_i,\tilde{\beta})}{\{\pi(Z_i,\tilde{\beta})\}^2} = \mathbb{E}\frac{\pi(Z)\pi_1(Z,\beta_0)}{(\pi(Z,\beta_0))^2} + o_p(1) = \mathbb{E}\frac{\pi_1(Z,\beta_0)}{\pi(Z,\beta_0)} + o_p(1) = B_{\dot{\pi}1} + o_p(1).$$

Therefore

$$J_1 \;=\; \frac{1}{N}\sum_{i=1}^{N} D_i \left\{1 - \frac{1}{\pi(Z_i,\beta_0)}\frac{n}{N}\right\} + \alpha_0 B_{\dot{\pi}1}(\hat{\beta}-\beta_0) + o_p(N^{-1/2}) = O_p(N^{-1/2}).$$

For J_2, it can be seen that

$$|J_2| \;\geq\; \frac{1}{1+|\hat{\gamma}|\max_{1\leq i\leq N} D_i/\pi(Z_i,\hat{\beta})}\cdot\frac{n}{N}\cdot\frac{1}{N}\sum_{i=1}^{N}\frac{D_i}{\pi(Z_i,\hat{\beta})}\left\{\frac{1}{\pi(Z_i,\hat{\beta})}-1\right\}.$$

Because under Condition 2(ii), $\mathbb{E}\{\pi(Z)/(\bar{\pi}(Z))^2\}<\infty$ and $\bar{\pi}(Z)\leq\inf_\beta \pi(Z,\beta)$, we have

$$\sup_{\beta:\|\beta-\beta_0\|<\varepsilon}\left|\frac{1}{N}\sum_{i=1}^{N}\frac{D_i}{\pi(Z_i,\beta)}\left\{\frac{1}{\pi(Z_i,\beta)}-1\right\}-\mathbb{E}\frac{\pi(Z)}{\pi(Z,\beta)}\left\{\frac{1}{\pi(Z_i,\beta)}-1\right\}\right| = o_p(1).$$

This together with the consistency of $\hat{\beta}$, implies that

$$\frac{1}{N}\sum_{i=1}^{N}\frac{D_i}{\pi(Z_i,\hat{\beta})}\left\{\frac{1}{\pi(Z_i,\hat{\beta})}-1\right\} = \mathbb{E}\frac{\pi(Z)}{\pi(Z,\beta_0)}\left\{\frac{1}{\pi(Z_i,\beta_0)}-1\right\} + o_p(1) = B_{11}-1+o_p(1),$$

where we have used $\pi(Z)=\pi(Z,\beta_0)$ and $B_{11}=\mathbb{E}\{1/\pi(Z)\}$. Therefore

$$|J_2| \;\geq\; \frac{1}{1+|\hat{\gamma}|\max_{1\leq i\leq N} D_i/\pi(Z_i,\hat{\beta})}\cdot O_p(1). \tag{9}$$

By (8), we have

$$\begin{aligned} O_p(N^{-1/2}) &= |J_1| = |J_2|\cdot|\hat{\gamma}| \\ &= \geq \frac{1}{1+|\hat{\gamma}|\max_{1\leq i\leq N} D_i/\pi(Z_i,\hat{\beta})}\cdot O_p(1)\cdot|\hat{\gamma}|. \end{aligned}$$

Because $\mathbb{E}\{\pi(Z)/(\bar{\pi}(Z))^2\}<\infty$, we have

$$\mathbb{E}\left\{\frac{D}{\bar{\pi}(Z)}\right\}^2 = \mathbb{E}\frac{\pi(Z)}{\{\bar{\pi}(Z)\}^2}<\infty.$$

Similar to Lemma 1, this implies that $\max_{1\leq i\leq N} D_i/\bar{\pi}(Z_i)=o(N^{1/2})$, and

$$\max_{1\leq i\leq N}\frac{D_i}{\pi(Z_i,\hat{\beta})}\leq\max_{1\leq i\leq N}\frac{D_i}{\bar{\pi}(Z_i)} = o(N^{1/2}). \tag{10}$$

It follows that $|\hat{\gamma}|\max_{1\leq i\leq N} D_i/\pi(Z_i,\hat{\beta})=o_p(1)$, therefore we have from inequality (9) that

$$O_p(n^{-1/2}) \;\geq\; \frac{1}{1+|\hat{\gamma}|\cdot o(N^{1/2})}\cdot O_p(1)\cdot|\hat{\gamma}|,$$

which implies $\hat{\gamma}=O_p(n^{-1/2})$. Because $n/N-\alpha_0=O_p(N^{-1/2})$, we have

$$\hat{\alpha}-\alpha_0 = \left(\frac{n}{N}-\alpha_0\right)-\hat{\gamma}\left(1-\frac{n}{N}\right) = O_p(N^{-1/2}). \tag{11}$$

4.2 Approximate of $\hat{\alpha}$

Restudying the discussion in the last subsection, we have

$$J_2 = -\alpha_0(B_{11} - 1) + o_p(1).$$

Then equation (8) implies

$$\begin{aligned}
\hat{\gamma} &= -\frac{1}{\alpha_0(B_{11}-1)}\left[\frac{1}{N}\sum_{i=1}^{N} D_i\left\{1 - \frac{1}{\pi(Z_i,\beta_0)}\frac{n}{N}\right\} + \alpha_0 B_{1\dot{\pi}}(\hat{\beta}-\beta_0)\right] + o_p(N^{-1/2}) \\
&= -\frac{1}{B_{11}-1}\left[\frac{1}{N}\sum_{i=1}^{N}\left\{1 - \frac{D_i}{\pi(Z_i,\beta_0)}\right\} + B_{1\dot{\pi}}(\hat{\beta}-\beta_0)\right] + o_p(N^{-1/2}).
\end{aligned}$$

By (11), we have

$$\begin{aligned}
\hat{\alpha} - \alpha_0 &= \frac{n}{N} - \alpha_0 - (1-\alpha_0)\,\hat{\gamma} + o_p(N^{-1/2}) \\
&= \frac{n}{N} - \alpha_0 + \frac{1-\alpha_0}{B_{11}-1}\frac{1}{N}\sum_{i=1}^{N}\left\{1 - \frac{D_i}{\pi(Z_i,\beta_0)}\right\} + \frac{1-\alpha_0}{B_{11}-1}B_{1\dot{\pi}}\delta_4 + o_p(N^{-1/2}) \\
&= \frac{n}{N} - \alpha_0 + \frac{1-\alpha_0}{B_{11}-1}\frac{1}{N}\sum_{i=1}^{N}\left\{1 - \frac{D_i}{\pi(Z_i,\beta_0)}\right\} + \frac{1-\alpha_0}{B_{11}-1}\frac{1}{N}\sum_{i=1}^{N} B_{1\dot{\pi}}h(D_i,Z_i) \\
&\quad + o_p(N^{-1/2}), \qquad (12)
\end{aligned}$$

where we have used the assumption $\hat{\beta} - \beta_0 = N^{-1}\sum_{i=1}^{N} h(D_i,Z_i) + o_p(N^{-1/2})$.

4.3 Approximate of $\lambda(\hat{\alpha}) = (N-n)/\{n(1-\hat{\alpha})\}$

Because $\delta_1 = \hat{\alpha} - \alpha_0 = O_p(N^{-1/2})$ and $\delta_2 = n/N - \alpha_0 = O_p(N^{-1/2})$, we have

$$\lambda(\hat{\alpha}) - \alpha_0^{-1} = \frac{(1-\alpha_0-\delta_2)}{(\alpha_0+\delta_2)(1-\alpha_0-\delta_1)} - \alpha_0^{-1} = \frac{\delta_1}{\alpha_0(1-\alpha_0)} - \frac{\delta_2}{\alpha_0^2(1-\alpha_0)} + o_p(N^{-1/2}).$$

Let $\delta_3 = \lambda(\hat{\alpha}) - \alpha_0^{-1}$ and $\delta_4 = \hat{\beta} - \beta_0$. Putting the approximations of δ_1 and δ_2 into the above equation gives

$$\begin{aligned}
\delta_3 &= \frac{1}{\alpha_0(1-\alpha_0)} \times \frac{1}{N}\sum_{i=1}^{N}(D_i - \alpha_0) + \frac{1}{\alpha_0(1-\alpha_0)} \times \frac{1-\alpha_0}{B_{11}-1}\frac{1}{N}\sum_{i=1}^{N}\left\{1 - \frac{D_i}{\pi(Z_i)}\right\} \\
&\quad + \frac{1}{\alpha_0(1-\alpha_0)}\frac{1-\alpha_0}{B_{11}-1}B_{1\dot{\pi}}\delta_4 - \frac{1}{\alpha_0^2(1-\alpha_0)}\frac{1}{N}\sum_{i=1}^{N}(D_i-\alpha_0) + o_p(N^{-1/2}) \qquad (13) \\
&= -\frac{1}{\alpha_0^2} \times \frac{1}{N}\sum_{i=1}^{N}(D_i - \alpha_0) + \frac{1}{\alpha_0(B_{11}-1)}\frac{1}{N}\sum_{i=1}^{N}\left\{1 - \frac{D_i}{\pi(Z_i)}\right\} \\
&\quad + \frac{B_{1\dot{\pi}}}{\alpha_0(B_{11}-1)}\sum_{i=1}^{N} h(D_i,Z_i) + o_p(N^{-1/2}) \\
&= \frac{1}{\alpha_0^2(B_{11}-1)}\frac{1}{N}\sum_{i=1}^{N}\left[\frac{D_i\{\pi(Z_i)-\alpha_0\}}{\pi(Z_i)} - B_{11}(D_i-\alpha_0) + \alpha_0 B_{1\dot{\pi}}h(D_i,Z_i)\right] + o_p(N^{-1/2}) \qquad (14)
\end{aligned}$$

4.4 Consistency of $\hat{\theta}_{\rm ELW}$

Recall that $\hat{\theta}_{\rm ELW}$ satisfies

$$0 \;=\; \sum_{i=1}^{N} \hat{p}_i g(Z_i, \hat{\theta}_{\rm ELW}) = \sum_{i=1}^{N} \frac{1}{n} \frac{D_i}{1+\lambda(\hat{\alpha})\{\pi(Z_i,\hat{\beta}) - \hat{\alpha}\}} g(Z_i, \hat{\theta}_{\rm ELW}). \tag{15}$$

Define

$$\Delta_n(\theta) = \sum_{i=1}^{N} \hat{p}_i g(Z_i, \theta) - \mathbb{E}\{g(Z,\theta)\} = \Delta_{n1}(\theta) + \Delta_{n2}(\theta),$$

where

$$\begin{aligned}
\Delta_{n1}(\theta) &= \sum_{i=1}^{N} \hat{p}_i g(Z_i,\theta) - \frac{1}{N}\sum_{i=1}^{N} \frac{D_i}{\pi(Z_i,\hat{\beta})} g(Z_i,\theta), \\
\Delta_{n2}(\theta) &= \frac{1}{N}\sum_{i=1}^{N} \frac{D_i}{\pi(Z_i,\hat{\beta})} g(Z_i,\theta) - \mathbb{E}\{g(Z,\theta)\}.
\end{aligned}$$

The uniformly convergence of $\Delta_{n2}(\theta)$ to zero follows from Condition 2, the compactness of Θ and the consistency of β. It remains to prove $\sup_{\theta\in\Theta} \|\Delta_{n1}(\theta)\| = o_p(1)$.

Note that

$$\begin{aligned}
\Delta_{n1}(\theta) &= \frac{1}{N}\sum_{i=1}^{N} \left\{ \frac{1}{n/N} \frac{D_i}{1+\lambda(\hat{\alpha})\{\pi(Z_i,\hat{\beta}) - \hat{\alpha}\}} - \frac{D_i}{\pi(Z_i,\hat{\beta})} \right\} g(Z_i,\theta) \\
&= \frac{1}{N}\sum_{i=1}^{N} \left\{ \frac{1}{\alpha_0+\delta_2} \frac{D_i}{1+(\alpha_0^{-1}+\delta_3)\{\pi(Z_i,\hat{\beta}) - (\alpha_0+\delta_1)\}} - \frac{D_i}{\pi(Z_i,\hat{\beta})} \right\} g(Z_i,\theta) \\
&= \frac{1}{N}\sum_{i=1}^{N} \xi_{Ni} \frac{D_i}{\pi(Z_i,\hat{\beta})} g(Z_i,\theta)
\end{aligned}$$

where

$$\xi_{Ni} \;=\; -D_i \cdot \frac{\delta_2\alpha_0^{-1} + \zeta_{Ni} + \delta_2\alpha_0^{-1}\zeta_{Ni}}{(1+\delta_2\alpha_0^{-1})(1+\zeta_{Ni})}$$

with

$$\zeta_{Ni} = \alpha_0\delta_3 - (\delta_1 + \alpha_0^2\delta_3 + \alpha_0\delta_2\delta_3)\frac{D_i}{\pi(Z_i,\hat{\beta})}. \tag{16}$$

We have shown in (10) that $\max_{1\le i\le N} D_i/\pi(Z_i,\hat{\beta}) = o(N^{1/2})$. In the meantime, δ_1, δ_2, and δ_2 are all of order $O_p(N^{-1/2})$, therefore

$$\max_{1\le i\le N} |\zeta_{Ni}| \le -D_i \cdot \frac{\delta_2\alpha_0^{-1} + \zeta_{Ni} + \delta_2\alpha_0^{-1}\zeta_{Ni}}{(1+\delta_2\alpha_0^{-1})(1+\zeta_{Ni})} = o_p(1).$$

Therefore

$$\max_{1\le i\le N} |\xi_{Ni}| \;\le\; \frac{O_p(N^{-1/2}) + O_p(N^{-1/2}) + O_p(N^{-1/2})\cdot o(N^{1/2})}{\{1-O_p(N^{-1/2})\}\left\{1 - O_p(N^{-1/2}) - O_p(N^{-1/2})\cdot o(N^{1/2})\right\}} = o_p(1).$$

The remaining proof is the same as that in section 3.4 and is omitted.

4.5 Asymptotic normality of $\hat{\theta}_{\mathrm{ELW}}$

By the same arguments to those in Section 3.5,

$$\hat{\theta}_{\mathrm{ELW}} - \theta_0 \quad = \quad -K^{-1} \sum_{i=1}^{N} \hat{p}_i g(Z_i, \theta_0) + o_p(N^{-1/2}).$$

In the proof of the consistency of $\hat{\theta}_{\mathrm{ELW}}$, we have shown that

$$\hat{p}_i = \frac{1}{N} \frac{D_i}{\pi(Z_i, \hat{\beta})} + \frac{1}{N} \frac{D_i}{\pi(Z_i, \hat{\beta})} \xi_{Ni},$$

where

$$\xi_{Ni} \quad = \quad -D_i \cdot \frac{\delta_2 \alpha_0^{-1} + \zeta_{Ni} + \delta_2 \alpha_0^{-1} \zeta_{Ni}}{(1 + \delta_2 \alpha_0^{-1})(1 + \zeta_{Ni})}$$

with ζ_{Ni} defined in (16).

Because $\max_{1 \le i \le N} |\zeta_{Ni}| = o_p(1)$, we have

$$\sum_{i=1}^{N} \hat{p}_i g(Z_i, \theta_0) \quad = \quad \sum_{i=1}^{N} \frac{1}{N} \frac{D_i}{\pi(Z_i, \hat{\beta})} g(Z_i, \theta_0) - \sum_{i=1}^{N} \frac{1}{N} \frac{D_i}{\pi(Z_i, \hat{\beta})} \cdot (\delta_2 \alpha_0^{-1} + \zeta_{Ni}) g(Z_i, \theta_0) + o_p(N^{-1/2}).$$

Since $\zeta_{Ni} = \alpha_0 \delta_3 - (\delta_1 + \alpha_0^2 \delta_3 + \alpha_0 \delta_2 \delta_3) D_i / \pi(Z_i, \hat{\beta})$ it follows that

$$\begin{aligned}
\sum_{i=1}^{N} \hat{p}_i g(Z_i, \theta_0) \quad &= \quad \sum_{i=1}^{N} \frac{1}{N} \frac{D_i g(Z_i, \theta_0)}{\pi(Z_i, \beta_0)} + \sum_{i=1}^{N} \frac{1}{N} \left\{ \frac{D_i}{\pi(Z_i, \hat{\beta})} - \frac{D_i}{\pi(Z_i, \beta_0)} \right\} g(Z_i, \theta_0) \\
&\quad - \sum_{i=1}^{N} \frac{1}{N} \frac{D_i g(Z_i, \theta_0)}{\pi(Z_i, \beta_0)} \cdot \left\{ \delta_2 \alpha_0^{-1} + \alpha_0 \delta_3 - (\delta_1 + \alpha_0^2 \delta_3) \frac{D_i}{\pi(Z_i, \beta_0)} \right\} + o_p(N^{-1/2}) \\
&= \quad \sum_{i=1}^{N} \frac{1}{N} \frac{D_i g(Z_i, \theta_0)}{\pi(Z_i, \beta_0)} - \sum_{i=1}^{N} \frac{1}{N} \frac{D_i g(Z_i, \theta_0) \pi_1(Z_i, \tilde{\beta})}{\{\pi(Z_i, \tilde{\beta})\}^2} (\hat{\beta} - \beta_0) \\
&\quad + B_{g1}(\delta_1 + \alpha_0^2 \delta_3) + o_p(N^{-1/2}) \\
&= \quad \sum_{i=1}^{N} \frac{1}{N} \frac{D_i g(Z_i, \theta_0)}{\pi(Z_i, \beta_0)} - B_{g\dot{\pi}}(\hat{\beta} - \beta_0) + B_{g1}\delta_1 + \alpha_0^2 B_{g1} \delta_3 + o_p(N^{-1/2}),
\end{aligned}$$

where the last equality holds under Condition 2 (ii).

Then using the approximate of $\delta - 1 = \hat{\alpha} - \alpha_0$ in (12), the approximate of δ_3 in (14), and

$$\hat{\beta} - \beta_0 = \frac{1}{N} \sum_{i=1}^{N} h(D_i, Z_i) + o_p(1),$$

we have

$$\begin{aligned}
\sum_{i=1}^{N}\hat{p}_i g(Z_i,\theta_0) &= \sum_{i=1}^{N}\frac{1}{N}\frac{D_i g(Z_i,\theta_0)}{\pi(Z_i,\beta_0)} - \frac{1}{N}\sum_{i=1}^{N}B_{g\dot\pi}h(D_i,Z_i) \\
&\quad + \frac{B_{g1}}{N}\sum_{i=1}^{N}(D_i-\alpha_0) + \frac{(1-\alpha_0)B_{g1}}{B_{11}-1}\frac{1}{N}\sum_{i=1}^{N}\left\{1-\frac{D_i}{\pi(Z_i,\beta_0)}\right\} \\
&\quad + \frac{(1-\alpha_0)B_{g1}B_{1\dot\pi}}{B_{11}-1}\frac{1}{N}\sum_{i=1}^{N}h(D_i,Z_i) + \frac{B_{g1}}{B_{11}-1}\frac{1}{N}\sum_{i=1}^{N}\frac{D_i\{\pi(Z_i)-\alpha_0\}}{\pi(Z_i)} \\
&\quad - \frac{B_{g1}B_{11}}{B_{11}-1}\frac{1}{N}\sum_{i=1}^{N}(D_i-\alpha_0) + \frac{\alpha_0 B_{g1}B_{1\dot\pi}}{B_{11}-1}\frac{1}{N}\sum_{i=1}^{N}h(D_i,Z_i) + o_p(N^{-1/2}) \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{D_i g(Z_i,\theta_0)}{\pi(Z_i,\beta_0)} + \frac{1}{N}\sum_{i=1}^{N}\left(\frac{B_{g1}B_{1\dot\pi}}{B_{11}-1} - B_{g\dot\pi}\right)h(D_i,Z_i) \\
&\quad + \frac{1}{N}\sum_{i=1}^{N}\frac{B_{g1}}{B_{11}-1}\left(1-\frac{D_i}{\pi(Z_i,\beta_0)}\right) + o_p(N^{-1/2}).
\end{aligned}$$

Consequently

$$\sqrt{N}\sum_{i=1}^{N}\hat{p}_i g(Z_i,\theta_0) \quad \xrightarrow{d} \quad N(0,\Omega),$$

where

$$\Omega = \mathbb{V}\mathrm{ar}\left\{\frac{D_i g(Z_i,\theta_0)}{\pi(Z_i,\beta_0)} + \left(\frac{B_{g1}B_{1\dot\pi}}{B_{11}-1} - B_{g\dot\pi}\right)h(D_i,Z_i) + \frac{B_{g1}}{B_{11}-1}\left(1-\frac{D_i}{\pi(Z_i,\beta_0)}\right)\right\}.$$

Finally, we have

$$\hat\theta_{\mathrm{ELW}} - \theta_0 \quad = \quad -K^{-1}\sqrt{N}\sum_{i=1}^{N}\hat{p}_i g(Z_i,\theta_0) \xrightarrow{d} N(0,\Sigma_{\mathrm{ELW}}),$$

where $\Sigma_{\mathrm{ELW}} = K^{-1}\Omega(K^{-1})^{\top}$.

5 Extension to unequal probability samplings

We have shown that the ELW method works for missing data problems, where data are independent and identically distributed. However, Theorem 2.1 does not hold any longer under general unequal probability samplings without replacement (UPS-WOR), where the sample size is fixed and the data are correlated, or under unequal probability sampling with replacement (UPS-WR), where a piece of data may be sampled multiple times. In this section, we extend the ELW method to UPS-WOR and UPS-WR.

To study the asymptotic properties of the ELW and IPW estimators under unequal probability samplings (UPS), we suppose that there is a sequence of large datasets, $\{Z_{\nu 1},\ldots,Z_{\nu N_\nu}\}$, of size N_ν,

$\nu = 1, 2, \ldots$, all of which are regarded as finite sets of non-random numbers. The finite population under study is one of them. For a particular large dataset $\{Z_{\nu 1}, \ldots, Z_{\nu N_\nu}\}$, a sample of size n_ν is taken from it with a prespecified UPS. For notational simplicity, we shall suppress the subscript ν and use $N \to \infty$ instead of $\nu \to \infty$. We denote the true value of α by $\alpha_0 = n/N$, and assume that the parameter of interest θ_0 is defined as the solution to $U_0(\theta) = 0$, where $U_0(\theta) = N^{-1} \sum_{i=1}^N g(Z_i, \theta)$ and g does not vary with N. As in the main paper, we assume that the dimension of g is equal to that of θ, and both are equal to r in this section. Hereafter a quantity with a star in the subscript stands for the limit of the counterpart without a star as $N \to \infty$, e.g. $\alpha_{0*} = \lim_{N\to\infty} \alpha_0$ and $\theta_{0*} = \lim_{N\to\infty} \theta_0$. We assume that the problem under study satisfies the following regularity conditions.

Condition 3. *(i) The parameter space is* $\Theta \subset \mathbb{R}^r$ *and* $\alpha_{0*} \in (0, 1)$. *(ii) The function* $g(Z, \theta)$ *has a continuous partial derivative* $g_1(Z, \theta) = \partial g(Z, \theta)/\partial \theta^\top$ *in a neighborhood* M_0 *of* θ_{0*} *for each* Z. *(iii)* $\theta_0 \in \Theta$ *and* $\theta_{0*} \in \Theta$ *are the unique solutions to* $U_0(\theta) = 0$ *and* $U_{0*} = 0$*, respectively. (iv)* $\sup_{\theta\in\Theta} |U_0(\theta) - U_{0*}(\theta)| = o(1)$. *(v)* $K_* = \partial U_{0*}(\theta_{0*})/\partial \theta^\top$ *is nonsingular.*

Condition 3 (i), (ii) and (v) are trivial. Condition 3 (iii) guarantees the identifiability of the parameter of interest. Condition 3 (iv) implies that $U_0(\theta)$ converges to $U_{0*}(\theta)$ uniformly for all $\theta \in \Theta$. This together with the uniqueness of θ_{0*} implies $\theta_0 \to \theta_{0*}$, i.e. the parameter of interest is stable.

5.1 ELW for UPS-WOR

Denote the inclusion probability of Z_i under a UPS-WOR by π_i. The population size N and all the π_i are known a priori. Define $D_k = 1$ if Z_k is selected, and 0 otherwise for $1 \le k \le N$. Then, the sample size $n = \sum_{k=1}^N D_k$ and the SIPW estimator $\hat\theta_{\mathrm{SIPW}}$ is the solution to $N^{-1} \sum_{i=1}^N g(Z_i, \theta) D_i/\pi_i = 0$. When $g(Z, \theta) = f(Z) - \theta$, we have the original IPW estimator or the famous Horvitz–Thompson estimator (Horvitz and Thompson, 1952) $\hat\theta_{\mathrm{IPW}} = N^{-1} \sum_{i=1}^N f(Z_i) D_i/\pi_i$. With π_i in place of $\pi(Z_i)$, we propose to directly apply the ELW method for missing data problems in the main paper and denote the ELW estimator of θ_0 by $\hat\theta_{\mathrm{ELW}}$. Although the ELW estimators for UPS-WOR and for missing data problems have the same form, their random behaviours are totally different because the joint distributions of the observed data are different. It is worth noting that under UPS-WOR, the Z_k are non-random and only the D_k are random variables. Unlike the cases in missing data problems, the ELW weights $\hat p_k$ do not lead to an estimate of a certain distribution function; they are simply taken as weights for parameter estimation.

The asymptotic normality of the IPW estimator, although difficult, has been established for many commonly used samplings, including simple random sampling with or without replacement (Erdös and Rényi, 1959; Hájek, 1960), rejective sampling with unequal probabilities (Hájek, 1964), stratified unequal probability sampling with or without replacement (Krewski and Rao, 1981; Bickel and Freedman, 1984), and two-phase sampling (Chen and Rao, 2007). See Wu and Thompson

(2020) for a comprehensive review. We establish the asymptotic normality of the ELW estimator under the so-called linearly negatively dependence assumption.

Definition 1 (Patterson et al. (2001)). *A sequence of random variables, $\{X_i\}$, is said to be linearly negatively dependent (LIND) if for any disjoint subsets of indices A, B and positive constants λ_j, $P(\sum_{k\in A}\lambda_k X_k \le s, \sum_{j\in B}\lambda_j X_j \le t) \le P(\sum_{k\in A}\lambda_k X_k \le s)\cdot P(\sum_{j\in B}\lambda_j X_j \le t)$ for any real numbers s and t.*

Many commonly used πps samplings, including conditional Poisson sampling, Sampford sampling, Pareto sampling, and pivotal sampling satisfy the LIND property (Patterson et al., 2001; Brändén and Jonasson, 2012).

Condition 4. *(i) Let $\{Z_k, 1\le k\le N\}$ be constant vectors and $\{D_k, 1\le k\le N\}$ be LIND random variables with $\pi_k = \mathbb{E}(D_k)$. (ii) There exists a function $\bar g(z)\ge 1$ satisfying*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\frac{1-\pi_i}{\pi_i}\{\bar g(Z_i)\}^2<\infty,\quad \lim_{N\to\infty}\frac{1}{N}\sum_{1\le i\ne j\le N}\frac{\mathbb{C}\mathrm{ov}(D_i,D_j)}{\pi_i\pi_j}\bar g(Z_i)\bar g(Z_j)=0,$$

such that $\sup_{\theta\in\Theta}\|g(Z,\theta)\|\le \bar g(Z)$ and $\sup_{\theta\in M_0}\|g_1(Z,\theta)\|_F\le \bar g(Z)$, where M_0 is the neighborhood in Condition 3.

Under Condition 4 (ii), $\bar g(Z)$ is a control function for 1, $g(Z,\theta)$ and $g_1(Z,\theta)$ and the quantity $N^{-1}\sum_{i=1}^N \bar g(Z_i)(D_i-\pi_i)/\pi_i$ satisfies the weak law of large numbers. Therefore when $\bar g(Z_i)$ is replaced by 1, $g(Z,\theta)$ and $g_1(Z,\theta)$, the corresponding quantities also satisfy the weak law of large numbers. Define $B_{gg} = N^{-1}\sum_{i=1}^N\{g(Z_i,\theta_0)\}^{\otimes 2}/\pi_i$, $B_{g1} = N^{-1}\sum_{i=1}^N g(Z_i,\theta_0)/\pi_i$, and $C_{gg} = N^{-1}\sum_{i=1}^N\{g(Z_i,\theta_0)\}^{\otimes 2}$. In the case of $g(Z,\theta)=f(Z)-\theta$, we write these quantities as B_{ff}, B_{f1} and C_{ff} after $g(Z,\theta_0)$ is replaced by $f(Z)$.

Theorem 5.1 below establishes the asymptotic normalities of $\hat\theta_{\mathrm{ELW}}$, $\hat\theta_{\mathrm{SIPW}}$, and $\hat\theta_{\mathrm{IPW}}$ under UPS-WOR. A proof of Theorem 5.1 is given in Section 6.

Theorem 5.1. *Assume Conditions 3 and 4. Let $W_k = (g^\top(Z_k,\theta_0),1)^\top/\pi_k$, and $V_0 = N^{-1}\sum_{k=1}^N W_kW_k^\top(1-\pi_k)\pi_k$. Suppose that $\lim_{N\to\infty}N^{-1}\sum_{k=1}^N\|W_k\|^2P(\|W_k\||D_k-\pi_k|\ge\varepsilon\sqrt{N})=0$ for any $\varepsilon>0$ and $\lim_{N\to\infty}V_0=V_{0*}$ is positive definite. As $N\to\infty$, we have*

(a) $\sqrt{N}(\hat\theta_{\mathrm{SIPW}}-\theta_0)\xrightarrow{d}N(0,\Sigma_{\mathrm{SIPW}})$, where $\Sigma_{\mathrm{SIPW}} = K_*^{-1}(B_{gg*}-C_{gg*})(K_*^{-1})^\top$;

(b) $\sqrt{N_\nu}(\hat\theta_{\mathrm{ELW}}-\theta_0)\xrightarrow{d}N(0,\Sigma_{\mathrm{ELW}})$, where $\Sigma_{\mathrm{ELW}} = K_*^{-1}\left(B_{gg*}-C_{gg*}-\frac{B_{g1*}B_{g1*}^\top}{B_{11*}-1}\right)K_*^{-1}$; the ELW estimator $\hat\theta_{\mathrm{ELW}}$ is more efficient than $\hat\theta_{\mathrm{SIPW}}$, i.e. $\Sigma_{\mathrm{ELW}}\le\Sigma_{\mathrm{SIPW}}$, where the equalities hold only if the π_k are all equal.

(c) in the case of $g(z,\theta)=f(z)-\theta$, $N^{1/2}(\hat\theta_{\mathrm{IPW}}-\theta_0)\xrightarrow{d}N(0,\Sigma_{\mathrm{IPW}})$, where $\Sigma_{\mathrm{IPW}} = B_{ff*}-C_{ff*}$; the ELW estimator $\hat\theta_{\mathrm{ELW}}$ is also more efficient than $\hat\theta_{\mathrm{IPW}}$.

By Theorem 5.1, again the ELW, IPW, and SIPW estimators are asymptotically normal, and the ELW estimator is asymptotically the most efficient. A consistent variance estimator for Σ_{ELW} can be constructed in a similar way to equation (14) in the main paper.

5.2 ELW for UPS-WR

Suppose a sample of size n is drawn by a UPS-WR from $\{Z_1, \ldots, Z_N\}$. Let the sampling probability of the ith individual at each sampling occasion be $\tilde{q}_i$ ($i = 1, 2, \ldots, N$). Clearly, $\tilde{q}_i > 0$ and $\sum_{i=1}^N \tilde{q}_i = 1$. Denote the observed data by z_i and the corresponding sampling probability by q_i for $i = 1, 2, \ldots, n$.

To apply the IPW and ELW methods, we define $\tilde{\pi}_i = n\tilde{q}_i$ ($1 \le i \le N$) and $\pi_i = nq_i$ ($1 \le i \le n$), which may be greater than 1. The SIPW estimator $\hat{\theta}_{\mathrm{SIPW}}$ of θ_0 under UPW-WR is the solution to $\hat{U}_{\mathrm{IPW}}(\theta) = 0$, where

$$\hat{U}_{\mathrm{IPW}}(\theta) = \frac{1}{nN}\sum_{i=1}^n \frac{g(z_i, \theta)}{q_i} = \frac{1}{N}\sum_{i=1}^n \frac{g(z_i, \theta)}{\pi_i}.$$

In the case of $g(z, \theta) = f(z) - \theta$, the original IPW estimator of $\theta_0 = N^{-1}\sum_{i=1}^N f(Z_i)$ is the Hansen–Hurwitz estimator (Hansen and Hurwitz, 1943), $\hat{\theta}_{\mathrm{IPW}} = N^{-1}\sum_{i=1}^n f(z_i)/\pi_i$.

The ELW method from missing data problems was introduced with data $\{(D_i, D_iZ_i, \pi(Z_i)) : 1 \le i \le N\}$. However, the notation D_i is not appropriate under UPS-WR, because a unit from the population may be drawn multiple times. As the ELW estimator depends only on N and $\{(z_i, \pi_i) : 1 \le i \le n\}$, it is applicable to UPS-WR through the following Algorithm.

Algorithm 1: ELW estimation procedure under UPS-WR

Input: The sample drawn, $\{(z_i, \pi_i) : i = 1, 2, \ldots, n\}$; The size of the finite population, N.

Output: The ELW estimate, $\hat{\theta}_{\mathrm{ELW}}$.

Step 1. Calculate $\xi_i = n/N + (1 - n/N)\pi_i$ ($1 \le i \le n$), $\zeta_l = \min_{1\le i\le n} \pi_i$, and $\zeta_u = \min_{1\le i\le n} \xi_i$.

Step 2. Calculate $\hat{\alpha}$ by solving $0 = \sum_{i=1}^n (\pi_i - \alpha)/(\xi_i - \alpha)$ in the interval $[\zeta_l, \zeta_u)$, and calculate $\lambda(\hat{\alpha}) = (N - n)/\{n(1 - \hat{\alpha})\}$.

Step 3. Calculate $\hat{p}_i = n^{-1}\{1 + \lambda(\hat{\alpha})(\pi_i - \hat{\alpha})\}^{-1}, i = 1, 2, \ldots, n$.

Step 4. Calculate the ELW estimate $\hat{\theta}_{\mathrm{ELW}}$ by solving $\sum_{i=1}^n \hat{p}_i g(z_i, \theta) = 0$.

Condition 5. *(i)* $\lim_{N\to\infty} \max_{1\le i\le N} \tilde{\pi}_i < 1$. *(ii) There exists a function* $\bar{g}(z) \ge 1$ *satisfying* $\lim_{N\to\infty} N^{-1}\sum_{i=1}^N \{\bar{g}(Z_i)\}^2/\tilde{\pi}_i < \infty$ *such that* $\sup_{\theta\in\Theta} \|g(Z, \theta)\| \le \bar{g}(Z)$ *and* $\sup_{\theta\in M_0} \|g_1(Z, \theta)\|_F \le \bar{g}(Z)$*, where* M_0 *is the neighborhood in Condition 3.*

Condition 5 (1) excludes the case with $\tilde{\pi}_i \ge 1$ at least when N is large. In UPS-WR, Condition 5 (ii) plays the role of Condition 4 (ii) in UPW-WOR. Define $B_{gg}, B_{g1}, B_{11}, B_{ff}$ and B_{f1} and their limits as $N \to \infty$ in the same way as in Section 3.1 except that $\tilde{\pi}_k$ takes place of π_k. Theorem 5.2 establishes the asymptotic normalities of $\hat{\theta}_{\mathrm{ELW}}$, $\hat{\theta}_{\mathrm{SIPW}}$, and $\hat{\theta}_{\mathrm{IPW}}$ under UPS-WR. A proof of Theorem 5.2 is given in Section 7.

Theorem 5.2. *Assume Conditions 3 and 5. As* $N \to \infty$,

(a) $\sqrt{N}(\hat{\theta}_{\mathrm{SIPW}} - \theta_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{SIPW}})$, *with* $\Sigma_{\mathrm{SIPW}} = K_*^{-1}B_{gg*}(K_*^{-1})^\top$;

(b) $\sqrt{B}(\hat\theta_{\rm ELW}-\theta_0)\xrightarrow{d} N(0,\Sigma_{\rm ELW})$, *where*

$$\Sigma_{\rm ELW} \;=\; K_*^{-1}\left\{B_{gg*}-\frac{B_{g1*}^{\otimes 2}}{(B_{11*}-1)^2}(B_{11*}+\alpha_{0*}^{-1}-2)\right\}(K_*^{-1})^\top;$$

the ELW estimator is more efficient than the SIPW estimator, i.e. $\Sigma_{\rm ELW}\le\Sigma_{\rm SIPW}$*, and the equality holds if and only if the* $\tilde\pi_k$ *are all equal;*

(c) in the case of $g(z,\theta)=f(z)-\theta$, $\sqrt{N}(\hat\theta_{\rm IPW}-\theta_0)\xrightarrow{d} N(0,\Sigma_{\rm IPW})$ *with* $\Sigma_{\rm IPW}=B_{ff*}-\theta_{0*}^{\otimes 2}\alpha_{0*}^{-1}$.

In Theorem 5.2, we establish the asymptotic normality of the SIPW, IPW and ELW estimators, and show that the ELW estimator is more efficient than the SIPW estimator. Unfortunately, we do not have an affirmative conclusion about the efficiency comparison between the ELW estimator and the IPW estimator under UPW-WR. As $\hat p_i\approx 1/(N\pi_i)$ $(1\le i\le n)$, we propose to estimate $\Sigma_{\rm ELW}$ by

$$\widehat\Sigma_{\rm ELW}=\hat K^{-1}\left\{\hat B_{gg*}-\frac{\hat B_{g1}^{\otimes 2}}{(\hat B_{11}-1)^2}(\hat B_{11}+\alpha_0^{-1}-2)\right\}(\hat K^{-1})^\top,$$

where $\hat B_{11}=N\sum_{k=1}^n(\hat p_k)^2$, $\hat K=N\sum_{k=1}^n(\hat p_k)^2 g_1(z_k,\hat\theta_{\rm IPW})$, $\hat B_{g1}=N\sum_{k=1}^n(\hat p_k)^2 g(z_k,\hat\theta_{\rm IPW})$, and $\hat B_{gg}=N\sum_{k=1}^n(\hat p_k)^2\{g(z_k,\hat\theta_{\rm IPW})\}^{\otimes 2}$.

The EL method originated from Hartley and Rao (1968)'s "scale-load" approach in survey sampling and was formulated by Owen (1988, 1990) as a general nonparametric tool for statistical inferences. There has been a growing body of literature proposing the use of EL for finite-population inference in survey sampling (Berger, 2018; Zhao and Wu, 2019). Most of the recent developments can be classified into two groups. One group is pseudo-EL approach (Chen and Sitter, 1999; Wu and Rao, 2010; Rao and Wu, 2010), which replaces the usual empirical log-likelihood function by its IPW version and retains the constraints. The pseudo empirical log-likelihood ratio statistic usually follows a weighted chisquare distribution (Wu and Rao, 2010), which limits its wide application. The other group of ELs retain the empirical log-likelihood function but replace the equality constraints by their IPW version. Popular examples include the EL under Poisson sampling (Kim, 2009), the sample EL (Chen and Kim, 2014; Zhao et al., 2022), and the unequal probability EL approach (Berger and Torres, 2016; Berger, 2018, 2020). An advantage of the second group of ELs over the pseudo EL is that their empirical log-likelihood ratio statistics often have central chisquare limiting distributions. Our biased-sample EL has a close relationship with the second group of ELs. As the foundation of our EL, the full likelihood $\tilde L$ in equation (4) in the main paper is proportional to $L_m\times L_c$, where $L_m=\binom{N}{n}\alpha^n(1-\alpha)^{N-n}$ is a marginal likelihood, and $L_c=\prod_{i=1}^n\{\pi(Z_i)p_i/\alpha\}$ is a conditional likelihood. Under suitable transformations, the second group of ELs which use sampling weights in the constraints are generally equivalent to the conditional log-likelihood $\log(L_c)$. With the additional L_m, our biased-sample EL automatically makes use of the information carried by N. In addition, the biased-sample EL ratio statistics also often have central chisquare limiting distributions, which is convenient for interval estimation and hypothesis testing. See Section 9.

In addition to unequal probabilities, stratification, clustering and missing data are also frequently encountered in survey sampling and raise many new interesting research problems. With slight modifications, Berger and Torres (2016)'s EL is able to accommodate general sampling designs with the complex structures (Berger, 2020). The key idea behind the modification is to transform the complex structures into suitable estimating equations, which can be flexibly integrated by the usual EL. We believe that the proposed EL is also applicable to these complex sampling designs after suitable estimating equations are constructed. This problem and whether the resulting likelihood ratio statistic is asymptotically pivotal are interesting topics for future research.

6 Proof of Theorem 5.1

We begin by two lemmas. The second lemma, i.e. Lemma 3, is crucial in our proof of Theorem 5.1. This lemma is proved based on the first lemma, i.e. Lemma 2, which is copied from Proposition 2.1 of Patterson et al. (2001), and is a modified version of Newman (1980)'s Theorem 1.

Lemma 2 (Newman, 1980)**.** *Suppose $D_1, \dots, D_N$ are LIND random variables and let* i *denote the imaginary unit $\sqrt{-1}$. Then*

$$\left| \mathbb{E}\left\{ \exp\left(\mathrm{i} \sum_{k=1}^{N} r_k D_k \right) \right\} - \prod_{k=1}^{N} \mathbb{E}\left(e^{\mathrm{i} r_k D_k} \right) \right| \le - \sum_{1 \le k \ne l \le N} |r_k r_l| \mathbb{C}\mathrm{ov}(D_k, D_l).$$

Please note that we suppress the subscript ν and use $N \to \infty$ instead of $\nu \to \infty$.

Lemma 3. *Let $\{Z_k, 1 \le k \le N\}$ be constant vectors and $\{D_k, 1 \le k \le N\}$ be LIND random variables. Let $\pi_k = \mathbb{E}(D_k)$ and $W_k = (g^\top(Z_k, \theta_0), 1)^\top / \pi_k$, and define $V_0 = N^{-1} \sum_{k=1}^{N} W_k W_k^\top \pi_k (1 - \pi_k)$. Suppose that*

$$\begin{aligned} &(1) \quad \lim_{N\to\infty} V_0 = V_0^* \quad \textit{is a positive definite matrix,} \\ &(2) \quad \lim_{N\to\infty} N^{-1} \sum_{1 \le k \ne l \le N} \|W_k\| \|W_l\| \mathbb{C}\mathrm{ov}(D_k, D_l) = 0, \quad \textit{and} \\ &(3) \quad \lim_{N\to\infty} N^{-1} \sum_{k=1}^{N} \|W_k\|^2 P(\|W_k\| |D_k - \pi_k| \ge \varepsilon \sqrt{N}) = 0 \textit{ for any } \varepsilon > 0. \end{aligned}$$

Then $(V_0^*)^{-1/2} N^{-1/2} \sum_{k=1}^{N} W_k (D_k - \pi_k) \xrightarrow{d} N(0, I_{r+1})$ as $N \to \infty$.

Proof of Lemma 3. We prove this Lemma along the same line of the proof of Patterson et al. (2001)'s Theorem 2.4. Let $\{R_k\}$ be a series of *independent* random variables, where R_k and D_k are identically distributed for each $k \ge 1$. By Linderberg-Feller central limit theorem, $V_0^{-1/2} N^{-1/2} \sum_{k=1}^{N} W_k (R_k - \pi_k) \xrightarrow{d} N(0, I_{r+1})$.

By Lemma 2, for any $(r+1)$-dimensional constant vector β,

$$\begin{aligned}
&\left|\mathbb{E}\left[\exp\left\{\mathrm{i}t\beta^\top V_0^{-\frac{1}{2}}N^{-\frac{1}{2}}\sum_{k=1}^N W_k(D_k-\pi_k)\right\}\right]-\prod_{k=1}^N \mathbb{E}[\exp\{\mathrm{i}\beta^\top V_0^{-\frac{1}{2}}N^{-\frac{1}{2}}W_k(D_k-\pi_k)\}]\right| \\
\leq\ & -\left|\exp\left\{-\mathrm{i}t\beta^\top V_0^{-\frac{1}{2}}N^{-\frac{1}{2}}\sum_{k=1}^N W_k\pi_k\right\}\right|\cdot\frac{1}{N}\sum_{1\leq k\neq l\leq N}|\beta^\top V_0^{-\frac{1}{2}}W_k|\cdot|\beta^\top V_0^{-\frac{1}{2}}W_l|\cdot\mathbb{C}\mathrm{ov}(D_k,D_l) \\
\leq\ & -\beta^\top V_0^{-1}\beta\cdot\frac{1}{N}\sum_{1\leq k\neq l\leq N}\|W_k\|\cdot\|W_l\|\cdot\mathbb{C}\mathrm{ov}(D_k,D_l),
\end{aligned}$$

which converges to $-\beta^\top V_{0*}^{-1}\beta\times 0=0$ as $N\to\infty$. This implies that the characteristic function of $\beta^\top V_0^{-\frac{1}{2}}N^{-\frac{1}{2}}\sum_{k=1}^N W_k(D_k-\pi_k)$, namely

$$\mathbb{E}\left[\exp\left\{\mathrm{i}t\beta^\top V_0^{-\frac{1}{2}}N^{-\frac{1}{2}}\sum_{k=1}^N W_k(D_k-\pi_k)\right\}\right],$$

has the same limit as $\prod_{k=1}^N\mathbb{E}[\exp\{\mathrm{i}\beta^\top V_0^{-\frac{1}{2}}N^{-\frac{1}{2}}W_k(D_k-\pi_k)\}]$, which is the characteristic function of $\beta^\top V_0^{-\frac{1}{2}}N^{-\frac{1}{2}}\sum_{k=1}^N W_k(R_k-\pi_k)$. We have shown that $V_0^{-\frac{1}{2}}N^{-\frac{1}{2}}\sum_{k=1}^N W_k(R_k-\pi_k)\xrightarrow{d}W$, an $(r+1)$-dimensional standard normal random vector. Therefore

$$\beta^\top V_0^{-\frac{1}{2}}N^{-\frac{1}{2}}\sum_{k=1}^N W_k(D_k-\pi_k)\xrightarrow{d}\beta^\top W.$$

Because of the arbitrariness of β, we have $V_0^{-\frac{1}{2}}N^{-\frac{1}{2}}\sum_{k=1}^N W_k(D_k-\pi_k)\xrightarrow{d}W$. $\square$

Under condition (2), we have

$$\lim_{N\to\infty}N^{-1}\sum_{1\leq k\neq l\leq N}\|W_k\|\|W_l\|\mathbb{C}\mathrm{ov}(D_k,D_l)=0.$$

This together with other conditions in Theorem 5.1 makes the conditions in Lemma 3 all satisfied, therefore the result in Lemma 3 holds under the conditions in Theorem 5.1.

6.1 Proof of result (a) of Theorem 5.1

Our proof of result (a) of Theorem 5.1 consists of two steps. In step I, we show that $\hat\theta_{\mathrm{SIPW}}$ is consistent. In step II, we derive a linear approximate of it, from which result (a) immediately follows.

6.1.1 Step I: consistency of $\hat\theta_{\mathrm{SIPW}}$

Condition 3 (iii) and (iv) imply that $\theta_0-\theta_{0*}=o(1)$. To prove the consistency of $\hat\theta_{\mathrm{SIPW}}$, i.e. $\hat\theta_{\mathrm{SIPW}}-\theta_0=o_p(1)$, it suffices to show $\hat\theta_{\mathrm{SIPW}}-\theta_{0*}=o(1)$. Recall that $\hat\theta_{\mathrm{SIPW}}$ is the solution to $\hat U_{\mathrm{IPW}}(\theta)=0$, where $\hat U_{\mathrm{IPW}}(\theta)=N^{-1}\sum_{i=1}^n g(Z_i,\theta)/\pi_i$. Condition 3 (iii) guarantees that $\inf_{\theta:\|\theta-\theta_{0*}\|>\varepsilon}\|U_{0*}(\theta)\|>$

0 for every $\varepsilon > 0$. By Theorem 5.9 of van de Vaart (2000), to prove $\hat{\theta}_{\text{SIPW}} - \theta_{0*} = o(1)$ it remains to show $\sup_{\theta\in\Theta} \left\|\hat{U}_{\text{IPW}}(\theta) - U_{0*}(\theta)\right\| = o_p(1)$.

It can be seen that

$$\sup_{\theta\in\Theta} \left\|\hat{U}_{\text{IPW}}(\theta) - U_{0*}(\theta)\right\| \leq \sup_{\theta\in\Theta} \left\|\hat{U}_{\text{IPW}}(\theta) - U_{0}(\theta)\right\| + \sup_{\theta\in\Theta} \left\|U_{0}(\theta) - U_{0*}(\theta)\right\|.$$

About the first term on the right-hand side of the above inequality, for any unit vector γ

$$\begin{aligned}\mathbb{V}\text{ar}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{D_i}{\pi_i}\gamma^\top g(Z_i,\theta)\right) &= \frac{1}{N^2}\sum_{i,j=1}^{N}\frac{\mathbb{C}\text{ov}(D_i,D_j)}{\pi_i\pi_j}\gamma^\top g(Z_i,\theta)g(Z_j,\theta)\gamma \\ &\leq \frac{1}{N^2}\sum_{i=1}^{N}\frac{1-\pi_i}{\pi_i}\{\bar{g}(Z_i)\}^2 + \left|\frac{1}{N^2}\sum_{1\leq i\neq j\leq N}\frac{\mathbb{C}\text{ov}(D_i,D_j)}{\pi_i\pi_j}\bar{g}(Z_i)\bar{g}(Z_j)\right|,\end{aligned}$$

where the last inequality holds because $\mathbb{C}\text{ov}(D_i, D_j) \leq 0$ for all $i \neq j$ under the LIND assumption. Under Condition 4 (ii), both the two terms on the right-hand side of the last inequality are $o(1)$. Since the last inequality holds uniformly for all θ and all unit vector γ and $\mathbb{E}\hat{U}_{\text{IPW}}(\theta) = U_0(\theta)$, we have

$$\sup_{\theta\in\Theta} \|\hat{U}_{\text{IPW}}(\theta) - \mathbb{E}\hat{U}_{\text{IPW}}(\theta)\| = \sup_{\theta\in\Theta} \|\hat{U}_{\text{IPW}}(\theta) - U_0(\theta)\| = o_p(1) \tag{17}$$

as N goes to infinity. In the meanwhile, under condition 3 (iv), $\sup_{\theta\in\Theta} \|U_0(\theta) - U_{0*}(\theta)\| = o_p(1)$. This proves $\sup_{\theta\in\Theta} \left\|\hat{U}_{\text{IPW}}(\theta) - U_{0*}(\theta)\right\| = o_p(1)$ and therefore proves the consistency of $\hat{\theta}_{\text{SIPW}}$.

6.1.2 Step II: asymptotic normality of $\hat{\theta}_{\text{SIPW}}$

Under Condition 3 (ii), we have

$$0 = \hat{U}_{\text{IPW}}(\hat{\theta}_{\text{SIPW}}) = \hat{U}_{\text{IPW}}(\theta_0) + \hat{U}^{(1)}_{\text{IPW}}(\breve{\theta})(\hat{\theta}_{\text{SIPW}} - \theta_0). \tag{18}$$

where $\hat{U}^{(1)}_{\text{IPW}}(\theta) = \partial\hat{U}_{\text{IPW}}(\theta)/\partial\theta^\top = N^{-1}\sum_{i=1}^{N} g_1(Z_i,\theta)D_i/\pi_i$, and $\breve{\theta} = \eta\hat{\theta}_{\text{SIPW}} + (1-\eta)\theta_0$ for some $\eta \in [0,1]$.

We wish to show that $\hat{U}^{(1)}_{\text{IPW}}(\breve{\theta}) = U^{(1)}_0(\theta_0) + o_p(1) = U^{(1)}_{0*}(\theta_0) + o_p(1)$, so that equation (26) can be simplified. Note that

$$\left\|\hat{U}^{(1)}_{\text{IPW}}(\breve{\theta}) - U^{(1)}_0(\theta_0)\right\|_F \leq \left\|\hat{U}^{(1)}_{\text{IPW}}(\breve{\theta}) - U^{(1)}_0(\breve{\theta})\right\|_F + \left|U^{(1)}_0(\breve{\theta}) - U^{(1)}_0(\theta_0)\right|_F \equiv J_1 + J_2.$$

By a proof similar to that of (17), we have $\sup_{\theta\in\Theta} \left\|\hat{U}^{(1)}_{\text{IPW}}(\theta) - \hat{U}^{(1)}_0(\theta)\right\|_F = o_p(1)$, which means

$$J_1 \leq \sup_{\theta\in\Theta} \left\|\hat{U}^{(1)}_{\text{IPW}}(\theta) - \hat{U}^{(1)}_0(\theta)\right\|_F = o_p(1).$$

For J_2, because $\hat{\theta}_{\text{SIPW}} - \theta_0 = o_p(1)$, under Condition 4 (ii), we have $J_2 = o_p(1)$ by dominated convergence theorem. Therefore

$$\hat{U}^{(1)}_{\text{IPW}}(\breve{\theta}) = U^{(1)}_0(\theta_0) + o_p(1) = K_* + o_p(1),$$

where $K_* = \partial U_{0*}(\theta_{0*})/\partial\theta^\top$.

Then it follows from (26) that

$$\hat{\theta}_{\text{SIPW}} - \theta_0 = -K_*^{-1}\hat{U}_{\text{IPW}}(\theta_0) + o_p(\|\hat{\theta}_{\text{SIPW}} - \theta_0\|)$$

By Lemma 3, $\sqrt{N}\hat{U}_{\text{IPW}}(\theta_0) \xrightarrow{d} N(0, B_{gg*} - C_{gg*})$. Therefore

$$\sqrt{N}(\hat{\theta}_{\text{SIPW}} - \theta_0) \xrightarrow{d} N(0, \Sigma_{\text{SIPW}}), \quad \Sigma_{\text{SIPW}} = K_*^{-1}(B_{gg*} - C_{gg*})(K_*^{-1})^\top.$$

6.2 Proof of result (b) of Theorem 5.1

We first present a technical lemma, which can help simplify our proof. Recall that $n = \sum_{k=1}^N D_k$ is fixed, D_k's are not independent and that the true value of α is $\alpha_0 = n/N$.

Lemma 4. *Let D_k's be LIND random variables with $\pi_k = \mathbb{E}(D_k) > 0$ for $k = 1, 2, \ldots,$. Suppose that* $\max_{1\le i\le N}(1/\pi_i) = o(N^{1/2})$, $B_{11} = N^{-1}\sum_{k=1}^N 1/\pi_k$, $B_{11*} = \lim_{N\to\infty} B_{11} < \infty$, *and*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{1\le k\ne l\le N} \frac{\mathbb{C}\text{ov}(D_k, D_l)}{\pi_k\pi_l} = 0.$$

Then $N^{-1}\sum_{i=1}^N D_i(1-\pi_i)/\pi_i^2 = B_{11} - 1 + o_p(1)$ as $N \to \infty$.

Proof. Direct calculations give

$$\mathbb{V}\text{ar}\left(\frac{1}{N}\sum_{i=1}^N \frac{D_i(1-\pi_i)}{\pi_i^2}\right) = \frac{1}{N^2}\sum_{i=1}^N \left(\frac{1-\pi_i}{\pi_i}\right)^3 + \frac{1}{N^2}\sum_{1\le k\ne l\le N} \frac{\mathbb{C}\text{ov}(D_k, D_l)}{\pi_k\pi_l}\frac{(1-\pi_k)(1-\pi_l)}{\pi_k\pi_l}.$$

Because D_k's are LIND random variables, $\mathbb{C}\text{ov}(D_k, D_l) \le 0$ for all $1 \le k \ne l \le N$. This, together with $0 < \pi_i \le 1$, implies

$$\begin{aligned}
\mathbb{V}\text{ar}\left(\frac{1}{N}\sum_{i=1}^N \frac{D_i(1-\pi_i)}{\pi_i^2}\right) &\le \left|\frac{1}{N^2}\sum_{i=1}^N \frac{1}{\pi_i^3}\right| + \left|\frac{1}{N^2}\sum_{1\le k\ne l\le N} \frac{\mathbb{C}\text{ov}(D_k, D_l)}{\pi_k\pi_l}\frac{1}{\pi_k\pi_l}\right| \\
&\le \max_{1\le i\le N}\frac{N}{\pi_i^2}\cdot\left\{\left|\frac{1}{N}\sum_{i=1}^N\frac{1}{\pi_i}\right| + \left|\frac{1}{N}\sum_{1\le k\ne l\le N}\frac{\mathbb{C}\text{ov}(D_k, D_l)}{\pi_k\pi_l}\right|\right\}.
\end{aligned}$$

Then by the conditions in the lemma, we have

$$\mathbb{V}\text{ar}\left(\frac{1}{N}\sum_{i=1}^N \frac{D_i(1-\pi_i)}{\pi_i^2}\right) \le o(1)\cdot\{O(1) + o(1)\} = o(1),$$

which means

$$\frac{1}{N}\sum_{i=1}^N \frac{D_i(1-\pi_i)}{\pi_i^2} = \frac{1}{N}\sum_{i=1}^N \frac{1-\pi_i}{\pi_i} + o_p(1) = B_{11} - 1 + o_p(1).$$

This completes the proof. ◻

Our proof of result (b) of Theorem 5.1 consists of five steps. In Step I, we show the consistency of $\hat{\alpha}$ and derive a linear approximate for it, based on which we derive a linear approximate for $\lambda(\hat{\alpha})$ in Step II. We prove the consistency of $\hat{\theta}_{\text{ELW}}$ in Step III, and derive a linear approximate for $\hat{\theta}_{\text{ELW}}$ and prove its asymptotic normality in Step IV. Step V shows that $\hat{\theta}_{\text{ELW}}$ is more efficient than $\hat{\theta}_{\text{SIPW}}$.

6.2.1 Step I: consistency and approximate of $\hat{\alpha}$

We first prove the consistency of $\hat{\alpha}$, i.e. $\hat{\alpha} - \alpha_0 = O_p(N^{-1/2})$. By definition, $\hat{\alpha}$ satisfies

$$\frac{1}{N}\sum_{k=1}^{N} \frac{D_k(\pi_k - \hat{\alpha})}{\alpha_0 + (1-\alpha_0)\pi_k - \hat{\alpha}} = 0 \tag{19}$$

and $\min_{1\le i\le N, D_i=1} \pi_i < \hat{\alpha} < \alpha_0 + (1-\alpha_0)\min_{1\le i\le N, D_i=1}\pi_i$.

Let $\hat{\gamma} = (\alpha_0 - \hat{\alpha})/(1-\alpha_0)$. Equation (19) can be rewritten as

$$\begin{aligned}
0 &= \frac{1}{N}\sum_{i=1}^{N} \frac{D_i}{\pi_i + \hat{\gamma}} \{\pi_i - \alpha_0 + (1-\alpha_0)\hat{\gamma}\} \\
&= \frac{1}{N}\sum_{i=1}^{N} \frac{D_i}{1+\hat{\gamma}D_i/\pi_i} \left\{1 - \frac{\alpha_0}{\pi_i} + \frac{1}{\pi_i}(1-\alpha_0)\hat{\gamma}\right\} \\
&= \frac{1}{N}\sum_{i=1}^{N} D_i\left(1 - \frac{\alpha_0}{\pi_i}\right) + \alpha_0 \cdot \frac{1}{N}\sum_{i=1}^{N} \frac{D_i/\pi_i}{1+\hat{\gamma}D_i/\pi_i}\left(\frac{1}{\pi_i} - 1\right)\hat{\gamma} \\
&= \alpha_0 \frac{1}{N}\sum_{i=1}^{N} \frac{\pi_i - D_i}{\pi_i} + \alpha_0 \cdot \frac{1}{N}\sum_{i=1}^{N} \frac{D_i/\pi_i}{1+\hat{\gamma}D_i/\pi_i}\left(\frac{1}{\pi_i} - 1\right)\hat{\gamma},
\end{aligned} \tag{20}$$

where we have used $\sum_{i=1}^{N} D_i/N = \alpha_0$. This implies

$$\left|\frac{1}{N}\sum_{i=1}^{N} \frac{\pi_i - D_i}{\pi_i}\right| = \left|\frac{1}{N}\sum_{i=1}^{N} \frac{D_i/\pi_i}{1+\hat{\gamma}D_i/\pi_i}\frac{1-\pi_i}{\pi_i}\hat{\gamma}\right|. \tag{21}$$

By Lemma 3, the left-hand side of equation (21) is clearly $O_p(N^{-1/2})$. The conditions in Lemma 4 are all satisfied under the conditions in Theorem 5.1. It follows from equation (21) and Lemma 4 that

$$\begin{aligned}
O_p(N^{-1/2}) &\ge \frac{1}{1+|\hat{\gamma}|\max_{1\le i\le N} D_i/\pi_i} \cdot \frac{1}{N}\sum_{i=1}^{N} \frac{D_i}{\pi_i}\frac{1-\pi_i}{\pi_i} \cdot |\hat{\gamma}| \\
&= \frac{1}{1+|\hat{\gamma}|\cdot o(N^{1/2})} \cdot O_p(1) \cdot |\hat{\gamma}|,
\end{aligned}$$

which implies $\hat{\gamma} = O_p(N^{-1/2})$. Because $\alpha_0 = \alpha_{0*} + o(1)$ and $\alpha_{0*} \in (0,1)$, we have

$$\hat{\alpha} - \alpha_0 = -\hat{\gamma}(1-\alpha_0) = O_p(N^{-1/2}).$$

The order $\hat{\gamma} = O_p(N^{-1/2})$ together with $\max_{1\le i\le N} D_i/\pi_i = o(N^{1/2})$ implies

$$\hat{\gamma} \max_{1\le i\le N} \frac{D_i}{\pi_i} = o_p(1).$$

It then follows from equation (20) that

$$\begin{aligned} 0 &= \frac{1}{N}\sum_{i=1}^{N} \frac{\pi_i - D_i}{\pi_i} + \frac{1}{N}\sum_{i=1}^{N} \frac{D_i}{\pi_i}\left(\frac{1}{\pi_i} - 1\right)\frac{\alpha_0 - \hat{\alpha}}{1-\alpha_0} + o_p(N^{-1/2}) \\ &= \frac{1}{N}\sum_{i=1}^{N} \frac{\pi_i - D_i}{\pi_i} - (B_{11}-1)\frac{\hat{\alpha} - \alpha_0}{1-\alpha_0} + o_p(N^{-1/2}), \end{aligned}$$

where the second equality follows from Lemma 4. Thus we have

$$\hat{\alpha} - \alpha_0 = -\frac{1-\alpha_0}{B_{11}-1}\frac{1}{N}\sum_{i=1}^{N}\frac{D_i - \pi_i}{\pi_i} + o_p(N^{-1/2}).$$

6.2.2 Step II: approximate of $\lambda(\hat{\alpha})$.

By definition,

$$\lambda(\hat{\alpha}) = \frac{1-\alpha_0}{\alpha_0}\cdot\frac{1}{1-\hat{\alpha}}.$$

Because $\hat{\alpha} - \alpha_0 = O_p(N^{-1/2})$, by Taylor expansion we have

$$\begin{aligned} \lambda(\hat{\alpha}) - \alpha_0^{-1} &= \frac{1-\alpha_0}{\alpha_0}\cdot\frac{\hat{\alpha}-\alpha_0}{(1-\alpha_0)^2} + o_p(N^{-1/2}) \\ &= \frac{\hat{\alpha}-\alpha_0}{\alpha_0(1-\alpha_0)} + o_p(N^{-1/2}). \end{aligned}$$

6.2.3 Step III: consistency of $\hat{\theta}_{\mathrm{ELW}}$

The ELW estimator $\hat{\theta}_{\mathrm{ELW}}$ of θ_0 is the solution to $\hat{U}_{\mathrm{ELW}}(\theta) = 0$, where

$$\hat{U}_{\mathrm{ELW}}(\theta) = \sum_{k=1}^{N} \hat{p}_k g(Z_k, \theta)$$

with

$$\hat{p}_k = \frac{D_k}{n}\cdot\frac{1}{1+\lambda(\hat{\alpha})(\pi_k - \hat{\alpha})}.$$

The proof is similar to that of the consistency of $\hat{\theta}_{\mathrm{SIPW}}$ in Section 6.1.1. We need only prove

$$\sup_{\theta\in\Theta}\|\hat{U}_{\mathrm{ELW}}(\theta) - \hat{U}_0(\theta)\| = o_p(1).$$

Observing that

$$\sup_{\theta\in\Theta}\|\hat{U}_{\mathrm{ELW}}(\theta) - \hat{U}_0(\theta)\| \le \sup_{\theta\in\Theta}\|\hat{U}_{\mathrm{ELW}}(\theta) - \hat{U}_{\mathrm{IPW}}(\theta)\| + \sup_{\theta\in\Theta}\|\hat{U}_{\mathrm{IPW}}(\theta) - \hat{U}_0(\theta)\|,$$

because of equation (17), it suffices to show

$$\sup_{\theta\in\Theta} \|\hat{U}_{\mathrm{ELW}}(\theta) - \hat{U}_{\mathrm{IPW}}(\theta)\| = o_p(1). \tag{22}$$

Let $\delta_1 = \hat{\alpha} - \alpha_0$ and $\delta_2 = \lambda(\hat{\alpha}) - \alpha_0^{-1}$. We have

$$\begin{aligned}
\hat{p}_k &= \frac{D_k}{N\alpha_0}\frac{1}{1 + (\alpha_0^{-1} + \delta_2)(\pi_k - \alpha_0 - \delta_1)} && (23)\\
&= \frac{1}{N}\frac{D_k}{\pi_k} \cdot \frac{1}{1 - \pi_k^{-1}\delta_1 + \delta_2\alpha_0 - \delta_2\alpha_0^2\pi_k^{-1} - \delta_1\delta_2\alpha_0\pi_k^{-1}}\\
&= \frac{D_k}{N\pi_k} \cdot (1 + \zeta_{k1} + \zeta_{k2}), && (24)
\end{aligned}$$

where $\zeta_{k1} = \pi_k^{-1}\delta_1 - \delta_2\alpha_0 + \delta_2\alpha_0^2\pi_k^{-1}$ and

$$\zeta_{k2} = \frac{\zeta_{k1}^2 - \zeta_{k1}\delta_1\delta_2\alpha_0\pi_k^{-1} - \delta_1\delta_2\alpha_0\pi_k^{-1}}{1 - \zeta_{k1} - \delta_1\delta_2\alpha_0\pi_k^{-1}}.$$

It follows that

$$\begin{aligned}
\sup_{\theta\in\Theta} \|\hat{U}_{\mathrm{ELW}}(\theta) - \hat{U}_{\mathrm{IPW}}(\theta)\| &= \sup_{\theta\in\Theta} \left\| \sum_{k=1}^N \left(\hat{p}_k - \frac{D_k}{N\pi_k}\right) g(Z_k, \theta) \right\|\\
&\le \sup_{\theta\in\Theta} \left\| \sum_{k=1}^N \frac{D_k}{N\pi_k} \cdot (\zeta_{k1} + \zeta_{k2}) g(Z_k, \theta) \right\|\\
&\le \sum_{k=1}^N \frac{D_k}{N\pi_k} \|\bar{g}(Z_k)\| \cdot (\max_{1\le k\le N} |\zeta_{k1}| + \max_{1\le k\le N} |\zeta_{k2}|).
\end{aligned}$$

The facts that $\delta_1 = O_p(N^{-1/2})$, $\delta_2 = O_p(N^{-1/2})$, and $\max_{1\le i\le N} 1/\pi_i = o(N^{1/2})$ imply $\max_{1\le k\le N} |\zeta_{k1}| = o_p(1)$ and $\max_{1\le k\le N} |\zeta_{k2}| = o_p(1)$. Condition 4 (ii) guarantees that $\sum_{k=1}^N \|\bar{g}(Z_k)\| D_k/(N\pi_k) = O_p(1)$. Therefore equation (22) follows.

6.2.4 Step IV: approximate and asymptotic normality of $\hat{\theta}_{\mathrm{ELW}}$

By mean value theorem, we have

$$0 = \hat{U}_{\mathrm{ELW}}(\hat{\theta}_{\mathrm{ELW}}) = \hat{U}_{\mathrm{ELW}}(\theta_0) + \hat{U}_{\mathrm{ELW}}^{(1)}(\breve{\theta})(\hat{\theta}_{\mathrm{ELW}} - \theta_0),$$

where

$$\hat{U}_{\mathrm{ELW}}^{(1)}(\theta) = \frac{\partial \hat{U}_{\mathrm{ELW}}(\theta)}{\partial \theta^\top} = \sum_{k=1}^N \hat{p}_k g_1(Z_k, \theta)$$

and $\breve{\theta} = \rho\hat{\theta}_{\mathrm{ELW}} + (1-\rho)\theta_0$ for some $\rho \in [0,1]$. We have shown that $\hat{\theta}_{\mathrm{ELW}} - \theta_0 = o_p(1)$, therefore $\breve{\theta} - \theta_0 = o_p(1)$

We approximate $\hat{U}_{\mathrm{ELW}}(\theta_0)$ and $\hat{U}^{(1)}_{\mathrm{ELW}}(\breve{\theta})$, separately. For, $\hat{U}^{(1)}_{\mathrm{ELW}}(\breve{\theta})$, using the approximate of $\hat{p}_k$ in (24), we have

$$\begin{aligned}
\hat{U}^{(1)}_{\mathrm{ELW}}(\breve{\theta}) &= \sum_{i=1}^{N} \frac{D_k}{N\pi_k} \cdot (1+\zeta_{k1}+\zeta_{k2}) g_1(Z_k,\breve{\theta}) \\
&= \sum_{i=1}^{N} \frac{D_k}{N\pi_k} g_1(Z_k,\theta_0) + o_p(1) \\
&= K_* + o_p(1),
\end{aligned}$$

where the second equality follows from $\max_{1\le k\le N}|\zeta_{k1}| = o_p(1)$, $\max_{1\le k\le N}|\zeta_{k2}| = o_p(1)$, and $\sup_{\theta\in N_0}\|g_1(Z,\theta)\|_F \le \bar{g}(Z)$. Therefore

$$\hat{\theta}_{\mathrm{ELW}} - \theta_0 = -K_*^{-1}\hat{U}_{\mathrm{ELW}}(\theta_0) + o_p(N^{-1/2}).$$

For $\hat{U}_{\mathrm{ELW}}(\theta_0)$, again using the approximate of $\hat{p}_k$ in (23), we have

$$\hat{U}_{\mathrm{ELW}}(\theta_0) = \sum_{i=1}^{N} \frac{D_k}{N\alpha_0} \frac{g(Z_k,\theta_0)}{1+(\alpha_0^{-1}+\delta_2)(\pi_k-\alpha_0-\delta_1)}.$$

By first-order Taylor expansion with respect to (δ_1,δ_2) at $(0,0)$, we have

$$\hat{U}_{\mathrm{ELW}}(\theta_0) = \sum_{i=1}^{N} \frac{D_k g(Z_k,\theta_0)}{N\pi_k} + \sum_{i=1}^{N} \frac{D_k}{N}\frac{g(Z_k,\theta_0)}{\pi_k^2}\delta_1 - \alpha_0 \sum_{i=1}^{N} \frac{D_k}{N}\frac{g(Z_k,\theta_0)}{\pi_k^2}(\pi_k-\alpha_0)\delta_2 + o_p(N^{-1/2}).$$

By a proof similar to that of Lemma 4, we can show that under Condition 4 (ii),

$$\sum_{i=1}^{N} \frac{D_k}{N}\frac{g(Z_k,\theta_0)}{\pi_k^2} = B_{g1} + o_p(1).$$

Using this result and $\delta_2 = \delta_1/\{\alpha_0(1-\alpha_0)\} + o_p(N^{-1/2})$, we further have

$$\hat{U}_{\mathrm{ELW}}(\theta_0) = \sum_{i=1}^{N} \frac{D_k}{N\pi_k} g(Z_k,\theta_0) + \frac{B_{g1}}{1-\alpha_0}\delta_1 + o_p(N^{-1/2}).$$

Putting

$$\delta_1 = \hat{\alpha} - \alpha_0 = -\frac{1-\alpha_0}{B_{11}-1}\frac{1}{N}\sum_{i=1}^{N}\frac{D_i-\pi_i}{\pi_i} + o_p(N^{-1/2}).$$

into the above equation gives

$$\begin{aligned}
\hat{U}_{\mathrm{ELW}}(\theta_0) &= \sum_{i=1}^{N} \frac{D_k}{N\pi_k} g(Z_k,\theta_0) - \frac{B_{g1}}{B_{11}-1}\frac{1}{N}\sum_{i=1}^{N}\frac{D_i-\pi_i}{\pi_i} + o_p(N^{-1/2}) \\
&= \sum_{i=1}^{N} \frac{D_k-\pi_k}{N\pi_k}\left\{g(Z_k,\theta_0) - \frac{B_{g1}}{B_{11}-1}\right\} + o_p(N^{-1/2}) \\
&= \sum_{i=1}^{N} \frac{D_k-\pi_k}{N\pi_k}\left\{g(Z_k,\theta_0) - \frac{B_{g1*}}{B_{11*}-1}\right\} + o_p(N^{-1/2}),
\end{aligned}$$

where we have used the fact that $\sum_{i=1}^N g(Z_k, \theta_0) = 0$. Let $G_* = B_{g1*}/(B_{11*} - 1)$. We can write

$$\sqrt{N}\hat{U}_{\mathrm{ELW}}(\theta_0) \quad = \quad (I_r,\ -G_*)\,\frac{1}{\sqrt{N}}\sum_{k=1}^N (D_k - \pi_k)W_k + o_p(1),$$

where I_r is an $r \times r$ identity matrix and $W_k = (g^\top(Z_k, \theta_0), 1)^\top/\pi_k$ is defined in Lemma 3. It follows from Lemma 3 that

$$\sqrt{N}\hat{U}_{\mathrm{ELW}}(\theta_0) \xrightarrow{d} N(0, \Omega_{\mathrm{ELW}}),$$

where

$$\begin{aligned}
\Omega_{\mathrm{ELW}} &= (I_r, -G_*^\top)\, V_{0*}\, (I_r, -G^\top)^\top \\
&= (I_r, -G_*^\top) \begin{pmatrix} B_{gg*} - C_{gg*} & B_{g1*} \\ B_{g1*}^\top & B_{11*} - 1 \end{pmatrix} (I_r, -G^\top)^\top \\
&= B_{gg*} - C_{gg*} - \frac{B_{g1*}B_{g1*}^\top}{B_{11*} - 1}.
\end{aligned}$$

Consequently, we have

$$\hat{\theta}_{\mathrm{ELW}} - \theta_0 = -K_*^{-1}\hat{U}_{\mathrm{ELW}}(\theta_0) + o_p(N^{-1/2}) \xrightarrow{d} N(0, \Sigma_{\mathrm{ELW}}),$$

where

$$\Sigma_{\mathrm{ELW}} \quad = \quad K_*^{-1}\Omega_{\mathrm{ELW}}K_*^{-1} = K_*^{-1}\left(B_{gg*} - C_{gg*} - \frac{B_{g1*}B_{g1*}^\top}{B_{11*} - 1}\right)K_*^{-1}.$$

6.2.5 Step V: efficiency comparison

We have shown

$$\sqrt{N}(\hat{\theta}_{\mathrm{SIPW}} - \theta_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{SIPW}}), \quad \Sigma_{\mathrm{SIPW}} = K_*^{-1}(B_{gg*} - C_{gg*})(K_*^{-1})^\top.$$

Because

$$\Sigma_{\mathrm{ELW}} = \Sigma_{\mathrm{SIPW}} - K_*^{-1}\frac{B_{g1*}B_{g1*}^\top}{B_{11*} - 1}K_*^{-1}$$

and

$$K_*^{-1}\frac{B_{g1*}B_{g1*}^\top}{B_{11*} - 1}K_*^{-1}$$

is a nonnegative definite matrix, we conclude that

$$\Sigma_{\mathrm{ELW}} \le \Sigma_{\mathrm{SIPW}}$$

and that the ELW estimator is more efficient than the SIPW estimator.

6.3 Proof of result (c) of Theorem 5.1

In the case of $g(z,\theta) = f(z) - \theta$, $\hat{\theta}_{\mathrm{IPW}} = N^{-1}\sum_{i=1}^{N} f(Z_i)D_i/\pi_i$. It can be verified that $K_* = 1$, and B_{gg*} and C_{gg*} reduce to $B_{ff*} - B_{f1*}\theta_{0*}^{\top} - \theta_{0*}B_{f1*}^{\top} + \theta_{0*}\theta_{0*}^{\top}B_{11*}$ and $C_{ff*} - \theta_{0*}\theta_{0*}^{\top}$. Similar to the proof of the asymptotic normality of $\hat{\theta}_{\mathrm{SIPW}}$, we immediately have

$$\sqrt{N}(\hat{\theta}_{\mathrm{IPW}} - \theta_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{IPW}}), \quad \Sigma_{\mathrm{IPW}} = B_{ff*} - C_{ff*}.$$

When $g(z,\theta) = f(z) - \theta$, Σ_{ELW} becomes

$$B_{ff*} - C_{ff*} - \frac{(B_{f1*} - \theta_{0*})^{\otimes 2}}{B_{11*} - 1}$$

which clearly implies that

$$\Sigma_{\mathrm{ELW}} \le \Sigma_{\mathrm{IPW}}$$

and that the ELW estimator is more efficient than the IPW estimator.

7 Proof of Theorem 5.2

7.1 Proof of result (a) of Theorem 5.2

We shall first prove the consistency of $\hat{\theta}_{\mathrm{SIPW}}$ and then derive a linear approximate for it, from which its asymptotic normality can be established.

7.1.1 Consistency of $\hat{\theta}_{\mathrm{SIPW}}$

Recall that the SIPW estimator $\hat{\theta}_{\mathrm{SIPW}}$ of θ_0 under UPW-WR is the solution to $\hat{U}_{\mathrm{IPW}}(\theta) = 0$, where

$$\hat{U}_{\mathrm{IPW}}(\theta) = \frac{1}{nN}\sum_{i=1}^{n}\frac{g(z_i,\theta)}{q_i} = \frac{1}{N}\sum_{i=1}^{n}\frac{g(z_i,\theta)}{\pi_i}.$$

Similar to the proof of $\hat{\theta}_{\mathrm{SIPW}}$ under UPS-WOR, it suffices to show

$$\sup_{\theta\in\Theta}\left\|\hat{U}_{\mathrm{IPW}}(\theta) - U_0(\theta)\right\| = o_p(1).$$

Because for fixed N, (z_i, π_i) $(1 \le i \le n)$ are iid random vectors, we have

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left(\frac{1}{N}\sum_{i=1}^{n}\frac{\gamma^{\top}g(z_i,\theta)}{\pi_i}\right) &= \frac{n}{N^2}\mathbb{V}\mathrm{ar}\left(\frac{\gamma^{\top}g(z_i,\theta)}{\pi_i}\right) \\
&\le \frac{n}{N^2}\mathbb{E}\left(\frac{\gamma^{\top}g(z_i,\theta)}{\pi_i}\right)^2 \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\frac{\{\gamma^{\top}g(Z_i,\theta)\}^2}{\tilde{\pi}_i} \\
&\le \frac{1}{N^2}\sum_{i=1}^{N}\frac{\{\bar{g}(Z_i)\}^2}{\tilde{\pi}_i} = o(1).
\end{aligned}$$

Because the last inequality holds uniformly for all θ and all unit vector γ and $\mathbb{E}\hat{U}_{\mathrm{IPW}}(\theta) = U_0(\theta)$, we have

$$\sup_{\theta\in\Theta} \|\hat{U}_{\mathrm{IPW}}(\theta) - \mathbb{E}\hat{U}_{\mathrm{IPW}}(\theta)\| = \sup_{\theta\in\Theta} \|\hat{U}_{\mathrm{IPW}}(\theta) - U_0(\theta)\| = o_p(1) \tag{25}$$

as N goes to infinity. This completes the proof.

7.1.2 Step II: asymptotic normality of $\hat{\theta}_{\mathrm{SIPW}}$

Under Condition 4 (ii), we have

$$0 = \hat{U}_{\mathrm{IPW}}(\hat{\theta}_{\mathrm{SIPW}}) = \hat{U}_{\mathrm{IPW}}(\theta_0) + \hat{U}^{(1)}_{\mathrm{IPW}}(\breve{\theta})(\hat{\theta}_{\mathrm{SIPW}} - \theta_0). \tag{26}$$

where $\hat{U}^{(1)}_{\mathrm{IPW}}(\theta) = \partial\hat{U}_{\mathrm{IPW}}(\theta)/\partial\theta^\top = N^{-1}\sum_{i=1}^{n} g_1(z_i,\theta)/\pi_i$, and $\breve{\theta} = \eta\hat{\theta}_{\mathrm{SIPW}} + (1-\eta)\theta_0$ for some $\eta \in [0,1]$.

We wish to show that $\hat{U}^{(1)}_{\mathrm{IPW}}(\breve{\theta}) = U^{(1)}_0(\theta_0) + o_p(1) = U^{(1)}_{0*}(\theta_0) + o_p(1)$, so that equation (26) can be simplified. Note that

$$\left\|\hat{U}^{(1)}_{\mathrm{IPW}}(\breve{\theta}) - U^{(1)}_0(\theta_0)\right\|_F \quad \leq \quad \left\|\hat{U}^{(1)}_{\mathrm{IPW}}(\breve{\theta}) - U^{(1)}_0(\breve{\theta})\right\|_F + \left|U^{(1)}_0(\breve{\theta}) - U^{(1)}_0(\theta_0)\right|_F \equiv J_1 + J_2.$$

By a proof similar to that of (25), we have $\sup_{\theta\in\Theta}\left\|\hat{U}^{(1)}_{\mathrm{IPW}}(\theta) - \hat{U}^{(1)}_0(\theta)\right\|_F = o_p(1)$, which means

$$J_1 \leq \sup_{\theta\in N_0}\left\|\hat{U}^{(1)}_{\mathrm{IPW}}(\theta) - \hat{U}^{(1)}_0(\theta)\right\|_F = o_p(1).$$

For J_2, because $\hat{\theta}_{\mathrm{SIPW}} - \theta_0 = o_p(1)$, under Condition 4(ii), we have $J_2 = o_p(1)$ by dominated convergence theorem. Therefore

$$\hat{U}^{(1)}_{\mathrm{IPW}}(\tilde{\theta}) = U^{(1)}_0(\theta_0) + o_p(1) = K_* + o_p(1),$$

where $K_* = \partial U_{0*}(\theta_{0*})/\partial\theta^\top$ is defined in Condition 3 (v).

Then it follows from (26) that

$$\hat{\theta}_{\mathrm{SIPW}} - \theta_0 = -K_*^{-1}\hat{U}_{\mathrm{IPW}}(\theta_0) + o_p(\|\hat{\theta}_{\mathrm{SIPW}} - \theta_0\|).$$

Because for fixed N, (z_i, π_i) $(1 \leq i \leq n)$ are iid random vectors, by Lindeberg-Feller central limit theorem we have $\sqrt{N}\hat{U}_{\mathrm{IPW}}(\theta_0) \xrightarrow{d} N(0, B_{gg*})$, where $B_{gg*} = \lim_{N\to\infty} B_{gg}$ and

$$B_{gg} = \frac{1}{N}\sum_{i=1}^{N}\frac{\{g(Z_i,\theta_0)\}^{\otimes 2}}{\tilde{\pi}_i}.$$

Therefore

$$\sqrt{N}(\hat{\theta}_{\mathrm{SIPW}} - \theta_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{SIPW}}), \quad \Sigma_{\mathrm{SIPW}} = K_*^{-1}B_{gg*}(K_*^{-1})^\top.$$

7.2 Proof of result (b) of Theorem 5.2

Recall that $\alpha_0 = n/N$ and $\hat{\alpha}$ satisfies

$$0 \;=\; \frac{1}{N}\sum_{i=1}^{n}\frac{\pi_i - \hat{\alpha}}{\alpha_0 + (1-\alpha_0)\pi_i - \hat{\alpha}}, \tag{27}$$

and $\min_{1\le i\le n}\pi_i \le \hat{\alpha} \le \alpha_0 + (1-\alpha_0)\min_{1\le i\le n}\pi_i$. The ELW weights are

$$\hat{p}_i = \frac{1}{n}\frac{1}{1+\lambda(\hat{\alpha})(\pi_i - \hat{\alpha})}$$

with $\lambda(\hat{\alpha}) = (N-n)/\{n(1-\hat{\alpha})\} = (1-\alpha_0)/\{\alpha_0(1-\hat{\alpha})\}$.

We prove result (b) in five steps. In step I, we first prove the $O_p(N^{-1/2})$ and then derive linear approximates for $\hat{\alpha}$ and $\lambda(\hat{\alpha})$. In step II, we prove the consistency of $\hat{\theta}_{\mathrm{ELW}}$. Based on its consistency, we derive a linear approximate for $\hat{\theta}_{\mathrm{ELW}}$ and prove its asymptotic normality in step III. In step IV, we show that $\hat{\theta}_{\mathrm{ELW}}$ is more efficient than $\hat{\theta}_{\mathrm{SIPW}}$.

7.2.1 Step I: approximate of $\hat{\alpha}$ and $\lambda(\hat{\alpha})$

We first prove $\hat{\alpha} - \alpha_0 = O_p(N^{-1/2})$. Let $\hat{\gamma} = (\alpha_0 - \hat{\alpha})/(1-\alpha_0)$. Equation (27) can be rewritten as

$$\begin{aligned}
0 \;&=\; \frac{1}{N}\sum_{i=1}^{n}\frac{1}{\pi_i + \hat{\gamma}}\{\pi_i - \alpha_0 + (1-\alpha_0)\,\hat{\gamma}\}\\
&=\; \frac{1}{N}\sum_{i=1}^{n}\frac{1}{1+\hat{\gamma}/\pi_i}\left\{1 - \frac{\alpha_0}{\pi_i} + \frac{1}{\pi_i}(1-\alpha_0)\,\hat{\gamma}\right\}\\
&=\; \frac{1}{N}\sum_{i=1}^{n}\left(1 - \frac{\alpha_0}{\pi_i}\right) + \alpha_0\cdot\frac{1}{N}\sum_{i=1}^{n}\frac{1/\pi_i}{1+\hat{\gamma}/\pi_i}\left(\frac{1}{\pi_i} - 1\right)\hat{\gamma}\\
&=\; \frac{\alpha_0}{N}\sum_{i=1}^{n}\left(\frac{1}{\alpha_0} - \frac{1}{\pi_i}\right) + \frac{\alpha_0}{N}\sum_{i=1}^{n}\frac{1/\pi_i}{1+\hat{\gamma}/\pi_i}\left(\frac{1}{\pi_i} - 1\right)\hat{\gamma},,
\end{aligned} \tag{28}$$

where $\sum_1$ stands for the summation over i such that p_i This implies

$$\left|\frac{1}{N}\sum_{i=1}^{n}\left(\frac{1}{\alpha_0} - \frac{1}{\pi_i}\right)\right| = \left|\frac{1}{N}\sum_{i=1}^{n}\frac{1/\pi_i}{1+\hat{\gamma}/\pi_i}\left(\frac{1}{\pi_i} - 1\right)\hat{\gamma}\right|. \tag{29}$$

The left-hand side of equation (29) is clearly $O_p(N^{-1/2})$. We have assumed $\alpha_0 \to \alpha_{0*} \in (0,1)$. If

$$\alpha_0\mathbb{E}(1/\pi_i^2) = B_{11} = N^{-1}\sum_{i=1}^{N}1/\tilde{\pi}_i \to B_{11*} \in (0,1),$$

then $\max_{1\le i\le n}1/\pi_i = o(n^{1/2}) = o(N^{1/2})$. It follows from equation (29) that

$$\begin{aligned}
O_p(N^{-1/2}) \;&\ge\; \frac{1}{1+|\hat{\gamma}|\max_{1\le i\le n}1/\pi_i}\cdot\frac{1}{N}\sum_{i=1}^{n}\frac{1}{\pi_i}\frac{1-\pi_i}{\pi_i}\cdot|\hat{\gamma}|\\
&=\; \frac{1}{1+|\hat{\gamma}|\cdot o(N^{1/2})}\cdot O_p(1)\cdot|\hat{\gamma}|,
\end{aligned}$$

which implies $\hat{\gamma} = O_p(N^{-1/2})$. Because $\alpha_0 = \alpha_{0*} + o(1)$ and $\alpha_{0*} \in (0,1)$, we have

$$\hat{\alpha} - \alpha_0 = -\hat{\gamma}\,(1-\alpha_0) = O_p(N^{-1/2}).$$

The order $\hat{\gamma} = O_p(N^{-1/2})$ together with $\max_{1\le i\le n} 1/\pi_i = o(N^{1/2})$ implies

$$\hat{\gamma} \max_{1\le i\le n} \frac{1}{\pi_i} = o_p(1).$$

It then follows from equation (28) that

$$\begin{aligned} 0 &= \frac{1}{N}\sum_{i=1}^{n}\left(\frac{1}{\alpha_0} - \frac{1}{\pi_i}\right) + \frac{1}{N}\sum_{i=1}^{n}\frac{1}{\pi_i}\left(\frac{1}{\pi_i} - 1\right)\frac{\alpha_0 - \hat{\alpha}}{1-\alpha_0} + o_p(N^{-1/2}) \\ &= \frac{1}{N}\sum_{i=1}^{n}\left(\frac{1}{\alpha_0} - \frac{1}{\pi_i}\right) - (B_{11}-1)\frac{\hat{\alpha} - \alpha_0}{1-\alpha_0} + o_p(N^{-1/2}). \end{aligned}$$

Thus we have

$$\hat{\alpha} - \alpha_0 = \frac{1-\alpha_0}{B_{11}-1}\frac{1}{N}\sum_{i=1}^{n}\left(\frac{1}{\alpha_0} - \frac{1}{\pi_i}\right) + o_p(N^{-1/2}).$$

which implies

$$\lambda(\hat{\alpha}) - \alpha_0^{-1} = \frac{\hat{\alpha} - \alpha_0}{\alpha_0(1-\alpha_0)} + o_p(N^{-1/2}).$$

7.2.2 Step II: consistency of $\hat{\theta}_{\mathrm{ELW}}$

The ELW estimator $\hat{\theta}_{\mathrm{ELW}}$ of θ_0 is the solution to $\hat{U}_{\mathrm{ELW}}(\theta) = 0$, where

$$\hat{U}_{\mathrm{ELW}}(\theta) = \sum_{k=1}^{n} \hat{p}_k g(z_k, \theta)$$

with

$$\hat{p}_k = \frac{1}{n}\cdot\frac{1}{1+\lambda(\hat{\alpha})(\pi_k - \hat{\alpha})}.$$

The proof is similar to that of the consistency of $\hat{\theta}_{\mathrm{SIPW}}$ in Section 6.1.1. We need only prove

$$\sup_{\theta\in\Theta} \|\hat{U}_{\mathrm{ELW}}(\theta) - \hat{U}_0(\theta)\| = o_p(1).$$

Observing that

$$\sup_{\theta\in\Theta} \|\hat{U}_{\mathrm{ELW}}(\theta) - \hat{U}_0(\theta)\| \le \sup_{\theta\in\Theta} \|\hat{U}_{\mathrm{ELW}}(\theta) - \hat{U}_{\mathrm{IPW}}(\theta)\| + \sup_{\theta\in\Theta} \|\hat{U}_{\mathrm{IPW}}(\theta) - \hat{U}_0(\theta)\|,$$

because of equation (25), it suffices to show

$$\sup_{\theta\in\Theta} \|\hat{U}_{\mathrm{ELW}}(\theta) - \hat{U}_{\mathrm{IPW}}(\theta)\| = o_p(1). \tag{30}$$

Let $\delta_1 = \hat{\alpha} - \alpha_0$ and $\delta_2 = \lambda(\hat{\alpha}) - \alpha_0^{-1}$. We have

$$\begin{aligned} \hat{p}_k &= \frac{1}{N\alpha_0} \frac{1}{1 + (\alpha_0^{-1} + \delta_2)(\pi_k - \alpha_0 - \delta_1)} && (31) \\ &= \frac{1}{N} \frac{1}{\pi_k} \cdot \frac{1}{1 - \pi_k^{-1}\delta_1 + \delta_2\alpha_0 - \delta_2\alpha_0^2\pi_k^{-1} - \delta_1\delta_2\alpha_0\pi_k^{-1}} \\ &= \frac{1}{N\pi_k} \cdot (1 + \zeta_{k1} + \zeta_{k2}), && (32) \end{aligned}$$

where $\zeta_{k1} = \pi_k^{-1}\delta_1 - \delta_2\alpha_0 + \delta_2\alpha_0^2\pi_k^{-1}$ and

$$\zeta_{k2} = \frac{\zeta_{k1}^2 - \zeta_{k1}\delta_1\delta_2\alpha_0\pi_k^{-1} - \delta_1\delta_2\alpha_0\pi_k^{-1}}{1 - \zeta_{k1} - \delta_1\delta_2\alpha_0\pi_k^{-1}}.$$

It follows that

$$\begin{aligned} \sup_{\theta\in\Theta} \|\hat{U}_{\mathrm{ELW}}(\theta) - \hat{U}_{\mathrm{IPW}}(\theta)\| &= \sup_{\theta\in\Theta} \left\| \sum_{k=1}^n \left(\hat{p}_k - \frac{1}{N\pi_k} \right) g(z_k, \theta) \right\| \\ &\le \sup_{\theta\in\Theta} \left\| \sum_{k=1}^n \frac{1}{N\pi_k} \cdot (\zeta_{k1} + \zeta_{k2}) g(z_k, \theta) \right\| \\ &\le \sum_{k=1}^n \frac{1}{N\pi_k} \bar{g}(z_k) \cdot (\max_{1\le k\le n} |\zeta_{k1}| + \max_{1\le k\le n} |\zeta_{k2}|). \end{aligned}$$

The facts that $\delta_1 = O_p(N^{-1/2})$, $\delta_2 = O_p(N^{-1/2})$, and $\max_{1\le i\le n} 1/\pi_i = o(N^{1/2})$ imply $\max_{1\le k\le n} |\zeta_{k1}| = o_p(1)$ and $\max_{1\le k\le n} |\zeta_{k2}| = o_p(1)$. Under condition 5(i), $\mathbb{E}\sum_{k=1}^n \bar{g}(Z_k)/(N\pi_k) = (1/N)\sum_{k=1}^N \bar{g}(Z_k) \le (1/N)\sum_{k=1}^N \{\bar{g}(Z_k)\}^2/\tilde{\pi}_k = O(1)$ when N is large. Therefore $\sum_{k=1}^n \bar{g}(z_k)/(N\pi_k) = O_p(1)$ and equation (30) follows.

7.2.3 Step III: approximate and asymptotic normality of $\hat{\theta}_{\mathrm{ELW}}$

By mean value theorem, we have

$$0 = \hat{U}_{\mathrm{ELW}}(\hat{\theta}_{\mathrm{ELW}}) = \hat{U}_{\mathrm{ELW}}(\theta_0) + \hat{U}^{(1)}_{\mathrm{ELW}}(\breve{\theta})(\hat{\theta}_{\mathrm{ELW}} - \theta_0),$$

where

$$\hat{U}^{(1)}_{\mathrm{ELW}}(\theta) = \frac{\partial \hat{U}_{\mathrm{ELW}}(\theta)}{\partial \theta^\top} = \sum_{k=1}^n \hat{p}_k g_1(z_k, \theta)$$

and $\breve{\theta} = \rho\hat{\theta}_{\mathrm{ELW}} + (1-\rho)\theta_0$ for some $\rho \in [0, 1]$. We have shown that $\hat{\theta}_{\mathrm{ELW}} - \theta_0 = o_p(1)$, therefore $\breve{\theta} - \theta_0 = o_p(1)$

We approximate $\hat{U}_{\mathrm{ELW}}(\theta_0)$ and $\hat{U}^{(1)}_{\mathrm{ELW}}(\breve{\theta})$, separately. For, $\hat{U}^{(1)}_{\mathrm{ELW}}(\breve{\theta})$, using the approximate of $\hat{p}_k$ in (32), we have

$$\begin{aligned} \hat{U}^{(1)}_{\mathrm{ELW}}(\breve{\theta}) &= \sum_{i=1}^n \frac{1}{N\pi_k} \cdot (1 + \zeta_{k1} + \zeta_{k2}) g_1(z_k, \breve{\theta}) \\ &= \sum_{i=1}^n \frac{1}{N\pi_k} g_1(z_k, \theta_0) + o_p(1) \\ &= K_* + o_p(1), \end{aligned}$$

where the second and third equalities follow from $\max_{1\le k\le n}|\zeta_{k1}| = o_p(1)$, $\max_{1\le k\le n}|\zeta_{k2}| = o_p(1)$, and $\sup_{\theta\in N_0}\|g_1(Z,\theta)\|_F \le \bar{g}(Z)$. Therefore

$$\hat{\theta}_{\rm ELW} - \theta_0 \;\; = \;\; -K_*^{-1}\hat{U}_{\rm ELW}(\theta_0) + o_p(N^{-1/2}).$$

For $\hat{U}_{\rm ELW}(\theta_0)$, again using the approximate of $\hat{p}_k$ in (31), we have

$$\hat{U}_{\rm ELW}(\theta_0) \;\; = \;\; \sum_{i=1}^n \frac{1}{N\alpha_0}\frac{g(z_k,\theta_0)}{1+(\alpha_0^{-1}+\delta_2)(\pi_k-\alpha_0-\delta_1)}.$$

By first-order Taylor expansion with respect to (δ_1,δ_2) at $(0,0)$, we have

$$\hat{U}_{\rm ELW}(\theta_0) = \sum_{i=1}^n \frac{g(z_k,\theta_0)}{N\pi_k} + \sum_{i=1}^n \frac{1}{N}\frac{g(z_k,\theta_0)}{\pi_k^2}\delta_1 - \alpha_0\sum_{i=1}^n \frac{1}{N}\frac{g(z_k,\theta_0)}{\pi_k^2}(\pi_k-\alpha_0)\delta_2 + o_p(N^{-1/2}).$$

Using

$$\sum_{i=1}^n \frac{1}{N}\frac{g(z_k,\theta_0)}{\pi_k^2} = B_{g1} + o_p(1),$$

and $\delta_2 = \delta_1/\{\alpha_0(1-\alpha_0)\} + o_p(N^{-1/2})$, we further have

$$\hat{U}_{\rm ELW}(\theta_0) \;\; = \;\; \sum_{i=1}^n \frac{1}{N\pi_k}g(z_k,\theta_0) + \frac{B_{g1}}{1-\alpha_0}\delta_1 + o_p(N^{-1/2}).$$

Putting

$$\delta_1 = \hat{\alpha}-\alpha_0 = \frac{1-\alpha_0}{B_{11}-1}\frac{1}{N}\sum_{i=1}^n\left(\frac{1}{\alpha_0}-\frac{1}{\pi_i}\right) + o_p(N^{-1/2}).$$

into the above equation gives

$$\hat{U}_{\rm ELW}(\theta_0) \;\; = \;\; \sum_{i=1}^n \frac{g(z_i,\theta_0)}{N\pi_i} + \frac{B_{g1}}{B_{11}-1}\frac{1}{N}\sum_{i=1}^n\left(\frac{1}{\alpha_0}-\frac{1}{\pi_i}\right) + o_p(N^{-1/2}).$$

Let $G_* = B_{g1*}/(B_{11*}-1)$. It can be verified that It follows from Theorem 3.1 that

$$\sqrt{N}\hat{U}_{\rm ELW}(\theta_0) \xrightarrow{d} N(0,\Omega_{\rm ELW}),$$

where

$$\begin{aligned}\Omega_{\rm ELW} \;\; &= \;\; B_{gg*} - 2G_*^{\otimes 2}(B_{11*}-1) + G_*^{\otimes 2}(B_{11*}-\alpha_{0*}^{-1}) \\ &= \;\; B_{gg*} - \frac{B_{g1*}^{\otimes 2}}{(B_{11*}-1)^2}(B_{11*}+\alpha_{0*}^{-1}-2).\end{aligned}$$

Consequently, we have

$$\hat{\theta}_{\rm ELW} - \theta_0 = -K_*^{-1}\hat{U}_{\rm ELW}(\theta_0) + o_p(N^{-1/2}) \xrightarrow{d} N(0,\Sigma_{\rm ELW}),$$

where

$$\Sigma_{\rm ELW} \;\; = \;\; K_*^{-1}\Omega_{\rm ELW}K_*^{-1} = K_*^{-1}\left\{B_{gg*} - \frac{B_{g1*}^{\otimes 2}}{(B_{11*}-1)^2}(B_{11*}+\alpha_{0*}^{-1}-2)\right\}(K_*^{-1})^\top.$$

7.2.4 Step IV: Efficiency comparison

We have derived $\Sigma_{\text{SIPW}} = K_*^{-1} B_{gg*} (K_*^{-1})^\top$. Thus

$$\Sigma_{\text{SIPW}} - \Sigma_{\text{ELW}} \;=\; \frac{K_*^{-1} B_{g1*}^{\otimes 2} (K_*^{-1})^\top}{(B_{11*}-1)^2} (B_{11*} + \alpha_{0*}^{-1} - 2).$$

Under condition 5, $\max_{1\le i\le N} \tilde{\pi}_i < 1$ when N is large enough. This implies that $B_{11*} > 1$. Therefore $B_{11*} + \alpha_{0*}^{-1} - 2 > 0$ and $\Sigma_{\text{ELW}} \le \Sigma_{\text{SIPW}}$.

7.3 Proof of result (c) of Theorem 5.2

In the case of $g(z,\theta) = f(z) - \theta$, similar to the proof of result (a) of Theorem 5.2, we can show that $\sqrt{N}(\hat{\theta}_{\text{IPW}} - \theta_0) \xrightarrow{d} N(0, \Sigma_{\text{IPW}})$ with $\Sigma_{\text{IPW}} = B_{ff*} - \theta_{0*}^{\otimes 2} \alpha_{0*}^{-1}$.

In the meanwhile, when $g(z,\theta) = f(z) - \theta$, we have $K_*^{-1} = I_r$ and

$$\Sigma_{\text{ELW}} \;=\; B_{ff*} - B_{f1*}\theta_{0*}^\top - \theta_{0*} B_{f1*}^\top + \theta_{0*}^{\otimes 2} B_{11} - \frac{(B_{f1*} - \theta_{0*} B_{11})^{\otimes 2}}{(B_{11*}-1)^2} (B_{11*} + \alpha_{0*}^{-1} - 2).$$

For efficiency comparison, we may write

$$\Sigma_{\text{ELW}} \;=\; B_{gg*} - \frac{B_{g1*}^{\otimes 2}}{(B_{11*}-1)^2} (B_{11*} + \alpha_{0*}^{-1} - 2).$$

and

$$\Sigma_{\text{IPW}} = B_{gg*} + B_{g1*}\theta_{0*}^\top + \theta_{0*} B_{g1*}^\top + \theta_{0*}^{\otimes 2} (B_{11*} - \alpha_{0*}^{-1}).$$

Therefore

$$\begin{aligned}
\Sigma_{\text{IPW}} - \Sigma_{\text{ELW}} \;&=\; B_{g1*}^{\otimes 2} \frac{B_{11*} + \alpha_{0*}^{-1} - 2}{(B_{11*}-1)^2} + B_{g1*}\theta_{0*}^\top + \theta_{0*} B_{g1*}^\top + \theta_{0*}^{\otimes 2} (B_{11} - \alpha_{0*}^{-1}) \\
&=\; \left(B_{g1*} \frac{\sqrt{B_{11*} + \alpha_{0*}^{-1} - 2}}{B_{11*} - 1} + \theta_{0*} \frac{B_{11*} - 1}{\sqrt{B_{11*} + \alpha_{0*}^{-1} - 2}} \right)^{\otimes 2} \\
&\quad - \theta_{0*}^{\otimes 2} \frac{(\alpha_{0*}^{-1} - 1)^2}{B_{11*} + \alpha_{0*}^{-1} - 2}.
\end{aligned}$$

When $B_{g1*} = c\theta_{0*}$ for some $c > 0$, we have

$$\begin{aligned}
\Sigma_{\text{IPW}} - \Sigma_{\text{ELW}} \;&=\; \left\{ \left(c \frac{\sqrt{B_{11*} + \alpha_{0*}^{-1} - 2}}{B_{11*} - 1} + \frac{B_{11*} - 1}{\sqrt{B_{11*} + \alpha_{0*}^{-1} - 2}} \right)^2 - \frac{(\alpha_{0*}^{-1} - 1)^2}{B_{11*} + \alpha_{0*}^{-1} - 2} \right\} \theta_{0*}^{\otimes 2} \\
&\ge\; \left\{ \frac{(B_{11*} - 1)^2}{B_{11*} + \alpha_{0*}^{-1} - 2} - \frac{(\alpha_{0*}^{-1} - 1)^2}{B_{11*} + \alpha_{0*}^{-1} - 2} \right\} \theta_{0*}^{\otimes 2},
\end{aligned}$$

where for two symmetric matrices A_1 and A_2, $A_1 \ge A_2$ means $A_2 - A_1$ is nonnegative definite. By Jensen's inequality, we have

$$B_{11} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\pi_k} \ge \frac{1}{\frac{1}{N}\sum_{i=1}^{N} \pi_k} = \frac{1}{\alpha_0} > 1,$$

which implies $B_{11*} = \lim_{N\to\infty} \geq \alpha_{0*}^{-1} > 1$. Consequently, $\Sigma_{\rm IPW} \geq \Sigma_{\rm ELW}$ when $B_{g1*} = c\theta_{0*}$ for some $c > 0$.

However, when

$$B_{g1*} = -\frac{(B_{11*} - 1)^2}{B_{11*} + \alpha_{0*}^{-1} - 2}\theta_{0*},$$

we have

$$\Sigma_{\rm IPW} - \Sigma_{\rm ELW} \quad = \quad -\frac{(\alpha_{0*}^{-1} - 1)^2}{B_{11*} + \alpha_{0*}^{-1} - 2} \cdot \theta_{0*}^{\otimes 2},$$

which is nonpositive definite. In summary, we have no affirmative conclusion about the efficiency comparison between the IPW estimator and the ELW estimator.

8 Additional simulation results

8.1 Additional simulation results related to Example 1

Regarding Example 1 of the main paper, we consider the scenarios with $\gamma = 1.3, 1.5, 1.9, 2.5$, and $N = 50, 500$, 2000. Tables 1–3 present the simulated RMSE results for all the estimation methods under comparison with sample size being 50, 500, and 2000, respectively. In all cases, the ELW estimator either has the smallest RMSE or its RMSE is very close to the smallest of the seven estimators. This observations are confirmed by the corresponding boxplots, which are displayed in Figures 1-3. The advantage of ELW over the rest six estimators is remarkably obvious in the scenarios with $c = 0.1$ under models 3 and 4.

The coverage probabilities and average lengths of the interval estimators under comparison with $N = 2000$, and $\gamma = 1.3$ and 1.9 are reported in Table 5. When $\gamma = 1.3$, all intervals have severe under-coverages in almost all cases. Even so, the ELW-re has the best overall performance in terms of coverage probability, whereas MW1-re and MW2-re have unstable performance under different settings. Except SIPW-re, which is comparable with ELW-re, the rest intervals usually have much undercoverages. When γ increases to 1.9, the IPW-re, SIPW-re, and ELW-re intervals are desirable, however MW1-re and MW2-re still have much under-coverages.

8.2 Simulation results when data were generated from Example 2

Example 2 (Unequal probability sampling). *The finite population consists of $N = 3000$ observations $\{(x_i, y_i) : 1 \leq i \leq N\}$. We generate x_i from the uniform distribution on $[0, 2]$, and $y_i = \mu(x_i) + \sqrt{3(1-\rho^2)}e_i$, where $e_i \sim N(0, 1)$ are independent of each other and of x_i. We consider four models for $\mu(x)$, namely $\sqrt{3}\rho x$ (Model 1), $\sqrt{3}\rho(x + x^2)$ (Model 2), $\sqrt{3}\rho x + 5$ (Model 3), and $\sqrt{3}\rho(x + x^2) + 5$ (Model 4), and three unequal probability samplings, namely Poisson sampling, pivotal sampling, and PPS (probability proportional to pize) sampling. We set n to 500 and 250,* *take $\pi_i = nx_i / \sum_{j=1}^{N} x_j$ for Poisson sampling and pivotal sampling, and take x_i as the size for PPS sampling. The goal is to estimate $\theta = (1/N) \sum_{i=1}^{N} y_i$.*

Example 2 is a modified version of Example 2 from Zong et al. (2019) and is designed to represent unequal probability sampling. Poisson sampling and pivotal sampling are UPS-WORs, with the sample size being random in the former and fixed in the latter. PPS sampling is a popular UPS-WR. In particular, pivotal sampling satisfies the LIND condition (Patterson et al., 2001) required by Theorem 5.1.

We exclude CHIM, MW1, and MW2 from the subsequent comparison because they are all designed not for unequal probability sampling but for missing data problems. Tables 6 and Table 8 present the simulated RMSEs for the IPW, SIPW, ZZZ, and ELW estimators.

ELW always has the smallest RMSEs under Poisson sampling. It still perform the best under both pivotal and PPS samplings, except for the scenarios with $\rho = 0.8$ for Models 1 and 2. When the RMSE of ELW is not the smallest, the boxplots of ZZZ and ELW are close to each other. In the remaining scenarios, the competitors of ELW have either much larger variances or much larger biases than ELW. Similar to Example 1 in the main paper, Models 3 and 4 in Example 2 are also Models 1 and 2 with a mean shift. Again, the performance of ELW and SIPW is equivariant under such a mean shift; however, the IPW and ZZZ estimators are very sensitive, and their performances become much worse when the mean of y_i moves away from the origin. The performances of the four estimators seem to be insensitive to the choice of ρ.

For interval estimation under unequal probability sampling, no valid resampling-based confidence intervals are available as far as we know. Therefore, we compare only IPW-an, SIPW-a, and ELW-an, which are the confidence intervals based on the asymptotic normalities of the IPW, SIPW, and ELW estimators. Tables 9 and 7 present their simulated coverage probabilities and average lengths. Under both Poisson and pivotal samplings, overall ELW-an has not only more accurate coverage probabilities but also shorter average lengths than SIPW-an. Under PPS sampling, when $\rho = 0.8$, ELW-an still wins, although both ELW-an and SIPW-an have undercoverages of 3% or more. However, when $\rho = 0.2$, SIPW-an has better and more desirable coverage probabilities. This implies that ELW-an still has room for improvement.

In summary, the asymptotic-normality-based ELW interval estimator has desirable performance under UPW-WORs such as Poisson and pivotal samplings, and acceptable performance under UPS-WRs such as PPS sampling.

8.3 Simulation results when data were generated from Example 3

In our simulation study regarding Example 1, we directly generated propensity scores when generating data. The actual not the estimated propensity scores are then used in the parameter estimation by all the seven methods under comparison. We may wonder how the proposed ELW method performs compared with the other methods when the estimated propensity scores are used. To this end, we conduct simulations by generating data from Example 3.

Example 3 (Missing data with estimated propensity scores)**.** *Let X follow the exponential distribution with rate $\lambda = 1$ or $1/2$. Given $X = x$, we generate D from the Bernoulli distribution with success probability $\pi(x) = e^{-x}/(1 + e^{-x})$, and generate Y from $Y = \mu(x) + \varepsilon \times \sigma$, where $\varepsilon \sim N(0, 1)$ and $\sigma = 1$ or $1/4$. Four choices of $\mu(x)$ are considered: $\mu(x) = \log(1 + x)$ (Model 1), $\mu(x) = \{(x - 1)^2 + (x - 3)^2\}/4$ (Model 2), $\mu(x) = 5 + \log(1 + x)$ (Model 3), and $\mu(x) = 5 + \{(x - 1)^2 + (x - 3)^2\}/4$ (Model 4). The parameter of interest is $\theta = \mathbb{E}(Y)$.*

When generating data from Example 3, we set the full data size to $N = 500$ or 2000. Let $\{(X_i, D_i, D_iY_i) : 1 \le i \le N\}$ be a generic sample we generated. When analyzing data, we assume that Y_i are missing at random, and we postulate a linear logistic model for the propensity score, namely

$$P(D = 1|X, Y) = P(D = 1|X) = \exp(\beta_1 + \beta_2 x)/\{1 + \exp(\beta_1 + \beta_2 x)\}.$$

Based on $\{(X_i, R_i) : 1 \le i \le N\}$, we estimate $\beta = (\beta_1, \beta_2)$ by its maximum likelihood estimate, and calculate the estimated propensity scores

$$\pi(X_i, \hat{\beta}) = \exp(\hat{\beta}_1 + \hat{\beta}_2 X_i)/\{1 + \exp(\hat{\beta}_1 + \hat{\beta}_2 X_i)\}.$$

Then we take them as if they were the true propensity scores and apply the estimation methods under comparison to estimate θ. We tabulate the RMSEs of the seven estimation methods under comparison based on 5000 samples in Table 10. We see that even when the propensity scores are estimated, the ELW method still has an obvious advantage over the IPW method and its variants. Out of the 32 simulation scenarios, the ELW method has the smallest RMSEs among the seven methods in 26 scenarios. In the rest 6 scenarios, the RMSEs of the ELW method are nearly the closet to the smallest RMSEs.

Table 1: Simulated RMSEs of the seven estimators under comparison when the full data size $N = 50$ and data are generated from Example 1.

γ	c	Model	IPW	SIPW	ZZZ	CHIM	MW1	MW2	ELW
1.3	1.0	1	18.55	4.58	3.83	4.61	28.81	32.82	4.82
1.3	1.0	2	16.18	3.47	3.81	3.47	19.92	22.86	3.80
1.3	1.0	3	167.94	4.62	18.69	4.62	9.83	7.96	4.94
1.3	1.0	4	76.39	3.50	18.59	3.50	21.22	23.34	3.89
1.3	0.1	1	41.04	3.34	2.96	3.35	2.80	3.03	3.15
1.3	0.1	2	13.52	1.43	2.95	1.43	3.22	3.58	0.88
1.3	0.1	3	120.61	3.35	18.18	3.35	7.62	4.28	3.14
1.3	0.1	4	357.10	1.44	18.66	1.44	6.99	3.75	0.88
1.5	1.0	1	28.03	3.57	3.04	3.57	11.57	15.10	3.61
1.5	1.0	2	18.64	2.76	2.91	2.76	4.66	5.33	2.86
1.5	1.0	3	46.17	3.62	13.14	3.62	8.21	5.98	3.67
1.5	1.0	4	57.44	2.89	13.55	2.89	8.70	6.91	2.95
1.5	0.1	1	6.71	2.46	2.08	2.46	1.68	1.73	2.29
1.5	0.1	2	20.73	1.08	2.09	1.08	1.24	0.98	0.67
1.5	0.1	3	41.40	2.40	12.96	2.40	7.12	3.70	2.24
1.5	0.1	4	47.84	1.09	13.32	1.09	7.21	4.19	0.68
1.9	1.0	1	3.62	2.59	2.22	2.59	2.66	2.91	2.55
1.9	1.0	2	3.25	2.11	2.12	2.11	3.09	3.87	2.09
1.9	1.0	3	17.61	2.64	8.37	2.64	6.19	4.61	2.59
1.9	1.0	4	14.96	2.16	8.84	2.16	6.90	6.13	2.09
1.9	0.1	1	3.40	1.62	1.31	1.62	1.19	1.26	1.54
1.9	0.1	2	2.63	0.77	1.28	0.77	0.87	0.59	0.50
1.9	0.1	3	20.52	1.61	8.20	1.61	5.71	3.65	1.56
1.9	0.1	4	14.74	0.77	8.80	0.77	6.24	4.26	0.50
2.5	1.0	1	2.12	1.91	1.76	1.91	2.03	2.26	1.88
2.5	1.0	2	1.98	1.64	1.65	1.64	1.92	2.24	1.63
2.5	1.0	3	7.54	1.91	5.57	1.91	4.61	3.84	1.91
2.5	1.0	4	8.16	1.64	6.30	1.64	5.19	4.42	1.61
2.5	0.1	1	1.28	1.11	0.99	1.11	1.00	1.08	1.10
2.5	0.1	2	1.18	0.56	0.81	0.56	2.09	2.91	0.40
2.5	0.1	3	7.28	1.10	5.40	1.10	4.27	3.23	1.09
2.5	0.1	4	7.69	0.56	6.08	0.56	4.95	3.92	0.40

Table 2: Simulated RMSEs of the seven estimators under comparison when the full data size $N = 500$ and data are generated from Example 1.

γ	c	Model	IPW	SIPW	ZZZ	CHIM	MW1	MW2	ELW
1.3	1.0	1	35.05	7.60	7.16	7.60	5.63	5.38	7.02
1.3	1.0	2	16.66	6.23	7.24	6.23	5.38	4.83	6.66
1.3	1.0	3	287.99	7.86	35.45	7.83	16.38	8.22	7.03
1.3	1.0	4	187.03	6.09	35.34	6.08	16.40	8.05	6.77
1.3	0.1	1	20.06	5.02	5.76	5.02	3.08	2.69	1.67
1.3	0.1	2	17.46	2.14	5.82	2.14	2.54	1.06	0.86
1.3	0.1	3	164.15	5.06	34.81	5.06	15.37	6.60	1.67
1.3	0.1	4	102.58	2.18	35.37	2.18	15.26	6.64	0.88
1.5	1.0	1	9.95	5.30	4.53	5.30	4.08	3.79	4.31
1.5	1.0	2	33.06	4.57	4.52	4.57	3.96	3.47	4.21
1.5	1.0	3	60.29	5.58	20.69	5.58	13.56	7.41	4.51
1.5	1.0	4	80.85	4.76	21.27	4.75	13.82	7.91	4.17
1.5	0.1	1	11.41	3.68	3.43	3.68	2.22	1.81	1.64
1.5	0.1	2	20.22	1.64	3.39	1.64	2.08	1.04	0.66
1.5	0.1	3	81.50	3.68	20.63	3.68	12.91	6.77	1.62
1.5	0.1	4	69.71	1.64	20.76	1.64	13.35	7.15	0.67
1.9	1.0	1	4.66	3.51	2.72	3.51	2.67	2.47	2.81
1.9	1.0	2	6.34	2.84	2.65	2.84	2.57	2.39	2.36
1.9	1.0	3	17.44	3.37	10.89	3.37	9.03	6.12	2.80
1.9	1.0	4	25.61	2.73	11.24	2.73	9.40	6.75	2.42
1.9	0.1	1	3.00	2.26	1.71	2.26	1.43	1.14	1.46
1.9	0.1	2	5.23	1.05	1.71	1.05	1.35	0.89	0.52
1.9	0.1	3	24.14	2.27	10.80	2.27	8.69	5.76	1.48
1.9	0.1	4	19.14	1.05	10.96	1.05	9.12	6.45	0.52
2.5	1.0	1	2.13	2.08	1.89	2.08	1.90	1.89	1.99
2.5	1.0	2	2.00	1.74	1.80	1.74	1.81	1.75	1.66
2.5	1.0	3	7.34	2.03	6.31	2.03	5.75	4.76	1.97
2.5	1.0	4	8.06	1.82	6.90	1.82	6.38	5.46	1.67
2.5	0.1	1	1.29	1.26	1.09	1.26	1.02	0.92	1.16
2.5	0.1	2	1.19	0.65	0.93	0.65	0.84	0.66	0.41
2.5	0.1	3	7.38	1.27	6.04	1.27	5.48	4.40	1.17
2.5	0.1	4	7.72	0.63	6.76	0.63	6.23	5.21	0.41

Table 3: Simulated RMSEs of the seven estimators under comparison when the full data size $N = 2000$ and data are generated from Example 1.

γ	c	Model	IPW	SIPW	ZZZ	CHIM	MW1	MW2	ELW
1.3	1.0	1	64.24	11.64	10.67	11.62	7.17	6.55	10.14
1.3	1.0	2	85.88	9.65	10.53	9.65	6.81	5.30	9.61
1.3	1.0	3	551.35	11.77	51.40	11.23	24.07	10.77	9.85
1.3	1.0	4	360.02	9.80	51.91	9.44	24.23	10.34	9.80
1.3	0.1	1	26.76	6.87	8.51	6.84	4.51	4.14	1.67
1.3	0.1	2	192.71	2.98	8.48	2.96	3.86	1.49	1.10
1.3	0.1	3	468.48	6.98	51.37	6.94	23.46	9.47	1.80
1.3	0.1	4	502.15	3.00	52.02	2.98	24.17	9.18	1.29
1.5	1.0	1	24.72	8.05	6.05	8.00	4.96	4.35	5.51
1.5	1.0	2	17.89	6.17	5.95	6.17	4.84	3.78	5.13
1.5	1.0	3	69.08	7.49	27.27	7.49	18.29	9.59	5.21
1.5	1.0	4	110.80	6.49	27.11	6.49	18.31	9.83	5.21
1.5	0.1	1	14.76	4.89	4.48	4.87	3.00	2.52	1.60
1.5	0.1	2	26.23	2.16	4.44	2.15	2.88	1.36	0.71
1.5	0.1	3	68.12	4.74	27.04	4.73	17.81	8.94	1.61
1.5	0.1	4	140.05	2.21	26.86	2.19	18.02	9.03	0.74
1.9	1.0	1	6.96	4.21	2.98	4.21	2.89	2.73	2.92
1.9	1.0	2	5.97	3.14	2.98	3.14	2.83	2.46	2.59
1.9	1.0	3	17.61	3.85	12.49	3.85	10.60	7.38	2.98
1.9	1.0	4	19.28	3.02	12.86	3.02	11.10	8.12	2.61
1.9	0.1	1	9.63	2.68	2.05	2.59	1.70	1.35	1.47
1.9	0.1	2	3.32	1.29	1.96	1.29	1.65	1.09	0.52
1.9	0.1	3	19.43	2.62	12.16	2.62	10.36	7.08	1.45
1.9	0.1	4	26.20	1.27	12.65	1.27	10.81	7.70	0.54
2.5	1.0	1	2.11	2.11	1.97	2.11	1.93	1.87	2.02
2.5	1.0	2	2.06	1.81	1.90	1.81	1.89	1.82	1.72
2.5	1.0	3	7.64	2.15	6.77	2.15	6.33	5.32	2.05
2.5	1.0	4	8.14	1.85	7.31	1.85	6.87	6.01	1.70
2.5	0.1	1	1.49	1.33	1.14	1.33	1.07	0.98	1.17
2.5	0.1	2	1.22	0.69	1.01	0.69	0.93	0.77	0.42
2.5	0.1	3	7.63	1.31	6.60	1.31	6.21	5.14	1.18
2.5	0.1	4	8.26	0.68	7.13	0.68	6.80	5.85	0.42

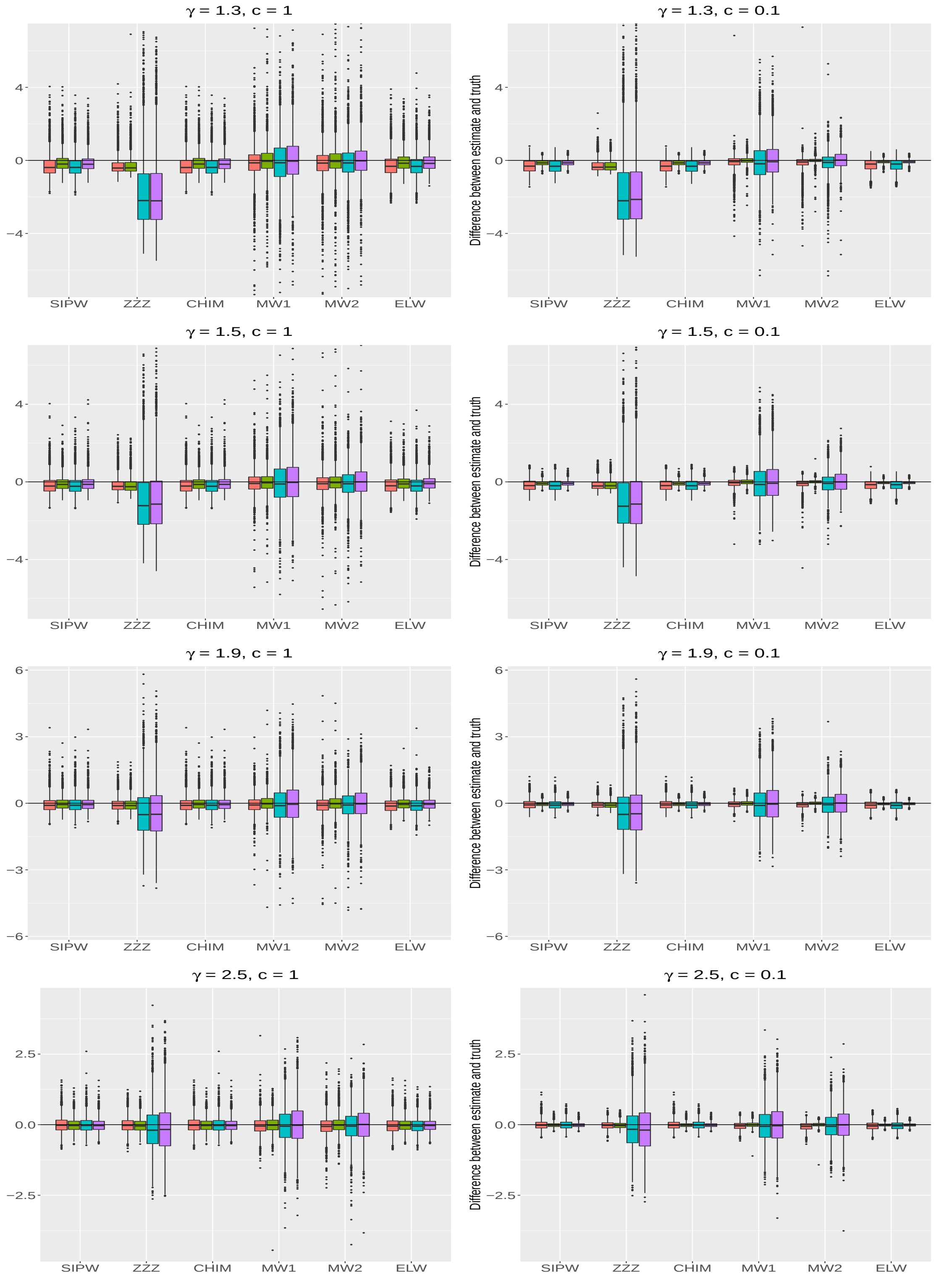

Figure 1: Boxplots of the SIPW, ZZZ, CHIM, MW1, MW2, and ELW estimators (minus the true parameter values) when data were generated from Example 1 with $N = 50$. For each scenario and each method, the four boxplots from left to right and in red, green, blue and purple correspond to models 1-4, respectively.

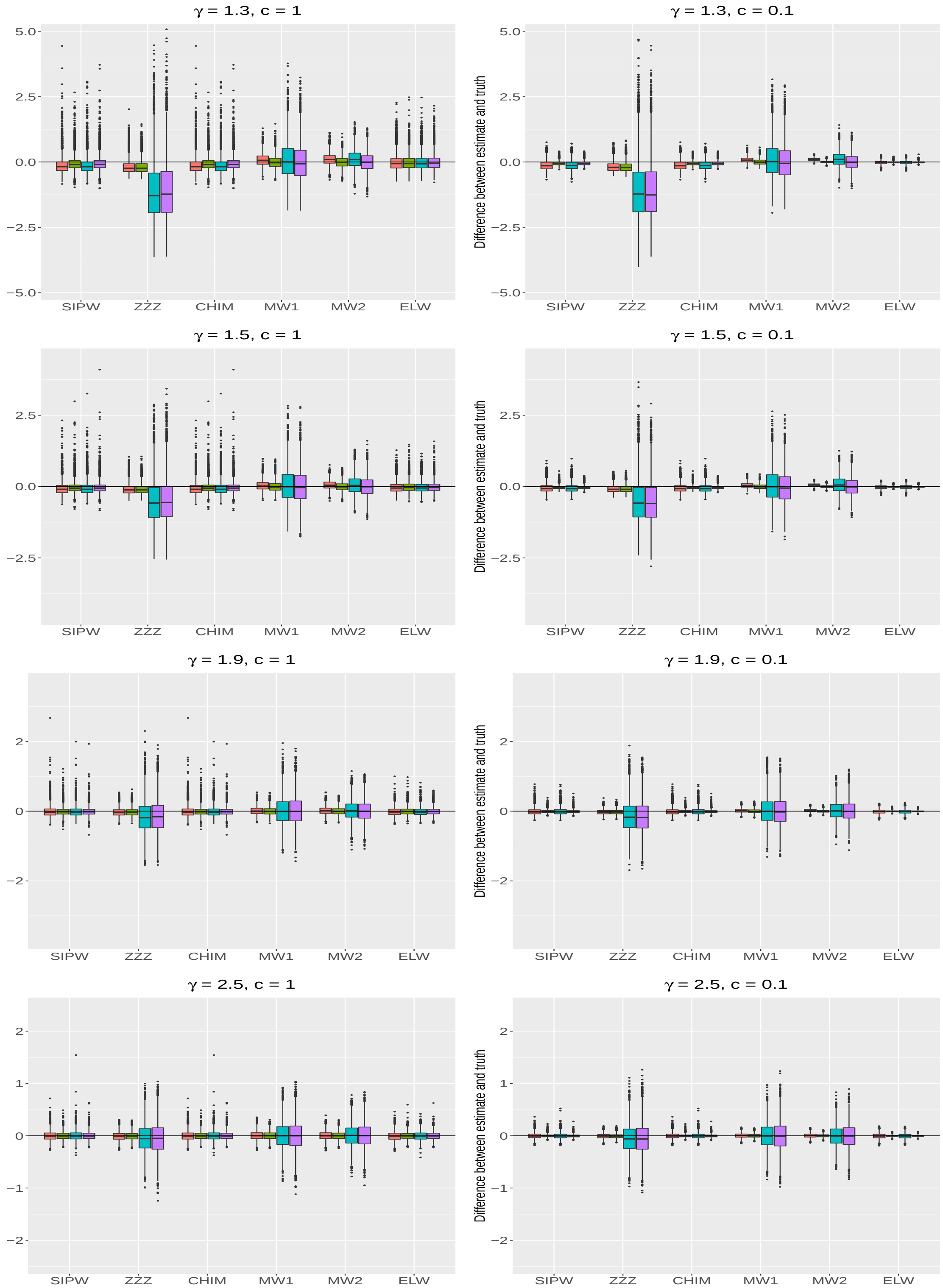

Figure 2: Boxplots of the SIPW, ZZZ, CHIM, MW1, MW2, and ELW estimators (minus the true parameter values) when data were generated from Example 1 with $N = 500$. For each scenario and each method, the four boxplots from left to right and in red, green, blue and purple correspond to models 1-4, respectively.

Table 4: Simulated coverage probabilities (%) of the interval estimators under comparison when data were generated from Example 1 with full data size $N = 2000$ with $\gamma = 1.5$ or 1.5 and $c = 1.0$. The numbers in parentheses are average lengths.

γ	Model	IPW-an	IPW-re	SIPW-an	SIPW-re	MW1-re	MW2-re	ELW-an	ELW-re
1.5	1	76.14 (0.556)	84.52 (3.029)	78.42 (0.436)	88.86 (1.455)	89.44 (1.060)	92.86 (0.780)	82.58 (0.329)	91.48 (1.044)
1.5	2	78.24 (1.593)	87.24 (6397.733)	85.08 (0.366)	91.14 (1.159)	89.68 (1.165)	94.52 (0.831)	81.62 (0.312)	91.04 (1.047)
1.5	3	76.76 (2.948)	86.40 (18.447)	77.84 (0.446)	88.94 (1.575)	82.74 (3.325)	81.44 (1.507)	82.02 (0.333)	90.94 (1.061)
1.5	4	78.74 (3.282)	86.32 (121.859)	85.36 (0.370)	91.60 (1.160)	82.32 (3.276)	83.32 (1.551)	82.54 (0.312)	92.38 (1.046)
2.5	1	94.02 (0.178)	91.40 (0.341)	93.60 (0.177)	92.46 (0.351)	91.92 (0.323)	93.14 (0.323)	93.32 (0.169)	93.20 (0.343)
2.5	2	93.52 (0.169)	93.48 (0.346)	93.72 (0.150)	93.54 (0.303)	92.38 (0.324)	93.40 (0.321)	93.72 (0.141)	93.60 (0.287)
2.5	3	94.14 (0.630)	93.68 (1.261)	93.86 (0.177)	92.74 (0.348)	89.40 (1.029)	88.16 (0.837)	93.56 (0.169)	93.40 (0.344)
2.5	4	94.26 (0.682)	93.34 (1.346)	93.80 (0.150)	94.06 (0.299)	90.24 (1.118)	88.96 (0.948)	93.46 (0.141)	93.94 (0.286)

Table 5: Simulated coverage probabilities (%) when the full data size $N = 2000$ and data are generated from Example 1 with $\gamma = 1.3$ or 1.9. The numbers in parentheses are average lengths.

γ	c	Model	IPW-an	IPW-re	SIPW-an	SIPW-re	MW1-re	MW2-re	ELW-an	ELW-re
1.3	1.0	1	60.64 (1.117)	78.06 (110.757)	65.26 (0.559)	85.20 (2.565)	90.28 (2.009)	89.20 (1.186)	69.68 (0.448)	89.12 (2.152)
1.3	1.0	2	59.68 (0.979)	78.70 (29.428)	76.16 (0.484)	88.30 (2.022)	92.34 (2.124)	94.90 (1.264)	67.86 (0.436)	88.96 (2.228)
1.3	1.0	3	60.82 (7.225)	78.88 (5683.808)	67.84 (0.558)	86.44 (2.526)	77.70 (4.806)	80.60 (1.664)	70.14 (0.453)	88.92 (2.183)
1.3	1.0	4	60.16 (6.386)	77.20 (840.508)	75.94 (0.479)	88.90 (1.951)	77.10 (4.673)	82.76 (1.668)	68.14 (0.438)	89.02 (2.251)
1.3	0.1	1	60.58 (0.983)	82.76 (195.461)	64.90 (0.353)	86.94 (1.453)	74.00 (0.771)	33.14 (0.276)	90.06 (0.120)	93.78 (0.298)
1.3	0.1	2	59.08 (0.826)	80.20 (24.051)	61.04 (0.146)	84.48 (0.607)	76.04 (0.798)	78.84 (0.227)	75.22 (0.056)	85.90 (0.183)
1.3	0.1	3	60.24 (5.066)	79.32 (91.168)	63.32 (0.352)	87.66 (1.451)	74.66 (4.424)	72.54 (1.224)	85.80 (0.119)	90.12 (0.298)
1.3	0.1	4	60.94 (5.712)	77.58 (216.103)	60.78 (0.148)	84.36 (0.616)	75.70 (4.317)	76.06 (1.274)	68.26 (0.057)	77.74 (0.186)
1.9	1.0	1	89.62 (0.289)	90.56 (0.950)	89.48 (0.272)	92.48 (0.667)	89.74 (0.517)	93.04 (0.467)	90.96 (0.225)	93.86 (0.525)
1.9	1.0	2	91.24 (0.289)	93.82 (0.777)	92.98 (0.226)	94.70 (0.536)	90.64 (0.554)	93.12 (0.490)	91.84 (0.197)	93.76 (0.467)
1.9	1.0	3	90.24 (1.287)	92.88 (3.695)	91.40 (0.277)	93.74 (0.669)	86.40 (1.823)	84.48 (1.173)	91.98 (0.226)	94.48 (0.529)
1.9	1.0	4	90.28 (1.301)	91.82 (3.372)	91.38 (0.225)	93.06 (0.524)	86.88 (1.851)	85.16 (1.260)	90.64 (0.197)	92.80 (0.469)
1.9	0.1	1	88.76 (0.203)	91.16 (0.565)	89.10 (0.189)	93.40 (0.488)	86.30 (0.287)	87.24 (0.201)	92.46 (0.120)	95.20 (0.269)
1.9	0.1	2	89.02 (0.203)	93.34 (0.668)	88.96 (0.089)	93.38 (0.235)	85.58 (0.295)	85.92 (0.181)	92.44 (0.042)	93.84 (0.090)
1.9	0.1	3	89.86 (1.234)	92.64 (3.398)	89.12 (0.188)	93.48 (0.482)	85.24 (1.760)	82.12 (1.080)	92.66 (0.120)	95.18 (0.270)
1.9	0.1	4	90.24 (1.286)	91.76 (3.160)	90.10 (0.089)	94.24 (0.234)	86.30 (1.788)	85.26 (1.177)	93.28 (0.042)	93.42 (0.091)

Table 6: Simulated RMSEs of the IPW, SIPW, ZZZ, and ELW estimators where data were generated from Example 2 with $n = 500$.

		IPW	SIPW	ZZZ	ELW	IPW	SIPW	ZZZ	ELW	IPW	SIPW	ZZZ	ELW
ρ	Model	Poisson sampling				Pivotal sampling				PPS sampling			
0.2	1	9.04	4.34	6.28	**3.93**	40.67	4.38	5.03	**3.78**	7.41	4.39	5.11	**3.87**
0.2	2	11.38	5.36	6.95	**4.16**	40.64	5.36	4.89	**3.92**	7.28	5.31	4.94	**4.03**
0.2	3	36.21	4.59	23.56	**3.89**	179.21	4.38	17.48	**3.78**	27.46	4.39	18.01	**3.87**
0.2	4	36.55	5.36	24.06	**4.16**	179.18	5.36	17.27	**3.92**	27.29	5.31	17.76	**4.03**
0.8	1	8.07	5.66	5.85	**3.39**	24.90	5.88	3.07	**2.91**	4.53	5.74	3.13	**3.00**
0.8	2	10.79	12.84	9.24	**6.56**	24.84	13.29	**2.79**	5.03	4.27	13.14	**2.79**	5.28
0.8	3	33.58	5.66	23.09	**3.39**	163.52	5.88	15.86	**2.91**	24.97	5.74	16.36	**3.00**
0.8	4	35.35	12.84	25.53	**6.56**	163.41	13.29	15.03	**5.03**	24.32	13.14	15.40	**5.28**

Table 7: Coverage probabilities and average length of Wald intervals based on the IPW, SIPW, and ELW estimators when data were generated from Example 2 when $n = 500$.

		Coverage probability			Average length		
ρ	Model	IPW-an	SIPW-an	ELW-an	IPW-an	SIPW-an	ELW-an
		Poisson sampling					
0.2	1	93.28	95.34	93.86	0.476	0.280	0.251
0.2	2	94.26	93.18	93.52	0.507	0.333	0.275
0.2	3	93.06	95.16	93.96	1.771	0.278	0.252
0.2	4	93.94	93.78	93.58	1.811	0.334	0.274
0.8	1	94.72	90.90	93.52	0.431	0.343	0.232
0.8	2	96.04	91.84	95.34	0.684	0.776	0.479
0.8	3	93.80	90.26	93.54	1.752	0.340	0.233
0.8	4	94.66	91.78	94.92	1.928	0.778	0.479
		Pivotal sampling					
0.2	1	96.82	94.34	92.84	0.494	0.270	0.247
0.2	2	98.66	92.52	93.68	0.528	0.313	0.262
0.2	3	97.86	94.84	93.40	1.789	0.267	0.246
0.2	4	98.52	92.48	93.56	1.828	0.311	0.263
0.8	1	99.98	89.40	94.90	0.438	0.317	0.214
0.8	2	100.00	91.46	97.48	0.698	0.729	0.436
0.8	3	98.98	88.80	94.46	1.734	0.310	0.214
0.8	4	99.88	89.78	97.10	1.936	0.724	0.435
		PPS sampling					
0.2	1	92.34	94.48	91.34	0.433	0.289	0.253
0.2	2	93.54	92.96	91.80	0.451	0.333	0.262
0.2	3	91.60	94.52	91.28	1.560	0.288	0.253
0.2	4	91.70	93.06	92.28	1.558	0.334	0.263
0.8	1	97.52	89.46	90.64	0.311	0.333	0.188
0.8	2	100.00	90.86	91.94	0.419	0.762	0.330
0.8	3	92.18	89.58	91.32	1.471	0.335	0.187
0.8	4	94.34	90.56	91.04	1.468	0.775	0.330

Table 8: Simulated RMSEs of the IPW, SIPW, ZZZ and ELW estimators with data generated from Example 2 when $n = 250$.

		IPW	SIPW	ZZZ	ELW	IPW	SIPW	ZZZ	ELW	IPW	SIPW	ZZZ	ELW
ρ	Model	Poisson sampling				Pivotal sampling				PPS sampling			
0.2	1	12.62	5.95	8.63	5.45	10.73	5.87	6.60	5.32	80.60	6.02	6.80	5.32
0.2	2	12.06	6.93	9.51	5.91	10.58	6.94	6.38	5.56	80.58	7.23	6.56	5.55
0.2	3	42.33	5.78	31.78	5.48	38.14	5.87	22.46	5.32	355.73	6.02	23.13	5.32
0.2	4	42.93	6.93	32.59	5.91	37.92	6.94	22.12	5.56	355.70	7.23	22.78	5.55
0.8	1	9.45	7.33	8.27	4.93	6.57	7.26	4.04	4.23	49.36	7.75	4.16	4.26
0.8	2	14.29	16.84	13.55	9.92	6.27	16.63	3.63	7.62	49.30	17.74	3.72	7.84
0.8	3	40.46	7.33	31.53	4.93	34.47	7.26	20.38	4.23	324.56	7.75	20.95	4.26
0.8	4	43.52	16.84	35.42	9.92	33.65	16.63	19.07	7.62	324.45	17.74	19.60	7.84

Table 9: Coverage probabilities and average length of Wald intervals based on the SIPW and ELW estimators with data generated from Example 2 when $n = 250$.

		Coverage probability			Average length		
ρ	Model	IPW-an	SIPW-an	ELW-an	IPW-an	SIPW-an	ELW-an
		Poisson sampling					
0.2	1	93.24	93.82	91.94	0.661	0.376	0.341
0.2	2	94.12	92.18	92.04	0.727	0.446	0.373
0.2	3	92.40	94.26	92.60	2.512	0.374	0.335
0.2	4	92.98	92.18	92.48	2.572	0.453	0.374
0.8	1	95.50	88.50	91.04	0.622	0.449	0.320
0.8	2	95.58	89.84	93.66	1.006	1.039	0.680
0.8	3	93.06	88.68	90.98	2.459	0.449	0.320
0.8	4	94.32	91.12	93.50	2.724	1.025	0.679
		Pivotal sampling					
0.2	1	97.38	94.54	93.14	0.671	0.365	0.339
0.2	2	98.98	92.54	93.26	0.746	0.422	0.365
0.2	3	98.10	94.56	92.96	2.767	0.370	0.342
0.2	4	98.90	92.82	93.18	2.525	0.421	0.364
0.8	1	100.00	89.24	94.06	0.634	0.414	0.301
0.8	2	100.00	89.96	96.54	1.004	0.975	0.642
0.8	3	99.26	89.38	94.40	2.484	0.416	0.302
0.8	4	100.00	89.44	96.84	2.726	0.972	0.641
		PPS sampling					
0.2	1	89.98	94.46	91.66	0.569	0.385	0.339
0.2	2	90.98	91.60	90.34	0.557	0.449	0.353
0.2	3	88.38	94.38	91.64	1.976	0.382	0.338
0.2	4	88.30	92.44	91.14	1.969	0.447	0.354
0.8	1	94.46	88.38	89.50	0.382	0.436	0.260
0.8	2	99.72	89.48	90.58	0.464	1.021	0.486
0.8	3	90.26	88.34	88.88	1.930	0.438	0.260
0.8	4	90.02	89.32	90.60	1.839	1.021	0.485

Table 10: Simulated RMSEs of the seven estimators under comparison when data are generated from Example 3.

N	λ	σ	Model	IPW	SIPW	ZZZ	CHIM	MW1	MW2	ELW
500	1.0	1.00	1	4.00	2.93	2.57	2.93	2.39	2.36	2.45
500	1.0	1.00	2	23.55	7.77	3.50	7.77	4.04	4.34	3.50
500	1.0	1.00	3	11.69	2.89	5.70	2.89	4.15	3.91	2.39
500	1.0	1.00	4	19.25	6.23	6.84	6.23	4.96	5.10	3.55
500	1.0	0.25	1	3.42	1.63	1.41	1.63	1.03	0.93	0.95
500	1.0	0.25	2	42.84	7.72	2.91	7.72	3.49	3.83	2.86
500	1.0	0.25	3	13.75	1.68	5.21	1.68	3.59	3.30	0.97
500	1.0	0.25	4	16.16	4.35	6.35	4.35	4.36	4.39	2.53
500	0.5	1.00	1	24.60	5.62	5.86	5.62	4.01	3.81	4.55
500	0.5	1.00	2	164.77	32.92	28.84	32.92	32.29	33.50	26.46
500	0.5	1.00	3	60.77	5.74	17.71	5.74	7.56	5.34	4.54
500	0.5	1.00	4	378.83	34.80	40.33	34.80	33.07	33.59	26.60
500	0.5	0.25	1	13.15	3.94	4.92	3.94	2.39	2.07	2.19
500	0.5	0.25	2	106.84	30.86	28.78	30.86	32.32	33.57	26.21
500	0.5	0.25	3	111.43	3.81	17.54	3.78	6.80	4.39	2.19
500	0.5	0.25	4	134.39	30.71	40.49	30.71	32.23	33.16	26.06
2000	1.0	1.00	1	4.94	3.42	2.96	3.42	2.63	2.51	2.67
2000	1.0	1.00	2	14.68	9.59	4.93	9.59	5.76	6.84	4.72
2000	1.0	1.00	3	17.44	3.68	6.64	3.68	4.70	4.00	2.59
2000	1.0	1.00	4	63.04	16.74	8.96	16.74	6.88	7.16	4.89
2000	1.0	0.25	1	5.25	2.48	1.77	2.48	1.27	1.08	1.07
2000	1.0	0.25	2	18.69	10.48	4.34	10.48	5.21	6.37	4.11
2000	1.0	0.25	3	16.06	2.29	6.40	2.29	4.23	3.39	1.06
2000	1.0	0.25	4	27.11	10.50	8.76	10.50	6.51	6.82	4.07
2000	0.5	1.00	1	48.22	9.10	8.09	8.91	5.19	4.77	5.61
2000	0.5	1.00	2	426.28	70.79	49.41	67.50	56.49	62.57	43.49
2000	0.5	1.00	3	75.96	9.07	24.90	9.07	10.97	6.81	5.64
2000	0.5	1.00	4	234.17	65.41	65.43	65.41	57.97	62.84	43.61
2000	0.5	0.25	1	36.09	6.57	7.01	6.50	3.70	3.37	2.54
2000	0.5	0.25	2	687.50	74.54	49.69	65.31	56.79	62.92	43.67
2000	0.5	0.25	3	77.50	6.33	24.37	6.33	10.42	5.78	2.58
2000	0.5	0.25	4	1186.83	75.96	65.42	68.23	57.91	62.84	43.68

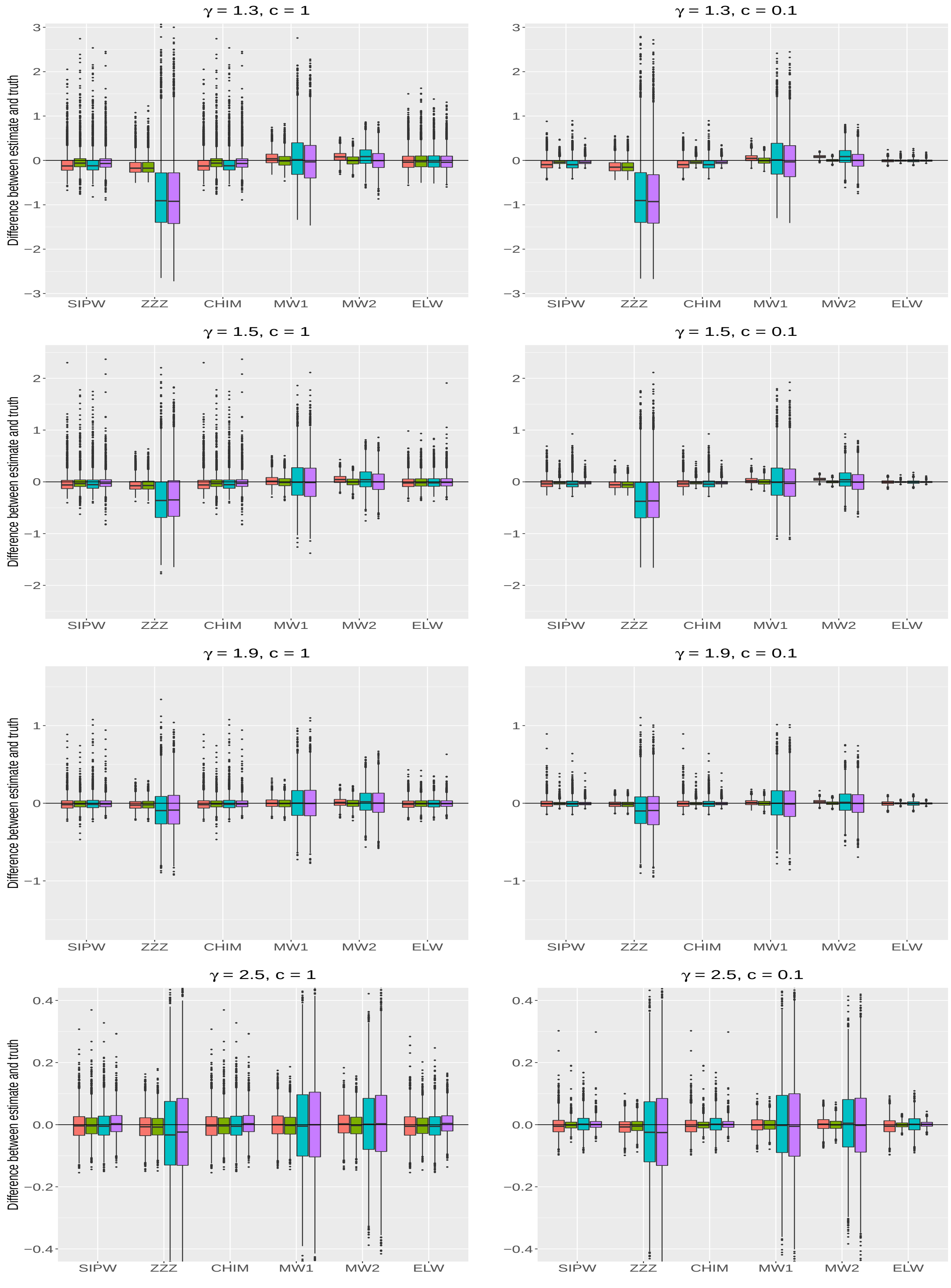

Figure 3: Boxplots of the SIPW, ZZZ, CHIM, MW1, MW2, and ELW estimators (minus the true parameter values) when data were generated from Example 1 with $N = 2000$. For each scenario and each method, the four boxplots from left to right and in red, green, blue and purple correspond to models 1-4, respectively.

9 ELW for over-identified estimating equations

In this section, we ignore the calculation issues and establish the limiting distributions of the ELW estimator and the ELW likelihood ratio statistic (LRT) under general over-identified estimating equations, which include just-identified estimating equations as special cases. We provide sketchy derivations and ignore rigorous regularity conditions that are needed. Suppose that the dimension s of the estimating function $g(z, \theta)$ is no less than the dimension r of θ. In this section, we assume that the estimating equations are correctly specified, namely their expectations are exactly equal to zero when the involved parameter takes its true value. When the estimating equations are mis-specified, namely their expectations are not equal to zero when the parameter takes its true value, the ELW estimator is inconsistent, as is the usual EL estimator. In this situation, the results in this section do not hold any longer.

9.1 Missing data problems with known propensity scores

We adopt the notation in Section 2 in the main paper. The empirical log-likelihood is

$$\tilde{\ell} = \sum_{i=1}^{N} [D_i \log(p_i) + D_i \log\{\pi(Z_i)\} + (1 - D_i) \log(1 - \alpha)],$$

and feasible p_i satisfy

$$p_i \geq 0, \quad \sum_{i=1}^{N} p_i = 1, \quad \sum_{i=1}^{N} p_i\{\pi(Z_i) - \alpha\} = 0, \quad \sum_{i=1}^{N} p_i g(Z_i, \theta) = 0.$$

After profiling out p_i's, we have the profile empirical log-likelihood function of (α, θ),

$$\ell(\alpha, \theta) = \sum_{i=1}^{N} [-D_i \log\{1 + \lambda_1(\pi(Z_i) - \alpha) + \lambda_2^\top g(Z_i, \theta)\} + (1 - D_i) \log(1 - \alpha)], \tag{33}$$

where we have ignored the terms not dependent of (α, θ), and (λ_1, λ_2) is the solution to

$$\sum_{i=1}^{N} \frac{D_i(\pi(Z_i) - \alpha)}{1 + \lambda_1(\pi(Z_i) - \alpha) + \lambda_2^\top g(Z_i, \theta)} = 0, \tag{34}$$

and

$$\sum_{i=1}^{N} \frac{D_i g(Z_i, \theta)}{1 + \lambda_1(\pi(Z_i) - \alpha) + \lambda_2^\top g(Z_i, \theta)} = 0. \tag{35}$$

We estimate (α, θ) by the maximum empirical likelihood estimator (MLE)

$$(\hat{\alpha}, \hat{\theta}) = \arg\max_{\alpha, \theta} \ell(\alpha, \theta), \tag{36}$$

and construct confidence regions or conduct hypothesis testing for θ using the likelihood ratio function

$$R(\theta) = 2\{\sup_{\alpha, \theta} \ell(\alpha, \theta) - \sup_{\alpha} \ell(\alpha, \theta)\}. \tag{37}$$

9.1.1 Asymptotic normality of the MLE

Let $\lambda_{10} = \alpha_0^{-1}, \lambda_{20} = 0_{s\times 1}$, and $(\hat{\lambda}_1, \hat{\lambda}_2)$ be the solution to equations (34) and (35) when $(\hat{\alpha}, \hat{\theta})$ is in place of (α, θ). Denote $\psi = (\alpha, \theta^\top, \lambda_1, \lambda_2^\top)^\top$, $\hat{\psi} = (\hat{\alpha}, \hat{\theta}^\top, \hat{\lambda}_1, \hat{\lambda}_2^\top)^\top$, and $\psi_0 = (\alpha_0, \theta_0^\top, \lambda_{10}, \lambda_{20}^\top)^\top$. Define

$$H(\psi) = (1/N)\sum_{i=1}^{N}[-D_i \log\{1 + \lambda_1(\pi(Z_i) - \alpha) + \lambda_2^\top g(Z_i, \theta)\} + (1 - D_i)\log(1 - \alpha)]. \tag{38}$$

Then $\hat{\psi}$ is a stationary point of $H(\psi)$ and satisfies $\nabla_\psi H(\hat{\psi}) = 0$, where we use $\nabla_\psi H$ to denote $\partial H/\partial\psi$. Under certain regularity conditions, using a proof similar to that of Lemma 1 of Qin and Lawless (1994), we have $\hat{\psi} = \psi_0 + O_p(N^{-1/2})$.

The first-order Taylor expansion of $0 = \nabla_\psi H(\hat{\psi})$ gives

$$0 = \nabla_\psi H(\psi_0) + \nabla_{\psi\psi^\top} H(\psi_0)(\hat{\psi} - \psi_0) + o_p(N^{-1/2}). \tag{39}$$

Define $u \equiv (u_a, u_b) = \nabla_\psi H(\psi_0)$. It can be found that $u_a = 0$ and

$$u_b = \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix} \frac{D_i}{\pi(Z_i)} - \frac{1-D_i}{1-\alpha_0} \\ \frac{D_i\alpha_0^2}{\pi(Z_i)} - D_i\alpha_0 \\ -\frac{\alpha_0 D_i g(Z_i,\theta_0)}{\pi(Z_i)} \end{pmatrix},$$

and $\nabla_{\psi\psi^\top} H(\psi_0) = \Upsilon + o_p(1)$, where

$$\Upsilon \equiv \begin{pmatrix} \Upsilon_{aa} & \Upsilon_{ab} \\ \Upsilon_{ba} & \Upsilon_{bb} \end{pmatrix} = \left(\begin{array}{c|ccc} 0 & 0 & 0 & -\alpha_0 K^\top \\ \hline 0 & B_{11} - (1-\alpha_0)^{-1} & \alpha_0^2 B_{11} & -\alpha_0 B_{g1}^\top \\ 0 & \alpha_0^2 B_{11} & \alpha_0^4 B_{11} - \alpha_0^3 & -\alpha_0^3 B_{g1}^\top \\ -\alpha_0 K & -\alpha_0 B_{g1} & -\alpha_0^3 B_{g1} & \alpha_0^2 B_{gg} \end{array}\right).$$

Suppose that Υ is invertible. It follows from (39) that $\hat{\psi} - \psi_0 = -\Upsilon^{-1}u + o_p(N^{-1/2})$. Since $\sqrt{N}u \xrightarrow{d} N(0, \Xi)$, where

$$\Xi \equiv \begin{pmatrix} \Xi_{aa} & \Xi_{ab} \\ \Xi_{ba} & \Xi_{bb} \end{pmatrix} = \left(\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \hline 0 & B_{11} + (1-\alpha_0)^{-1} & \alpha_0^2 B_{11} - \alpha_0 & -\alpha_0 B_{g1}^\top \\ 0 & \alpha_0^2 B_{11} - \alpha_0 & \alpha_0^4 B_{11} - \alpha_0^3 & -\alpha_0^3 B_{g1}^\top \\ 0 & -\alpha_0 B_{g1} & -\alpha_0^3 B_{g1} & \alpha_0^2 B_{gg} \end{array}\right),$$

we have $\sqrt{N}(\hat{\psi} - \psi_0) \xrightarrow{d} N(0, \Sigma)$ with

$$\Sigma = \Upsilon^{-1}\Xi\Upsilon^{-1} \equiv \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix},$$

from which we have $\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma_{aa})$, where

$$\Sigma_{aa} = \left[K^\top \left\{B_{gg} - \frac{B_{g1}B_{g1}^\top}{B_{11} - 1}\right\}^{-1} K\right]^{-1}. \tag{40}$$

In the case of just-identified estimating equations, K is a square matrix. If it is invertible, we have

$$\Sigma_{aa} = K^{-1}\left(B_{gg} - \frac{B_{g1}B_{g1}^\top}{B_{11}-1}\right)(K^{-1})^\top.$$

This is exactly equal to Σ_{ELW} in Section 2.3 in the main paper.

9.1.2 Limiting distribution of the LRT

Let $\tilde\alpha = \arg\max_\alpha \ell(\alpha,\theta_0)$ and $(\tilde\lambda_1,\tilde\lambda_2)$ be the solution to equations (34) and (35) when (α,θ) is replaced by $(\tilde\alpha,\theta_0)$. The LRT can be expressed as

$$R(\theta_0) = 2\{\ell(\hat\alpha,\hat\theta) - \ell(\tilde\alpha,\theta_0)\} = 2N\{H(\hat\psi) - H(\theta_0,\tilde\alpha,\tilde\lambda_1,\tilde\lambda_2)\}.$$

It can be found that

$$\begin{aligned} H(\hat\psi) &= H(\psi_0) - \frac{1}{2}u^\top\Upsilon^{-1}u + o_p(N^{-1}),\\ H(\theta_0,\tilde\alpha,\tilde\lambda_1,\tilde\lambda_2) &= H(\psi_0) - \frac{1}{2}u_b^\top\Upsilon_{bb}^{-1}u_b + o_p(N^{-1}). \end{aligned}$$

Write

$$\Upsilon^{-1} = \begin{pmatrix} \Upsilon^{aa} & \Upsilon^{ab}\\ \Upsilon^{ba} & \Upsilon^{bb}\end{pmatrix}.$$

Because $u=(u_a,u_b)$ with $u_a=0$, we have

$$R(\theta_0) \;=\; Nu_b^\top(\Upsilon_{bb}^{-1} - \Upsilon^{bb})u_b + o_p(1).$$

where $\Upsilon_{bb}^{-1} - \Upsilon^{bb} = \Upsilon_{bb}^{-1}\Upsilon_{ba}(\Upsilon_{ab}\Upsilon_{bb}^{-1}\Upsilon_{ba})^{-1}\Upsilon_{ab}\Upsilon_{bb}^{-1}$.

We have shown $\sqrt{N}u_b \xrightarrow{d} N(0,\Xi_{bb})$, $\Xi_{bb} = \Upsilon_{bb} + \tilde\Upsilon_{bb}$ with

$$\tilde\Upsilon^{bb} = \begin{pmatrix} 2(1-\alpha_0)^{-1} & -\alpha_0 & 0\\ -\alpha_0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}.$$

It can be verified that $\Upsilon_{ab}\Upsilon_{bb}^{-1}\tilde\Upsilon_{bb}\Upsilon_{bb}^{-1}\Upsilon_{ba}=0$, therefore

$$(\Upsilon_{bb}^{-1} - \Upsilon^{bb})\Xi_{bb}(\Upsilon_{bb}^{-1} - \Upsilon^{bb}) = (\Upsilon_{bb}^{-1} - \Upsilon^{bb}),$$

which means $\Xi_{bb}^{1/2}(\Upsilon_{bb}^{-1} - \Upsilon^{bb})\Xi_{bb}^{1/2}$ is idempotent. By Cochran's theorem, $R(\theta_0)\xrightarrow{d}\chi_r^2$ as $N\to\infty$.

9.2 Missing data problems with unknown propensity scores

We adopt the notation and assumptions in Section 2.3 in the main paper. The proof is similar to that in the case of known propensity score, and we only highlight their differences.

Define $\ell(\alpha,\theta)$, $(\hat{\alpha},\hat{\theta})$, $R(\theta)$ and $H(\psi)$ the same as in before after replacing $\pi(Z_i)$ by $\pi(\cdot,\hat{\beta})$. The u_b here is

$$u_b = \frac{1}{N}\sum_{i=1}^N \begin{pmatrix} \frac{D_i}{\pi(Z_i,\hat{\beta})} - \frac{1-D_i}{1-\alpha_0} \\ \frac{D_i\alpha_0^2}{\pi(Z_i,\hat{\beta})} - D_i\alpha_0 \\ -\frac{\alpha_0 D_i g(Z_i,\theta_0)}{\pi(Z_i,\hat{\beta})} \end{pmatrix},$$

which is different from that in the case of known propensity score. It can be approximated as

$$u_b = \frac{1}{N}\sum_{i=1}^N \begin{pmatrix} \frac{D_i}{\pi(Z_i)} - \frac{1-D_i}{1-\alpha_0} - B_{\dot{\pi}} h(D_i,Z_i) \\ \frac{D_i\alpha_0^2}{\pi(Z_i)} - D_i\alpha_0 - \alpha_0^2 B_{\dot{\pi}} h(D_i,Z_i) \\ -\frac{\alpha_0 D_i g(Z_i,\theta_0)}{\pi(Z_i)} + \alpha_0 B_{g\dot{\pi}} h(D_i,Z_i) \end{pmatrix} + o_p(N^{-1/2}).$$

This implies $\sqrt{N}u_b \xrightarrow{d} N(0,\Xi_{bb})$, where

$$\Xi_{bb} = \mathbb{V}\text{ar} \begin{pmatrix} \frac{D}{\pi(Z)} - \frac{1-D}{1-\alpha_0} - B_{\dot{\pi}} h(D,Z) \\ \frac{D\alpha_0^2}{\pi(Z)} - D\alpha_0 - \alpha_0^2 B_{\dot{\pi}} h(D,Z) \\ -\frac{\alpha_0 D g(Z,\theta_0)}{\pi(Z)} + \alpha_0 B_{g\dot{\pi}} h(D,Z) \end{pmatrix} \tag{41}$$

is different from that in the case of known propensity score. As the matrix Υ is the same as before, we have $\sqrt{N}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0,\Sigma_{aa})$, where $\Sigma_{aa} = \Upsilon^{ab}\Xi_{bb}\Upsilon^{ba}$ is different from the Σ_{aa} in the case of known propensity score.

The LRT $R(\theta_0)$ can still be approximated as

$$R(\theta_0) = N u_b^\top (\Upsilon_{bb}^{-1} - \Upsilon^{bb}) u_b + o_p(1).$$

However in general we have

$$(\Upsilon_{bb}^{-1} - \Upsilon^{bb})\Xi_{bb}(\Upsilon_{bb}^{-1} - \Upsilon^{bb}) \neq (\Upsilon_{bb}^{-1} - \Upsilon^{bb}),$$

which means that $\Xi_{bb}^{1/2}(\Upsilon_{bb}^{-1} - \Upsilon^{bb})\Xi_{bb}^{1/2}$ is not idempotent. Consequently in the case of unknown propensity score, asymptotically the LRT $R(\theta_0)$ follows a weighted chisquare distribution $\sum_{i=1}^r \rho_i Z_i^2$, where $\rho_1,\dots,\rho_r$ are the eigenvalues of $\Xi_{bb}^{1/2}(\Upsilon_{bb}^{-1} - \Upsilon^{bb})\Xi_{bb}^{1/2}$, and Z_i are independent and identically distributed random variables following $N(0,1)$.

9.3 UPS-WOR with non-negligible sampling fraction

We adopt the notation and assumptions in Section 3.1 in the main paper. For convenience, we do not distinguish the quantities for fixed N and their limits when $N\to\infty$. Define $\ell(\alpha,\theta)$, $(\hat{\alpha},\hat{\theta})$, $R(\theta)$ and $H(\psi)$ the same as in Section 9.1 after replacing $\pi(Z_i)$ by π. We have $u = (u_a, u_b^\top)^\top$ with $u_a = 0$ and

$$u_b = \frac{1}{N}\sum_{i=1}^N \begin{pmatrix} \frac{D_i}{\pi_i} - 1 \\ \frac{D_i\alpha_0^2}{\pi_i} - \alpha_0^2 \\ -\frac{\alpha_0 D_i g(Z_i,\theta_0)}{\pi_i} \end{pmatrix},$$

and

$$\Upsilon \equiv \begin{pmatrix} \Upsilon_{aa} & \Upsilon_{ab} \\ \Upsilon_{ba} & \Upsilon_{bb} \end{pmatrix} = \left(\begin{array}{c:ccc} 0 & 0 & 0 & -\alpha_0 K^\top \\ \hdashline 0 & B_{11} - (1-\alpha_0)^{-1} & \alpha_0^2 B_{11} & -\alpha_0 B_{g1}^\top \\ 0 & \alpha_0^2 B_{11} & \alpha_0^4 B_{11} - \alpha_0^3 & -\alpha_0^3 B_{g1}^\top \\ -\alpha_0 K & -\alpha_0 B_{g1} & -\alpha_0^3 B_{g1} & \alpha_0^2 B_{gg} \end{array}\right),$$

It can be verified that $\sqrt{N} u_b \xrightarrow{d} N(0, \Xi_{bb})$, where

$$\Xi_{bb} = \begin{pmatrix} B_{11} - 1 & \alpha_0^2 (B_{11} - 1) & -\alpha_0 B_{g1}^\top \\ \alpha_0^2 (B_{11} - 1) & \alpha_0^4 (B_{11} - 1) & -\alpha_0^3 B_{g1}^\top \\ -\alpha_0 B_{g1} & -\alpha_0^3 B_{g1} & \alpha_0^2 (B_{gg} - C_{gg}) \end{pmatrix}.$$

Therefore $\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{ELW}})$ where

$$\Sigma_{\mathrm{ELW}} = (K E_{gg}^{-1} K^\top)^{-1} - (K E_{gg}^{-1} K^\top)^{-1} K E_{gg}^{-1} C_{gg} E_{gg}^{-1} K^\top (K E_{gg}^{-1} K^\top)^{-1}$$

and $E_{gg} = B_{gg} - B_{g1} B_{g1}^\top / (B_{11} - 1)$.

As before, the likelihood ratio statistic can be approximated as $R(\theta_0) = N u_n^\top (\Upsilon_{bb}^{-1} - \Upsilon^{bb}) u_b + o_p(1)$. However, because

$$\begin{aligned} & (\Upsilon_{bb}^{-1} - \Upsilon^{bb}) \Xi_{bb} (\Upsilon_{bb}^{-1} - \Upsilon^{bb}) - (\Upsilon_{bb}^{-1} - \Upsilon^{bb}) \\ = \; & -\Upsilon_{bb}^{-1} \Upsilon_{ba} (\Upsilon_{ab} \Upsilon_{bb}^{-1} \Upsilon_{ba})^{-1} K E_{gg}^{-1} C_{gg} E_{gg}^{-1} K^\top (\Upsilon_{ab} \Upsilon_{bb}^{-1} \Upsilon_{ba})^{-1} \Upsilon_{ab} \Upsilon_{bb}^{-1} \end{aligned}$$

does not vanish if $C_{gg} \neq 0$, we conclude that the limiting distribution of the LRT is not a central chisquare distribution, but a weighted chisquare distribution.

The only difference in form between the matrices Υ and Ξ_{bb} in this case and in the case of known propensity score is that the Ξ_{bb} in this case has an extra C_{gg}. We conjecture that if C_{gg} is negligible compared with B_{gg} (e.g. the sampling fraction α_0 is negligible), then the limiting distribution of the LRT will restore to the central χ_r^2 distribution.

9.4 UPS-WR with non-negligible sampling fraction

We adopt the notation and assumptions in Section 3.2 in the main paper. For convenience, we do not distinguish the quantities for fixed N and their limits when $N \to \infty$.

Define $\ell(\alpha, \theta)$, $(\hat{\alpha}, \hat{\theta})$, $R(\theta)$ and $H(\psi)$ the same as in Section 9.1 after replacing $\pi(Z_i)$, Z_i, and $\sum_{i=1}^N$ by π, z_i, and $\sum_{i=1}^n$, respectively. We have

$$\Upsilon \equiv \begin{pmatrix} \Upsilon_{aa} & \Upsilon_{ab} \\ \Upsilon_{ba} & \Upsilon_{bb} \end{pmatrix} = \left(\begin{array}{c:ccc} 0 & 0 & 0 & -\alpha_0 K^\top \\ \hdashline 0 & B_{11} - (1-\alpha_0)^{-1} & \alpha_0^2 B_{11} & -\alpha_0 B_{g1}^\top \\ 0 & \alpha_0^2 B_{11} & \alpha_0^4 B_{11} - \alpha_0^3 & -\alpha_0^3 B_{g1}^\top \\ -\alpha_0 K & -\alpha_0 B_{g1} & -\alpha_0^3 B_{g1} & \alpha_0^2 B_{gg} \end{array}\right),$$

and $u = (u_a, u_b^\top)^\top$ with $u_a = 0$ and

$$u_b = \frac{1}{N} \sum_{i=1}^{n} \begin{pmatrix} \frac{1}{\pi_i} - 1 \\ \frac{\alpha_0^2}{\pi_i} - \alpha_0^2 \\ -\frac{\alpha_0 g(z_i,\theta_0)}{\pi_i} \end{pmatrix}.$$

It can be verified that $\sqrt{N} u_b \xrightarrow{d} N(0, \Xi_{bb})$, where

$$\Xi_{bb} = \begin{pmatrix} B_{11} - 1 & \alpha_0^2(B_{11} - 1) & -\alpha_0 B_{g1}^\top \\ \alpha_0^2(B_{11} - 1) & \alpha_0^4(B_{11} - 1) & -\alpha_0^3 B_{g1}^\top \\ -\alpha_0 B_{g1} & -\alpha_0^3 B_{g1} & \alpha_0^2 B_{gg} \end{pmatrix}.$$

Please note that the matrices Υ in the cases of UPS-WOR and UPS-WR have the same expression, and the only difference between the Ξ in these cases is that the matrix C_{gg} appears in the case of UPS-WOR but vanishes in the case of UPS-WR. Consequently, using the same proof in the case of UPW-WOR, we have $\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma_{\mathrm{ELW}})$ with

$$\Sigma_{\mathrm{ELW}} = \Upsilon^{ab} \Xi_{bb} \Upsilon^{ba} = (K E_{gg}^{-1} K^\top)^{-1} = \left\{ K \left(B_{gg} - \frac{B_{g1} B_{g1}^\top}{B_{11} - 1} \right)^{-1} K^\top \right\}^{-1},$$

and the LRT $R(\theta_0) = N u_n^\top (\Upsilon_{bb}^{-1} - \Upsilon_{bb}^{-1}) u_b + o_p(1)$ asymptotically follows the central χ_r^2 distribution.

References

Berger, Y. G. (2018) Empirical likelihood approaches in survey sampling. *Survey Statistician*, **78**, 22–31.

Berger, Y. G. and Torres, O. D. L. R. (2016) Empirical likelihood confidence intervals for complex sampling dvesigns. *Journal of the Royal Statistical Society: Series B*, **78**, 319–341.

Berger, Y. G. (2020) An empirical likelihood approach under cluster sampling with missing observations. *Annals of the Institute of Statistical Mathematics*, **72**, 91–121.

Bickel, P. J. and Freedman, D. A. (1984) Asymptotic normality and the bootstrap in stratified sampling. *Annals of Statistics*, **12**, 470–482.

Brändén, P. and Jonasson, J. (2012) Negative dependence in sampling. *Scandinavian Journal of Statistics*, **39**(4), 830–838.

Chen, J. and Rao, J. N. K. (2007) Asymptotic normality under two-phase sampling designs. *Statistica Sinica*, **17**, 1047–1064.

Chen, S., and Kim, J. K. (2014). Population empirical likelihood for nonparametric inference in survey sampling. *Statistica Sinica*, **24**, 335–355.

Chen, J. and Sitter, R. R. (1999) A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statistica Sinica*, **9**, 385–406.

Erdös, P. and Rényi, A. (1959) On the central limit theorem for samples from a finite population. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, **4**, 49–61.

Hájek, J. (1960) Limiting distributions in simple random sampling from a finite population. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, **5**, 361–374.

Hájek, J. (1964) Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Annals of Mathematical Statistics*, **35**, 1491–1523.

Hansen, M. and Hurwitz, W. (1943) On the theory of sampling from finite populations. *Annals of Mathematical Statistics*, **14**, 333–362.

Hartley, H. O. and Rao, J. N. K. (1968). A new estimation theory for sample surveys. *Biometrika*, **55**, 547–557.

Horvitz, D. G. and Thompson, D. J. (1952) A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, **47**, 663–685.

Kim, J. K. (2009) Calibration estimation using empirical likelihood in survey sampling. *Statistica Sinica*, **19**, 145–157.

Krewski, D. and Rao, J. N. K. (1981) Inference from stratified samples: properties of linearization, jackknife and balanced repeated replication methods. *Annals of Statistics*, **9**, 1010–1019.

Newman, C. M. (1980). Normal fluctuations and the FKG inequalities. *Communication in Mathematical Physics*, **91**, 75–90.

Owen, A. B. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249.

Owen, A. B. (1990) Empirical likelihood ratio confidence regions. *Annals of Statistics*, **18**, 90–120.

Owen, A. B. (2001) *Empirical Likelihood.* New York: Chapman and Hall.

Patterson, R. F., Smith, W. D., Taylor, R. L. and Bozorgnia, A. (2001) Limit theorems for negatively dependent random variables. *Nonlinear Analysis*, **47**, 1283–1295.

Qin, J. and Lawless, J. (1994) Empirical likelihood and general equations. *Annals of Statistics*, **22**, 300–325.

Rao, J. N. K. and Wu, C. (2010). Pseudo-empirical likelihood inference for multiple frame surveys. *Journal of the American Statistical Association*, **105**(492), 1494-1503.

Taylor, R. L., Patterson, R. F and Bozorgnia, A. (2001). Weak laws of large numbers for arrays of rowwise negatively dependent random variables. *Journal of Applied Mathematics and Stochastic Analysis*, **14**(3), 227–236.

van de Vaart, A. W. (2000). *Asymptotic Statistics.* Cambridge University Press, Cambridge, United Kingdom.

Wu, C., and Rao, J. N. K. (2006). Pseudo-empirical likelihood ratio confidence intervals for complex surveys. *Canadian Journal of Statistics*, **34**, 359–375.

Wu, C. and Thompson, M. E. (2020) *Sampling Theory and Practice.* Switzerland: Springer.

Zhao, P., Haziza, D. and Wu, C. (2022). Sample empirical likelihood and the design-based oracle variable selection theory. *Statistica Sinica*, **32**(2). DOI: 10.5705/ss.202019.0058.

Zhao, P. and Wu, C. (2019) Some theoretical and practical aspects of empirical likelihood methods for complex surveys. *International Statistical Review*, **87**, S1, S239–S256.

Zong, X., Zhu, R. and Zou, G. (2019) Improved Horvitz–Thompson estimator in survey sampling. *Survey Methodology*, **45**(1), 165–184.