SUPPLEMENTARY MATERIAL FOR "WEIGHTED LIKELIHOOD ESTIMATION UNDER TWO-PHASE SAMPLING"

BY TAKUMI SAEGUSA* AND JON A. WELLNER†

University of Washington

In this supplement we present the proofs for [6]. Equation and theorem references made to the main document do not contain letters.

A. Appendix. We repeatedly use the notation for empirical measures and processes introduced in Section 2 following [2]. The fundamental idea of [2] is to view $\mathbb{G}^{\xi}_{j,N_j}$ as the exchangeably weighted bootstrap empirical process corresponding to $\mathbb{G}_{j,N_j} \equiv \sqrt{N_j}\left(\mathbb{P}_{j,N_j} - P_{0|j}\right)$ for $j = 1, \ldots, J$. The processes $\mathbb{G}^{\xi}_{j,N_j}$ converge weakly to $\sqrt{p_j(1-p_j)}\mathbb{G}_j$ for independent $P_{0|j}$-Brownian bridge processes $\mathbb{G}_j$, $j = 1, \ldots, J$, in $\ell^\infty(\mathcal{F})$ for Donsker classes $\mathcal{F}$.

Asymptotic linearity and the limiting distributions of $\hat{\alpha}_N$ in binary regression and (modified and centered) calibration are given by the following proposition. The proof requires a Glivenko-Cantelli theorem for $\mathbb{P}^{\pi}_N$ whose proof is independent of Proposition A.1.

PROPOSITION A.1. *Under the Condition 3.1, $\hat{\alpha}_N$ is consistent for α_0, and*

$$\begin{aligned}
&\sqrt{N}(\hat{\alpha}_N - \alpha_0) \\
&= S_0^{-1}\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{\dot{G}_e(Z_i^T\alpha_0)Z_i}{\pi_0(V_i)(1-\pi_0(V_i))}(\xi_i - \pi_0(V_i)) + o_P^*(1) \\
&\rightsquigarrow S_0^{-1}\sum_{j=1}^{J}\sqrt{\frac{\nu_j}{p_j(1-p_j)}}\mathbb{G}_j\dot{G}_e(Z^T\alpha_0)Z,
\end{aligned}$$

where $\mathbb{G}_j$ are independent $P_{0|j}$-Brownian bridge processes.

*Supported by NIH/NIAID R01 AI089341

†Supported in part by NSF Grant DMS-1104832, NI-AID grant 2R01 AI291968-04, and the Alexander von Humboldt Foundation

Under the Condition 3.2, both $\hat{\alpha}_N^c$, $\hat{\alpha}_N^{mc}$ and $\hat{\alpha}_N^{cc}$ are consistent, and

$$\begin{aligned}
&\sqrt{N}(\hat{\alpha}_N^c - \alpha_0) \\
&= -\frac{1}{\sqrt{N}} \sum_{i=1}^N \dot{G}(0)^{-1} \left\{P_0 Z^{\otimes 2}\right\}^{-1} Z_i \left(\frac{\xi_i - \pi_0(V_i)}{\pi_0(V_i)}\right) + o_{P*}(1) \\
&\rightsquigarrow -\dot{G}(0)^{-1} \left\{P_0 Z^{\otimes 2}\right\}^{-1} \sum_{j=1}^J \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j Z,
\end{aligned}$$

$$\begin{aligned}
&\sqrt{N}(\hat{\alpha}_N^{mc} - \alpha_0) \\
&= -\frac{1}{\sqrt{N}} \sum_{i=1}^N \dot{G}(0)^{-1} \left\{P_0 \frac{1-\pi_0(V)}{\pi_0(V)} Z^{\otimes 2}\right\}^{-1} Z_i \left(\frac{\xi_i - \pi_0(V_i)}{\pi_0(V_i)}\right) + o_{P*}(1) \\
&\rightsquigarrow -\dot{G}(0)^{-1} \left\{P_0 \frac{1-\pi_0(V)}{\pi_0(V)} Z^{\otimes 2}\right\}^{-1} \sum_{j=1}^J \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j Z,
\end{aligned}$$

and

$$\begin{aligned}
&\sqrt{N}(\hat{\alpha}_N^{cc} - \alpha_0) \\
&= -\dot{G}(0)^{-1} \left\{P_0 \frac{1-\pi_0(V)}{\pi_0(V)} (Z-\mu_Z)^{\otimes 2}\right\}^{-1} \\
&\quad \times \frac{1}{\sqrt{N}} \sum_{i=1}^N (Z_i - \mu_Z) \left(\frac{\xi_i - \pi_0(V_i)}{\pi_0(V_i)}\right) + o_{P*}(1) \\
&\rightsquigarrow -\dot{G}(0)^{-1} \left\{P_0 \frac{1-\pi_0(V)}{\pi_0(V)} (Z-\mu_Z)^{\otimes 2}\right\}^{-1} \sum_{j=1}^J \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j (Z-\mu_Z),
\end{aligned}$$

where the $P_{0|j}$-Brownian bridge processes, $\mathbb{G}_j$, are independent.

PROOF. We first consider estimated weights. Define $M_N(\alpha) \equiv \mathbb{P}_N m_\alpha$ and $M(\alpha) = P_{\alpha_0} m_\alpha$ where $m_\alpha(Z,\xi) = \log\left(\{p_\alpha(\xi|Z) + p_{\alpha_0}(\xi|Z)\}/2\right)$. We again apply Theorem 5.7 of [7] for a consistency proof. Because $p_\alpha(\xi|Z)$ is a valid marginal density of a single observation ξ given Z, the argument of [7], page 66, can be used to verify the second condition of the theorem. We verify the first condition of Theorem 5.7 of [7]. Let $\tilde{G}_e(z;\alpha) \equiv \{G_e(z^T\alpha) + G_e(z^T\alpha_0)\}/2$. Then $m_\alpha(z,\xi) = \xi \log \tilde{G}_e(z;\alpha) + (1-\xi)\log(1-\tilde{G}_e(z;\alpha))$. We rewrite $\mathbb{P}_N m_\alpha$ as

$$\begin{aligned}\mathbb{P}_N m_\alpha &= \frac{1}{N}\sum_{i=1}^N \xi_i \log \tilde{G}_e(Z_i;\alpha) + (1-\xi_i)\log\left(1-\tilde{G}_e(Z_i;\alpha)\right)\\ &= \sum_{j=1}^J \left\{\frac{N_j}{N}\frac{n_j}{N_j}\left[\frac{1}{N_j}\sum_{i=1}^N \frac{\xi_{j,i}}{n_j/N_j}\log\tilde{G}_e(Z_{j,i};\alpha)\right]\right\}\\ &\quad+\sum_{j=1}^J\left\{\frac{N_j}{N}\left(1-\frac{n_j}{N_j}\right)\left[\frac{1}{N_j}\sum_{i=1}^N\frac{1-\xi_{j,i}}{1-n_j/N_j}\log\left(1-\tilde{G}_e(Z_{j,i};\alpha)\right)\right]\right\}.\end{aligned}$$

Thus, if we establish that both $\mathcal{S}_{0,j}\equiv\left\{\log\left(1-\tilde{G}_e(z^T\alpha)\right):\alpha\in\mathbb{R}^{J+k}, V\in\mathcal{V}_j\right\}$ and $\mathcal{S}_{1,j}\equiv\left\{\log\tilde{G}_e(z^T\alpha):\alpha\in\mathbb{R}^{J+k}, V\in\mathcal{V}_j\right\}$ are P_0-Glivenko-Cantelli for $j=1,\ldots,J$, it follows from Theorem 5.1 applied to sampled subjects and non-sampled subjects in each stratum separately that $\mathbb{P}_N m_\alpha$ converges in probability to

$$\begin{aligned}P_0 m_\alpha &= \sum_{j=1}^J \nu_j p_j P_0\left(\log\tilde{G}_e(Z^T\alpha)\Big| V\in\mathcal{V}_j\right)\\ &\quad+\sum_{j=1}^J \nu_j(1-p_j)P_0\left(\log\left(1-\tilde{G}_e(Z^T\alpha)\right)\Big| V\in\mathcal{V}_j\right),\end{aligned}$$

uniformly in α. Note that the method of estimated weights does not estimate the sampling probability for the subjects in a stratum if the sampling probability is 1. Thus, we can assume that $G_e(Z^T\alpha_0)\le\sigma'<1$. Hence we have $\log(\sigma/2)\le\log\tilde{G}_e(Z^T\alpha)\le 0$ and $\log(\{1-\sigma'\}/2)\le\log\left(1-\tilde{G}_e(Z^T\alpha)\right)\le 0$ for all $j=1,\ldots,J$ and $\alpha\in\mathbb{R}^{J+k}$. This implies that all sets $\mathcal{S}_{k,j}, k=0,1$, have integrable envelopes. Now it suffices to show that all sets are VC subgraph classes. Note first that $\{z^T\alpha:\alpha\in\mathbb{R}^{J+k}\}$ is a VC subgraph class by Lemma 2.6.15 of [10]. Note also that G_e and the logarithm are monotone functions. Because a map by a monotone function, addition and multiplication all preserve the property of the VC subgraph class by Lemma 2.6.17 of [10], our claim follows and hence the first condition is verified. Since we have by concavity of the logarithm and the property of $\hat{\alpha}_N$ that

$$\begin{aligned}M_N(\hat{\alpha}_N) &\ge \frac{1}{2}\mathbb{P}_N\log p_{\hat{\alpha}_N}(\xi|V)+\frac{1}{2}\mathbb{P}_N\log p_{\alpha_0}(\xi|V)\\ &\ge \frac{1}{2}\mathbb{P}_N\log p_{\alpha_0}(\xi|V)+\frac{1}{2}\mathbb{P}_N\log p_{\alpha_0}(\xi|V)=M_N(\alpha_0),\end{aligned}$$

consistency follows from Theorem 5.7 of [7].

We apply Theorem 3.3.1 of [10] to show asymptotic normality of $\hat\alpha_N$. Define

$$\Phi_{N,e}(\alpha) = \frac{1}{N}\sum_{i=1}^N \frac{\dot G_e(Z_i^T\alpha)Z_i}{G_e(Z_i^T\alpha)(1-G_e(Z_i^T\alpha))}\left(\xi_i - G_e(Z_i^T\alpha)\right) \equiv \mathbb{P}_N\phi_\alpha(\xi,V),$$

and

$$\Phi_e(\alpha) = P_0\left\{\frac{\dot G_e(Z^T\alpha)Z}{G_e(Z^T\alpha)(1-G_e(Z^T\alpha))}\left(\sum_{j=1}^J p_j I(V\in\mathcal{V}_j) - G_e(Z^T\alpha)\right)\right\}.$$

Note that $\Phi_{N,e}(\hat\alpha_N)=0$ because $(\partial/\partial\alpha)\mathbb{P}_N\log p_\alpha = \Phi_{N,e}(\alpha)$. Note also that $\Phi_e(\alpha_0)=0$ since $G_e(Z^T\alpha_0)=p_j$ when $V\in\mathcal{V}_j$. It follows by the decomposition (10) of the inverse probability weighted empirical processes in [2] that

$$\begin{aligned}
&\sqrt{N}(\Phi_{N,e}(\alpha_0)-\Phi_e(\alpha_0)) = \sqrt{N}\Phi_{N,e}(\alpha_0)\\
&= \sqrt{N}\mathbb{P}_N\frac{\dot G_e(Z^T\alpha_0)Z}{G_e(Z^T\alpha_0)(1-G_e(Z^T\alpha_0))}(\xi - G_e(Z^T\alpha_0))\\
&= \sqrt{N}\mathbb{P}_N^\pi\frac{\pi_0(V)}{G_e(Z^T\alpha_0)}\frac{\dot G_e(Z^T\alpha_0)Z}{1-G_e(Z^T\alpha_0)} - \sqrt{N}\mathbb{P}_N\frac{\dot G_e(Z^T\alpha_0)Z}{1-G_e(Z^T\alpha_0)}\\
&= \sum_{j=1}^J\sqrt{\frac{N_j}{N}}\frac{N_j}{n_j}\mathbb{G}_j^\xi\frac{\pi_0(V)}{G_e(Z^T\alpha_0)}\frac{\dot G_e(Z^T\alpha_0)Z}{1-G_e(Z^T\alpha_0)}\\
&\quad+\sqrt{N}\mathbb{P}_N\left(\frac{\pi_0(V)}{G_e(Z^T\alpha_0)}-1\right)\frac{\dot G_e(Z^T\alpha_0)Z}{1-G_e(Z^T\alpha_0)}.
\end{aligned}$$

Since $\pi_0(V)=n_j/N_j$ and $G_e(Z^T\alpha_0)=p_j$ when $V\in\mathcal{V}_j$, the first term converges to

$$\sum_{j=1}^J\sqrt{\frac{N_j}{N}}\frac{N_j}{n_j}\frac{n_j/N_j}{p_j(1-p_j)}\mathbb{G}_j^\xi\dot G_e(Z^T\alpha_0)Z \rightsquigarrow \sum_{j=1}^J\sqrt{\frac{\nu_j}{p_j(1-p_j)}}\mathbb{G}_j\dot G_e(Z^T\alpha_0)Z.$$

The second term can be written as

$$\sum_{j=1}^J\sqrt{N_j}\left(\frac{n_j}{N_j}-p_j\right)\sqrt{\frac{N_j}{N}}\frac{1}{p_j(1-p_j)}\frac{1}{N_j}\sum_{i=1}^{N_j}\dot G_e(Z_{j,i}^T\alpha_0)Z_{j,i}.$$

Since $n_j=[N_jp_j]$ by assumption, it is easy to see that $-N_j^{-1/2}\le\sqrt{N_j}(n_j/N_j-p_j)\le 0$, and hence $\sqrt{N_j}(n_j/N_j-p_j)\to 0$. Since $N_j^{-1}\sum_{i=1}^{N_j}\dot G_e(Z_{j,i}^T\alpha_0)Z_{j,i} =$

$O_{P^*}(1)$ by the weak law of large numbers and $\sqrt{N_j/N} \to \sqrt{\nu_j}$, the second term converges to zero in probability. The weak convergence of $\sqrt{N}(\Phi_{N,e} - \Phi_e)(\alpha_0)$ follows from Slutsky's theorem.

For asymptotic equicontinuity of the process, it suffices to consider a compact subset $\mathcal{A}_{e,0} \in \mathbb{R}^{J+k}$ where α_0 is its interior point since $\hat{\alpha}_N$ is consistent. Let

$$\phi_{\alpha,1}(v) \equiv \frac{\pi_0(v) z^{\otimes 2}}{G_e(z^T\alpha)\left\{1 - G_e(z^T\alpha)\right\}} \left(\ddot{G}_e(z^T\alpha) - \frac{\left\{\dot{G}_e(z^T\alpha)\right\}^2}{G_e(z^T\alpha)} \right),$$

$$\phi_{\alpha,2}(v) \equiv \frac{z^{\otimes 2}}{1 - G_e(z^T\alpha)} \left(\ddot{G}_e(z^T\alpha) - \frac{\left\{\dot{G}_e(z^T\alpha)\right\}^2}{1 - G_e(z^T\alpha)} \right).$$

Taylor's theorem gives

$$\phi_\alpha(\xi, v) - \phi_{\alpha_0}(\xi, v) = \phi_{\alpha_1^*,1}(v) \frac{\xi}{\pi_0(v)} (\alpha - \alpha_0) + \phi_{\alpha_2^*,2}(v)(\alpha - \alpha_0),$$

where $\alpha_j^*, j = 1, 2$, are some convex combinations of α and α_0. Thus,

$$\begin{aligned}
&\sqrt{N}(\Phi_{N,e} - \Phi_e)(\alpha) - \sqrt{N}(\Phi_{N,e} - \Phi_e)(\alpha_0) \\
&= \sqrt{N}(\mathbb{P}_N - P_0)(\phi_\alpha - \phi_{\alpha_0}) + \sqrt{N} P_0(\phi_\alpha - \phi_{\alpha_0}) \\
&\qquad - \sqrt{N}\Phi_e(\alpha) + \sqrt{N}\Phi_e(\alpha_0) \\
&= (\mathbb{P}_N^\pi - P_0)\phi_{\alpha_1^*,1}\sqrt{N}(\alpha - \alpha_0) + (\mathbb{P}_N - P_0)\phi_{\alpha_2^*,2}\sqrt{N}(\alpha - \alpha_0) \\
&\qquad + P_0 \phi_{\alpha_1^*,1} \left(\xi - \sum_{j=1}^J p_j I(V \in \mathcal{V}_j) \right) \sqrt{N}(\alpha - \alpha_0).
\end{aligned} \tag{A.1}$$

To show this is $o_{P^*}(1 + \sqrt{N}(\alpha - \alpha_0))$, we first show that the set $\mathcal{T}_k = \{\phi_{\alpha,k} : \alpha \in \mathcal{A}_{e,0}\}, k = 1, 2$, are Glivenko-Cantelli. It is easy to see that $\{z^T\alpha : \alpha \in \mathcal{A}_{e,0}\}$ is Glivenko-Cantelli. Since $G_e \in \mathcal{C}^2$ by assumption, $\phi_{\alpha,k}, k = 1, 2$, are uniformly bounded in $\alpha \in \mathcal{A}_{e,0}$. Thus, the sets $\mathcal{T}_k$, $k = 1, 2$ are both Glivenko-Cantelli by the Glivenko-Cantelli preservation theorem [9]. For the third term in (A.1), apply the dominated convergence theorem with $P_0(\xi|V) = \sum_{j=1}^J (n_j/N_j) I(V \in \mathcal{V}_j) \to \sum_{j=1}^J p_j I(V \in \mathcal{V}_j)$.

Since $\dot{\Phi}(\alpha_0) = -S_0$, apply Theorem 3.3.1 of [10] to obtain

$$\begin{aligned}
&\sqrt{N}(\hat{\alpha}_N - \alpha_0) \\
&= S_0^{-1}\sqrt{N}\frac{1}{N}\sum_{i=1}^{N}\frac{\dot{G}_e(Z_i^T\alpha_0)Z_i}{G_e(Z_i^T\alpha_0)(1-G_e(Z_i^T\alpha_0))}\left(\xi_i - G_e(Z_i^T\alpha_0)\right) + o_{P^*}(1) \\
&\rightsquigarrow S_0^{-1}\sum_{j=1}^{J}\sqrt{\frac{\nu_j}{p_j(1-p_j)}}\mathbb{G}_j\dot{G}_e(Z^T\alpha_0)Z.
\end{aligned}$$

This completes the proof.

Next we consider modified calibration with $\hat{\alpha}_N = \hat{\alpha}_N^{mc}$. The cases for (centered) calibration (i.e., $\hat{\alpha}_N = \hat{\alpha}_N^{c}$ and $\hat{\alpha}_N = \hat{\alpha}_N^{cc}$) are similar. Define $\Phi_{N,mc}(\alpha) \equiv \mathbb{P}_N^{\pi}G_{mc}(V;\alpha)Z - \mathbb{P}_N Z$ and $\Phi_{mc}(\alpha) \equiv P_0[(G_{mc}(V;\alpha)-1)Z]$. Note that $\Phi_{N,mc}(\hat{\alpha}_N) = 0$ and $\Psi_{mc}(0) = 0$. We apply Theorem 5.7 of [7] for a consistency proof. For the first condition of the theorem, we have

$$\begin{aligned}
&\sup_{\alpha\in\mathbb{R}^k}\|\Phi_{N,mc}(\alpha) - \Phi_{mc}(\alpha)\| \\
&= \sup_{\alpha\in\mathbb{R}^k}\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\xi_i}{\pi_0(V_i)}G_{mc}(V;\alpha) - 1\right)Z_i - P_0\left\{G_{mc}(V;\alpha)-1\right\}Z\right\| \\
&\leq \sup_{\alpha\in\mathbb{R}^k}\left\|\frac{1}{N}\sum_{i=1}^{N}\frac{\xi_i}{\pi_0(V_i)}G_{mc}(V_i;\alpha)Z_i - P_0G_{mc}(;\alpha)Z\right\| \\
&\quad + \sup_{\alpha\in\mathbb{R}^k}\left\|\frac{1}{N}\sum_{i=1}^{N}Z_i - P_0Z\right\|,
\end{aligned}$$

where $\|\cdot\|$ is the Euclidean norm. Since α is a vector in $\mathbb{R}^k$ and G is monotone, $\{G_{mc}(\cdot;\alpha) : \alpha \in \mathbb{R}^k\}$ is a VC subgraph by Lemmas 2.6.15 and 2.6.18 of [10]. Boundedness of G implies that the set $\left\{G_{mc}(v;\alpha)z : \alpha \in \mathbb{R}^k\right\}$ is P_0-Glivenko-Cantelli by the Glivenko-Cantelli preservation theorem [9]. Then the first term is $o_{P^*}(1)$ by Theorem 5.1. The second term is $o_{P^*}(1)$ by the weak law of large numbers.

The second condition of the theorem is that for any $\epsilon > 0$, $\inf_{|\alpha|>\epsilon}\|\Phi_{mc}(\alpha)\| > 0$. Suppose, to the contrary, that $\inf_{|\alpha|>\epsilon}\|\Phi_{mc}(\alpha)\| = 0$ for some $\epsilon > 0$. Then there exists a sequence $\{\alpha^{(m)}\} \subset \mathbb{R}^k$ with $|\alpha^{(m)}| > \epsilon$ for each $m = 1, 2, \ldots$, such that

$$\|\Phi_{mc}(\alpha^{(m)})\| \to 0.$$

Let $\Phi_{j,c}(\alpha)$, $j = 1, \ldots, k$, be the jth element of $\Phi_{mc}(\alpha)$. Since the norm $\|\cdot\|$ is the Euclidean norm, each element $\Phi_{j,c}(\alpha^{(m)})$ converges to zero. If $\alpha^{(m)}$

converges to $\alpha^{(\infty)}$ with $|\alpha^{(\infty)}| < \infty$, then by the dominated convergence theorem and Taylor's theorem,

$$0 = P_0 \left[\left\{ G_{mc}(V;\alpha^{(\infty)}) - 1 \right\} Z \right] = P_0 \left[(\pi_0(V)^{-1} - 1)\dot{G}_{mc}(V;\alpha^*)Z^{\otimes 2}\alpha^{(\infty)} \right]$$

for some α^* with $|\alpha^*| \leq |\alpha^{(\infty)}|$. Because $P_0(\pi_0(V)^{-1} - 1)\dot{G}_{mc}(V;\alpha^*)Z^{\otimes 2}$ is positive definite by assumption, $\alpha^{(\infty)}$ must be zero, which contradicts the fact that $|\alpha^{(\infty)}| \geq \epsilon$.

We assume that some elements of $\alpha^{(m)}$ diverge. Then, a further subsequence $\alpha^{(m')}$ converges to some $\alpha^{(\infty)}$ whose elements are extended real numbers. Define a unit vector $\beta^{(\infty)} \equiv \lim_{m' \to \infty} \alpha^{(m')}/\|\alpha^{(m')}\|$. Then we have for each Z on the set $\{\pi_0(V) < 1\}$ that

$$\begin{aligned} G_{mc}^{(\infty)}(Z)\, Z^T\beta^{(\infty)} &\equiv \lim_{m'\to\infty} G\left(\frac{1-\pi_0(V)}{\pi_0(V)} Z^T \frac{\alpha^{(m')}}{\|\alpha^{m'}\|}\|\alpha^{m'}\| \right) Z^T\beta^{(\infty)} \\ &= \begin{cases} M_1 Z^T\beta^{(\infty)} & \text{if } Z^T\beta^{(\infty)} > 0 \\ m_1 Z^T\beta^{(\infty)} & \text{if } Z^T\beta^{(\infty)} < 0 \\ 0 & \text{if } Z^T\beta^{(\infty)} = 0 \end{cases} . \end{aligned}$$

It follows by the dominated convergence theorem applied to each element of the vector of $\Phi_{mc}(\alpha)$ that

$$\begin{aligned} &0 = \lim_{m'\to\infty} \Phi_{mc}\left(\alpha^{(m')}\right)^T \beta^{(\infty)} = P_0 \lim_{m'\to\infty} \left\{ G_{mc}\left(V;\alpha^{(m')}\right) - 1 \right\} Z^T\beta^{(\infty)} \\ &= (M_1 - 1)P_0 I_{\left\{Z^T\beta^{(\infty)}>0,\pi_0(V)<1\right\}} Z^T\beta^{(\infty)} \\ &\quad +(m_1 - 1)P_0 I_{\left\{Z^T\beta^{(\infty)}<0,\pi_0(V)<1\right\}} Z^T\beta^{(\infty)}. \end{aligned}$$

However, this is strictly positive since $m_1 < 1$ and $M_1 > 1$, which is a contradiction. This completes the proof that $\hat{\alpha}_N \to_{P^*} 0$.

We apply Theorem 3.3.1 of [10] to show the asymptotic normality of $\hat{\alpha}_N$. For asymptotic equicontinuity condition, it follows by Taylor's theorem that

$$\begin{aligned} &\sqrt{N}(\Phi_{N,mc} - \Phi_{mc})(\hat{\alpha}_N) - \sqrt{N}(\Phi_{N,mc} - \Phi_{mc})(\alpha_0) \\ &= \mathbb{G}_N^\pi[G_{mc}(V;\hat{\alpha}_N)Z - G_{mc}(V;\alpha_0)Z] \\ &= (\mathbb{P}_N^\pi - P_0)(\pi_0(V)^{-1} - 1)\dot{G}_{mc}(V;\alpha^*)Z^{\otimes 2}\sqrt{N}(\hat{\alpha} - \alpha_0) \end{aligned}$$

for some α^* with $|\alpha^* - \alpha_0| \leq |\hat{\alpha}_N - \alpha_0|$. This term is $o_P(1 + \sqrt{N}|\hat{\alpha} - \alpha_0)|)$ if $(\mathbb{P}_N^\pi - P_0)(\pi_0(V)^{-1} - 1)Z^{\otimes 2}\dot{G}_{mc}(V;\alpha) \to_{P^*} 0$, uniformly in α. Let $\mathcal{A}_{mc,1} \subset \mathbb{R}^k$ be a compact neighborhood of zero. Since $\hat{\alpha}_N$ is consistent, it suffices to show that the set $\{(\pi_0^{-1}(V) - 1)Z^{\otimes 2}\dot{G}_{mc}(Z;\alpha) : \alpha \in \mathcal{A}_{mc,1}\}$ is Glivenko-Cantelli. Since $|\pi_0^{-1}(V) - 1|$ and Z are bounded, the VC subgraph class

$\{(\pi_0^{-1}(V)-1)Z\alpha : \alpha \in \mathcal{A}_{mc,1}\}$ (Lemma 2.6.15 of [10]) is P_0-Glivenko-Cantelli. Because $\dot{G}$ is continuous and bounded, the set $\{\dot{G}_{mc}(Z;\alpha) : \alpha \in \mathcal{A}_{mc,1}\}$ is Glivenko-Cantelli by the Glivenko-Cantelli preservation theorem of [9]. Apply the Glivenko-Cantelli preservation theorem of [9] again to conclude $\{(\pi_0^{-1}(V)-1)Z^{\otimes 2}\dot{G}_{mc}(Z;\alpha) : \alpha \in \mathcal{A}_{mc,1}\}$ is Glivenko-Cantelli. Hence, asymptotic equicontinuity follows from Theorem 5.1. We show the weak convergence of the process $\sqrt{N}(\Phi_{N,mc}-\Phi_{mc})(\alpha)$ at $\alpha_0 = 0$. Since $G_{mc}(v;\alpha_0) = 1$, it follows from the decomposition (10) of the inverse probability weighted empirical processes in [2] that

$$
\begin{aligned}
&\sqrt{N}(\Phi_{N,mc}-\Phi_{mc})(\alpha_0) = \sqrt{N}\Phi_{N,mc}(0) = \sqrt{N}(\mathbb{P}_N^\pi - \mathbb{P}_N)Z \\
&= \sum_{j=1}^J \sqrt{\frac{N_j}{N}}\frac{N_j}{n_j}\mathbb{G}_{j,N_j}^\xi Z \\
&\rightsquigarrow \sum_{j=1}^J \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j Z \quad (\text{ by Theorem 5.3}).
\end{aligned}
$$

The Fréchet derivative of $\Phi_{mc}(\alpha_0)$ is

$$
\dot{\Phi}_{mc}(\alpha)|_{\alpha=\alpha_0} = \frac{\partial}{\partial\alpha}P_0(G_{mc}(V;\alpha)-1)Z\bigg|_{\alpha=\alpha_0} = \dot{G}(0)P_0(\pi_0(V)^{-1}-1)Z^{\otimes 2}.
$$

Thus, by Theorem 3.3.1 of [10] we obtain

$$
\begin{aligned}
\sqrt{N}\hat{\alpha}_N &= -\dot{\Phi}_{mc}(0)\sqrt{N}(\Phi_{N,mc}-\Phi_{mc})(0) + o_{P*}(1) \\
&\rightsquigarrow -\dot{G}(0)^{-1}\left\{P_0(\pi_0(V)^{-1}-1)Z^{\otimes 2}\right\}^{-1}\sum_{j=1}^J \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j Z.
\end{aligned}
$$

□

Here we give proofs of the theorems in Section 5.

Proof of Theorem 5.1. First consider $\mathbb{P}_N^\pi$. By the decomposition (10) of the inverse probability weighted empirical processes in [2], we have

$$
\|\mathbb{P}_N^\pi - P_0\|_{\mathcal{F}} \le \|\mathbb{P}_N - P_0\|_{\mathcal{F}} + \sum_{j=1}^J \frac{N_j}{N}\frac{N_j}{n_j}\left\|\mathbb{P}_{j,N_j}^\xi - \frac{n_j}{N_j}\mathbb{P}_{j,N_j}\right\|_{\mathcal{F}}.
$$

The first term is $o_{P^*}(1)$ since $\mathcal{F}$ is Glivenko-Cantelli. Since $(N_j/N)(N_j/n_j) \to_{P^*} \nu_j/p_j$, each summand in the second term is $o_{P^*}(1)$ by the bootstrap Glivenko-Cantelli theorem, which is an easy corollary to Lemma 3.6.16 of [10].

Consider $\mathbb{P}_N^{\pi,e}$. Because $\hat{\alpha}_N \to_{P^*} \alpha_0$ by Proposition A.1, it suffices to consider a compact neighborhood $K \subset \mathbb{R}^{J+k}$ of α_0. Since Z is bounded and G_e is continuous, $\{\pi_\alpha(V)\}^{-1} = \{G_e(\alpha^T Z)\}^{-1}$ is bounded in this neighborhood. Because α is a vector in $\mathbb{R}^{J+k}$ and G_e is monotone, $\{\{G_e(\alpha)\}^{-1} : \alpha \in K\}$ is a VC subgraph class by Lemmas 2.6.15 and 2.6.18 of [10]. Boundedness of G_e implies that the set

$$\left\{\pi_0\{G_e(\cdot\alpha)\}^{-1} f : f \in \mathcal{F}, \alpha \in K\right\}$$

is P_0-Glivenko-Cantelli by the Glivenko-Cantelli preservation theorem [9]. Since $\hat{\alpha}_N \to_{P^*} \alpha_0$, we have by (5.14) that
$\left\|\mathbb{P}_N^{\pi,e} - P_0\right\|_{\mathcal{F}} \to_{P^*} 0$, by recognizing that

$$\mathbb{P}_N^{\pi,e} = \frac{1}{N}\sum_{i=1}^{N} \frac{\xi_i}{\pi_0(V_i)} \left\{\frac{\pi_0(V_i)}{G_e(\hat{\alpha}_N^T Z_i)} \delta_{X_i}\right\}.$$

Consider $\mathbb{P}_N^{\pi,mc}$. The cases for $\mathbb{P}_N^{\pi,c}$ and $\mathbb{P}_N^{\pi,cc}$ are similar. We verified in the proof of Proposition A.1 that $\{G_{mc}(\cdot;\alpha) : \alpha \in \mathbb{R}^k\}$ is a VC subgraph class. Boundedness of G implies that the set

$$\left\{G_{mc}(\cdot;\alpha) f : f \in \mathcal{F}, \alpha \in \mathbb{R}^k\right\}$$

is P_0-Glivenko-Cantelli by the Glivenko-Cantelli preservation theorem [9]. Since $\hat{\alpha}_N$ converges to zero in probability by Proposition A.1, the result follows by (5.14).

□

Several lemmas are required for the proof of Theorem 5.2.

LEMMA A.1. *Let $\mathcal{F}$ be a class of functions with $P_0|f| < \infty$ for every $f \in \mathcal{F}$. Then,*

$$E^* \left\| \sqrt{\frac{N_j}{N}} I(N_j > 0) \mathbb{G}_{j,N_j} \right\|_{\mathcal{F}} \lesssim E^* \left\|\mathbb{G}_N\right\|_{\mathcal{F}}, \quad \textit{for each} \ \ j = 1, \ldots, J.$$

PROOF. Let ϵ_i, $i = 1, \ldots, N$, be independent Rademacher variables, independent of $X_i, i = 1, \ldots, N$, and N_j. It follows from the symmetrization inequality (Lemma 2.3.6) of [10]

$$E^*\|\mathbb{G}_N\|_{\mathcal{F}} \quad \gtrsim \quad E^* \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \epsilon_i f(X_i) \right\|_{\mathcal{F}}.$$

Rewrite this and use Jensen's inequality again with $E[\epsilon f(X)] = 0$ to obtain

$$\begin{aligned} &E^* \left\| \sum_{j=1}^J I(N_j > 0) \sqrt{\frac{N_j}{N}} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \epsilon_{j,i} f(X_{j,i}) \right\|_{\mathcal{F}} \\ &\gtrsim E^* \left\| I(N_j > 0) \sqrt{\frac{N_j}{N}} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} \epsilon_{j,i} f(X_{j,i}) \right\|_{\mathcal{F}}. \end{aligned}$$

Here we implicitly change the law. This can be justified by Proposition A.1 of [2].

Now applying the Lemma 2.3.6 of [10] to the jth stratum, this is further bounded below, up to some constant, by

$$\begin{aligned} &E^* \left\| I(N_j > 0) \sqrt{\frac{N_j}{N}} \frac{1}{\sqrt{N_j}} \sum_{i=1}^{N_j} (f(X_{j,i}) - P_{0|j} f) \right\|_{\mathcal{F}} \\ &= E^* \left\| I(N_j > 0) \sqrt{N_j/N} \mathbb{G}_{j,N_j} \right\|_{\mathcal{F}}. \end{aligned}$$

□

The following is a multiplier inequality for bounded exchangeable weights. Note that the sum of stochastic processes in the second term is divided by $n^{1/2}$ rather than $k^{1/2}$.

PROOF OF LEMMA 5.1. This follows the proof of Lemma 3.6.7 of [10] up to the last line. Since the ξ_i's can be split into their positive and negative parts, we only consider the case where they are nonnegative. Thus for any $1 \le n_0 \le n$,

$$\begin{aligned} E \left\| \sum_{i=1}^n \xi_i Z_i \right\|_{\mathcal{F}}^* &\le E \left\| \sum_{i=1}^{n_0 - 1} \xi_{(i)} Z_i \right\|_{\mathcal{F}}^* + E \left\| \sum_{j=n_0}^n \xi_{(i)} Z_i \right\|_{\mathcal{F}}^* \\ &\le E \left(\max_{1 \le i \le n} \xi_i \right) \frac{n_0 - 1}{n} \sum_{i=1}^n E^* \|Z_i\|_{\mathcal{F}} + E \left\| \sum_{i=n_0}^n \xi_{(i)} Z_i \right\|_{\mathcal{F}}^*, \end{aligned}$$

where $\xi_{(i)}, i = 1, \ldots, n$, are the reverse order statistics of $\xi_i, i = 1, \ldots, n$. To bound the second term, we substitute $\xi_{(i)} = \sum_{k=i}^n (\xi_{(k)} - \xi_{(k+1)})$ with

$\xi_{(n+1)} = 0$, and change the order of summation to obtain

$$\begin{aligned} E\left\|\sum_{i=n_0}^{n} \xi_{(i)} Z_i\right\|_{\mathcal{F}}^{*} &= E\left\|\sum_{i=n_0}^{n}\sum_{k=i}^{n}(\xi_{(k)} - \xi_{(k+1)}) Z_i\right\|_{\mathcal{F}}^{*} \\ &= E\left\|\sum_{k=n_0}^{n}\left(\xi_{(k)} - \xi_{(k+1)}\right)\sum_{i=n_0}^{k} Z_i\right\|_{\mathcal{F}}^{*}. \end{aligned}$$

It follows from the triangle inequality and the independence of the ξ's and Z_i's that this is bounded by

$$\begin{aligned} &\sum_{k=n_0}^{n} E^{*}\left\|\left(\xi_{(k)} - \xi_{(k+1)}\right)\sum_{i=n_0}^{k} Z_i\right\|_{\mathcal{F}}^{*} \\ &\quad = \sum_{k=n_0}^{n} E^{*}\left\{\left(\xi_{(k)} - \xi_{(k+1)}\right)\left\|\sum_{i=n_0}^{k} Z_i\right\|_{\mathcal{F}}^{*}\right\} \\ &\quad = \sum_{k=n_0}^{n} E^{*}(\xi_{(k)} - \xi_{(k+1)}) E^{*}\left\|\sum_{i=n_0}^{k} Z_i\right\|_{\mathcal{F}}^{*} \\ &\quad \leq \sum_{k=n_0}^{n} E^{*}(\xi_{(k)} - \xi_{(k+1)}) \max_{n_0 \leq k \leq n} E^{*}\left\|\sum_{i=n_0}^{k} Z_i\right\|_{\mathcal{F}}^{*} \\ &\quad = E^{*}\sum_{k=n_0}^{n}\left(\xi_{(k)} - \xi_{(k+1)}\right) \max_{n_0 \leq k \leq n} E^{*}\left\|\sum_{i=n_0}^{k} Z_{R_i}\right\|_{\mathcal{F}}^{*} \\ &\quad \leq (u-l) \max_{n_0 \leq k \leq n} E^{*}\left\|\sum_{i=n_0}^{k} Z_{R_i}\right\|_{\mathcal{F}}^{*} \end{aligned}$$

using the boundedness of the ξ_i's in the last line. The proof for the negative parts of the ξ_i's is similar and the inequality follows. □

LEMMA A.2. *For an arbitrary set $\mathcal{F}$ of integrable functions,*

$$E^{*}\left\|\mathbb{G}_N^{\pi}\right\|_{\mathcal{F}} \lesssim E^{*}\left\|\mathbb{G}_N\right\|_{\mathcal{F}}.$$

PROOF. We decompose $\mathbb{G}_N^\pi$ as in (2.1): thus

$$\begin{aligned} E^*\|\mathbb{G}_N^\pi\|_{\mathcal{F}} &= E^*\left\|\mathbb{G}_N + \sum_{j=1}^J \sqrt{\frac{N_j}{N}}\left(\frac{N_j}{n_j}\right)\mathbb{G}_{j,N_j}^\xi\right\|_{\mathcal{F}} \\ &\leq E^*\left\|\mathbb{G}_N\right\|_{\mathcal{F}} + \sum_{j=1}^J E^*\left\|\sqrt{\frac{N_j}{N}}\left(\frac{N_j}{n_j}\right)\mathbb{G}_{j,N_j}^\xi\right\|_{\mathcal{F}}. \end{aligned}$$

It therefore suffices to show that each $E^*\left\|m_j\mathbb{G}_{j,N_j}\right\|_{\mathcal{F}}$ is bounded up to some constant by $E^*\left\|\mathbb{G}_N\right\|_{\mathcal{F}}$ where $m_j \equiv (N_j/N)^{1/2}(N_j/n_j)$.

Rewrite $\mathbb{G}_{j,N_j}^\xi$ as

$$\mathbb{G}_{j,N_j}^\xi = \sqrt{N_j}\left(\mathbb{P}_{j,N_j}^\xi - \frac{n_j}{N_j}\mathbb{P}_{j,N_j}\right) = \frac{1}{\sqrt{N_j}}\sum_{i=1}^{N_j}\xi_{j,i}\left(\delta_{X_{j,i}} - \mathbb{P}_{j,N_j}\right).$$

Now we condition on $\underline{N} \equiv (N_1,\ldots,N_J)$, and write $E_{\underline{N}}$ for $E(\cdot|\underline{N})$. Since $\xi_{j,i} \in \{0,1\}$, it follows by the multiplier inequality of Lemma 5.1 applied conditionally with $n_0 = 1$ and $Z_i = m_j(\delta_{X_{j,i}} - \mathbb{P}_{j,N_j})$ that $E_{\underline{N}}\|m_j\mathbb{G}_{j,N_j}^\xi\|_{\mathcal{F}}$ is bounded by

$$\begin{aligned} &(1-0)\max_{1\leq k\leq N_j} E_{\underline{N}}\left\|\frac{1}{\sqrt{N_j}}\sum_{i=1}^k m_j(\delta_{X_{j,i}} - \mathbb{P}_{j,N_j})\right\|_{\mathcal{F}}^* \\ &= \max_{1\leq k\leq N_j} E_{\underline{N}}\left[\frac{N_j}{n_j}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^k(\delta_{X_{j,i}} - \mathbb{P}_{j,N_j})\right\|_{\mathcal{F}}^*\right]. \end{aligned}$$

Note that $N_j/n_j \leq \sigma^{-1}$ for some $\sigma > 0$ by assumption so that we can replace N_j/n_j by σ^{-1} in the last display to obtain an upper bound. Then, apply the triangle inequality to further bound this by

$$\max_{1\leq k\leq N_j} E_{\underline{N}}^*\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^k(\delta_{X_{j,i}} - P_{0|j})\right\|_{\mathcal{F}}^* + \max_{1\leq k\leq N_j} E_{\underline{N}}^*\left[\frac{k}{\sqrt{N}}\left\|(\mathbb{P}_{j,N_j} - P_{0|j})\right\|_{\mathcal{F}}\right].$$

Since $\delta_{X_{j,i}} - P_{0|j}$ has mean zero, it follows by Jensen's inequality that the first term is bounded by

$$E_{\underline{N}}^*\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N_j}(\delta_{X_{j,i}} - P_{0|j})\right\|_{\mathcal{F}}^* = E_{\underline{N}}^*\left\|\sqrt{\frac{N_j}{N}}\mathbb{G}_{j,N_j}\right\|_{\mathcal{F}}.$$

The second term is bounded by $E^*_{\underline{N}} \left\| \sqrt{N_j/N}\mathbb{G}_{j,N_j} \right\|_{\mathcal{F}}$. Now compute unconditionally and apply Lemma A.1 to find that both terms are bounded by $E^*\|\mathbb{G}_N\|_{\mathcal{F}}$. □

PROOF OF THEOREM 5.2. It follows by Lemma A.2 and the assumption on $E^*\|\mathbb{G}_N\|_{\mathcal{M}_\delta}$ that

$$E^* \left\|\mathbb{G}_N^\pi\right\|_{\mathcal{M}_\delta} \lesssim E^* \left\|\mathbb{G}_N\right\|_{\mathcal{M}_\delta} \leq \phi_N(\delta).$$

By application of Theorem 3.2.5 of [10], we conclude that the conclusion of (1) of the theorem holds.

For the second statement, note that Theorem 3.2 of [4] holds in a general setting where $P_0 m_{\theta,\eta}$ and $\mathbb{P}_n m_{\theta,\eta}$ are replaced by the deterministic function $\mathbb{M}(\theta,\eta)$ and the stochastic process $\mathbb{M}_n(\theta,\eta)$, respectively. Our parameters α and θ play roles of their θ and η, respectively. Our choice of $\mathbb{M}$ and $\mathbb{M}_N$ is $P_0 G_{mc}(V;\alpha)m_\theta$ and $\mathbb{P}_N^{\pi,mc} m_\theta$. The condition 5.17 corresponds to (3.5) of [4]. The condition 5.18 together with Lemma A.2 verifies their (3.6). Apply their Theorem 3.2 to obtain $d(\hat{\theta}_{N,mc},\theta_0) \leq O_{P^*}(\delta_N^{-1} + |\hat{\alpha}_N - \alpha_0|) = O_{P^*}(\delta_N^{-1})$. The cases for $\hat{\theta}_{N,e}$, $\hat{\theta}_{N,c}$ and $\hat{\theta}_{N,cc}$ are similar. □

PROOF OF LEMMA 5.2. We consider modified calibration. Other three cases are similar. Because $G(0) = 1$ and Z is bounded, consistency of $\hat{\alpha}_N$ implies that there exists $\mathcal{A}_{mc,2} \subset \mathcal{A}_{mc}$ such that for some fixed constant $C > 0$, $G_{mc}(v;\alpha) \geq C$ and $\dot{G}_{mc}(v;\alpha) \geq C$ for every $\alpha \in \mathcal{A}_{mc,2}$ and $P(\hat{\alpha}_N \in \mathcal{A}_{mc,2}) \to 1$. Then, for arbitrary $\alpha \in \mathcal{A}_{mc,2}$,

$$\begin{aligned}
&P_0 G_{mc}(V;\alpha)(m_\theta - m_{\theta_0}) = P_0 G_{mc}(V;\alpha)\log\frac{p_\theta}{p_{\theta_0}} \\
&\leq 2P_0 G_{mc}(V;\alpha)\left(\sqrt{\frac{p_\theta}{p_{\theta_0}}} - 1\right) \\
&= \int G_{mc}(v;\alpha)\left\{-(p_\theta^{1/2} - p_{\theta_0}^{1/2})^2 + p_\theta - p_{\theta_0}\right\} d\mu \\
&\leq -C\int (p_\theta^{1/2} - p_{\theta_0}^{1/2})^2 d\mu + \int \{G_{mc}(v;\alpha) - 1\}(p_\theta - p_{\theta_0}) d\mu \\
&= -Ch^2(p_\theta, p_{\theta_0}) + \int \dot{G}_{mc}(v;\alpha^*)(\pi_0^{-1}(v) - 1)v^T(p_\theta - p_{\theta_0})d\mu(\alpha - \alpha_0),
\end{aligned}$$

where α^* is some convex combination of α and α_0. Because the integral in the last display is a bounded row vector, the second term in the last display is bounded by $|\alpha - \alpha_0|^2$ up to some constant. Thus, the condition (5.17) holds. □

The following lemma is useful when showing asymptotic equicontinuity of processes involving $\mathbb{P}_N^{\pi,e}$, $\mathbb{P}_N^{\pi,c}$, $\mathbb{P}_N^{\pi,mc}$ and $\mathbb{P}_N^{\pi,cc}$.

LEMMA A.3. *Suppose Conditions 3.2 and 3.1 hold. Let $\mathcal{F}$ be a Glivenko-Cantelli class. Then*

$$\sup_{f\in\mathcal{F}}\left|\sqrt{N}(\mathbb{P}_N - P_0)\left\{\frac{\xi}{\pi_{\hat{\alpha}_N}(V)}f - \frac{\xi}{\pi_{\alpha_0}(V)}f\right\}\right| = o_{P^*}(1), \tag{A.2}$$

where $\pi_{\hat{\alpha}_N}$ is either an estimated or calibrated probability (with modified or centered calibration).

PROOF. We only consider modified calibration. The cases for estimated weights and (centered) calibration are similar. It follows by Taylor's theorem that

$$\begin{aligned}
&\sup_{f\in\mathcal{F}}\left|\sqrt{N}(\mathbb{P}_N - P_0)\left\{\frac{\xi}{\pi_{\hat{\alpha}_N}(V)}f - \frac{\xi}{\pi_{\alpha_0}(V)}f\right\}\right| \\
&= \sup_{f\in\mathcal{F}}\left|(\mathbb{P}_N^{\pi} - P_0)\left((\pi_0^{-1}(V) - 1)Z^T\dot{G}_{mc}(Z;\alpha^*)f\right)\right|\sqrt{N}|\hat{\alpha}_N - \alpha_0|,
\end{aligned}$$

for some α^* with $|\alpha^* - \alpha_0| \leq |\hat{\alpha}_N - \alpha_0|$. Because $\sqrt{N}(\hat{\alpha}_N - \alpha_0) = O_{P^*}(1)$ by Proposition A.1, it follows that (A.2) is $o_{P^*}(1)$ by Theorem 5.1 and Proposition A.1 if the set $\{(\pi_0(V)^{-1} - 1)Z^T\dot{G}\{(\pi_0^{-1}(V) - 1)Z^T\alpha\} : \alpha \in \mathcal{A}_{mc,3}, f \in \mathcal{F}\}$ is P_0-Glivenko-Cantelli where $\mathcal{A}_{mc,3} \subset \mathcal{A}_{mc}$ is some compact set containing $\alpha_0 = 0$. This is easily verified in the same way as in the proof of Proposition A.1. □

PROOF OF THEOREM 5.3. The result (5.19) follows from [2]. Consider the IPW empirical process with modified calibration. It follows by Taylor's theorem that

$$\begin{aligned}
&\mathbb{G}_N^{\pi,mc}f - \mathbb{G}_N^{\pi}f \\
&= \mathbb{G}_N\left(\frac{\xi}{\pi_{\hat{\alpha}_N}(V)} - \frac{\xi}{\pi_{\alpha_0}(V)}\right)f + \sqrt{N}P_0\left(\frac{\xi}{\pi_{\hat{\alpha}_N}(V)} - \frac{\xi}{\pi_{\alpha_0}(V)}\right)f \\
&= \mathbb{G}_N\left(\frac{\xi}{\pi_{\hat{\alpha}_N}(V)} - \frac{\xi}{\pi_{\alpha_0}(V)}\right)f \\
&\quad + P_0\left(\frac{1-\pi_0(V)}{\pi_0(V)}Z^T\dot{G}_{mc}(V;\alpha^*)f\right)\sqrt{N}(\hat{\alpha}_N - \alpha_0),
\end{aligned} \tag{A.3}$$

where α^* is some convex combination of $\hat{\alpha}_N$ and α_0. The first term is $o_{P^*}(1)$ by Lemma A.3. Since $(\pi_0(V)^{-1} - 1)Z^T\dot{G}_{mc}$ is bounded and f is integrable, it follows from the dominated convergence theorem that

$$P_0\left(\frac{1-\pi_0(V)}{\pi_0(V)}Z^T\dot{G}_{mc}(V;\alpha^*)f\right) \to P_0\left(\frac{1-\pi_0(V)}{\pi_0(V)}Z^T\dot{G}(0)f\right).$$

Apply the result (5.19) and Proposition A.1 to conclude the finite-dimensional convergence

$$\begin{aligned}
&\mathbb{G}_N^{\pi,mc}f = \mathbb{G}_N^{\pi}f + (\mathbb{G}_N^{\pi,mc} - \mathbb{G}_N^{\pi})f \\
&\to_d \mathbb{G}f + \sum_{j=1}^J \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j f \\
&\quad - P_0\left(\frac{1-\pi_0(V)}{\pi_0(V)}Z^T\dot{G}(0)f\right)\dot{G}(0)^{-1}\left\{P_0\frac{1-\pi_0(V)}{\pi_0(V)}Z^{\otimes 2}\right\}^{-1} \\
&\quad \times \sum_{j=1}^J \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j Z \\
&= \mathbb{G}f + \sum_{j=1}^J \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j f - \sum_{j=1}^J \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j Q_{mc} f \\
&= \mathbb{G}f + \sum_{j=1}^J \sqrt{\nu_j}\sqrt{\frac{1-p_j}{p_j}}\mathbb{G}_j (f - Q_{mc} f).
\end{aligned}$$

Next, we prove asymptotic equicontinuity of $\mathbb{G}_N^{\pi,mc}$ with respect to the metric ρ_{mc} defined by

$$\rho_{mc}^2(f,g) = P_0(f-g)^2 + \sum_{j=1}^J \nu_j \frac{1-p_j}{p_j}\mathrm{Var}_{0|j}(f-g).$$

First recall that $\mathbb{G}_N^{\pi}$ is asymptotically equicontinuous with respect to the metric ρ defined by

$$\rho^2(f,g) = \sigma_{P_0}^2(f-g) + \sum_{j=1}^J \nu_j \frac{1-p_j}{p_j}\mathrm{Var}_{0|j}(f-g).$$

The part $\sigma_{P_0}^2(f-g)$ corresponds to the empirical process $\mathbb{G}_N \equiv \sqrt{N}(\mathbb{P}_N - P_0)$ in the decomposition (2.1) of the inverse probability weighted empirical processes. However, this empirical process $\mathbb{G}_N$ is asymptotically equicontinuous with respect to the $L_2(P)$-metric with an assumption $\|P_0\|_{\mathcal{F}} < \infty$ in view of Problem 2.1.2 of [10]. Thus, $\mathbb{G}_N^{\pi}$ is asymptotically equicontinuous with respect to ρ_{mc}. Now, it remains to verify the asymptotic equicontinuity of $\mathbb{G}_N^{\pi,mc} - \mathbb{G}_N^{\pi}$. Let $h_N \in \mathcal{F}_{\delta_N} \equiv \{f - g : f, g \in \mathcal{F}, \rho_{mc}(f,g) \leq \delta_N\}$ for an arbitrary sequence $\delta_N \downarrow 0$. In view of (A.3)

$$(\mathbb{G}_N^{\pi,mc} - \mathbb{G}_N^{\pi})h_N = o_{P^*}(1) + P_0\left(\frac{1-\pi_0(V)}{\pi_0(V)} Z^T \dot{G}_{mc}(V;\alpha^*)h_N\right) O_{P^*}(1),$$

where α^* is some convex combination of $\hat{\alpha}_N$ and α_0. Because each element of a vector $(\pi_0(V)^{-1}-1)Z^T\dot{G}_{mc}(V;\alpha^*)$ is bounded, it follows from the Cauchy-Schwarz inequality that each element of $P_0\{(\pi_0(V)^{-1}-1)Z^T\dot{G}_{mc}(V;\alpha^*)h_N\}$ is bounded up to some constant by $P_0(h_N^2)$. Since $\rho_{mc}(f,g) \to 0$ implies $P_0(f-g)^2 \to 0$, we have $P_0 h_N^2 \to 0$ as $N \to \infty$. This verifies the asymptotic equicontinuity of $\mathbb{G}_N^{\pi,mc}$ and hence completes showing its weak convergence.

The cases for $\mathbb{G}_N^{\pi,e}$, $\mathbb{G}_N^{\pi,c}$ and $\mathbb{G}_N^{\pi,cc}$ follow analogously. □

PROOF OF THEOREM 5.4. Since $\mathcal{F}$ is Donsker, it follows by Lemma 2.3.11 of [10] that $E^*\|\mathbb{G}_N\|_{\mathcal{F}_{\delta_N}} \to 0$ for every sequence $\delta_N \downarrow 0$. Thus, the result follows from Lemma A.2. Apply Markov's inequality to obtain $\|\mathbb{G}_N^{\pi}\|_{\mathcal{F}_{\delta_N}} = o_{P^*}(1)$. For the second statement, consider the expansion (A.3) of $\mathbb{G}_N^{\pi,mc}f - \mathbb{G}_N^{\pi}f$ with $f \in \mathcal{F}_{\delta_N}$. The first term is $o_{P^*}(1)$ by Lemma A.3. Since f converges to zero in $L_2(P_0)$, the second term is $o_{P^*}(1)$ by the dominated convergence theorem and Proposition A.1. Apply the triangle inequality to conclude $\left\|\mathbb{G}_N^{\pi,mc}\right\|_{\mathcal{F}_{\delta_N}} = o_{P^*}(1)$.

The proofs for $\mathbb{G}_N^{\pi,e}$, $\mathbb{G}_N^{\pi,c}$ and $\mathbb{G}_N^{\pi,cc}$ are similar. □

PROOF OF LEMMA 5.3. Without loss of generality, assume that $\hat{\theta}_N$ takes its values in $\Theta_\delta \equiv \{\theta \in \Theta : \|\theta - \theta_0\| < \delta\}$ because of consistency of $\hat{\theta}_N$ to θ_0. Define a function $f : \ell^\infty(\Theta_\delta \times \mathcal{H}) \times \Theta_\delta \mapsto \ell^\infty(\mathcal{H})$ by $f(z,\theta)h = z(\theta,h)$. Note that f is continuous at every point (z,θ_0) such that $\|z(\theta,h) - z(\theta_0,h)\|_{\mathcal{H}} \to 0$, as $\theta \to \theta_0$. To see this, suppose $z_N \to z$ and $\theta_N \to \theta_0$. Then, for a fixed $\epsilon > 0$, there exists n_0 such that $\|z_N - z\| < \epsilon$ and $\|\theta_N - \theta_0\| < \epsilon$ for $N \geq N_0$. For $N \geq N_0$, we have

$$\begin{aligned}
&\|f(z_N,\theta_N) - f(z,\theta_0)\|_{\mathcal{H}} \\
&\leq \|f(z_N,\theta_N) - f(z_0,\theta_N)\|_{\mathcal{H}} + \|f(z_0,\theta_N) - f(z_0,\theta_0)\| \\
&\leq \sup_{\theta\in\Theta_\delta, h\in\mathcal{H}} |z_N(\theta,h) - z(\theta,h)| + \|z(\theta_N,h) - z(\theta_0,h)\|_{\mathcal{H}} \\
&< 2\epsilon.
\end{aligned}$$

Define a stochastic process $\mathbb{Z}_N$ indexed by $\Theta_\delta \times \mathcal{H}$ by

$$\mathbb{Z}_N(\theta,h) = \mathbb{G}_N^{\pi}\left(\psi_{\theta,h} - \psi_{\theta_0,h}\right).$$

Because $\{\psi_{\theta,h} - \psi_{\theta_0,h} : \|\theta - \theta_0\| < \delta, \theta \in \Theta, h \in \mathcal{H}\}$ is Donsker, Theorem 5.3 implies that the sequence $\mathbb{Z}_N$ converges in $\ell^\infty(\Theta_\delta \times \mathcal{H})$ to a tight Gaussian process $\mathbb{Z}$ given by

$$\mathbb{Z} = \mathbb{G} + \sum_{j=1}^{J} \sqrt{\nu_j} \sqrt{\frac{1-p_j}{p_j}} \mathbb{G}_j .$$

This process has continuous sample paths with respect to the semimetric ρ given by

$$\rho^2\left((\theta_1, h_1), (\theta_2, h_2)\right) = P\left(\psi_{\theta_1,h_1} - \psi_{\theta_0,h_1} - \psi_{\theta_2,h_2} + \psi_{\theta_0,h_2}\right)^2$$

because $(\Theta_\delta \times \mathcal{H}, \rho)$ is totally bounded and $\mathbb{Z}$ is uniformly ρ-continuous. To see the latter, note that

$$\rho^2\left((\theta_1, h_1), (\theta_2, h_2)\right) \geq P\left\{ \left(\psi_{\theta_1,h_1} - \psi_{\theta_0,h_1} - \psi_{\theta_2,h_2} + \psi_{\theta_0,h_2}\right)^2 \middle| V \in \mathcal{V}_j \right\} \nu_j$$

for each $j = 1, \ldots, J$. By assumption

$$\sup_{h\in\mathcal{H}} \rho^2\left((\theta, h), (\theta_0, h)\right) = \sup_{h\in\mathcal{H}} P\left(\psi_{\theta,h} - \psi_{\theta_0,h} + 0\right)^2 \to 0,$$

as $\theta \to \theta_0$. Thus, f is continuous at almost all sample paths of $\mathbb{Z}$.

By Slutsky's theorem, $(\mathbb{Z}_N, \hat{\theta}_N) \rightsquigarrow (\mathbb{Z}, \theta_0)$. By the continuous mapping theorem, $\mathbb{Z}_N(\hat{\theta}_N) = f(\mathbb{Z}_N, \hat{\theta}_N) \rightsquigarrow f(\mathbb{Z}, \theta_0) = 0$ in $\ell^\infty(\mathcal{H})$.

The other cases for $\mathbb{G}_N^{\pi,e}$, $\mathbb{G}_N^{\pi,c}$, $\mathbb{G}_N^{\pi,mc}$ and $\mathbb{G}_N^{\pi,cc}$ follow analogously; see the proof of Theorem 5.3. □

With the results of Section 5 in hand, we are ready to prove the main theorems.

PROOF OF THEOREM 3.1. The asymptotic distributions of $\hat{\theta}_N$ is derived in [2]. Here we derive the asymptotic distribution of $\hat{\theta}_{N,mc}$ that is a solution of the calibrated weighted likelihood equations with modified calibration

$$\Psi^\pi_{N,1,mc}(\theta, \eta, \alpha) = \mathbb{P}^\pi_N G_{mc}(V; \alpha) \dot{\ell}_{\theta,\eta} = 0,$$
$$\Psi^\pi_{N,2,mc}(\theta, \eta, \alpha) h = \mathbb{P}^\pi_N G_{mc}(V; \alpha)(B_{\theta,\eta} h - P_{\theta,\eta} B_{\theta,\eta} h) = 0,$$

for all $h \in \mathcal{H}$ with $\alpha = \hat{\alpha}_N$. Let $\Psi_{mc}(\theta, \eta, \alpha) = (\Psi_{1,mc}(\theta, \eta, \alpha), \Psi_{2,mc}(\theta, \eta, \alpha))$

$$\Psi_{1,mc}(\theta, \eta, \alpha) = P_0 G_{mc}(V; \alpha) \dot{\ell}_{\theta,\eta},$$
$$\Psi_{2,mc}(\theta, \eta, \alpha) = P_0 G_{mc}(V; \alpha)(B_{\theta,\eta} h - P_{\theta,\eta} B_{\theta,\eta} h).$$

The derivative map of Ψ_{mc} with respect to (θ, η) at $(\theta_0, \eta_0, \alpha)$ has components $P_0\{G_{mc}(V;\alpha)\dot{\psi}_{ij,\theta_0,\eta_0,h}\}, i,j = 1,2$.

Our proof proceed by verifying the conditions of Theorem 1 of [3]. The weak convergence of $\sqrt{N}(\Psi_{N,j,mc} - \Psi_{j,mc})(\theta_0, \eta_0, \alpha_0)$ follows from Theorem 5.3. The asymptotic equicontinuity conditions

$$\sup_{\theta\in\Theta,\eta\in H} \left\| \sqrt{N}(\Psi^{\pi}_{N,j,mc} - \Psi_{j,mc})(\theta,\eta,\hat{\alpha}_N) - \sqrt{N}(\Psi^{\pi}_{N,j,mc} - \Psi_{j,mc})(\theta,\eta,\alpha_0)\right\|_{\mathcal{H}} = o_{P^*}(1),$$

for $j = 1, 2$, follows from Lemma A.3. The other asymptotic equicontinuity condition

$$\left\| \sqrt{N}(\Psi^{\pi}_{N,j,mc} - \Psi_{j,mc})(\hat{\theta}_{N,mc},\hat{\eta}_{N,mc},\alpha_0) - \sqrt{N}(\Psi^{\pi}_{N,j,mc} - \Psi_{j,mc})(\theta_0,\eta_0,\alpha_0)\right\|_{\mathcal{H}} = o_{P^*}(1),$$

for $j = 1, 2$, follows from the Condition 3.4 and Lemma 5.3. Thus conditions (2) and (3) of [3] are satisfied.

The Fréchet differentiability of the map $(\theta, \eta) \mapsto \Phi_{j,mc}(\theta, \eta, \alpha)$ uniformly over the neighborhood of α_0 follows by the Condition 3.5 and boundedness of G;

$$\begin{aligned}
&\left\|\Psi_{mc}(\theta,\eta,\alpha)h - \Psi_{mc}(\theta_0,\eta_0,\alpha)h - \dot{\Psi}_{mc}((\theta,\eta)-(\theta_0,\eta_0))\right\|_H \\
&= \sup_{h\in\mathcal{H}} \left| E\left\{ G_{mc}(V;\alpha)\left(\psi_{\theta,\eta,h} - \psi_{\theta_0,\eta_0,h} - \dot{\psi}_{\theta_0,\eta_0,h}((\theta,\eta)-(\theta_0,\eta_0))\right)\right\}\right| \\
&\le \left\{EG^2_{mc}(V;\alpha)\right\}^{1/2} \sup_{h\in\mathcal{H}} \left[E\left\{\psi_{\theta,\eta,h} - \psi_{\theta_0,\eta_0,h} - \dot{\psi}_{\theta_0,\eta_0,h}((\theta,\eta)-(\theta_0,\eta_0))\right\}^2\right]^{1/2} \\
&= o_{P^*}\left(\|(\theta,\eta)-(\theta_0,\eta_0)\|\right).
\end{aligned}$$

The Fréchet derivative $\dot{\Psi}_{\alpha,mc}$ of the map $\alpha \mapsto \{\Psi_{mc}(\theta,\eta,\alpha)h : h \in \mathcal{H}\}$ is

$$\frac{\partial}{\partial\alpha}\Psi_{mc}(\theta,\eta,\alpha)h = \frac{\partial}{\partial\alpha}E\left[G_{mc}(V;\alpha)\psi_{\theta,\eta,h}\right] = E\left[\frac{1-\pi_0(V)}{\pi_0(V)}Z^T\dot{G}_{mc}(V;\alpha)\psi_{\theta,\eta,h}\right].$$

Now proceed in the same way as [3] to obtain

$$\begin{aligned}
&\sqrt{N}(\hat{\theta}_{N,mc} - \theta_0) \\
&= \sqrt{N}(\hat{\theta}_N - \theta_0) + E\left[\tilde{\ell}_{\theta_0,\eta_0}\frac{1-\pi_0(V)}{\pi_0(V)}Z^T\dot{G}(0)\right]\sqrt{N}(\hat{\alpha}_N - \alpha_0) + o_{P^*}(1).
\end{aligned}$$

Because $\sqrt{N}(\hat{\theta}_N - \theta_0) = \mathbb{G}^{\pi}_N\tilde{\ell}_{\theta_0,\eta_0} + o_{P^*}(1)$ ((16) of [2]), it follows from (A.3) and consistency and asymptotic normality of $\hat{\alpha}_N$ that $\sqrt{N}(\hat{\theta}_{N,mc} - \theta_0) = \mathbb{G}^{\pi,mc}_N\tilde{\ell}_{\theta_0,\eta_0} + o_{P^*}(1)$. Apply Theorem 5.3 to complete the proof.

The other three cases are similar. □

LEMMA A.4. *Let* $\mathbb{Z}_1, \mathbb{Z}_2, \ldots$ *be i.i.d. stochastic processes indexed by* $\mathcal{F}_N$ *with* $E^*\|\mathbb{Z}_1\|_{\mathcal{F}_N}$ *uniformly bounded in* N. *Suppose that* $\|\mathbb{S}_N\|_{\mathcal{F}_N} \equiv \|\sum_{i=1}^N \mathbb{Z}_i\|_{\mathcal{F}_N} = o_{P^*}(1)$. *Then*

$$E^*\|\mathbb{S}_N\|_{\mathcal{F}_N} \to 0, \quad N \to \infty.$$

PROOF. Fix $\epsilon > 0$. Let $\mathbb{Y}_i$ be independent copies of $\mathbb{Z}_i$ and define $\mathbb{T}_N = \sum_{i=1}^N \mathbb{Y}_i$, and $\mathbb{U}_N = \mathbb{T}_N - \mathbb{S}_N$. Since $\|\mathbb{U}_N\|_{\mathcal{F}_N} = o_{P^*}(1)$, $\limsup_N P(\|\mathbb{U}_N\|_{\mathcal{F}_N} \geq x\sqrt{N}) \leq \limsup_N P(\|\mathbb{U}_N\|_{\mathcal{F}_N} \geq x) = 0$ by the portmanteau theorem. This implies that there exists N_0 such that for $N \geq N_0$

$$P^*(\|\mathbb{U}_N\|_{\mathcal{F}_N} > x\sqrt{N}) \leq \epsilon/x^2.$$

Since $\mathbb{U}_N$ is a sum of independent symmetric processes, we can apply Lévy's inequality to obtain

$$P^*\left(\max_{1\leq i\leq n}\|\mathbb{Z}_i - \mathbb{Y}_i\|_{\mathcal{F}_N} > x\sqrt{N}\right) \leq 2P^*(\|\mathbb{U}_N\|_{\mathcal{F}_N} > x\sqrt{N}) \leq 2\epsilon/x^2.$$

In view of Problem 2.3.2 of [10], for every $N \geq N_0$,

$$x^2 N P^*(\|\mathbb{Z}_1 - \mathbb{Y}_1\|_{\mathcal{F}_N} > x\sqrt{N}) \leq 4\epsilon.$$

Note that on the event that $\|\mathbb{Z}_1\|_{\mathcal{F}_N} > x$, we have

$$\beta_N(x) \equiv P_Y^*(\|\mathbb{Y}_1\|_{\mathcal{F}_N} < x/2) \leq P_Y^*(\|\mathbb{Z}_1 - \mathbb{Y}_1\|_{\mathcal{F}_N} > x/2).$$

Integrating both sides with respect to $\mathbb{Z}$ gives

$$\beta_N(x) P^*(\|\mathbb{Z}_1\|_{\mathcal{F}_N} > x) \leq P^*(\|\mathbb{Z}_1 - \mathbb{Y}_1\|_{\mathcal{F}_N} > x/2).$$

By Markov's inequality,

$$\beta_N(x) = 1 - P^*(\|\mathbb{Y}_1\|_{\mathcal{F}_N} \geq x/2) \geq 1 - 2x^{-1}E\|\mathbb{Y}_1\|_{\mathcal{F}_N}$$

Since $E\|\mathbb{Y}_1\|_{\mathcal{F}_N}$ is uniformly bounded in N, it follows that, for x sufficiently large, $\beta_N(x)^{-1}$ is uniformly bounded in N and, therefore, $P^*(\|\mathbb{Z}_1\|_{\mathcal{F}_N} > x\sqrt{N})$ is bounded by $P^*(\|\mathbb{Z}_1 - \mathbb{Y}_1\|_{\mathcal{F}_N} > x\sqrt{N})$ up to some constant for every N. Hence this proves that $P^*(\|\mathbb{Z}_1\|_{\mathcal{F}_N} > x) = o(x^{-2})$.

Now we apply the Hoffmann-Jørgensen inequality to obtain

$$E^*\|\mathbb{S}_N\|_{\mathcal{F}_N} \lesssim E^* \max_{i\leq N}\|\mathbb{Z}_i\|_{\mathcal{F}_N} + G_N^{-1}(u)$$

for an absolute constant u where

$$G_N(t) = P^*(\|\mathbb{S}_N\|_{\mathcal{F}_N} \leq t).$$

Since $P^*(\|\mathbb{Z}_1\|_{\mathcal{F}_N} > x) = o(x^{-2})$, $E^* \max_{i\leq N}\|\mathbb{Z}_i\|_{\mathcal{F}_N} \to 0$ in view of Problem 2.3.3 of [10]. The second term goes to zero since $\|\mathbb{S}_N\|_{\mathcal{F}_N} = o_{P^*}(1)$. This completes the proof. □

PROOF OF LEMMA 5.4. Define $\mathcal{G}_N = \{N^{-1/2}f : f \in \mathcal{F}_N\}$. We apply Lemma A.4 with $\mathbb{Z}_i$ and $\mathcal{F}_N$ in Lemma A.4 replaced by $\delta_{X_i} - P_0$ and $\mathcal{G}_N$, respectively. The uniform boundedness condition of Lemma A.4 is satisfied, because $E^*\|\delta_{X_1} - P_0\|_{\mathcal{F}_N} < \infty$ for $N \geq N_0$, and this expectation is decreasing in $N \geq N_0$. Thus, $E^*\|\mathbb{G}_N\|_{\mathcal{F}_N} = E^*\|\sum_{i=1}^N(\delta_{X_i} - P)\|_{\mathcal{G}_N} \to 0$. Apply Lemma A.2, and Markov's inequality to obtain $\|\mathbb{G}_N^\pi\|_{\mathcal{F}_N} = o_{P^*}(1)$.

For the IPW process with modified calibration, consider the expansion (A.3) of $(\mathbb{G}_N^{\pi,mc} - \mathbb{G}_N^\pi)f$. Then the first term is $o_{P^*}(1)$ by Lemma A.3. Suppose that $f = f_N \in \mathcal{F}_N$ converges to zero pointwise. Since $(\pi_0(V)^{-1} - 1)Z\dot{G}_{mc}$ is bounded, the second term in the expansion (A.3) is $o_{P^*}(1)$ by the dominated convergence theorem and Proposition A.1. Suppose instead that $f = f_N \in \mathcal{F}_N$ converges to zero in $L_1(P_0)$. Then the same conclusion that the second term in the expansion (A.3) is $o_{P^*}(1)$ follows directly. Apply the triangle inequality to conclude $\left\|\mathbb{G}_N^{\pi,mc}\right\|_{\mathcal{F}_{\delta_N}} = o_{P^*}(1)$.

The proofs for $\mathbb{G}_N^{\pi,e}$, $\mathbb{G}_N^{\pi,c}$ and $\mathbb{G}_N^{\pi,cc}$ are similar. □

PROOF OF THEOREM 3.2. We only consider the WLE with modified calibration, $\hat{\theta}_{N,mc}$. The other four cases are similar.

We evaluate the stochastic order of $\sqrt{N}\mathbb{P}_N^{\pi,mc}\dot{\ell}_{\theta_0,\eta_0} + \sqrt{N}P_0\dot{\ell}_{\hat{\theta}_{N,mc},\hat{\eta}_{N,mc}}$. Because $\mathbb{P}_N^{\pi,mc}\dot{\ell}_{\hat{\theta}_{N,mc},\hat{\eta}_{N,mc}} = o_{P^*}(N^{-1/2})$ by assumption and $P_0\dot{\ell}_{\theta_0,\eta_0} = 0$, we have $\sqrt{N}\mathbb{P}_N^{\pi,mc}\dot{\ell}_{\theta_0,\eta_0} + \sqrt{N}P_0\dot{\ell}_{\hat{\theta}_{N,mc},\hat{\eta}_{N,mc}} = -\mathbb{G}_N^{\pi,mc}(\dot{\ell}_{\hat{\theta}_{N,mc},\hat{\eta}_{N,mc}} - \dot{\ell}_{\theta_0,\eta_0}) + o_{P^*}(1)$. Let $\delta_N \downarrow 0$ be arbitrary and define $\mathcal{F}_N \equiv \{\dot{\ell}_{\theta,\eta} - \dot{\ell}_{\theta_0,\eta_0} : |\theta - \theta_0| \leq \delta_N, \|\eta - \eta_0\| \leq N^{-\beta}\}$. Then $f \in \mathcal{F}_N$ converges to zero either pointwise pointwise or in $L_1(P_0)$ by Condition 3.8 as $N \to \infty$. Moreover, it follows from Condition 3.8 that $\|\mathbb{G}_N\|_{\mathcal{F}_N} = o_{P^*}(1)$ and that there exists some N_0 that $\mathcal{F}_N$ is Glivenko-Cantelli for $N \geq N_0$. Apply Lemma 5.4 to obtain $\|\mathbb{G}_N^{\pi,mc}\|_{\mathcal{F}_N} = o_{P^*}(1)$ and conclude $\sqrt{N}\mathbb{P}_N^{\pi,mc}\dot{\ell}_{\theta_0,\eta_0} + \sqrt{N}P_0\dot{\ell}_{\hat{\theta}_{N,mc},\hat{\eta}_{N,mc}} = o_{P^*}(1)$. Similarly, $\sqrt{N}\mathbb{P}_N^{\pi,mc}B_{\theta_0,\eta_0}[\underline{h}^*] + \sqrt{N}P_0B_{\hat{\theta}_{N,mc},\hat{\eta}_{N,mc}}[\underline{h}^*] = o_{P^*}(1)$. These stochastic orders and Condition 3.9 imply that

$$
\begin{aligned}
&P_0\left\{-\dot{\ell}_{\theta_0,\eta_0}(\dot{\ell}_{\theta_0,\eta_0}^T(\hat{\theta}_{N,mc} - \theta_0) + B_{\theta_0,\eta_0}[\hat{\eta}_{N,mc} - \eta_0])\right\} \\
&\quad + o\left(|\hat{\theta}_{N,mc} - \theta_0|\right) + O\left(\|\hat{\eta}_{N,mc} - \eta_0\|^\alpha\right) + \mathbb{P}_N^{\pi,mc}\dot{\ell}_{\theta_0,\eta_0} \\
&= P_0\{-\dot{\ell}_{\theta_0,\eta_0}(\dot{\ell}_{\theta_0,\eta_0}^T(\hat{\theta}_{N,mc} - \theta_0) + B_{\theta_0,\eta_0}[\hat{\eta}_{N,mc} - \eta_0]) - \dot{\ell}_{\hat{\theta}_{N,mc},\hat{\eta}_{N,mc}} + \dot{\ell}_{\theta_0,\eta_0}\} \\
&\quad + o\left(|\hat{\theta}_{N,mc} - \theta_0|\right) + O\left(\|\hat{\eta}_{N,mc} - \eta_0\|^\alpha\right) + P_0\dot{\ell}_{\hat{\theta}_{N,mc},\hat{\eta}_{N,mc}} + \mathbb{P}_N^{\pi,mc}\dot{\ell}_{\theta_0,\eta_0} \\
&= o_{P^*}(N^{-1/2}), \qquad \text{(A.4)}
\end{aligned}
$$

and, furthermore, that

$$
\begin{aligned}
&P_0 \left\{ -B_{\theta_0,\eta_0}\left[\underline{h}^*\right] (\dot{\ell}^T_{\theta_0,\eta_0}(\hat{\theta}_{N,mc} - \theta_0) + B_{\theta_0,\eta_0}[\hat{\eta}_{N,mc} - \eta_0]) \right\} \\
&\quad + o\left(|\hat{\theta}_{N,mc} - \theta_0|\right) + O\left(\|\hat{\eta}_{N,mc} - \eta_0\|^{\alpha}\right) + \mathbb{P}_N^{\pi,mc} B_{\theta_0,\eta_0}\left[\underline{h}^*\right] \\
&= o_{P^*}(N^{-1/2}). \qquad \text{(A.5)}
\end{aligned}
$$

By Condition 3.6 and $\alpha\beta > 1/2$, $\sqrt{N} O_{P^*}(\|\hat{\eta}_N - \eta_0\|^{\alpha}) = o_{P^*}(1)$. So by Condition 3.7 and taking the difference of (A.4) and (A.5), we have

$$
\begin{aligned}
&-P_0 \left(\left\{ \dot{\ell}_{\theta_0,\eta_0} - B_{\theta_0,\eta_0}\left[\underline{h}^*\right] \right\} \dot{\ell}^T_{\theta_0,\eta_0} \right) \left(\hat{\theta}_{N,mc} - \theta_0\right) + \ o\left(|\hat{\theta}_{N,mc} - \theta_0|\right) \\
&\quad + o_P(N^{-1/2}) - o_P(N^{-1/2}) + \mathbb{P}_N^{\pi,mc}\left(\dot{\ell}_{\theta_0,\eta_0} - B_{\theta_0,\eta_0}\left[\underline{h}^*\right] \right) \\
&= o_P(N^{-1/2}) - o_P(N^{-1/2}),
\end{aligned}
$$

or

$$
-I_0(\hat{\theta}_{N,mc} - \theta_0) = \mathbb{P}_N^{\pi,mc}\left(\dot{\ell}_{\theta_0,\eta_0} - B_{\theta_0,\eta_0}\left[\underline{h}^*\right] \right) + o_{P^*}(N^{-1/2}).
$$

It follows by the invertibility of I_0 that

$$
\sqrt{N}\left(\hat{\theta}_{N,mc} - \theta_0\right) = -\sqrt{N}\mathbb{P}_N^{\pi,mc} I_0^{-1}\left(\dot{\ell}_{\theta_0,\eta_0} - B_{\theta_0,\eta_0}\left[\underline{h}^*\right] \right) + o_P(1).
$$

Now, we recognize that the summand inside $\mathbb{P}_N^{\pi,mc}$ is the efficient influence function for θ and apply Theorem 5.3. ☐

PROOF OF THEOREM 3.3. Theorem 3.1 for cases for $\hat{\theta}_N^{Bern}$ and $\hat{\theta}_{N,e}^{Bern}$ are proved in [2; 3]. We only consider the WLE with modified calibration, $\hat{\theta}_{N,mc}$. The other four estimators for both theorems are similar.

Under stratified Bernoulli sampling, independence of sampling indicators allows us to proceed in the same as in the proofs of Theorems 3.1 and 3.2 to conclude $\sqrt{N}(\hat{\theta}_{N,mc}^{Bern} - \theta_0) = \sqrt{N}\mathbb{P}_N^{\pi,mc}\tilde{\ell}_0 + o_{P^*}(1)$ and asymptotic linearity of $\hat{\alpha}_N$ in Proposition A.1. In view of (A.3), $\sqrt{N}(\hat{\theta}_{N,mc}^{Bern} - \theta_0) = \sqrt{N}\mathbb{P}_N f + o_{P^*}(1)$ where

$$
f(X, V, \xi) = \frac{\xi}{\pi_0(V)}\tilde{\ell}_0 - \frac{\xi - \pi_0(V)}{\pi_0(V)} Q_{mc}\tilde{\ell}_0. \qquad \text{(A.6)}
$$

Apply the central limit theorem and compute

$$
\begin{aligned}
\Sigma_{mc}^{Bern} &= \mathrm{Var}(f) \\
&= \mathrm{Var}\left(E\left[\frac{\xi}{\pi_0(V)}\tilde{\ell}_0 - \frac{\xi-\pi_0(V)}{\pi_0(V)}Q_{mc}\tilde{\ell}_0 \middle| X,V\right]\right) \\
&\quad + E\left[\mathrm{Var}\left(\frac{\xi}{\pi_0(V)}\tilde{\ell}_0 - \frac{\xi-\pi_0(V)}{\pi_0(V)}Q_{mc}\tilde{\ell}_0 \middle| X,V\right)\right] \\
&= \mathrm{Var}(\tilde{\ell}_0) + E\left[\mathrm{Var}\left(\frac{\xi}{\pi_0(V)}(I-Q_{mc})\tilde{\ell}_0 \middle| X,V\right)\right] \\
&= I_0^{-1} + E\left[\frac{1-\pi_0(V)}{\pi_0(V)}\{(I-Q_{mc})\tilde{\ell}_0\}^{\otimes 2}\right] \\
&= I_0^{-1} + \sum_{j=1}^{J}\nu_j\frac{1-p_j}{p_j}P_{0|j}\{(I-Q_{mc})\tilde{\ell}_0\}^{\otimes 2}.
\end{aligned}
$$

□

Proof of Corollary 3.2. We only consider the WLE with modified calibration, $\hat{\theta}_{N,mc}$. The other two cases are similar.

Let $Q_{mc}\tilde{\ell}_0 \equiv AZ$ where $A = A_1A_2$ with $A_1 \equiv P_0[(\pi_0^{-1}(V)-1)\tilde{\ell}_0 Z^T]$ and $A_2 \equiv \{P_0[(\pi_0^{-1}(V)-1)Z^{\otimes 2}]\}^{-1}$. Recall that $\Sigma^{Bern} = \mathrm{Var}\{(\xi/\pi_0(V))\tilde{\ell}_0\}$. In view of (A.6), it suffices to show that $\mathrm{Cov}\{(\xi/\pi_0(V))\tilde{\ell}_0, (\xi/\pi_0(V)-1)AZ\}$ is equal to $\mathrm{Var}\left((\xi/\pi_0(V)-1)AZ\right)$. This is true since

$$
\begin{aligned}
&\mathrm{Cov}\left\{\frac{\xi}{\pi_0(V)}\tilde{\ell}_0, \frac{\xi-\pi_0(V)}{\pi_0(V)}AZ\right\} = E\left\{\frac{\xi}{\pi_0(V)}\tilde{\ell}_0\frac{\xi-\pi_0(V)}{\pi_0(V)}Z\right\}A^T \\
&= E\left[\tilde{\ell}_0 ZE\left\{\frac{\xi}{\pi_0(V)}\frac{\xi-\pi_0(V)}{\pi_0(V)}\middle| X,V\right\}\right]A^T \\
&= E\left[\frac{1-\pi_0(V)}{\pi_0(V)}\tilde{\ell}_0 Z\right]A^T = A_1A_2A_1^T,
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Var}\left(\frac{\xi-\pi_0(V)}{\pi_0(V)}AZ\right) &= A\mathrm{Var}\left(\frac{\xi-\pi_0(V)}{\pi_0(V)}Z\right)A^T \\
&= AE\left[\mathrm{Var}\left(\frac{\xi-\pi_0(V)}{\pi_0(V)}Z\middle| X,V\right)\right]A^T \\
&\quad + A\mathrm{Var}\left(ZE\left[\frac{\xi-\pi_0(V)}{\pi_0(V)}\middle| X,V\right]\right)A^T \\
&= AE\left[Z^{\otimes 2}\frac{1-\pi_0(V)}{\pi_0(V)}\right]A^T + 0 = A_1A_2A_1^T.
\end{aligned}
$$

□

PROOF OF COROLLARY 3.3. (1). We first consider stratified Bernoulli sampling. The case for $\hat{\theta}_{N,c}$ was proved in [1]. We only consider the WLE with modified calibration, $\hat{\theta}_{N,mc}$. The other two cases, $\hat{\theta}_{N,e}$ and $\hat{\theta}_{N,cc}$, are similar.

For $\tilde{Z} \equiv (Z^{(1)}, \dots, Z^{(J)})^T$ with $Z^{(j)} \equiv I(V \in \mathcal{V}_j)Z^T$, we compute $\tilde{A}_1 \equiv P_0[(\pi_0^{-1}(V)-1)\tilde{\ell}_0\tilde{Z}^T]$ and $\tilde{A}_2 \equiv \{P_0[(\pi_0^{-1}(V)-1)\tilde{Z}^{\otimes 2}]\}^{-1}$. Note that $Q_{mc}\tilde{\ell}_0 = \tilde{A}_1\tilde{A}_2\tilde{Z}$. The matrix $\tilde{A}_1 = [\tilde{A}_{1,1}, \dots, \tilde{A}_{1,J}]$ is a partitioned matrix where

$$\tilde{A}_{1,j} \equiv P_0\left(\frac{1-\pi_0(V)}{\pi_0(V)}\tilde{\ell}_0 Z^{(j)}\right) = \nu_j P_{0|j}\left(\frac{1-p_j}{p_j}\tilde{\ell}_0 Z^T\right) \in \mathbb{R}^{p\times k}.$$

and the matrix $\tilde{A}_2$ is the block diagonal matrix the jth block of which is

$$\tilde{A}_{2,j} \equiv \left\{P_0\frac{1-\pi_0(V)}{\pi_0(V)}[(Z^{(j)})^T]^{\otimes 2}\right\}^{-1} = \left\{\nu_j P_{0|j}\frac{1-p_j}{p_j}Z^{\otimes 2}\right\}^{-1} \in \mathbb{R}^{k\times k}.$$

Thus, the matrix $\tilde{A} \equiv \tilde{A}_1\tilde{A}_2$ is a partitioned matrix $\tilde{A} = [\tilde{A}_1, \dots, \tilde{A}_J]$ where

$$\tilde{A}_j = \tilde{A}_{1,j}\tilde{A}_{2,j} = P_{0|j}\left(\tilde{\ell}_0 Z^T\right)\left\{P_{0|j}Z^{\otimes 2}\right\}^{-1}.$$

It follows by the definition of the $Z^{(j)}$'s that

$$\begin{aligned} P_{0|j}\left\{(I-Q_{mc})\tilde{\ell}_0\right\}^{\otimes 2} &= P_{0|j}\left\{\tilde{\ell}_0 - \tilde{A}\tilde{Z}\right\}^{\otimes 2} \\ &= P_{0|j}\left\{\tilde{\ell}_0 - \tilde{A}_j Z\right\}^{\otimes 2} = P_{0|j}\left\{(I-Q_c^{(j)})\tilde{\ell}_0\right\}^{\otimes 2}. \end{aligned}$$

Since

$$P_{0|j}\left(\tilde{A}_j Z\right)^{\otimes 2} = \tilde{A}_j P_{0|j}Z^{\otimes 2}\tilde{A}_j^T = P_{0|j}\left(\tilde{\ell}_0 Z^T\right)\left\{P_{0|j}Z^{\otimes 2}\right\}^{-1} P_{0|j}\left(\tilde{\ell}_0 Z^T\right)^T,$$

and

$$P_{0|j}\left(\tilde{\ell}_0 Z^T\right)\tilde{A}_j^T = P_{0|j}\left(\tilde{\ell}_0 Z^T\right)\left\{P_{0|j}Z^{\otimes 2}\right\}^{-1} P_{0|j}\left(\tilde{\ell}_0 Z^T\right)^T,$$

it follows that

$$P_{0|j}\left\{(I-Q_c^{(j)})\tilde{\ell}_0\right\}^{\otimes 2} = P_{0|j}\tilde{\ell}_0^{\otimes 2} - P_{0|j}\{Q_c^{(j)}\tilde{\ell}_0\}^{\otimes 2}.$$

Substitution of this into (3.11) gives (3.12).

(2). Next, we consider the second part of Corollary 3.3 concerning stratified sampling without replacement. For $\tilde{Z} \equiv (Z^{(1)}, \ldots, Z^{(J)})^T$ with $Z^{(j)} \equiv I(V \in \mathcal{V}_j) Z^T$, we compute $\tilde{B}_1 \equiv P_0[(\pi_0^{-1}(V) - 1)\tilde{\ell}_0(\tilde{Z} - \mu_{\tilde{Z}})^T]$ and $\tilde{B}_2 \equiv \{P_0[(\pi_0^{-1}(V) - 1)(\tilde{Z} - \mu_{\tilde{Z}})^{\otimes 2}]\}^{-1}$. Note that $Q_{cc}\tilde{\ell}_0 = \tilde{B}_1\tilde{B}_2\tilde{Z}$ and $\mu_{\tilde{Z}} = (\mu_{Z,1}^T, \ldots, \mu_{Z,J}^T)^T$. The matrix $\tilde{B}_1 = [\tilde{B}_{1,1}, \ldots, \tilde{B}_{1,J}]$ is a partitioned matrix where

$$\tilde{B}_{1,j} \equiv P_0\left(\frac{1-\pi_0(V)}{\pi_0(V)}\tilde{\ell}_0(Z^{(j)} - \mu_{Z,j}^T)\right) = \nu_j P_{0|j}\left(\frac{1-p_j}{p_j}\tilde{\ell}_0(Z - \mu_{Z,j})^T\right).$$

and the matrix $\tilde{B}_2$ is the block diagonal matrix the jth block of which is

$$\tilde{B}_{2,j} \equiv \left\{P_0\frac{1-\pi_0(V)}{\pi_0(V)}[(Z^{(j)})^T - \mu_{Z,j}]^{\otimes 2}\right\}^{-1} = \left\{\nu_j P_{0|j}\frac{1-p_j}{p_j}(Z - \mu_{Z,j})^{\otimes 2}\right\}^{-1}.$$

Thus, the matrix $\tilde{B} \equiv \tilde{B}_1\tilde{B}_2$ is a partitioned matrix $\tilde{B} = [\tilde{B}_1, \ldots, \tilde{B}_J]$ where

$$\tilde{B}_j = \tilde{B}_{1,j}\tilde{B}_{2,j} = P_{0|j}\left(\tilde{\ell}_0(Z - \mu_{Z,j})^T\right)\left\{P_{0|j}(Z - \mu_{Z,j})^{\otimes 2}\right\}^{-1}.$$

It follows by the definition of $Z^{(j)}$'s that

$$\begin{aligned}
&\mathrm{Var}_{0|j}\left\{(I - Q_{cc})\tilde{\ell}_0\right\} = \mathrm{Var}_{0|j}\left\{\tilde{\ell}_0 - \tilde{B}(\tilde{Z} - \mu_{\tilde{Z}})\right\} \\
&= \mathrm{Var}_{0|j}\left\{\tilde{\ell}_0 - \tilde{B}_j(Z - \mu_{Z,j})\right\} = \mathrm{Var}_{0|j}\left\{(I - Q_{cc}^{(j)})\tilde{\ell}_0\right\}.
\end{aligned}$$

Then, since

$$\begin{aligned}
&\mathrm{Var}_{0|j}\left(\tilde{B}_j(Z - \mu_{Z,j})\right) = \tilde{B}_j \mathrm{Var}_{0|j}(Z)\tilde{B}_j^T \\
&= P_{0|j}\left(\tilde{\ell}_0(Z - \mu_{Z,j})^T\right)\left\{\mathrm{Var}_{0|j}(Z)\right\}^{-1} P_{0|j}\left(\tilde{\ell}_0(Z - \mu_{Z,j})^T\right)^T,
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{Cov}_{0|j}\left(\tilde{\ell}_0, \tilde{B}_j(Z - \mu_{Z,j})\right) = P_{0|j}\left(\tilde{\ell}_0(Z - \mu_{Z,j})^T\right)\tilde{B}_j^T \\
&= P_{0|j}\left(\tilde{\ell}_0(Z - \mu_{Z,j})^T\right)\left\{\mathrm{Var}_{0|j}(Z)\right\}^{-1} P_{0|j}\left(\tilde{\ell}_0(Z - \mu_{Z,j})^T\right)^T,
\end{aligned}$$

it follows that

$$\mathrm{Var}_{0|j}\left\{(I - Q_c^{(j)})\tilde{\ell}_0\right\} = \mathrm{Var}_{0|j}\left(\tilde{\ell}_0\right) - \mathrm{Var}_{0|j}\{Q_c^{(j)}\tilde{\ell}_0\}.$$

Substitution of this last identity into (3.8) gives (3.13). □

Now we give the proof of Theorem 4.1.

PROOF OF THEOREM 4.1. We only consider the WLE with modified calibration. Proofs for the other four estimators are similar. Our proof closely follows the consistency proof for the MLE for complete data in [8].

Because of the assumption on τ, we restrict our attention to the interval $[0,\tau]$. For a bounded function $h \in L_2(\Lambda)$, define a perturbation $d\hat{\Lambda}_{N,mc,t} = (1+th)d\hat{\Lambda}_{N,mc}$ of $\hat{\Lambda}_{N,mc}$. The weighted log likelihood with modified calibration, $\mathbb{P}_N^{\pi,mc}\ell_{\theta,\Lambda}$, evaluated at $(\hat{\theta}_{N,mc}, \hat{\Lambda}_{N,mc,t})$ viewed as a function of t is maximal at $t=0$ by the definition of the WLE with modified calibration. Thus, differentiating at $t=0$ yields $\mathbb{P}_N^{\pi,mc}B_{\hat{\theta}_{N,mc},\hat{\Lambda}_{N,mc}}h = 0$, or

$$\begin{aligned}\mathbb{P}_N^{\pi,mc}\Delta h(Y) &= \mathbb{P}_N^{\pi,mc}e^{\hat{\theta}_{N,mc}^T X}\int_{[0,Y]} h d\hat{\Lambda}_{N,mc} \\ &= \int \mathbb{P}_N^{\pi,mc}\left\{e^{\hat{\theta}_{N,mc}^T X} I_{[Y\geq s]}\right\} h(s)d\hat{\Lambda}_{N,mc}(s).\end{aligned}$$

Let $\hat{M}_{N,0}(s) = \mathbb{P}_N^{\pi,mc}e^{\hat{\theta}_{N,mc}^T X}I(Y\geq s)$. Replacing h in the above display by $h/\hat{M}_{N,0}$ yields

$$\hat{\Lambda}_{N,mc}h = \int \frac{h(s)}{\hat{M}_{N,0}(s)}\mathbb{P}_N^{\pi,mc}\left\{e^{\hat{\theta}_{N,mc}^T X}I(Y\geq s)\right\} d\hat{\Lambda}_{N,mc}(s) = \mathbb{P}_N^{\pi,mc}\frac{\Delta h(Y)}{\hat{M}_{N,0}(Y)}.$$

Similar reasoning via $P_0 B_0 h = 0$ leads to $\Lambda_0 h = P_0\Delta h(Y)/M_0(Y)$. Let $\tilde{\Lambda}_N h = \mathbb{P}_N^{\pi,mc}\Delta h(Y)/M_0(Y)$. Since $P(T>\tau)>0$ and $P(C=\tau)>0$, we have for $s\leq\tau$ that $M_0(s)\geq M_0(\tau)>0$. The function $(y,\delta)\mapsto \delta h(y)/M_0(y)$ is bounded, and hence $\{\delta h(y)/M_0(y) : h\in\mathcal{H}\}$ is Glivenko-Cantelli by the Glivenko-Cantelli preservation theorem [9] and the fact that $\mathcal{H}$ is Glivenko-Cantelli. Thus, $\|\tilde{\Lambda}_N\|_{\mathcal{H}} \to_{P^*} \|P_{\theta_0,\Lambda_0}\Delta h(Y)/M_0(Y)\|_{\mathcal{H}} = \|\Lambda_0\|_{\mathcal{H}}$. Moreover, since $\hat{\Lambda}_{N,mc}\{Y_i\} = \hat{\Lambda}_{N,mc}\delta_{Y_i} = N^{-1}(\xi_i/\pi_{\hat{\alpha}_N}(V_i))(\Delta_i/\hat{M}_{N,0}(Y_i))$, and similarly $\tilde{\Lambda}_N\{Y_i\} = N^{-1}(\xi_i/\pi_{\hat{\alpha}_N}(V_i))(\Delta_i/M_0(Y_i))$, we have $\hat{\Lambda}_{N,mc}\{Y_i\}/\tilde{\Lambda}_N\{Y_i\} = M_0(Y_i)/\hat{M}_{N,0}(Y_i)$.

Since the weighted log likelihood with modified calibration evaluated at $(\hat{\theta}_{N,mc},\hat{\Lambda}_{N,mc})$ is larger than at $(\theta_0,\tilde{\Lambda}_N)$, we have

$$\begin{aligned}0 &\leq \mathbb{P}_N^{\pi,mc}(\ell_{\hat{\theta}_{N,mc},\hat{\Lambda}_{N,mc}} - \ell_{\theta_0,\tilde{\Lambda}_N}) \\ &= (\hat{\theta}_{N,mc}-\theta_0)^T\mathbb{P}_N^{\pi,mc}\Delta X - \mathbb{P}_N^{\pi,mc}(e^{\hat{\theta}_{N,mc}^T X}\hat{\Lambda}_N(Y) - e^{\theta_0^T X}\tilde{\Lambda}_N(Y)) \\ &\quad + \mathbb{P}_N^{\pi,mc}\Delta\log\{M_0(Y)/\hat{M}_{N,0}(Y)\}.\end{aligned}$$

We take the limit of this on N. Because Θ is compact, there is a subsequence of $\{\hat{\theta}_N\}$ that converges to $\theta_\infty\in\Theta$. It follows by Theorem 5.1

that along the convergent subsequence of $\{\hat\theta_N\}$, $(\hat\theta_N-\theta_0)^T\mathbb{P}_N^{\pi,mc}\Delta X \to_{P^*} (\theta_\infty-\theta_0)^T P_{\theta_0,\Lambda_0}\Delta X$.

For the second term, note that $\hat\Lambda_N(\tau)$ is uniformly bounded, because $e^{\theta^T X}$ is uniformly bounded in θ and X, and $\hat\Lambda_N(\tau)\mathbb{P}_N^{\pi,mc}e^{\hat\theta_N^T X}I(Y=\tau)\le \mathbb{P}_N^{\pi,mc}e^{\hat\theta_{N,mc}^T X}\hat\Lambda_N(Y)=\mathbb{P}_N^{\pi,mc}\Delta\le 1$. Here we use the weighted likelihood equation with $h=1$ above. Since $\{\hat\Lambda_{N,mc}\}$ and $\{\tilde\Lambda_N\}$ are both subsets of the class of monotone, bounded cadlag functions that is Glivenko-Cantelli, it follows by the Glivenko-Cantelli preservation theorem [9] and Theorem 5.1 that

$$
\begin{aligned}
&\mathbb{P}_N^{\pi,mc}\{e^{\hat\theta_{N,mc}^T X}\hat\Lambda_N(Y)-e^{\theta_0^T X}\tilde\Lambda_N(Y)\}\\
&\quad = P_{\theta_0,\Lambda_0}\{e^{\theta_\infty^T X}\hat\Lambda_N(Y)-e^{\theta_0^T X}\tilde\Lambda_N(Y)\}+o_{P^*}(1),
\end{aligned}
\tag{A.7}
$$

along a subsequence of $\hat\theta_{N,mc}$.

For the third term, note that $\{\hat M_{N,0}\}$ is a subset of the class of monotone, bounded, cadlag functions, which is Glivenko-Cantelli, and hence so is it. Note also that $\hat M_{N,0}(\tau)=\mathbb{P}_N^{\pi,mc}e^{\hat\theta_{N,mc}^T X}I(Y=\tau)$ is bounded away from zero with probability tending to 1 since $P(T>\tau)>0$ and $P(C=\tau)>0$. Since $\hat M_{N,0}(t)\ge \hat M_{N,0}(\tau)$ for $t\le\tau$, the set $\{\delta\log(M_0(y)/\hat M_{N,0}(y))\}$ is Glivenko-Cantelli by the Glivenko-Cantelli preservation theorem again so that

$$
\begin{aligned}
&\mathbb{P}_N^{\pi,mc}\Delta\log(M_0(Y)/\hat M_{N,0}(Y))\\
&\quad = P_{\theta_0,\Lambda_0}\Delta\log(M_0(Y)/\hat M_N(Y))+o_{P^*}(1)
\end{aligned}
\tag{A.8}
$$

by Theorem 5.1.

The set $\{\delta h(y)/\hat M_{N,0}(y):h\in\mathcal{H}\}$ is Glivenko-Cantelli by the Glivenko-Cantelli preservation theorem [9] so that $\|\hat\Lambda_N\|_{\mathcal{H}}=\|P_{\theta_0,\Lambda_0}\Delta h(Y)/\hat M_{N,0}(Y)\|_{\mathcal{H}}+o_{P^*}(1)$ by Theorem 5.1. Since we have by Theorem 5.1 that

$$
\hat M_{N,0}(s)=\mathbb{P}_N^{\pi,mc}e^{\hat\theta_{N,mc}^T X}I(Y\ge s)\to_{P^*} P_{\theta_0,\Lambda_0}e^{\theta_\infty^T X}I(Y\ge s)\equiv M_{\infty,0}(s)
$$

uniformly in s, it follows by the dominated convergence theorem that

$$
\begin{aligned}
\|\hat\Lambda_N\|_{\mathcal{H}} &= \|P_{\theta_0,\Lambda_0}\Delta h(Y)/\hat M_{N,0}(Y)\|_{\mathcal{H}}+o_{P^*}(1)\\
&\to_{P^*} \|P_{\theta_0,\Lambda_0}\Delta h(Y)/M_{\infty,0}(Y)\|_{\mathcal{H}}\equiv\|\Lambda_\infty\|_{\mathcal{H}},
\end{aligned}
$$

along a subsequence of $\hat\theta_N$.

Apply the dominated convergence theorem to replace $\hat\Lambda_{N,mc}$, $\tilde\Lambda_N$, and $\hat M_{N,0}$ by Λ_∞, Λ_0 and $M_{\infty,0}$ in (A.7) and (A.8) and conclude

$$
\begin{aligned}
0 &\le (\theta_\infty-\theta_0)^T P_{\theta_0,\Lambda_0}\Delta X - P_{\theta_0,\Lambda_0}\left(e^{\theta_\infty^T X}\Lambda_\infty(Y)-e^{\theta_0^T X}\Lambda_0(Y)\right)\\
&\quad + P_{\theta_0,\Lambda_0}\Delta\log\{M_0(Y)/M_\infty(Y)\}.
\end{aligned}
\tag{A.9}
$$

Since $M_0/M_\infty = d\Lambda_\infty/d\Lambda_0$, (A.9) is in fact minus one times the Kullback-Leibler divergence

$$K(P_{\theta_0,\Lambda_0}, P_{\theta_\infty,\Lambda_\infty}) \equiv P_{\theta_0,\Lambda_0} \log\{p_{\theta_0,\Lambda_0}/p_{\theta_\infty,\Lambda_\infty}\} \geq 0,$$

for the complete data model. Thus, (A.9) is exactly zero. But since $K(P_{\theta_0,\Lambda_0}, P_{\theta,\Lambda})$ is strictly positive unless $(\theta,\Lambda) = (\theta_0,\Lambda_0)$ by the identifiability of parameters, we must have $(\theta_\infty,\Lambda_\infty) = (\theta_0,\Lambda_0)$. This is true for any subsequence of $\hat{\theta}_{N,mc}$, and the result follows. □

We give the characterization of WLE's for the Cox model with interval censoring. Let $n = \sum_{i=1}^N \xi_i$ be the number of observations sampled at phase II. Let $Y_{(1)},\ldots,Y_{(n)}$ be the order statistics of $Y_1,\ldots,Y_N$ with $\xi_i = 1, i = 1,\ldots,N$. Let $\Delta_{(i)}$, $X_{(i)}$, $U_{(i)}$, and $\xi_{(i)}$ correspond to $Y_{(i)}$; for example, if $Y_{(i)} = Y_j$, then $\Delta_{(i)} = \Delta_j$. Let $\pi_{(i)} = \pi_0(V_{(i)})$. Because only fully observed subjects contribute to the weighted likelihood, $\hat{\Lambda}_N(Y_i)$ for subjects with $\xi_i = 0$ does not matter in the maximization. In fact, $\hat{\Lambda}_N(Y_{(i)}) = \hat{\Lambda}_N(Y_{(i-1)})$ for subjects with $\xi_{(i)} = 0$ for $i \geq 2$. The WLE $\hat{\Lambda}_N$ of Λ corresponds to $\underline{x} = (\hat{\Lambda}_{(1)},\ldots,\hat{\Lambda}_{(N)})$ that maximizes

$$\phi(\theta,\underline{x}) = \sum_{i=1}^n \frac{1}{\pi_{(i)}}\left[\log\left\{1-\exp\left(-e^{\theta^T X_{(i)}}\right)x_i\right\} - (1-\Delta_{(i)})e^{\theta^T X_{(i)}}x_i\right]$$

at $\hat{\theta}_N$ subject to $0 \leq x_1 \leq \cdots \leq x_n$. The monotonicity constraint on $\underline{x}$ is imposed to guarantee that an estimate of Λ is nondecreasing. Note that $\phi(\theta,\underline{x})$ is concave in $\underline{x}$.

Without loss of generality, we can assume that $\Delta_{(1)} = 1$ and $\Delta_{(n)} = 0$. If $\Delta_{(1)} = 0$ or $\Delta_{(n)} = 1$, then $\hat{\Lambda}_N(Y_{(1)}) = 0$ or $\hat{\Lambda}_N(Y_{(n)}) = \infty$, so that the first or the last summand in ϕ is zero. Hence ignoring these terms does not change the maximization of the weighted likelihood.

LEMMA A.5. *Assume that* $\Delta_{(1)} = 1$ *and* $\Delta_{(n)} = 0$. *Then the WLE* $(\hat{\theta}_N,\hat{\Lambda}_N)$ *satisfies*

$$\mathbb{P}_N^\pi \hat{\Lambda}_N(Y)\exp(\hat{\theta}_N^T X)XQ(Y,\Delta,X,\hat{\theta}_N,\hat{\Lambda}_N(Y)) = 0,$$

$$\sum_{j\geq i}\frac{\xi_{(j)}}{\pi_{(j)}}Q(Y_{(j)},\Delta_{(j)},X_{(j)};\hat{\theta}_N,\hat{\Lambda}_N)\exp(\hat{\theta}_N^T X_{(j)}) \leq 0, \ \textit{for } i = 1,\ldots,n,$$

$$\mathbb{P}_N^\pi Q(Y,\Delta,X;\hat{\theta}_N,\hat{\Lambda}_N)\exp(\hat{\theta}_N^T X)\hat{\Lambda}_N(Y) = 0.$$

Moreover, the corresponding (in)equalities holds for the WLE's with estimated weights and (modified and centered) calibration.

PROOF. The first equation is simply the weighted score equation for θ.

For the second inequality, let 1_j be the vector which has 1's as its last j components and zeros as its first $n-j$ components. Let $\underline{\hat{\Lambda}}_N = (\hat{\Lambda}_N(Y_{(i)}))_{i=1}^n$. For $\epsilon > 0$, the vector $\underline{\hat{\Lambda}}_N + \epsilon 1_j$ satisfies the monotonicity constraint. It follows by the definition of the WLE that

$$0 \geq \lim_{\epsilon \downarrow 0} \frac{\phi(\hat{\theta}_N, \underline{\hat{\Lambda}}_N + \epsilon 1_j) - \phi(\hat{\theta}_N, \underline{\hat{\Lambda}}_N)}{\epsilon}$$
$$= \sum_{i=1}^n \frac{1}{\pi_{(i)}} \left[\Delta_{(i)} \frac{e^{-e^{\hat{\theta}_N^T X_{(i)}} \hat{\Lambda}_N(Y_{(i)}) + \hat{\theta}_N^T X_{(i)}}}{1 - e^{-e^{\hat{\theta}_N^T X_{(i)}} \hat{\Lambda}_N(Y_{(i)})}} - (1 - \Delta_{(i)}) e^{\hat{\theta}_N^T X_{(i)}} \right] I(i \geq j).$$

Relabeling i and j gives the desired result. Note that the assumption that $\Delta_{(1)} = 1$ and $\Delta_{(n)} = 0$ guarantees that the above derivative is finite.

The last equality follows for the same reason that

$$\lim_{h \to 0} \frac{\phi(\hat{\theta}_N, \underline{\hat{\Lambda}}_N + h\underline{\hat{\Lambda}}_N) - \phi(\hat{\theta}_N, \underline{\hat{\Lambda}}_N)}{h} = 0.$$

Note that adding terms associated with $\xi_i = 0$ does not contribute to the sum in the above derivative.

For the other four estimators, change weights appropriately. □

References.

[1] BRESLOW, N. E., LUMLEY, T., BALLANTYNE, C., CHAMBLESS, L. and KULICH, M. (2009). Improved Horvitz-Thompson estimation of model parameters from two-phase stratified samples: applications in epidemiology. *Stat. Biosc.* **1** 32-49.

[2] BRESLOW, N. E. and WELLNER, J. A. (2007). Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. *Scand. J. Statist.* **34** 86–102. MR2325244

[3] BRESLOW, N. E. and WELLNER, J. A. (2008). A Z-theorem with estimated nuisance parameters and correction note for: "Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression" [Scand. J. Statist. **34** (2007), no. 1, 86–102; MR2325244]. *Scand. J. Statist.* **35** 186–192. MR2391566

[4] MURPHY, S. A. and VAN DER VAART, A. W. (1999). Observed information in semiparametric models. *Bernoulli* **5** 381–412. MR1693616 (2000c:62040)

[5] SAEGUSA, T. and WELLNER, J. A. (2011). Weighted likelihood estimation under two-phase sampling Technical Report. available at arXiv:1112.4951.

[6] SAEGUSA, T. and WELLNER, J. A. (2012). Weighted likelihood estimation under two-phase sampling *Ann. Statist.* submitted.

[7] VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge University Press, Cambridge. MR1652247 (2000c:62003)

[8] van der Vaart, A. (2002). Semiparametric statistics. In *Lectures on probability theory and statistics (Saint-Flour, 1999). Lecture Notes in Math.* **1781** 331–457. Springer, Berlin.

[9] van der Vaart, A. and Wellner, J. A. (2000). Preservation theorems for Glivenko-Cantelli and uniform Glivenko-Cantelli classes. In *High Dimensional Probability, II (Seattle, WA, 1999). Progr. Probab.* **47** 115–133. Birkhäuser Boston, Boston, MA. MR1857319 (2002j:60053)

[10] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics*. Springer-Verlag, New York. MR1385671 (97g:60035)

Department of Biostatistics
University of Washington
Seattle, WA 98195-7232
E-mail: tsaegusa@uw.edu

Department of Statistics
University of Washington
Seattle, WA 98195-4322,
E-mail: jaw@stat.washington.edu