Covariate Distribution Balance via Propensity Scores: Supplemental Appendix

Pedro H. C. Sant'Anna* Xiaojun Song† Qi Xu‡

Vanderbilt University Peking University Vanderbilt University

August 28, 2018

This supplemental appendix contains additional computational details, auxiliary lemmas, and proofs of the main theorems presented in the main text. Appendix A discusses closed-form representation of the integrated propensity score (IPS) objective function using the three families of weighting functions under Assumption 3 in the main text. Appendix B presents auxiliary lemmas. Finally, Appendix C collects all the proofs of the main results of the paper.

*Department of Economics, Vanderbilt University. E-mail: pedro.h.santanna@vanderbilt.edu.

†Department of Business Statistics and Econometrics, Guanghua School of Management and Center for Statistical Science, Peking University, Beijing, 100871, China. E-mail: sxj@gsm.pku.edu.cn.

‡Department of Economics, Vanderbilt University. E-mail: qi.xu.1@vanderbilt.edu.

A Closed-form Representation of the IPS Objective Functions

In this section, we derive the closed-form representation of the IPS objective function $Q_{n,w}(\boldsymbol{\beta})$ using the three families of weighting functions under Assumption 3 in the main text.

First, recall from Section 2.2 that

$$\widehat{\boldsymbol{\beta}}^{ips}_{n,w} = \arg\min_{\boldsymbol{\beta}\in\Theta} Q_{n,w}(\boldsymbol{\beta}),$$

where $Q_{n,w}(\boldsymbol{\beta}) = \int \left\|\mathbf{H}_{n,w}(\boldsymbol{\beta},\mathbf{u})\right\|^2\ \Psi_n(d\mathbf{u})$, Ψ_n is an uniformly consistent estimator of Ψ, $\mathbf{H}_{n,w}(\boldsymbol{\beta},\mathbf{u}) = \mathbb{E}_n\left[\mathbf{h}_n(D,\mathbf{X};\boldsymbol{\beta})\, w(\mathbf{X};\mathbf{u})\right]$, with $\mathbf{h}_n(D,\mathbf{X};\boldsymbol{\beta}) = \left(h_{n,1}(D,\mathbf{X};\boldsymbol{\beta}), h_{n,0}(D,\mathbf{X};\boldsymbol{\beta})\right)'$, $h_{n,d}(D,\mathbf{X};\boldsymbol{\beta}) = w^{ps}_{n,d}(D,\mathbf{X};\boldsymbol{\beta}) - 1$, $d\in\{0,1\}$, and

$$\begin{aligned} w^{ps}_{n,1}(D,\mathbf{X};\boldsymbol{\beta}) &= \frac{D}{p(\mathbf{X};\boldsymbol{\beta})}\Bigg/\mathbb{E}_n\left[\frac{D}{p(\mathbf{X};\boldsymbol{\beta})}\right], \\ w^{ps}_{n,0}(D,\mathbf{X};\boldsymbol{\beta}) &= \frac{1-D}{1-p(\mathbf{X};\boldsymbol{\beta})}\Bigg/\mathbb{E}_n\left[\frac{1-D}{1-p(\mathbf{X};\boldsymbol{\beta})}\right]. \end{aligned}$$

In the following, we derive the closed-form representation of the IPS objective function of the estimators in (10)-(12). In light of Remark 4, we emphasize the role played by $h_{n,1}$ and $h_{n,0}$ in the computation of the objective functions.

Case 1: Indicator Weights

As introduced in (10), in this case we have $\Psi_n(\mathbf{u}) = F_{n,\mathbf{X}}(\mathbf{u})$ and $w(\mathbf{X};\mathbf{u}) = 1\{\mathbf{X}\leq\mathbf{u}\}$. For any $\mathbf{g}(\mathbf{x})$, $\int \mathbf{g}(\mathbf{u})\,F_{n,\mathbf{X}}(d\mathbf{u}) = n^{-1}\sum_{j=1}^n \mathbf{g}(\mathbf{X}_j)$. Since $1\{\mathbf{X}\leq\mathbf{u}\}$ is real-valued, conjugate transpose reduces to direct transpose. Hence,

$$\begin{aligned} \mathbf{H}_{n,ind}(\boldsymbol{\beta},\mathbf{u}) &= \frac{1}{n}\sum_{j=1}^n \mathbf{h}_n(D_j,\mathbf{X}_j;\boldsymbol{\beta})\,1(\mathbf{X}_j\leq\mathbf{u}) \\ &= \frac{1}{n}\sum_{j=1}^n \left(h_{n,1}(D_j,\mathbf{X}_j;\boldsymbol{\beta}), h_{n,0}(D_j,\mathbf{X}_j;\boldsymbol{\beta})\right)' 1(\mathbf{X}_j\leq\mathbf{u}) \end{aligned}$$

and

$$
\begin{aligned}
Q_{n,ind}(\boldsymbol{\beta}) &= \int_{[-\infty,\infty]^k} \left\| \mathbb{E}_n \left[\mathbf{h}_n(D,\mathbf{X};\boldsymbol{\beta}) 1(\mathbf{X} \leq \mathbf{u}) \right] \right\|^2 F_{n,\mathbf{X}}(d\mathbf{u}); \\
&= \frac{1}{n^3} \sum_{l=1}^{n} \left(\sum_{j=1}^{n} \mathbf{h}_n(D_j,\mathbf{X}_j;\boldsymbol{\beta}) 1(\mathbf{X}_j \leq \mathbf{X}_l) \right)' \left(\sum_{j=1}^{n} \mathbf{h}_n(D_j,\mathbf{X}_j;\boldsymbol{\beta}) 1(\mathbf{X}_j \leq \mathbf{X}_l) \right) \\
&= \frac{1}{n^3} \sum_{l=1}^{n} \left(\sum_{j=1}^{n} h_{n,1}(D_j,\mathbf{X}_j;\boldsymbol{\beta}) 1(\mathbf{X}_j \leq \mathbf{X}_l) \right)^2 \\
&+ \frac{1}{n^3} \sum_{l=1}^{n} \left(\sum_{j=1}^{n} h_{n,0}(D_j,\mathbf{X}_j;\boldsymbol{\beta}) 1(\mathbf{X}_j \leq \mathbf{X}_l) \right)^2 .
\end{aligned}
$$

Case 2: Projection Matrix Weights

As introduced in (11), in this case we have $\Psi_n(\mathbf{u}) = n^{-1} \sum_{i=1}^{n} 1(\boldsymbol{\gamma}'\mathbf{X}_j \leq u) \times \boldsymbol{\gamma}$ and $w(\mathbf{X};\mathbf{u}) = 1\{\boldsymbol{\gamma}'\mathbf{X} \leq u\}$. Before proceeding to the detailed derivations, from Appendix B of Escanciano (2006), we have that

$$
\begin{aligned}
&\int_{[-\infty\times\infty]\times\mathbb{S}_k} 1(\boldsymbol{\gamma}'\mathbf{X}_j \leq u) 1(\boldsymbol{\gamma}'\mathbf{X}_s \leq u) F_{n,\boldsymbol{\gamma}'\mathbf{X}}(du)\, d\boldsymbol{\gamma} \\
&= \frac{1}{n} \sum_{r=1}^{n} \int_{\mathbb{S}_k} 1(\boldsymbol{\gamma}'\mathbf{X}_j \leq \boldsymbol{\gamma}'\mathbf{X}_r) 1(\boldsymbol{\gamma}'\mathbf{X}_s \leq \boldsymbol{\gamma}'\mathbf{X}_r)\, d\boldsymbol{\gamma} \\
&= \frac{1}{n} \sum_{r=1}^{n} A_{jsr} \\
&\equiv A_{js},
\end{aligned}
$$

where A_{jsr} is proportional to the volume of a spherical wedge and can be computed as

$$
A_{jsr} \equiv A_{jsr}^{(0)} \frac{\pi^{(k/2)-1}}{\Gamma\left(\dfrac{k}{2}+1\right)},
$$

where $\Gamma(\cdot)$ is the gamma function and

$$
A_{jsr}^{(0)} \equiv \left| \pi - \arccos\left(\frac{(\mathbf{X}_j - \mathbf{X}_r)'(\mathbf{X}_s - \mathbf{X}_r)}{|(\mathbf{X}_j - \mathbf{X}_r)| \cdot |(\mathbf{X}_s - \mathbf{X}_r)|} \right) \right| .
$$

With this result in hand, and the fact that $1(\boldsymbol{\gamma}'\mathbf{X} \leq u)$ is real-valued, the objective function $Q_{n,proj}(\boldsymbol{\beta})$ can be written as

$$
\begin{aligned}
Q_{n,proj}(\boldsymbol{\beta}) &= \int_{[-\infty,\infty]\times\mathbb{S}_k} \left\|\mathbb{E}_n\left[\mathbf{h}_n(D,\mathbf{X};\boldsymbol{\beta})1(\boldsymbol{\gamma}'\mathbf{X}\le u)\right]\right\|^2 F_{n,\boldsymbol{\gamma}}(du)d\boldsymbol{\gamma} \\
&= \frac{1}{n^2}\int_{\mathbb{R}\times\mathbb{S}_k}\left(\sum_{j=1}^{n} h_{n,1}(D_j,\mathbf{X}_j;\boldsymbol{\beta})1(\boldsymbol{\gamma}'\mathbf{X}_j\le u)\right)^2 F_{n,\boldsymbol{\gamma}'\mathbf{X}}(du)\,d\boldsymbol{\gamma} \\
&+ \frac{1}{n^2}\int_{\mathbb{R}\times\mathbb{S}_k}\left(\sum_{j=1}^{n} h_{n,0}(D_j,\mathbf{X}_j;\boldsymbol{\beta})1(\boldsymbol{\gamma}'\mathbf{X}_j\le u)\right)^2 F_{n,\boldsymbol{\gamma}'\mathbf{X}}(du)\,d\boldsymbol{\gamma} \\
&= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{s=1}^{n} h_{n,1}(D_j,\mathbf{X}_j;\boldsymbol{\beta})\,h_{n,1}(D_s,\mathbf{X}_s;\boldsymbol{\beta})\,A_{js} \\
&+ \frac{1}{n^2}\sum_{j=1}^{n}\sum_{s=1}^{n} h_{n,0}(D_j,\mathbf{X}_j;\boldsymbol{\beta})\,h_{n,0}(D_s,\mathbf{X}_s;\boldsymbol{\beta})\,A_{js}.
\end{aligned}
$$

Case 3: Exponential Weights

Finally, as introduced in (12), now we have $w(\mathbf{X};\mathbf{u}) = exp(i\mathbf{u}'\Phi(\mathbf{X}))$ and

$$
\Psi_n(d\mathbf{u}) = \Psi(d\mathbf{u}) = \frac{\exp(-\mathbf{u}'\mathbf{u}/2)}{(2\pi)^{k/2}}d\mathbf{u}.
$$

For notational convenience, let $\mathbf{h}_{n,j}(\boldsymbol{\beta}) \equiv \mathbf{h}_n(D_j,\mathbf{X}_j;\boldsymbol{\beta})$. Hence

$$
\begin{aligned}
&Q_{n,exp}(\boldsymbol{\beta}) \\
&= \int_{\mathbb{R}^k}\left\|\mathbb{E}_n\left[\mathbf{h}_n(D,\mathbf{X};\boldsymbol{\beta})\exp(i\mathbf{u}'\Phi(\mathbf{X}))\right]\right\|^2 \frac{\exp\left(-\frac{1}{2}\mathbf{u}'\mathbf{u}\right)}{(2\pi)^{k/2}}d\mathbf{u} \\
&= \frac{1}{n^2}\int_{\mathbb{R}^k}\left(\sum_{j=1}^{n}\mathbf{h}_{n,j}(\boldsymbol{\beta})\exp(-i\mathbf{u}'\Phi(\mathbf{X}_j))\right)'\left(\sum_{s=1}^{n}\mathbf{h}_{n,s}(\boldsymbol{\beta})\exp(i\mathbf{u}'\Phi(\mathbf{X}_j))\right)\Psi(d\mathbf{u}) \\
&= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{s=1}^{n}\left\{\mathbf{h}'_{n,j}(\boldsymbol{\beta})\,\mathbf{h}_{n,s}(\boldsymbol{\beta})\int_{\mathbb{R}^k}\exp\left(i\mathbf{u}'\left(\Phi(\mathbf{X}_j)-\Phi(\mathbf{X}_s)\right)\right)\frac{\exp(-\mathbf{u}'\mathbf{u}/2)}{(2\pi)^{k/2}}d\mathbf{u}\right\} \\
&= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{s=1}^{n}\mathbf{h}'_{n,j}(\boldsymbol{\beta})\,\mathbf{h}_{n,s}(\boldsymbol{\beta})\exp\left\{-\frac{1}{2}||\Phi(\mathbf{X}_j)-\Phi(\mathbf{X}_s)||^2\right\}. \qquad \text{(A.1)}
\end{aligned}
$$

To get the last equality, we exploit that

$$\int_{\mathbb{R}^k} \exp\left(i\mathbf{u}'\mathbf{t}\right) \cdot \frac{\exp\left(-\mathbf{u}'\mathbf{u}/2\right)}{(2\pi)^{k/2}} d\mathbf{u} = \mathbb{E}_{\mathbf{U}}[\exp\left(i\mathbf{U}'\mathbf{t}\right)]$$

$$= \exp\left\{-\mathbf{t}'\mathbf{t}/2\right\},$$

where we use the definition of characteristic function for the random variable $\mathbf{U}$, and exploits that $\mathbf{U}$ follows a standard k-variate normal distribution. Letting $\mathbf{t} = \Phi\left(\mathbf{X}_j\right) - \Phi\left(\mathbf{X}_s\right)$, (A.1) follows immediately.

Thus, from (A.1) and the definition of $\mathbf{h}_{n,j}\left(\boldsymbol{\beta}\right)$, we have

$$Q_{n,exp}\left(\boldsymbol{\beta}\right) = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{s=1}^{n} h_{n,1}\left(D_j, \mathbf{X}_j; \boldsymbol{\beta}\right) h_{n,1}\left(D_s, \mathbf{X}_s; \boldsymbol{\beta}\right) \exp\left\{-\frac{1}{2}||\Phi\left(\mathbf{X}_j\right) - \Phi\left(\mathbf{X}_s\right)||^2\right\}$$
$$+ \frac{1}{n^2}\sum_{j=1}^{n}\sum_{s=1}^{n} h_{n,0}\left(D_j, \mathbf{X}_j; \boldsymbol{\beta}\right) h_{n,0}\left(D_s, \mathbf{X}_s; \boldsymbol{\beta}\right) \exp\left\{-\frac{1}{2}||\Phi\left(\mathbf{X}_j\right) - \Phi\left(\mathbf{X}_s\right)||^2\right\}.$$

B Auxiliary Lemmas

In this Section, we present and prove some auxiliary lemmas that helps on proving the main results of the paper.

Lemma B.1 *Let Π be a compact, convex subset of $\mathbb{R}^k$ with a non-empty interior. Then*

$$\mathcal{W}_{ind} = \left\{\mathbf{x} \in \mathcal{X} \mapsto 1(\mathbf{x} \leq \mathbf{u}) : \mathbf{u} \in [-\infty, \infty]^k\right\},$$

$$\mathcal{W}_{proj} = \left\{\mathbf{x} \in \mathcal{X} \mapsto 1\left\{\boldsymbol{\gamma}'\mathbf{x} \leq u\right\} : (\boldsymbol{\gamma}, u) \in \mathbb{S}_k \times [-\infty, \infty]\right\},$$

$$\mathcal{W}_{exp} = \left\{\mathbf{x} \in \mathcal{X} \mapsto \exp(i\mathbf{u}'\Phi\left(\mathbf{x}\right)) : \mathbf{u} \in \Pi\right\},$$

are uniformly bounded Donsker classes of functions.

Proof of Lemma B.1: The uniform boundedness property follows from the fact that $1(x \leq u) \leq 1$, $1\left\{\boldsymbol{\gamma}'\mathbf{x} \leq u\right\} \leq 1$ and $|\exp(i\mathbf{u}'\Phi\left(\mathbf{x}\right))| = |\cos\left(\mathbf{u}'\Phi\left(\mathbf{x}\right)\right) + i\sin\left(\mathbf{u}'\Phi\left(\mathbf{x}\right)\right)| \leq 1$. From Example 2.5.4 in van der Vaart and Wellner (1996), $\mathcal{W}_{ind}$ is Donsker. From Theorems 2.5.2, 2.6.7 and Problem 14 on page 152 in van der Vaart and Wellner (1996),

$\mathcal{W}_{\text{proj}}$ is Donsker. Finally, since $\exp(i\mathbf{u}'\Phi(\mathbf{x}))$ is infinitely differentiable with respect to $\mathbf{u}$, and all derivatives are uniformly bounded on Π, the Donsker property of $\mathcal{W}_{\exp}$ follows from Theorem 2.5.6 and Corollary 2.7.2 in van der Vaart and Wellner (1996).■

Lemma B.2 *Under Assumption 2(i) − (iii), the classes of functions*

$$\mathcal{F}_1 \equiv \{(d, \mathbf{x}) \in \{0,1\} \times \mathcal{X} \mapsto d/p(\mathbf{x};\boldsymbol{\beta}) : \boldsymbol{\beta} \in \Theta\},$$

$$\mathcal{F}_2 \equiv \{(d, \mathbf{x}) \in \{0,1\} \times \mathcal{X} \mapsto (1-d)/(1-p(\mathbf{x};\boldsymbol{\beta})) : \boldsymbol{\beta} \in \Theta\},$$

$$\mathcal{F}_3 \equiv \mathcal{F}_1 \cdot \mathcal{W},$$

$$\mathcal{F}_4 \equiv \mathcal{F}_2 \cdot \mathcal{W},$$

where $\mathcal{W}$ is either equal to $\mathcal{W}_{ind}$, $\mathcal{W}_{proj}$ or $\mathcal{W}_{exp}$, are Glivenko-Cantelli.

Proof of Lemma B.2: By Example 19.8 in van der Vaart (1998), $\mathcal{F}_1$ and $\mathcal{F}_2$ are Glivenko-Cantelli (GC) classes under Assumption 2$(i)-(iii)$. By Lemma B.1, $\mathcal{W}_{ind}$, $\mathcal{W}_{\text{proj}}$ and $\mathcal{W}_{\exp}$ are uniformly bounded Donsker classes of functions, and therefore they are also GC. Finally, by Corollary 9.26 in Kosorok (2008), $\mathcal{F}_3$ and $\mathcal{F}_4$ are GC. ■

Let

$$\widehat{C}_{ind,F_{n,\mathbf{x}}} = 2\int_{[-\infty\times\infty]^k} \dot{\mathbf{H}}'_{n,ind}(\widehat{\boldsymbol{\beta}}^{ips}_{n,ind}, \mathbf{u})\dot{\mathbf{H}}_{n,ind}(\tilde{\boldsymbol{\beta}}, \mathbf{u})F_{n,\mathbf{x}}(d\mathbf{u}),$$

$$\widehat{C}_{proj,F_{n,\gamma}} = 2\int_{[-\infty\times\infty]\times\mathbb{S}_k} \dot{\mathbf{H}}'_{n,proj}(\widehat{\boldsymbol{\beta}}^{ips}_{n,proj}, \mathbf{u})\dot{\mathbf{H}}_{n,proj}(\tilde{\boldsymbol{\beta}}, \mathbf{u})F_{n,\gamma}(du)d\boldsymbol{\gamma},$$

and

$$\widehat{C}_{\exp,\Phi} = \int_{\mathbb{R}^k} \dot{\mathbf{H}}^c_{n,\exp}(\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp}, \mathbf{u})\dot{\mathbf{H}}_{n,\exp}(\tilde{\boldsymbol{\beta}}, \mathbf{u})\phi(\mathbf{u})\,d\mathbf{u} + \int_{\mathbb{R}^k} \dot{\mathbf{H}}'_{n,\exp}(\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp}, \mathbf{u})\left(\dot{\mathbf{H}}'_{n,\exp}(\tilde{\boldsymbol{\beta}}, \mathbf{u})\right)^c\phi(\mathbf{u})\,d\mathbf{u},$$

where $\phi(\mathbf{u})$ is the standard k-variate normal density function and $\tilde{\boldsymbol{\beta}}$ satisfies $\left\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\| \leq$

$\left\|\widehat{\boldsymbol{\beta}}_{n,w}^{ips}-\boldsymbol{\beta}_0\right\|$. Furthermore, write

$$C_{ind,F_{\mathbf{X}}}=2\int_{[-\infty,\infty]^k}\dot{\mathbf{H}}_{ind}'(\boldsymbol{\beta}_0,\mathbf{u})\dot{\mathbf{H}}_{ind}(\boldsymbol{\beta}_0,\mathbf{u})F_{\mathbf{X}}(d\mathbf{u}),$$

$$C_{proj,F_\gamma}=2\int_{[-\infty,\infty]\times\mathbb{S}_k}\dot{\mathbf{H}}_{proj}'(\boldsymbol{\beta}_0,\mathbf{u})\dot{\mathbf{H}}_{proj}(\boldsymbol{\beta}_0,\mathbf{u})F_\gamma(du)d\boldsymbol{\gamma},$$

$$C_{\exp,\Phi}=\int_{\mathbb{R}^k}\left(\dot{\mathbf{H}}_{\exp}^c(\boldsymbol{\beta}_0,\mathbf{u})\dot{\mathbf{H}}_w(\boldsymbol{\beta}_0,\mathbf{u})+\dot{\mathbf{H}}_{\exp}'(\boldsymbol{\beta}_0,\mathbf{u})\left(\dot{\mathbf{H}}_{\exp}'(\boldsymbol{\beta}_0,\mathbf{u})\right)^c\right)\phi\left(\mathbf{u}\right)d\mathbf{u}.$$

Lemma B.3 *Let $\mathcal{W}$ be equal to either $\mathcal{W}_{ind}$, $\mathcal{W}_{proj}$ or $\mathcal{W}_{exp}$. Then, under Assumption 2,*

$$\mathcal{F}_5\equiv\left\{(d,\mathbf{x})\in\{0,1\}\times\mathcal{X}\mapsto\frac{d}{p\left(\mathbf{x};\boldsymbol{\beta}\right)^2}\dot{p}\left(\mathbf{x};\boldsymbol{\beta}\right),\boldsymbol{\beta}\in\Theta_0\right\},$$

$$\mathcal{F}_6\equiv\left\{(d,\mathbf{x})\in\{0,1\}\times\mathcal{X}\mapsto\frac{1-d}{\left(1-p\left(\mathbf{x};\boldsymbol{\beta}\right)\right)^2}\dot{p}\left(\mathbf{x};\boldsymbol{\beta}\right),\boldsymbol{\beta}\in\Theta_0\right\},$$

$$\mathcal{F}_7\equiv\mathcal{F}_5\cdot\mathcal{W},$$

$$\mathcal{F}_8\equiv\mathcal{F}_6\cdot\mathcal{W},$$

are Glivenko-Cantelli classes of functions. Furthermore,

$$\widehat{C}_{ind,F_{n,\mathbf{X}}}-C_{ind,F_{\mathbf{X}}}=o_p\left(1\right),$$

$$\widehat{C}_{proj,F_{n,\gamma}}-C_{proj,F_\gamma}=o_p\left(1\right),$$

$$\widehat{C}_{\exp,\Phi}-C_{\exp,\Phi}=o_p\left(1\right).$$

Proof of Lemma B.3: By Example 19.8 in van der Vaart (1998), $\mathcal{F}_5$ and $\mathcal{F}_6$ are Glivenko-Cantelli (GC) classes under Assumption 2. By Lemma B.1, $\mathcal{W}_{ind}$, $\mathcal{W}_{\text{proj}}$ and $\mathcal{W}_{\text{exp}}$ are uniformly bounded Donsker classes of functions, and therefore they are also GC. Finally, by Corollary 9.27 in Kosorok (2008), $\mathcal{F}_3$ and $\mathcal{F}_4$ are GC.

Next, from the first part of Theorem 1, we have that $\widehat{\boldsymbol{\beta}}_{n,w}^{ips}\xrightarrow{p}\boldsymbol{\beta}_0$, which in turn implies that $\widehat{\boldsymbol{\beta}}_{n,w}^{ips},\tilde{\boldsymbol{\beta}}\in\Theta_0$ with probability approaching one. Thus, from Lemma B.1 and a direct application of the CMT, we conclude that

$$\widehat{C}_{ind,F_{n,\mathbf{X}}}-C_{ind,F_{\mathbf{X}}}=o_p\left(1\right),$$

$$\widehat{C}_{proj,F_{n,\gamma}} - C_{proj,F_\gamma} = o_p\left(1\right).$$

To conclude the proof of this lemma, we need to show that

$$\widehat{C}_{\exp,\Phi} - C_{\exp,\Phi} = o_p\left(1\right).$$

Toward this end, as in the consistency proof of Theorem 1, fix an arbitrarily small $\epsilon > 0$ and choose a compact and convex set Π such that

$$\left| \int_{\mathbb{R}^k \backslash \Pi} \phi\left(\mathbf{u}\right) d\mathbf{u} \right| \leq \epsilon. \tag{B.1}$$

Then, write

$$\begin{aligned}
&\int_{\mathbb{R}^k} A_{n,1}\left(\mathbf{u};\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp},\tilde{\boldsymbol{\beta}}\right)\phi\left(\mathbf{u}\right) d\mathbf{u} \\
&= \int_{\Pi} A_{n,1}\left(\mathbf{u};\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp},\tilde{\boldsymbol{\beta}}\right)\phi\left(\mathbf{u}\right) d\mathbf{u} + \int_{\mathbb{R}^k \backslash \Pi} A_{n,1}\left(\mathbf{u};\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp},\tilde{\boldsymbol{\beta}}\right)\phi\left(\mathbf{u}\right) d\mathbf{u}, \\
&\equiv J_{3n} + J_{4n}.
\end{aligned}$$

with

$$A_{n,1}\left(\mathbf{u};\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp},\tilde{\boldsymbol{\beta}}\right) \equiv \dot{\mathbf{H}}^{c}_{n,\exp}(\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp},\mathbf{u})\dot{\mathbf{H}}_{n,\exp}(\tilde{\boldsymbol{\beta}},\mathbf{u}) + \dot{\mathbf{H}}^{'}_{n,\exp}(\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp},\mathbf{u})\left(\dot{\mathbf{H}}^{\prime}_{n,\exp}(\tilde{\boldsymbol{\beta}},\mathbf{u})\right)^{c}.$$

Let

$$A_1\left(\mathbf{u};\boldsymbol{\beta}_0,\boldsymbol{\beta}\right) \equiv \dot{\mathbf{H}}^{c}_{\exp}(\boldsymbol{\beta}_0,\mathbf{u})\dot{\mathbf{H}}_{\exp}(\boldsymbol{\beta},\mathbf{u}) + \dot{\mathbf{H}}^{'}_{\exp}(\boldsymbol{\beta}_0,\mathbf{u})\left(\dot{\mathbf{H}}^{\prime}_{\exp}(\boldsymbol{\beta},\mathbf{u})\right)^{c}.$$

From the GC results above and the CMT, we have that

$$\sup_{\mathbf{u}\in\Pi}\left\| A_{n,1}\left(\mathbf{u};\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp},\tilde{\boldsymbol{\beta}}\right) - A_1\left(\mathbf{u};\boldsymbol{\beta}_0,\boldsymbol{\beta}_0\right)\right\| \xrightarrow{p} 0.$$

Thus, by another application of the CMT, it follows that

$$J_{3n} = \int_{\Pi} A_1\left(\mathbf{u};\boldsymbol{\beta}_0,\boldsymbol{\beta}_0\right)\phi\left(\mathbf{u}\right) d\mathbf{u} + o_p\left(1\right).$$

For J_{4n}, since $|\exp(i\mathbf{u}'\Phi\left(\mathbf{x}\right)| \leq 1$, we have that under Assumption 2, for all $\boldsymbol{\beta} \in \Theta_0$,

$$\left\|\dot{\mathbf{H}}_{n,\exp}(\boldsymbol{\beta},\mathbf{u})\right\| \leq \mathbb{E}_n\left[b\left(\mathbf{X}\right)\right] = O_p\left(1\right)$$

for some integrable function $b(\mathbf{X})$. Hence, by (B.1) we have that $J_{4n} = O_p(\epsilon)$. Since $\epsilon > 0$ is arbitrary, this concludes the proof. ■

Lemma B.4 *Let Π be a compact, convex subset of $\mathbb{R}^k$ with a non-empty interior. Then, under Assumption 2,*

$$\mathcal{F}_{ind} \equiv \left\{(d,\mathbf{x}) \in \{0,1\} \times \mathcal{X} \mapsto \mathbf{h}\left(d,\mathbf{x};\boldsymbol{\beta}_0\right) 1(\mathbf{x} \leq \mathbf{u}) : \mathbf{u} \in [-\infty,\infty]^k\right\},$$

$$\mathcal{F}_{proj} \equiv \left\{(d,\mathbf{x}) \in \{0,1\} \times \mathcal{X} \mapsto \mathbf{h}\left(d,\mathbf{x};\boldsymbol{\beta}_0\right) 1\left\{\boldsymbol{\gamma}'\mathbf{x} \leq u\right\} : (\boldsymbol{\gamma},u) \in \mathbb{S}_k \times [-\infty,\infty]\right\}$$

$$\mathcal{F}_{\exp} \equiv \left\{(d,\mathbf{x}) \in \{0,1\} \times \mathcal{X} \mapsto \mathbf{h}\left(d,\mathbf{x};\boldsymbol{\beta}_0\right) \exp(i\mathbf{u}'\Phi\left(\mathbf{x}\right)) : \mathbf{u} \in \Pi\right\},$$

are Donsker classes of functions.

Proof of Lemma B.4: The Donsker properties follow directly from Lemma B.1, Assumption 2(ii), and Corollary 9.32 in Kosorok (2008). ■

Define

$$A_{n,2,ind}\left(\mathbf{x}\right) = 2 \cdot \int_{[-\infty,\infty]^k} \dot{\mathbf{H}}_{n,ind}'(\hat{\boldsymbol{\beta}}_{n,ind}^{ips},\mathbf{u}) 1(\mathbf{x} \leq \mathbf{u}) F_{n,\mathbf{X}}(d\mathbf{u}),$$

$$A_{n,2,proj}\left(\mathbf{x}\right) = 2 \cdot \int_{[-\infty,\infty]\times\mathbb{S}_k} \dot{\mathbf{H}}_{n,proj}'(\hat{\boldsymbol{\beta}}_{n,proj}^{ips},(u,\boldsymbol{\gamma})) 1\left\{\boldsymbol{\gamma}'\mathbf{x} \leq u\right\} F_{n,\boldsymbol{\gamma}}(du)d\boldsymbol{\gamma},$$

$$A_{n,2,\exp}\left(\mathbf{x}\right) = \int_{\mathbb{R}^k} \left(\dot{\mathbf{H}}_{n,\exp}^{c}(\hat{\boldsymbol{\beta}}_{n,\exp}^{ips},\mathbf{u}) \exp(i\mathbf{u}'\Phi\left(\mathbf{x}\right)) + \dot{\mathbf{H}}_{n,\exp}'(\hat{\boldsymbol{\beta}}_{n,\exp}^{ips},\mathbf{u}) \exp(-i\mathbf{u}'\Phi\left(\mathbf{x}\right))\right) \phi\left(\mathbf{u}\right) d\mathbf{u},$$

and let

$$A_{2,ind}\left(\mathbf{x}\right) = 2 \cdot \int_{[-\infty,\infty]^k} \left(\dot{\mathbf{H}}_{ind}'(\boldsymbol{\beta}_0,\mathbf{u}) 1(\mathbf{x} \leq \mathbf{u})\right) F_{\mathbf{X}}(d\mathbf{u}),$$

$$A_{2,proj}\left(\mathbf{x}\right) = 2 \cdot \int_{[-\infty,\infty]\times\mathbb{S}_k} \dot{\mathbf{H}}_{proj}'(\boldsymbol{\beta}_0,(u,\boldsymbol{\gamma})) 1\left\{\boldsymbol{\gamma}'\mathbf{x} \leq u\right\} F_{\boldsymbol{\gamma}}(du)d\boldsymbol{\gamma},$$

$$A_{2,\exp}\left(\mathbf{x}\right) = \int_{\mathbb{R}^k} \left(\dot{\mathbf{H}}_{\exp}^{c}(\boldsymbol{\beta}_0,\mathbf{u}) \exp(i\mathbf{u}'\Phi\left(\mathbf{x}\right)) + \dot{\mathbf{H}}_{\exp}'(\boldsymbol{\beta}_0,\mathbf{u}) \exp(-i\mathbf{u}'\Phi\left(\mathbf{x}\right))\right) \phi\left(\mathbf{u}\right) d\mathbf{u}.$$

Lemma B.5 *Under Assumption 2,*

$$\mathbb{E}_n\left[A_{n,2,ind}\left(\mathbf{X}\right) \cdot \mathbf{h}_n\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\right] = \mathbb{E}_n\left[A_{2,ind}\left(\mathbf{X}\right) \cdot \mathbf{h}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\right] + o_p\left(n^{-1/2}\right). \tag{B.2}$$

$$\mathbb{E}_n\left[A_{n,2,proj}\left(\mathbf{X}\right)\cdot\mathbf{h}_n\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\right]=\mathbb{E}_n\left[A_{2,proj}\left(\mathbf{X}\right)\cdot\mathbf{h}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\right]+o_p\left(n^{-1/2}\right),\tag{B.3}$$

$$\mathbb{E}_n\left[A_{n,2,\exp}\left(\mathbf{X}\right)\cdot\mathbf{h}_n\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\right]=\mathbb{E}_n\left[A_{2,\exp}\left(\mathbf{X}\right)\cdot\mathbf{h}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\right]+o_p\left(n^{-1/2}\right).\tag{B.4}$$

Proof of Lemma B.5: First note that

$$\sqrt{n}\mathbb{E}_n\left[A_{n,2,ind}\left(\mathbf{X}\right)\cdot\mathbf{h}_n\left(D,X;\boldsymbol{\beta}_0\right)\right]=2\int\dot{\mathbf{H}}_{n,ind}\left(\hat{\boldsymbol{\beta}}^{ips}_{n,ind},\mathbf{u}\right)\sqrt{n}\mathbf{H}_{n,ind}\left(\boldsymbol{\beta}_0,\mathbf{u}\right)F_{n,\mathbf{X}}(d\mathbf{u}).$$

Then, from Lemma B.3, Lemma B.4, the CMT, and the fact that $\mathbf{H}_{ind}\left(\boldsymbol{\beta}_0,\mathbf{u}\right)=0$ $a.e$, it follows that, uniformly in $\mathbf{u}\in[-\infty,\infty]^k$,

$$\dot{\mathbf{H}}^{'}_{n,ind}\left(\hat{\boldsymbol{\beta}}^{ips}_{n,ind},\mathbf{u}\right)\cdot\sqrt{n}\mathbf{H}_{n,ind}\left(\boldsymbol{\beta}_0,\mathbf{u}\right)=\dot{\mathbf{H}}^{'}_{ind}\left(\boldsymbol{\beta}_0,\mathbf{u}\right)\cdot\sqrt{n}\mathbb{E}_n\left[\mathbf{h}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)1(\mathbf{X}\leq\mathbf{u})\right]+o_p\left(1\right).$$

Furthermore, the process

$$\dot{\mathbf{H}}^{'}_{ind}\left(\boldsymbol{\beta}_0,\mathbf{u}\right)\cdot\sqrt{n}\mathbb{E}_n\left[\mathbf{h}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)1(\mathbf{X}\leq\mathbf{u})\right]$$

is asymptotically tight in $\ell^\infty\left([-\infty,\infty]^k\right)$. Given Lemma B.1 and the Glivenko-Cantelli theorem, we apply Lemma 3.1 of Chang (1990) to conclude (B.2).

The proof of (B.3) follows exactly the same steps and is therefore omitted. To prove (B.4), fix an arbitrarily small $\epsilon>0$ and choose a compact and convex set Π such that

$$\left|\int_{\mathbb{R}^k\backslash\Pi}\phi\left(\mathbf{u}\right)d\mathbf{u}\right|\leq\epsilon.\tag{B.5}$$

Then write

$$\begin{aligned}
&\sqrt{n}\mathbb{E}_n\left[A_{n,2,\exp}\left(\mathbf{X}\right)\cdot\mathbf{h}_n\left(D,X;\boldsymbol{\beta}_0\right)\right]\\
&=\int_{\mathbb{R}^k}\dot{\mathbf{H}}^{c}_{n,\exp}(\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp},\mathbf{u})\sqrt{n}\mathbf{H}_{n,\exp}\left(\boldsymbol{\beta}_0,\mathbf{u}\right)\phi\left(\mathbf{u}\right)d\mathbf{u}\\
&+\int_{\mathbb{R}^k}\dot{\mathbf{H}}^{'}_{n,\exp}(\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp},\mathbf{u})\sqrt{n}\left(\mathbf{H}^{'}_{n,\exp}\left(\boldsymbol{\beta}_0,\mathbf{u}\right)\right)^c\phi\left(\mathbf{u}\right)d\mathbf{u}\\
&=\int_{\Pi}\widehat{A}_{n,3}\left(\mathbf{u};\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp},\boldsymbol{\beta}_0\right)\phi\left(\mathbf{u}\right)d\mathbf{u}+\int_{\mathbb{R}^k\backslash\Pi}\widehat{A}_{n,3}\left(\mathbf{u};\widehat{\boldsymbol{\beta}}^{ips}_{n,\exp},\boldsymbol{\beta}_0\right)\phi\left(\mathbf{u}\right)d\mathbf{u},\\
&\equiv J_{5n}+J_{6n}.
\end{aligned}$$

with

$$\widehat{A}_{n,3}\left(\mathbf{u};\widehat{\boldsymbol{\beta}}_{n,\exp}^{ips},\boldsymbol{\beta}_0\right)\equiv\dot{\mathbf{H}}_{n,\exp}^{c}(\widehat{\boldsymbol{\beta}}_{n,\exp}^{ips},\mathbf{u})\sqrt{n}\mathbf{H}_{n,\exp}\left(\boldsymbol{\beta}_0,\mathbf{u}\right)$$
$$+\dot{\mathbf{H}}_{n,\exp}^{'}(\widehat{\boldsymbol{\beta}}_{n,\exp}^{ips},\mathbf{u})\sqrt{n}\left(\mathbf{H}_{n,\exp}^{'}\left(\boldsymbol{\beta}_0,\mathbf{u}\right)\right)^{c}.$$

Let

$$A_{n,3}\left(\mathbf{u};\boldsymbol{\beta}_0,\boldsymbol{\beta}_0\right)\equiv\dot{\mathbf{H}}_{\exp}^{c}(\boldsymbol{\beta}_0,\mathbf{u})\sqrt{n}\mathbf{H}_{n,\exp}\left(\boldsymbol{\beta}_0,\mathbf{u}\right)+\dot{\mathbf{H}}_{\exp}^{'}(\boldsymbol{\beta}_0,\mathbf{u})\left(\mathbf{H}_{n,\exp}^{'}\left(\boldsymbol{\beta}_0,\mathbf{u}\right)\right)^{c}.$$

From Lemma B.3, Lemma B.4, the CMT, and the fact that $\mathbf{H}_{\exp}\left(\boldsymbol{\beta}_0,\mathbf{u}\right)=0$ $a.e$, it follows that, uniformly in $\mathbf{u}\in\Pi$,

$$\widehat{A}_{n,3}\left(\mathbf{u};\widehat{\boldsymbol{\beta}}_{n,\exp}^{ips},\boldsymbol{\beta}_0\right)=A_{n,3}\left(\mathbf{u};\boldsymbol{\beta}_0,\boldsymbol{\beta}_0\right)+o_p\left(1\right).$$

Furthermore, the process $A_{n,3}\left(\mathbf{u};\boldsymbol{\beta}_0,\boldsymbol{\beta}\right)$ is asymptotically tight in $\ell^{\infty}\left(\Pi\right)$. Given that $\int\left(\cdot\right)\phi\left(\mathbf{u}\right)d\mathbf{u}$ is a nonrandom continuous functional, by the CMT, we conclude that

$$J_{5n}=\int_{\Pi}A_{n,3}\left(\mathbf{u};\boldsymbol{\beta}_0,\boldsymbol{\beta}_0\right)\phi\left(\mathbf{u}\right)d\mathbf{u}+o_p\left(1\right).$$

For J_{6n}, since $|\exp(i\mathbf{u}'\Phi\left(\mathbf{x}\right)|\leq 1$, we have that under Assumption 2,

$$\left\|\dot{\mathbf{H}}_{n,\exp}(\widehat{\boldsymbol{\beta}}_{n,\exp}^{ips},\mathbf{u})\right\|\leq C\cdot\mathbb{E}_n\left[b\left(\mathbf{X}\right)\right]=O_p\left(1\right),$$

for some $C<\infty$ and some integrable function $b\left(\mathbf{X}\right)$. Furthermore, $\mathbb{E}\left[\left\|\mathbf{h}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\right\|\right]<\infty$,

$$\sqrt{n}\left\|\mathbf{H}_{n,\exp}\left(\boldsymbol{\beta}_0,\mathbf{u}\right)\right\|\leq C\cdot\sqrt{n}\mathbb{E}_n\left[\left\|\mathbf{h}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\right\|\right]=O_p\left(n^{1/2}\right)$$

Hence, by (B.5) we have that $J_{6n}=O_p(\epsilon\cdot n^{1/2})$. Since $\epsilon>0$ is arbitrary, we can pick ϵ such that, for some $\delta>0$, $\epsilon=o\left(n^{-1/2-\delta}\right)$, which concludes the proof. ■

C Proofs of Main Results

Proof of Lemma 1: The first part, $Q_w(\boldsymbol{\beta})\geq 0$, follows directly from the definition. Next, as discussed in Section 2.2, the covariate balancing condition (1) is equivalent to (4), implying that $Q_w(\boldsymbol{\beta}_0) = 0$. On the other hand, if $Q_w(\boldsymbol{\beta}) = 0$, it must be that $\mathbf{H}_w(\boldsymbol{\beta}, \boldsymbol{u}) = 0$ *a.e.* on Π, because $||\cdot|| \geq 0$ and the integrating probability measure Ψ is absolutely continuous with respect to a dominating measure *on* Π. However, $\mathbf{H}_w(\boldsymbol{\beta}, \boldsymbol{u}) = 0$ *a.e.* on Π if and only if $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ since, from (1), and (3), it follows that $\mathbb{P}\left(\mathbb{E}[\mathbf{h}(D, \mathbf{X}; \boldsymbol{\beta})|\mathbf{X}] = 0\right) < 1$ when $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$. This concludes the proof. ■

Proof of Theorem 1: We first show that $\widehat{\boldsymbol{\beta}}^{ips}_{n,w} - \boldsymbol{\beta}_0 = o_p(1)$ using M-estimator theory, see e.g. Theorem 5.7 in van der Vaart (1998). Since Lemma 1 already established that $Q_w(\boldsymbol{\beta})$ achieves the unique minimum value at $\boldsymbol{\beta}_0$, and by Assumption 2 we have that $\mathbf{H}_w(\boldsymbol{\beta}, \mathbf{u})$ is continuous at each $\boldsymbol{\beta} \in \Theta$, Θ is compact, we have that by Exercise 5.27 in van der Vaart (1998) for every $\varepsilon > 0$

$$\inf_{\boldsymbol{\beta}: \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \geq \varepsilon} Q_w(\boldsymbol{\beta}) > Q_w(\boldsymbol{\beta}_0).$$

Thus, to establish consistency of $\widehat{\boldsymbol{\beta}}^{ips}_{n,w}$ it suffices to show that, as $n \to \infty$,

$$\sup_{\boldsymbol{\beta} \in \Theta} |Q_{n,w}(\boldsymbol{\beta}) - Q_w(\boldsymbol{\beta})| \xrightarrow{p} 0.$$

From Lemma B.1 we have that $F_{n,\mathbf{X}}$ and $F_{n,\boldsymbol{\gamma}}$ are uniformly consistent for $F_{\mathbf{X}}$ and $F_{\boldsymbol{\gamma}}$, respectively, whereas by Lemma B.2 and the continuous mapping theorem (CMT), we have that

$$\sup_{(\boldsymbol{\beta}, \mathbf{u}) \in \Theta \times [-\infty, \infty]^k} \|\mathbf{H}_{n,ind}(\boldsymbol{\beta}, \mathbf{u}) - \mathbf{H}_{ind}(\boldsymbol{\beta}, \mathbf{u})\| \xrightarrow{p} 0,$$

$$\sup_{(\boldsymbol{\beta}, \mathbf{u}) \in \Theta \times ([-\infty, \infty] \times \mathbb{S}_k)} \|\mathbf{H}_{n,proj}(\boldsymbol{\beta}, \mathbf{u}) - \mathbf{H}_{proj}(\boldsymbol{\beta}, \mathbf{u})\| \xrightarrow{p} 0.$$

Given that integration is a linear functional, by the CMT we have that

$$\sup_{\boldsymbol{\beta} \in \Theta} |Q_{n,ind}(\boldsymbol{\beta}) - Q_{ind}(\boldsymbol{\beta})| \xrightarrow{p} 0,$$

$$\sup_{\boldsymbol{\beta} \in \Theta} |Q_{n,proj}(\boldsymbol{\beta}) - Q_{proj}(\boldsymbol{\beta})| \xrightarrow{p} 0.$$

To complete the consistency proof, we need to show that

$$\sup_{\boldsymbol{\beta} \in \Theta} \left| \int_{\mathbb{R}^k} \left(||\mathbf{H}_{n,\exp}(\boldsymbol{\beta}, \mathbf{u})||^2 - \|\mathbf{H}_{\exp}(\boldsymbol{\beta}, \mathbf{u})\|^2 \right) \phi\left(\mathbf{u}\right) d\mathbf{u} \right| \xrightarrow{p} 0,$$

where $\phi(\mathbf{u})$ is the standard k-variate normal density function. Fix an arbitrarily small $\epsilon > 0$ and choose a compact and convex set Π such that

$$\left| \int_{\mathbb{R}^k \backslash \Pi} \phi(\mathbf{u})\, d\mathbf{u} \right| \leq \epsilon. \tag{C.1}$$

Then, write

$$\begin{aligned}
&\int_{\mathbb{R}^k} ||\mathbf{H}_{n,\exp}(\boldsymbol{\beta}, \mathbf{u})||^2 \phi(\mathbf{u})\, d\mathbf{u} \\
&= \int_{\Pi} ||\mathbf{H}_{n,\exp}(\boldsymbol{\beta}, \mathbf{u})||^2 \phi(\mathbf{u})\, d\mathbf{u} + \int_{\mathbb{R}^k \backslash \Pi} ||\mathbf{H}_{n,\exp}(\boldsymbol{\beta}, \mathbf{u})||^2 \phi(\mathbf{u})\, d\mathbf{u} \\
&\equiv J_{1n} + J_{2n}.
\end{aligned}$$

From Lemma B.2 and the CMT, we have that

$$\sup_{(\boldsymbol{\beta}, \mathbf{u}) \in \Theta \times \Pi} \|\mathbf{H}_{n,\exp}(\boldsymbol{\beta}, \mathbf{u}) - \mathbf{H}_{\exp}(\boldsymbol{\beta}, \mathbf{u})\| \xrightarrow{p} 0.$$

Thus, by another application of the CMT, it follows that

$$J_{1n} = \int_{\Pi} \|\mathbf{H}_{\exp}(\boldsymbol{\beta}, \mathbf{u})\|^2 \phi(\mathbf{u})\, d\mathbf{u} + o_p(1)$$

uniformly in $\boldsymbol{\beta} \in \Theta$. For J_{2n}, since $|\exp(i\mathbf{u}'\Phi(\mathbf{x})| \leq 1$, we have that under Assumption 2(ii)

$$\|\mathbf{H}_{n,\exp}(\boldsymbol{\beta}, \mathbf{u})\| \leq C$$

for some $C < \infty$. Hence, by (C.1) we have that $J_{2n} = O_p(\epsilon)$. Since $\epsilon > 0$ is arbitrary, we conclude that

$$\sup_{\boldsymbol{\beta} \in \Theta} |Q_{n,\exp}(\boldsymbol{\beta}) - Q_{\exp}(\boldsymbol{\beta})| \xrightarrow{p} 0.$$

Next, we derive the asymptotic linear representation of $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{ips}_{n,w} - \boldsymbol{\beta}_0\right)$. Towards this end, note that the first order condition of $\min_{\beta} Q_{n,w}(\boldsymbol{\beta})$ is given as follows

$$\int \left\{ \dot{\mathbf{H}}^c_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w}, \mathbf{u}) \mathbf{H}_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w}, \mathbf{u}) + \left(\mathbf{H}^c_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w}, \mathbf{u}) \dot{\mathbf{H}}_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w}, \mathbf{u}) \right)' \right\} \Psi_n(d\mathbf{u}) = 0. \tag{C.2}$$

By the mean value theorem (after properly extending to the case of complex-valued functions of real variables), and Assumption 2(i), we have that

$$\mathbf{H}_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u}) = \mathbf{H}_{n,w}(\boldsymbol{\beta}_0,\mathbf{u}) + \dot{\mathbf{H}}_{n,w}(\tilde{\boldsymbol{\beta}},\mathbf{u})(\widehat{\boldsymbol{\beta}}^{ips}_{n,w} - \boldsymbol{\beta}_0), \tag{C.3}$$

where $\tilde{\boldsymbol{\beta}}$ satisfies $\left\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\| \leq \left\|\widehat{\boldsymbol{\beta}}^{ips}_{n,w} - \boldsymbol{\beta}_0\right\|$. Plugging (C.3) into (C.2), we can write

$$\int \left(\dot{\mathbf{H}}^c_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u})\mathbf{H}_{n,w}(\boldsymbol{\beta}_0,\mathbf{u}) + \dot{\mathbf{H}}'_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u})\left(\mathbf{H}'_{n,w}(\boldsymbol{\beta}_0,\mathbf{u})\right)^c\right)\Psi_n(d\mathbf{u})$$
$$+ \widehat{C}_{w,\Psi_n}\left(\widehat{\boldsymbol{\beta}}^{ips}_{n,w} - \boldsymbol{\beta}_0\right) = 0$$

where

$$\widehat{C}_{w,\Psi_n} = \int \left(\dot{\mathbf{H}}^c_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u})\dot{\mathbf{H}}_{n,w}(\tilde{\boldsymbol{\beta}},\mathbf{u}) + \dot{\mathbf{H}}'_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u})\left(\dot{\mathbf{H}}'_{n,w}(\tilde{\boldsymbol{\beta}},\mathbf{u})\right)^c\right)\Psi_n(d\mathbf{u}).$$

Therefore

$$\begin{aligned}&\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{ips}_{n,w} - \boldsymbol{\beta}_0\right)\\ &= -\widehat{C}^{-1}_{w,\Psi_n}\cdot\sqrt{n}\int \left(\dot{\mathbf{H}}^c_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u})\mathbf{H}_{n,w}(\boldsymbol{\beta}_0,\mathbf{u}) + \dot{\mathbf{H}}'_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u})\left(\mathbf{H}'_{n,w}(\boldsymbol{\beta}_0,\mathbf{u})\right)^c\right)\Psi_n(d\mathbf{u}).\end{aligned} \tag{C.4}$$

By exploiting that $\mathbf{H}_{n,w}(\boldsymbol{\beta}_0,\mathbf{u}) = \mathbb{E}_n\left[\mathbf{h}_n(D,\mathbf{X};\boldsymbol{\beta}_0)\,w(\mathbf{X},\mathbf{u})\right]$, we can express (C.4) as

$$\begin{aligned}&\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{ips}_{n,w} - \boldsymbol{\beta}_0\right)\\ &\quad= -\widehat{C}^{-1}_{w,\Psi_n}\cdot\sqrt{n}\mathbb{E}_n\left[\int \left(\dot{\mathbf{H}}^c_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u})w(\mathbf{X};\mathbf{u}) + \dot{\mathbf{H}}'_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u})w^c(\mathbf{X},\mathbf{u})\right)\Psi_n(d\mathbf{u})\right.\\ &\qquad\left.\cdot\mathbf{h}_n(D,\mathbf{X};\boldsymbol{\beta}_0)\right]\end{aligned} \tag{C.5}$$

From Lemma B.3 we have that

$$\widehat{C}_{w,\Psi_n} = C_{w,\Psi} + o_p(1), \tag{C.6}$$

whereas by Lemma B.5 we have that

$$\sqrt{n}\mathbb{E}_n\left[\int \left(\dot{\mathbf{H}}^c_{w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u})w(\mathbf{X};\mathbf{u}) + \dot{\mathbf{H}}'_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w},\mathbf{u})w^c(\mathbf{X},\mathbf{u})\right)\Psi_n(d\mathbf{u})\cdot\mathbf{h}_n(D,\mathbf{X};\boldsymbol{\beta}_0)\right]$$

$$= \sqrt{n}\mathbb{E}_n \left[\int \left(\dot{\mathbf{H}}^c_w(\boldsymbol{\beta_0}, \mathbf{u}) w(\mathbf{X};\mathbf{u}) + \dot{\mathbf{H}}'_w(\boldsymbol{\beta_0}, \mathbf{u}) w^c(\mathbf{X}, \mathbf{u}) \right) \Psi(d\mathbf{u}) \cdot \mathbf{h}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right) \right] + o_p\left(1\right) \tag{C.7}$$

Thus, from (C.5)-(C.7), we conclude that

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{ips}_{n,w} - \boldsymbol{\beta}_0\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l_{w,\Psi}\left(D_i, \mathbf{X}_i; \boldsymbol{\beta}_0\right) + o_p\left(1\right),$$

with

$$l_{w,\Psi}\left(D_i, \mathbf{X}_i; \boldsymbol{\beta}_0\right) = -C^{-1}_{w,\Psi} \cdot \int \left(\dot{\mathbf{H}}^c_w(\boldsymbol{\beta}_0, \mathbf{u}) w(\mathbf{X};\mathbf{u}) + \dot{\mathbf{H}}'_w(\boldsymbol{\beta}_0, \mathbf{u}) w^c(\mathbf{X}, \mathbf{u}) \right) \Psi(d\mathbf{u}) \cdot \mathbf{h}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right).$$

Since $\dot{\mathbf{H}}^c_w(\boldsymbol{\beta}_0, \mathbf{u}) w(\mathbf{X};\mathbf{u}) + \dot{\mathbf{H}}'_w(\boldsymbol{\beta}_0, \mathbf{u}) w^c(\mathbf{X};\mathbf{u})$ is real-valued, under Assumptions 2-3, as long as $\mathbb{E}[||l_{w,\Psi}(D, X; \beta_0)||^2] < \infty$, the asymptotic normality result follows directly from the application of the standard central limit theorem. Next, we show that $l_{w,\Psi}(D, \mathbf{X}; \boldsymbol{\beta}_0)$ is indeed square integrable when $C_{w,\Psi}$ is positive definite. Here, it suffices to show that $\mathbb{E}\left[||s(D, \mathbf{X}; \boldsymbol{\beta_0})||^2\right] < \infty$ where

$$s(D, \mathbf{X}; \boldsymbol{\beta_0}) \equiv \int \left\{ \dot{\mathbf{H}}^c_w(\boldsymbol{\beta}_0, \mathbf{u}) w(\mathbf{X};\mathbf{u}) + \dot{\mathbf{H}}'_w(\boldsymbol{\beta}_0, \mathbf{u}) w^c(\mathbf{X};\mathbf{u}) \right\} \Psi(d\mathbf{u}) \cdot \mathbf{h}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right).$$

Let $K(\mathbf{x}, \mathbf{u};\boldsymbol{\beta_0}) \equiv \dot{\mathbf{H}}^c_w(\boldsymbol{\beta}_0, \mathbf{u}) w(\mathbf{x};\mathbf{u}) + \dot{\mathbf{H}}'_w(\boldsymbol{\beta}_0, \mathbf{u}) w^c(\mathbf{x};\mathbf{u})$. Then, for an arbitrary $C < \infty$,

$$\begin{aligned} ||s(d, \mathbf{x};\boldsymbol{\beta_0})||^2 &\le \int ||K(\mathbf{x}, \mathbf{u};\boldsymbol{\beta_0}) \mathbf{h}\left(d, \mathbf{x};\boldsymbol{\beta}_0\right)||^2 \Psi(d\mathbf{u}) \\ &\le \int ||K(\mathbf{x}, \mathbf{u};\boldsymbol{\beta_0})||^2 \cdot ||\mathbf{h}\left(d, \mathbf{x};\boldsymbol{\beta}_0\right)||^2 \Psi(d\mathbf{u}) \\ &\le C \cdot \int ||K(\mathbf{x}, \mathbf{u};\boldsymbol{\beta_0})||^2 \Psi(d\mathbf{u}), \end{aligned}$$

where the first inequality follows from Jensen's inequality, the second inequality follows from the Cauchy-Schwarz inequality, and the third from Assumption 2(ii). Finally, from Lemma B.1 we have $w(\mathbf{X};\mathbf{u})$ is uniformly bounded, and therefore,

$$\mathbb{E}\left[||K(\mathbf{X}, \mathbf{u};\boldsymbol{\beta_0})||^2\right] \le C \cdot ||\dot{\mathbf{H}}_w(\boldsymbol{\beta}_0, \mathbf{u})||^2$$

$$\leq C \cdot \left(\mathbb{E}\left[\left\|\dot{\mathbf{h}}\left(\mathbf{X};\boldsymbol{\beta}_0\right)\right\|\right]\right)^2$$

$$< \infty,$$

where the last inequality follows from Assumption 2(iii). This concludes our proof.■

Proof of Theorem 2: The proof is divided into three parts.

Part 1: Asymptotic Properties of the Average Treatment Effect.

It suffices to show that

$$\sqrt{n}\left(\widehat{ATE}_n^{ips} - ATE\right) = \sqrt{n}\mathbb{E}_n\left[\psi_{w,\Psi}^{ate}\left(Y,D,\mathbf{X}\right)\right] + o_p\left(1\right) \tag{C.8}$$

where $\psi_{w,\Psi}^{ate}\left(Y,D,\mathbf{X}\right) = g^{ate}(Y,X,D) - l_{w,\Psi}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)' \mathbf{G}_{\boldsymbol{\beta}}^{ate}$,

$$\mathbb{E}\left[\psi_{w,\Psi}^{ate}\left(Y,D,\mathbf{X}\right)\right] = 0$$

and

$$\mathbb{E}\left[\psi_{w,\Psi}^{ate}\left(Y,D,\mathbf{X}\right)^2\right] < \infty,$$

i.e., that $\sqrt{n}\left(\widehat{ATE}_n^{ips} - ATE\right)$ admits an asymptotically linear representation. We show this by combining the mean value theorem, continuos mapping theorem, and the results in Theorem 1.

We first show that

$$\mathbb{E}_n[w_{1,n}^{ps}(D,\mathbf{X};\hat{\boldsymbol{\beta}}_{n,w}^{ips})Y - \mathbb{E}\left[Y\left(1\right)\right]]$$
$$= \mathbb{E}_n\left[w_1^{ps}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\cdot\left(Y - \mathbb{E}\left[Y\left(1\right)\right]\right) - l_{w,\Psi}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)' \mathbf{G}_{\boldsymbol{\beta},1}^{ate}\right] + o_p\left(n^{-1/2}\right), \tag{C.9}$$

where

$$\mathbf{G}_{\boldsymbol{\beta},1}^{ate} = \mathbb{E}\left[\frac{g_1^{ate}}{p(\mathbf{X};\boldsymbol{\beta}_0)}\cdot\dot{p}\left(\mathbf{X};\boldsymbol{\beta}_0\right)\right],$$

and $g_1^{ate}\left(Y,D,\mathbf{X}\right) = w_1^{ps}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\cdot\left(Y - \mathbb{E}\left[Y\left(1\right)\right]\right)$. By the mean value theorem and some manipulations, we have that

$$\mathbb{E}_n[w_{1,n}^{ps}(D,\mathbf{X};\hat{\boldsymbol{\beta}}_{n,w}^{ips})Y]$$

$$= \mathbb{E}_n\left[w_{1,n}^{ps}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right) Y\right]$$

$$- \mathbb{E}_n\left[\frac{w_{1,n}^{ps}\left(D, \mathbf{X}; \tilde{\boldsymbol{\beta}}\right)\left(Y - \mathbb{E}_n\left[w_{1,n}^{ps}\left(D, \mathbf{X}; \tilde{\boldsymbol{\beta}}\right) Y\right]\right) \cdot \dot{p}\left(\mathbf{X}; \tilde{\boldsymbol{\beta}}\right)'}{p\left(\mathbf{X}; \tilde{\boldsymbol{\beta}}\right)}\right]\left(\widehat{\boldsymbol{\beta}}_{n,w}^{ips} - \boldsymbol{\beta}_0\right),$$

where $\tilde{\boldsymbol{\beta}}$ satisfies $\left\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\| \leq \left\|\widehat{\boldsymbol{\beta}}_{n,w}^{ips} - \boldsymbol{\beta}_0\right\|$. From Theorem 1, we have that

$$\begin{aligned}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{n,w}^{ips} - \boldsymbol{\beta}_0\right) &= \sqrt{n}\mathbb{E}_n\left[l_{w,\Psi}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right)\right] + o_p(1) \\ &= O_p(1),\end{aligned}$$

and therefore, by the CMT, under Assumptions 2-3,

$$\begin{aligned}&\mathbb{E}_n[w_{1,n}^{ps}(D, \mathbf{X}; \hat{\boldsymbol{\beta}}_{n,w}^{ips})Y] \\ &\qquad = \mathbb{E}_n\left[w_{1,n}^{ps}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right) Y\right] - \mathbb{E}_n\left[l_{w,\Psi}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right)' \cdot \mathbf{G}_{\boldsymbol{\beta},1}^{ate}\right] + o_p\left(n^{-1/2}\right). \quad \text{(C.10)}\end{aligned}$$

Given that, under Assumption 2,

$$\mathbb{E}_n\left[\frac{D}{p\left(\mathbf{X}; \boldsymbol{\beta}_0\right)}\right] - \mathbb{E}\left[\frac{D}{p\left(\mathbf{X}; \boldsymbol{\beta}_0\right)}\right] = O_p\left(n^{-1/2}\right),$$

we have that, by the CMT,

$$\begin{aligned}\mathbb{E}_n\left[w_{1,n}^{ps}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right) Y\right] &= \mathbb{E}_n\left[w_1^{ps}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right)\left(Y - \mathbb{E}\left[w_1^{ps}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right) Y\right]\right)\right] \\ &\quad + \mathbb{E}\left[w_1^{ps}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right) Y\right] + o_p\left(n^{-1/2}\right), \\ &= \mathbb{E}_n\left[w_1^{ps}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right)\left(Y - \mathbb{E}\left[w_1^{ps}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right) Y\right]\right)\right] \\ &\quad + \mathbb{E}\left[Y(1)\right] + o_p\left(n^{-1/2}\right), \qquad \text{(C.11)}\end{aligned}$$

where the last step follows from Assumption 1. Hence, from (C.10) and (C.11), we conclude the proof of (C.9).

By symmetry, we have that

$$\mathbb{E}_n[w_{0,n}^{ps}(D, \mathbf{X}; \hat{\boldsymbol{\beta}}_{n,w}^{ips})Y - \mathbb{E}\left[Y(0)\right]]$$

$$= \mathbb{E}_n \left[w_0^{ps} \left(D, \mathbf{X}; \boldsymbol{\beta}_0 \right) \cdot \left(Y - \mathbb{E}\left[Y\left(0\right)\right]\right) + l_{w,\Psi}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right)' \mathbf{G}^{ate}_{\boldsymbol{\beta},0} \right] + o_p\left(n^{-1/2}\right), \quad \text{(C.12)}$$

where

$$\mathbf{G}^{ate}_{\boldsymbol{\beta},0} = \mathbb{E}\left[\frac{g_0^{ate}}{1 - p(\mathbf{X}; \boldsymbol{\beta}_\mathbf{0})} \cdot \dot{p}\left(\mathbf{X}; \boldsymbol{\beta}_\mathbf{0}\right)\right],$$

and $g_0^{ate}\left(Y, D, \mathbf{X}\right) = w_0^{ps}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right) \cdot \left(Y - \mathbb{E}\left[Y\left(0\right)\right]\right)$.

By combining (C.9) with (C.12), we have that

$$\sqrt{n}\left(\widehat{ATE}_n^{ips} - ATE\right) = \mathbb{E}_n\left[\psi^{ate}_{w,\Psi}\left(Y, D, \mathbf{X}\right)\right] + o_p\left(1\right),$$

where $\mathbb{E}\left[\psi^{ate}_{w,\Psi}\left(Y, D, \mathbf{X}\right)\right] = 0$ follows from the law of iterated expectations and Assumption 1. Next, note that

$$\begin{aligned}
\mathbb{E}[\psi^{ate}_{w,\Psi}(Y, D, \mathbf{X})^2] &= \mathbb{E}[(g^{ate}\left(Y, D, \mathbf{X}\right) - l_{w,\Psi}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right)' \mathbf{G}^{ate}_{\boldsymbol{\beta}})^2] \\
&= \mathbb{E}\left[g^{ate2} - 2 g^{ate} l'_{w,\Psi} \mathbf{G}^{ate}_{\boldsymbol{\beta}} + (l'_{w,\Psi}\mathbf{G}^{ate}_{\boldsymbol{\beta}})^2 \right] \\
&\leq \mathbb{E}[g^{ate2}] + \mathbb{E}[(l'_{w,\Psi}\mathbf{G}^{ate}_{\boldsymbol{\beta}})^2] + 2\mathbb{E}[|g^{ate} l'_{w,\Psi}\mathbf{G}^{ate}_{\boldsymbol{\beta}}|] \qquad \text{(C.13)}
\end{aligned}$$

Let $C_1 \equiv \sup_{d,\mathbf{x}} w_1^{ps}\left(d, \mathbf{x}; \boldsymbol{\beta}_0\right)^2$ and $C_2 \equiv \sup_{d,\mathbf{x}} w_0^{ps}\left(d, \mathbf{x}; \boldsymbol{\beta}_0\right)^2$, and note that, under Assumption 2(ii), $1 \leq C_1, C_2 < \infty$. Then

$$\begin{aligned}
\mathbb{E}[g^{ate2}] &= \mathbb{E}[w_1^{ps2}(Y - \mathbb{E}[Y(1)])^2 + w_0^{ps2}(Y - \mathbb{E}[Y(0)])^2] \\
&\leq C_1 \mathbb{E}[(Y\left(1\right) - \mathbb{E}[Y(1)])^2] + C_2 \mathbb{E}[(Y\left(0\right) - \mathbb{E}[Y(0)])^2] \\
&< \infty, \qquad \text{(C.14)}
\end{aligned}$$

where the first equality follows from $w_1^{ps} \cdot w_0^{ps} = 0 \; a.s.$, the first inequality follows from Assumption (1) and Assumption 2(ii), whereas the last inequality follows from Assumption 4(i).

Next, by Cauchy-Schwarz inequality, Theorem 1, and Assumption 4(ii),

$$\begin{aligned}
\mathbb{E}[(l'_{w,\Psi}\mathbf{G}^{ate}_{\boldsymbol{\beta}})^2] &\leq ||\mathbf{G}^{ate}_{\boldsymbol{\beta}}||^2 \cdot \mathbb{E}[||l_{w,\Psi}||^2] \\
&< \infty, \qquad \text{(C.15)}
\end{aligned}$$

whereas, by Cauchy-Schwarz inequality, (C.14) and (C.15),

$$\mathbb{E}[|g^{ate} l'_{w,\Psi} \mathbf{G}^{ate}_{\boldsymbol{\beta}}|] \leq \mathbb{E}[|g^{ate}|^2]^{1/2} \cdot \mathbb{E}[|l'_{w,\Psi} \mathbf{G}^{ate}_{\boldsymbol{\beta}}|^2]^{1/2}$$

$$< \infty. \tag{C.16}$$

Hence, $\mathbb{E}[\psi^{ate}_{w,\Psi}(Y,D,\mathbf{X})^2] < \infty$ follows from (C.13)-(C.16), which concludes the proof of (C.8).

Part 2: Asymptotic Properties of the Distribution Treatment Effects.

The (uniform) asymptotic linear representation for the Distribution Treatment Effect parameter follows from exactly the same steps as in Part 1 and is therefore omitted. Next, we show that the classes of functions

$$\mathcal{F}_{1,dte} \equiv \left\{ (z,d,\mathbf{x}) \in \mathbb{R} \times \{0,1\} \times \mathcal{X} \mapsto \psi^{dte}_{1,w,\Psi}(z,d,\mathbf{x};y) : y \in [-\infty,\infty] \right\},$$

$$\mathcal{F}_{0,dte} \equiv \left\{ (z,d,\mathbf{x}) \in \mathbb{R} \times \{0,1\} \times \mathcal{X} \mapsto \psi^{dte}_{0,w,\Psi}(z,d,\mathbf{x};y) : y \in [-\infty,\infty] \right\}$$

are Donsker, where

$$\psi^{dte}_{1,w,\Psi}(z,d,\mathbf{x};y) = g^{dte}_1(z,d,\mathbf{x};y) - l_{w,\Psi}(d,\mathbf{x};\boldsymbol{\beta}_0)' \cdot \mathbf{G}^{dte}_{1,\boldsymbol{\beta}}(y),$$

$$\psi^{dte}_{0,w,\Psi}(z,d,\mathbf{x};y) = g^{dte}_0(z,d,\mathbf{x};y) + l_{w,\Psi}(d,\mathbf{x};\boldsymbol{\beta}_0)' \cdot \mathbf{G}^{dte}_{0,\boldsymbol{\beta}}(y),$$

and

$$\mathbf{G}^{dte}_{1,\boldsymbol{\beta}}(y) = \mathbb{E}\left[\frac{g^{dte}_1(Y,D,\mathbf{X};y)}{p(\mathbf{X};\boldsymbol{\beta}_0)} \cdot \dot{p}(\mathbf{X};\boldsymbol{\beta}_0) \right],$$

$$\mathbf{G}^{dte}_{0,\boldsymbol{\beta}}(y) = \mathbb{E}\left[\frac{g^{dte}_0(Y,D,\mathbf{X};y)}{1-p(\mathbf{X};\boldsymbol{\beta}_0)} \cdot \dot{p}(\mathbf{X};\boldsymbol{\beta}_0) \right].$$

Toward this end, note that the classes of functions $\left\{ l_{w,\Psi}(d,\mathbf{x};\boldsymbol{\beta}_0)' \cdot \mathbf{G}^{dte}_{1,\boldsymbol{\beta}}(y) : y \in [-\infty,\infty] \right\}$ and $\left\{ l_{w,\Psi}(d,\mathbf{x};\boldsymbol{\beta}_0)' \cdot \mathbf{G}^{dte}_{0,\boldsymbol{\beta}}(y) : y \in [-\infty,\infty] \right\}$are Donsker since they depend on y in a deterministic manner, $\mathbf{G}^{dte}_{d,\boldsymbol{\beta}}(y) < \infty$, $d \in \{0,1\}$, and, by Theorem 1, $\mathbb{E}[||l_{w,\Psi}||^2] < \infty$. The Donsker property of $\left\{ g^{dte}_1(z,d,\mathbf{x};y) : y \in [-\infty,\infty] \right\}$ and $\left\{ g^{dte}_0(z,d,\mathbf{x};y) : y \in [-\infty,\infty] \right\}$ follows from Lemma B.1, Assumption 2(ii), and Corollary 9.32 in Kosorok (2008). Thus, from Corollary 9.32 in Kosorok (2008), we conclude that $\mathcal{F}_{1,dte}$ and $\mathcal{F}_{0,dte}$ are Donsker.

Let $\boldsymbol{\lambda}^{dte}_{w,\Psi}(z,d,\mathbf{x};\cdot) = \left(\psi^{dte}_{1,w,\Psi}(z,d,\mathbf{x};\cdot), \psi^{dte}_{0,w,\Psi}(z,d,\mathbf{x};\cdot)\right)'$, and denote

$$\mathbb{G}^{dte,(1,0)}_{n,w,\Psi}(\cdot) = \sqrt{n}\mathbb{E}_n\left[\boldsymbol{\lambda}^{dte}_{w,\Psi}(Y,D,\mathbf{X};\cdot)\right].$$

Thus, under Assumptions 1-4,

$$\mathbb{G}^{dte,(1,0)}_{n,w,\Psi}(\cdot) \Rightarrow \mathbb{G}^{dte,(1,0)}_{\infty,w,\Psi}(\cdot) \text{ in } \ell^{\infty}([-\infty,\infty]) \times \ell^{\infty}([-\infty,\infty]), \tag{C.17}$$

where $\Rightarrow$ denotes weak convergence in the sense of J. Hoffman-Jørgensen, see e.g. van der Vaart and Wellner (1996), $\ell^{\infty}(T)$ is the collection of all bounded functions $f : T \mapsto \mathbb{R}$, and $\mathbb{G}^{dte,(1,0)}_{\infty,w,\Psi}(\cdot)$ is a tight, two-dimensional mean zero Gaussian process with covariance kernel

$$\Gamma(\mathbf{y}_1,\mathbf{y}_2) = \mathbb{E}\left[\boldsymbol{\lambda}^{dte}_{w,\Psi}(Y,D,\mathbf{X};\mathbf{y}_1)\,\boldsymbol{\lambda}^{dte}_{w,\Psi}(Y,D,\mathbf{X};\mathbf{y}_2)'\right],$$

in which

$$\boldsymbol{\lambda}^{dte}_{w,\Psi}(z,d,\mathbf{x};\mathbf{y}) = \left(\psi^{dte}_{1,w,\Psi}(z,d,\mathbf{x};y_1), \psi^{dte}_{0,w,\Psi}(z,d,\mathbf{x};y_2)\right)'.$$

By the CMT, we have that

$$\sqrt{n}\left(\widehat{DTE}^{ips}_n - DTE\right)(\cdot) = (1,-1)\,\mathbb{G}^{dte,(1,0)}_{n,w,\Psi}(\cdot) + o_p(1),$$

$$\Rightarrow \mathbb{G}^{dte}_{\infty,w,\Psi}(\cdot) \text{ in } \ell^{\infty}([-\infty,\infty])$$

where $\mathbb{G}^{dte}_{\infty,w,\Psi}(\cdot)$ is a tight, univariate mean zero Gaussian process with covariance kernel

$$\Gamma_{dte}(y_1,y_2) = \mathbb{E}\left[\psi^{dte}_{w,\Psi}(Y,D,\mathbf{X};y_1)\,\psi^{dte}_{w,\Psi}(Y,D,\mathbf{X};y_2)\right].$$

The asymptotic normality result now follows by simply fixing y.

Part 3: Asymptotic Properties of the Quantile Treatment Effects.

Define $\widehat{\mathbf{q}}^{ips}_n(\boldsymbol{\tau}) = \left(\widehat{q}^{ips}_{n,Y(1)}(\tau_1), \widehat{q}^{ips}_{n,Y(0)}(\tau_2)\right)'$, $\mathbf{q}(\boldsymbol{\tau}) = \left(q_{Y(1)}(\tau_1), q_{Y(0)}(\tau_2)\right)'$, $\mathbf{f}^{-1}(\tau) = \left(f^{-1}_{Y(1)}\left(q_{Y(1)}(\tau_1)\right), f^{-1}_{Y(0)}\left(q_{Y(0)}(\tau_2)\right)\right)'$ and $\boldsymbol{\tau} = (\tau_1,\tau_2) \in [a_1,a_2]^2$, where a_1 and a_2 satisfy $0 < a_1 \le a_2 < 1$. Under Assumptions 1-4, we have that, from (C.17), Lemma 21.4 in van der Vaart (1998), and the functional delta method, see e.g. Theorem 20.8 in van der

Vaart (1998),

$$\sqrt{n}\left(\widehat{\mathbf{q}}_n^{ips}-\mathbf{q}\right)(\cdot)=-\mathbf{f}^{-1}(\cdot)'\cdot\mathbb{G}_{n,w,\Psi}^{dte,(1,0)}\left(\mathbf{q}(\cdot)\right)+o_p(1)$$

$$\Rightarrow -\mathbf{f}^{-1}(\cdot)'\cdot\mathbb{G}_{\infty,w,\Psi}^{dte,(1,0)}\left(\mathbf{q}(\cdot)\right)\text{ in }\ell^{\infty}\left([a_1,a_2]\right)\times\ell^{\infty}\left([a_1,a_2]\right).$$

Then, by the CMT,

$$\sqrt{n}\left(\widehat{QTE}_n^{ips}-QTE\right)(\cdot)=(1,-1)\cdot\left(-\mathbf{f}^{-1}(\cdot)'\cdot\mathbb{G}_{n,w,\Psi}^{dte,(1,0)}\left(\mathbf{q}(\cdot)\right)\right)+o_p(1)$$

$$\Rightarrow \mathbb{G}_{\infty,w,\Psi}^{qte}(\cdot)\text{ in }\ell^{\infty}[a_1,a_2]$$

where $\mathbb{G}_{\infty,w,\Psi}^{qte}(\cdot)$ is a tight, mean zero Gaussian process with covariance kernel

$$\Gamma_{qte}(\tau_1,\tau_2)=\mathbb{E}\left[\psi_{w,\Psi}^{qte}(Y,D,\mathbf{X};\tau_1)\,\psi_{w,\Psi}^{dte}(Y,D,\mathbf{X};\tau_2)\right].$$

The asymptotic normality result now follows by simply fixing τ.■

Proof of Theorem 3: The proof goes along the exact same lines as that of Theorem 2 and is therefore omitted.

References

Chang, M. N. (1990), "Weak Convergence of a Self-Consistent Estimator of the Survival Function with Doubly Censored Data," *The Annals of Statistics*, 18, 391–404.

Escanciano, J. C. (2006), "A consistent diagnostic test for regression models using projections," *Econometric Theory*, 22, 1030–1051.

Kosorok, M. R. (2008), *Introduction to empirical processes and semiparametric inference*, New York, NY: Springer.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge: Cambridge University Press.

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer.