Covariate Distribution Balance via Propensity Scores: Supplemental Appendix

Pedro H. C. Sant'Anna* Vanderbilt University

Xiaojun Song† Peking University

Qi Xu‡ Vanderbilt University

April 3, 2020

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. Appendix C collects all the proofs of the main results of the paper, while Appendix D collects the proofs of the results when treatment allocation is endogenous. Finally, Appendix E presents results for the case when one is interested in average, distributional and quantile treatment effects on the treated subpopulation.

*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

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

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, $\mathbf{H}_{n,w}(\boldsymbol{\beta},\mathbf{u}) = \mathbb{E}_n\left[\mathbf{h}_n\left(D,\mathbf{X};\boldsymbol{\beta}\right)w(\mathbf{X};\mathbf{u})\right]$, with $\mathbf{h}_n\left(D,\mathbf{X};\boldsymbol{\beta}\right) = \left(h_{n,1}\left(D,\mathbf{X};\boldsymbol{\beta}\right), h_{n,0}\left(D,\mathbf{X};\quad\boldsymbol{\beta}\right)\right)'$, $h_{n,d}\left(D,\mathbf{X};\boldsymbol{\beta}\right) = \varpi_{n,d}^{ps}\left(D,\mathbf{X};\boldsymbol{\beta}\right) - 1$, $d\in\{0,1\}$, and

$$\varpi_{n,1}^{ps}\left(D,\mathbf{X};\boldsymbol{\beta}\right) = \frac{D}{p\left(\mathbf{X};\boldsymbol{\beta}\right)}\Bigg/\ \mathbb{E}_n\left[\frac{D}{p\left(\mathbf{X};\boldsymbol{\beta}\right)}\right],$$

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

In the following, we derive the closed-form representation of the IPS objective function of the estimators in (2.10)-(2.12). In light of Remark E.1, 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 (2.10), in this case we have $\Psi_n\left(\mathbf{u}\right) = F_{n,\mathbf{X}}\left(\mathbf{u}\right)$ and $w\left(\mathbf{X};\mathbf{u}\right) = 1\{\mathbf{X}\le\mathbf{u}\}$. For any $\mathbf{g}\left(\mathbf{x}\right)$, $\int \mathbf{g}\left(\mathbf{u}\right)F_{n,\boldsymbol{X}}\left(d\mathbf{u}\right) = n^{-1}\sum_{j=1}^{n}\mathbf{g}\left(\mathbf{X}_j\right)$. Since $1\{\mathbf{X}\le\mathbf{u}\}$ is real-valued, conjugate transpose reduces to direct transpose. Hence,

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

and

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

$$+\frac{1}{n^3}\sum_{l=1}^{n}\left(\sum_{j=1}^{n}h_{n,0}\left(D_j,\mathbf{X}_j;\boldsymbol{\beta}\right)1\left(\mathbf{X}_j\leq\mathbf{X}_l\right)\right)^2.$$

Case 2: Projection Weights

As introduced in (2.11), in this case we have $\Psi_n\left(\mathbf{u}\right)=n^{-1}\sum_{i=1}^{n}1\left(\boldsymbol{\gamma}'\mathbf{X}_j\leq u\right)\times\boldsymbol{\gamma}$ and $w\left(\mathbf{X};\mathbf{u}\right)=1\{\boldsymbol{\gamma}'\mathbf{X}\leq u\}$. Before proceeding to the detailed derivations, from Appendix B of Escanciano (2006), we have that

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

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

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

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

$$A_{jsr}^{(0)}=\begin{cases}2\pi & if\ X_i=X_r=X_j,\\ \pi & if\ X_i=X_j,\ X_i=X_r\ or\ X_j=X_r,\\ \left|\pi-\arccos\left(\frac{\left(X_i-X_r\right)'\left(X_j-X_r\right)}{\left\|X_i-X_r\right\|\left\|X_j-X_r\right\|}\right)\right| & otherwise.\end{cases}$$

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

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

$$= \frac{1}{n^2}\sum_{j=1}^{n}\sum_{s=1}^{n} h_{n,1}\left(D_j,\mathbf{X}_j;\boldsymbol{\beta}\right) h_{n,1}\left(D_s,\mathbf{X}_s;\boldsymbol{\beta}\right) A_{js}$$

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

Case 3: Exponential Weights

Finally, as introduced in (2.12), now we have $w\left(\mathbf{X};\mathbf{u}\right) = exp\left(i\mathbf{u}'\Phi\left(\mathbf{X}\right)\right)$ and

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

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

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

To get the last equality, we exploit that

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

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 help 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 $1(\boldsymbol{x}\leq\boldsymbol{u})\leq 1$, $1\left\{\boldsymbol{\gamma}'\mathbf{x}\leq u\right\}\leq 1$ 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 $\exp(i\mathbf{u}'\Phi\left(\mathbf{x}\right))$ is infinitely differentiable with respect to $\mathbf{u}$, and all derivatives are uniformly bounded on Π, the Donsker property of $\mathcal{W}_{\text{exp}}$ 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*

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

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

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

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

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 $\mathcal{W}_{\text{exp}}$ are uniformly bounded Donsker classes of functions, and therefore they are also GC, see ,e.g., Lemma 8.17 in Kosorok (2008). Finally, by Corollary 9.26 in Kosorok (2008), $\mathcal{F}_3$ and $\mathcal{F}_4$ are GC. ■

Let

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

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

and

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

where $\phi\left(\mathbf{u}\right)$ is the standard k-variate normal density function and $\widetilde{\boldsymbol{\beta}}$ satisfies $\left\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\| \leq \left\|\widehat{\boldsymbol{\beta}}^{ips}_{n,w}-\boldsymbol{\beta}_0\right\|$. Furthermore, write

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

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

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

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

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

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

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

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

Toward this end, as in the consistency proof of Theorem 3.1, fix an arbitrarily small $\epsilon > 0$ and choose a compact and convex set Π such that

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

Then, write

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

with

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

Let

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

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

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

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

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

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

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

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,*

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

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

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

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

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

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

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

and let

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

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

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

Lemma B.5 *Under Assumption 2,*

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

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

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

Proof of Lemma B.5: First note that

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

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 $\mathbf{u}\in[-\infty,\infty]^k$,

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

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

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

Let

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

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

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

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

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

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

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

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

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

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 2.1: 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 (2.1) is equivalent to (2.4), implying that $Q_w(\boldsymbol{\beta}_0) = 0$.

To complete the proof we then need to show that if $Q_w(\boldsymbol{\beta}) = 0$, then $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. Towards this end, recall that 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 $\mathbb{E}\left[\mathbf{h}(D, \mathbf{X}; \boldsymbol{\beta})\,|\mathbf{X}\right] = 0$ $a.s.$, which is equivalent to

$$\mathbb{E}\left[\frac{D}{p(\mathbf{X}; \boldsymbol{\beta})}\middle| \mathbf{X}\right] = \mathbb{E}\left[\frac{D}{p(\mathbf{X}; \boldsymbol{\beta})}\right] \;\; a.s. \text{ and } \mathbb{E}\left[\frac{1-D}{1-p(\mathbf{X}; \boldsymbol{\beta})}\middle| \mathbf{X}\right] = \mathbb{E}\left[\frac{1-D}{1-p(\mathbf{X}; \boldsymbol{\beta})}\right] a.s..$$

The above is further equivalent to

$$\frac{p(\mathbf{X}; \boldsymbol{\beta}_0)}{p(\mathbf{X}; \boldsymbol{\beta})} = \mathbb{E}\left[\frac{p(\mathbf{X}; \boldsymbol{\beta}_0)}{p(\mathbf{X}; \boldsymbol{\beta})}\right] \;\; a.s. \text{ and } \frac{1-p(\mathbf{X}; \boldsymbol{\beta}_0)}{1-p(\mathbf{X}; \boldsymbol{\beta})} = \mathbb{E}\left[\frac{1-p(\mathbf{X}; \boldsymbol{\beta}_0)}{1-p(\mathbf{X}; \boldsymbol{\beta})}\right] a.s..$$

That is, there exist some constants c_1 and c_0, potentially depending on β, such that $p(\mathbf{X}; \boldsymbol{\beta}_0) = c_1 p(\mathbf{X}; \boldsymbol{\beta})$ $a.s.$ and $1 - p(\mathbf{X}; \boldsymbol{\beta}_0) = c_0(1 - p(\mathbf{X}; \boldsymbol{\beta}))$ $a.s.$, which imply

$$(c_1 - c_0)\, p(\mathbf{X}; \boldsymbol{\beta}) = 1 - c_0 \quad a.s. \tag{C.1}$$

If $c_1 \neq c_0$, then $p(\mathbf{X}; \boldsymbol{\beta}) = (1 - c_0)/(c_1 - c_0)$ $a.s.$ so that $p(\mathbf{X}; \boldsymbol{\beta})$ degenerates to a constant.

This leads to a contradiction. In light of (C.1), we then conclude that $c_1 = c_0 = 1$ and that $\mathbf{H}_w(\boldsymbol{\beta}, \boldsymbol{u}) = 0$ *a.e.* on Π is equivalent to $p(\mathbf{X}; \boldsymbol{\beta}) = p(\mathbf{X}; \boldsymbol{\beta}_0)$ *a.s.*. Given that $p(\mathbf{X}; \boldsymbol{\beta}) = p(\mathbf{X}; \boldsymbol{\beta}_0)$ *a.s.* for a unique $\boldsymbol{\beta}_0$, we must have $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. This concludes the proof. ■

Proof of Theorem 3.1: 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 2.1 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

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

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

$$\left| \int_{\mathbb{R}^k\setminus\Pi} \phi(\mathbf{u})\, d\mathbf{u} \right| \leq \epsilon. \tag{C.2}$$

Then, write

$$\int_{\mathbb{R}^k} ||\mathbf{H}_{n,\exp}(\boldsymbol{\beta}, \mathbf{u})||^2 \phi(\mathbf{u})\, d\mathbf{u}$$

$$= \int_{\Pi} ||\mathbf{H}_{n,\exp}(\boldsymbol{\beta}, \mathbf{u})||^2 \phi(\mathbf{u})\, d\mathbf{u} + \int_{\mathbb{R}^k \setminus \Pi} ||\mathbf{H}_{n,\exp}(\boldsymbol{\beta}, \mathbf{u})||^2 \phi(\mathbf{u})\, d\mathbf{u}$$

$$\equiv J_{1n} + J_{2n}.$$

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.2) 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

$$\int \left\{ \dot{\mathbf{H}}^c_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w}, \mathbf{u}) \mathbf{H}_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w}, \mathbf{u}) + \left(\mathbf{H}^c_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w}, \mathbf{u}) \dot{\mathbf{H}}_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w}, \mathbf{u}) \right)' \right\} \Psi_n(d\mathbf{u}) = 0. \tag{C.3}$$

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

$$\mathbf{H}_{n,w}(\widehat{\boldsymbol{\beta}}^{ips}_{n,w}, \mathbf{u}) = \mathbf{H}_{n,w}(\boldsymbol{\beta}_0, \mathbf{u}) + \dot{\mathbf{H}}_{n,w}(\widetilde{\boldsymbol{\beta}}, \mathbf{u})(\widehat{\boldsymbol{\beta}}^{ips}_{n,w} - \boldsymbol{\beta}_0), \tag{C.4}$$

where $\widetilde{\boldsymbol{\beta}}$ satisfies $\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\| \leq \left\|\widehat{\boldsymbol{\beta}}^{ips}_{n,w} - \boldsymbol{\beta}_0\right\|$. Plugging (C.4) into (C.3), 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

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

Therefore

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{n,w}^{ips}-\boldsymbol{\beta}_0\right)$$
$$=-\widehat{C}_{w,\Psi_n}^{-1}\cdot\sqrt{n}\int\left(\dot{\mathbf{H}}_{n,w}^{c}(\widehat{\boldsymbol{\beta}}_{n,w}^{ips},\mathbf{u})\mathbf{H}_{n,w}(\boldsymbol{\beta}_0,\mathbf{u})+\dot{\mathbf{H}}_{n,w}'(\widehat{\boldsymbol{\beta}}_{n,w}^{ips},\mathbf{u})\left(\mathbf{H}_{n,w}'(\boldsymbol{\beta}_0,\mathbf{u})\right)^{c}\right)\Psi_n(d\mathbf{u}). \quad \text{(C.5)}$$

By exploiting that $\mathbf{H}_{n,w}\left(\boldsymbol{\beta}_0,\mathbf{u}\right)=\mathbb{E}_n\left[\mathbf{h}_n\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)w(\mathbf{X};\mathbf{u})\right]$, we can express (C.5) as

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{n,w}^{ips}-\boldsymbol{\beta}_0\right)$$
$$=-\widehat{C}_{w,\Psi_n}^{-1}\cdot\sqrt{n}\mathbb{E}_n\left[\int\left(\dot{\mathbf{H}}_{n,w}^{c}(\widehat{\boldsymbol{\beta}}_{n,w}^{ips},\mathbf{u})w(\mathbf{X};\mathbf{u})+\dot{\mathbf{H}}_{n,w}'(\widehat{\boldsymbol{\beta}}_{n,w}^{ips},\mathbf{u})w^{c}(\mathbf{X};\mathbf{u})\right)\Psi_n(d\mathbf{u})\right.$$
$$\left.\cdot\mathbf{h}_n\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\right] \quad \text{(C.6)}$$

From Lemma B.3 we have that

$$\widehat{C}_{w,\Psi_n}=C_{w,\Psi}+o_p\left(1\right), \quad \text{(C.7)}$$

whereas by Lemma B.5 we have that

$$\sqrt{n}\mathbb{E}_n\left[\int\left(\dot{\mathbf{H}}_{w}^{c}(\widehat{\boldsymbol{\beta}}_{n,w}^{ips},\mathbf{u})w(\mathbf{X};\mathbf{u})+\dot{\mathbf{H}}_{n,w}'(\widehat{\boldsymbol{\beta}}_{n,w}^{ips},\mathbf{u})w^{c}(\mathbf{X};\mathbf{u})\right)\Psi_n(d\mathbf{u})\cdot\mathbf{h}_n\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\right]$$
$$=\sqrt{n}\mathbb{E}_n\left[\int\left(\dot{\mathbf{H}}_{w}^{c}(\boldsymbol{\beta}_{\mathbf{0}},\mathbf{u})w(\mathbf{X};\mathbf{u})+\dot{\mathbf{H}}_{w}'(\boldsymbol{\beta}_{\mathbf{0}},\mathbf{u})w^{c}(\mathbf{X},\mathbf{u})\right)\Psi(d\mathbf{u})\cdot\mathbf{h}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\right]+o_p\left(1\right). \quad \text{(C.8)}$$

Thus, from (C.6)-(C.8), we conclude that

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

with

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

Since $\dot{\mathbf{H}}_{w}^{c}(\boldsymbol{\beta}_0,\mathbf{u})w(\mathbf{X};\mathbf{u})+\dot{\mathbf{H}}_{w}'(\boldsymbol{\beta}_0,\mathbf{u})w^{c}(\mathbf{X};\mathbf{u})$ 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 $\mathbb{E}\left[||s(D,\mathbf{X};\boldsymbol{\beta}_{\mathbf{0}})||^2\right]<\infty$ where

$$s(D,\mathbf{X};\boldsymbol{\beta}_{\mathbf{0}})\equiv\int\left\{\dot{\mathbf{H}}_{w}^{c}(\boldsymbol{\beta}_0,\mathbf{u})w(\mathbf{X};\mathbf{u})+\dot{\mathbf{H}}_{w}'(\boldsymbol{\beta}_0,\mathbf{u})w^{c}(\mathbf{X};\mathbf{u})\right\}\Psi(d\mathbf{u})\cdot\mathbf{h}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right).$$

Let $K(\mathbf{x},\mathbf{u};\boldsymbol{\beta}_{\mathbf{0}})\equiv\dot{\mathbf{H}}_{w}^{c}(\boldsymbol{\beta}_0,\mathbf{u})w(\mathbf{x};\mathbf{u})+\dot{\mathbf{H}}_{w}'(\boldsymbol{\beta}_0,\mathbf{u})w^{c}(\mathbf{x};\mathbf{u})$. Then, for some $C<\infty$,

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

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,

$$
\begin{aligned}
\mathbb{E}\left[||K(\mathbf{X},\mathbf{u};\boldsymbol{\beta_0})||^2\right] &\leq C\cdot||\dot{\mathbf{H}}_w(\boldsymbol{\beta}_0,\mathbf{u})||^2 \\
&\leq C\cdot\left(\mathbb{E}\left[\left\|\dot{\mathbf{h}}\left(\mathbf{X};\boldsymbol{\beta}_0\right)\right\|\right]\right)^2 \\
&< \infty,
\end{aligned}
$$

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

Proof of Theorem 3.2: The proof is divided into three parts.

Part 1: Asymptotic Properties of the Average Treatment Effect.

It suffices to show that

$$
\sqrt{n}\left(\widehat{ATE}_n^{ips}-ATE\right)=\sqrt{n}\mathbb{E}_n\left[\psi_{w,\Psi}^{ate}\left(Y,D,\mathbf{X}\right)\right]+o_p\left(1\right) \tag{C.9}
$$

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 3.1.

We first show that

$$
\begin{aligned}
&\mathbb{E}_n[\varpi_{n,1}^{ps}(D,\mathbf{X};\widehat{\boldsymbol{\beta}}_{n,w}^{ips})Y-\mathbb{E}\left[Y\left(1\right)\right]] \\
&\qquad=\mathbb{E}_n\left[\varpi_1^{ps}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)\cdot\left(Y-\mathbb{E}\left[Y\left(1\right)\right]\right)-l_{w,\Psi}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)'\mathbf{G}_{\boldsymbol{\beta},1}^{ate}\right]+o_p\left(n^{-1/2}\right),
\end{aligned} \tag{C.10}
$$

where

$$\mathbf{G}_{\boldsymbol{\beta},1}^{ate} = \mathbb{E}\left[\frac{g_1^{ate}}{p(\mathbf{X};\boldsymbol{\beta}_\mathbf{0})} \cdot \dot{p}\left(\mathbf{X};\boldsymbol{\beta}_\mathbf{0}\right)\right],$$

and $g_1^{ate}\left(Y, D, \mathbf{X}\right) = \varpi_1^{ps}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right) \cdot \left(Y - \mathbb{E}\left[Y\left(1\right)\right]\right)$. By the mean value theorem and some manipulations, we have that

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

where $\widetilde{\boldsymbol{\beta}}$ satisfies $\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\| \leq \left\|\widehat{\boldsymbol{\beta}}_{n,w}^{ips} - \boldsymbol{\beta}_0\right\|$. From Theorem 3.1, we have that

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

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

$$\begin{aligned}
&\mathbb{E}_n[\varpi_{n,1}^{ps}(D, \mathbf{X}; \widehat{\boldsymbol{\beta}}_{n,w}^{ips})Y] \\
&\qquad = \mathbb{E}_n\left[\varpi_{n,1}^{ps}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right) Y\right] - \mathbb{E}_n\left[l_{w,\Psi}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right)' \cdot \mathbf{G}_{\boldsymbol{\beta},1}^{ate}\right] + o_p\left(n^{-1/2}\right). \quad \text{(C.11)}
\end{aligned}$$

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,

$$\begin{aligned}
\mathbb{E}_n\left[\varpi_{n,1}^{ps}\left(D, \mathbf{X}; \boldsymbol{\beta}_0^{lte}\right) Y\right] &= \mathbb{E}_n\left[\varpi_1^{ps}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right)\left(Y - \mathbb{E}\left[\varpi_1^{ps}\left(D, \mathbf{X}; \boldsymbol{\beta}_0\right) Y\right]\right)\right] \\
&+ \mathbb{E}\left[\varpi_1^{ps}\left(D, \mathbf{X}; \boldsymbol{\beta}_0^{lte}\right) Y\right] + o_p\left(n^{-1/2}\right), \\
&= \mathbb{E}_n\left[\varpi_1^{ps}\left(D, \mathbf{X}; \boldsymbol{\beta}_0^{lte}\right)\left(Y - \mathbb{E}\left[\varpi_1^{ps}\left(D, \mathbf{X}; \boldsymbol{\beta}_0^{lte}\right) Y\right]\right)\right] \\
&+ \mathbb{E}\left[Y\left(1\right)\right] + o_p\left(n^{-1/2}\right), \qquad \text{(C.12)}
\end{aligned}$$

where the last step follows from Assumption 1. Hence, from (C.11) and (C.12), we conclude the proof of (C.10).

By symmetry, we have that

$$\mathbb{E}_n[\varpi^{ps}_{n,0}(D,\mathbf{X};\widehat{\boldsymbol{\beta}}^{ips}_{n,w})Y-\mathbb{E}\left[Y\left(0\right)\right]]$$
$$=\mathbb{E}_n\left[\varpi^{ps}_{0}\left(D,\mathbf{X};\boldsymbol{\beta}^{lte}_{0}\right)\cdot\left(Y-\mathbb{E}\left[Y\left(0\right)\right]\right)+l_{w,\Psi}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)'\mathbf{G}^{ate}_{\boldsymbol{\beta},0}\right]+o_p\left(n^{-1/2}\right), \quad (C.13)$$

where

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

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

By combining (C.10) with (C.13), we have that

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

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

$$\begin{aligned}\mathbb{E}[\psi^{ate}_{w,\Psi}(Y,D,\mathbf{X})^2]&=\mathbb{E}[(g^{ate}\left(Y,D,\mathbf{X}\right)-l_{w,\Psi}\left(D,\mathbf{X};\boldsymbol{\beta}_0\right)'\mathbf{G}^{ate}_{\boldsymbol{\beta}})^2]\\&=\mathbb{E}\left[g^{ate2}-2g^{ate}l'_{w,\Psi}\mathbf{G}^{ate}_{\boldsymbol{\beta}}+(l'_{w,\Psi}\mathbf{G}^{ate}_{\boldsymbol{\beta}})^2\right]\\&\leq\mathbb{E}[g^{ate2}]+\mathbb{E}[(l'_{w,\Psi}\mathbf{G}^{ate}_{\boldsymbol{\beta}})^2]+2\mathbb{E}[|g^{ate}l'_{w,\Psi}\mathbf{G}^{ate}_{\boldsymbol{\beta}}|] \qquad (C.14)\end{aligned}$$

Let $C_1\equiv\sup_{d,\mathbf{x}}\varpi^{ps}_1\left(d,\mathbf{x};\boldsymbol{\beta}_0\right)^2$ and $C_2\equiv\sup_{d,\mathbf{x}}\varpi^{ps}_0\left(d,\mathbf{x};\boldsymbol{\beta}_0\right)^2$, and note that, under Assumption 2(ii), $1\leq C_1,C_2<\infty$. Then

$$\begin{aligned}\mathbb{E}[g^{ate2}]&=\mathbb{E}[\varpi^{ps2}_1(Y-\mathbb{E}[Y(1)])^2+\varpi^{ps2}_0(Y-\mathbb{E}[Y(0)])^2]\\&\leq C_1\mathbb{E}[(Y\left(1\right)-\mathbb{E}[Y(1)])^2]+C_2\mathbb{E}[(Y\left(0\right)-\mathbb{E}[Y(0)])^2]\\&<\infty, \qquad (C.15)\end{aligned}$$

where the first equality follows from $\varpi^{ps}_1\cdot\varpi^{ps}_0=0$ $a.s.$, 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 3.1, and Assumption 4(ii),

$$\begin{aligned}\mathbb{E}[(l'_{w,\Psi}\mathbf{G}^{ate}_{\boldsymbol{\beta}})^2]&\leq||\mathbf{G}^{ate}_{\boldsymbol{\beta}}||^2\cdot\mathbb{E}[||l_{w,\Psi}||^2]\\&<\infty, \qquad (C.16)\end{aligned}$$

whereas, by Cauchy-Schwarz inequality, (C.15) and (C.16),

$$\begin{aligned}\mathbb{E}[|g^{ate}l'_{w,\Psi}\mathbf{G}^{ate}_{\boldsymbol{\beta}}|]&\leq\mathbb{E}[|g^{ate}|^2]^{1/2}\cdot\mathbb{E}[|l'_{w,\Psi}\mathbf{G}^{ate}_{\boldsymbol{\beta}}|^2]^{1/2}\\&<\infty. \qquad (C.17)\end{aligned}$$

Hence, $\mathbb{E}[\psi_{w,\Psi}^{ate}(Y,D,\mathbf{X})^2]<\infty$ follows from (C.14)-(C.17), which concludes the proof of (C.9).

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

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

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

are Donsker, where

$$\psi_{1,w,\Psi}^{dte}\left(z,d,\mathbf{x};y\right) = g_1^{dte}\left(z,d,\mathbf{x};y\right) - l_{w,\Psi}\left(d,\mathbf{x};\boldsymbol{\beta}_0\right)'\cdot\mathbf{G}_{1,\boldsymbol{\beta}}^{dte}\left(y\right),$$

$$\psi_{0,w,\Psi}^{dte}\left(z,d,\mathbf{x};y\right) = g_0^{dte}\left(z,d,\mathbf{x};y\right) + l_{w,\Psi}\left(d,\mathbf{x};\boldsymbol{\beta}_0\right)'\cdot\mathbf{G}_{0,\boldsymbol{\beta}}^{dte}\left(y\right),$$

and

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

$$\mathbf{G}_{0,\boldsymbol{\beta}}^{dte}\left(y\right) = \mathbb{E}\left[\frac{g_0^{dte}\left(Y,D,\mathbf{X};y\right)}{1-p\left(\mathbf{X};\boldsymbol{\beta}_0\right)}\cdot\dot{p}\left(\mathbf{X};\boldsymbol{\beta}_0\right)\right].$$

Toward this end, note that the classes of functions $\left\{l_{w,\Psi}\left(d,\mathbf{x};\boldsymbol{\beta}_0\right)'\cdot\mathbf{G}_{1,\boldsymbol{\beta}}^{dte}\left(y\right) : y\in[-\infty,\infty]\right\}$ and $\left\{l_{w,\Psi}\left(d,\mathbf{x};\boldsymbol{\beta}_0\right)'\cdot\mathbf{G}_{0,\boldsymbol{\beta}}^{dte}\left(y\right) : y\in[-\infty,\infty]\right\}$are Donsker since they depend on y in a deterministic manner, $\mathbf{G}_{d,\boldsymbol{\beta}}^{dte}\left(y\right)<\infty$, $d\in\{0,1\}$, and, by Theorem 3.1, $\mathbb{E}[||l_{w,\Psi}||^2]<\infty$. The Donsker property of $\left\{g_1^{dte}\left(z,d,\mathbf{x};y\right) : y\in[-\infty,\infty]\right\}$ and $\left\{g_0^{dte}\left(z,d,\mathbf{x};y\right) : y\in[-\infty,\infty]\right\}$ 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 $\boldsymbol{\lambda}_{w,\Psi}^{dte}\left(z,d,\mathbf{x};\cdot\right) = \left(\psi_{1,w,\Psi}^{dte}\left(z,d,\mathbf{x};\cdot\right),\psi_{0,w,\Psi}^{dte}\left(z,d,\mathbf{x};\cdot\right)\right)'$, and denote

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

Thus, under Assumptions 1-4,

$$\mathbb{G}_{n,w,\Psi}^{dte,(1,0)}\left(\cdot\right) \Rightarrow \mathbb{G}_{\infty,w,\Psi}^{dte,(1,0)}\left(\cdot\right) \text{ in } \ell^{\infty}\left([-\infty,\infty]\right)\times\ell^{\infty}\left([-\infty,\infty]\right), \tag{C.18}$$

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

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

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 $\widehat{\mathbf{q}}^{ips}_n(\boldsymbol{\tau}) = \left(\widehat{q}^{ips}_{n,Y(1)}(\tau_1), \widehat{q}^{ips}_{n,Y(0)}(\tau_2)\right)'$, $\mathbf{q}(\boldsymbol{\tau}) = \left(q_{Y(1)}(\tau_1), q_{Y(0)}(\tau_2)\right)'$, $\mathbf{f}^{-1}(\tau) = \left(f^{-1}_{Y(1)}\left(q_{Y(1)}(\tau_1)\right), f^{-1}_{Y(0)}\right.$ and $\boldsymbol{\tau} = (\tau_1,\tau_2) \in [a_1,a_2]^2$, where a_1 and a_2 satisfy $0 < a_1 < a_2 < 1$. Under Assumptions 1-4, we have that, from (C.18), Lemma 21.4 in van der Vaart (1998), and the functional delta method, see e.g. Theorem 20.8 in van der Vaart (1998),

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

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

Then, by the CMT,

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

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

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

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

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

D Proofs of the Results under Endogeneity

In this section, we will introduce additional notations for and present the proof of Theorem 4.1 and 4.2. First of all, it is straightforward to show that the LIPS objective function $Q^{lte}_{n,w}(\boldsymbol{\beta}) = \int_\Pi \left\|\mathbf{H}^{lte}_{n,w}(\boldsymbol{\beta},\mathbf{u})\right\|^2 \Psi_n(d\mathbf{u})$, has closed-form representation analogous to the IPS case. All the proofs in Appendix A follow through by replacing $\mathbf{h}_n(D,\mathbf{X};\boldsymbol{\beta})$ with $\mathbf{h}^{lte}_n(D,Z,\mathbf{X};\boldsymbol{\beta})$.

D.I Lemmas for Section 4

Before proving Theorem 4.1 and 4.2, we point out that results similar to Lemma 2.1 continue to hold and so do the auxiliary lemmas analogous to those in Appendix B.

Lemma D.1 *(Lemma 2.1) Let $\Theta \subset \mathbb{R}^k$ be the parameter space, and assume that (4.3) is satisfied for a unique $\boldsymbol{\beta}_0^{lte} \in \Theta$. Then $Q_w^{lte}(\boldsymbol{\beta}) \geq 0$, $\forall \boldsymbol{\beta} \in \Theta$, and $Q_w^{lte}(\boldsymbol{\beta}_0^{lte}) = 0$ if and only if the the covariate balancing condition (4.3) holds.*

Proof of Lemma D.1: Note that $Q_w^{lte}(\boldsymbol{\beta}) \geq 0$ follows trivially from the definition. Next, analogous to the discussion in Section 2.2, the covariate balancing condition among compliers (4.3) is equivalent to (4.4), implying that $Q_w^{lte}(\boldsymbol{\beta}_0^{lte}) = 0$.

To complete the proof we then need to show that if $Q_w^{lte}(\boldsymbol{\beta}) = 0$, then $\boldsymbol{\beta} = \boldsymbol{\beta}_0^{lte}$. Towards this end, recall that if $Q_w^{lte}(\boldsymbol{\beta}) = 0$, it must be that $\mathbf{H}_w^{lte}(\boldsymbol{\beta}, \boldsymbol{u}) = 0$ *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, $\mathbf{H}_w^{lte}(\boldsymbol{\beta}, \boldsymbol{u}) = 0$ *a.e.* on Π if and only if $\mathbb{E}\left[\mathbf{h}^{lte}(D, Z, \mathbf{X}; \boldsymbol{\beta}) | \mathbf{X}\right] = 0$ *a.s.*, which is equivalent to

$$\frac{1}{\kappa_1(\boldsymbol{\beta})}\mathbb{E}\left[\frac{DZ}{q(\mathbf{X};\boldsymbol{\beta})} - \frac{D(1-Z)}{1-q(\mathbf{X};\boldsymbol{\beta})}\bigg|\mathbf{X}\right] = \frac{1}{\kappa(\boldsymbol{\beta})}\mathbb{E}\left[1 - \frac{(1-D)Z}{q(\mathbf{X};\boldsymbol{\beta})} - \frac{D(1-Z)}{1-q(\mathbf{X};\boldsymbol{\beta})}\bigg|\mathbf{X}\right] \quad a.s.,$$

and

$$\frac{1}{\kappa_0(\boldsymbol{\beta})}\mathbb{E}\left[\frac{(1-D)Z}{q(\mathbf{X};\boldsymbol{\beta})} - \frac{(1-D)(1-Z)}{1-q(\mathbf{X};\boldsymbol{\beta})}\bigg|\mathbf{X}\right] = \frac{1}{\kappa(\boldsymbol{\beta})}\mathbb{E}\left[1 - \frac{(1-D)Z}{q(\mathbf{X};\boldsymbol{\beta})} - \frac{D(1-Z)}{1-q(\mathbf{X};\boldsymbol{\beta})}\bigg|\mathbf{X}\right] \quad a.s..$$

Let $p_{dz}(\mathbf{X}) = \mathbb{P}(D = d|\mathbf{X}, Z = z)$ for $d, z \in \{0,1\}$. By straightforward calculation, the two equations given above are further equivalent to

$$\begin{aligned}\frac{1}{\kappa_1(\boldsymbol{\beta})}&\left[p_{11}(\mathbf{X})\frac{q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)}{q(\mathbf{X};\boldsymbol{\beta})} - p_{10}(\mathbf{X})\frac{1-q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)}{1-q(\mathbf{X};\boldsymbol{\beta})}\right] \\ &= \frac{1}{\kappa(\boldsymbol{\beta})}\left[1 - p_{01}(\mathbf{X})\frac{q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)}{q(\mathbf{X};\boldsymbol{\beta})} - p_{10}(\mathbf{X})\frac{1-q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)}{1-q(\mathbf{X};\boldsymbol{\beta})}\right] \quad a.s., \qquad \text{(D.1)}\end{aligned}$$

and

$$\begin{aligned}\frac{1}{\kappa_0(\boldsymbol{\beta})}&\left[p_{01}(\mathbf{X})\frac{q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)}{q(\mathbf{X};\boldsymbol{\beta})} - p_{00}(\mathbf{X})\frac{1-q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)}{1-q(\mathbf{X};\boldsymbol{\beta})}\right] \\ &= \frac{1}{\kappa(\boldsymbol{\beta})}\left[1 - p_{01}(\mathbf{X})\frac{q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)}{q(\mathbf{X};\boldsymbol{\beta})} - p_{10}(\mathbf{X})\frac{1-q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)}{1-q(\mathbf{X};\boldsymbol{\beta})}\right] \quad a.s.. \qquad \text{(D.2)}\end{aligned}$$

Denote $c_1(\mathbf{X}) = q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)/q(\mathbf{X};\boldsymbol{\beta})$ and $c_0(\mathbf{X}) = \left(1 - q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\right)/(1 - q(\mathbf{X};\boldsymbol{\beta}))$. Note that

$p_{1z}(\mathbf{X}) + p_{0z}(\mathbf{X}) = 1$ $a.s.$ for $z \in \{0,1\}$. Then dividing (D.1) by (D.2) and rearranging yield

$$\begin{aligned}\frac{\kappa_0(\boldsymbol{\beta})}{\kappa_1(\boldsymbol{\beta})} =& \frac{(1-p_{11}(\mathbf{X}))\, c_1(\mathbf{X}) - (1-p_{10}(\mathbf{X}))\, c_0(\mathbf{X})}{p_{11}(\mathbf{X})\, c_1(\mathbf{X}) - p_{10}(\mathbf{X})\, c_0(\mathbf{X})} \\ =& \frac{c_1(\mathbf{X}) - c_0(\mathbf{X})}{p_{11}(\mathbf{X})\, c_1(\mathbf{X}) - p_{10}(\mathbf{X})\, c_0(\mathbf{X})} - 1 \;\; a.s.,\end{aligned}$$

which implies $c_1(\mathbf{X}) = c_0(\mathbf{X})$ $a.s.$ by noting that $\kappa_0(\boldsymbol{\beta})/\kappa_1(\boldsymbol{\beta})$ is a constant. Otherwise, it will lead to a contradiction since the right-hand side of the above equation is a non-degenerate measurable function of $\mathbf{X}$. With the help of $c_1(\mathbf{X}) = c_0(\mathbf{X})$ $a.s.$, (D.1) is equivalent to

$$\begin{aligned}\frac{\kappa(\boldsymbol{\beta})}{\kappa_1(\boldsymbol{\beta})} =& \frac{1-(1-p_{11}(\mathbf{X}))\, c_1(\mathbf{X}) - p_{10}(\mathbf{X})\, c_1(\mathbf{X})}{(p_{11}(\mathbf{X}) - p_{10}(\mathbf{X}))\, c_1(\mathbf{X})} \\ =& \frac{1-c_1(\mathbf{X})}{(p_{11}(\mathbf{X}) - p_{10}(\mathbf{X}))\, c_1(\mathbf{X})} + 1 \;\; a.s.,\end{aligned}$$

which implies $c_1(\mathbf{X}) = 1$ $a.s.$ by noting that $\kappa(\boldsymbol{\beta})/\kappa_1(\boldsymbol{\beta})$ is a constant. Otherwise, it will lead to a contradiction. Combing the previous results, we have $c_1(\mathbf{X}) = c_0(\mathbf{X}) = 1$ $a.s.$. We then conclude that $\mathbf{H}_w^{lte}(\boldsymbol{\beta}, \boldsymbol{u}) = 0$ $a.e.$ on Π is equivalent to $q(\mathbf{X};\boldsymbol{\beta}) = q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)$ $a.s.$. Given that $q(\mathbf{X};\boldsymbol{\beta}) = q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)$ $a.s.$ for a unique $\boldsymbol{\beta}_0^{lte}$, we must have $\boldsymbol{\beta} = \boldsymbol{\beta}_0^{lte}$. This concludes the proof. ■

Lemma D.2 *(Lemma B.2) Under Assumption 6$(i)-(iii)$, the classes of functions*

$$\begin{aligned}&\mathcal{F}_1^{lte} \equiv \{(d,z,\mathbf{x}) \in \{0,1\}\times\{0,1\}\times\mathcal{X} \mapsto d\left(\frac{z}{q(\mathbf{x};\boldsymbol{\beta})} - \frac{(1-z)}{1-q(\mathbf{x};\boldsymbol{\beta})}\right) : \boldsymbol{\beta}\in\Theta\},\\ &\mathcal{F}_2^{lte} \equiv \{(d,z,\mathbf{x}) \in \{0,1\}\times\{0,1\}\times\mathcal{X} \mapsto \left(1 - \frac{(1-d)z}{q(\mathbf{x};\boldsymbol{\beta})} - \frac{d(1-z)}{1-q(\mathbf{x};\boldsymbol{\beta})}\right) : \boldsymbol{\beta}\in\Theta\},\\ &\mathcal{F}_3^{lte} \equiv \mathcal{F}_1^{lte}\cdot\mathcal{W},\\ &\mathcal{F}_4^{lte} \equiv \mathcal{F}_2^{lte}\cdot\mathcal{W},\end{aligned}$$

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

Proof of Lemma D.2 The Glivenko-Cantelli property of $\mathcal{F}_1^{lte}$ and $\mathcal{F}_2^{lte}$ follows from Example 19.8 in van der Vaart (1998) under Assumption 6$(i)-(iii)$. ■

The family of functions associated with $\mathbf{h}_0^{lte}$,

$$\mathcal{F}_0^{lte} \equiv \{(d,z,\mathbf{x}) \in \{0,1\}\times\{0,1\}\times\mathcal{X} \mapsto (1-d)\left(z/q(\mathbf{x};\boldsymbol{\beta}) - (1-z)/(1-q(\mathbf{x};\boldsymbol{\beta}))\right) : \boldsymbol{\beta}\in\Theta\}.$$

is also Glivenko-Cantelli as a result of Lemma D.2.

Let

$$\widehat{C}_{ind,F_{n,\mathbf{X}}}^{lte} = 2\int_{[-\infty,\infty]^k} \dot{\mathbf{H}}_{n,ind}^{lte}(\widehat{\boldsymbol{\beta}}_{n,ind}^{lips}, \mathbf{u})'\dot{\mathbf{H}}_{n,ind}^{lte}(\widetilde{\boldsymbol{\beta}}, \mathbf{u}) F_{n,\mathbf{X}}(d\mathbf{u}),$$

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

and

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

where $\phi\left(\mathbf{u}\right)$ is the standard k-variate normal density function and $\widetilde{\boldsymbol{\beta}}$ satisfies $\left\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right\| \leq \left\|\widehat{\boldsymbol{\beta}}^{lips}_{n,w}-\boldsymbol{\beta}^{lte}_0\right\|$. Furthermore, write

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

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

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

Lemma D.3 *(Lemma B.3) Let $\mathcal{W}$ be equal to either $\mathcal{W}_{ind}$, $\mathcal{W}_{proj}$ or $\mathcal{W}_{exp}$. Then, under Assumption 6,*

$$\mathcal{F}^{lte}_5 \equiv \left\{(d,z,\mathbf{x})\in\{0,1\}\times\{0,1\}\times\mathcal{X}\mapsto d\left(\frac{z}{q\left(\mathbf{x};\boldsymbol{\beta}\right)^2}+\frac{1-z}{(1-q\left(\mathbf{x};\boldsymbol{\beta}\right))^2}\right)\dot{q}\left(\mathbf{x};\boldsymbol{\beta}\right),\boldsymbol{\beta}\in\Theta^{lte}_0\right\},$$

$$\mathcal{F}^{lte}_6 \equiv \left\{(d,z,\mathbf{x})\in\{0,1\}\times\{0,1\}\times\mathcal{X}\mapsto \left(\frac{(1-d)z}{q\left(\mathbf{x};\boldsymbol{\beta}\right)^2}-\frac{d(1-z)}{(1-q\left(\mathbf{x};\boldsymbol{\beta}\right))^2}\right)\dot{q}\left(\mathbf{x};\boldsymbol{\beta}\right),\boldsymbol{\beta}\in\Theta^{lte}_0\right\},$$

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

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

are Glivenko-Cantelli classes of functions. Furthermore,

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

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

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

Proof of Lemma D.3: The follows from the same steps as in the proof of Lemma B.3 and is therefore omitted. ■

As in Lemma D.2, it follows trivially from Lemma D.3 that

$$\mathcal{F}_9^{lte} \equiv \left\{ (d, z, \mathbf{x}) \in \{0,1\} \times \{0,1\} \times \mathcal{X} \mapsto (1-d) \cdot \left(\frac{z}{q(\mathbf{x};\boldsymbol{\beta})^2} + \frac{1-z}{(1-q(\mathbf{x};\boldsymbol{\beta}))^2} \right) \dot{q}(\mathbf{x};\boldsymbol{\beta}), \boldsymbol{\beta} \in \Theta_0^{lte} \right\},$$

is Glivenko-Cantelli.

Lemma D.4 *(Lemma B.4) Let Π be a compact, convex subset of $\mathbb{R}^k$ with a non-empty interior. Then, under Assumption 6,*

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

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

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

are Donsker classes of functions.

Proof of Lemma D.4: The follows from the same steps as in the proof of Lemma B.4 and is therefore omitted. ■

Next, define

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

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

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

and let $A_{n,2,ind}^{lte}(\mathbf{x})$, $A_{n,2,proj}^{lte}(\mathbf{x})$ and $A_{n,2,exp}^{lte}(\mathbf{x})$ be the counterparts in the sample.

Lemma D.5 *(Lemma B.5) Under Assumption 6,*

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

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

$$\mathbb{E}_n\left[A_{n,2,\exp}^{lte}(\mathbf{X}) \cdot \mathbf{h}_n\left(D, Z, \mathbf{X}; \boldsymbol{\beta}_0^{lte}\right)\right] = \mathbb{E}_n\left[A_{2,\exp}^{lte}(\mathbf{X}) \cdot \mathbf{h}\left(D, Z, \mathbf{X}; \boldsymbol{\beta}_0^{lte}\right)\right] + o_p\left(n^{-1/2}\right). \quad \text{(D.5)}$$

Proof of Lemma D.5: The follows from the same steps as in the proof of Lemma B.5 and is therefore omitted. ■

D.II Proof of Main Results

Proof of Theorem 4.1: We first show consistency of $\widehat{\boldsymbol{\beta}}_{n,w}^{lips}$. From Lemma D.1 we know that $Q_w^{lte}(\boldsymbol{\beta})$ is uniquely minimized at $\boldsymbol{\beta}_0^{lte}$, and that, under Assumption 6, $\mathbf{H}_w^{lte}(\boldsymbol{\beta}, \mathbf{u})$ 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 $\varepsilon > 0$

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

Therefore, the consistency of $\widehat{\boldsymbol{\beta}}_{n,w}^{lips}$ follows immediately from the uniform convergence of $Q_{n,w}^{lte}(\boldsymbol{\beta})$ over Θ as $n \to \infty$. Lemma B.1, D.2 and CMT ensure that

$$\sup_{(\boldsymbol{\beta}, \mathbf{u}) \in \Theta \times \Pi} \left\| \mathbf{H}_{n,w}^{lte}(\boldsymbol{\beta}, \mathbf{u}) - \mathbf{H}_w^{lte}(\boldsymbol{\beta}, \mathbf{u}) \right\| \xrightarrow{p} 0,\ w \in \{ind, proj\}.$$

and for $w = exp$, the same arguments as in the proof of Theorem 4.1 can be applied to show the uniform convergence.

To derive the asymptotic linear representation of $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{n,w}^{lips} - \boldsymbol{\beta}_0^{lte}\right)$, we first apply Taylor expansion to the first order condition of $Q_{n,w}^{lte}(\boldsymbol{\beta})$, which gives

$$\begin{aligned}
&\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{n,w}^{lips} - \boldsymbol{\beta}_0^{lte}\right) \\
&= -(\widehat{C}_{w,\Psi_n}^{lte})^{-1} \cdot \sqrt{n} \int \left(\dot{\mathbf{H}}_{n,w}^{lte}(\widehat{\boldsymbol{\beta}}_{n,w}^{lips}, \mathbf{u})^c \mathbf{H}_{n,w}^{lte}(\boldsymbol{\beta}_0^{lte}, \mathbf{u}) + \dot{\mathbf{H}}_{n,w}^{lte}(\widehat{\boldsymbol{\beta}}_{n,w}^{lips}, \mathbf{u})' \left(\mathbf{H}_{n,w}^{lte}(\boldsymbol{\beta}_0^{lte}, \mathbf{u})' \right)^c \right) \Psi_n(d\mathbf{u}) \\
&\quad = -(\widehat{C}_{w,\Psi_n}^{lte})^{-1} \cdot \sqrt{n}\mathbb{E}_n \left[\int \left(\dot{\mathbf{H}}_{n,w}^{lte}(\widehat{\boldsymbol{\beta}}_{n,w}^{lips}, \mathbf{u})^c w(\mathbf{X}; \mathbf{u}) + \dot{\mathbf{H}}_{n,w}^{lte}(\widehat{\boldsymbol{\beta}}_{n,w}^{lips}, \mathbf{u})' w^c(\mathbf{X}, \mathbf{u}) \right) \Psi_n(d\mathbf{u}) \right. \\
&\qquad\qquad \left. \cdot \mathbf{h}_n\left(D, Z, \mathbf{X}; \boldsymbol{\beta}_0^{lte}\right) \right]
\end{aligned} \tag{D.6}$$

From Lemma D.3 and Lemma D.5 we have that

$$\widehat{C}_{w,\Psi_n}^{lte} = C_{w,\Psi}^{lte} + o_p(1), \tag{D.7}$$

$$\mathbb{E}_n\left[A_{n,2,ind}^{lte}(\mathbf{X}) \cdot \mathbf{h}_n\left(D, Z, \mathbf{X}; \boldsymbol{\beta}_0^{lte}\right)\right] = \mathbb{E}_n\left[A_{2,ind}^{lte}(\mathbf{X}) \cdot \mathbf{h}\left(D, Z, \mathbf{X}; \boldsymbol{\beta}_0^{lte}\right)\right] + o_p(1). \tag{D.8}$$

Consistency of $\widehat{\boldsymbol{\beta}}_{n,w}^{lips}$ follows from (D.7) and (D.8) with $l_{w,\Psi}^{lte}\left(D, Z, \mathbf{X}; \boldsymbol{\beta}_0^{lte}\right)$ given by (4.6). Asymptotic normality results from the square integrability of $l_{w,\Psi}^{lte}\left(D, Z, \mathbf{X}; \boldsymbol{\beta}_0^{lte}\right)$, which is further guaranteed by the uniform boundedness of $w(\mathbf{X}; \mathbf{u})$ and Assumption 6(ii). ■

Before presenting the proof of Theorem 4.2, we define some quantities related to the influence function of $\psi_{w,\Psi}^j$, for $w \in \{ind, exp, proj\}$. Let

$$\psi_{w,\Psi}^{late}(Y, D, Z, \mathbf{X}) = g^{late}(Y, D, Z, \mathbf{X}) - l_{w,\Psi}^{lte}\left(D, Z, \mathbf{X}; \boldsymbol{\beta}_0^{lte}\right)' \cdot \mathbf{G}_{\boldsymbol{\beta}}^{late}, \tag{D.9}$$

$$\psi_{w,\Psi}^{ldte}\left(Y,D,Z,\mathbf{X};y\right)=g^{ldte}\left(Y,D,Z,\mathbf{X};y\right)-l_{w,\Psi}^{lte}\left(D,Z,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)'\cdot\mathbf{G}_{\boldsymbol{\beta}}^{ldte}\left(y\right), \tag{D.10}$$

$$\psi_{w,\Psi}^{lqte}\left(Y,D,Z,\mathbf{X};\tau\right)=-\left(g^{lqte}\left(Y,D,Z,\mathbf{X};\tau\right)-l_{w,\Psi}^{lte}\left(D,Z,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)'\cdot\mathbf{G}_{\boldsymbol{\beta}^{lte}}^{lqte}\left(\tau\right)\right), \tag{D.11}$$

where, for $j\in\{late,ldte,lqte\}$, $g^j\left(Y,D,Z,\mathbf{X}\right)=g_1^j\left(Y,D,Z,\mathbf{X}\right)-g_0^j\left(Y,D,Z,\mathbf{X}\right)$, with

$$g_d^{late}\left(Y,D,Z,\mathbf{X}\right)=\varpi_d^{lte}\left(D,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\cdot\left(Y-\mathbb{E}\left[Y\left(d\right)|\mathcal{C}\right]\right),$$

$$g_d^{ldte}\left(Y,D,Z,\mathbf{X};y\right)=\varpi_d^{lte}\left(D,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\cdot\left(1\left\{Y\leq y\right\}-F_{Y(d)|\mathcal{C}}\left(y\right)\right),$$

$$g_d^{lqte}\left(Y,D,Z,\mathbf{X};\tau\right)=\frac{\varpi_d^{lte}\left(D,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\cdot\left(1\left\{Y\leq q_{Y(d)|\mathcal{C}}\left(\tau\right)\right\}-\tau\right)}{f_{Y(d)|\mathcal{C}}\left(q_{Y(d)|\mathcal{C}}\left(\tau\right)\right)},$$

and

$$\mathbf{G}_{\boldsymbol{\beta}}^{late}=\mathbb{E}\left[\left(\sum_{d=0,1}\frac{1\{D=d\}(Y-\mathbb{E}[Y(d)|\mathcal{C}])}{(-1)^{d+1}\kappa_d(\boldsymbol{\beta}_0^{lte})}\right)\cdot\left(\frac{}{Z}q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)^2+\frac{}{1-Z}(1-q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right))^2\right)\dot{q}\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\right],$$

$$\mathbf{G}_{\boldsymbol{\beta}}^{ldte}\left(y\right)=\mathbb{E}\left[\left(\sum_{d=0,1}\frac{1\{D=d\}(1\{Y\leq y\}-F_{Y(d)|\mathcal{C}}(y))}{(-1)^{d+1}\kappa_d(\boldsymbol{\beta}_0^{lte})}\right)\cdot\left(\frac{Z}{q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)^2}+\frac{1-Z}{(1-q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right))^2}\right)\dot{q}\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\right],$$

$$\mathbf{G}_{\boldsymbol{\beta}}^{qte}\left(\tau\right)=\mathbb{E}\left[\left(\sum_{d=0,1}\frac{1\{D=d\}(1\{Y\leq q_{Y(d)|\mathcal{C}}(\tau)\}-\tau)}{(-1)^{d+1}\kappa_d(\boldsymbol{\beta}_0^{lte})}\right)\cdot\left(\frac{Z}{q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)^2}+\frac{1-Z}{(1-q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right))^2}\right)\dot{q}\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\right].$$

The functions g^{late}, g^{ldte} and g^{lqte} are the influence functions of the LATE, LDTE and LQTE estimators, respectively, when the instrumental propensity score parameters $\boldsymbol{\beta}_0^{lte}$ are known. We denote $\Omega_{w,\Psi}^{late}=\mathbb{E}\left[\psi_{w,\Psi}^{late}\left(Y,D,Z,\mathbf{X}\right)^2\right]$, $\Omega_{w,\Psi,y}^{ldte}=\mathbb{E}\left[\psi_{w,\Psi}^{ldte}\left(Y,D,Z,\mathbf{X};y\right)^2\right]$, and $\Omega_{w,\Psi,\tau}^{lqte}=\mathbb{E}\left[\psi_{w,\Psi}^{lqte}\left(Y,D,Z,\mathbf{X};\tau\right)^2\right]$.

Part 1: Asymptotic Properties of the Local Average Treatment Effect.

As in the proof of Theorem 3.2, we show that

$$\sqrt{n}\left(\widehat{LATE}_n^{lips}-LATE\right)=\sqrt{n}\mathbb{E}_n\left[\psi_{w,\Psi}^{late}\left(Y,D,Z,\mathbf{X}\right)\right]+o_p\left(1\right) \tag{D.12}$$

where $\mathbb{E}\left[\psi_{w,\Psi}^{late}\left(Y,D,Z,\mathbf{X}\right)\right]=0$ and $\mathbb{E}\left[\psi_{w,\Psi}^{late}\left(Y,D,Z,\mathbf{X}\right)^2\right]<\infty$.

The key is to show that

$$\begin{aligned}&\mathbb{E}_n[\varpi_{n,d}^{lte}(D,Z,\mathbf{X};\widehat{\boldsymbol{\beta}}_{n,w}^{lips})Y]-\mathbb{E}\left[Y\left(d\right)|\mathcal{C}\right]\\&\quad=\mathbb{E}_n\left[\varpi_d^{lte}\left(D,Z,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\cdot\left(Y-\mathbb{E}\left[Y\left(d\right)|\mathcal{C}\right]\right)-l_{w,\Psi}^{lte}\left(D,Z,\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)'\mathbf{G}_{\boldsymbol{\beta},d}^{late}\right]+o_p\left(n^{-1/2}\right),\end{aligned} \tag{D.13}$$

where $\mathbf{G}^{late}_{\boldsymbol{\beta},d} = \mathbb{E}\left[\frac{1\{D=d\}(Y-\mathbb{E}[Y(d)|\mathcal{C}])}{\kappa_d(\boldsymbol{\beta}^{lte}_0)} \cdot \left(\frac{Z}{q\left(\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)^2} + \frac{1-Z}{(1-q\left(\mathbf{X};\boldsymbol{\beta}^{lte}_0\right))^2}\right) \dot{q}\left(\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)\right]$

Without loss of generality, we focus on $d=1$. By Taylor expanding $\mathbb{E}_n[\varpi^{lte}_{n,d}(D,Z,\mathbf{X};\widehat{\boldsymbol{\beta}}^{lips}_{n,w})Y]$ around $\boldsymbol{\beta}^{lte}_0$ and some algebraic manipulations, we have that

$$
\begin{aligned}
&\mathbb{E}_n[\varpi^{lte}_{n,1}(D,Z,\mathbf{X};\widehat{\boldsymbol{\beta}}^{lips}_{n,w})Y] \\
&= \mathbb{E}_n\left[\varpi^{lte}_{n,1}\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)Y\right] \\
&- \mathbb{E}_n\left[\left(\frac{Z}{q\left(\mathbf{X};\widetilde{\boldsymbol{\beta}}\right)^2} + \frac{1-Z}{\left(1-q\left(\mathbf{X};\widetilde{\boldsymbol{\beta}}\right)\right)^2}\right)\left(Y - \mathbb{E}_n\left[\varpi^{lte}_{n,1}\left(D,Z,\mathbf{X};\widetilde{\boldsymbol{\beta}}\right)Y\right]\right)\cdot \dot{q}\left(\mathbf{X};\widetilde{\boldsymbol{\beta}}\right)'\right]\left(\widehat{\boldsymbol{\beta}}^{lips}_{n,w} - \boldsymbol{\beta}^{lte}_0\right),
\end{aligned}
$$

where $\widetilde{\boldsymbol{\beta}}$ satisfies $\left\|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^{lte}_0\right\| \leq \left\|\widehat{\boldsymbol{\beta}}^{lips}_{n,w} - \boldsymbol{\beta}^{lte}_0\right\|$. From Theorem 4.1, we have that

$$
\sqrt{n}\left(\widehat{\boldsymbol{\beta}}^{lips}_{n,w} - \boldsymbol{\beta}_0\right) = \sqrt{n}\mathbb{E}_n\left[l^{lte}_{w,\Psi}\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)\right] + o_p(1)
$$

and therefore by CMT,

$$
\begin{aligned}
&\mathbb{E}_n[\varpi^{lte}_{n,1}(D,Z,\mathbf{X};\widehat{\boldsymbol{\beta}}^{lips}_{n,w})Y] \\
&\quad = \mathbb{E}_n\left[\varpi^{lte}_{n,1}\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)Y\right] - \mathbb{E}_n\left[l^{lte}_{w,\Psi}\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)'\cdot \mathbf{G}^{late}_{\boldsymbol{\beta},1}\right] + o_p\left(n^{-1/2}\right). \quad \text{(D.14)}
\end{aligned}
$$

To get the influence function of the first term, we use the fact that $\varpi^{lte}_{n,1}$ is normalized with mean equal to 1 and that $\varpi^{lte}_{n,1} \xrightarrow{a.s.} \varpi^{lte}_1$ by Lemma D.2,

$$
\begin{aligned}
\mathbb{E}_n\left[\varpi^{lte}_{n,1}\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)Y\right] - \mathbb{E}[Y(1)|\mathcal{C}] &= \mathbb{E}_n\left[\varpi^{lte}_{n,1}\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)(Y-\mathbb{E}[Y(1)|\mathcal{C}])\right] && \text{(D.15)} \\
&= \mathbb{E}_n\left[\varpi^{lte}_1\left(D,Z,\mathbf{X};\boldsymbol{\beta}^{lte}_0\right)(Y-\mathbb{E}[Y(1)|\mathcal{C}])\right] + o_p(n^{-1/2}) && \text{(D.16)}
\end{aligned}
$$

We can also show $\mathbb{E}[\varpi^{lte}_{n,0}Y]$ admits an asymptotic linear representation and (D.12) follows by CMT and the orthogonality between $(g^{late}_1, \mathbf{G}^{late}_{\boldsymbol{\beta},1})$ and $(g^{late}_0, \mathbf{G}^{late}_{\boldsymbol{\beta},0})$.

Lastly, we need to show $\psi^{late}_{w,\Psi}$ is square integrable. By Assumption 6 (uniform boundedness of ϖ^{lte}_d), Assumption 7 (conditional square integrability of $Y(d)$) and Theorem 4.1 (square integrability of $l^{lte}_{w,\Psi}$), a standard application of Cauchy-Schwartz inequality would lead to the desired result. ■

Part 2: Asymptotic Properties of the Local Distribution Treatment Effects.

The (uniform) asymptotic linear representation for the Local Distribution Treatment Effect parameter without rearrangement can be derived as in Part 1.

We define the rearranging operator of CDF as $\mathcal{R}_{\mathcal{F}d}: F_{Y(d)|\mathcal{C}}(\cdot) \mapsto F^r_{Y(d)|\mathcal{C}}(\cdot) \equiv \int 1\{F^{-1}_{Y(d)|\mathcal{C}}(\tau) \leq \cdot\}d\tau$, for $d=0,1$. Then, by the monotonicity of $F_{Y(d)|\mathcal{C}}(\cdot)$ and Corollary 3 of Chernozhukov et al.

(2010),

$$\sqrt{n}(\widehat{F^r}_{n,\varpi_d^{lte}\cdot Y}(\cdot) - F_{Y(d)|\mathcal{C}}(\cdot)) = \sqrt{n}(\widehat{F}_{n,\varpi_d^{lte}\cdot Y}(\cdot) - F_{Y(d)|\mathcal{C}}(\cdot)) + o_p(1), \tag{D.17}$$

uniformly over $[q_{Y(d)|\mathcal{C}}(a_1) - \epsilon, q_{Y(d)|\mathcal{C}}(a_2) + \epsilon]$, for $d = 0, 1$.

Therefore, as in the proof of Theorem 3.2 we only need to show that the classes of functions

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

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

are Donsker, where

$$\psi_{1,w,\Psi}^{ldte}(v, d, z, \mathbf{x};y) = g_1^{ldte}(v, d, z, \mathbf{x}; y) - l_{w,\Psi}^{lte}(d, z, \mathbf{x}; \boldsymbol{\beta}_0)' \cdot \mathbf{G}_{1,\boldsymbol{\beta}}^{ldte}(y),$$

$$\psi_{0,w,\Psi}^{ldte}(v, d, z, \mathbf{x};y) = g_0^{ldte}(v, d, z, \mathbf{x}; y) - l_{w,\Psi}^{lte}(d, z, \mathbf{x}; \boldsymbol{\beta}_0)' \cdot \mathbf{G}_{0,\boldsymbol{\beta}}^{ldte}(y),$$

and

$$\mathbf{G}_{\boldsymbol{\beta},d}^{ldte} = \mathbb{E}\left[\frac{1\{D=d\}(1\{Y \le y\} - F_{Y(d)|\mathcal{C}}(y))}{\kappa_d(\boldsymbol{\beta}_0^{lte})} \cdot \left(\frac{Z}{q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)^2} + \frac{1-Z}{(1-q\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right))^2}\right)\dot{q}\left(\mathbf{X};\boldsymbol{\beta}_0^{lte}\right)\right].$$

First note that $\left\{l_{w,\Psi}^{lte}\left(d, z, \mathbf{x}; \boldsymbol{\beta}_0^{lte}\right)' \cdot \mathbf{G}_{d,\boldsymbol{\beta}}^{ldte}(y) : y \in [-\infty, \infty]\right\}$ is Donsker since they are deterministic functions of y, $\mathbf{G}_{d,\boldsymbol{\beta}}^{ldte}(y) < \infty$, $d \in \{0,1\}$, and, by Theorem 4.1, the square integrability of $l_{w,\Psi}^{lte}$. The Donsker property of $\left\{g_d^{ldte}(v, d, z, \mathbf{x}; y) : y \in [-\infty, \infty]\right\}$ follows from Lemma B.1, Assumption 6, and Corollary 9.32 in Kosorok (2008). Thus, from Corollary 9.32 in Kosorok (2008), we conclude that $\mathcal{F}_{1,ldte}$ and $\mathcal{F}_{0,ldte}$ are Donsker.

Hence, under Assumptions 3, 5-7,

$$\mathbb{G}_{n,w,\Psi}^{ldte,(1,0)}(\cdot) \equiv \sqrt{n}\mathbb{E}_n\left[\left(\psi_{1,w,\Psi}^{ldte}(v, d, z, \mathbf{x};\cdot), \psi_{0,w,\Psi}^{ldte}(v, d, z, \mathbf{x};\cdot)\right)'\right] \tag{D.18}$$

$$\Rightarrow \mathbb{G}_{\infty,w,\Psi}^{ldte,(1,0)}(\cdot) \text{ in } \ell^\infty([-\infty,\infty]) \times \ell^\infty([-\infty,\infty]), \tag{D.19}$$

where $\mathbb{G}_{\infty,w,\Psi}^{ldte,(1,0)}(\cdot)$ is a tight, two-dimensional mean zero Gaussian process with covariance kernel $\Gamma(y_1, y_2) = \mathbb{E}[(\psi_{1,w,\Psi}^{ldte}(y_1), \psi_{0,w,\Psi}^{ldte}(y_1))(\psi_{1,w,\Psi}^{ldte}(y_2), \psi_{0,w,\Psi}^{ldte}(y_2))']$, Applying CMT, it follows that

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

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

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

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

Fixing y leads to asymptotic normality result of $\widehat{LDTE}^{lips}$ in Theorem 4.2.

Part 3: Asymptotic Properties of the Local Quantile Treatment Effects.

Likewise, by Corollary 3 of Chernozhukov et al. (2010) and the monotonicity of $F^{-1}_{Y(d)|\mathcal{C}}(\cdot)$, the rearranged quantile estimator $\widehat{F^r}^{-1}_{n,\varpi^{lte}_d\cdot Y}(\cdot)$ have the same first order asymptotic distribution as $\widehat{F}^{-1}_{n,\varpi^{lte}_d\cdot Y}(\cdot)$, for $d=0,1$. Therefore, we can focus on deriving the asymptotic property of the original quantile estimator.

Under Assumptions 3, 5-7, we can use Lemma 21.4 in van der Vaart (1998), and the functional delta method to show that (D.19) leads to

$$\begin{aligned}\sqrt{n}\left(\widehat{\mathbf{q}}^{lips}_n-\mathbf{q}^{lte}\right)(\cdot)&=-\mathbf{f}^{-1}_{lte}(\cdot)'\cdot\mathbb{G}^{ldte,(1,0)}_{n,w,\Psi}\left(\mathbf{q}^{lte}(\cdot)\right)+o_p(1)\\&\Rightarrow-\mathbf{f}^{-1}_{lte}(\cdot)'\cdot\mathbb{G}^{ldte,(1,0)}_{\infty,w,\Psi}\left(\mathbf{q}^{lte}(\cdot)\right)\ \text{in}\ \ell^\infty([a_1,a_2])\times\ell^\infty([a_1,a_2]),\end{aligned}$$

where $\widehat{\mathbf{q}}^{lips}_n(\boldsymbol{\tau})=\left(\widehat{q}^{lips}_{n,Y(1)|\mathcal{C}}(\tau_1),\widehat{q}^{lte}_{n,Y(0)|\mathcal{C}}(\tau_2)\right)'$, $\mathbf{q}^{lte}(\boldsymbol{\tau})=\left(q_{Y(1)|\mathcal{C}}(\tau_1),q_{Y(0)|\mathcal{C}}(\tau_2)\right)'$, $\mathbf{f}^{-1}_{lte}(\tau)=\left(f^{-1}_{Y(1)|\mathcal{C}}\left(q_{Y(1)|\mathcal{C}}(\tau_1)\right),f^{-1}_{Y(0)|\mathcal{C}}\left(q_{Y(0)|\mathcal{C}}(\tau_2)\right)\right)'$, $\boldsymbol{\tau}=(\tau_1,\tau_2)\in[a_1,a_2]^2$, and for all $0<a_1<a_2<1$.

Applying CMT again yields

$$\begin{aligned}\sqrt{n}\left(\widehat{LQTE}^{lips}_n-LQTE\right)(\cdot)&=(1,-1)\cdot\left(-\mathbf{f}^{-1}_{lte}(\cdot)'\cdot\mathbb{G}^{ldte,(1,0)}_{n,w,\Psi}\left(\mathbf{q}^{lte}(\cdot)\right)\right)+o_p(1)\\&\Rightarrow\mathbb{G}^{lqte}_{\infty,w,\Psi}(\cdot)\ \text{in}\ \ell^\infty[a_1,a_2],\end{aligned}$$

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

$$\Gamma_{lqte}(\tau_1,\tau_2)=\mathbb{E}\left[\psi^{lqte}_{w,\Psi}(Y,D,Z,\mathbf{X};\tau_1)\,\psi^{lqte}_{w,\Psi}(Y,D,Z,\mathbf{X};\tau_2)\right].$$

Now, fixing τ leads to asymptotic normality result of $\widehat{LQTE}^{lips}$ in Theorem 4.2. ∎

E Estimating Treatment Effects on the Treated

In this section we focus on treatment effect parameters for the treated subpopulation instead of the overall population. Heckman et al. (1997) argue that analyzing treatment effects on the treated instead of overall treatment effects is more interesting when the policy intervention is directed at individuals with certain characteristics. For instance, if an employment program (or a clinical treatment) is directed at individuals who face barriers to employment (or who have some specific symptoms), perhaps there is little interest in analyzing the effect of this intervention on individuals with strong labor market attachment (or on individual who does not have these symptoms). Another potential advantage of focusing on the treated subpopulation is that one can weaken the overlap condition in Assumption 2(ii) by allowing the PS to be close or even exactly equal to zero. This is particularly important in one of our applications in Section 6.

Analogous to the discussion in the previous section, here the goal is to make inference about the average, distributional and quantile treatment effect on the treated, defined as $ATT = \mathbb{E}\left[Y\left(1\right)|D=1\right] - \mathbb{E}\left[Y\left(0\right)|D=1\right]$, $DTT\left(y\right) = F_{Y(1)|D=1}\left(y\right) - F_{Y(0)|D=1}\left(y\right)$, and $QTT\left(\tau\right) = q_{Y(1)|D=1}\left(\tau\right) - q_{Y(0)|D=1}\left(\tau\right)$, respectively, where, for $d \in \{0,1\}$, $F_{Y(d)|D=1}\left(y\right) = \mathbb{E}\left[1\left\{Y\left(d\right) \leq y\right\}|D=1\right]$, $y \in \mathbb{R}$, and $q_{Y(d)|D=1}\left(\tau\right) = \inf\left\{y : F_{Y(d)|D=1}\left(y\right) \geq \tau\right\}$, $\tau \in (0,1)$.

Let $w_1^{tt,ps}\left(D, \mathbf{X}\right) = D/\mathbb{E}\left[D\right]$ and

$$w_0^{tt,ps}\left(D, \mathbf{X}\right) = \frac{\left(1-D\right)p\left(\mathbf{X}\right)}{1-p\left(\mathbf{X}\right)} \Bigg/ \mathbb{E}\left[\frac{\left(1-D\right)p\left(\mathbf{X}\right)}{1-p\left(\mathbf{X}\right)}\right].$$

Note that functionals of the distribution of $Y\left(1\right)$ for the treated subpopulation can be directly estimated from the data using the analogy principle. Thus, when analyzing treatment effects on the treated, the main challenge faced is to identify and make inference about functionals of the distribution of $Y\left(0\right)$ for the treated subpopulation. Towards this end, we make the following assumptions.

Assumption E.1 *(a) Given* $\mathbf{X}$, $Y\left(0\right)$ *is independent from* D*; and (b) for all* $\mathbf{x} \in \mathcal{X}$, $p\left(\mathbf{x}\right)$ *is uniformly bounded away from one.*

Assumption E.2 *For* $d \in \{0,1\}$, *(i)* $\mathbb{E}\left[Y\left(d\right)^2|D=1\right] < M$ *for some* $0 < M < \infty$, *(ii)*

$$\mathbb{E}\left[\sup_{\beta \in \Theta_0}\left\|\frac{w_0^{tt,ps}\left(D, \mathbf{X}; \boldsymbol{\beta}\right)\left(Y - \mathbb{E}[Y(0)|D=1]\right)}{p\left(\mathbf{X}; \boldsymbol{\beta}\right)\left(1-p\left(\mathbf{X}; \boldsymbol{\beta}\right)\right)} \cdot \dot{p}\left(\mathbf{X}; \boldsymbol{\beta}\right)\right\|\right] < \infty,$$

and (iii) for some $\varepsilon > 0$, $0 < a_1 < a_2 < 1$, $F_{Y(d)|D=1}$ *is continuously differentiable on* $\left[q_{Y(d)|D=1}\left(a_1\right) - \varepsilon, q_{Y(d)|D=1}\left(\right.\right.$ *with strictly positive derivative* $f_{Y(d)|D=1}$.

Assumption E.1 is a weaker version of Assumption 1, where we do not impose any lower bound on the PS, nor make any assumption about the relationship of $Y\left(1\right)$, D, and $\mathbf{X}$. Assumption E.2 is the analogue of Assumption 4.

As shown by Heckman et al. (1997), under Assumptions E.1 - E.2, the ATT is identified by

$$ATT = \mathbb{E}\left[\left(w_1^{tt,ps}\left(D, \mathbf{X}\right) - w_0^{tt,ps}\left(D, \mathbf{X}\right)\right)Y\right].$$

Analogously, $F_{Y(0)|D=1}\left(y\right)$ is identified by

$$F_{Y(0)|D=1}\left(y\right) = \mathbb{E}\left[w_0^{tt,ps}\left(D, \mathbf{X}\right)1\left\{Y \leq y\right\}\right],$$

implying that both $DTT\left(y\right)$ and $QTT\left(\tau\right)$ can also be written as functionals of the observed data; see e.g. Firpo (2007). Such identification results suggest that we can estimate the ATT, $DTT\left(y\right)$ and $QTT\left(\tau\right)$ by

$$\widehat{ATT}_n^{ips} = \mathbb{E}_n\left[\left(w_{n,1}^{tt,ps}\left(D, \mathbf{X}\right) - w_{n,0}^{tt,ps}\left(D, \mathbf{X}; \widehat{\boldsymbol{\beta}}_{n,w}^{ips}\right)\right)Y\right],$$

$$\widehat{DTT}_n^{ips}\left(y\right) = \mathbb{E}_n\left[\left(w_{n,1}^{tt,ps}\left(D, \mathbf{X}\right) - w_{n,0}^{tt,ps}\left(D, \mathbf{X}; \widehat{\boldsymbol{\beta}}_{n,w}^{ips}\right)\right)1\left\{Y \leq y\right\}\right],$$

$$\widehat{QTT}_n^{ips}(\tau) = \widehat{q}_{n,Y(1)|D=1}(\tau) - \widehat{q}_{n,Y(0)|D=1}^{ips}(\tau),$$

where

$$\widehat{q}_{n,Y(1)|D=1} = \arg\min_{q\in\mathbb{R}} \mathbb{E}_n\left[w_{n,1}^{tt,ps}(D,\mathbf{X})\cdot\rho_\tau(Y-q)\right],$$

$$\widehat{q}_{n,Y(0)|D=1}^{ips} = \arg\min_{q\in\mathbb{R}} \mathbb{E}_n\left[w_{n,0}^{tt,ps}\left(D,\mathbf{X};\widehat{\boldsymbol{\beta}}_{n,w}^{ips}\right)\cdot\rho_\tau(Y-q)\right],$$

$w_{n,1}^{tt,ps}(D,\mathbf{X}) = D/\,\mathbb{E}_n[D]$, and

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

The next theorem summarizes the asymptotic properties of the IPW estimators for the treatment effect on the treated based on the IPS. For $j\in\{att, dtt, qtt\}$, let $g^j(Y,D,\mathbf{X}) = g_1^j(Y,D,\mathbf{X}) - g_0^j(Y,D,\mathbf{X})$, with, for $d\in\{0,1\}$,

$$g_d^{att}(Y,D,\mathbf{X}) = w_d^{tt,ps}(D,\mathbf{X};\boldsymbol{\beta}_0)\cdot(Y-\mathbb{E}[Y(d)\,|D=1]),$$

$$g_d^{dtt}(Y,D,\mathbf{X};y) = w_d^{tt,ps}(D,\mathbf{X};\boldsymbol{\beta}_0)\cdot\left(1\{Y\le y\} - F_{Y(d)|D=1}(y)\right),$$

$$g_d^{qtt}(Y,D,\mathbf{X};\tau) = \frac{w_d^{tt,ps}(D,\mathbf{X};\boldsymbol{\beta}_0)\cdot\left(1\left\{Y\le q_{Y(d)|D=1}(\tau)\right\}-\tau\right)}{f_{Y(d)|D=1}\left(q_{Y(d)|D=1}(\tau)\right)}.$$

Finally, let $\Omega_{w,\Psi}^{att} = \mathbb{E}\left[\psi_{w,\Psi}^{att}(Y,D,\mathbf{X})^2\right]$, $\Omega_{w,\Psi,y}^{dtt} = \mathbb{E}\left[\psi_{w,\Psi}^{dtt}(Y,D,\mathbf{X};y)^2\right]$, and $\Omega_{w,\Psi,\tau}^{qtt} = \mathbb{E}\left[\psi_{w,\Psi}^{qtt}(Y,D,\mathbf{X};\tau)^2\right]$, where $\psi_{w,\Psi}^{att}$, $\psi_{w,\Psi}^{dtt}$, and $\psi_{w,\Psi}^{qtt}$ are defined analogously to (3.5)-(3.7), but with g^{att}, g^{dtt}, g^{qtt} playing the role of g^{ate}, g^{dte}, g^{qte}, respectively, and

$$\mathbf{G}_{\boldsymbol{\beta}}^{att} = \mathbb{E}\left[\frac{g_0^{att}(Y,D,\mathbf{X})}{p(\mathbf{X};\boldsymbol{\beta}_0)(1-p(\mathbf{X};\boldsymbol{\beta}_0))}\cdot\dot{p}(\mathbf{X};\boldsymbol{\beta}_0)\right],$$

$$\mathbf{G}_{\boldsymbol{\beta}}^{dtt}(y) = \mathbb{E}\left[\frac{g_0^{dtt}(Y,D,\mathbf{X};y)}{p(\mathbf{X};\boldsymbol{\beta}_0)(1-p(\mathbf{X};\boldsymbol{\beta}_0))}\cdot\dot{p}(\mathbf{X};\boldsymbol{\beta}_0)\right],$$

$$\mathbf{G}_{\boldsymbol{\beta}}^{qtt}(\tau) = \mathbb{E}\left[\frac{g_0^{qtt}(Y,D,\mathbf{X};\tau)}{p(\mathbf{X};\boldsymbol{\beta}_0)(1-p(\mathbf{X};\boldsymbol{\beta}_0))}\cdot\dot{p}(\mathbf{X};\boldsymbol{\beta}_0)\right],$$

playing the role of $\mathbf{G}_{\boldsymbol{\beta}}^{ate}$, $\mathbf{G}_{\boldsymbol{\beta}}^{dte}$, and $\mathbf{G}_{\boldsymbol{\beta}}^{qte}$, respectively.

Theorem E.1 *Under Assumptions 2, 3, E.1, and E.2, for each $y\in\mathbb{R}$, $\tau\in[\varepsilon,1-\varepsilon]$, we have that, as $n\to\infty$,*

$$\sqrt{n}\left(\widehat{ATT}_n^{ips} - ATT\right)\xrightarrow{d} N\left(0,\Omega_{w,\Psi}^{att}\right),$$

$$\sqrt{n}\left(\widehat{DTT}_n^{ips} - DTT\right)(y)\xrightarrow{d} N\left(0,\Omega_{w,\Psi,y}^{dtt}\right),$$

$$\sqrt{n}\left(\widehat{QTT}_{n}^{ips}-QTT\right)(\tau)\xrightarrow{d}N\left(0,\Omega_{w,\Psi,\tau}^{qtt}\right).$$

Remark E.1 When average, distributional and quantile treatment effect on the treated are the main parameters of interest, instead of using (2.7), one may wish to estimate $\boldsymbol{\beta}_0$ such that, for every measurable, integrable function $f(\mathbf{X})$ of the covariates,

$$\mathbb{E}\left[\left(\left(\frac{(1-D)\,p(\mathbf{X};\boldsymbol{\beta}_0)}{1-p(\mathbf{X};\boldsymbol{\beta}_0)}\Big/\mathbb{E}\left[\frac{(1-D)\,p(\mathbf{X};\boldsymbol{\beta}_0)}{1-p(\mathbf{X};\boldsymbol{\beta}_0)}\right]\right)-\frac{D}{\mathbb{E}[D]}\right)f(\mathbf{X})\right]=\mathbf{0}. \tag{E.1}$$

From the discussion in Section 2, and the fact that

$$\frac{(1-D)\,p(\mathbf{X};\boldsymbol{\beta}_0)}{1-p(\mathbf{X};\boldsymbol{\beta}_0)}-D=\frac{(1-D)}{1-p(\mathbf{X};\boldsymbol{\beta}_0)}-1,$$

and $\mathbb{E}\left[(1-D)\,p(\mathbf{X};\boldsymbol{\beta}_0)/\left(1-p(\mathbf{X};\boldsymbol{\beta}_0)\right)\right]=\mathbb{E}[D]$, we can conclude that one can use

$$H_{0,w}(\boldsymbol{\beta},\mathbf{u})=\mathbb{E}\left[\left(\left(\frac{(1-D)}{1-p(\mathbf{X};\boldsymbol{\beta})}\Big/\mathbb{E}\left[\frac{(1-D)}{1-p(\mathbf{X};\boldsymbol{\beta})}\right]\right)-1\right)w(\mathbf{X};\mathbf{u})\right]$$

to construct a minimum distance estimator for $\boldsymbol{\beta}_0$ analogous to (2.5). In order to avoid additional cumbersome notation, the results stated in Theorem E.1 do not use this alternative IPS estimator, though such a modification is straightforward.

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.

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010), "Quantile and Probability Curves Without Crossing," *Econometrica*, 78(3), 1093–1125.

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

Firpo, S. (2007), "Efficient semiparametric estimation of quantile treatment effects," *Econometrica*, 75(1), 259–276.

Heckman, J. J., Ichimura, H., and Todd, P. (1997), "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme," *The Review of Economic Studies*, 64(4), 605–654.

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.