The following document is intended for online publication only (authors' webpage).

Supplement to "Identification and Estimation of Distributional Impacts of Interventions Using Changes in Inequality Measures", Part I: Proofs, derivations and some additional technical material

Sergio Firpo (sergio.firpo@fgv.br)
Cristine Pinto (cristine.pinto@fgv.br)

December 2014

1 Introduction

In the accompanying paper to this supplement we presented several theoretical results. We did not, however, provide formal proofs for their validity in that paper. In this supplement we establish formally and in a very detailed manner the validity of those results. We also present for the sake of completeness some specific formulas of influence functions used in that paper. In particular, we present the formulas for the Coefficient of Variation, the Interquartile Range, the Theil Index and the Gini Coefficient.

2 Lemma 1

Proof of Lemma 1. A proof of this Lemma has two parts. The first concerns the identification of the marginal (and conditional on $T = 1$) CDFs of potential outcomes and is omitted because it follows trivially from Firpo (2007), Lemma 1. The second part follows by the definition of ITE parameters. They are differences in the functionals of these marginal (conditional on $T = 1$) identified distributions and therefore they can be expressed as functions of the observable data (Y, X, T). ■

3 Lemma 2

Proof of Lemma 2. The proof of this second lemma is based on a similar decomposition to that presented in the appendix of Hirano, Imbens and Ridder (2003, henceforth HIR) and therefore we use the same notation to what had been used in that paper. Thus, our estimator of the propensity score is $\widehat{p}(\cdot) = \widehat{p}_K(\cdot) \equiv \Lambda\left(H_K(\cdot)^\top \widehat{\pi}_K\right)$. In fact, we set notation as in HIR to simplify the comparison of the decomposition presented below with results presented in their appendix. We break down the difference

$\sqrt{N}\left(\widehat{F}_Y^{\widehat{\omega}}(y)-F_Y^{\omega}(y)\right)$ into six components:

$$
\begin{aligned}
& \sqrt{N}\left(\widehat{F}_Y^{\widehat{\omega}}(y)-F_Y^{\omega}(y)\right) && (1)\\
= \quad & \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\omega\left(T_i,\widehat{p}_K\left(X_i\right)\right)\mathbb{I}\left\{Y_i\leq y\right\}-\omega\left(T_i,p\left(X_i\right)\right)\mathbb{I}\left\{Y_i\leq y\right\}\right. && (2)\\
& \left.-\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\mathbb{I}\left\{Y_i\leq y\right\}\left(\widehat{p}_K\left(X_i\right)-p\left(X_i\right)\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
& +\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\mathbb{I}\left\{Y_i\leq y\right\}\left(\widehat{p}_K\left(X_i\right)-p\left(X_i\right)\right)\right.\\
& \left.-\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T_i,p\left(x\right)\right)}{\partial p\left(x\right)}\cdot\mathbb{I}\left\{Y_i\leq y\right\}\middle|x\right]\left(\widehat{p}_K\left(x\right)-p\left(x\right)\right)dF\left(x\right)\right) && (3)
\end{aligned}
$$

$$
\begin{aligned}
& +\sqrt{N}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\cdot\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(\widehat{p}_K\left(x\right)-p\left(x\right)\right)dF\left(x\right)\\
& -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{\Psi}_K\left(X_i\right)\left(\frac{T_i-p_K\left(X_i\right)}{\sqrt{p_K\left(X_i\right)\left(1-p_K\left(X_i\right)\right)}}\right) && (4)
\end{aligned}
$$

$$
\begin{aligned}
& +\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\widetilde{\Psi}_K\left(X_i\right)-\Psi_K\left(X_i\right)\right)\left(\frac{T_i-p_K\left(X_i\right)}{\sqrt{p_K\left(X_i\right)\left(1-p_K\left(X_i\right)\right)}}\right) && (5)\\
& +\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\Psi_K\left(X_i\right)\left(\frac{T_i-p_K\left(X_i\right)}{\sqrt{p_K\left(X_i\right)\left(1-p_K\left(X_i\right)\right)}}\right)-\Psi_0\left(X_i\right)\left(\frac{T_i-p\left(X_i\right)}{\sqrt{p\left(X_i\right)\left(1-p\left(X_i\right)\right)}}\right)\right) && (6)\\
& +\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\omega\left(T_i,p\left(X_i\right)\right)\cdot\mathbb{I}\left\{Y_i\leq y\right\}-F_Y^{\omega}\left(y\right)\right)+\Psi_0\left(X_i\right)\left(\frac{T_i-p\left(X_i\right)}{\sqrt{p\left(X_i\right)\left(1-p\left(X_i\right)\right)}}\right) && (7)
\end{aligned}
$$

with

$$
\widetilde{\Psi}_K\left(x\right)=\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T_i,p\left(z\right)\right)}{\partial p\left(z\right)}\cdot\mathbb{I}\left\{Y_i\leq y\right\}\middle|z\right]\Lambda^{'}\left(H_K\left(z\right)^{\top}\widetilde{\pi}_K\right)H_K\left(z\right)^{\top}dF\left(z\right)\widetilde{\Sigma}_K^{-1}\sqrt{\Lambda^{'}\left(H_K\left(x\right)^{\top}\pi_K\right)}H_K\left(x\right) \quad (8)
$$

$$
\Psi_K\left(x\right)=\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T_i,p\left(z\right)\right)}{\partial p\left(z\right)}\cdot\mathbb{I}\left\{Y_i\leq y\right\}\middle|z\right]\Lambda^{'}\left(H_K\left(z\right)^{\top}\pi_K\right)H_K\left(z\right)^{\top}dF\left(z\right)\Sigma_K^{-1}\sqrt{\Lambda^{'}\left(H_K\left(x\right)^{\top}\pi_K\right)}H_K\left(x\right) \quad (9)
$$

$$
\Psi_0\left(X_i\right)=\mathbb{E}\left[\frac{\partial\omega\left(T_i,p\left(x\right)\right)}{\partial p\left(x\right)}\cdot\mathbb{I}\left\{Y_i\leq y\right\}\middle|x\right]\sqrt{p\left(X_i\right)\left(1-p\left(X_i\right)\right)} \quad (10)
$$

and

$$
\Sigma_K=\mathbb{E}\left[\Lambda^{'}\left(H_K\left(X\right)^{\top}\pi_K\right)H_K\left(X\right)H_K\left(X\right)^{\top}\right]
$$

$$
\widetilde{\Sigma}_K=\frac{1}{N}\sum_{i=1}^{N}\Lambda^{'}\left(H_K\left(X_i\right)^{\top}\widetilde{\pi}_K\right)H_K\left(X_i\right)H_K\left(X_i\right)^{\top},
$$

where $\widetilde{\pi}_K$ lies between $\widehat{\pi}_K$ and π_K.
Now we show that each term can be bounded uniformly in y.

By Taylor expansion,

$$\omega\left(T_i, \widehat{p}_K\left(X_i\right)\right)=\omega\left(T_i, p\left(X_i\right)\right)+\frac{\partial \omega\left(T_i, p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\left(\widehat{p}_K\left(X_i\right)-p\left(X_i\right)\right)+O_p\left(\left\|\widehat{p}_K\left(X_i\right)-p\left(X_i\right)\right\|^2\right).$$

We can rewrite (2) as,

$$\begin{aligned} & \frac{1}{\sqrt{N}} \frac{\partial \omega\left(T_i, p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\left(\widehat{p}_K\left(X_i\right)-p\left(X_i\right)\right)-\frac{1}{\sqrt{N}} \sum_{i=1}^N\left(\frac{\partial \omega\left(T_i, p\left(X_i\right)\right)}{\partial p\left(X_i\right)} \mathbb{I}\left\{Y_i \leq y\right\}\left(\widehat{p}_K\left(X_i\right)-p\left(X_i\right)\right)\right) \\ = & O_p\left(\sqrt{N}\left\|\widehat{p}_K\left(X_i\right)-p\left(X_i\right)\right\|^2\right), \end{aligned}$$

where by triangle inequality

$$\left\|\widehat{p}_K\left(X_i\right)-p\left(X_i\right)\right\|^2 \leq\left\|\widehat{p}_K\left(X_i\right)-p_K\left(X_i\right)\right\|^2+\left\|p_K\left(X_i\right)-p\left(X_i\right)\right\|^2. \tag{11}$$

By the mean value theorem,

$$\widehat{p}_K(x)-p_K(x)=\Lambda^{\prime}\left(H_K(x)^{\top} \widetilde{\pi}_K\right) H_K(x)^{\top}\left(\widehat{\pi}_K-\pi_K\right).$$

Using this result, the first component of (11) is therefore

$$\left\|\widehat{p}_K\left(X_i\right)-p_K\left(X_i\right)\right\|^2 \leq C \zeta(K(N))^2\left\|\widehat{\pi}_K-\pi_K\right\|^2$$

where $\zeta(K)=\sup_x\left\|H_K(x)\right\|$. Next, using Lemma 1 in the appendix of HIR, the first term of the sum in (11) is bounded by

$$\left\|\widehat{p}_K\left(X_i\right)-p_K\left(X_i\right)\right\|^2=O_p\left(C \zeta(K(N))^2 K^{-\frac{s}{r}}\right).$$

Using again Lemma 2 in the appendix of HIR, we can bound the last term in the expansion,

$$\left\|p_K\left(X_i\right)-p\left(X_i\right)\right\|^2=O_p\left(\frac{\zeta(K(N))}{N}\right).$$

At the end,

$$O_p\left(\sqrt{N}\left\|\widehat{p}_K\left(X_i\right)-p\left(X_i\right)\right\|^2\right)=O_p\left(\frac{\zeta(K(N))}{\sqrt{N}}\right)+O_p\left(\sqrt{N} \zeta(K(N))^2 K^{-\frac{s}{r}}\right).$$

Now, we find a bound on (3). First, we rewrite this term as

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\mathbb{I}\left\{Y_i\leq y\right\}\left(\widehat{p}_K\left(X_i\right)-p_K\left(X_i\right)\right)\right. \tag{12}$$
$$\left.-\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\cdot\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(\widehat{p}_K\left(x\right)-p_K\left(x\right)\right)dF\left(x\right)\right)$$
$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\mathbb{I}\left\{Y_i\leq y\right\}\left(p_K\left(X_i\right)-p\left(X_i\right)\right)\right. \tag{13}$$
$$\left.-\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\cdot\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(p_K\left(x\right)-p\left(x\right)\right)dF\left(x\right)\right)$$

and denote (13) by V_K. Note that $E\left[V_K\right]=0$ and that

$$\begin{aligned}
\mathbb{V}ar\left[V_K\right] &= \mathbb{E}\left[\mathbb{V}ar\left[\left(\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\right)\mathbb{I}\left\{Y\leq y\right\}\middle|X\right]\left(p_K\left(X\right)-p\left(X\right)\right)^2\right] \\
&\quad+\mathbb{V}ar\left[\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|X\right]\left(p_K\left(X\right)-p\left(X\right)\right)\right] \\
&\leq \mathbb{E}\left[\mathbb{E}\left[\left(\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\right)^2\mathbb{I}\left\{Y\leq y\right\}\middle|X\right]\left(p_K\left(X\right)-p\left(X\right)\right)^2\right] \\
&\leq C\zeta\left(K\left(N\right)\right)^2K^{-\frac{s}{r}}
\end{aligned}$$

and

$$\mathbb{E}\left[\left|V_K\right|\right]\leq\sqrt{\mathbb{V}ar\left[V_K\right]}\leq C\zeta\left(K\left(N\right)\right)K^{-\frac{s}{2r}}$$

and

$$\sup_{y\in\mathcal{Y}}\left|V_K\right|=O_p\left(\zeta\left(K\left(N\right)\right)K^{-\frac{s}{2r}}\right).$$

Now consider (12), by the mean value theorem this term can be rewritten as

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\mathbb{I}\left\{Y_i\leq y\right\}\Lambda'\left(H_K\left(X_i\right)^\top\widetilde{\pi}_K\right)H_K\left(X_i\right)^\top\right.$$
$$\left.-\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\Lambda'\left(H_K\left(x\right)^\top\widetilde{\pi}_K\right)H_K\left(x\right)^\top dF\left(x\right)\right]\left(\widehat{\pi}_K-\pi_K\right).$$

By a second application of the mean value theorem, we write the first term in the expression above as

$$W_{1K(N)}+W_{2K(N)}-W_{3K(N)}$$

with

$$\begin{aligned}
W_{1K(N)} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\mathbb{I}\left\{Y_i\leq y\right\}\Lambda'\left(H_K\left(X_i\right)^\top\pi_K\right)H_K\left(X_i\right)^\top\right. \\
&\quad\left.-\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\Lambda'\left(H_K\left(x\right)^\top\pi_K\right)H_K\left(x\right)^\top dF\left(x\right)\right],
\end{aligned}$$

$$W_{2K(N)} = \frac{1}{2\sqrt{N}} \sum_{i=1}^{N} \frac{\partial \omega\left(T_i, p\left(X_i\right)\right)}{\partial p\left(X_i\right)} \mathbb{I}\left\{Y_i \leq y\right\} \Lambda^{''}\left(H_K\left(X_i\right)^{\top} \widetilde{\widetilde{\pi}}_K\right) H_K\left(X_i\right) H_K\left(X_i\right)^{\top}\left(\widetilde{\pi}_K - \pi_K\right)$$

and

$$W_{3K(N)} = \frac{1}{2\sqrt{N}} \sum_{i=1}^{N} \mathbb{E}\left[\left.\frac{\partial \omega\left(T, p\left(x\right)\right)}{\partial p\left(x\right)} \mathbb{I}\left\{Y \leq y\right\}\right| x\right] \Lambda^{''}\left(H_K\left(x\right)^{\top} \widetilde{\widetilde{\pi}}_K\right) H_K\left(x\right) H_K\left(x\right)^{\top} dF\left(x\right)\left(\widetilde{\pi}_K - \pi_K\right)$$

where $\widetilde{\widetilde{\pi}}_K$ lies between $\widetilde{\pi}_K$ and π_K. First, we calculate the variance of $W_{1K(N)}$,

$$\begin{aligned}
\mathbb{V}ar\left[W_{1K(N)}\right] &= \mathbb{E}\left[\mathbb{V}ar\left[\left.\frac{\partial \omega\left(T, p\left(X\right)\right)}{\partial p\left(X\right)} \mathbb{I}\left\{Y \leq y\right\}\right| X\right]\left(\Lambda^{'}\left(H_K\left(X\right)^{\top} \pi_K\right)\right)^2 H_K\left(X\right) H_K\left(X\right)^{\top}\right] \\
&\quad + \mathbb{V}ar\left[\mathbb{E}\left[\left.\frac{\partial \omega\left(T, p\left(X\right)\right)}{\partial p\left(X\right)} \mathbb{I}\left\{Y \leq y\right\}\right| X\right] \Lambda^{'}\left(H_K\left(X\right)^{\top} \pi_K\right) H_K\left(X\right)^{\top}\right] \\
&\leq \mathbb{E}\left[\mathbb{E}\left[\left.\left(\frac{\partial \omega\left(T, p\left(X\right)\right)}{\partial p\left(X\right)}\right)^2 \mathbb{I}\left\{Y \leq y\right\}\right| X\right]\left(\Lambda^{'}\left(H_K\left(X\right)^{\top} \pi_K\right)\right)^2 H_K\left(X\right) H_K\left(X\right)^{\top}\right] \\
&\leq C\mathbb{E}\left[H_K\left(X\right) H_K\left(X\right)^{\top}\right]
\end{aligned}$$

and hence

$$\begin{aligned}
\mathbb{E}\left[\left\|W_{1K(N)}\right\|\right] &\leq \sqrt{tr\left(\mathbb{V}ar\left[W_{1K(N)}\right]\right)} \\
&\leq C\sqrt{tr\left(\mathbb{E}\left[H_K\left(X\right) H_K\left(X\right)^{\top}\right]\right)} \\
&\leq C\zeta\left(K\left(N\right)\right). \qquad (14)
\end{aligned}$$

Now, working with $W_{2K(N)}$, we first notice that for (Y, X, T) and all y,

$$\begin{aligned}
&\left\|\frac{\partial \omega\left(T, p\left(X\right)\right)}{\partial p\left(X\right)} \mathbb{I}\left\{Y \leq y\right\} \Lambda^{''}\left(H_K\left(X\right)^{\top} \widetilde{\widetilde{\pi}}_K\right) H_K\left(X\right) H_K\left(X\right)^{\top}\left(\widetilde{\pi}_K - \pi_K\right)\right\| \\
\leq &\left|\frac{\partial \omega\left(T, p\left(X\right)\right)}{\partial p\left(X\right)} \mathbb{I}\left\{Y \leq y\right\} \Lambda^{''}\left(H_K\left(X\right)^{\top} \widetilde{\widetilde{\pi}}_K\right)\right| \left\|H_K\left(X\right) H_K\left(X\right)^{\top}\right\| \left\|\widetilde{\pi}_K - \pi_K\right\|
\end{aligned}$$

and note that,

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{\partial \omega\left(T, p\left(X\right)\right)}{\partial p\left(X\right)} \mathbb{I}\left\{Y \leq y\right\} \Lambda^{''}\left(H_K\left(X\right)^{\top} \widetilde{\widetilde{\pi}}_K\right)\right| \left\|H_K\left(X\right) H_K\left(X\right)^{\top}\right\| \left\|\widetilde{\pi}_K - \pi_K\right\|\right] \\
\leq\ & C\frac{\zeta\left(K\left(N\right)\right)^{\frac{3}{2}}}{\sqrt{N}}
\end{aligned}$$

and,

$$\mathbb{E}\left[\left\|W_{2K(N)}\right\|\right] \leq C\zeta\left(K\left(N\right)\right)^{\frac{3}{2}}.$$

By analogy we work with $W_{3K(N)}$,

$$\begin{aligned}
&\left\|\frac{\sqrt{N}}{2} \int_{\mathcal{X}} \mathbb{E}\left[\left.\frac{\partial \omega\left(T, p\left(x\right)\right)}{\partial p\left(x\right)} \mathbb{I}\left\{Y \leq y\right\}\right| x\right] \Lambda^{''}\left(H_K\left(x\right)' \widetilde{\widetilde{\pi}}_K\right) H_K\left(x\right) H_K\left(x\right)^{\top} dF\left(x\right)\left(\widetilde{\pi}_K - \pi_K\right)\right\| \\
\leq\ & C\zeta\left(K\left(N\right)\right)^{\frac{3}{2}}.
\end{aligned}$$

Using the triangle and Cauchy-Schwartz inequality,

$$\mathbb{E}\left[\left|\left(W_{1K(N)}+W_{2K(N)}-W_{3K(N)}\right)^{\top}\left(\widehat{\pi}_K-\pi_K\right)\right|\right]\leq C\frac{\zeta\left(K\left(N\right)\right)^2}{\sqrt{N}}.$$

Combining the bounds above, we have

$$\begin{aligned}
&\sup_{y\in\mathcal{Y}}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\mathbb{I}\left\{Y_i\leq y\right\}\left(\widehat{p}_K\left(X_i\right)-p\left(X_i\right)\right)\right.\right.\\
&\left.\left.-\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(\widehat{p}_K\left(x\right)-p\left(x\right)\right)dF\left(x\right)\right)\right|\\
=\;&O_p\left(\zeta\left(K\left(N\right)\right)K^{-\frac{s}{2r}}\right)+O_p\left(\frac{\zeta\left(K\left(N\right)\right)^2}{\sqrt{N}}\right).
\end{aligned}$$

Now, we work with (4). Note that

$$\begin{aligned}
&\sqrt{N}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(\widehat{p}_K\left(x\right)-p\left(x\right)\right)dF\left(x\right)\\
=\;&\sqrt{N}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(\widehat{p}_K\left(x\right)-p_K\left(x\right)\right)dF\left(x\right)\\
&+\sqrt{N}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(p_K\left(x\right)-p\left(x\right)\right)dF\left(x\right).
\end{aligned}$$

Using the mean value expansion we obtain,

$$\begin{aligned}
&\sqrt{N}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(\widehat{p}_K\left(x\right)-p_K\left(x\right)\right)dF\left(x\right)\\
=\;&\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\Lambda'\left(H_K\left(x\right)^{\top}\widetilde{\pi}_K\right)H_K\left(x\right)^{\top}dF\left(x\right)\sqrt{N}\left(\widehat{\pi}_K-\pi_K\right)
\end{aligned}$$

and define

$$\widetilde{\Xi}_K\equiv\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\Lambda'\left(H_K\left(x\right)^{\top}\widetilde{\pi}_K\right)H_K\left(x\right)^{\top}dF\left(x\right).$$

Using this definition, we can rewrite (8) as

$$\begin{aligned}
\widetilde{\Psi}_K\left(x\right)\;&=\;\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(z\right)\right)}{\partial p\left(z\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|z\right]\Lambda'\left(H_K\left(z\right)^{\top}\widetilde{\pi}_K\right)H_K\left(z\right)^{\top}dF\left(z\right)\Sigma_K^{-1}\sqrt{\Lambda'\left(H_K\left(x\right)^{\top}\pi_K\right)}H_K\left(x\right)\\
&=\;\widetilde{\Xi}_K\widetilde{\Sigma}_K^{-1}\sqrt{\Lambda'\left(H_K\left(x\right)^{\top}\pi_K\right)}H_K\left(x\right).
\end{aligned}$$

The first order condition of the pseudo-maximum likelihood approach to calculate $\widehat{\pi}_K$ is

$$\sum_{i=1}^{N}H_K\left(X_i\right)\cdot\left(T_i-\Lambda\left(H_K\left(X_i\right)^{\top}\widehat{\pi}_K\right)\right)=0$$

and by the mean value theorem,

$$\begin{aligned}\sqrt{N}\left(\widehat{\pi}_K-\pi_K\right) &= \left(\frac{1}{N}\sum_{i=1}^{N}H_K\left(X_i\right)H_K\left(X_i\right)^{\top}\Lambda'\left(H_K\left(X_i\right)^{\top}\widetilde{\pi}_K\right)\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}H_K\left(X_i\right)\cdot\left(T_i-p_K\left(X_i\right)\right)\right)\\ &= \widetilde{\Sigma}_K^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}H_K\left(X_i\right)\cdot\left(T_i-p_K\left(X_i\right)\right)\right)\end{aligned}$$

and

$$\begin{aligned}&\sqrt{N}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(\widehat{p}_K\left(x\right)-p_K\left(x\right)\right)dF\left(x\right)\\ =&\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\Lambda'\left(H_K\left(x\right)^{\top}\widetilde{\pi}_K\right)H_K\left(x\right)^{\top}dF\left(x\right)\sqrt{N}\left(\widehat{\pi}_K-\pi_K\right)\\ =&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{\Xi}_K\cdot\widetilde{\Sigma}_K^{-1}\cdot\sqrt{\Lambda'\left(H_K\left(X_i\right)^{\top}\pi_K\right)}\cdot H_K\left(X_i\right)\cdot\frac{\left(T_i-p_K\left(X_i\right)\right)}{\sqrt{\Lambda'\left(H_K\left(X_i\right)^{\top}\pi_K\right)}}\\ =&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{\Psi}_K\left(X_i\right)\cdot\frac{\left(T_i-p_K\left(X_i\right)\right)}{\sqrt{\Lambda'\left(H_K\left(X_i\right)^{\top}\pi_K\right)}}\\ =&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{\Psi}_K\left(X_i\right)\cdot\frac{\left(T_i-p_K\left(X_i\right)\right)}{\sqrt{p_K\left(X_i\right)\left(1-p_K\left(X_i\right)\right)}}\end{aligned}$$

and hence

$$\begin{aligned}&\left|\sqrt{N}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(\widehat{p}_K\left(x\right)-p\left(x\right)\right)dF\left(x\right)-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{\Psi}_K\left(X_i\right)\cdot\frac{\left(T_i-p_K\left(X_i\right)\right)}{\sqrt{p_K\left(X_i\right)\left(1-p_K\left(X_i\right)\right)}}\right|\\ =&\left|\sqrt{N}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(p_K\left(x\right)-p\left(x\right)\right)dF\left(x\right)\right|.\end{aligned}$$

Since $E\left[\frac{\partial\omega(T,p(x))}{\partial p(x)}\middle|x\right]$ is bounded away on X we have that

$$\begin{aligned}&\sup_{y\in\mathcal{Y}}\left|\sqrt{N}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|X\right]\left(p_K\left(x\right)-p\left(x\right)\right)dF\left(x\right)\right|\\ \leq\ & C\sqrt{N}\zeta\left(K\left(N\right)\right)K^{-\frac{s}{2r}}=O\left(\sqrt{N}\zeta\left(K\left(N\right)\right)K^{-\frac{s}{2r}}\right).\end{aligned}$$

Now we work with (5). Let us define

$$\Xi_K\equiv\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\Lambda'\left(H_K\left(x\right)^{\top}\pi_K\right)H_K\left(x\right)^{\top}dF\left(x\right).$$

Using this definition, we can rewrite (9) as

$$\begin{aligned}
\Psi_K(x) &= \int_{\mathcal{X}} \mathbb{E}\left[\frac{\partial \omega(T, p(z))}{\partial p(z)} \mathbb{I}\{Y \leq y\} \middle| z\right] \Lambda^{'}\left(H_K(z)^\top \pi_K\right) H_K(z)^\top dF(z) \Sigma_K^{-1} \sqrt{\Lambda'\left(H_K(x)^\top \pi_K\right)} H_K(x) \\
&= \Xi_K \Sigma_K^{-1} \sqrt{\Lambda'\left(H_K(x)^\top \pi_K\right)} H_K(x).
\end{aligned}$$

Therefore, (5) can be rewritten as:

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\widetilde{\Psi}_K(X_i) - \Psi_K(X_i)\right) \left(\frac{T_i - p_K(X_i)}{\sqrt{p_K(X_i)(1 - p_K(X_i))}}\right) = -\left(\Xi_K \Sigma_K^{-1} - \widetilde{\Xi}_K \widetilde{\Sigma}_K^{-1}\right) \cdot B_{K(N)}$$

with

$$B_{K(N)} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} H_K(X_i)(T_i - p_K(X_i)).$$

Note that

$$\left(\Xi_K \Sigma_K^{-1} - \widetilde{\Xi}_K \widetilde{\Sigma}_K^{-1}\right) \cdot B_{K(N)} = \left(\Xi_K - \widetilde{\Xi}_K\right) \widetilde{\Sigma}_K^{-1} B_{K(N)} + \Xi_K \left(\Sigma_K^{-1} - \widetilde{\Sigma}_K^{-1}\right) B_{K(N)}$$

and working with the first term we obtain

$$\left|\left(\Xi_K - \widetilde{\Xi}_K\right) \widetilde{\Sigma}_K^{-1} B_{K(N)}\right| \leq \frac{1}{\lambda_{\min}\left(\widetilde{\Sigma}_K\right)} \left\|B_{K(N)}\right\| \left\|\Xi_K - \widetilde{\Xi}_K\right\|$$

where $\lambda_{\min}(A)$ is the smallest eigenvalue of a matrix A. By the mean value theorem and using the fact that $p(x)$ is bounded away from zero,

$$\begin{aligned}
&\left\|\Xi_K - \widetilde{\Xi}_K\right\| \\
&= \left\|\int_{\mathcal{X}} \mathbb{E}\left[\frac{\partial \omega(T, p(z))}{\partial p(z)} \mathbb{I}\{Y \leq y\} \middle| z\right] \left(\Lambda^{'}\left(H_K(z)^\top \pi_K\right) - \Lambda^{'}\left(H_K(z)^\top \widetilde{\pi}_K\right)\right) H_K(z)^\top dF(z)\right\| \\
&\leq C \int_{\mathcal{X}} \left|\Lambda^{\acute{}}\left(H_K(z)^\top \widetilde{\widetilde{\pi}}_K\right)\right| \|H_K(z)\|^2 dF(z) \|\pi_K - \widetilde{\pi}_K\| \\
&\leq C \int_{\mathcal{X}} \|H_K(z)\|^2 dF(z) \|\pi_K - \widetilde{\pi}_K\| \leq C\zeta(K(N))^2 \|\pi_K - \widetilde{\pi}_K\|
\end{aligned}$$

and

$$\left|\left(\Xi_K - \widetilde{\Xi}_K\right) \widetilde{\Sigma}_K^{-1} B_{K(N)}\right| \leq C\zeta(K(N))^2 \|\pi_K - \widetilde{\pi}_K\| \left\|B_{K(N)}\right\|.$$

Now, we work with the second term

$$\begin{aligned}
\left|\Xi_K \left(\Sigma_K^{-1} - \widetilde{\Xi}_K\right) B_{K(N)}\right| &= \left|\Xi_K \Sigma_K^{-1} \left(\widetilde{\Sigma}_K - \Sigma_K\right) \widetilde{\Sigma}_K^{-1} B_{K(N)}\right| \\
&\leq \frac{1}{\lambda_{\min(\widetilde{\Sigma}_K)}} \left\|B_{K(N)}\right\| \left\|\left(\widetilde{\Sigma}_K - \Sigma_K\right) \Sigma_K^{-1} \Xi_K\right\|.
\end{aligned}$$

Define $\widehat{\Sigma}_K = \frac{1}{N}\sum_{i=1}^{N}\Lambda'\left(H_K(X_i)^\top \pi_K\right) H_K(X_i) H_K(X_i)^\top$ and $W_K = \Sigma_K^{-1}\Xi_K$,

$$\begin{aligned}
& \left\|\left(\widetilde{\Sigma}_K - \Sigma_K\right) W_K\right\| \\
\leq\ & \left\|\left(\widetilde{\Sigma}_K - \widehat{\Sigma}_K\right) W_K\right\| + \left\|\left(\widehat{\Sigma}_K - \Sigma_K\right) W_K\right\| \\
\leq\ & \frac{1}{N}\sum_{i=1}^{N}\left\|H_K(X_i)^\top(\widetilde{\pi}_K - \pi_K)\Lambda''\left(H_K(X_i)^\top \widetilde{\widetilde{\pi}}_K\right) H_K(X_i) H_K(X_i)^\top W_K\right\| + \\
& \frac{1}{\sqrt{N}}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\Lambda'\left(H_K(X_i)^\top \pi_K\right) H_K(X_i) H_K(X_i)^\top - \mathbb{E}\left[\Lambda'\left(H_K(X)^\top \pi_K\right) H_K(X) H_K(X)^\top\right]\right) W_K\right\| \\
\leq\ & C\zeta(K(N))^3 \|\pi_K - \widetilde{\pi}_K\| \|W_K\| + \\
& \frac{1}{\sqrt{N}}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\Lambda'\left(H_K(X_i)^\top \pi_K\right) H_K(X_i) H_K(X_i)^\top - \mathbb{E}\left[\Lambda'\left(H_K(X)^\top \pi_K\right) H_K(X) H_K(X)^\top\right]\right) W_K\right\|.
\end{aligned}$$

Note that

$$\begin{aligned}
\|W_K\| &\leq C\|\Xi_K\| \\
&\leq C\zeta(K(N))
\end{aligned}$$

and that

$$\begin{aligned}
& \left\|\mathbb{V}ar\left[\Lambda'\left(H_K(X)^\top \pi_K\right) H_K(X) H_K(X)^\top W_K\right]\right\| \\
=\ & W_K^\top \mathbb{V}ar\left[\Lambda'\left(H_K(X)^\top \pi_K\right) H_K(X) H_K(X)^\top\right] W_K \\
\leq\ & W_K^\top \mathbb{E}\left[\left(\Lambda'\left(H_K(X)^\top \pi_K\right) H_K(X) H_K(X)^\top\right)^2\right] W_K \\
\leq\ & C W_K^\top \mathbb{E}\left[\left(H_K(X) H_K(X)^\top\right)^2\right] W_K \\
\leq\ & C\zeta(K(N))^4.
\end{aligned}$$

Now, we look at $B_{K(N)}$

$$B_{K(N)} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} H_K(X_i)(T_i - p(X_i)) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N} H_K(X_i)(p(X_i) - p_K(X_i)).$$

Since $p(x)$ is bounded away from zero and one,

$$\begin{aligned}
\mathbb{V}ar\left[B_{K(N)}\right] &\leq \mathbb{E}\left[p(X)(1-p(X)) H_K(X) H_K(X)^\top\right] + \mathbb{E}\left[(p(X) - p_K(X))^2 H_K(X) H_K(X)^\top\right] \\
&\leq C_1\zeta(K(N))^2 + C_2\zeta(K(N))^4 K^{-\frac{s}{r}}
\end{aligned}$$

and

$$\mathbb{E}\left[\left\|B_{K(N)}\right\|\right] \leq C\zeta(K(N))^2.$$

By Cauchy-Schwartz inequality we have

$$\mathbb{E}\left[\left|\left(\Xi_K - \widetilde{\Xi}_K\right)\widetilde{\Sigma}_K^{-1} B_{K(N)}\right|\right] \leq \frac{C\zeta(K(N))^{\frac{7}{2}}}{\sqrt{N}}$$

and

$$\mathbb{E}\left[\left|\Xi_K\left(\Sigma_K^{-1}-\widetilde{\Xi}_K\right)B_{K(N)}\right|\right]\leq C_1\frac{\zeta\left(K\left(N\right)\right)^{\frac{11}{2}}}{\sqrt{N}}+C_2\frac{\zeta\left(K\left(N\right)\right)^3}{\sqrt{N}}.$$

At the end,

$$\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\widetilde{\Psi}_K\left(X_i\right)-\Psi_K\left(X_i\right)\right)\left(\frac{T_i-p_K\left(X_i\right)}{\sqrt{p_K\left(X_i\right)\left(1-p_K\left(X_i\right)\right)}}\right)\right|=O_p\left(\frac{\zeta\left(K\left(N\right)\right)^{\frac{11}{2}}}{\sqrt{N}}\right).$$

Now we work with (6). Since (6) is a sum of iid random variables with mean different than zero, we bound this term by deriving the order of its second moment,

$$\mathbb{E}\left[\left(\Psi_K\left(X\right)\left(\frac{T-p_K\left(X\right)}{\sqrt{p_K\left(X\right)\left(1-p_K\left(X\right)\right)}}\right)-\Psi_0\left(X\right)\left(\frac{T-p\left(X\right)}{\sqrt{p\left(X\right)\left(1-p\left(X\right)\right)}}\right)\right)^2\right]$$

$$\begin{aligned}
= \; & \mathbb{E}\left[\left(\Psi_K\left(X\right)\left(\frac{T-p_K\left(X\right)}{\sqrt{p_K\left(X\right)\left(1-p_K\left(X\right)\right)}}\right)-\Psi_K\left(X\right)\left(\frac{T-p\left(X\right)}{\sqrt{p\left(X\right)\left(1-p\left(X\right)\right)}}\right)\right)^2\right] \\
& +\mathbb{E}\left[\left(\Psi_K\left(X\right)\left(\frac{T-p\left(X\right)}{\sqrt{p\left(X\right)\left(1-p\left(X\right)\right)}}\right)-\Psi_0\left(X\right)\left(\frac{T-p\left(X\right)}{\sqrt{p\left(X\right)\left(1-p\left(X\right)\right)}}\right)\right)^2\right] \\
& +2\cdot\mathbb{E}\left[\left(\Psi_K\left(X\right)\left(\frac{T_i-p_K\left(X\right)}{\sqrt{p_K\left(X\right)\left(1-p_K\left(X\right)\right)}}\right)-\Psi_K\left(X\right)\left(\frac{T-p\left(X\right)}{\sqrt{p\left(X\right)\left(1-p\left(X\right)\right)}}\right)\right)\cdot\right. \\
& \left.\left(\Psi_K\left(X\right)\left(\frac{T-p\left(X\right)}{\sqrt{p\left(X\right)\left(1-p\left(X\right)\right)}}\right)-\Psi_0\left(X\right)\left(\frac{T-p\left(X\right)}{\sqrt{p\left(X\right)\left(1-p\left(X\right)\right)}}\right)\right)\right]
\end{aligned}$$

$$\begin{aligned}
= \; & \mathbb{E}\left[\Psi_K\left(X\right)^2\left(\frac{T-p_K\left(X\right)}{\sqrt{p_K\left(X\right)\left(1-p_K\left(X\right)\right)}}-\frac{T-p\left(X\right)}{\sqrt{p\left(X\right)\left(1-p\left(X\right)\right)}}\right)^2\right] \\
& +\mathbb{E}\left[\left(\Psi_K\left(X\right)-\Psi_0\left(X\right)\right)^2\left(\frac{T-p\left(X\right)}{\sqrt{p\left(X\right)\left(1-p\left(X\right)\right)}}\right)^2\right] \\
& +2\mathbb{E}\left[\Psi_K\left(X\right)\left(\Psi_K\left(X\right)-\Psi_0\left(X\right)\right)\left(\frac{T-p_K\left(X\right)}{\sqrt{p_K\left(X\right)\left(1-p_K\left(X\right)\right)}}-\frac{T-p\left(X\right)}{\sqrt{p\left(X\right)\left(1-p\left(X\right)\right)}}\right)\left(\frac{T-p\left(X\right)}{\sqrt{p\left(X\right)\left(1-p\left(X\right)\right)}}\right)\right].
\end{aligned}$$

We can approximate $\Psi_K\left(x\right)$ as the least squares projection of $\Psi_0\left(x\right)$ on $H_K\left(x\right)\sqrt{p_K\left(x\right)\left(1-p_K\left(x\right)\right)}$. If we assume that $\Psi_0\left(x\right)$ is t times continuously differentiable, and using Lemma 1 of HIR we have,

$$\sup_{x\in\mathcal{X}}\left|\Psi_0\left(x\right)-\Psi_K\left(x\right)\right|<CK^{-\frac{t}{r}}$$

and hence

$$\begin{aligned}
\mathbb{E}\left[\left(\Psi_K\left(X\right)-\Psi_0\left(X\right)\right)^2\left(\frac{T-p\left(X\right)}{\sqrt{p\left(X\right)\left(1-p\left(X\right)\right)}}\right)^2\right] &= \mathbb{E}\left[\left(\Psi_K\left(X\right)-\Psi_0\left(X\right)\right)^2\right] \\
&\leq CK^{-\frac{2t}{r}}.
\end{aligned}$$

Now, consider the first term in the expansion,

$$
\begin{aligned}
& \mathbb{E}\left[\Psi_K(X)^2\left(\frac{T-p_K(X)}{\sqrt{p_K(X)(1-p_K(X))}}-\frac{T-p(X)}{\sqrt{p(X)(1-p(X))}}\right)^2\right] \\
= \; & \mathbb{E}\left[\Psi_K(X)^2\frac{(p_K(X)-p(X))^2}{p_K(X)(1-p_K(X))}\right]+\mathbb{E}\left[\Psi_K(X)^2\left(\frac{\sqrt{p(X)(1-p(X))}}{\sqrt{p_K(X)(1-p_K(X))}}-1\right)^2\right].
\end{aligned}
$$

Using the approximation of $\Psi_K(x)$,

$$
\Psi_K(x)^2\leq\Psi_0(x)^2+|\Psi_0(x)|\,CK^{-\frac{t}{r}}
$$

and therefore

$$
\begin{aligned}
& \mathbb{E}\left[\Psi_K(X)^2\frac{(p_K(X)-p(X))^2}{p_K(X)(1-p_K(X))}\right] \\
\leq \; & \mathbb{E}\left[\Psi_0(X)^2\frac{(p_K(X)-p(X))^2}{p_K(X)(1-p_K(X))}\right]+CK^{-\frac{s}{r}}\mathbb{E}\left[|\Psi_0(X)|\frac{(p_K(X)-p(X))^2}{p_K(X)(1-p_K(X))}\right] \\
\leq \; & \mathbb{E}\left[\mathbb{E}\left[\frac{\partial\omega(T,p(X))}{\partial p(X)}\cdot\mathbb{I}\{Y\leq y\}\middle|X\right]^2\frac{p(X)(1-p(X))}{p_K(X)(1-p_K(X))}(p_K(X)-p(X))^2\right] \\
& +CK^{-\frac{s}{r}}\mathbb{E}\left[\left|\mathbb{E}\left[\frac{\partial\omega(T,p(X))}{\partial p(X)}\cdot\mathbb{I}\{Y\leq y\}\middle|X\right]\right|\frac{\sqrt{p(X)(1-p(X))}(p_K(X)-p(X))^2}{p_K(X)(1-p_K(X))}\right].
\end{aligned}
$$

Since $p(x)$ is bounded from 0 and 1 on X and using Lemma 1 in the appendix of HIR,

$$
\begin{aligned}
& \mathbb{E}\left[\left|\mathbb{E}\left[\frac{\partial w(T_i,p(x))}{\partial p(x)}\cdot\mathbb{I}\{Y_i\leq y\}\middle|X\right]\right|\frac{\sqrt{p(X_i)(1-p(X_i))}(p_K(X_i)-p(X_i))^2}{p_K(X_i)(1-p_K(X_i))}\right] \\
\leq \; & C\mathbb{E}\left[\frac{(p_K(X_i)-p(X_i))^2}{p_K(X_i)(1-p_K(X_i))}\right]\leq C\zeta(K(N))^2K^{-\frac{s}{r}}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}\left[\Psi_K(X)^2\frac{(p_K(X)-p(X))^2}{p_K(X)(1-p_K(X))}\right] & \leq C_1\zeta(K(N))^2K^{-\frac{s}{r}}+C_2\zeta(K(N))^2K^{-\frac{s}{r}-\frac{t}{r}} \\
& \leq C\zeta(K(N))^2K^{-\frac{s}{r}}.
\end{aligned}
$$

By analogy,

$$
\mathbb{E}\left[\Psi_K(X)^2\left(\frac{\sqrt{p(X)(1-p(X))}}{\sqrt{p_K(X)(1-p_K(X))}}-1\right)^2\right]\leq C\zeta(K(N))^2K^{-\frac{s}{r}}
$$

and since $\Psi_K(x)$ is bounded and $p_K(x)$ and $p(x)$ are bounded away from 0 and 1,

$$
\begin{aligned}
& \mathbb{E}\left[\Psi_K(X)(\Psi_K(X)-\Psi_0(X))\left(\frac{T-p_K(X)}{\sqrt{p_K(X)(1-p_K(X))}}-\frac{T-p(X)}{\sqrt{p(X)(1-p(X))}}\right)\left(\frac{T-p(X)}{\sqrt{p(X)(1-p(X))}}\right)\right] \\
\leq \; & C\cdot\mathbb{E}[|\Psi_K(X)-\Psi_0(X)|] \\
\leq \; & CK^{-\frac{t}{r}}.
\end{aligned}
$$

Finally,

$$\begin{aligned}
&\sup_{y\in\mathcal{Y}}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\Psi_K\left(X_i\right)\left(\frac{T_i-p_K\left(X_i\right)}{\sqrt{p_K\left(X_i\right)\left(1-p_K\left(X_i\right)\right)}}\right)-\Psi_0\left(X_i\right)\left(\frac{T_i-p\left(X_i\right)}{\sqrt{p\left(X_i\right)\left(1-p\left(X_i\right)\right)}}\right)\right)\right| \\
= \;& O_p\left(\max\left(K^{-\frac{t}{2r}},\zeta\left(K\left(N\right)\right)K^{-\frac{s}{2r}}\right)\right).
\end{aligned}$$

Combining the bounds, we have

$$\begin{aligned}
&\sup_{y\in\mathcal{Y}}\left|\sqrt{N}\left(\widehat{F}_Y^{\widehat{\omega}}\left(y\right)-F_Y^{\omega}\left(y\right)\right)-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\left(\omega\left(T_i,p\left(X_i\right)\right)\cdot\mathbb{I}\left\{Y_i\leq y\right\}-F_Y^{\omega}\left(y\right)\right)-\Psi_0\left(X_i\right)\left(\frac{T_i-p\left(X_i\right)}{\sqrt{p\left(X_i\right)\left(1-p\left(X_i\right)\right)}}\right)\right\}\right| \\
= \;& O_p\left(\frac{\zeta\left(K\left(N\right)\right)^3}{\sqrt{N}}\right)+O_p\left(\sqrt{N}\zeta\left(K\left(N\right)\right)^2K^{-\frac{s}{r}}\right) \\
&+O_p\left(\zeta\left(K\left(N\right)\right)K^{-\frac{s}{2r}}\right)+O_p\left(\frac{\zeta\left(K\left(N\right)\right)^2}{\sqrt{N}}\right) \\
&+O\left(\sqrt{N}\zeta\left(K\left(N\right)\right)K^{-\frac{s}{2r}}\right)+O_p\left(\frac{\zeta\left(K\left(N\right)\right)^{\frac{11}{2}}}{\sqrt{N}}\right)+O_p\left(\max\left(K^{-\frac{t}{2r}},\zeta\left(K\left(N\right)\right)K^{-\frac{s}{2r}}\right)\right) \\
= \;& O_p\left(\sqrt{N}\zeta\left(K\left(N\right)\right)^2K^{-\frac{s}{r}}\right)+O_p\left(\zeta\left(K\left(N\right)\right)^{\frac{5}{2}}K^{-\frac{s}{2r}}\right)+O_p\left(\frac{\zeta\left(K\left(N\right)\right)^{\frac{11}{2}}}{\sqrt{N}}\right).
\end{aligned}$$

And under the assumptions on $\zeta\left(K\left(N\right)\right)$ and s, this sum is $o_p\left(1\right)$. ■

4 Proof of Theorem 1

This proof is divided into two parts. First we demonstrate that $\mathcal{H}$ is P-Donsker, where $\mathcal{H}$ is the collection of measurable functions from $\left(\mathcal{Y}\times\mathcal{X}\times\{0,1\}\right)\rightarrow\mathbb{R}$, $\mathcal{H}=\left\{\psi\left(Y,X,T,y\right)\right|y\in\mathcal{Y}\}$ that are right continuous and whose limits from the left exist everywhere in $\left(\mathcal{Y}\times\mathcal{X}\times\{0,1\}\right)$. The space $D[\mathcal{Y}\times\mathcal{X}\times\{0,1\}]$ is equipped with the uniform norm. This first part of the result be shown in an intermediate lemma. Then in the second part we apply Donsker's Theorem.

Lemma S. 1 *(Donsker) Under assumptions 2, 3 and 5, for* $j=A,B$, $\mathcal{H}_j=\left\{\psi_j\left(Y,X,T,y\right)\right|y\in\mathcal{Y}\}$ *is Donsker with a finite envelope function.*

Proof of Lemma S.1. For notational simplicity we fix and omit subscript j. The measurable collection of functions $\mathcal{W}=\left\{\mathbb{I}\left\{Y\leq y\right\}\right|y\in\mathcal{Y}\}$ is Donsker since the bracketing number $N_{[]}\left(\sqrt{\varepsilon},\mathcal{W},L_2\left(P\right)\right)\leq\frac{2}{\varepsilon}$ are of the polynomial order $\left(\frac{1}{\varepsilon}\right)^2$. The bracketing integral is finite since it converges at a slower rate than $\left(\frac{1}{\varepsilon}\right)^2$. Note that the collection of indicators functions $\mathcal{W}$ has a finite envelope function such that $\left|\mathbb{I}\left\{Y\leq y\right\}\right|\leq\Upsilon\left(y\right)<\infty$.
Now, we work the measurable functions $\mathcal{K}=\left\{F_Y\left(y\right|X\right)\left|y\in\mathcal{Y}\right\}$. Using the proof of Lemma A.2 in Donald and Hsu (2014), we can claim that $\mathcal{K}$ is Donsker. In this proof, Donald and Hsu (2014) show that $N_{[]}\left(\varepsilon,\mathcal{K},L_2\left(P\right)\right)\leq1+\left(\frac{1}{\varepsilon}\right)^2$, and the bracketing integral is finite since it converges at a slower rate than $\left(\frac{1}{\varepsilon}\right)^2$. Donald and Hsu (2014) argue that we can find a collection of $0=y_0<y_1<...<y_k$ such that $F_y\left(y_j\right|X\right)-F_y\left(y_{j-1}\right|X\right)\leq\varepsilon^2$ for all $1\leq j\leq k$. In this case, the approximating functions satisfy the inequality $\left|F_y\left(y_{j-1}\right|X\right)\right|\leq F$, and the collection $\mathcal{K}$ has a finite envelope function F.
Since $p\left(x\right)$ is bounded away from zero and one in $\mathcal{X}$, $d_1\left(T,X\right)\equiv\omega\left(T,p\left(X\right)\right)$ is a uniformly bounded measurable function, and $d_1\left(T,X\right)\cdot\mathcal{W}$ is Donske with a finite envelope function. Similarly, define $d_2\left(T,X\right)\equiv\mathbb{E}\left[\left.\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\right|X\right]\left(T-p\left(X\right)\right)$, $d_2\left(T,X\right)$ is a uniformly bounded measurable function, and $d_2\left(T,X\right)\cdot\mathcal{K}$ is

Donsker with a finite envelope function.
Hence, $\mathcal{H} = \{ d_1 (T, X) \cdot \mathcal{W} + d_2 (T, X) \cdot \mathcal{K} | \, y \in \mathcal{Y}\}$ is Donsker with a finite envelope function.We can find a finite envelope function for the class $\mathcal{H}_j$ such that $\left|\psi_j (Y, X, T, y)\right| \leq \Psi (y) < \infty$ for every y and ψ_j. ■
Proof of Theorem 1. In the second part of the proof, we let $\mathbb{P}_N$ be the empirical measure of the sample (Y, X, T). Using Lemma 1, $\mathcal{H}$ is P-Donsker. Now we apply Donsker's Theorem (Theorem 19.3 at page 266 of van der Vaart, 1998) to $\sqrt{N} \begin{pmatrix} \widehat{F}_Y^{\widehat{\omega}_A} - F_Y^{\omega_A} \\ \widehat{F}_Y^{\widehat{\omega}_B} - F_Y^{\omega_B} \end{pmatrix}$, which shall converge to a zero mean Gaussian process, defined by $\mathbb{G}^{\omega_{A,B}}$, with variance-covariance matrix

$$\mathbb{E}\left[\mathbb{G}^{\omega_j} (s) \mathbb{G}^{\omega_k} (t)\right] = \mathbb{E}\left[\left(\psi_j (Y, X, T, s) - F_Y^{\omega_j} (s)\right) \cdot \left(\psi_k (Y, X, T, t) - F_Y^{\omega_k} (t)\right)\right],$$

for $(s, t) \in \mathcal{Y} \times \mathcal{Y}$, where j and $k = A, B$. ■

5 Proof of Theorem 2

The proof is divided into two parts. In the first part, presented as an intermediate lemma, we fix $j = A, B$ and (i) derive the asymptotic distribution of $v\left(\widehat{F}_Y^{\widehat{w}}\right)$; (ii) demonstrate that for a fixed y, $\widehat{F}_Y^{\widehat{w}}$ is efficient for F_Y^{ω}; and (iii) therefore $v\left(\widehat{F}_Y^{\widehat{w}}\right)$ will be efficient because v is Hadamard. Finally, in the second part, we derive the asymptotic distribution of the difference $v\left(\widehat{F}_Y^{\widehat{w}_A}\right) - v\left(\widehat{F}_Y^{\widehat{w}_B}\right)$ and demonstrate that it is efficient. In what follows, let $\rightsquigarrow$ denote convergence in law.

Lemma S. 2 (*Semiparametric Efficiency*) *For $j = A, B$, if assumptions 1-6 hold then,*

$$\begin{aligned} & \sqrt{N}\left(\nu\left(\widehat{F}_Y^{\widehat{\omega}_j}\right) - \nu\left(F_Y^{\omega_j}\right)\right) \\ \rightsquigarrow \; & \psi^{\nu}\left(\mathbb{G}_N^{\omega_j}; F_Y^{\omega_j}\right) \\ = \; & \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \omega_j (T_i, p(X_i)) \cdot \phi^{\nu}\left(Y_i; F_Y^{\omega_j}\right) \\ & + \mathbb{E}\left[\frac{\partial \omega_j (T, p(X_i))}{\partial p(X_i)} \cdot \phi^{\nu}\left(Y; F_Y^{\omega_j}\right) \middle| X_i\right] (T_i - p(X_i)) + o_p(1) \end{aligned}$$

where ψ^{ν} is the functional ν's Hadamard derivative, $\phi^{\nu}(Y_i; \cdot) = \psi^{\nu}(\delta_{Y_i}; \cdot)$ and δ_{Y_i} is the Dirac measure at observation i. Moreover, $\nu\left(\widehat{F}_Y^{\widehat{\omega}_j}\right)$ is asymptotically efficient.

Proof of Lemma S.2. Again, we fix j and therefore drop the subscript. Using results in Theorem 1,

$$\sqrt{N}\left(\widehat{F}_Y^{\widehat{\omega}} - F_Y^{\omega}\right) = \mathbb{G}_N^{\omega} + o_p(1) \Rightarrow \mathbb{G}^{\omega}$$

where $\mathbb{G}^{\omega}$ is a Gaussian process with variance-covariance matrix given by

$$\mathbb{E}\left[\mathbb{G}^{\omega}\mathbb{G}^{\omega}\right](s, t) = \mathbb{E}\left[\left(\psi (Y, X, T, s) - F_Y^{\omega}(s)\right) \cdot \left(\psi (Y, X, T, t) - F_Y^{\omega}(t)\right)\right]$$

for $(s, t) \in \mathcal{Y} \times \mathcal{Y}$. Because the map $\nu : \mathcal{F}_{\nu} \to \mathbb{R}$ is Hadamard differentiable at $F_Y^{\omega} \in \mathcal{F}_{\nu}$ we can apply van der Vaart's (1998) Theorem 20.8:

$$\sqrt{N}\left(\nu\left(\widehat{F}_Y^{\widehat{\omega}}\right) - \nu\left(F_Y^{\omega}\right)\right) = \psi^{\nu}\left(\mathbb{G}_N^{\omega}; F_Y^{\omega}\right) + o_p(1).$$

And since ν is Hadamard differentiable, its functional derivative $\psi^{\nu}\left(\cdot;F_Y^{\omega}\right)$ is linear, implying that

$$\begin{aligned}\psi^{\nu}\left(\mathbb{G}_N^{\omega};F_Y^{\omega}\right) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega\left(T_i,p\left(X_i\right)\right)\cdot\phi^{\nu}\left(Y_i;F_Y^{\omega}\right)\\ &\quad+\mathbb{E}\left[\left.\frac{\partial\omega\left(T,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\cdot\phi^{\nu}\left(Y;F_Y^{\omega}\right)\right|X_i\right]\left(T_i-p\left(X_i\right)\right).\end{aligned}$$

We now establish efficiency for $\nu\left(\widehat{F}_Y^{\widehat{\omega}}\right)$ as an estimator for $\nu\left(F_Y^{\omega}\right)$. Consider a (regular) parametric submodel of the joint distribution of (Y,X,T) with cdf $F\left(y,x,t;\theta\right)$. The log-likelihood is

$$\begin{aligned}\ln f\left(y,x,t|\theta\right) &= t\left[\ln f_1\left(y|x,\theta\right)+\ln p\left(x|\theta\right)\right]+\left(1-t\right)\left[\ln f_0\left(y|x,\theta\right)+\ln\left(1-p\left(x|\theta\right)\right)\right]+\ln f\left(x|\theta\right)\\ &= t\ln f_1\left(y|x,\theta\right)+\left(1-t\right)\ln f_0\left(y|x,\theta\right)+t\ln p\left(x|\theta\right)+\left(1-t\right)\ln\left(1-p\left(x|\theta\right)\right)+\ln f\left(x|\theta\right)\end{aligned}$$

where for $j=0,1$ we use the ignorability assumption to write $f\left(y|x,T=j;\theta\right)$ as $f_j\left(y|x,\theta\right)$, which is the conditional density of $Y\left(j\right)$ given x for parameter value θ. Following results in Hahn (1998), we have that the corresponding score function:

$$S\left(y,x,t|\theta\right)=ts_1\left(y|x;\theta\right)+\left(1-t\right)s_0\left(y|x;\theta\right)+\frac{dp\left(x|\theta\right)}{d\theta}\frac{\left(t-p\left(x|\theta\right)\right)}{p\left(x|\theta\right)\left(1-p\left(x|\theta\right)\right)}+s\left(x|\theta\right)$$

where for $j=0,1,\ s_j\left(y|x;\theta\right)\equiv d\ln f_j\left(y|x;\theta\right)/d\theta$ and $s\left(x|\theta\right)\equiv d\ln f\left(x|\theta\right)/d\theta$. The tangent space for this model is:

$$\mathcal{L}=\left\{S\left(y,x,t\right):S\left(y,x,t\right)=ts_1\left(y|x\right)+\left(1-t\right)s_0\left(y|x\right)+a\left(x\right)\left(t-p\left(x\right)\right)+s\left(x\right)\right\}$$

where $a\left(x\right)$ is a square-integrable function of x,

$$\int s_j\left(y|x\right)f_j\left(y|x;\theta_0\right)=0,\ \forall x,\ j=0,1,$$

$$\int s\left(x|\theta\right)f\left(x|\theta_0\right)=0,$$

and for notational simplicity, $p\left(\cdot|\theta_0\right)=p\left(\cdot\right)$.
A parameter $\rho\left(\theta\right)$ is pathwise differentiable if there is a differentiable zero-mean function $F_{\rho}\left(Y,X,T\right)$ such that $F_{\rho}\left(Y,X,T\right)\in\mathcal{L}$ and for all regular parametric models:

$$\left.\frac{\partial\rho\left(\theta\right)}{\partial\theta}\right|_{\theta=\theta_0}=\mathbb{E}\left[F_{\rho}\left(Y,X,T\right)\cdot S\left(Y,X,T|\theta_0\right)\right]$$

We specialize $\rho\left(\theta\right)$ to $F_Y^{\omega}\left(y;\theta\right)$, the weighted distribution of Y at a given $y\in\mathcal{Y}$, which can be written as

$$\begin{aligned}&F_Y^{\omega}\left(y;\theta\right)\\ =&\int\left(\int\mathbb{I}\{z\leq y\}f_1\left(z|x,\theta\right)dz\right)\omega\left(1,p\left(x|\theta\right)\right)p\left(x|\theta\right)f\left(x|\theta\right)dx\\ &+\int\left(\int\mathbb{I}\{z\leq y\}f_0\left(z|x,\theta\right)dz\right)\omega\left(0,p\left(x|\theta\right)\right)\left(1-p\left(x|\theta\right)\right)f\left(x|\theta\right)dx\end{aligned}$$

and calculate its derivative with respect to θ, which can be written as

$$
\begin{aligned}
& \frac{\partial F_Y^{\omega}(y;\theta)}{\partial \theta} \\
= & \int \left(\int \mathbb{I}\{z \le y\} s_1(z|x,\theta) f_1(z|x,\theta)\, dz \right) \omega(1, p(x|\theta))\, p(x|\theta)\, f(x|\theta)\, dx \\
& + \int \left(\int \mathbb{I}\{z \le y\} s_0(z|x,\theta) f_0(z|x,\theta)\, dz \right) \omega(0, p(x|\theta))\, (1-p(x|\theta))\, f(x|\theta)\, dx \\
& + \int \left(\int \mathbb{I}\{z \le y\} f_1(z|x,\theta)\, dz \right) \omega(1, p(x|\theta)) \frac{dp(x|\theta)}{d\theta} f(x|\theta)\, dx \\
& - \int \left(\int \mathbb{I}\{z \le y\} f_0(z|x,\theta)\, dz \right) \omega(0, p(x|\theta)) \frac{dp(x|\theta)}{d\theta} f(x|\theta)\, dx \\
& + \int \left(\int \mathbb{I}\{z \le y\} f_1(z|x,\theta)\, dz \right) \omega(1, p(x|\theta))\, p(x|\theta)\, s(x|\theta)\, f(x|\theta)\, dx \\
& + \int \left(\int \mathbb{I}\{z \le y\} f_0(z|x,\theta)\, dz \right) \omega(0, p(x|\theta))\, (1-p(x|\theta))\, s(x|\theta)\, f(x|\theta)\, dx \\
& + \int \left(\int \mathbb{I}\{z \le y\} f_1(z|x,\theta)\, dz \right) \frac{\partial \omega(1, p(x|\theta))}{\partial p(x|\theta)} \frac{dp(x|\theta)}{d\theta} p(x|\theta)\, f(x|\theta)\, dx \\
& + \int \left(\int \mathbb{I}\{z \le y\} f_0(z|x,\theta)\, dz \right) \frac{\partial \omega(0, p(x|\theta))}{\partial p(x|\theta)} \frac{dp(x|\theta)}{d\theta} (1-p(x|\theta))\, f(x|\theta)\, dx.
\end{aligned}
$$

Now, evaluating that derivative at θ_0 we have

$$
\begin{aligned}
\left. \frac{\partial F_Y^{\omega}(y;\theta)}{\partial \theta} \right|_{\theta=\theta_0} &= \mathbb{E}\left[\omega(T, p(X))\, \mathbb{I}\{Y \le y\} S(Y, X, T|\theta_0)\right] \\
&+ \mathbb{E}\left[\mathbb{E}\left[\frac{\partial \omega(T, p(X))}{\partial p(X)} \cdot \mathbb{I}\{Y \le y\} \middle| X \right] \left. \frac{dp(X|\theta)}{d\theta} \right|_{\theta=\theta_0} \right]
\end{aligned}
$$

and after some tedious algebra one can show that

$$
\begin{aligned}
& \mathbb{E}\left[\mathbb{E}\left[\frac{\partial \omega(T, p(X))}{\partial p(X)} \cdot \mathbb{I}\{Y \le y\} \middle| X \right] \left. \frac{dp(X|\theta)}{d\theta} \right|_{\theta=\theta_0} \right] \\
= & \mathbb{E}\left[\mathbb{E}\left[\frac{\partial \omega(T, p(X))}{\partial p(X)} \cdot \mathbb{I}\{Y \le y\} \middle| X \right] (T - p(X))\, S(Y, X, T|\theta_0) \right]
\end{aligned}
$$

and therefore

$$
\left. \frac{\partial F_Y^{\omega}(y;\theta)}{\partial \theta} \right|_{\theta=\theta_0} = \mathbb{E}\left[F_{\rho}(Y, X, T) \cdot S(Y, X, T|\theta_0)\right]
$$

where, for $\rho(\theta) = F_Y^{\omega}(y;\theta)$, we have that

$$\begin{aligned} & F_{\rho}(Y,X,T) \\ = & \ \omega(T,p(X))\,\mathbb{I}\{Y\leq y\} - F_Y^{\omega}(y;\theta_0) \\ & +\mathbb{E}\left[\left.\frac{\partial\omega(T,p(X))}{\partial p(X)}\mathbb{I}\{Y\leq y\}\right| X\right](T-p(X)) \\ = & \ \psi(Y,X,T,y) - F_Y^{\omega}(y;\theta_0). \end{aligned}$$

We now show that $F_{\rho}(Y,X,T)\in\mathcal{L}$. We rewrite $\psi(Y,X,T,y)$ as

$$\begin{aligned} & \psi(Y,X,T,y) \\ = & \ T\omega(1,p(X))\left(\mathbb{I}\{Y\leq y\} - F_{Y|X,T}(y|X,1;\theta_0)\right) \\ & +(1-T)\,\omega(0,p(X))\left(\mathbb{I}\{Y\leq y\} - F_{Y|X,T}(y|X,0;\theta_0)\right) \\ & +(\mathbb{E}\left[\left.\frac{\partial\omega(T,p(X))}{\partial p(X)}\mathbb{I}\{Y\leq y\}\right| X\right] \\ & +\omega(1,p(X))\,F_{Y|X,T}(y|X,1;\theta_0) - \omega(0,p(X))\,F_{Y|X,T}(y|X,0;\theta_0)\,)\,(T-p(X)) \\ & +p(X)\,\omega(1,p(X))\,F_{Y|X,T}(y|X,1;\theta_0) + (1-p(X))\,\omega(0,p(X))\,F_{Y|X,T}(y|X,0;\theta_0) \end{aligned}$$

and we can easily check that for $j=0,1$:

$$\mathbb{E}\left[\omega(j,p(X))\left(\mathbb{I}\{Y\leq y\} - F_{Y|X,T}(y|X,j;\theta_0)\right)|X,T=j\right] = 0$$

and

$$F_Y^{\omega}(y;\theta_0) = \mathbb{E}\left[F_{Y|X,T}(y|X,1;\theta_0)\,\omega(1,p(X))\,p(X) + F_{Y|X,T}(y|X,0;\theta_0)\,\omega(0,p(X))\,(1-p(X))\right].$$

According to Bickel, Klaassen, Ritov and Wellner (1993), because the estimator for F_Y^{ω}, $\widehat{F}_Y^{\widehat{\omega}}$, is asymptotically linear with influence function $F_{\rho}(Y,X,T)\in\mathcal{L}$, $\widehat{F}_Y^{\widehat{\omega}}$ is also asymptotically efficient at $\mathcal{P}$, the joint distribution of (Y,X,T). Therefore, since ν is Hadamard differentiable, by theorem 25.47 of van der Vaart (1998) $\nu\left(\widehat{F}_Y^{\widehat{\omega}}\right)$ is asymptotically efficient at $\mathcal{P}$ for estimating $\nu(F_Y^{\omega})$. ∎

Proof of Theorem 2. This result follows mechanically from previous lemma. By definition,

$$\begin{aligned} \sqrt{N}\left(\widehat{\Delta}-\Delta\right) & \equiv \sqrt{N}\left(\nu\left(\widehat{F}_Y^{\widehat{\omega}_A}\right) - \nu\left(\widehat{F}_Y^{\widehat{\omega}_B}\right) - \left(\nu\left(F_Y^{\omega_A}\right) - \nu\left(F_Y^{\omega_B}\right)\right)\right) \\ & \rightsquigarrow \psi^{\nu}\left(\mathbb{G}_N^{\omega A}; F_Y^{\omega_A}\right) - \psi^{\nu}\left(\mathbb{G}_N^{\omega_B}; F_Y^{\omega_B}\right) \end{aligned}$$

Now, for the efficiency part, let us define the functional vector $\boldsymbol{\nu}:\mathcal{F}_v\to\mathbb{R}^2$. Thus,

$$\begin{aligned} & \sqrt{N}\left(\nu\left(\widehat{F}_Y^{\widehat{\omega}_A}\right) - \nu\left(\widehat{F}_Y^{\widehat{\omega}_B}\right) - \left(\nu\left(F_Y^{\omega_A}\right) - \nu\left(F_Y^{\omega_B}\right)\right)\right) \\ = & \ \sqrt{N}\begin{bmatrix}1\\-1\end{bmatrix}^{\top}\left(\boldsymbol{\nu}\left(\begin{bmatrix}\widehat{F}_Y^{\widehat{\omega}_A}\\ \widehat{F}_Y^{\widehat{\omega}_B}\end{bmatrix}\right) - \boldsymbol{\nu}\left(\begin{bmatrix}F_Y^{\omega_A}\\ F_Y^{\omega_B}\end{bmatrix}\right)\right). \end{aligned}$$

Result from previous lemma allows us to write

$$\partial\begin{bmatrix}F_Y^{\omega_A}(y;\theta)\\ F_Y^{\omega_B}(y;\theta)\end{bmatrix}\Big/\partial\theta\Bigg|_{\theta=\theta_0} = \mathbb{E}\left[\begin{bmatrix}(\psi_A(Y,X,T,y) - F_Y^{\omega_A}(y;\theta_0))\\ (\psi_B(Y,X,T,y) - F_Y^{\omega_B}(y;\theta_0))\end{bmatrix} S(Y,X,T|\theta_0)\right]$$

and therefore we have that $\boldsymbol{v}\left(\left[\begin{array}{c}\widehat{F}_Y^{\widehat{\omega}_A}\\ \widehat{F}_Y^{\widehat{\omega}_B}\end{array}\right]\right)$ is efficient for $\boldsymbol{v}\left(\left[\begin{array}{c}F_Y^{\omega_A}\\ F_Y^{\omega_B}\end{array}\right]\right)$. Moreover, $\left[\begin{array}{c}1\\ -1\end{array}\right]^\top\left(\boldsymbol{v}\left(\left[\begin{array}{c}\widehat{F}_Y^{\widehat{\omega}_A}\\ \widehat{F}_Y^{\widehat{\omega}_B}\end{array}\right]\right)\right)$ is efficient for $\left[\begin{array}{c}1\\ -1\end{array}\right]^\top\left(\boldsymbol{v}\left(\left[\begin{array}{c}F_Y^{\omega_A}\\ F_Y^{\omega_B}\end{array}\right]\right)\right)$. ■

6 The Influence Functions of Inequality Measures

For completion, we specify the format of the function $\phi^{\nu}\left(Y;F_Y^{\omega}\right)$ for each one of four inequality measures considered in this paper: the Gini coefficient, the Theil Index, the Coefficient of Variation and the Interquartile Range.

6.1 Gini Coefficient

The influence function of the Gini coefficient was derived by Hoeffding (1948) as an example of the results on the asymptotic distribution of U-Statistics. We follow a more recent literature on the statistical properties of inequality measures (Cowell, 2000 and Schluter and Trede, 2003), and we write the influence function of the Gini coefficient as:

$$\phi^{Gini}(y;F_Y^{\omega}) = A_G\left(F_Y^{\omega}\right) + B_G\left(F_Y^{\omega}\right) + C_G\left(y;F_Y^{\omega}\right)$$

where

$$A_G\left(F_Y^{\omega}\right) = 2R_G\left(F_Y^{\omega}\right)/\mu\left(F_Y^{\omega}\right)$$

$$B_G\left(F_Y^{\omega}\right) = A_G\left(F_Y^{\omega}\right)/\mu\left(F_Y^{\omega}\right)$$

$$C_G\left(y;F_Y^{\omega}\right) = -2\mu\left(F_Y^{\omega}\right)^{-1}\left[y\cdot\left[1-F_Y^{\omega}\left(y\right)\right] + GL\left(F_Y^{\omega}\left(y\right),F_Y^{\omega}\right)\right]$$

with

$$\mu\left(F_Y^{\omega}\right) = \int_{\mathcal{Y}} y\cdot dF_Y^{\omega}\left(y\right)$$

$$F_Y^{\omega}\left(y\right) = E\left[\omega\left(T,p\left(X\right)\right)\cdot\mathbb{I}\{Y\leq y\}\right]$$

$$R\left(F_Y^{\omega}\right) = \int_0^1 GL\left(p,F_Y^{\omega}\right)dp$$

$$GL\left(p,F_Y^{\omega}\right) = \int_{-\infty}^{\left(F_Y^{\omega}\right)^{-1}(p)} y\cdot dF_Y^{\omega}\left(y\right)$$

and

$$p = F_Y^{\omega}\left(\left(F_Y^{\omega}\right)^{-1}\left(p\right)\right).$$

6.2 Theil Index

Applying the Delta Method, we obtain de influence function of the Theil index,

$$\phi^{Theil}(y;F_Y^{\omega}) = \frac{1}{\mu^{\omega}}\left(y\cdot\log\left(y\right)-v^{\omega}\right) - \frac{v^{\omega}+\mu^{\omega}}{\left(\mu^{\omega}\right)^2}\left(y-\mu^{\omega}\right)$$

where

$$\mu^{\omega} = \mu\left(F_Y^{\omega}\right)$$

$$v^{\omega} = \int_{\mathcal{Y}} y\cdot\log\left(y\right)\cdot dF_Y^{\omega}\left(y\right)$$

6.3 Coefficient of Variation

In addition, by an application of the Delta Method, we also obtain the influence function of the Coefficient of Variation,

$$\phi^{CV}(y;F_Y^{\omega}) = \frac{1}{2}\frac{(y-\mu^{\omega})^2-\sigma_{\omega}^2}{\mu^{\omega}\cdot\sqrt{\sigma_{\omega}^2}} - \frac{\sqrt{\sigma_{\omega}^2}}{(\mu^{\omega})^2}(y-\mu^{\omega})$$

where

$$\sigma_{\omega}^2 = \int_{\mathcal{Y}} y^2 \cdot dF_Y^{\omega}(y) - (\mu^{\omega})^2$$

6.4 Interquartile Range

Following Ferguson (1996) and Van de Vaart (1998), we define the influence function of the τ-th quantile of the weighted distribution of Y as

$$\phi^{q_{\tau}^{\omega}}(y;F_Y^{\omega}) = \frac{\tau - \mathbf{1}\{y \leq q_{\tau}^{\omega}\}}{f_Y^{\omega}(q_{\tau}^{\omega})}$$

where $f_Y^{\omega}(q_{\tau}^{\omega}) = dF_Y^{\omega}(q_{\tau}^{\omega})$ is the density evaluated at the quantile q_{τ}^{ω}, where $\tau = F_Y^{\omega}(q_{\tau}^{\omega}) \in [0,1]$. The influence function of the interquartile range is the difference of the two quantile influence functions,

$$\phi^{IQR}(y;F_Y^{\omega}) = \frac{0.75 - \mathbf{1}\{y \leq q_{0.75}^{\omega}\}}{f_Y^{\omega}(q_{0.75}^{\omega})} - \frac{0.25 - \mathbf{1}\{y \leq q_{0.25}^{\omega}\}}{f_Y^{\omega}(q_{0.25}^{\omega})}$$

7 Proof of Theorem 3

We divide the proof of this theorem into two parts. We first show in an intermediate lemma that for $j = A, B$, $\sup_{x\in\mathcal{X}}\left|g_j(x) - H_{K_j}(x)^{\top}\widehat{\gamma}_K^{\omega jl}\right| = o_p(1)$. Then, we sow that because all of the nonparametric components of the variance estimator converge uniformly in probability to their population counterparts, the variance estimator is consistent for the asymptotic variance.

Lemma S. 3 (*Uniform Consistency of Regression Component*) *For $j = A, B$ and under assumptions 1-7*

$$\sup_{x\in\mathcal{X}}\left|g_j(x) - H_{K_j}(x)^{\top}\widehat{\gamma}_K^{\omega_j}\right| = o_p(1)$$

Proof of Lemma S.3. Again, we fix j and omit the subscript. Let us define $\widetilde{\gamma}_K^{\omega}$, $\overline{\gamma}_K^{\omega}$ and γ_K^{ω}

$$\widetilde{\gamma}_K^{\omega} = \arg\min_{\gamma}\sum_{i=1}^{N}\left(\left.\frac{\partial\omega(T_i,p(X))}{\partial p(X)}\right|_{p(X)=p(X_i)}\cdot\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}}\right) - H_K(X_i)^{\top}\gamma\right)^2$$

$$\overline{\gamma}_K^{\omega} = \arg\min_{\gamma}\mathbb{E}\left[\left(\frac{\partial\omega(T,p(X))}{\partial p(X)}\cdot\phi^{\nu}\left(Y;\widehat{F}_Y^{\widehat{\omega}}\right) - H_K(X)^{\top}\gamma\right)^2\right]$$

$$\gamma_K^{\omega} = \arg\min_{\gamma}\mathbb{E}\left[\left(\frac{\partial\omega(T,p(X))}{\partial p(X)}\cdot\phi^{\nu}(Y;F_Y^{\omega}) - H_K(X)^{\top}\gamma\right)^2\right].$$

Using triangle inequality,

$$\begin{aligned}\sup_{x\in\mathcal{X}}\left|g(x) - H_K(x)^{\top}\widehat{\gamma}_K^{\omega}\right| \quad\leq\quad & \sup_{x\in\mathcal{X}}\left|g(x) - H_K(x)^{\top}\gamma_K^{\omega}\right| \\ & +\zeta(K)\left(\|\gamma_K^{\omega}-\overline{\gamma}_K^{\omega}\| + \|\overline{\gamma}_K^{\omega}-\widetilde{\gamma}_K^{\omega}\| + \|\widetilde{\gamma}_K^{\omega}-\widehat{\gamma}_K^{\omega}\|\right)\end{aligned}$$

where $\zeta(K) = \sup_x \|H_K(x)\|$. First, under the assumption that the function $g(\cdot)$ is s times continuously differentiable we have that for a fixed K:

$$\sup_{x\in\mathcal{X}} \left| g(x) - H_K(x)^\top \gamma_K^\omega \right| \leq CK^{-\frac{s}{r}}.$$

We then work with differences in the coefficients:

$$\begin{aligned} \|\gamma_K^\omega - \overline{\gamma}_K^\omega\| &= \left\| \mathbb{E}\left[H_K(X) \frac{\partial\omega(T, p(X))}{\partial p(X)} \left(\phi^\nu(Y; F_Y^\omega) - \phi^\nu\left(Y; \widehat{F}_Y^{\widehat{\omega}}\right)\right)\right]\right\| \\ &\leq C \cdot \zeta(K) \cdot \left| \mathbb{E}\left[\phi^\nu\left(Y; \widehat{F}_Y^{\widehat{\omega}}\right) - \phi^\nu(Y; F_Y^\omega)\right]\right|. \end{aligned}$$

Note that because $p(\cdot)$ is bounded away from zero and is less than one, for $l = 0, 1$

$$\sup_{x\in\mathcal{X}} \left| \frac{\partial\omega(l, p(x))}{\partial p(x)} \right| \leq C_l \leq \sup_l C_l = C$$

and as a result of Theorem 2

$$\begin{aligned} \left| \mathbb{E}\left(\phi^\nu\left(Y; \widehat{F}_Y^{\widehat{\omega}}\right) - \phi^\nu(Y; F_Y^\omega)\right)\right| &\leq \mathbb{E}\left[\left| \frac{\partial\phi^\nu(Y; z)}{\partial z}\right|_{z=F_Y^\omega}\right] \sup_{y\in\mathcal{Y}} \left| \widehat{F}_Y^{\widehat{\omega}}(y) - F_Y^\omega(y)\right| + \left|\left(\widehat{F}_Y^{\widehat{\omega}} - F_Y^\omega\right)\right|^2 \\ &\leq CN^{-1/2} \end{aligned}$$

thus

$$\|\gamma_K^\omega - \overline{\gamma}_K^\omega\| \leq C\zeta(K) N^{-1/2}.$$

Now, the difference

$$\begin{aligned} & \|\overline{\gamma}_K^\omega - \widetilde{\gamma}_K^\omega\| \\ = & \left\| N^{-1} \sum_{i=1}^N \left(H_K(X_i) \frac{\partial\omega(T_i, p(X_i))}{\partial p(X_i)} \phi^\nu\left(Y_i; \widehat{F}_Y^{\widehat{\omega}}\right)\right) - \mathbb{E}\left[H_K(X) \frac{\partial\omega(T, p(X))}{\partial p(X)} \phi^\nu\left(Y; \widehat{F}_Y^{\widehat{\omega}}\right)\right]\right\| \\ \leq & CN^{-1/2} \left(\mathbb{V}\left[H_K(X) \frac{\partial\omega(T, p(X))}{\partial p(X)} \phi^\nu\left(Y; \widehat{F}_Y^{\widehat{\omega}}\right)\right]\right)^{1/2} \\ = & CN^{-1/2} \left(\mathbb{V}\left[H_K(X) \frac{\partial\omega(T, p(X))}{\partial p(X)} \left(\phi^\nu(Y; F_Y^\omega) + \left(\phi^\nu\left(Y; \widehat{F}_Y^{\widehat{\omega}}\right) - \phi^\nu(Y; F_Y^\omega)\right)\right)\right]\right)^{1/2} \end{aligned}$$

and working with the variance

$$\begin{aligned}
& \mathbb{V}\left[H_K(X)\frac{\partial\omega(T,p(X))}{\partial p(X)}\left(\phi^{\nu}(Y;F_Y^{\omega})+\left(\phi^{\nu}\left(Y;\widehat{F}_Y^{\widehat{\omega}}\right)-\phi^{\nu}(Y;F_Y^{\omega})\right)\right)\right] \\
\leq\ & \mathbb{V}\left[H_K(X)\frac{\partial\omega(T,p(X))}{\partial p(X)}\phi^{\nu}(Y;F_Y^{\omega})\right] \\
& +\mathbb{V}\left[H_K(X)\frac{\partial\omega(T_i,p(X))}{\partial p(X)}\left(\phi^{\nu}\left(Y;\widehat{F}_Y^{\widehat{\omega}}\right)-\phi^{\nu}(Y;F_Y^{\omega})\right)\right] \\
& +2\left|\mathbb{C}ov\left[H_K(X)\frac{\partial\omega(T,p(X))}{\partial p(X)}\phi^{\nu}(Y;F_Y^{\omega}),H_K(X)\frac{\partial\omega(T,p(X))}{\partial p(X)}\left(\phi^{\nu}\left(Y;\widehat{F}_Y^{\widehat{\omega}}\right)-\phi^{\nu}(Y;F_Y^{\omega})\right)\right]\right| \\
\leq\ & C\zeta^2(K)E\left[(\phi^{\nu}(Y;F_Y^{\omega}))^2\right] \\
& +C\zeta^2(K)E\left[\left(\phi^{\nu}\left(Y;\widehat{F}_Y^{\widehat{\omega}}\right)-\phi^{\nu}(Y;F_Y^{\omega})\right)^2\right] \\
& +C\zeta^2(K)\mathbb{C}ov\left[\phi^{\nu}(Y;F_Y^{\omega}),\left(\phi^{\nu}\left(Y;\widehat{F}_Y^{\widehat{\omega}}\right)-\phi^{\nu}(Y;F_Y^{\omega})\right)\right] \\
=\ & C\zeta^2(K)N^{-1}.
\end{aligned}$$

Therefore we have that

$$\|\overline{\gamma}_K^{\omega}-\widetilde{\gamma}_K^{\omega}\|=CN^{-1/2}\left(\zeta^2(K)N^{-1}\right)^{1/2}=CN^{-1}\zeta(K).$$

Finally,

$$\begin{aligned}
& \|\widetilde{\gamma}_K^{\omega}-\widehat{\gamma}_K^{\omega}\| \\
\leq\ & N^{-1}\left\|\sum_{i=1}^{N}H_K(X_i)\left(\left(\frac{\partial\omega(T_i,\widehat{p}(X_i))}{\partial p(X)}-\frac{\partial\omega(T_i,p(X_i))}{\partial p(X)}\right)\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}}\right)\right)\right\| \\
\leq\ & C\zeta(K)N^{-1}\left\|\sum_{i=1}^{N}\left(\left(\frac{\partial\omega(T_i,\widehat{p}(X_i))}{\partial p(X)}-\frac{\partial\omega(T_i,p(X_i))}{\partial p(X)}\right)\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}}\right)\right)\right\| \\
\leq\ & C\zeta(K)\sup_{t\in\{0,1\},x\in\mathcal{X}}\left|\frac{\partial^2\omega(t,p(x))}{\partial p^2(x)}\right|N^{-1}\left\|\sum_{i=1}^{N}(\widehat{p}(X_i)-p(X_i))\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}}\right)\right\|
\end{aligned}$$

and since the second derivative of with respect to $p(x)$ is bounded,we have that

$$\begin{aligned}
\|\widetilde{\gamma}_K^{\omega}-\widehat{\gamma}_K^{\omega}\| \leq\ & C\zeta(K)N^{-1}\left\|\sum_{i=1}^{N}(\widehat{p}(X_i)-p(X_i))\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}}\right)\right\| \\
\leq\ & C\zeta(K)N^{-1}\left\|\sum_{i=1}^{N}\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}}\right)\right\|\left(\sup_{x\in\mathcal{X}}\left(\Lambda'\left(H_{K_\pi}(x)^{\top}\widetilde{\pi}_K\right)H_{K_\pi}(x)^{\top}\right)\|(\widehat{\pi}_K-\pi_K)\|+\sup_{x\in\mathcal{X}}|p_K(x)-p(x)|\right).
\end{aligned}$$

Now let us work first with

$$\begin{aligned}
& \sup_{x\in\mathcal{X}}\left(\Lambda'\left(H_{K_\pi}(x)^{\top}\widetilde{\pi}_K\right)H_{K_\pi}(x)^{\top}\right)\|(\widehat{\pi}_K-\pi_K)\|+\sup_{x\in\mathcal{X}}|p_K(x)-p(x)| \\
\leq\ & \zeta(K_\pi)\left(K_\pi^{1/2}N^{-1/2}+K_\pi^{-s_p/2r}\right)
\end{aligned}$$

and then with

$$\begin{aligned} N^{-1}\left\|\sum_{i=1}^{N}\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}}\right)\right\| &\leq N^{-1}\left\|\sum_{i=1}^{N}\phi^{\nu}\left(Y_i;F_Y^{\omega}\right)\right\|+N^{-1}\left\|\sum_{i=1}^{N}\left(\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}}\right)-\phi^{\nu}\left(Y_i;F_Y^{\omega}\right)\right)\right\| \\ &\leq O_p(1)+O_p\left(N^{-1/2}\right). \end{aligned}$$

We reach that

$$\|\widetilde{\gamma}_K^{\omega}-\widehat{\gamma}_K^{\omega}\|=C\zeta(K)\,\zeta(K_\pi)\left(K_\pi^{1/2}N^{-1/2}+K_\pi^{-s_p/2r}\right).$$

Therefore, we have that

$$\begin{aligned} & \sup_{x\in\mathcal{X}}\left|g(x)-H_K(x)^{\top}\widehat{\gamma}_K^{\omega}\right| \\ = {} & O_p\left(K^{-\frac{s}{r}}(N)\right)+O_p\left(\zeta^2(K)\,N^{-1/2}\right) \\ & +O_p\left(\zeta^2(K(N))\,N^{-1}\right)+ \\ & +O_p\left(\zeta^2(K(N))\,\zeta(K_\pi(N))\,K_\pi^{1/2}N^{-1/2}\right) \\ & +O_p\left(\zeta^2(K(N))\,\zeta(K_\pi(N))\,K_\pi^{-s_p/2r}\right) \\ = {} & o_p(1). \end{aligned}$$

■

Proof of Theorem 3. We have that $\sup_x|\widehat{p}(x)-p(x)|=o_p(1)$, for $j=A,B$, $\sup_x|\widehat{g}_j(x)-g_j(x)|=o_p(1)$ and $\sup_{y\in\mathcal{Y}}\left|\widehat{F}_Y^{\widehat{\omega}_j}(y)-F_Y^{\omega_j}(y)\right|=o_p(1)$. We can rewrite $\widehat{V}_{AB}$ as

$$\widehat{V}_{AB}=\frac{1}{N}\sum_{i=1}^{N}h\left(T_i,Y_i,\widehat{p}(X_i),\widehat{g}_A(X_i),\widehat{g}_B(X_i);\widehat{F}_Y^{\widehat{\omega}_A},\widehat{F}_Y^{\widehat{\omega}_B}\right)$$

where h is a continuously differentiable function with respect to $W=\left[p(X),g_A(X),g_B(X),F_Y^{\omega_A},F_Y^{\omega_B}\right]^{\top}$. For convenience, define $\widehat{W}=\left[\widehat{p}(X),\widehat{g}_A(X),\widehat{g}_B(X),\widehat{F}_Y^{\widehat{\omega}_A},\widehat{F}_Y^{\widehat{\omega}_B}\right]^{\top}$. Thus, a simple linearization of $\widehat{V}_{AB}$ yields

$$\begin{aligned} \left|\widehat{V}_{AB}-V_{AB}\right| \leq {} & \left\|\mathbb{E}\frac{\partial h}{\partial Z}(T_i,Y_i,W_i)\right\| \\ & \cdot\sup_{x\in\mathcal{X}}|\widehat{p}(x)-p(x)|\sup_{x\in\mathcal{X}}|\widehat{g}_A(x)-g_A(x)|\sup_{x\in\mathcal{X}}|\widehat{g}_B(x)-g_B(x)| \\ & \cdot\sup_{y\in\mathcal{Y}}\left|\widehat{F}_Y^{\widehat{\omega}_A}(y)-F_Y^{\omega_A}(y)\right|\sup_{y\in\mathcal{Y}}\left|\widehat{F}_Y^{\widehat{\omega}_B}(y)-F_Y^{\omega_B}(y)\right|+o_p(1)=o_p(1). \end{aligned}$$

■

8 Proof of Theorem 4

We divide this proof into two parts. Consider the bootstrap scheme in the main text. First, we show that given the sample Z, $\sqrt{N}\left(\widehat{\Delta}_b-\widehat{\Delta}\right)$ converges conditionally in distribution to the same limit as $\sqrt{N}\left(\widehat{\Delta}-\Delta\right)$. Based on that, we show that we can use the percentile bootstrap to construct confidence intervals for Δ.

Proof of Theorem 4. First part. Fix $j = A, B$ and omit that subscript. In the proof of Theorem 1, we show that the class of measurable functions $\mathcal{H} = \{\psi(Y, X, T, y) | y \in \mathcal{Y}\}$ is Donsker with a finite envelope function. Define $\ell^{\infty}(\mathcal{H})$ as the space of bounded functions on $\mathcal{H}$ with supremum norm. From theorems 1 and 2, $\mathbb{G}_N^{\omega}$ is a sequence of mappings with values onto the normed space, $\ell^{\infty}(\mathcal{H})$, converging in distribution to the Gaussian Process $\mathbb{G}^{\omega}$. Following van der Vaart (1998), section 23.2.1, $\mathbb{G}_N^{\omega*}$ is a sequence of mappings with values onto the normed space, $\ell^{\infty}(\mathcal{H})$, converging (conditionally on Z) in distribution to $\mathbb{G}^{\omega}$. Putting it more formally, we use van der Vaart's theorem 23.7 to write

$$\sup_{h \in \mathrm{BL}_1(\ell^{\infty}(\mathcal{H}))} \left| \mathbb{E}_Z \left[h\left(\mathbb{G}_N^{\omega*}\right) \right] - \mathbb{E}\left[h\left(\mathbb{G}^{\omega}\right) \right] \right| \to_p 0$$

where $\mathbb{E}_Z$ denotes the expectation conditionally on $Z = \{(Y_i, X_i, T_i) : i = 1, ..., N\}$.[1]
Because $\nu(\cdot) : \mathcal{F}_{\nu} \to \mathbb{R}$ is Hadamard differentiable tangentially to the subset $\mathcal{F}_{\nu}^{\omega}$, and $\mathcal{F}_{\nu} \subset \ell^{\infty}(\mathcal{H})$, by the Delta-method for bootstrap in probability (Theorem 23.9 in van der Vaart, 1998), conditionally on Z, the sequence $\sqrt{N}\left(\nu\left(\widehat{F}_{Y_b}^{\widehat{\omega}}\right) - \nu\left(\widehat{F}_{Y}^{\widehat{\omega}}\right)\right)$ should converge in distribution to the same limit as $\sqrt{N}\left(\nu\left(\widehat{F}_{Y}^{\widehat{\omega}}\right) - \nu\left(F_Y^{\omega}\right)\right)$. Consequently and given Z, $\sqrt{N}\left(\widehat{\Delta}_b - \widehat{\Delta}\right)$ converges conditionally in distribution to the same limit as $\sqrt{N}\left(\widehat{\Delta} - \Delta\right)$.
Second part. Fix α with $0 < \alpha < 1$. Define the following quantities $F_{\widehat{\Delta}}(r)$ and $F_{\widehat{\Delta}^*}(r)$ for some $r \in \mathbb{R}$ as:

$$\begin{aligned} F_{\widehat{\Delta}}(r) &= \mathbb{E}\left[\mathbf{1}\left\{\sqrt{N}\left(\widehat{\Delta} - \Delta\right) \leq r\right\}\right] \\ F_{\widehat{\Delta}^*}(r) &= \mathrm{B}^{-1} \sum_{b=1}^{N} \mathbf{1}\left\{\sqrt{N}\left(\widehat{\Delta}_b - \widehat{\Delta}\right) \leq r\right\} \end{aligned}$$

and their inverses evaluated at α as[2]

$$\begin{aligned} d_{\alpha} &= F_{\widehat{\Delta}}^{-1}(\alpha) \\ d_{\alpha}^* &= F_{\widehat{\Delta}^*}^{-1}(\alpha). \end{aligned}$$

Note that for an appropriate choice of $h(\cdot)$, we could write $d_{\alpha} = \mathbb{E}\left[h\left(\mathbb{G}^{\omega}\right)\right]$ and $d_{\alpha}^* = \mathbb{E}_Z\left[h\left(\mathbb{G}_N^{\omega*}\right)\right]$. Therefore, from the first part of the theorem, $d_{\alpha}^* - d_{\alpha} \to_p 0$ uniformly.
Now, we show that the bootstrap confidence interval for Δ, $CI^*(\Delta, (1-\alpha)\%)$, can be rewritten using $d_{1-\alpha/2}^*$ and $d_{\alpha/2}^*$:

$$\begin{aligned} & CI^*(\Delta, (1-\alpha)\%) \\ = & \left(2\widehat{\Delta} - \widehat{\Delta}_{[(1-\alpha/2)\cdot \mathrm{B}]}, 2\widehat{\Delta} - \widehat{\Delta}_{[(\alpha/2)\cdot \mathrm{B}]}\right) \\ = & \left(\widehat{\Delta} - \frac{d_{1-\alpha/2}^*}{\sqrt{N}}, \widehat{\Delta} - \frac{d_{\alpha/2}^*}{\sqrt{N}}\right). \end{aligned}$$

---

[1] The set $\mathrm{BL}_1(\ell^{\infty}(\mathcal{H}))$ consists of all functions $h : \ell^{\infty}(\mathcal{H}) \to [-1, 1]$ that are uniformly Lipschitz. See van der Vaart (1998), page 332.

[2] If the cdfs admit flat regions, we then define their inverses (the $\alpha-$quantiles) as being:

$$\begin{aligned} d_{\alpha} &= \inf_{d \in \mathbb{R}} F_{\widehat{\Delta}}(d) \geq \alpha \\ d_{\alpha}^* &= \inf_{d \in \mathbb{R}} F_{\widehat{\Delta}^*}(d) \geq \alpha. \end{aligned}$$

That happens because

$$
\begin{aligned}
\alpha &= \mathrm{B}^{-1}\sum_{b=1}^{N}\mathbf{1}\left\{\widehat{\Delta}_b \leq \widehat{\Delta}_{[\alpha\cdot \mathrm{B}]}\right\} \\
&= \mathrm{B}^{-1}\sum_{b=1}^{N}\mathbf{1}\left\{\sqrt{N}\left(\widehat{\Delta}_b - \widehat{\Delta}\right) \leq \sqrt{N}\left(\widehat{\Delta}_{[\alpha\cdot \mathrm{B}]} - \widehat{\Delta}\right)\right\} \\
&= F_{\widehat{\Delta}^*}\left(\sqrt{N}\left(\widehat{\Delta}_{[\alpha\cdot \mathrm{B}]} - \widehat{\Delta}\right)\right) = F_{\widehat{\Delta}^*}\left(d_\alpha^*\right).
\end{aligned}
$$

Then,

$$
CI^*\left(\Delta,(1-\alpha)\,\%\right) = \left(\widehat{\Delta} - \frac{d^*_{1-\alpha/2}}{\sqrt{N}}, \widehat{\Delta} - \frac{d^*_{\alpha/2}}{\sqrt{N}}\right)
$$

and therefore

$$
\begin{aligned}
\Pr\left[\Delta \in CI^*\left(\Delta,(1-\alpha)\,\%\right)\right] &= \Pr\left[\widehat{\Delta} - \frac{d^*_{1-\alpha/2}}{\sqrt{N}} \leq \Delta \leq \widehat{\Delta} - \frac{d^*_{\alpha/2}}{\sqrt{N}}\right] \\
&= \Pr\left[d^*_{\alpha/2} \leq \sqrt{N}\left(\widehat{\Delta} - \Delta\right) \leq d^*_{1-\alpha/2}\right] \\
&= F_{\widehat{\Delta}}\left(d^*_{1-\alpha/2}\right) - F_{\widehat{\Delta}}\left(d^*_{\alpha/2}\right) \\
&= 1-\alpha/2 + F_{\widehat{\Delta}}\left(d^*_{1-\alpha/2}\right) - F_{\widehat{\Delta}}\left(d_{1-\alpha/2}\right) - \left(\alpha/2 + F_{\widehat{\Delta}}\left(d^*_{\alpha/2}\right) - F_{\widehat{\Delta}}\left(d_{\alpha/2}\right)\right) \\
&\to {}_p 1-\alpha,
\end{aligned}
$$

which holds since $F_{\widehat{\Delta}}(\cdot)$ is continuous, implying that $F_{\widehat{\Delta}}\left(d^*_\alpha\right) - F_{\widehat{\Delta}}\left(d_\alpha\right) \to_p 0$ uniformly. ■

REFERENCES

Bickel, P, Klaassen, C, Ritov, Y, Wellner J. 1993. *Efficient and Adaptive Inference in Semiparametric Models*. Johns Hopkins University Press, Baltimore.

Cowell, F. 2000. *Measurement of Inequality," in Handbook of Income Distribution.* North Holland, Amsterdam.

Donald, S, HSU, Y. 2014. Estimation and Inference for Distribution Functions and Quantile Functions in Treatment Effect Models. *Journal of Econometrics* 178: 383-397.

Ferguson, T. 1996. *A Course in Large Sample Theory.* Chapman & Hall Texts in Statistical Science Series, Taylor & Francis.

Firpo, S. 2007. Efficient Semiparametric Estimation of Quantile Treatment Effects. *Econometrica* 75: 259-276.

Hahn, J. 1998. On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects. *Econometrica* 66: 315-331.

Hirano, K, Imbens, G, Ridder G. 2003. Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. *Econometrica* 71: 1161-1189.

Hoeffding, W. 1948. A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics* 19: 293–325.

Schluter, C, Trede, M. 2003. Local Versus Global Assessment of Mobility. *International Economic Review* 44: 1313-1335.

Tarozzi, A. 2007. Calculating Comparable Statistics from Incomparable Surveys, with an Application to Poverty in India. *Journal of Business and Economic Statistics* 25(3): 314-336.

van der Vaart, A. 1998. *Asymptotic Statistics.* Cambridge, Cambridge University Press.