Counterfactual Treatment Effects: Estimation and Inference Supplementary Appendix

Yu-Chin Hsu[a,b,c], Tsung-Chih Lai[d], and Robert P. Lieli[e]

[a]Institute of Economics, Academia Sinica, Taiwan
[b]Department of Finance, National Central University, Taiwan
[c]Department of Economics, National Chengchi University, Taiwan
[d]Department of Economics, Feng Chia University, Taiwan
[e]Department of Economics and Business, Central European University, Budapest

July 21, 2020

This supplementary appendix contains three parts. Appendix A extends the theoretical analysis to the treated cases. Appendix B collects the proofs of lemmas, theorems, and corollaries. Appendix C provides robustness checks of the empirical results.

A Treated Cases

As mentioned in the main text, treatment effects for the treated subgroup are sometimes more interesting than for the overall population. In this section, we then consider the average counterfactual treatment effect for the treated (ACTT) as

$$\delta_t^* = \mathrm{E}(Y_1^*|D^* = 1) - \mathrm{E}(Y_0^*|D^* = 1), \tag{A.1}$$

and the quantile counterfactual treatment effect for the treated (QCTT) as

$$\delta_t^*(\tau) = \mathrm{Q}_{Y_1^*|D^*}(\tau|1) - \mathrm{Q}_{Y_0^*|D^*}(\tau|1), \tag{A.2}$$

where the expectation and quantile operators are taken with respect to the conditional distribution of Y_d^* given $D^* = 1$ for $d = 0, 1$, where Y_d^* and D^* denote the potential outcomes and treatment indicator in the counterfactual environment with $Y^* = D^*Y_1^* + (1 - D^*)Y_0^*$. Note that since the counterfactual treatment assignment D^* is not observable in our framework, a different set of assumptions is needed to identify (A.1) and (A.2). Define $p^*(x) = \Pr(D^* = 1|X^* = x)$ be the counterfactual propensity score for all $x \in \mathcal{X}^*$.

Assumption A.1 (Unconfoundedness for the Untreated)**.**

(i) $Y_0 \perp\!\!\!\perp D|X$.

(ii) $p(X) > 0$.

Assumption A.2 (Invariance of Conditional Distributions for the Treated)**.**

(i) $F_{Y_d^*|X^*,D^*}(y|x,1) = F_{Y_d|X,D}(y|x,1)$ for all $x \in \mathcal{X}^*$, $d = 0,1$.

(ii) $\mathcal{X}^* \subseteq \mathcal{X}$.

Assumption A.3 (Invariance of Propensity Scores)**.** $p^*(x) = p(x)$ for all $x \in \mathcal{X}^*$.

Clearly, Assumptions A.1 and A.2 are weaker than their counterparts Assumptions 2.1 and 2.3. In fact, they can be further weakened for the ACTT case similar to Section 7 in the main text. To identify treated parameters, however, we need to invoke Assumption A.3 so that the counterfactual treatment assignment can be determined. This assumption requires that the probabilities of receiving treatment must be the same for individuals who are observationally equivalent between counterfactual and status quo populations. Given these assumptions, ACTT and QCTT can be identified as follows.

Lemma A.1. Suppose Assumptions A.1–A.3 hold. Then ACTT and QCTT defined in (A.1) and (A.2) are identified by

$$\begin{aligned}
\delta_t^* &= \int_{\mathcal{X}} \frac{p(x)}{\mathrm{E}[p(X^*)]}[\mathrm{E}(Y|X=x,D=1) - \mathrm{E}(Y|X=x,D=0)]\,\mathrm{d}F_{X^*}(x),\\
\delta_t^*(\tau) &= \inf\left\{y \in \mathcal{Y} : \int_{\mathcal{X}} \frac{p(x)}{\mathrm{E}[p(X^*)]} F_{Y|X,D}(y|x,1)\,\mathrm{d}F_{X^*}(x) \geq \tau\right\}\\
&\quad - \inf\left\{y \in \mathcal{Y} : \int_{\mathcal{X}} \frac{p(x)}{\mathrm{E}[p(X^*)]} F_{Y|X,D}(y|x,0)\,\mathrm{d}F_{X^*}(x) \geq \tau\right\}.
\end{aligned}$$

According to Lemma A.1, ACTT and QCTT estimators are given respectively by

$$\begin{aligned}
\widehat{\delta}_t^* &= \sum_{j=1}^{n^*} \widehat{p}(X_j^*)\Big[\widehat{\mathrm{E}}(Y_1|X=X_j^*) - \widehat{\mathrm{E}}(Y_0|X=X_j^*)\Big]\Big/\sum_{j=1}^{n^*}\widehat{p}(X_j^*),\\
\widehat{\delta}_t^*(\tau) &= \widehat{\mathrm{Q}}_{Y_1^*|D^*}(\tau|1) - \widehat{\mathrm{Q}}_{Y_0^*|D^*}(\tau|1),
\end{aligned}$$

where $\widehat{p}(x)$ is given in (4.9) of the main text, $\widehat{\mathrm{E}}(Y_d|X=x)$ is the Nadaraya-Watson estimator for $d = 0,1$, i.e.,

$$\widehat{\mathrm{E}}(Y_d|X=x) = \frac{\sum_{i=1}^n Y_i\,1\{D_i = d\}K_{x,h}(X_i - x)}{\sum_{i=1}^n 1\{D_i = d\}K_{x,h}(X_i - x)},$$

and $\widehat{\mathrm{Q}}_{Y_d^*|D^*}(\tau|1) = \inf\{y \in \mathcal{Y} : \widehat{F}_{Y_d^*|D^*}(y|1) \geq \tau\}$ with

$$\widehat{F}_{Y_d^*|D^*}(y|1) = \sum_{j=1}^{n^*}\widehat{p}(X_j^*)\widehat{F}_{Y_d|X}(y|X_j^*)\Big/\sum_{j=1}^{n^*}\widehat{p}(X_j^*),$$

where $\widehat{F}_{Y_d|X}(y|x)$ is also given in (4.7) of the main text. Similar to the overall cases, the asymptotic properties of ACTT and QCTT estimators can be derived under a modified version of Assumption 3.3:

Assumption A.4 (Distribution of Y_d^* for the Treated)**.**

(i) $F_{Y_d^*|D^*}(y|1)$ has a compact support $[y_{d\ell}^*, y_{du}^*] \subseteq \mathcal{Y}$.

(ii) $F_{Y_d^*|D^*}(y|1)$ is continuous on $\mathcal{Y}$.

(iii) $f_{Y_d^*|D^*}(y|1)$ is bounded away from 0 and is two-times differentiable on $\mathcal{Y}$.

Corollary A.1. Suppose Assumptions 3.1, 3.2, 3.4–3.6, and A.1–A.4 hold. Then,

$$\sqrt{n}\Big(\widehat{\delta}_t^* - \delta_t^*\Big) \xrightarrow{d} \mathcal{N}\Big(0, \mathrm{Var}(\widehat{\delta}_t^*)\Big),$$

where $\mathrm{Var}(\widehat{\delta}_t^*) = \mathrm{E}[\varrho_{\delta_t^*}(Z)^2] + \mathrm{E}[\varphi_{\delta_t^*}(X^*)^2]$ with

$$\varrho_{\delta_t^*}(Z) = \frac{p(X)}{\mathrm{E}[p(X^*)]}\left\{\frac{D[Y - \mathrm{E}(Y_1|X)]}{p(X)} - \frac{(1-D)[Y - \mathrm{E}(Y_0|X)]}{1-p(X)}\right\}\frac{f_{X^*}(X)}{f_X(X)},$$
$$\varphi_{\delta_t^*}(X^*) = \sqrt{\lambda}\frac{p(X^*)}{\mathrm{E}[p(X^*)]}[\mathrm{E}(Y_1|X^*) - \mathrm{E}(Y_0|X^*) - \delta_t^*].$$

Moreover, we have

$$\sqrt{n}\Big(\widehat{\delta}_t^*(\cdot) - \delta_t^*(\cdot)\Big) \Rightarrow \Delta_t(\cdot),$$

where $\Delta_t(\tau)$ is a Gaussian process with mean zero and covariance function $\Psi_t(\tau_1, \tau_2) = \mathrm{E}[\psi_t(\tau_1)\psi_t(\tau_2)]$, where the variance function $\psi_t(\tau) = \mathrm{E}[\varrho_t(\tau, Z)^2] + \mathrm{E}[\varphi_t(\tau, X^*)^2]$ with

$$\varrho_t(\tau, Z) = -\left[\frac{\varrho_{1,t}^F(\mathrm{Q}_{Y_1^*|D^*}(\tau|1), Z)}{f_{Y_1^*|D^*}(\mathrm{Q}_{Y_1^*|D^*}(\tau|1)|1)} - \frac{\varrho_{0,t}^F(\mathrm{Q}_{Y_0^*|D^*}(\tau|1), Z)}{f_{Y_0^*|D^*}(\mathrm{Q}_{Y_0^*|D^*}(\tau|1)|1)}\right],$$
$$\varphi_t(\tau, X^*) = -\left[\frac{\varphi_{1,t}^F(\mathrm{Q}_{Y_1^*|D^*}(\tau|1), X^*)}{f_{Y_1^*|D^*}(\mathrm{Q}_{Y_1^*|D^*}(\tau|1)|1)} - \frac{\varphi_{0,t}^F(\mathrm{Q}_{Y_0^*|D^*}(\tau|1), X^*)}{f_{Y_0^*|D^*}(\mathrm{Q}_{Y_0^*|D^*}(\tau|1)|1)}\right],$$

where $\varrho_{d,t}^F(y, Z)$ and $\varphi_{d,t}^F(y, X^*)$ are given by

$$\varrho_{d,t}^F(y, Z) = \frac{p(X)}{\mathrm{E}[p(X^*)]}\frac{1\{D=d\}\big[1\{Y \le y\} - F_{Y_d|X}(y|X)\big]}{p(X)^d[1-p(X)]^{1-d}}\frac{f_{X^*}(X)}{f_X(X)},$$
$$\varphi_{d,t}^F(y, X^*) = \sqrt{\lambda}\frac{p(X^*)}{\mathrm{E}[p(X^*)]}\big[F_{Y_d|X}(y|X^*) - F_{Y_d^*}(y)\big],$$

and the convergence takes place in $\ell^\infty([0,1])$.

For uniform inference, we again propose to use multiplier bootstrap to approximate $\Delta_t(\cdot)$. That is,

$$\Delta_t^u(\tau) = \begin{cases} \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^n U_i[\widehat{\varrho}_t(\tau, Z_i) + \widehat{\varphi}_t(\tau, X_i^*)] & \text{if} \quad X^* = \pi(X), \\ \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^n U_i\widehat{\varrho}_t(\tau, Z_i) + \dfrac{1}{\sqrt{n^*}}\displaystyle\sum_{j=1}^{n^*} U_j^*\widehat{\varphi}_t(\tau, X_j^*) & \text{if} \quad X^* \perp\!\!\!\perp X, \end{cases}$$

where $\widehat{\varrho}_t$ and $\widehat{\varphi}_t$ can be estimated given $\widehat{f}_{Y_d^*|D^*}(y|1) = \max\{\widetilde{f}_{Y_d^*|D^*}(y|1), b_n\}$ with

$$\widetilde{f}_{Y_d^*|D^*}(y|1) = \sum_{j=1}^{n^*}\widehat{p}(X_j^*)\widetilde{f}_{Y_d|X}(y|X_j^*) \Big/ \sum_{j=1}^{n^*}\widehat{p}(X_j^*),$$

where $\widehat{p}(x)$ and $\widetilde{f}_{Y_d|X}(y|x)$ are given in main text. One can show $\Delta_t^u(\cdot) \overset{p}{\Rightarrow} \Delta_t(\cdot)$ similar to Theorem 4.1 and then conduct uniform inference accordingly. We omit the details for brevity.

B Proofs

Proof of Lemma 2.1:

By the law of iterated expectations, Assumption 2.3, Assumption 2.1(i), and $Y = Y_d$ for $D = d$, we have

$$\begin{aligned} F_{Y_d^*}(y) &= \int_{\mathcal{X}^*} F_{Y_d^*|X^*}(y|x)\,\mathrm{d}F_{X^*}(x) = \int_{\mathcal{X}} F_{Y_d|X}(y|x)\,\mathrm{d}F_{X^*}(x) \\ &= \int_{\mathcal{X}} F_{Y_d|D,X}(y|d,x)\,\mathrm{d}F_{X^*}(x) = \int_{\mathcal{X}} F_{Y|D,X}(y|d,x)\,\mathrm{d}F_{X^*}(x), \end{aligned}$$

where $F_{Y|D,X}(y|d,x)$ is well defined for all d and x under Assumption 2.3(ii). Since X^* is defined on the same sample space as X that takes values inside $\mathcal{X}$ with probability 1 by Assumption 2.3(ii), $F_{Y_d^*}(y)$ is identified. Accordingly, the corresponding quantile function and the QCTE are identified as well. □

Proof of Lemma 3.1:

The proof consists of two parts. First, we show that $\sqrt{n}(\widetilde{F}_{Y_d^*}(y) - F_{Y_d^*}(y))$ is asymptotically linear with the following influence function representation:

$$\begin{aligned} \sqrt{n}\Big(\widetilde{F}_{Y_d^*}(y) - F_{Y_d^*}(y)\Big) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \frac{1\{D_i = d\}\big[1\{Y_i \le y\} - F_{Y_d|X}(y|X_i)\big]}{p(X_i)^d[1-p(X_i)]^{1-d}}\frac{f_{X^*}(X_i)}{f_X(X_i)} \\ &+ \frac{1}{\sqrt{n^*}}\sum_{j=1}^{n^*}\sqrt{\lambda}\big[F_{Y_d|X}(y|X_j^*) - F_{Y_d^*}(y)\big] + o_p(1) \qquad \text{(B.1)} \\ &\equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varrho_d^F(y, Z_i) + \frac{1}{\sqrt{n^*}}\sum_{j=1}^{n^*}\varphi_d^F(y, X_j^*) + o_p(1). \end{aligned}$$

Since $\varrho_d^F(y,\cdot)$ and $\varphi_d^F(y,\cdot)$ belong to Donsker classes for all $y \in \mathcal{Y}$ and the Cartesian product of two Donsker classes of functions is still a Donsker class as in van der Vaart (2000), Lemma 3.1 holds by the functional central limit theorem for $\widetilde{\boldsymbol{F}} = (\widetilde{F}_{Y_0^*}, \widetilde{F}_{Y_1^*})^T$ in place of $\widehat{\boldsymbol{F}} = (\widehat{F}_{Y_0^*}, \widehat{F}_{Y_1^*})^T$. Next, we complete the proof by establishing the first-order asymptotic equivalence between $\widehat{F}_{Y_d^*}(y)$ and $\widetilde{F}_{Y_d^*}(y)$, i.e.,

$$\sup_{y\in\mathcal{Y}}\Big|\widehat{F}_{Y_d^*}(y) - \widetilde{F}_{Y_d^*}(y)\Big| = o_p(n^{-1/2}). \qquad \text{(B.2)}$$

The derivation of (B.1) is similar to Theorem 1 of Rothe (2010). For simplicity, let $n^* = n$ so that $\lambda = 1$. Let P and P^* be the distribution function of X and X^*, respectively. Denote $\mathcal{G}_n \equiv \sqrt{n}(\mathcal{P}_n - \mathcal{P})$, where $\mathcal{P}$ is the expectation under P and $\mathcal{P}_n$ is the empirical distribution under P such that for every measurable function $\phi : \mathcal{X} \to \mathbb{R}$, $\mathcal{P}\phi = \int \phi\,\mathrm{d}P$ and $\mathcal{P}_n\phi = n^{-1}\sum_{i=1}^n \phi(X_i)$. Define $\mathcal{G}_n^*$, $\mathcal{P}^*$ and $\mathcal{P}_n^*$ similarly under P^*.

To begin with, we rewrite $\sqrt{n}(\widehat{F}_{Y_d^*}(y) - F_{Y_d^*}(y))$ as

$$\begin{aligned} \sqrt{n}\Big(\widehat{F}_{Y_d^*}(y) - F_{Y_d^*}(y)\Big) =& \mathcal{G}_n^*\Big(\widehat{F}_{Y_d|X}(y|x) - F_{Y_d|X}(y|x)\Big) \qquad \text{(B.3)} \\ &+ \sqrt{n}\mathcal{P}^*\Big(\widehat{F}_{Y_d|X}(y|x) - F_{Y_d|X}(y|x)\Big) \qquad \text{(B.4)} \\ &+ \frac{1}{\sqrt{n}}\sum_{j=1}^{n^*}\big(F_{Y_d|X}(y|X_j^*) - F_{Y_d^*}(y)\big). \qquad \text{(B.5)} \end{aligned}$$

It is true that (B.3) is $o_p(1)$ uniformly over $y \in \mathcal{Y}$ by Lemma 1 of Rothe (2010) and Lemma 19.24 of van der Vaart (2000). Next, we show that uniformly over $y \in \mathcal{Y}$,

$$(\text{B.4}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1\{D_i = d\}\big[1\{Y_i \le y\} - F_{Y_d|X}(y|X_i)\big]}{p(X_i)^d[1-p(X_i)]^{1-d}} \frac{f_{X^*}(X_i)}{f_X(X_i)} + o_p(1).$$

Define $G_d(y,x) \equiv \mathrm{E}[1\{Y \le y\}\, 1\{D = d\}|X = x] f_X(x)$, $\widehat{G}_d(y,x) \equiv n^{-1}\sum_{i=1}^{n} 1\{Y_i \le y\}\, 1\{D_i = d\} K_{x,h}(X_i - x)$, $g_d(x) \equiv \mathrm{E}[1\{D = d\}|X = x] f_X(x)$, and $\widehat{g}_d(x) \equiv n^{-1}\sum_{i=1}^{n} 1\{D_i = d\} K_{x,h}(X_i - x)$. It is easy to see that $\widehat{F}_{Y_d|X}(y|x) = \widehat{G}_d(y,x)/\widehat{g}_d(x)$. Moreover, we have

$$\begin{aligned}
&\mathcal{P}^*\Big(\widehat{F}_{Y_d|X}(y|x) - F_{Y_d|X}(y|x)\Big) \\
&= \int \frac{1}{n}\sum_{i=1}^{n}\big[1\{Y_i \le y\} - F_{Y_d|X}(y|X_i)\big] \frac{1\{D_i = d\}K_{x,h}(X_i - x)}{\widehat{g}_d(x)} f_{X^*}(x)\,\mathrm{d}x && \text{(B.6)} \\
&+ \int \frac{1}{n}\sum_{i=1}^{n}\big[F_{Y_d|X}(y|X_i) - F_{Y_d|X}(y|x)\big] \frac{1\{D_i = d\}K_{x,h}(X_i - x)}{\widehat{g}_d(x)} f_{X^*}(x)\,\mathrm{d}x. && \text{(B.7)}
\end{aligned}$$

Since $K_{x,h}(X_i - x)$ is differentiable in x by Assumption 3.5(v), we can apply a second-order Taylor expansion of $\widehat{g}_d(x)$ around $g_d(x)$ in (B.6),

$$\begin{aligned}
(\text{B.6}) &= \int \frac{1}{n}\sum_{i=1}^{n}\big[1\{Y_i \le y\} - F_{Y_d|X}(y|X_i)\big] \frac{1\{D_i = d\}K_{x,h}(x - X_i)}{g_d(x)} f_{X^*}(x)\,\mathrm{d}x && \text{(B.8)} \\
&\quad - \int \frac{1}{n}\sum_{i=1}^{n}\big[1\{Y_i \le y\} - F_{Y_d|X}(y|X_i)\big] \frac{1\{D_i = d\}K_{x,h}(x - X_i)}{g_d^2(x)} [\widehat{g}_d(x) - g_d(x)] f_{X^*}(x)\,\mathrm{d}x && \text{(B.9)} \\
&\quad + o_p(n^{-1/2}),
\end{aligned}$$

where the remainder term is $o_p(n^{-1/2})$ uniformly in both x and y since (i) $\sup_{y \in \mathcal{Y}} |\, 1\{Y_i \le y\} - F_{Y_d|X}(y|X_i)| \le 1$; (ii) $K_{x,h}(x - X_i)$ and $g_d(x)$ are bounded for all $x \in \mathcal{X}$ by Assumptions 3.5 and 3.2(ii); (iii) $\sup_{x \in \mathcal{X}} |\widehat{g}_d(x) - g_d(x)| = O_p((\log n/nh^k)^{1/2} + h^r) = o_p(n^{-1/4})$ by Assumption 3.6 and Lemma B.3 of Newey (1994); (iv) $f_{X^*}(x)$ is bounded by Assumption 3.2(iii); (v) the dominated convergence theorem.

To derive an expression for (B.8), we note that $g_d(x) = p(x)^d[1-p(x)]^{1-d} f_X(x)$ and define $\mu_d(x) \equiv f_{X^*}(x)/g_d(x)$ which is r-times differentiable under Assumptions 3.2(iii) and 3.4(i). Also denote $K_x^{(\gamma)}(u) = \partial^{|\gamma|}/(\partial^{\gamma_1} u_1, \ldots, \partial^{\gamma_k} u_k) K_x(u)$ and $\mu_d^{(\gamma)}(x) = \partial^{|\gamma|}/(\partial^{\gamma_1} x_1, \ldots, \partial^{\gamma_k} x_k)\mu_d(x)$. By a standard change of variables $x = uh + X_i$ and a rth-order Taylor expansion of $K_{uh+X_i}(u)$ and $\mu_d(uh + X_i)$ around $K_{X_i}(u)$ and $\mu_d(X_i)$, respectively, we have uniformly over $y \in \mathcal{Y}$ that

$$
\begin{aligned}
(\text{B.8}) &= \int \frac{1}{n}\sum_{i=1}^{n}\big[1\{Y_i \le y\} - F_{Y_d|X}(y|X_i)\big]\, 1\{D_i = d\} K_{x,h}(x - X_i)\mu_d(x)\,\mathrm{d}x \\
&= \frac{1}{n}\sum_{i=1}^{n}\big[1\{Y_i \le y\} - F_{Y_d|X}(y|X_i)\big]\, 1\{D_i = d\} \int K_{uh+X_i}(u)\mu_d(uh+X_i)\,\mathrm{d}u \\
&= \frac{1}{n}\sum_{i=1}^{n}\big[1\{Y_i \le y\} - F_{Y_d|X}(y|X_i)\big]\, 1\{D_i = d\} \int \Big[K_{X_i}(u) + \cdots + (uh)^r K_{\xi}^{(r)}(u)\Big]\cdot \\
&\qquad\qquad \Big[\mu_d(X_i) + \cdots + (uh)^r e_d^{(r)}(\xi)\Big]\,\mathrm{d}u \\
&= \frac{1}{n}\sum_{i=1}^{n}\big[1\{Y_i \le y\} - F_{Y_d|X}(y|X_i)\big]\, 1\{D_i = d\}\mu_d(X_i) + O_p(h^r) \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1\{D_i = d\}}{p(X_i)^d[1-p(X_i)]^{1-d}}\big[1\{Y_i \le y\} - F_{Y_d|X}(y|X_i)\big]\frac{f_{X^*}(X_i)}{f_X(X_i)} + o_p(n^{-1/2}),
\end{aligned}
$$

where ξ is some value between $uh + X_i$ and X_i. The fourth equality follows from interchanging the differentiation and integration (which is true by the dominated convergence theorem) and Assumption 3.5. The last equality holds because $g_d(x) = p(x)^d[1-p(x)]^{1-d} f_X(x)$ and $O_p(h^r) = o_p(n^{-1/2})$ by Assumption 3.6.

Equation (B.9) can be derived in a similar manner. To be more specific, define $\nu_d(x) \equiv f_{X^*}(x)/g_d^2(x)$ which is also r-times differentiable in x. By the definition of $\widehat{g}_d(x)$,

$$
\begin{aligned}
(\text{B.9}) &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} 1\{D_i = d\}\big[1\{Y_i \le y\} - F_{Y_d|X}(y|X_i)\big]\cdot \\
&\qquad \int [1\{D_j = d\}K_{x,h}(X_j - x) - g_d(x)]K_{x,h}(x - X_i)\nu_d(x)\,\mathrm{d}x \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} 1\{D_i = d\}\big[1\{Y_i \le y\} - F_{Y_d|X}(y|X_i)\big]\cdot \\
&\qquad \left\{\int 1\{D_j = d\}K_{x,h}(X_j - x)K_{x,h}(x - X_i)\nu_d(x)\,\mathrm{d}x - \int K_{x,h}(x - X_i)\mu_d(x)\,\mathrm{d}x\right\} \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} 1\{D_i = d\}\big[1\{Y_i \le y\} - F_{Y_d|X}(y|X_i)\big]\cdot \\
&\qquad \left\{\Big[1\{D_j = d\}K_{X_i,h}(X_j - X_i)\nu_d(X_i) + o_p(n^{-1/2})\Big] - \Big[\mu_d(X_i) + o_p(n^{-1/2})\Big]\right\} \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} 1\{D_i = d\}\big[1\{Y_i \le y\} - F_{Y_d|X}(y|X_i)\big]\nu_d(X_i)\cdot \\
&\qquad [1\{D_j = d\}K_{X_i,h}(X_j - X_i) - g(X_i)] + o_p(n^{-1/2}) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} 1\{D_i = d\}\big[1\{Y_i \le y\} - F_{Y_d|X}(y|X_i)\big]\nu_d(X_i)\cdot \\
&\qquad \{1\{D_j = d\}K_{X_i,h}(X_j - X_i) - \mathrm{E}[1\{D_j = d\}K_{X_i,h}(X_j - X_i)]\} + o_p(n^{-1/2}),
\end{aligned}
\tag{B.10}
$$

where the second equality holds by $\mu_d(x) = g_d(x)\nu_d(x)$, the third holds by applying a similar argument as for (B.8), the fourth holds again by $\mu_d(x) = g_d(x)\nu_d(x)$, and the last equality holds because $\mathrm{E}[1\{D_j = d\}K_{x,h}(X_j - x)] - g(x) = O_p(h^r) = o_p(n^{-1/2})$ uniformly in x by Lemma B.2 of Newey (1994). Moreover, the leading term in (B.10) is a degenerate second-order U-process as pointed out by Rothe (2010). We therefore apply the uniform law of large numbers for U-processes (Nolan and Pollard, 1987; Sherman, 1994) to show that (B.10) is $O_p(h^{-k}n^{-1}) + o_p(n^{-1/2}) = o_p(n^{-1/2})$ under Assumption 3.6.

Combining all the results obtained above, (B.6) can be rewritten as

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1\{D_i=d\}\big[1\{Y_i\leq y\}-F_{Y_d|X}(y|X_i)\big]}{p(X_i)^d[1-p(X_i)]^{1-d}}\frac{f_{X^*}(X_i)}{f_X(X_i)}+o_p(n^{-1/2}).$$

One can also show that (B.7) is $o_p(n^{-1/2})$ through similar arguments. As a result, (B.4) is equivalent to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1\{D_i=d\}\big[1\{Y_i\leq y\}-F_{Y_d|X}(y|X_i)\big]}{p(X_i)^d[1-p(X_i)]^{1-d}}\frac{f_{X^*}(X_i)}{f_X(X_i)}+o_p(1),$$

and we have the asymptotic linear representation in (B.1). Since $1\{Y\leq y\}$ is a type I function and the other functions in (B.1) are type II functions defined in Andrews (1994), $\varrho_d^F(y,\cdot)$ and $\varphi_d^F(y,\cdot)$ belong to some Donsker classes for all $y\in\mathcal{Y}$. By van der Vaart (2000, p.270) in which the Cartesian product of two Donsker classes of functions is still a Donsker class, Lemma 3.1 holds by the functional central limit theorem for $\widetilde{\boldsymbol{F}}$ in place of $\widehat{\boldsymbol{F}}$.

We now show the second part of the proof which claims that $\widehat{F}_{Y_d^*}(y)$ and $\widetilde{F}_{Y_d^*}(y)$ are asymptotically equivalent to the first-order approximation, or $\sup_{y\in\mathcal{Y}}|\widehat{F}_{Y_d^*}(y)-\widetilde{F}_{Y_d^*}(y)|=o_p(n^{-1/2})$ as stated in (B.2). For simplicity assume that $\widetilde{F}_{Y_d^*}(0)\geq 0$ so that $\phi_1(\widetilde{F}_{Y_d^*})(y)=\sup_{y'\leq y}\widetilde{F}_{Y_d^*}(y')$ for all $y\in\mathcal{Y}=[0,\bar{y}]$. From the first part of the proof, it is true that $\sup_{y\in\mathcal{Y}}|\sqrt{n}(\widetilde{F}_{Y_d^*}(y)-F_{Y_d^*}(y))|=O_p(1)$, implying that for any $\epsilon_1>0$, there exist an $M>0$ and a large $N=N_M$ such that for all $n>N$,

$$\mathrm{P}\left(\sup_{y\in\mathcal{Y}}\left|\sqrt{n}\Big(\widetilde{F}_{Y_d^*}(y)-F_{Y_d^*}(y)\Big)\right|\leq M\right)\geq 1-\epsilon_1. \tag{B.11}$$

Next, it can also be shown that $\sqrt{n}(\widetilde{F}_{Y_d^*}(y)-F_{Y_d^*}(y))$ is stochastic equicontinuous with respect to the pseudometric $\rho(y_1,y_2)=|F_{Y_d^*}(y_1)-F_{Y_d^*}(y_2)|^{1/2}$ for all $(y_1,y_2)\in\mathcal{Y}$ from Theorem 3.1 of Hsu, Lai, and Lieli (2019), meaning that for any $\epsilon_2>0$ and $\epsilon_3>0$, there exist a $\delta>0$ small enough and an N_δ large enough such that for all $n>N_\delta$,

$$\mathrm{P}\left(\sup_{\rho(y_1,y_2)\leq\delta}\left|\sqrt{n}\Big(\widetilde{F}_{Y_d^*}(y_1)-F_{Y_d^*}(y_1)\Big)-\sqrt{n}\Big(\widetilde{F}_{Y_d^*}(y_2)-F_{Y_d^*}(y_2)\Big)\right|\leq\epsilon_2\right)\geq 1-\epsilon_3. \tag{B.12}$$

If we pick a large N such that

$$2M/\sqrt{N}<\delta^2, \tag{B.13}$$

for $y_1\leq y_2$ with $\rho(y_1,y_2)>\delta$ and for $n>N$ with $\sup_{y\in\mathcal{Y}}|\sqrt{n}(\widetilde{F}_{Y_d^*}(y)-F_{Y_d^*}(y))|\leq M$ almost surely,

$$\begin{aligned}\widetilde{F}_{Y_d^*}(y_1)-\widetilde{F}_{Y_d^*}(y_2)=&\Big(\widetilde{F}_{Y_d^*}(y_1)-F_{Y_d^*}(y_1)\Big)-\Big(\widetilde{F}_{Y_d^*}(y_2)-F_{Y_d^*}(y_2)\Big)-\big(F_{Y_d^*}(y_2)-F_{Y_d^*}(y_1)\big)\\ \leq&2M/\sqrt{n}-\delta^2<2M/\sqrt{N}-\delta^2<0\quad\text{a.s.,}\end{aligned} \tag{B.14}$$

where the inequality holds almost surely because $\sqrt{n}\Big(\widetilde{F}_{Y_d^*}(y_1)-F_{Y_d^*}(y_1)\Big)\leq M$, $\sqrt{n}\Big(\widetilde{F}_{Y_d^*}(y_2)-F_{Y_d^*}(y_2)\Big)\geq$

$-M$, and $F_{Y_d^*}(y_2) - F_{Y_d^*}(y_1) > \delta^2$ by the definition of $\rho(y_1, y_2)$. This implies that for $n > N$,

$$\phi_1(\widetilde{F}_{Y_d^*})(y) = \sup_{y' \le y} \widetilde{F}_{Y_d^*}(y') = \sup_{\{y': y' \le y, \rho(y', y) \le \delta\}} \widetilde{F}_{Y_d^*}(y'). \tag{B.15}$$

Therefore, for all $y \in \mathcal{Y}$ and for $n > N$ with $\sup_{y \in \mathcal{Y}} |\sqrt{n}(\widetilde{F}_{Y_d^*}(y) - F_{Y_d^*}(y))| \le M$, we have that

$$\begin{aligned}
0 &\le \sqrt{n}\Big(\phi_1(\widetilde{F}_{Y_d^*})(y) - \widetilde{F}_{Y_d^*}(y)\Big) \\
&= \sqrt{n}\Big(\phi_1(\widetilde{F}_{Y_d^*})(y) - F_{Y_d^*}(y) - \big(\widetilde{F}_{Y_d^*}(y) - F_{Y_d^*}(y)\big)\Big) \\
&= \sqrt{n}\left(\sup_{\{y': y' \le y, \rho(y', y) \le \delta\}} \Big(\widetilde{F}_{Y_d^*}(y') - F_{Y_d^*}(y) - \big(\widetilde{F}_{Y_d^*}(y) - F_{Y_d^*}(y)\big)\Big)\right) \\
&\le \sup_{\{y': y' \le y, \rho(y', y) \le \delta\}} \sqrt{n}\Big(\widetilde{F}_{Y_d^*}(y') - F_{Y_d^*}(y') - \big(\widetilde{F}_{Y_d^*}(y) - F_{Y_d^*}(y)\big)\Big) \\
&\le \sup_{\rho(y_1, y_2) \le \delta} \Big|\sqrt{n}\Big(\widetilde{F}_{Y_d^*}(y_1) - F_{Y_d^*}(y_1)\Big) - \sqrt{n}\Big(\widetilde{F}_{Y_d^*}(y_2) - F_{Y_d^*}(y_2)\Big)\Big| \quad \text{a.s.},
\end{aligned} \tag{B.16}$$

where the second-to-last inequality holds because $F_{Y_d^*}(y') \le F_{Y_d^*}(y)$ for $y' \le y$, and the last inequality holds because the last supremum is taken over all y_1 and y_2 such that $\rho(y_1, y_2) \le \delta$ instead of over y' such that $y' \le y$ and $\rho(y', y) \le \delta$. That is, $\sqrt{n}(\phi_1(\widetilde{F}_{Y_d^*})(y) - \widetilde{F}_{Y_d^*}(y))$ is $o_p(1)$ by (B.12).

Finally, since it is true that $\sup_{y \in \mathcal{Y}} \sqrt{n}(\phi_1(\widetilde{F}_{Y_d^*})(y) - 1) = \sup_{y \in \mathcal{Y}} \sqrt{n}(\widetilde{F}_{Y_d^*}(y) - 1) = o_p(1)$ by Theorem 3.1 of Hsu, Lai, and Lieli (2019), we have for all $y \in \mathcal{Y}$,

$$\begin{aligned}
\sqrt{n}\Big(\widehat{F}_{Y_d^*}(y) - \widetilde{F}_{Y_d^*}(y)\Big) =& \sqrt{n}\left(\frac{\phi_1(\widetilde{F}_{Y_d^*})(y)}{\sup_{y \in \mathcal{Y}} \phi_1(\widetilde{F}_{Y_d^*})(y)} - \widetilde{F}_{Y_d^*}(y)\right) \\
=& \sqrt{n}\Big(\phi_1(\widetilde{F}_{Y_d^*})(y) - \widetilde{F}_{Y_d^*}(y)\Big) - \phi_1(\widetilde{F}_{Y_d^*})(y)\sqrt{n}\Big(\sup_{y \in \mathcal{Y}} \phi_1(\widetilde{F}_{Y_d^*})(y) - 1\Big) + o_p(1) \\
=& \sqrt{n}\Big(\phi_1(\widetilde{F}_{Y_d^*})(y) - \widetilde{F}_{Y_d^*}(y)\Big) + o_p(1),
\end{aligned} \tag{B.17}$$

where the second equality follows from a mean-valued expansion of $\sup_{y \in \mathcal{Y}} \phi_1(\widetilde{F})_{Y_d^*}(y)$ around 1 and the last equality holds because $\phi_1(\widetilde{F}_{Y_d^*})(y) \xrightarrow{p} 1$ and $\sup_{y \in \mathcal{Y}} \sqrt{n}(\phi_1(\widetilde{F}_{Y_d^*})(y) - 1) = o_p(1)$. Put together, when conditions for (B.11), (B.12) and (B.13) hold, by (B.16) and (B.17),

$$\begin{aligned}
&\mathrm{P}\left(\sup_{y \in \mathcal{Y}} \Big|\sqrt{n}\Big(\widehat{F}_{Y_d^*}(y) - \widetilde{F}_{Y_d^*}(y)\Big)\Big| \le \epsilon_2\right) \\
\ge &\mathrm{P}\left(\sup_{\rho(y_1, y_2) \le \delta} \Big|\sqrt{n}\Big(\widetilde{F}_{Y_d^*}(y_1) - F_{Y_d^*}(y_1)\Big) - \Big(\widetilde{F}_{Y_d^*}(y_2) - F_{Y_d^*}(y_2)\Big)\Big| \le \epsilon_2\right) \ge 1 - \epsilon_3.
\end{aligned}$$

□

Proof of Theorem 3.1:

Given the quantile map is Hadamard differentiable, Theorem 3.1 follows immediately from Lemma 3.1 and the functional delta method. □

Proof of Theorem 4.1:

We consider only the independent case here since the proof for the transformed case is similar to that of Donald and Hsu (2014, Theorem 4.5). We first show that

$$\mathcal{F}_d^u(y) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i \widehat{\varrho}_d^F(y, Z_i) + \frac{1}{\sqrt{n^*}}\sum_{j=1}^{n^*} U_j^* \widehat{\varphi}_d^F(y, X_j^*) \overset{p}{\Rightarrow} \mathcal{F}_d(y).$$

To see this, note that

$$\mathcal{F}_d^u(y) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i \varrho_d^F(y, Z_i) + \frac{1}{\sqrt{n^*}}\sum_{j=1}^{n^*} U_j^* \varphi_d^F(y, X_j^*) \tag{B.18}$$

$$+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i \big[\widehat{\varrho}_d^F(y, Z_i) - \varrho_d^F(y, Z_i)\big] \tag{B.19}$$

$$+ \frac{1}{\sqrt{n^*}}\sum_{j=1}^{n^*} U_j^* \big[\widehat{\varphi}_d^F(y, X_j^*) - \varphi_d^F(y, X_j^*)\big]. \tag{B.20}$$

We now show (B.19) converges weakly to a zero process conditional on the sample path $\mathcal{Z} \equiv \{\omega \in Z_i : i = 1, 2, \dots\}$ with probability approaching one. That is,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i \big[\widehat{\varrho}_d^F(y, Z_i) - \varrho_d^F(y, Z_i)\big] \overset{p}{\Rightarrow} 0. \tag{B.21}$$

Note that (B.21) is true if and only if for any subsequence k_n of n, there exists a further subsequence ℓ_n of k_n such that

$$\frac{1}{\sqrt{\ell_n}}\sum_{i=1}^{\ell_n} U_i \big[\widehat{\varrho}_{d,\ell_n}^F(y, Z_i) - \varrho_d^F(y, Z_i)\big] \overset{\text{a.s.}}{\Rightarrow} 0, \tag{B.22}$$

where $\widehat{\varrho}_{d,\ell_n}^F(y,z)$ denotes the estimator at ℓ_n. By Lemma 4.1, we have $\sup_{y\in\mathcal{Y},z\in\mathcal{Z}} |\widehat{\varrho}_{d,\ell_n}^F(y,z) - \varrho_d^F(y,z)| \overset{\text{a.s.}}{\to} 0$ for any subsequence k_n of n and a further subsequence ℓ_n of k_n. We then define $\mathcal{Z}_{\ell_n} \equiv \{\omega \in \mathcal{Z} : \sup_{y\in\mathcal{Y},z\in\mathcal{Z}} |\widehat{\varrho}_{d,\ell_n}^F(y,z)(\omega) - \varrho_d^F(y,z)| \to 0\}$ where $\widehat{\varrho}_{d,\ell_n}^F(y,z)(\omega)$ denotes the realization at ω and $\mathrm{P}(\mathcal{Z}_{\ell_n}) = 1$. For any $\omega \in \mathcal{Z}_{\ell_n}$, define

$$t_{\ell_n,i}(U_i, y|\omega) = \frac{U_i}{\sqrt{\ell_n}}\big[\widehat{\varrho}_{d,\ell_n}^F(y, Z_i)(\omega) - \varrho_d^F(y, Z_i)\big].$$

Note that since we have conditioned $t_{\ell_n,i}(U_i, y|\omega)$ on the sample path ω, the randomness comes from the U_i's which is independent of the sample path ω.

Next, we claim that the triangular array $\{t_{\ell_n,i}(U_i, y|\omega), 1 \le i \le \ell_n, \ell_n \ge 1\}$ satisfies assumptions (i)–(v) of Theorem 10.6 in Pollard (1990). If this is the case, we can then apply the functional central limit theorem to show:

$$\frac{1}{\sqrt{\ell_n}}\sum_{i=1}^{\ell_n} U_i \big[\widehat{\varrho}_{d,\ell_n}^F(y, Z_i)(\omega) - \varrho_d^F(y, Z_i)\big] \Rightarrow 0,$$

meaning that (B.22) and (B.21) would follow accordingly. By Theorem 3.1 in Hsu (2016), it is sufficient to check that $\widehat{\varrho}^F_{d,\ell_n}(y, Z_i)$ satisfies (i)–(iii) of Assumption 3.2 in Hsu (2016):

(i) $\{\widehat{\varrho}^F_{d,\ell_n}(y, Z_i) : y \in \mathcal{Y}, 1 \leq i \leq \ell_n, \ell_n \geq 1\}$ is manageable with respect to the envelope function $\{\widehat{\Omega}_{\ell_n}(Z_i) : 1 \leq i \leq \ell_n, \ell_n \geq 1\}$ in the sense of Definition 7.9 of Pollard (1990).

(ii) $\sup_{y_1,y_2\in\mathcal{Y}} |\ell_n^{-1} \sum_{i=1}^{\ell_n} \widehat{\varrho}^F_{d,\ell_n}(y_1, Z_i)\widehat{\varrho}^F_{d,\ell_n}(y_2, Z_i) - \mathrm{E}[\varrho^F_d(y_1, Z)\varrho^F_d(y_2, Z)]| \xrightarrow{p} 0$.

(iii) There exists a $\delta > 0$ such that

$$\frac{1}{\ell_n}\sum_{i=1}^{\ell_n} \widehat{\Omega}^2_{\ell_n}(Z_i) - \frac{1}{\ell_n}\sum_{i=1}^{\ell_n} \Omega^2_{\ell_n}(Z_i) \xrightarrow{p} 0, \qquad \frac{1}{\ell_n}\sum_{i=1}^{\ell_n} \widehat{\Omega}^{2+\delta}_{\ell_n}(Z_i) - \frac{1}{\ell_n}\sum_{i=1}^{\ell_n} \Omega^{2+\delta}_{\ell_n}(Z_i) \xrightarrow{p} 0.$$

To check (i), recall that

$$\widehat{\varrho}^F_{d,\ell_n}(y, Z_i) = \frac{1\{D_i = d\}\Big[1\{Y_i \leq y\} - \widehat{F}_{Y_d|X,\ell_n}(y|X_i)\Big]}{\widehat{p}_{\ell_n}(X_i)^d[1 - \widehat{p}_{\ell_n}(X_i)]^{1-d}} \frac{\widehat{f}_{X^*,\ell_n}(X_i)}{\widehat{f}_{X,\ell_n}(X_i)},$$

where the subscript ℓ_n indicates estimators at ℓ_n. Since $1\{Y_i \leq y\}$ for all $y \in \mathcal{Y}$ forms a Vapnik-Chervonenkis class of functions, it is manageable with respect to the envelope function of 1's. In addition, due to monotonicity $\widehat{F}_{Y_d|X,\ell_n}(y|x)$ satisfies Pollard's entropy condition as in (4.2) of Andrews (1994) with the envelope function being $M_{\ell_n} \geq 1$. Next, by construction $a_{\ell_n} = \inf_{x\in\mathcal{X}} \widehat{p}_{\ell_n}(x) = \inf_{x\in\mathcal{X}} 1 - \widehat{p}_{\ell_n}(x)$ and $b_{\ell_n} = \inf_{x\in\mathcal{X}} \widehat{f}_{X,\ell_n}(x)$. Since $\widehat{f}_{X^*,\ell_n}(x)$ is uniformly bounded by, say B_{ℓ_n}, it belongs to a type II class of functions with the envelope function being B_{ℓ_n}. Taken all together, $\widehat{\varrho}^F_{d,\ell_n}(y, Z_i)$ is manageable with respect to a constant envelope function $\widehat{\Omega}_{\ell_n} = a_{\ell_n} b_{\ell_n} B_{\ell_n}(1 + M_{\ell_n}) > 0$, and hence (i) is satisfied.

To check (ii) and (iii), note that the functions involved in $\widehat{\varrho}^F_d(y, z)$ are uniformly consistent over $y \in \mathcal{Y}$ and $z \in \mathcal{Z}$ by Lemma 4.1. It is therefore easy to see (ii) and (iii) follow accordingly. In other words, the triangular array $\{t_{\ell_n,i}(U_i, y|\omega)\}$ for all $\omega \in \mathcal{Z}$ satisfies all requirements in Theorem 10.6 of Pollard (1990). We then argue that conditional on the sample path ω and given the randomness coming from the U_i's,

$$\frac{1}{\sqrt{\ell_n}}\sum_{i=1}^{\ell_n} U_i\big[\widehat{\varrho}^F_{d,\ell_n}(y, X_i)(\omega) - \varrho^F_d(y, X_i)\big] \Rightarrow 0.$$

By a similar argument, it can be shown that (B.20) also converges weakly to a zero process conditional on the sample path $\{\omega \in X^*_j : j = 1, 2, \dots\}$. Finally, by Corollary 2.9.3 in van der Vaart and Wellner (1996), it is true that (B.18) converges weakly to $\mathcal{F}_d(y)$ with probability approaching one.

We are now ready to show the conditional weak convergence of the simulated process for QCTE,

$$\Delta^u(\tau) = -\left[\frac{\mathcal{F}^u_1(\widehat{\mathrm{Q}}_{Y^*_1}(\tau))}{\widehat{f}_{Y^*_1}(\widehat{\mathrm{Q}}_{Y^*_1}(\tau))} - \frac{\mathcal{F}^u_0(\widehat{\mathrm{Q}}_{Y^*_0}(\tau))}{\widehat{f}_{Y^*_0}(\widehat{\mathrm{Q}}_{Y^*_0}(\tau))}\right] \overset{p}{\Rightarrow} \Delta(\tau). \tag{B.23}$$

Note that by Assumptions 3.3(iii) and 4.1, it follows that

$$
\begin{aligned}
&\sup_{\tau\in[0,1]}\left|\widehat{f}_{Y_d^*}(\widehat{\mathrm{Q}}_{Y_d^*}(\tau))-f_{Y_d^*}(\mathrm{Q}_{Y_d^*}(\tau))\right|\\
&\le \sup_{\tau\in[0,1]}\left|\widehat{f}_{Y_d^*}(\widehat{\mathrm{Q}}_{Y_d^*}(\tau))-f_{Y_d^*}(\widehat{\mathrm{Q}}_{Y_d^*}(\tau))\right|+\sup_{\tau\in[0,1]}\left|f_{Y_d^*}(\widehat{\mathrm{Q}}_{Y_d^*}(\tau))-f_{Y_d^*}(\mathrm{Q}_{Y_d^*}(\tau))\right|\\
&\le \sup_{\tau\in[0,1]}\left|\widehat{f}_{Y_d^*}(\widehat{\mathrm{Q}}_{Y_d^*}(\tau))-f_{Y_d^*}(\widehat{\mathrm{Q}}_{Y_d^*}(\tau))\right|+C\cdot\sup_{\tau\in[0,1]}\left|\widehat{\mathrm{Q}}_{Y_d^*}(\tau)-\mathrm{Q}_{Y_d^*}(\tau)\right|=o_p(1),
\end{aligned}
$$

for some constant C. Moreover, it is also true that $\sup_{\tau\in[0,1]}|\mathcal{F}_d^u(\widehat{\mathrm{Q}}_{Y_d^*}(\tau))-\mathcal{F}_d^u(\mathrm{Q}_{Y_d^*}(\tau))|=o_p(1)$ conditioning on the sample path with probability approaching one by the equicontinuity of $\mathcal{F}_d^u(y)$ and the uniform consistency of $\widehat{\mathrm{Q}}_{Y_d^*}(\tau)$. As a result, we have

$$
\begin{aligned}
&\sup_{\tau\in[0,1]}\left|\frac{\mathcal{F}_d^u(\widehat{\mathrm{Q}}_{Y_d^*}(\tau))}{\widehat{f}_{Y_d^*}(\widehat{\mathrm{Q}}_{Y_d^*}(\tau))}-\frac{\mathcal{F}_d^u(\mathrm{Q}_{Y_d^*}(\tau))}{f_{Y_d^*}(\mathrm{Q}_{Y_d^*}(\tau))}\right|\\
&\le \sup_{\tau\in[0,1]}\left|\frac{\mathcal{F}_d^u(\widehat{\mathrm{Q}}_{Y_d^*}(\tau))}{\widehat{f}_{Y_d^*}(\widehat{\mathrm{Q}}_{Y_d^*}(\tau))}-\frac{\mathcal{F}_d^u(\widehat{\mathrm{Q}}_{Y_d^*}(\tau))}{f_{Y_d^*}(\mathrm{Q}_{Y_d^*}(\tau))}\right|+\sup_{\tau\in[0,1]}\left|\frac{\mathcal{F}_d^u(\widehat{\mathrm{Q}}_{Y_d^*}(\tau))}{f_{Y_d^*}(\mathrm{Q}_{Y_d^*}(\tau))}-\frac{\mathcal{F}_d^u(\mathrm{Q}_{Y_d^*}(\tau))}{f_{Y_d^*}(\mathrm{Q}_{Y_d^*}(\tau))}\right|\\
&\le \sup_{\tau\in[0,1]}\left|\mathcal{F}_d^u(\widehat{\mathrm{Q}}_{Y_1^*}(\tau))\right|\sup_{\tau\in[0,1]}\left|\frac{\widehat{f}_{Y_d^*}(\widehat{\mathrm{Q}}_{Y_d^*}(\tau))-f_{Y_d^*}(\mathrm{Q}_{Y_d^*}(\tau))}{\widehat{f}_{Y_d^*}(\widehat{\mathrm{Q}}_{Y_d^*}(\tau))f_{Y_d^*}(\mathrm{Q}_{Y_d^*}(\tau))}\right|+\sup_{\tau\in[0,1]}\left|\frac{\mathcal{F}_1^u(\widehat{\mathrm{Q}}_{Y_1^*}(\tau))-\mathcal{F}_1^u(\mathrm{Q}_{Y_1^*}(\tau))}{f_{Y_d^*}(\mathrm{Q}_{Y_d^*}(\tau))}\right|\\
&\le C\cdot\sup_{\tau\in[0,1]}\left|\mathcal{F}_d^u(\widehat{\mathrm{Q}}_{Y_1^*}(\tau))\right|\sup_{\tau\in[0,1]}\left|\widehat{f}_{Y_d^*}(\widehat{\mathrm{Q}}_{Y_d^*}(\tau))-f_{Y_d^*}(\mathrm{Q}_{Y_d^*}(\tau))\right|+C'\cdot\sup_{\tau\in[0,1]}\left|\mathcal{F}_1^u(\widehat{\mathrm{Q}}_{Y_1^*}(\tau))-\mathcal{F}_1^u(\mathrm{Q}_{Y_1^*}(\tau))\right|\\
&=C\cdot O_p(1)\cdot o_p(1)+C'\cdot o_p(1)=o_p(1),
\end{aligned}
$$

where the last inequality holds by Assumption 3.3(iii) and the fact that $\widehat{f}_{Y_d^*}(y)$ is bounded from below for all y by the trimming method. Finally, (B.23) holds because

$$
\begin{aligned}
\Delta^u(\tau)=-\Bigg\{&\left[\frac{\mathcal{F}_1^u(\widehat{\mathrm{Q}}_{Y_1^*}(\tau))}{\widehat{f}_{Y_1^*}(\widehat{\mathrm{Q}}_{Y_1^*}(\tau))}-\frac{\mathcal{F}_1^u(\mathrm{Q}_{Y_1^*}(\tau))}{f_{Y_1^*}(\mathrm{Q}_{Y_1^*}(\tau))}\right]+\left[\frac{\mathcal{F}_0^u(\widehat{\mathrm{Q}}_{Y_0^*}(\tau))}{\widehat{f}_{Y_0^*}(\widehat{\mathrm{Q}}_{Y_0^*}(\tau))}-\frac{\mathcal{F}_0^u(\mathrm{Q}_{Y_0^*}(\tau))}{f_{Y_0^*}(\mathrm{Q}_{Y_0^*}(\tau))}\right]\\
&+\left[\frac{\mathcal{F}_1^u(\mathrm{Q}_{Y_1^*}(\tau))}{f_{Y_1^*}(\mathrm{Q}_{Y_1^*}(\tau))}-\frac{\mathcal{F}_0^u(\mathrm{Q}_{Y_0^*}(\tau))}{f_{Y_0^*}(\mathrm{Q}_{Y_0^*}(\tau))}\right]\Bigg\}\overset{p}{\Rightarrow}\Delta(\tau). \qquad \square
\end{aligned}
$$

Proof of Lemma 4.1:

It suffices to check that

$$
\begin{aligned}
&\sup_{y\in\mathcal{Y}}\left|\widehat{F}_{Y_d^*}(y)-F_{Y_d^*}(y)\right|+\sup_{y\in\mathcal{Y},x\in\mathcal{X}}\left|\widehat{F}_{Y_d|X}(y|x)-F_{Y_d|X}(y|x)\right|+\sup_{x\in\mathcal{X}}|\widehat{p}(x)-p(x)|\\
&+\sup_{x\in\mathcal{X}}\left|\widehat{f}_X(x)-f_X(x)\right|+\sup_{x\in\mathcal{X}}\left|\widehat{f}_{X^*}(x)-f_{X^*}(x)\right|+\sup_{y\in\mathcal{Y}}\left|\widehat{f}_{Y_d^*}(y)-f_{Y_d^*}(y)\right|=o_p(1).
\end{aligned} \tag{B.24}
$$

The first term in (B.24) has already been shown by Lemma 3.1. Regarding the second and third terms in (B.24), note that the uniform consistency of the unmodified estimators $\widetilde{F}_{Y_d|X}(y|x)$ and $\widetilde{p}(x)$ are established by Härdle, Jansson and Serfling (1988). We then follow Lemma 4.1 of Donald and Hsu (2014) to show that the monotonized estimator $\widehat{F}_{Y_d|X}(y|x)$ is also uniformly consistent for both y and x, i.e., the second term in (B.24) holds. For a fixed x, suppose y' is the first point at which $\widetilde{F}_{Y_d|X}(y|x)$ jumps down. Then for $y\in[y',y'+\epsilon)$ where $\epsilon>0$, $\widehat{F}_{Y_d|X}(y|x)=\widetilde{F}_{Y_d|X}(y'-\epsilon|x)>\widetilde{F}_{Y_d|X}(y|x)$. On the other hand, for $y\in[y'-\epsilon,y,y')$,

$\widehat{F}_{Y_d|X}(y|x) = \widetilde{F}_{Y_d|X}(y'-\epsilon|x)$. Now, focus on the case where $y' \leq y < y'+\epsilon$. If $\widehat{F}_{Y_d|X}(y|x) \leq F_{Y_d|X}(y|x)$, we have $F_{Y_d|X}(y|x) - \widetilde{F}_{Y_d|X}(y|x) > F_{Y_d|X}(y|x) - \widehat{F}_{Y_d|X}(y|x) \geq 0$. If $\widehat{F}_{Y_d|X}(y|x) > F_{Y_d|X}(y|x)$ and recall that $F_{Y_d|X}(y|x)$ is nondecreasing in y, we have $\widetilde{F}_{Y_d|X}(y'-\epsilon|x) - F_{Y_d|X}(y'-\epsilon|x) \geq \widehat{F}_{Y_d|X}(y|x) - F_{Y_d|X}(y|x) > 0$. These results imply that for $y' \leq y < y'+\epsilon$,

$$\Big|\widehat{F}_{Y_d|X}(y|x) - F_{Y_d|X}(y|x)\Big| \leq \max\Big\{\Big|\widetilde{F}_{Y_d|X}(y'-\epsilon|x) - F_{Y_d|X}(y'-\epsilon|x)\Big|, \Big|\widetilde{F}_{Y_d|X}(y|x) - F_{Y_d|X}(y|x)\Big|\Big\}.$$

From the above inequality and the fact that $\widehat{F}_{Y_d|X}(y|x) = \widetilde{F}_{Y_d|X}(y|x)$ for $0 \leq y < y'$, it is true that

$$\sup_{0\leq y\leq y'+\epsilon}\Big|\widehat{F}_{Y_d|X}(y|x) - F_{Y_d|X}(y|x)\Big| \leq \sup_{0\leq y\leq y'+\epsilon}\Big|\widetilde{F}_{Y_d|X}(y|x) - F_{Y_d|X}(y|x)\Big|.$$

Since this inequality holds for all x and by induction for y', we can then show that

$$\sup_{y\in\mathcal{Y},x\in\mathcal{X}}\Big|\widehat{F}_{Y_d|X}(y|x) - F_{Y_d|X}(y|x)\Big| \leq \sup_{y\in\mathcal{Y},x\in\mathcal{X}}\Big|\widetilde{F}_{Y_d|X}(y|x) - F_{Y_d|X}(y|x)\Big| = o_p(1).$$

For the third term in (B.24), since $|a_n| \leq 1$, it follows that $\sup_{x\in\mathcal{X}}|\widehat{p}(x) - p(x)| \leq \sup_{x\in\mathcal{X}}|\widetilde{p}(x) - p(x)| = o_p(1)$.

Regarding the fourth term in (B.24), note that $\sup_{x\in\mathcal{X}}|\widetilde{f}_X(x) - f_X(x)| = o_p(1)$ is already established by Jones (1993). Given the fact that b_n converges to 0 and by Assumption 3.2(ii), we have $\inf_{x\in\mathcal{X}} f_X(x) \geq b_n$ with probability approaching one. This implies that

$$\sup_{x\in\mathcal{X}}\Big|\widetilde{f}_X(x) - \widehat{f}_X(x)\Big| = o_p(1).$$

By triangular inequality we then have

$$\sup_{x\in\mathcal{X}}\Big|\widehat{f}_X(x) - f_X(x)\Big| \leq \sup_{x\in\mathcal{X}}\Big|\widehat{f}_X(x) - \widetilde{f}_X(x)\Big| + \sup_{x\in\mathcal{X}}\Big|\widetilde{f}_X(x) - f_X(x)\Big| = o_p(1).$$

For the other parts in (B.24), since $\sup_{y\in\mathcal{Y},x\in\mathcal{X}}|\widetilde{f}_{Y_d|X}(y|x) - f_{Y_d|X}(y|x)| = o_p(1)$ as shown by Hyndman, Bashtannyk and Grunwald (1996), the results regarding the fifth and the last terms follow similarly. □

Proof of Corollary 7.1:

The proof is omitted since it is similar to the proof of Lemma 3.1 by replacing $1\{Y_i \leq y\}$'s with Y_i's. □

Proof of Lemma A.1:

To see this, note that $\mathrm{Q}_{Y_d^*|D^*}(\tau|1) = \inf\{y \in \mathcal{Y} : F_{Y_d^*|D^*}(y|1) \geq \tau\}$ and

$$
\begin{aligned}
F_{Y_d^*|D^*}(y|1) &= \int_{\mathcal{X}^*} F_{Y_d^*|X^*,D^*}(y|x,1)\,\mathrm{d}F_{X^*|D^*}(x|1) \\
&= \int_{\mathcal{X}} F_{Y_d|X,D}(y|x,1) f_{X^*|D^*}(x|1)\,\mathrm{d}x \\
&= \int_{\mathcal{X}} F_{Y|X,D}(y|x,d) \frac{p^*(x) f_{X^*}(x)}{\mathrm{P}(D^*=1)}\,\mathrm{d}x \\
&= \int_{\mathcal{X}} F_{Y|X,D}(y|x,d) \frac{p(x)}{\int_{\mathcal{X}} p(x) f_{X^*}(x)\,\mathrm{d}x}\,\mathrm{d}F_{X^*}(x) \\
&= \int_{\mathcal{X}} F_{Y|X,D}(y|x,d) \frac{p(x)}{\mathrm{E}[p(X^*)]}\,\mathrm{d}F_{X^*}(x),
\end{aligned}
$$

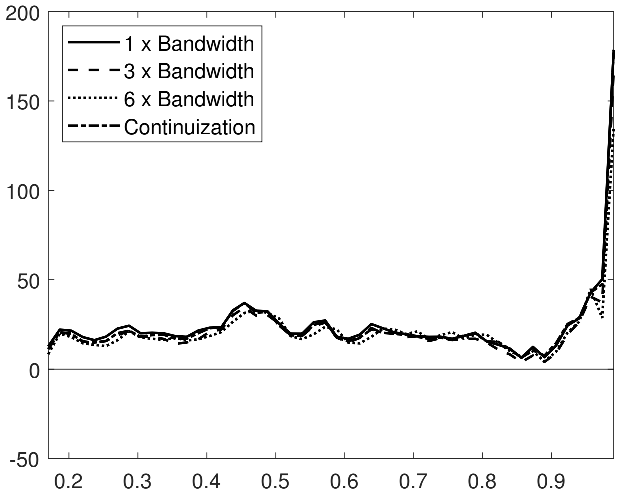
where the second equality follows from Assumption 7.6 and the third holds since $Y_1 = Y$ if $D = 1$, $F_{Y_0|X,D}(y|x,1) = F_{Y_0|X,D}(y|x,0) = F_{Y|X,D}(y|x,0)$ by Assumption 7.5(i), and by Bayes' theorem. The fourth equality is true given Assumption 7.7. Since $p(x) > 0$ for all $x \in \mathcal{X}$ by Assumption 7.5(ii) and Y, X, D, and X^* are all observable, the last line is well defined and is identified. Thus, ACTT and QCTT can be identified as well. □

Proof of Corollary A.1:

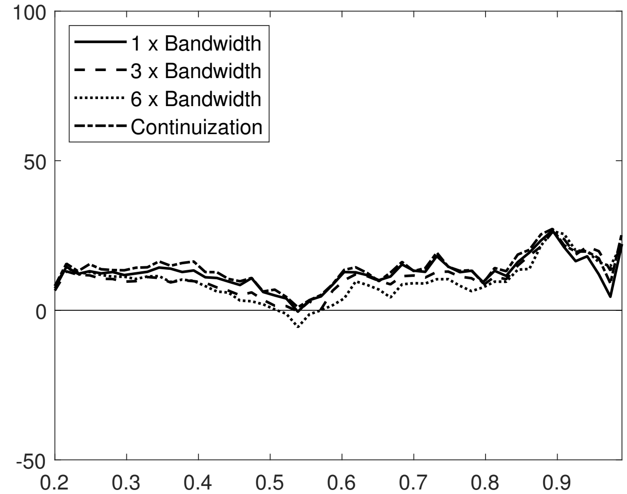
The proof follows the same line of reasoning as in Lemma 3.1 and Theorem 3.1 and so is omitted. □

C Robustness Checks

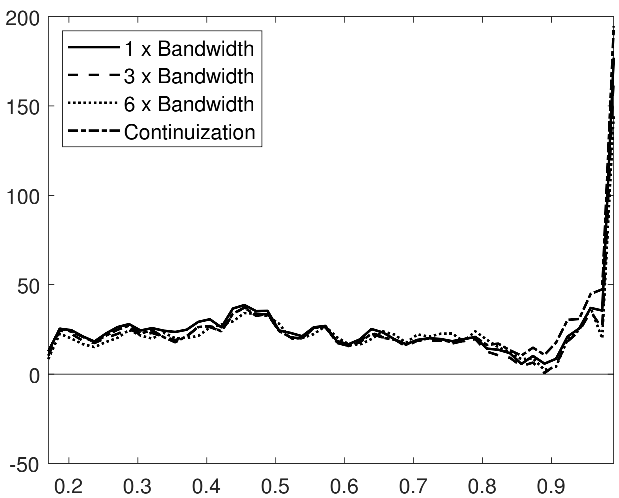
In this section, we undertake several sensitivity checks to examine the robustness of our empirical results. First, since age is recorded in years (i.e., an integer variable) in our dataset, we then vary the bandwidths from h_d to $3h_d$ and $6h_d$ for $d = 0, 1$ such that the bandwidth values are less than 1, between 1 and 2, and between 2 and 3, respectively, for both males and females. Next, we "continuize" the age variable by adding a small random noise. To be more specific, we add a uniformly distributed random number in the range $[-0.5, 0.5]$ to the integer-valued age to make it more continuous. The point estimates of QTEs and QCTEs under these conditions are presented in Figure C.1. As can be seen from the figure, the estimates are virtually identical to those reported in the main text, suggesting that our empirical results are not sensitive with respect to the choice of bandwidth. Thus, we conclude that these robustness checks confirm our main findings.



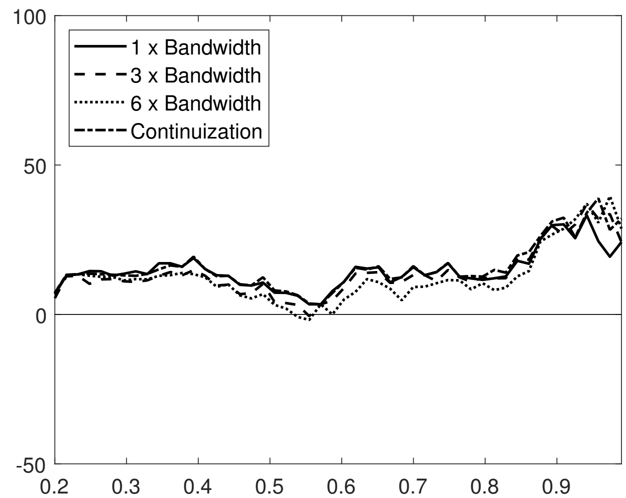
(a) QTEs of Job Corps on earnings for males.



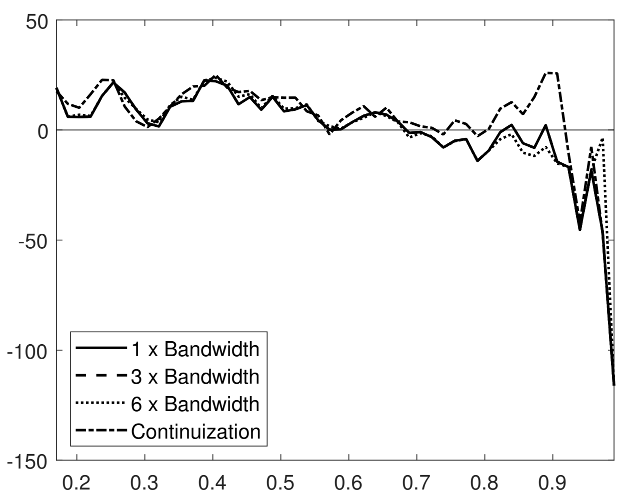
(b) QTEs of Job Corps on earnings for females.



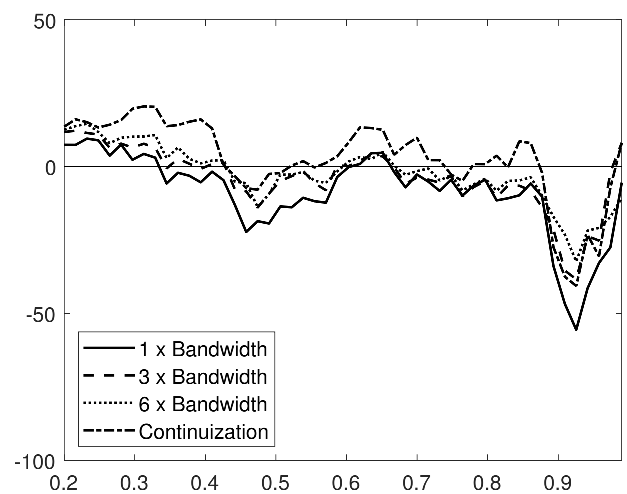
(c) QCTEs for females with male's earnings structure.



(d) QCTEs for males with female's earnings structure.



(e) QCTEs for males with increased education.



(f) QCTEs for females with increased education.

Figure C.1. Robustness checks.

References

[1] Andrews, D. W. (1993). *Empirical process methods in econometrics.* Cowles Foundation for Research in Economics.

[2] Donald, S. G., and Hsu, Y. C. (2014). "Estimation and inference for distribution functions and quantile functions in treatment effect models." *Journal of Econometrics*, 178, 383-397.

[3] Hardle, W., Janssen, P., and Serfling, R. (1988). "Strong uniform consistency rates for estimators of conditional functionals." *The Annals of Statistics*, 16(4), 1428-1449.

[4] Hsu, Y. C. (2016). "Multiplier bootstrap for empirical processes" (No. 16-A010). Institute of Economics, Academia Sinica, Taipei, Taiwan.

[5] Hsu, Y.-C., Lieli, R. P., and Lai, T.-C. (2019). "Estimation and inference for distribution functions and quantile functions in endogenous treatment effect models." Working Paper.

[6] Hyndman, R. J., Bashtannyk, D. M., and Grunwald, G. K. (1996). "Estimating and visualizing conditional densities." *Journal of Computational and Graphical Statistics*, 5(4), 315-336.

[7] Jones, M. C. (1993). "Simple boundary correction for kernel density estimation." *Statistics and computing*, 3(3), 135-146.

[8] Newey, W. K. (1994). "Kernel estimation of partial means and a general variance estimator." *Econometric Theory*, 10(2), 1-21.

[9] Nolan, D., and Pollard, D. (1987). "U-processes: rates of convergence." *The Annals of Statistics*, 780-799.

[10] Pollard, D. (1990). *Empirical processes: theory and applications.* In NSF-CBMS regional conference series in probability and statistics (pp. i-86). Institute of Mathematical Statistics and the American Statistical Association.

[11] Rothe, C. (2010). "Nonparametric estimation of distributional policy effects." *Journal of Econometrics*, 155(1), 56-70.

[12] Sherman, R. P. (1994). "Maximal inequalities for degenerate U-processes with applications to optimization estimators." *The Annals of Statistics*, 439-459.

[13] Van der Vaart, A. W. (2000). *Asymptotic statistics* (Vol. 3). Cambridge university press.

[14] Vaart, A. W., and Wellner, J. A. (1996). *Weak convergence and empirical processes: with applications to statistics.*