Online Appendix for "Multiple Testing and the Distributional Effects of Accountability Incentives in Education"

Steven F. Lehrer* — Queen's University and NBER

R. Vincent Pohl† — Mathematica

Kyungchul Song‡ — University of British Columbia

March 2021

Abstract

This is the online appendix for Lehrer, Pohl, and Song (2021). Three sections are included, which provide further details regarding (1) additional motivation for the testing procedure, (2) the asymptotic validity of the multiple testing procedure, and (3) robustness checks for the main result.

*Department of Economics, email: lehrers@queensu.ca.

†Division of Health Policy Assessment, email: vincent.pohl@gmail.com.

‡Vancouver School of Economics, email: kysong@mail.ubc.ca.

A Additional motivation for the testing procedure from an empirical perspective

The well-documented diverse and heterogeneous behavior in how individuals respond to a particular treatment or intervention has not only changed how economists think about econometric models and policy evaluation but also has profound consequences for the scientific evaluation of public policy. James Heckman stresses this point in his 2001 Nobel lecture, where he notes that conditional mean impacts including the average treatment effect may provide limited guidance for policy design and implementation (Heckman, 2001). Although the importance of heterogeneous treatment effects is widely recognized in the causal inference literature, common practice remains to report an average causal effect parameter. While an increasing number of studies account for possible treatment effect heterogeneity when evaluating programs or other interventions, most conduct statistical inference without allowing for dependence across subgroups.

As Fink, McConnell, and Vollmer (2014) point out, a majority of studies based on field experiments published in 10 specific journals estimate separate average causal parameters for different subgroups, but report traditional standard errors and p-values when testing for heterogeneous treatment effects through interaction terms or subgroup analyses. This is inappropriate because each interaction term represents a separate hypothesis beyond the original experimental design and results in a substantially increased type I error. The problem when testing multiple hypotheses jointly is the potential over-rejection of the null hypothesis. Intuitively, if the null hypothesis of no treatment effect is true, testing it across 100 subsamples, we expect about five rejections at the 95 percent confidence level. However, since the probability of a false positive equals 0.05 for each individual hypothesis, the probability of falsely rejecting at least one true null hypothesis may be much larger. Hence, the type I error exceeds the nominal size of the test.

A similar observation can be made for distributional treatment effects. A growing number of studies examine if treatment effects differ across quantiles of the outcome variable, i.e. they estimate quantile treatment effects (QTEs) (e.g., Heckman, Smith, and Clements, 1997; Friedlander and Robins, 1997; Abadie, 2002; Bitler, Gelbach, and Hoynes, 2006; Firpo, 2007). Testing for the presence of positive (or, generally, non-zero) QTEs involves a test of multiple hypotheses, for example 99 hypotheses in the case of percentile treatment effects. Therefore, the naive approach of comparing individual test results to find quantile groups with positive and statistically significant treatment effects inevitably suffers from the issue of data mining due to the reuse of the same data as emphasized by White (2000). As a result, the type I error rates can exceed the desired level of the test, which leads researchers to reject "too many" individual hypotheses.[1] To the best of our knowledge, no published study estimating distributional treatment effects makes such a correction. Among articles published in five high-impact economic journals between 2008 to 2017 that estimate distributional treatment effects none corrects inference for multiple testing (Allen, Clark, and Houde, 2014; Angrist, Lang, and Oreopoulos, 2009; Bandiera et al., 2017; Banerjee et al., 2015; Behaghel, de Chaisemartin, and Gurgand, 2017; Brown et al., 2014; Crepon et al., 2015; Evans and Garthwaite, 2012; Fack and Landais, 2010; Fairlie and Robinson, 2013; McKenzie, 2017; Meyer and Sullivan, 2008; Muralidharan, Niehaus, and Sukhtankar, 2016). The absence of these corrections may reflect that econometric testing procedures for QTEs were not previously developed. Lehrer, Pohl, and Song (2021) aims to fill that gap and also provide a formal result of asymptotic validity when the propensity scores are parametrically specified.

B Mathematical proofs

Throughout the proof, we assume that $B = \infty$ for simplicity, so that we ignore the bootstrap simulation errors in the proof. Our first result is the asymptotic linear representation of $\sqrt{n}(\hat{q}_d(\tau) - q_d(\tau))$ and its bootstrap version that is uniform over $\tau \in [\tau_L, \tau_U]$. Let us introduce some notation. Let

$$a_\tau(Y_i; q) = \tau - 1\{Y_i \leqslant q\}, \tag{1}$$

and for $u \in \mathbb{R}$,

$$\Delta(Y_i; q, u) = \sqrt{n} \int_0^1 \left(a_\tau(Y_i; q + n^{-1/2}us) - a_\tau(Y_i; q) \right) ds. \tag{2}$$

Note that the right hand side in (2) does not depend on τ.

Let $J_d(\tau_U, \tau_L) = \{q_d(\tau) : \tau \in [\tau_L, \tau_U]\}$, and assume that it is bounded in $\mathbb{R}$. (See Assumption 2.3(ii) of the main text.) For $q \in J_d(\tau_U, \tau_L)$ and $u \in \mathbb{R}$, let

$$\varphi_n(Y_i; q, u) = \frac{\Delta(Y_i; q, u)}{n^{3/4}} = -n^{-1/4} \int_0^1 1\{q < Y_i \leqslant q + n^{-1/2}us\} ds. \tag{3}$$

[1]In part as a response, statistical inference procedures developed in Heckman, Smith, and Clements (1997), Abadie, Angrist, and Imbens (2002), Rothe (2010), and Maier (2011), among others, focus on the whole distribution of potential outcomes to side-step multiple comparisons.

Let $\mathcal{B}$ be a class of bounded measurable functions $b : \mathbb{R} \times \mathcal{X} \times \{0,1\} \to \mathbb{R}$. Define for each $u \in \mathbb{R}$,

$$\mathcal{H}_n(u) = \{\varphi_n(\cdot; q, u)b(\cdot) : (q, b) \in J_d(\tau_U, \tau_L) \times \mathcal{B}\}. \tag{4}$$

We let $V_i = (Y_i, X_i', D_i) \in \mathbb{R}^{d_V}$ for brevity of notation.

We introduce a pseudo-norm $\|\cdot\|_{P,2}$ on the set of measurable functions on $\mathbb{R} \times \mathcal{X} \times \{0,1\}$: $\|f\|_{P,2} = (E|f(V_i)|^2)^{1/2}$, for any measurable map f. We also denote the sup norm by $\|\cdot\|_\infty$: $\|f\|_\infty = \sup_{v \in \mathbb{R}^{d_V}} |f(v)|$. For each $\varepsilon > 0$ and $u \in \mathbb{R}$, let $N_{[]}(\varepsilon, \mathcal{H}_n(u), \|\cdot\|_{P,2})$ denote the ε-bracketing number of $\mathcal{H}_n(u)$ with respect to $\|\cdot\|_{P,2}$ (see van der Vaart and Wellner, 1996, p. 83).

Lemma B.1 *For $u \in \mathbb{R}$, there exist constants $C_1, C_2, C_3, C_4 > 0$ such that for each $\varepsilon \in (0,1)$, there are brackets $[h_{L,j}, h_{U,j}]$ with $1 \leqslant j \leqslant N(\varepsilon)$, satisfying that the brackets cover $\mathcal{H}_n(u)$ and for each $k \geqslant 2$,*

$$E[|h_{L,j}(V_i) - h_{U,j}(V_i)|^k] \leqslant C_1(C_2 n^{-1/4})^{k-2}\varepsilon^2, \tag{5}$$

and

$$\log N(\varepsilon) \leqslant C_3 - C_3 \log(\varepsilon) + C_4 \log N_{[]}(C\varepsilon^2, \mathcal{B}, \|\cdot\|_2). \tag{6}$$

Proof: First, define for $\delta > 0$,

$$\begin{aligned} z_\delta(y; q, s, u) \quad &= \quad (1 - \min\{(y - q - n^{-1/2}us)/\delta, 1\})1\{0 < y - q - n^{-1/2}us\} \\ &\quad +1\{y - q - n^{-1/2}us \leqslant 0\}. \end{aligned} \tag{7}$$

Define

$$\varphi_{U,\delta}(y; q, u) = \int_0^1 \varphi_{U,\delta}(y; q, s, u)ds, \text{ and} \tag{8}$$

$$\varphi_{L,\delta}(y; q, u) = \int_0^1 \varphi_{L,\delta}(y; q, s, u)ds, \tag{9}$$

where

$$\varphi_{U,\delta}(y; q, s, u) \quad = \quad z_\delta(y; q, s) - z_\delta(y + \delta + n^{-1/2}us; q, s), \text{ and} \tag{10}$$

$$\varphi_{L,\delta}(y; q, s, u) \quad = \quad \min\left\{z_\delta(y + \delta; q, s), z_\delta(-y + 2q + \delta + n^{-1/2}us; q, s)\right\}. \tag{11}$$

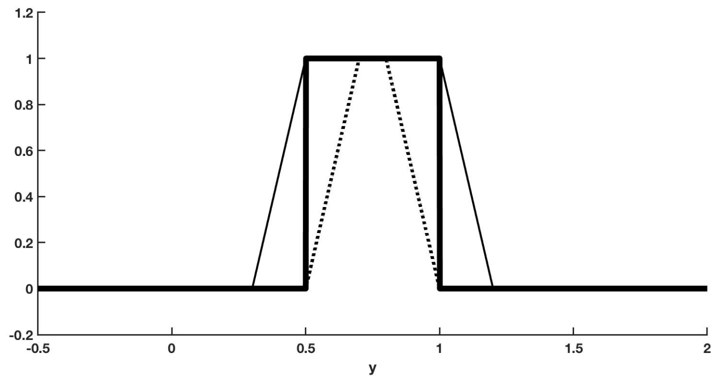


Figure 1: Illustration of $\varphi_{U,\delta}, \varphi_{L,\delta}$: The solid thick line depicts $\varphi(\cdot; q, s, u)$, the solid thin line $\varphi_{U,\delta}(\cdot; q, s, u)$ and the dotted line $\varphi_{L,\delta}(\cdot; q, s, u)$. Here we take $n^{-1/2}us = 0.5, q = 0.5$ and $\delta = 0.2$. The absolute slope of both maps $\varphi_{U,\delta}(\cdot; q, s, u)$ and $\varphi_{L,\delta}(\cdot; q, s, u)$ are bounded by $1/\delta$.

Let $\varphi(y; q, s, u) = 1\{y - q - n^{-1/2}us \leqslant 0\} - 1\{y - q \leqslant 0\}$ and define

$$\varphi_n(y; q, u) = n^{-1/4} \int_0^1 \varphi(y; q, s, u) ds. \tag{12}$$

Note that the definition (3) conforms with this.

Then, we have for all $y \in \mathbb{R}$, (see Figure 1)

$$\varphi_{L,\delta}(y; q, u) \leqslant n^{1/4} \varphi_n(y; q, u) \leqslant \varphi_{U,\delta}(y; q, u). \tag{13}$$

It is not hard to see that for all $q, q' \in \mathbb{R}$, and all $y \in \mathbb{R}$,

$$\begin{aligned} |\varphi_{U,\delta}(y; q, u) - \varphi_{U,\delta}(y; q', u)| &\leqslant |q - q'|/\delta, \text{ and} \\ |\varphi_{L,\delta}(y; q, u) - \varphi_{L,\delta}(y; q', u)| &\leqslant |q - q'|/\delta. \end{aligned} \tag{14}$$

Furthermore, for some constant $C > 0$,

$$E\left[(\varphi_{U,\delta}(Y_i; q, u) - \varphi_{L,\delta}(Y_i; q, u))^2 \,|D_i = d \right] \leqslant C\delta, \tag{15}$$

and

$$E\left[\varphi^2_{U,\delta}(Y_i; q, u)|D_i = d\right] \leqslant 1, \text{ and } E\left[\varphi^2_{L,\delta}(Y_i; q, u)|D_i = d\right] \leqslant 1. \tag{16}$$

Define

$$\mathcal{H}_{L,\delta}(u) = \{\varphi_{L,\delta}(\cdot; q, u)b(\cdot)/n^{1/4} : (q, b) \in J_d(\tau_U, \tau_L) \times \mathcal{B}\}, \text{ and} \tag{17}$$

$$\mathcal{H}_{U,\delta}(u) = \{\varphi_{U,\delta}(\cdot; q, u)b(\cdot)/n^{1/4} : (q, b) \in J_d(\tau_U, \tau_L) \times \mathcal{B}\}. \tag{18}$$

From (14) and using the fact that $n \geqslant 1$, $J_d(\tau_U, \tau_L)$ is bounded, and $\varphi_{L,\delta}(\cdot; q, u)$, $\varphi_{U,\delta}(\cdot; q, u)$ and $b(\cdot)$ are bounded maps, we find that

$$\begin{aligned} N_{[]}(\varepsilon, \mathcal{H}_{L,\delta}(u), \|\cdot\|_{P,2}) &\leqslant C(\varepsilon\delta)^{-1} \times N_{[]}(C\varepsilon, \mathcal{B}, \|\cdot\|_{P,2}), \text{ and} \\ N_{[]}(\varepsilon, \mathcal{H}_{U,\delta}(u), \|\cdot\|_{P,2}) &\leqslant C(\varepsilon\delta)^{-1} \times N_{[]}(C\varepsilon, \mathcal{B}, \|\cdot\|_{P,2}), \end{aligned} \tag{19}$$

for all $\varepsilon > 0$, for some constant $C > 0$. We take $\delta = \varepsilon^2$ and ε^2-brackets $[h_{L,a,j}, h_{L,b,j}]_{j=1}^N$ and $[h_{U,a,j}, h_{U,b,j}]_{j=1}^N$ such that the former set of brackets cover $\mathcal{H}_{L,\varepsilon^2}(u)$ and the latter $\mathcal{H}_{U,\varepsilon^2}(u)$, both with respect to $\|\cdot\|_{P,2}$, and

$$\|h_{U,a,j}\|_\infty + \|h_{U,b,j}\|_\infty + \|h_{L,a,j}\|_\infty + \|h_{L,b,j}\|_\infty \leqslant \frac{C}{n^{1/4}}, \tag{20}$$

for all $j = 1, ..., N$, for some constant $C > 0$. By (13) and (19), we lose no generality by taking brackets so that for each $h \in \mathcal{H}_n(u)$, there exists $j \in \{1, ..., N\}$ such that[2]

$$\min\{h_{U,b,j}, h_{L,a,j}\} \leqslant h \leqslant \max\{h_{L,a,j}, h_{U,b,j}\}, \tag{21}$$

and

$$\log N \leqslant C - C\log\varepsilon + C\log N_{[]}(C\varepsilon^2, \mathcal{B}, \|\cdot\|_{P,2}), \tag{22}$$

for some $C > 0$. We set

$$h_{L,j} = \min\{h_{U,b,j}, h_{L,a,j}\}, \text{ and } h_{U,j} = \max\{h_{U,b,j}, h_{L,a,j}\}. \tag{23}$$

[2]Since b can take negative values, the inequality (13) does not necessarily imply that $h_{L,a,j} \leqslant h \leqslant h_{U,b,j}$.

Therfore, by (20) and (15), for each $k \geqslant 2$,

$$\begin{aligned} E[|h_{L,j}(V_i) - h_{U,j}(V_i)|^k] &\leqslant C(Cn^{-1/4})^{k-2} E\left[(h_{L,a,j}(V_i) - h_{U,b,j}(V_i))^2\right] \\ &\leqslant 2C(Cn^{-1/4})^{k-2} E\left[(h_{L,a,j}(V_i) - h_{L,b,j}(V_i))^2\right] \\ &\quad +4C(Cn^{-1/4})^{k-2} E\left[(h_{L,b,j}(V_i) - h_{U,a,j}(V_i))^2\right] \\ &\quad +4C(Cn^{-1/4})^{k-2} E\left[(h_{U,a,j}(V_i) - h_{U,b,j}(V_i))^2\right] \\ &\leqslant C_1(C_2 n^{-1/4})^{k-2}\left(\varepsilon^4 + \varepsilon^2 + \varepsilon^4\right), \end{aligned} \tag{24}$$

for some constants $C, C_1, C_2 > 0$. The terms ε^4 are due to the choice of ε^2-brackets and the term ε^2 comes from (15) and $\delta = \varepsilon^2$. ■

Define for each $\tau \in [\tau_L, \tau_U]$ and $b \in \mathcal{B}$,

$$U(\tau, b; \delta) = \{(\tau_1, b_1) \in [\tau_L, \tau_U] \times \mathcal{B} : |\tau - \tau_1| + \|b - b_1\|_{P,2} \leqslant \delta\}. \tag{25}$$

Recall the definitions of $a_\tau(Y_i; q)$ and $\Delta(Y_i; q, u)$ in (1) and (2).

Lemma B.2 *Suppose that $\mathcal{B}$ and $\mathcal{H}_n(u)$ are as in Lemma B.1, for some $u \in \mathbb{R}$, and that for $d = 0, 1$, the density f_d of Y_{di} is bounded. Furthermore, assume that there exists $C > 0$ such that for all $\varepsilon > 0$,*

$$\log N_{[]}(\varepsilon, \mathcal{B}, \|\cdot\|_{P,2}) \leqslant C - C \log \varepsilon. \tag{26}$$

Then the following statements hold.

(i) There exist $s > 0$ and $C > 0$ such that for all $n \geqslant 1$, and for all $\delta > 0$,

$$E\left[\sup_{(\tau_1, b_1) \in U(\tau, b; \delta)} \left|\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(a_{\tau_1}(Y_i; q_d(\tau_1)) b_1(V_i) - E\left[a_{\tau_1}(Y_i; q_d(\tau_1)) b_1(V_i)\right]\right)\right|\right] \leqslant C\delta^s.$$

(ii) There exists $C > 0$ such that for all $n \geqslant 1$,

$$E\left[\sup_{h \in \mathcal{H}_n(u)} \left|\sum_{i=1}^n \left(h(V_i) - E\left[h(V_i)\right]\right)\right|\right] \leqslant C \log n.$$

Proof: (i) Let

$$\mathcal{B}_1 = \{a_\tau(\cdot; q) : (\tau, q) \in [\tau_L, \tau_U] \times \mathbb{R}\}. \tag{27}$$

Certainly, $\mathcal{B}_1$ is a VC class. Both classes $\mathcal{B}_1$ and $\mathcal{B}$ are classes of bounded functions. Hence, if we let $\mathcal{B}_2$ be the collection of functions $f(\cdot)g(\cdot)$ as we run $f \in \mathcal{B}_1$ and $g \in \mathcal{B}$, we have for some constant $C > 0$,

$$\log N_{[]}(\varepsilon, \mathcal{B}_2, \|\cdot\|_{P,2}) \leqslant C - C\log\varepsilon + \log N_{[]}(C\varepsilon, \mathcal{B}, \|\cdot\|_{P,2}). \tag{28}$$

Using (26), we obtain the finite integral bracketing entropy bound for the left hand side of (28). The desired result of (i) follows by the maximal inequality. (For example, see (1) in van der Vaart (1996).)

(ii) When $u = 0$, we have $\Delta(Y_i; q_d(\tau), u) = 0$, a.s. We focus on the case $u \neq 0$. Observe that since b is bounded and $|\varphi_n(Y_i; q, u)| \leqslant 1$, for some constant $C > 0$,

$$E\left[|\varphi_n(Y_i; q, u)b(V_i)|^k\right] \leqslant CE\left[|\varphi_n(Y_i; q, u)|^k\right] \leqslant CE\left[\varphi_n^2(Y_i; q, u)\right], \tag{29}$$

for all $k \geqslant 2$. Observe that

$$\begin{aligned}
E\left[\varphi_n^2(Y_i; q, u)\right] &\leqslant \frac{1}{\sqrt{n}}\int_0^1 P\{q \leqslant Y_i \leqslant q + n^{-1/2}us\}ds && (30)\\
&= \frac{1}{\sqrt{n}}\int_0^1 P\{q \leqslant Y_{1i} \leqslant q + n^{-1/2}us|D_i = 1\}P\{D_i = 1\}ds \\
&\quad + \frac{1}{\sqrt{n}}\int_0^1 P\{q \leqslant Y_{0i} \leqslant q + n^{-1/2}us|D_i = 0\}P\{D_i = 0\}ds \\
&\leqslant \frac{1}{\sqrt{n}}\int_0^1 P\{q \leqslant Y_{1i} \leqslant q + n^{-1/2}us\}ds && (31)\\
&\quad + \frac{1}{\sqrt{n}}\int_0^1 P\{q \leqslant Y_{0i} \leqslant q + n^{-1/2}us\}ds \\
&\leqslant \frac{2}{n}\max_{d=0,1}\sup_{q\in\mathbb{R}} f_d(q)u.
\end{aligned}$$

Using this in combination with Lemma B.1, we apply Theorem 6.8 of Massart (2007) (taking $b = 1$ and $\sigma = Cn^{-1/2}$ there) to obtain that

$$\begin{aligned}
E\left[\sup_{h\in\mathcal{H}_n(u)}\left|\sum_{i=1}^n (h(V_i) - Eh(V_i))\right|\right] &\leqslant C_1 + C_1\sqrt{n}\int_0^{C_1/\sqrt{n}}\sqrt{\log(1/z)}dz + C_1\log n \\
&\leqslant C_2\log n,
\end{aligned}$$

for some constants $C_1, C_2 > 0$ from large n on. Thus we obtain the desired result. ■

Lemma B.3 *For $d = 0, 1$,*

$$\sum_{i=1}^{n} \frac{(\tau - 1\{Y_i \leqslant \hat{q}_d(\tau)\})1\{D_i = d\}}{\hat{p}_d(X_i)} = 0. \tag{32}$$

Proof: Note Knight's identity (see (15) on page 1822 of Kato (2009))

$$\rho_\tau(x - y) - \rho_\tau(x) = -y(\tau - 1\{x \leqslant 0\}) + y \int_0^1 (1\{x \leqslant ys\} - 1\{x \leqslant 0\})ds. \tag{33}$$

Take any $\varepsilon > 0$. By the definition of $\hat{q}_d(\tau)$ as a minimizer of $\hat{Q}_d(q, \tau)$ over q defined in Section 2.2.3,

$$\begin{aligned}
0 &\leqslant \hat{Q}_d(\hat{q}_d(\tau) - \varepsilon, \tau) - \hat{Q}_d(\hat{q}_d(\tau), \tau) && (34)\\
&= -\varepsilon \sum_{i=1}^{n} \frac{(\tau - 1\{Y_i \leqslant \hat{q}_d(\tau)\})1\{D_i = d\}}{\hat{p}_d(X_i)} && (35)\\
&\quad + \varepsilon \sum_{i=1}^{n} \frac{\int_0^1 (1\{Y_i \leqslant \hat{q}_d(\tau) + \varepsilon s\} - 1\{Y_i \leqslant \hat{q}_d(\tau)\})ds1\{D_i = d\}}{\hat{p}_d(X_i)} && (36)
\end{aligned}$$

and

$$\begin{aligned}
0 &\leqslant \hat{Q}_d(\hat{q}_d(\tau) + \varepsilon, \tau) - \hat{Q}_d(\hat{q}_d(\tau), \tau) && (37)\\
&= \varepsilon \sum_{i=1}^{n} \frac{(\tau - 1\{Y_i \leqslant \hat{q}_d(\tau)\})1\{D_i = d\}}{\hat{p}_d(X_i)} && (38)\\
&\quad - \varepsilon \sum_{i=1}^{n} \frac{\int_0^1 (1\{Y_i \leqslant \hat{q}_d(\tau) - \varepsilon s\} - 1\{Y_i \leqslant \hat{q}_d(\tau)\})ds1\{D_i = d\}}{\hat{p}_d(X_i)}. && (39)
\end{aligned}$$

Hence

$$\begin{aligned}
&\sum_{i=1}^{n} \frac{\int_0^1 (1\{Y_i \leqslant \hat{q}_d(\tau) - \varepsilon s\} - 1\{Y_i \leqslant \hat{q}_d(\tau)\})ds1\{D_i = d\}}{\hat{p}_d(X_i)} && (40)\\
&\leqslant \sum_{i=1}^{n} \frac{(\tau - 1\{Y_i \leqslant \hat{q}_d(\tau)\})1\{D_i = d\}}{\hat{p}_d(X_i)} && (41)\\
&\leqslant \sum_{i=1}^{n} \frac{\int_0^1 (1\{Y_i \leqslant \hat{q}_d(\tau) + \varepsilon s\} - 1\{Y_i \leqslant \hat{q}_d(\tau)\})ds1\{D_i = d\}}{\hat{p}_d(X_i)}. && (42)
\end{aligned}$$

By sending $\varepsilon \to 0$, we obtain the desired result. ■

Theorem B.1 *Suppose that Assumptions 2.2 and 2.3 in the main text hold, and let*

$$\zeta_i = \psi(V_i) - E\psi(V_i), \text{ and } \zeta_i^* = \psi(V_i^*) - E[\psi(V_i^*)|\mathcal{F}_n]. \tag{43}$$

Then the following statements hold.

(i)

$$\begin{aligned}
&\sqrt{n}(\hat{q}_d(\tau) - q_d(\tau)) \\
&= \frac{1}{\sqrt{n} f_d(q_d(\tau))} \sum_{i=1}^{n} \frac{a_\tau(Y_i; q_d(\tau)) 1\{D_i = d\}}{p_d(X_i)} \\
&+ \frac{1}{\sqrt{n} f_d(q_d(\tau))} \sum_{j=1}^{n} E\left[\frac{a_\tau(Y_i; q_d(\tau)) g_d(X_i; \beta_0)' 1\{D_i = d\}}{p_d^2(X_i)}\right] \zeta_j + o_P(1),
\end{aligned}$$

uniformly over $\tau \in [\tau_L, \tau_U]$.

(ii)

$$\begin{aligned}
&\sqrt{n}(\hat{q}_d^*(\tau) - \hat{q}_d(\tau)) \\
&= \frac{1}{\sqrt{n} f_d(q_d(\tau))} \sum_{j=1}^{n} \left(\frac{a_\tau(Y_j^*; q_d(\tau)) 1\{D_j = d\}}{p_d(X_j^*)} - \frac{1}{n} \sum_{i=1}^{n} \frac{a_\tau(Y_i; q_d(\tau)) 1\{D_i = d\}}{p_d(X_i)}\right) \\
&+ \frac{1}{\sqrt{n} f_d(q_d(\tau))} \sum_{j=1}^{n} E\left[\frac{a_\tau(Y_i; q_d(\tau)) g_d(X_i; \beta_0)' 1\{D_i = d\}}{p_d^2(X_i)}\right] \zeta_j^* + o_P(1),
\end{aligned}$$

uniformly over $\tau \in [\tau_L, \tau_U]$.

Proof: (i) Note that by the definition of $\hat{q}_d(\tau)$,

$$\begin{aligned}
\sqrt{n}(\hat{q}_d(\tau) - q_d(\tau)) &= \arg\min_{u \in \mathbb{R}} \hat{Q}_d(q_d(\tau) + n^{-1/2}u; \tau) \\
&= \arg\min_{u \in \mathbb{R}} \left(\hat{Q}_d(q_d(\tau) + n^{-1/2}u; \tau) - \hat{Q}_d(q_d(\tau); \tau)\right).
\end{aligned}$$

Recall the definitions of $\hat{Q}_d(q; \tau)$ and $Q_d(q; \tau)$ in Section 2.2.3 of the main text. We write

$$\hat{Q}_d(q_d(\tau) + n^{-1/2}u; \tau) - \hat{Q}_d(q_d(\tau); \tau) = A_n + B_n, \tag{44}$$

where

$$
\begin{aligned}
A_n &= \hat{Q}_d(q_d(\tau)+n^{-1/2}u;\tau)-\hat{Q}_d(q_d(\tau);\tau) && (45)\\
&\quad -\left(Q_d(q_d(\tau)+n^{-1/2}u;\tau)-Q_d(q_d(\tau);\tau)\right), \text{ and} && (46)\\
B_n &= Q_d(q_d(\tau)+n^{-1/2}u;\tau)-Q_d(q_d(\tau);\tau).
\end{aligned}
$$

We can follow the same arguments as in the proof of Theorem 3 of Kato (2009), and show that

$$
B_n=-\frac{u}{\sqrt{n}}\sum_{i=1}^{n}\frac{a_\tau(Y_i;q_d(\tau))1\{D_i=d\}}{p_d(X_i)}+\frac{u^2}{2}f_d(q_d(\tau))+o_P(1), \tag{47}
$$

uniformly over $\tau\in[\tau_L,\tau_U]$. Let us focus on A_n. Using Knight's identity in (33), we write A_n as

$$
uZ_{n,d}^{(1)}(\tau)+Z_{n,d}^{(2)}(u,\tau), \tag{48}
$$

where

$$
\begin{aligned}
Z_{n,d}^{(1)}(\tau) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{a_\tau(Y_i;q_d(\tau))1\{D_i=d\}}{p_d(X_i)}\left(\frac{p_d(X_i)}{\hat{p}_d(X_i)}-1\right), \text{ and} && (49)\\
Z_{n,d}^{(2)}(u,\tau) &= -\frac{u}{n}\sum_{i=1}^{n}\left(\frac{p_d(X_i)}{\hat{p}_d(X_i)}-1\right)\frac{\Delta(Y_i;q_d(\tau),u)1\{D_i=d\}}{p_d(X_i)}. && (50)
\end{aligned}
$$

We write

$$
Z_{n,d}^{(1)}(\tau)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{a_\tau(Y_i;q_d(\tau))1\{D_i=d\}}{p_d(X_i)}\frac{p_d(X_i)-\hat{p}_d(X_i)}{p_d(X_i)}+R_n(\tau), \tag{51}
$$

where

$$
R_n(\tau)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{a_\tau(Y_i;q_d(\tau))1\{D_i=d\}}{p_d(X_i)}\frac{(p_d(X_i)-\hat{p}_d(X_i))^2}{p_d(X_i)\hat{p}_d(X_i)}. \tag{52}
$$

By expanding $G(x;\hat{\beta})$ around β_0 and using Assumption 2.3 (i), it is not hard to see that

$$
\hat{p}_d(x)-p_d(x)=g_d(x;\beta_0)'(\hat{\beta}-\beta_0)+O_P(n^{-1}), \tag{53}
$$

uniformly over $x \in \mathcal{X}$. Since $|a_\tau(Y_i; q_d(\tau))/p_d(X_i)| \leqslant \varepsilon^{-1}$ for all $\tau \in [\tau_L, \tau_U]$ by Assumption 2.1(ii), we find that

$$\sup_{\tau \in [\tau_L, \tau_U]} |R_n(\tau)| = O_P(n^{-1/2}). \tag{54}$$

Applying this and the expansion in (53) to the leading term on the right hand side of (51), we obtain that

$$\begin{aligned} Z_{n,d}^{(1)}(\tau) &= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{a_\tau(Y_i; q_d(\tau)) 1\{D_i = d\}}{p_d(X_i)} \frac{g_d(X_i; \beta_0)'(\hat{\beta} - \beta_0)}{p_d(X_i)} + o_P(1) \\ &= -\left(\frac{1}{n} \sum_{i=1}^{n} \frac{a_\tau(Y_i; q_d(\tau)) 1\{D_i = d\}}{p_d(X_i)} \frac{g_d(X_i; \beta_0)'}{p_d(X_i)} \right) \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_i + o_P(1) \right) + o_P(1) \\ &= -E\left[\frac{a_\tau(Y_i; q_d(\tau)) 1\{D_i = d\}}{p_d(X_i)} \frac{g_d(X_i; \beta_0)'}{p_d(X_i)} \right] \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_i + o_P(1), \end{aligned}$$

by Lemma B.2(i). Using the same arguments, we also obtain that

$$Z_{n,d}^{(2)}(u, \tau) = -u \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\Delta(Y_i; q_d(\tau), u) 1\{D_i = d\}}{p_d(X_i)} \frac{g_d(X_i; \beta_0)'}{p_d(X_i)} \right) \frac{1}{n} \sum_{i=1}^{n} \zeta_i + o_P(1) = o_P(1).$$

We define $b_0(V_i) = 1\{D_i = d\} g_{d,k}(X_i; \beta_0)/p_d^2(X_i)$, where $g_{d,k}(X_i; \beta_0)$ is the k-th entry of $g_d(X_i; \beta_0)$, and take $\mathcal{B} = \{b_0\}$, i.e., the singleton of b_0 in the definition of $\mathcal{H}_n(u)$ in (4). We bound

$$\begin{aligned} \sup_{\tau \in [\tau_L, \tau_U]} & \left| \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta(Y_i; q_d(\tau), u) 1\{D_i = d\}}{p_d(X_i)} \frac{g_{d,k}(X_i; \beta_0)}{p_d(X_i)} \right. \\ & \qquad \left. - E\left[\frac{\Delta(Y_i; q_d(\tau), u) 1\{D_i = d\}}{p_d(X_i)} \frac{g_{d,k}(X_i; \beta_0)}{p_d(X_i)} \right] \right| \\ \leqslant & \; n^{-1/4} \sup_{h \in \mathcal{H}_n(u)} \left| \sum_{i=1}^{n} (h(V_i) - E h(V_i)) \right|. \end{aligned}$$

By Lemma B.2(ii), we find that uniformly over $\tau \in [\tau_L, \tau_U]$,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta(Y_i; q_d(\tau), u) 1\{D_i = d\}}{p_d(X_i)} \frac{g_{d,k}(X_i; \beta_0)}{p_d(X_i)} &= E\left[\frac{\Delta(Y_i; q_d(\tau), u) 1\{D_i = d\}}{p_d(X_i)} \frac{g_{d,k}(X_i; \beta_0)}{p_d(X_i)} \right] \\ & \quad + O(n^{-1/4} \log n). \end{aligned}$$

Since $E[\zeta_i] = 0$, we find that

$$Z_{n,d}^{(2)}(u,\tau) = o_P(1), \text{ uniformly over } \tau \in [\tau_L, \tau_U]. \tag{55}$$

Therefore, we conclude that

$$A_n = -E\left[\frac{a_\tau(Y_i; q_d(\tau))1\{D_i = d\}}{p_d(X_i)} \frac{g_d(X_i;\beta_0)'}{p_d(X_i)}\right] \frac{1}{\sqrt{n}} \sum_{i=1}^n \zeta_i + o_P(1). \tag{56}$$

Combining this with (47), and applying Theorem 2 of Kato (2009), we obtain the desired result of (i).

(ii) The proof of the bootstrap version is similar to that of (i). First, we write

$$\sqrt{n}(\hat{q}_d^*(\tau) - \hat{q}_d(\tau)) = \arg\min_{u \in \mathbb{R}} \hat{Q}_d^*(\hat{q}_d(\tau) + n^{-1/2}u; \tau) - \hat{Q}_d^*(\hat{q}_d(\tau); \tau), \tag{57}$$

where

$$\hat{Q}_d^*(q;\tau) = \sum_{i=1}^n \frac{1\{D_i^* = d\}}{\hat{p}_d^*(X_i^*)} \rho_\tau(Y_i^* - q), \text{ and} \tag{58}$$

$$Q_d^*(q;\tau) = \sum_{i=1}^n \frac{1\{D_i^* = d\}}{\hat{p}_d(X_i^*)} \rho_\tau(Y_i^* - q). \tag{59}$$

Similarly as before, we write

$$\hat{Q}_d^*(\hat{q}_d(\tau) + n^{-1/2}u; \tau) - Q_d^*(\hat{q}_d(\tau); \tau) = A_n^* + B_n^*, \tag{60}$$

where

$$A_n^* \equiv \hat{Q}_d^*(\hat{q}_d(\tau) + n^{-1/2}u; \tau) - \hat{Q}_d^*(\hat{q}_d(\tau); \tau) \tag{61}$$

$$-(Q_d^*(\hat{q}_d(\tau) + n^{-1/2}u; \tau) - Q_d^*(\hat{q}_d(\tau), \text{ and} \tag{62}$$

$$B_n^* \equiv Q_d^*(\hat{q}_d(\tau) + n^{-1/2}u; \tau) - Q_d^*(\hat{q}_d(\tau); \tau).$$

Following the similar arguments as before, we obtain that

$$A_n^* = -uE\left[\frac{a_\tau(Y_i^*; \hat{q}_d(\tau))1\{D_i^* = d\}}{\hat{p}_d(X_i^*)} \frac{g_d(X_i^*;\hat{\beta})'}{\hat{p}_d(X_i^*)} |\mathcal{F}_n\right] \frac{1}{\sqrt{n}} \sum_{i=1}^n \zeta_i^* + o_P(1). \tag{63}$$

Note that from (i) of this theorem and Assumption 2.2(i),

$$\sup_{\tau\in[\tau_L,\tau_U]} |\hat{q}_d(\tau) - q_d(\tau)| = o_P(1), \text{ and } \hat{\beta} = \beta_0 + o_P(1). \tag{64}$$

Hence using Lemma B.2(i), we obtain that

$$\begin{aligned} & E\left[\frac{a_\tau(Y_i^*;\hat{q}_d(\tau))1\{D_i^*=d\}}{\hat{p}_d(X_i^*)}\frac{g_d(X_i^*;\hat{\beta})'}{\hat{p}_d(X_i^*)}|\mathcal{F}_n\right] \\ = \;& \frac{1}{n}\sum_{i=1}^n \frac{a_\tau(Y_i;\hat{q}_d(\tau))1\{D_i=d\}}{\hat{p}_d(X_i)}\frac{g_d(X_i;\hat{\beta})'}{\hat{p}_d(X_i)} \\ = \;& \frac{1}{n}\sum_{i=1}^n \frac{a_\tau(Y_i;q_d(\tau))1\{D_i=d\}}{p_d(X_i)}\frac{g_d(X_i;\beta_0)'}{p_d(X_i)} + o_P(1) \\ = \;& E\left[\frac{a_\tau(Y_i;q_d(\tau))1\{D_i=d\}}{p_d(X_i)}\frac{g_d(X_i;\beta_0)'}{p_d(X_i)}\right] + o_P(1). \end{aligned} \tag{65}$$

Let us turn to B_n^* defined in (61). Using Knight's identity, we write B_n^* as

$$uZ_{n,d}^{*(1)}(\tau) + Z_{n,d}^{*(2)}(u,\tau), \tag{66}$$

where

$$Z_{n,d}^{*(1)}(\tau) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{a_\tau(Y_i^*;\hat{q}_d(\tau))1\{D_i^*=d\}}{\hat{p}_d(X_i^*)}, \text{ and} \tag{67}$$

$$Z_{n,d}^{*(2)}(u,\tau) = -\frac{u}{n}\sum_{i=1}^n \frac{\hat{\Delta}(Y_i^*;\hat{q}_d(\tau),u)1\{D_i^*=d\}}{\hat{p}_d(X_i^*)}, \tag{68}$$

with

$$\hat{\Delta}(Y_i^*;\hat{q}_d(\tau),u) = \sqrt{n}\int_0^1 \left(a_\tau(Y_i^*;\hat{q}_d(\tau)+n^{-1/2}us) - a_\tau(Y_i^*;\hat{q}_d(\tau))\right) ds. \tag{69}$$

Let us first consider $Z_{n,d}^{*(2)}(u,\tau)$. Recall the notation $V_i = (Y_i, X_i', D_i)$ and $V_i^* = (Y_i^*, X_i^{*\prime}, D_i^*)$. We write $Z_{n,d}^{*(2)}(u,\tau)$ as

$$
\begin{aligned}
&-\frac{u}{n}\sum_{i=1}^{n}\left(\tilde{\eta}(V_i^*;\hat{q}_d(\tau),\hat{\beta}) - \frac{1}{n}\sum_{i=1}^{n}\tilde{\eta}(V_i;\hat{q}_d(\tau),\hat{\beta})\right) && (70)\\
&-\frac{u}{n}\sum_{i=1}^{n}\left(\tilde{\eta}(V_i;\hat{q}_d(\tau),\hat{\beta}) - \int \tilde{\eta}(v;\hat{q}_d(\tau),\hat{\beta})dF_V(v)\right)\\
&-u\left(\int \tilde{\eta}(v;\hat{q}_d(\tau),\hat{\beta})dF_V(v) - \int \tilde{\eta}(v;q_d(\tau),\beta_0)dF_V(v)\right)\\
&-uE\left[\frac{\Delta(Y_i;q_d(\tau))1\{D_i = d\}}{p_d(X_i)}\right], && (71)
\end{aligned}
$$

where F_V is the CDF of V_i, and

$$
\tilde{\eta}(V_i^*;\hat{q}_d(\tau),\hat{\beta}) = \frac{\Delta(Y_i^*;\hat{q}_d(\tau))1\{D_i^* = d\}}{\hat{p}_d(X_i^*)}. \tag{72}
$$

We show that the first two terms in (70) are $o_P(1)$ uniformly over $\tau \in [\tau_L, \tau_U]$. We will deal with the first term in (70). By Assumptions 2.1(ii) and 2.3(i) in the main text, we can find $\varepsilon > 0$ such that $G_d(x;\beta) > 0$ for all $x \in \mathcal{X}$ and all $\beta \in B(\beta_0;\varepsilon)$, where $B(\beta_0;\varepsilon) = \{\beta \in \Theta : \|\beta - \beta_0\| \leqslant \varepsilon\}$. Define

$$
b_\beta(V_i) = \frac{1\{D_i = d\}}{G_d(X_i;\beta)}, \tag{73}
$$

and let

$$
\mathcal{B} = \{b_\beta : \beta \in B(\beta_0;\varepsilon)\}, \tag{74}
$$

and define $\mathcal{H}_n(u)$ as in (4) using this $\mathcal{B}$. Since the set $B(\beta_0;\varepsilon)$ is bounded in $\mathbb{R}^{d_\beta}$, by Assumptions 2.1(ii) and 2.3(i) in the main text, we find that the bracketing condition in (26) is satisfied for this set $\mathcal{B}$. Furthermore, by Assumption 2.2(i) in the main text, we have

$\hat{\beta} \in B(\beta_0; \varepsilon)$ with probability approaching one. Now, observe that

$$\begin{aligned} &\sup_{\tau \in [\tau_L, \tau_U]} \left| \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\Delta(Y_i^*; \hat{q}_d(\tau), u) 1\{D_i^* = d\}}{\hat{p}_d(X_i^*)} \right.\right. \\ &\qquad \left.\left. - E\left[\frac{\Delta(Y_i^*; \hat{q}_d(\tau), u) 1\{D_i^* = d\}}{\hat{p}_d(X_i^*)} | \mathcal{F}_n \right] \right) \right| \\ &\leqslant n^{-1/4} \sup_{h \in \mathcal{H}_n(u)} \left| \sum_{i=1}^{n} (h(V_i^*) - E[h(V_i^*) | \mathcal{F}_n]) \right| = O_P(n^{-1/4} \log n), \end{aligned} \tag{75}$$

by Lemma B.2(ii). Thus the first term in (70) is $o_P(1)$ uniformly over $\tau \in [\tau_L, \tau_U]$. The second term can be dealt with in the same way.

Let us turn to the third term in (70). This term is also $o_P(1)$ because $\hat{q}_d(\tau) = q_d(\tau) + o_P(1)$ and $\hat{\beta} = \beta_0 + o_P(1)$. Following precisely the same argument in the proof of Theorem 3 in Kato (2009) used to deal with B_n in (45), we can show that

$$-uE\left[\frac{\Delta(Y_i; q_d(\tau)) 1\{D_i = d\}}{p_d(X_i)} \right] = \frac{u^2}{2} f_d(q_d(\tau)) + o(1), \tag{76}$$

uniformly over $\tau \in [\tau_L, \tau_U]$.

As for $Z_{n,d}^{*(1)}(\tau)$, we use Lemma B.3, and write

$$Z_{n,d}^{*(1)}(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{a_\tau(Y_i^*; \hat{q}_d(\tau)) 1\{D_i^* = d\}}{G_d(X_i^*; \hat{\beta})} - \frac{1}{n} \sum_{i=1}^{n} \frac{a_\tau(Y_i; \hat{q}_d(\tau)) 1\{D_i = d\}}{G_d(X_i; \hat{\beta})} \right).$$

Using Lemma B.2(i) and similar arguments used to show (75) above, we can show that

$$Z_{n,d}^{*(1)}(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{a_\tau(Y_i^*; q_d(\tau)) 1\{D_i^* = d\}}{G_d(X_i^*; \beta_0)} - \frac{1}{n} \sum_{i=1}^{n} \frac{a_\tau(Y_i; q_d(\tau)) 1\{D_i = d\}}{G_d(X_i; \beta_0)} \right) + o_P(1).$$

Hence collecting the results for $Z_{n,d}^{*(1)}(\tau)$ and $Z_{n,d}^{*(2)}(u, \tau)$, we conclude that

$$\begin{aligned} &\hat{Q}_d^*(\hat{q}_d(\tau) + n^{-1/2}u; \tau) - Q_d^*(\hat{q}_d(\tau); \tau) \\ &= -uE\left[\frac{a_\tau(Y_i^*; \hat{q}_d(\tau)) 1\{D_i^* = d\}}{\hat{p}_d(X_i^*)} \frac{g_d(X_i^*; \hat{\beta})'}{\hat{p}_d(X_i^*)} | \mathcal{F}_n \right] \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \zeta_i^* \\ &- u \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{a_\tau(Y_i^*; q_d(\tau)) 1\{D_i^* = d\}}{G_d(X_i^*; \beta_0)} - \frac{1}{n} \sum_{i=1}^{n} \frac{a_\tau(Y_i; q_d(\tau)) 1\{D_i = d\}}{G_d(X_i; \beta_0)} \right) \\ &+ \frac{u^2}{2} f_d(q_d(\tau)) + o_P(1). \end{aligned}$$

Now the desired result follows from Theorem 2 of Kato (2009) similarly as before. ■

Let us define

$$q^{\Delta}(\tau) = q_1(\tau) - q_0(\tau), \text{ and } \hat{q}^{\Delta}(\tau) = \hat{q}_1(\tau) - \hat{q}_0(\tau). \tag{77}$$

Similarly, we define a bootstrap version $\hat{q}_1^*(\tau) - \hat{q}_0^*(\tau)$. The following theorem gives the weak convergence of the process $\{\sqrt{n}(\hat{q}^{\Delta}(\tau) - q^{\Delta}(\tau)) : \tau \in [\tau_L, \tau_U]\}$. Let $\ell^{\infty}([\tau_L, \tau_U])$ be the collection of bounded and measurable functions on $[\tau_L, \tau_U]$. Let BL_1 be the bounded Lipschitz functionals on $\ell^{\infty}([\tau_L, \tau_U])$ with Lipschitz constant 1, i.e.,

$$\mathrm{BL}_1 = \{h \in \ell^{\infty}([\tau_L, \tau_U]) : |h(\tau_1) - h(\tau_2)| \leqslant |\tau_1 - \tau_2|, \tau_1, \tau_2 \in [\tau_L, \tau_U]\}. \tag{78}$$

For a sequence of stochastic processes $\mathbb{G}_n$ and a process $\mathbb{G}$ on $[\tau_L, \tau_U]$, we write

$$\mathbb{G}_n \rightsquigarrow \mathbb{G}, \text{ in } \ell^{\infty}[\tau_L, \tau_U]), \tag{79}$$

as $n \to \infty$, if

$$\sup_{h \in \mathrm{BL}_1} |E^* [h(\mathbb{G}_n)] - E[h(\mathbb{G})]| \to 0, \tag{80}$$

as $n \to \infty$, where E^* denotes the outer expectation. Let $\mathbb{G}_n^*$ be a stochastic process on $[\tau_L, \tau_U]$ such that for each $\tau \in [\tau_L, \tau_U]$, $\mathbb{G}_n^*(\tau)$ is a measurable map of the bootstrap sample (Y_i^*, X_i^*). Then if for any $\varepsilon > 0$,

$$P^* \left\{ \sup_{h \in \mathrm{BL}_1} |E[h(\mathbb{G}_n^*)|\mathcal{F}_n] - E[h(\mathbb{G})]| > \varepsilon \right\} \to 0, \tag{81}$$

as $n \to \infty$, for some we write $\mathbb{G}_n^* \rightsquigarrow_* \mathbb{G}$ in $\ell^{\infty}([\tau_L, \tau_U])$. Here P^* denotes the outer probability.

Theorem B.2 *Suppose that Assumptions 2.2 and 2.3 in the main text hold. Then the following statements hold.*

(i)

$$\sqrt{n}(\hat{q}^{\Delta} - q^{\Delta}) \rightsquigarrow \mathbb{G}, \text{ in } \ell^{\infty}([\tau_L, \tau_U]). \tag{82}$$

(ii)

$$\sqrt{n}(\hat{q}^{\Delta *} - q^{\Delta *}) \rightsquigarrow_* \mathbb{G}, \ \textit{in} \ \ell^\infty([\tau_L, \tau_U]). \tag{83}$$

Proof: Define

$$\mathcal{G} = \{\xi(\cdot; q, \tau) : (q, \tau) \in J_d(\tau_U, \tau_L) \times [\tau_L, \tau_U]\}, \tag{84}$$

where

$$\xi(V_j; q, \tau) = -\frac{a_\tau(Y_j; q)1\{D_i = d\}}{f_d(q)p_d(X_i)} \tag{85}$$

$$+E\left[\frac{a_\tau(Y_i; q)g_d(X_i; \beta_0)'1\{D_i = d\}}{f_d(q)p_d^2(X_i)}\right]\zeta_j + o_P(1). \tag{86}$$

For (i), it suffices to show that $\mathcal{G}$ is P-Donsker. This also implies (ii) by Theorem 2.2 of Giné (1997, p. 104). Define

$$\nu_n(\xi) = \frac{1}{\sqrt{n}}\sum_{i=1}^n(\xi(V_i) - E\xi(V_i)). \tag{87}$$

The convergence of finite dimensional distributions of $\{\nu_n(\xi) : \xi \in \mathcal{G}\}$ follow by the usual central limit theorem. In order to show that $\mathcal{G}$ is P-Donsker, it suffices to show that $\mathcal{G}$ is totally bounded with respect to a pseudo-norm ρ and $\{\nu_n(\xi) : \xi \in \mathcal{G}\}$ is asymptotically equicontinuous with respect to ρ. We take the norm ρ to be $\|\cdot\|_{P,2}$. The total boundedness follows by the same arguments as in the proof of (i) of Lemma B.2. It remains to show asymptotic equicontinuity of the process ν_n. For this, we write

$$\sqrt{n}(\hat{q}^\Delta(\tau) - q^\Delta(\tau)) = -A_{n,1}(q_d(\tau), \tau) + A_{n,2}(q_d(\tau), \tau) + o_P(1), \tag{88}$$

where

$$A_{n,1}(q, \tau) = \frac{1}{\sqrt{n}f_d(q)}\sum_{j=1}^n \frac{a_\tau(Y_j; q)1\{D_j = d\}}{p_d(X_j)}, \text{ and} \tag{89}$$

$$A_{n,2}(q, \tau) = \frac{1}{\sqrt{n}f_d(q)}\sum_{j=1}^n E\left[\frac{a_\tau(Y_i; q)g_d(X_i; \beta_0)'1\{D_i = d\}}{p_d^2(X_i)}\right]\zeta_j. \tag{90}$$

Stochastic equicontinuity of $\{A_{n,1}(q,\tau) : (q,\tau) \in J_d(\tau_U,\tau_L) \times [\tau_L,\tau_U]\}$ obviously follows from Lemma B.2(i) and the Lipshitz continuity of $1/f_d(q)$ in $q \in J_d(\tau_U,\tau_L)$. It is not hard to show similarly that $A_{n,2}$ is stochastically equicontinuous as well. ■

We are prepared to prove Theorem 2.1 in the main text.

Proof of Theorem 2.1 in the main text: Define for any $W' \subset W$,

$$\hat{c}_{1-\alpha}(W') = \inf\left\{c \in \mathbb{R} : \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left\{T_b^*(W') \leqslant c\right\} \geqslant 1-\alpha\right\}.$$

In light of Theorem 2.1 of Romano and Shaikh (2010) and the fact that the functional $\sup_{w\in W'} \Gamma(\cdot, S_w)$ is increasing in W', it suffices to show that

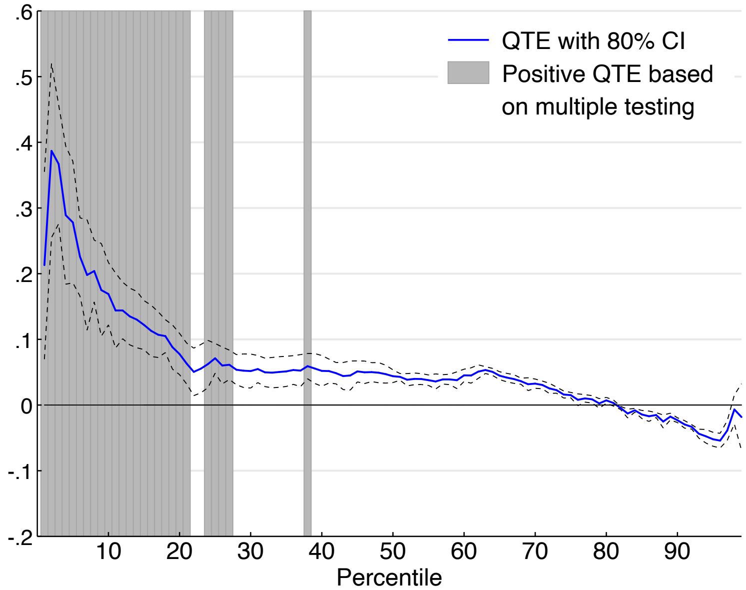
$$\limsup_{n\to\infty} P\left\{\sup_{w\in W_P} \Gamma(\sqrt{n}(\hat{q}^{\Delta} - q^{\Delta}); S_w) \geqslant \hat{c}_{1-\alpha}(W_P)\right\} \leqslant \alpha.$$

However, this follows immediately from Theorem B.2 and the Continuous Mapping Theorem, as $\sup_{w\in W_P} \Gamma(\cdot, S_w)$ is a continuous functional. (See e.g., Theorem 10.8 of Kosorok (2008).) ■

C Robustness checks

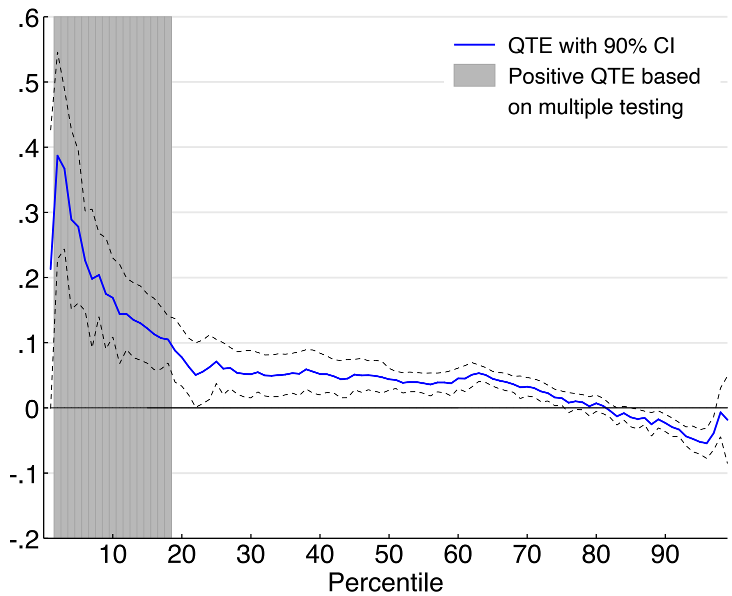
In this section, we present three analyses as robustness checks to our main multiple testing results without subgroups (see Figure 1 in the paper). First, we conduct the step-down procedure with an FWER of 0.1 instead of 0.05. A larger FWER captures the fact that our statistical inference does not adjust for clustering and stratification as in the original paper by Andrabi, Das, and Khwaja (2017). Note clustering leads to smaller p-values, so we counteract this by increasing the FWER. As expected, the results presented in Figure 2 indicate that we reject the null hypothesis of non-positive QTE at more percentiles than in the original result. Specifically, in addition to the percentiles in Figure 1 in the paper, the null hypothesis is violated for the 20th, 21st, 24th to 27th, and 38th percentiles. Hence, the results do not differ in a major way from our main result.

Second, to ensure that type I error rates and power are uniform across quantiles, we use the bootstrap interquartile range rescaling in Algorithm 3 of Chernozhukov, Fernandez-Val, and Melly (2013) before conducting the step-down procedure. The multiple testing results in Figure 3 reveal that, in contrast to our main results in Figure 1 of the paper, we cannot



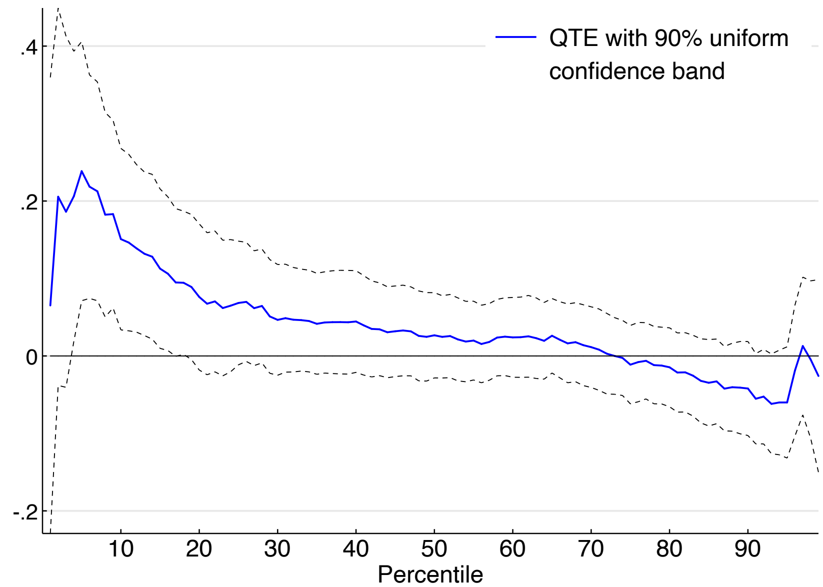
Note: Multiple testing results show quantiles for which the QTE is positive at an FWER of 10 percent (see hypothesis (H.3) in Section 2.3 of the main text).

Figure 2: QTE and multiple testing results with an FWER of 0.1, no subgroups



Note: Multiple testing results show quantiles for which the QTE is positive at an FWER of 5 percent (see hypothesis (H.3) in Section 2.3 of the main text).

Figure 3: QTE and multiple testing results using the bootstrap interquartile range rescaling, no subgroups



Note: QTEs and confidence set estimated following Chernozhukov, Fernandez-Val, and Melly (2013).

Figure 4: QTE and uniform confidence bands, no subgroups

reject the null hypothesis at the 1st and 19th percentiles. Hence, overall, the bootstrap interquartile range rescaling does not change our findings.

Finally, we implement the procedure proposed by Chernozhukov, Fernandez-Val, and Melly (2013), i.e. we estimate conditional QTE and construct uniform confidence bands, as an alternative way to determine for which quantiles the null hypothesis of non-positive QTE is violated. Figure 4 presents the results. While the shape of the estimated QTE is not identical to Figure 1 in the paper because the Chernozhukov, Fernandez-Val, and Melly (2013) procedure is based on conditional QTE, we observe a similar general pattern of treatment effects that decline in magnitude at higher quantiles that also become statistically insignificant. We find that the null hypothesis is violated for the 5th to the 17th and the 19th percentiles, which aligns well with our main result. In summary, these robustness checks reinforce our main findings that ignoring multiple testings leads to substantial inflation in the number of false positive conclusions and that the set of significantly positive QTE at lower percentiles supports the distributional effects predicted by the underlying theory.

References

Abadie, Alberto. 2002. "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models." *Journal of the American Statistical Association* 97 (457):284–292.

Abadie, Alberto, Joshua Angrist, and Guido Imbens. 2002. "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings." *Econometrica* 70 (1):91–117.

Allen, Jason, Robert Clark, and Jean-François Houde. 2014. "The Effect of Mergers in Search Markets: Evidence from the Canadian Mortgage Industry." *American Economic Review* 104 (10):3365–3396.

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja. 2017. "Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets." *American Economic Review* 107 (6):1535–63.

Angrist, Joshua, Daniel Lang, and Philip Oreopoulos. 2009. "Incentives and Services for College Achievement: Evidence from a Randomized Trial." *American Economic Journal: Applied Economics* 1 (1):136–163.

Bandiera, Oriana, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2017. "Labor Markets and Poverty in Village Economies." *Quarterly Journal of Economics* 132 (2):811–870.

Banerjee, Abhijit V., Esther Duflo, Rachel Glennerster, and Cynthia Kinnan. 2015. "The Miracle of Microfinance? Evidence from a Randomized Evaluation." *American Economic Journal: Applied Economics* 7 (1):22–53.

Behaghel, Luc, Clément de Chaisemartin, and Marc Gurgand. 2017. "Ready for Boarding? The Effects of a Boarding School for Disadvantaged Students." *American Economic Journal: Applied Economics* 9 (1):140–164.

Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes. 2006. "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments." *American Economic Review* 96 (4):988–1012.

Brown, Jason, Mark Duggan, Ilyana Kuziemko, and William Woolston. 2014. "How Does Risk Selection Respond to Risk Adjustment? New Evidence from the Medicare Advantage Program." *American Economic Review* 104 (10):3335–3364.

Chernozhukov, Victor, Ivan Fernandez-Val, and Blaise Melly. 2013. "Inference on Counterfactual Distributions." *Econometrica* 81 (6):2205–2268.

Crepon, Bruno, Florencia Devoto, Esther Duflo, and William Pariente. 2015. "Estimating the Impact of Microcredit on Those Who Take It Up: Evidence from a Randomized Experiment in Morocco." *American Economic Journal: Applied Economics* 7 (1):123–150.

Evans, William and Craig Garthwaite. 2012. "Estimating Heterogeneity in the Benefits of Medical Treatment Intensity." *Review of Economics and Statistics* 94 (3):635–649.

Fack, Gabrielle and Camille Landais. 2010. "Are Tax Incentives for Charitable Giving Efficient? Evidence from France." *American Economic Journal: Economic Policy* 2 (2):117–141.

Fairlie, Robert and Jonathan Robinson. 2013. "Experimental Evidence on the Effects of Home Computers on Academic Achievement among Schoolchildren." *American Economic Journal: Applied Economics* 5 (3):211–240.

Fink, Gunther, Margaret McConnell, and Sebastian Vollmer. 2014. "Testing for Heterogeneous Treatment Effects in Experimental Data: False Discovery Risks and Correction Procedures." *Journal of Development Effectiveness* 6 (1):44–57.

Firpo, Sergio. 2007. "Efficient Semiparametric Estimation of Quantile Treatment Effects." *Econometrica* 75 (1):259–276.

Friedlander, Daniel and Philip K. Robins. 1997. "The Distributional Impacts of Social Programs." *Evaluation Review* 21 (5):531–553.

Giné, Evarist. 1997. "Lecture Notes on Some Aspects of the Bootstrap." In *Ecole de Éte de Calcul de Probabilités de Saint-Flour. Lecture Notes in Mathematics*, vol. 1665.

Heckman, James J. 2001. "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture." *Journal of Political Economy* 109 (4):673–748.

Heckman, James J., Jeffrey Smith, and Nancy Clements. 1997. "Making The Most Out Of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts." *Review of Economic Studies* 64 (4):487–535.

Kato, Kengo. 2009. "Asymptotics for Argmin Processes: Convexity Arguments." *Journal of Multivariate Analysis* 100 (8):1816–1829.

Kosorok, Michael. R. 2008. *Introduction to Empirical Processes and Semiparametric Inference*. New York: Springer Verlag.

Lehrer, Steven F., R. Vincent Pohl, and Kyungchul Song. 2021. "Multiple Testing and the Distributional Effects of Accountability Incentives in Education."

Maier, Michael. 2011. "Tests For Distributional Treatment Effects Under Unconfoundedness." *Economics Letters* 110 (1):49–51.

Massart, Pascal. 2007. *Concentration Inequalities and Model Selection*. Berlin, Heidelberg: Springer.

McKenzie, David. 2017. "Identifying and Spurring High-Growth Entrepreneurship: Experimental Evidence from a Business Plan Competition." *American Economic Review* 107 (8):2278–2307.

Meyer, Bruce D and James X Sullivan. 2008. "Changes in the Consumption, Income, and Well-Being of Single Mother Headed Families." *American Economic Review* 98 (5):2221–2241.

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar. 2016. "Building State Capacity: Evidence from Biometric Smartcards in India." *American Economic Review* 106 (10):2895–2929.

Romano, Joseph P. and Azeem M. Shaikh. 2010. "Inference for the Identified Set in Partially Identified Econometric Models." *Econometrica* 78 (1):169–211.

Rothe, Christoph. 2010. "Nonparametric Estimation of Distributional Policy Effects." *Journal of Econometrics* 155 (1):56–70.

van der Vaart, Aad W. 1996. "New Donsker Classes." *Annals of Statistics* 24:2128–2140.

van der Vaart, Aad W. and Jon A. Wellner. 1996. *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.

White, Halbert. 2000. "A Reality Check for Data Snooping." *Econometrica* 68 (5):1097–1126.