Supplement to “Identification and Inference in a Quantile Regression Discontinuity Design under Rank Similarity with Covariates”

Zequn Jin Yu Zhang* Zhengyu Zhang* Yahong Zhou*

June 26, 2023

Abstract

This supplementary material contains the proofs of the lemmas in Appendix A (Section A), additional Monte Carlo simulation results (Section B), and a guide to bandwidth choice procedures in practice (Section C).

Section A: Proofs of Lemmas in Appendix

Lemma A.1. Under the same assumptions as Theorem 2, then for any given $x \in S_X$,

$$\sqrt{nhh_x}\begin{pmatrix} \widehat{m}^+\left(\mathbb{1}\left\{Y<y\right\}D\middle|x\right)-m^+\left(\mathbb{1}\left\{Y<y\right\}D\middle|x\right) \\ \widehat{m}^+\left(\mathbb{1}\left\{Y<y\right\}(1-D)\middle|x\right)-m^+\left(\mathbb{1}\left\{Y<y\right\}(1-D)\middle|x\right) \\ \widehat{m}^-\left(\mathbb{1}\left\{Y<y\right\}D\middle|x\right)-m^-\left(\mathbb{1}\left\{Y<y\right\}D\middle|x\right) \\ \widehat{m}^-\left(\mathbb{1}\left\{Y<y\right\}(1-D)\middle|x\right)-m^-\left(\mathbb{1}\left\{Y<y\right\}(1-D)\middle|x\right) \end{pmatrix} \to^d \begin{pmatrix} \mathbb{Z}_{m_D^+}(y|x) \\ \mathbb{Z}_{m_{1-D}^+}(y|x) \\ \mathbb{Z}_{m_D^-}(y|x) \\ \mathbb{Z}_{m_{1-D}^-}(y|x) \end{pmatrix}$$

where $\mathbb{Z}_{m_D^+}$, $\mathbb{Z}_{m_{1-D}^+}$, $\mathbb{Z}_{m_D^-}$ and $\mathbb{Z}_{m_{1-D}^-}$ are tight zero-mean Gaussian processes in $\ell^\infty\left(S_Y\right)^4$.

Proof. We only prove the first part, and the remaining terms can be similarly verified. Recall that $\nu_i(r_0,x)=\left(1,\left(R_i-r_0\right)/h,\left(R_i-r_0\right)^2/h^2,\left(X_i-x\right)/h_x,\left(X_i-x\right)^2/h_x^2,\left(R_i-r_0\right)\left(X_i-x\right)/hh_x\right)'$ and $m^+\left(\mathbb{1}\{Y<y\}D|x\right)=E[\mathbb{1}\{Y<y\}D|R=r_0^+,X=x]$. Let $\widetilde{m}_1\left(y,r,x\right)=E[\mathbb{1}\{Y<y\}D|R=r,X=x]$ and

$$\alpha_1^+=\left(\widetilde{m}_1,h\widetilde{m}_1^{(1,0)},\frac{h^2}{2}\widetilde{m}_1^{(2,0)},h_x\widetilde{m}_1^{(0,1)},\frac{h_x^2}{2}\widetilde{m}_1^{(0,2)},hh_x\widetilde{m}_1^{(1,1)}\right)'$$

with $\widetilde{m}_1^{(t,s)}=\widetilde{m}_1^{(t,s)}\left(y,r_0^+,x\right)=\lim_{\delta\to0^+}\frac{\partial^t}{\partial r}\frac{\partial^s}{\partial x}\widetilde{m}_1\left(y,r,x\right)\big|_{r=r_0+\delta}$. For each data point $R_i>r_0$, a Taylor expansion gives

$$\begin{aligned}\mathbb{1}\{Y_i\le y\}D_i=&\widetilde{m}_1\left(y,R_i,X_i\right)+\left(\mathbb{1}\{Y_i\le y\}D_i-\widetilde{m}_1\left(y,R_i,X_i\right)\right)=\widetilde{m}_1\left(y,R_i,X_i\right)+\epsilon_i(y)\\ =&\nu_i'(r_0,x)\alpha_1^++\sum_{t=0}^{3}\frac{h^th_x^{3-t}}{6/C_t^3}\left(\frac{R_i-r_0}{h}\right)^t\left(\frac{X_i-x}{h_x}\right)^{3-t}\widetilde{m}_1^{(t,3-t)}\left(y,r_i^*,x_i^*\right)+\epsilon_i(y),\end{aligned}$$

with r_i^*, x_i^* lying between R_i and r_0, X_i and x, respectively. Then by definition,

$$\widehat{m}^+\left(\mathbb{1}\{Y<y\}D|x\right)=m^+\left(\mathbb{1}\{Y<y\}D|x\right)+(a)+(b),$$

where

$$\begin{aligned}(a)=&e_1'\left(\frac{1}{nhh_x}\sum_{i=1}^n\nu_i(r_0,x)\nu_i'(r_0,x)K\left(\frac{X_i-x}{h_x}\right)K\left(\frac{R_i-r_0}{h}\right)\mathbb{1}\{R_i>r_0\}\right)^{-1}\\ &\times\frac{1}{nhh_x}\sum_{i=1}^n\sum_{t=0}^3\frac{h^th_x^{3-t}}{6/C_t^3}\left(\frac{R_i-r_0}{h}\right)^t\left(\frac{X_i-x}{h_x}\right)^{3-t}\widetilde{m}_1^{(t,3-t)}\left(y,r_i^*,x_i^*\right)\nu_i(r_0,x)\\ &\times K\left(\frac{X_i-x}{h_x}\right)K\left(\frac{R_i-r_0}{h}\right)\mathbb{1}\{R_i>r_0\},\end{aligned}$$

and

$$\begin{aligned}(b)=&e_1'\left(\frac{1}{nhh_x}\sum_{i=1}^n\nu_i(r_0,x)\nu_i'(r_0,x)K\left(\frac{X_i-x}{h_x}\right)K\left(\frac{R_i-r_0}{h}\right)\mathbb{1}\{R_i>r_0\}\right)^{-1}\\ &\times\frac{1}{nhh_x}\sum_{i=1}^n\epsilon_i(y)\nu_i(r_0,x)K\left(\frac{X_i-x}{h_x}\right)K\left(\frac{R_i-r_0}{h}\right)\mathbb{1}\{R_i>r_0\}.\end{aligned}$$

We will show that (a) and (b) denote the bias and variance, respectively.

Step 1. Firstly, we consider their common inverse factor. Define

$$\Gamma^+=\int(1\ u\ u^2\ s\ s^2\ us)'\cdot K(u)K(t)\cdot(1\ u\ u^2\ s\ s^2\ us)\mathbb{1}\{u>0\}duds.$$

By the triangle inequality,

$$\begin{aligned}
&\left|\frac{1}{nhh_x}\sum_{i=1}^{n}\nu_i(r_0,x)\nu_i'(r_0,x)K\left(\frac{X_i-x}{h_x}\right)K\left(\frac{R_i-r_0}{h}\right)\mathbb{1}\{R_i>r_0\}-\Gamma^+f_{RX}(r_0,x)\right| \\
\leq&\left|\frac{1}{nhh_x}\sum_{i=1}^{n}\nu_i(r_0,x)\nu_i'(r_0,x)K\left(\frac{X_i-x}{h_x}\right)K\left(\frac{R_i-r_0}{h}\right)\mathbb{1}\{R_i>r_0\}\right. \\
&\left.-E\frac{1}{nhh_x}\sum_{i=1}^{n}\nu_i(r_0,x)\nu_i'(r_0,x)K\left(\frac{X_i-x}{h_x}\right)K\left(\frac{R_i-r_0}{h}\right)\mathbb{1}\{R_i>r_0\}\right| \\
+&\left|E\frac{1}{nhh_x}\sum_{i=1}^{n}\nu_i(r_0,x)\nu_i'(r_0,x)K\left(\frac{X_i-x}{h_x}\right)K\left(\frac{R_i-r_0}{h}\right)\mathbb{1}\{R_i>r_0\}-\Gamma^+f_{RX}(r_0,x)\right|
\end{aligned}$$

uniformly in x. Regarding the deterministic part, by a Taylor expansion,

$$E\frac{1}{nhh_x}\sum_{i=1}^{n}\nu_i(r_0,x)\nu_i'(r_0,x)K\left(\frac{X_i-x}{h_x}\right)K\left(\frac{R_i-r_0}{h}\right)\mathbb{1}\{R_i>r_0\}-\Gamma^+f_{RX}(r_0,x)=O\left(h+h_x^2\right)$$

uniformly in x. For the stochastic part, applying similar arguments as proving Theorem 1.4 in Li and Racine (2007), this term has order $O_p^v\left(\sqrt{\log n/nhh_x}\right)$ uniformly in x. Consequently, we have

$$\frac{1}{nhh_x}\sum_{i=1}^{n}\nu_i(r_0,x)\nu_i'(r_0,x)K\left(\frac{X_i-x}{h_x}\right)K\left(\frac{R_i-r_0}{h}\right)\mathbb{1}\{R_i>r_0\}-\Gamma^+f_{RX}(r_0,x)=O_p^v\left(h+h_x^2+\sqrt{\frac{\log n}{nhh_x}}\right)$$

uniformly in x.

Step 2. For term (a), applying similar arguments in Step 1, we have

$$\begin{aligned}
(a)=&e_1'\left(\Gamma^+f_{RX}(r_0,x)+O_p^v\left(h+h_x+\sqrt{\frac{\log n}{nhh_x}}\right)\right)^{-1} \\
&\times\sum_{t=0}^{3}\left(\frac{h^th_x^{3-t}}{6/C_t^3}\Lambda_t^+f_{RX}(r_0,x)\widetilde{m}_1^{(t,3-t)}\left(y,r_0^+,x\right)+O_p^v\left(h^th_x^{3-t}\left(h+h_x^2\right)+h^th_x^{3-t}\sqrt{\frac{\log n}{nhh_x}}\right)\right)
\end{aligned}$$

uniformly in x, where

$$\Lambda_t^+=\int(1\ u\ u^2\ s\ s^2\ us)'u^ts^{3-t}K(u)K(s)\mathbb{1}\{u>0\}duds.$$

Thus, through the local quadratic estimation, the bias term (a) has order $O_p^v\left(\max\{h^3,h_x^3\}\right)$ uniformly in x, which can be neglected.

Step 3. For term (b), we will show that for any given $x\in S_X$, $\sqrt{nhh_x}(b)$ weakly converges to a tightly Gaussian process in $\ell^\infty\left(S_Y\right)$.[1] By construction, local quadratic regression satisfies $E\left[\epsilon_i(y)|R_i,X_i\right]=E\left[\mathbb{1}\{Y_i\leq y\}D_i-\widetilde{m}_1\left(y,R_i,X_i\right)|R_i,X_i\right]=0$; thus, by the law of iterated expectations, we have $E(b)=0$. Notice that (b) is formed as a uniform Lipschitz transform of sample averages. Let

$$\begin{aligned}
A_{n,t,s}(x)&=\frac{1}{nhh_x}\sum_{i=1}^{n}\left(\frac{R_i-r_0}{h}\right)^t\left(\frac{X_i-x}{h_x}\right)^sK\left(\frac{R_i-r_0}{h}\right)K\left(\frac{X_i-x}{h_x}\right)\mathbb{1}\{R_i>r_0\}, \\
B_{n,t,s}(y,x)&=\frac{1}{nhh_x}\sum_{i=1}^{n}\epsilon_i(y)\left(\frac{R_i-r_0}{h}\right)^t\left(\frac{X_i-x}{h_x}\right)^sK\left(\frac{R_i-r_0}{h}\right)K\left(\frac{X_i-x}{h_x}\right)\mathbb{1}\{R_i>r_0\},
\end{aligned}$$

[1] Throughout the proof of convergence as a process, the argument x is given and the set of functions is only indexed by y. Such a process is usually called a "local" empirical process (Deheuvels and Mason (1994), Einmahl and Mason (1997,1998), Mason (2004), Chernozhukov et al. (2014)). We omit the claim "for any given $x\in S_X$" in the remaining proof.

for $0 \leq t+s \leq 4$ in $A_{n,t,s}$ and $0 \leq t+s \leq 2$ in $B_{n,t,s}(y,x)$. To show the weak convergence of (b), we establish that each of the terms above converges weakly as processes and apply a functional delta method. We start by establishing the convergence as a process of

$$\sqrt{nhh_x}\left(B_{n,t,s}(y,x) - EB_{n,t,s}(y,x)\right) = \sqrt{nhh_x}B_{n,t,s}(y,x), \tag{L.1.1}$$

since the $A_{n,t,s}$ terms are trivial functions of y. Let $Z = (Y, D, R, X)$. Define the set of functions $\mathfrak{F}_n = \{f_{n,y} : y \in S_Y\}$ with[2]

$$f_{n,y}(Z_i) = \frac{1}{\sqrt{hh_x}}\epsilon_i(y)\left(\frac{R_i - r_0}{h}\right)^t\left(\frac{X_i - x}{h_x}\right)^s K\left(\frac{R_i - r_0}{h}\right)K\left(\frac{X_i - x}{h_x}\right)\mathbb{1}\{R_i > r_0\}.$$

Then, the process $(L.1.1)$ can be written as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f_{n,y}(Z_i) : y \in S_Y,$$

which corresponds to van der Vaart and Wellner's (1996) setup for Theorem 2.11.22 for convergence of processes indexed by classes of functions changing with n. Denote E^* as the outer expectation, and $\rho(y_1, y_2)$ as a pseudo-norm on $\mathbb{R}$. The conditions needed for convergence are the following

1. There exist envelope functions $F_n : |f_{n,y}(z)| \leq F_n(z), \forall z, f, n$ which satisfy
   (i) $E^*F_n^2 = O(1)$, and
   (ii) $E^*F_n^2\mathbb{1}\{F_n > \eta\sqrt{n}\} \to 0$ for every $\eta > 0$.
2. $\mathfrak{F}_{n,\delta} = \{f_{n,y_1} - f_{n,y_2} : \rho(y_1, y_2) < \delta\}$ and $\mathfrak{F}^2_{n,\delta}$ are P^{υ}-measurable for every $\delta > 0$.
3. $f_{n,y}$ satisfy
$$\sup_{\rho(y_1,y_2)<\delta_n} E\left[f_{n,y_1} - f_{n,y_2}\right]^2 \to 0 \quad \text{for every} \quad \delta_n \to 0.$$
4. The uniform entropy condition on page 220 of van der Vaart and Wellner (1996) holds.

Start with the first condition. Define a set of envelop functions to be

$$F_n = \left|\frac{1}{\sqrt{hh_x}}\left(\frac{R_i - r_0}{h}\right)^t\left(\frac{X_i - x}{h_x}\right)^s K\left(\frac{R_i - r_0}{h}\right)K\left(\frac{X_i - x}{h_x}\right)\mathbb{1}\{R_i > r_0\}\right|.$$

Condition 1(i) can be written as

$$\begin{aligned} EF_n^2 &= \int\left(\frac{1}{\sqrt{hh_x}}\left(\frac{R - r_0}{h}\right)^t\left(\frac{X - x}{h_x}\right)^s K\left(\frac{R - r_0}{h}\right)K\left(\frac{X - x}{h_x}\right)\right)^2\mathbb{1}\{R > r_0\}f(R,X)dRdX \\ &= \int\left(u^t v^s K(u)K(v)\right)^2\mathbb{1}\{u > 0\}f(r_0 + uh, x + vh_x)dudv, \end{aligned}$$

where $f(r,x)$ denotes the distribution of (R,X) evaluated at (r,x). Condition 1(i) then holds under the boundedness assumptions on R, X and $K(\cdot)$. Again, based on the boundedness assumption on $K(\cdot)$, F_n is bounded by $\frac{1}{\sqrt{hh_x}}\|K\|^2_\infty$ with $\|K\|_\infty = \sup_{z\in S_Z}|K(z)|$. Thus, condition 1(ii) is trivially satisfied. Condition 2 is implied by the assumption that $K(\cdot)$ is measurable.

To show condition 3, note that

$$\begin{aligned} &\sup_{\rho(y_1,y_2)<\delta_n} E\left[f_{n,y_1} - f_{n,y_2}\right]^2 \\ = &\sup_{\rho(y_1,y_2)<\delta_n}\int_{R>r_0}\left(\epsilon(y_1) - \epsilon(y_2)\right)^2\left(\frac{1}{\sqrt{hh_x}}\left(\frac{R - r_0}{h}\right)^t\left(\frac{X - x}{h_x}\right)^s K\left(\frac{R - r_0}{h}\right)K\left(\frac{X - x}{h_x}\right)\right)^2 f(Z)dZ. \end{aligned}$$

[2]The formal definition of $\mathfrak{F}_n$ and $f_{n,y}$ should also depend on both r_0 and x, such as $\mathfrak{F}_{n,r_0,x} = \{f_{n,y,r_0,x} : y \in S_Y\}$. We suppress the dependence on r_0 and x in the remaining proof to ease notation.

In view of condition 1(i), it suffices to show that

$$
\begin{aligned}
&\sup_{\rho(y_1,y_2)<\delta_n} \int_{R>r_0} (\epsilon(y_1)-\epsilon(y_2))^2 f(Y,D|R,X)dYdD \\
=&\sup_{\rho(y_1,y_2)<\delta_n} \int_{R>r_0} \Big(\mathbb{1}\{Y\le y_1\}-\mathbb{1}\{Y\le y_2\}\Big)^2 Df(Y,D|R,X)dYdD \\
-&\sup_{\rho(y_1,y_2)<\delta_n} \left[\int_{R>r_0} \Big(\mathbb{1}\{Y\le y_1\}-\mathbb{1}\{Y\le y_2\}\Big) Df(Y,D|R,X)dYdD\right]^2 \to 0
\end{aligned}
$$

for every $\delta_n \to 0$, where $f(y,d|r,x)$ denotes the distribution of (Y,D) conditional on $(R,X) = (r,x)$ evaluated at (y,d). By the Lipschitz continuous condition, the second term trivially converges to 0 as $\delta_n \to 0$. For the first term, notice that

$$
\begin{aligned}
&\sup_{\rho(y_1,y_2)<\delta_n} \int_{R>r_0} \Big(\mathbb{1}\{Y\le y_1\}-\mathbb{1}\{Y\le y_2\}\Big)^2 Df(Y,D|R,X)dYdD \\
=&\sup_{\rho(y_1,y_2)<\delta_n} \Big(F_{Y|D=1,R>r_0,X}(y_1) - 2F_{Y|D=1,R>r_0,X}(\min\{y_1,y_2\}) + F_{Y|D=1,R>r_0,X}(y_2)\Big) E[D|R>r_0,X] \\
=&\sup_{\rho(y_1,y_2)<\delta_n} \Big(F_{Y|D=1,R>r_0,X}(\max\{y_1,y_2\}) - F_{Y|D=1,R>r_0,X}(\min\{y_1,y_2\})\Big) E[D|R>r_0,X] \to 0
\end{aligned}
$$

as $\delta_n \to 0$ again by the Lipschitz continuous condition.

Finally, by example 2.11.24 on page 221 of van der Vaart and Wellner (1996), condition 4 is satisfied since $\mathfrak{F}_n$ is a VC class with finite VC-index. Notice that every one-point set is shattered, but a two-point set

$$
\{t_1,t_2\} = \left\{ \begin{pmatrix} y_1 \\ r_1 \\ x_1 \end{pmatrix}, \begin{pmatrix} y_2 \\ r_2 \\ x_2 \end{pmatrix} \right\}
$$

with, say, $y_1 < y_2$ is not shattered because the function cannot pick out $\{t_2\}$. Thus, $\mathfrak{F}_n$ is a VC class with finite VC-index of 2, which establishes the weak convergence of $B_{n,t,s}(y,x)$. A similar argument applies to the $A_{n,t,s}(x)$ terms. By the Cramér-Wold device, the terms converge jointly. Finally, by the functional delta method, $\sqrt{nhh_x}(b)$ weakly converges to a tight mean-zero Gaussian process $\mathbb{Z}_{m_D^+}$ with covariance function

$$
\frac{\lambda^+}{f_{RX}(r_0,x)} \mathrm{Cov}\left(\mathbb{1}\{Y\le y\}D, \mathbb{1}\{Y\le \widetilde{y}\}D \,\middle|\, R=r_0^+, X=x\right),
$$

where

$$
\lambda^+ = e_1 \left(\Gamma^+\right)^{-1} \Gamma_2^+ \left(\Gamma^+\right)^{-1} e_1'
$$

and

$$
\Gamma_2^+ = \int (1\ u\ u^2\ s\ s^2\ us)' \cdot K^2(u)K^2(t) \cdot (1\ u\ u^2\ s\ s^2\ us) \mathbb{1}\{u>0\} duds.
$$

Let

$$
W(y) = \begin{pmatrix} \mathbb{1}\{Y\le y\}D \\ \mathbb{1}\{Y\le y\}(1-D) \end{pmatrix}.
$$

Moreover, the covariance function for the Gaussian processes $\left(\mathbb{Z}_{m_D^+}, \mathbb{Z}_{m_{1-D}^+}, \mathbb{Z}_{m_D^-}, \mathbb{Z}_{m_{1-D}^-}\right)$ can be written as

$$
\Sigma^m(y,\widetilde{y},x) = \begin{bmatrix} \sigma^+(y,\widetilde{y},x) & 0 \\ 0 & \sigma^-(y,\widetilde{y},x) \end{bmatrix},
$$

where

$$\sigma^{+}(y, \widetilde{y}, x) = \frac{\lambda^{+}}{f_{RX}(r_0, x)} \text{Cov}\left(W(y), W\left(\widetilde{y}\right) \middle| R = r_0^{+}, X = x \right),$$

and $\sigma^{-}(y, \widetilde{y}, x)$ is the corresponding left limit. ■

Lemma A.2. Under the same assumptions as Theorem 2, then for any given $x \in S_X$,

$$\sqrt{nhh_x} \begin{pmatrix} \widehat{F}_{1|X}(y|x) - F_{1|X}(y|x) \\ \widehat{F}_{0|X}(y|x) - F_{0|X}(y|x) \end{pmatrix} \rightarrow^{d} \begin{pmatrix} \mathbb{Z}_{F_1}(y|x) \\ \mathbb{Z}_{F_0}(y|x) \end{pmatrix}$$

where $\mathbb{Z}_{F_1}$ and $\mathbb{Z}_{F_0}$ are tight zero-mean Gaussian processes in $\ell^{\infty}\left(S_Y\right)^2$.

Proof. We now establish Hadamard or Fréchet differentiability of the closed-form solution for the conditional potential outcome CDFs. First consider $\widehat{\widetilde{F}}_1(y, x)$ and $\widehat{\widetilde{F}}_0(y, x)$. By the functional delta method,

$$\sqrt{nhh_x} \begin{pmatrix} \widehat{\widetilde{F}}_1(y, x) - \widetilde{F}_1(y, x) \\ \widehat{\widetilde{F}}_0(y, x) - \widetilde{F}_0(y, x) \end{pmatrix} \rightarrow^{d} \begin{pmatrix} \mathbb{Z}_{\widetilde{F}_1}(y|x) \\ \mathbb{Z}_{\widetilde{F}_0}(y|x) \end{pmatrix}$$

where $\mathbb{Z}_{\widetilde{F}_1}$ and $\mathbb{Z}_{\widetilde{F}_0}$ are tight zero-mean Gaussian processes which can be written as

$$\mathbb{Z}_{\widetilde{F}_1}(y|x) = \mathbb{Z}_{m_D^{+}}(y|x) - \mathbb{Z}_{m_D^{-}}(y|x) \quad \textit{and} \quad \mathbb{Z}_{\widetilde{F}_0}(y|x) = \mathbb{Z}_{m_{1-D}^{-}}(y|x) - \mathbb{Z}_{m_{1-D}^{+}}(y|x).$$

Since the quantile operator is Hadamard differentiable for absolutely continuous functions, which is assumed in the Assumption 5.1.6 (also see, for instance, Section 2.2.4 in Kosorok (2008) for a definition of the functional delta method and an application to the quantile operator). Then, by the functional delta method,

$$\sqrt{nhh_x} \begin{pmatrix} \widehat{\widetilde{q}}_1(\tau, x) - \widetilde{q}_1(\tau, x) \\ \widehat{\widetilde{q}}_0(\tau, x) - \widetilde{q}_0(\tau, x) \end{pmatrix} \rightarrow^{d} \begin{pmatrix} \mathbb{Z}_{\widetilde{q}_1}(\tau|x) \\ \mathbb{Z}_{\widetilde{q}_0}(\tau|x) \end{pmatrix}$$

where $\mathbb{Z}_{\widetilde{q}_1}$ and $\mathbb{Z}_{\widetilde{q}_0}$ are tight zero-mean Gaussian processes in $\ell^{\infty}\left(\mathcal{T}\right)^2$, in which $\mathcal{T} \subset (0, 1)$ is a compact interval, and

$$\mathbb{Z}_{\widetilde{q}_1}(\tau|x) = -\mathbb{Z}_{\widetilde{F}_1}\left(\widetilde{q}_1(\tau|x)|x\right) / \widetilde{f}_1\left(\widetilde{q}_1(\tau|x), x\right),$$
$$\mathbb{Z}_{\widetilde{q}_0}(\tau|x) = -\mathbb{Z}_{\widetilde{F}_0}\left(\widetilde{q}_0(\tau|x)|x\right) / \widetilde{f}_0\left(\widetilde{q}_0(\tau|x), x\right).$$

Using Lemma 3.9.27 in Van der Vaart and Wellner (1996), or more specifically, Lemma A.1 in Callaway et al. (2018), by the functional delta method, we have

$$\sqrt{nhh_x} \begin{pmatrix} \widehat{\widetilde{q}}_0\left(\widehat{\widetilde{F}}_1(y, x), x\right) - \widetilde{q}_0\left(\widetilde{F}_1(y, x), x\right) \\ \widehat{\widetilde{q}}_1\left(\widehat{\widetilde{F}}_0(y, x), x\right) - \widetilde{q}_1\left(\widetilde{F}_0(y, x), x\right) \end{pmatrix} \rightarrow^{d} \begin{pmatrix} \mathbb{Z}_{\widetilde{q}_0 \circ \widetilde{F}_1}(y|x) \\ \mathbb{Z}_{\widetilde{q}_1 \circ \widetilde{F}_0}(y|x) \end{pmatrix}$$

where $\mathbb{Z}_{\widetilde{q}_0 \circ \widetilde{F}_1}$ and $\mathbb{Z}_{\widetilde{q}_1 \circ \widetilde{F}_0}$ are tight zero-mean Gaussian processes, and

$$\mathbb{Z}_{\widetilde{q}_0 \circ \widetilde{F}_1}(y|x) = \mathbb{Z}_{\widetilde{q}_0}\left(\widetilde{F}_1(y, x)|x\right) + \frac{\mathbb{Z}_{\widetilde{F}_1}(y|x)}{\widetilde{f}_0\left(\widetilde{q}_0\left(\widetilde{F}_1(y, x), x\right), x\right)},$$
$$\mathbb{Z}_{\widetilde{q}_1 \circ \widetilde{F}_0}(y|x) = \mathbb{Z}_{\widetilde{q}_1}\left(\widetilde{F}_0(y, x)|x\right) + \frac{\mathbb{Z}_{\widetilde{F}_0}(y|x)}{\widetilde{f}_1\left(\widetilde{q}_1\left(\widetilde{F}_0(y, x), x\right), x\right)}.$$

Define the map ϕ: $S_Y \times S_X \mapsto (0,1)$, by

$$\phi(y,x) = m^+\left(\mathbb{1}\{Y<y\}(1-D)|x\right) = P\left(Y \le y\middle|D=0, R=r_0^+, X=x\right)\left(1-p\left(r_0^+,x\right)\right).$$

It is straightforward to verify that ϕ is Fréchet-differentiable uniformly over y on S_Y, with the derivative

$$\phi'_y(t,x) = t\cdot f_{Y|DRX}(y,0,r_0^+,x)\left(1-p\left(r_0^+,x\right)\right).$$

Similarly, define the map ψ: $S_Y \times S_X \mapsto (0,1)$, by

$$\psi(y,x) = m^-\left(\mathbb{1}\{Y<y\}D|x\right) = P\left(Y \le y\middle|D=1, R=r_0^-, X=x\right)p\left(r_0^-,x\right).$$

Then, ψ is Fréchet-differentiable uniformly over y on S_Y, with the derivative

$$\psi'_y(t,x) = t\cdot f_{Y|DRX}(y,1,r_0^-,x)p\left(r_0^-,x\right).$$

Using Lemma 3.9.27 in Van der Vaart and Wellner (1996) and the functional delta method gives

$$\sqrt{nhh_x}\begin{pmatrix} \widehat{\phi}\left(\widehat{\widetilde{q}}_0\left(\widehat{\widetilde{F}}_1(y,x),x\right),x\right) - \phi\left(\widetilde{q}_0\left(\widetilde{F}_1(y,x),x\right),x\right) \\ \widehat{\psi}\left(\widehat{\widetilde{q}}_1\left(\widehat{\widetilde{F}}_0(y,x),x\right),x\right) - \psi\left(\widetilde{q}_1\left(\widetilde{F}_0(y,x),x\right),x\right)\end{pmatrix} \to^d \begin{pmatrix} \mathbb{Z}_{\phi\circ\widetilde{q}_0\circ\widetilde{F}_1}(y|x) \\ \mathbb{Z}_{\psi\circ\widetilde{q}_1\circ\widetilde{F}_0}(y|x)\end{pmatrix}$$

where $\mathbb{Z}_{\phi\circ\widetilde{q}_0\circ\widetilde{F}_1}$ and $\mathbb{Z}_{\psi\circ\widetilde{q}_1\circ\widetilde{F}_0}$ are tight zero-mean Gaussian processes, and

$$\mathbb{Z}_{\phi\circ\widetilde{q}_0\circ\widetilde{F}_1}(y|x) = \mathbb{Z}_{m^+_{1-D}}\left(\widetilde{q}_0\left(\widetilde{F}_1(y,x),x\right)|x\right) + \mathbb{Z}_{\widetilde{q}_0\circ\widetilde{F}_1}(y|x)\cdot f_{Y|DRX}\left(\widetilde{q}_0\left(\widetilde{F}_1(y,x)\right),0,r_0^+,x\right)\left(1-p\left(r_0^+,x\right)\right),$$
$$\mathbb{Z}_{\psi\circ\widetilde{q}_1\circ\widetilde{F}_0}(y|x) = \mathbb{Z}_{m^-_{D}}\left(\widetilde{q}_1\left(\widetilde{F}_0(y,x),x\right)|x\right) + \mathbb{Z}_{\widetilde{q}_1\circ\widetilde{F}_0}(y|x)\cdot f_{Y|DRX}\left(\widetilde{q}_1\left(\widetilde{F}_0(y,x),x\right),1,r_0^-,x\right)p\left(r_0^-,x\right).$$

Again by the functional delta method,

$$\sqrt{nhh_x}\begin{pmatrix}\widehat{F}_{1|X}(y|x) - F_{1|X}(y|x) \\ \widehat{F}_{0|X}(y|x) - F_{0|X}(y|x)\end{pmatrix} \to^d \begin{pmatrix}\mathbb{Z}_{F_1}(y|x) \\ \mathbb{Z}_{F_0}(y|x)\end{pmatrix}$$

where $\mathbb{Z}_{F_1}$ and $\mathbb{Z}_{F_0}$ are tight zero-mean Gaussian processes, with

$$\mathbb{Z}_{F_1}(y|x) = \mathbb{Z}_{m_D^+}(y|x) + \mathbb{Z}_{\phi\circ\widetilde{q}_0\circ\widetilde{F}_1}(y|x) \quad and \quad \mathbb{Z}_{F_0}(y|x) = \mathbb{Z}_{\psi\circ\widetilde{q}_1\circ\widetilde{F}_0}(y|x) + \mathbb{Z}_{m^-_{1-D}}(y|x).$$

Combining these results above, we can represent $\mathbb{Z}_{F_1}$ and $\mathbb{Z}_{F_0}$ as a linear combination of $\mathbb{Z}_{m_D^+}$, $\mathbb{Z}_{m^+_{1-D}}$, $\mathbb{Z}_{m_D^-}$ and $\mathbb{Z}_{m^-_{1-D}}$, that is,

$$\begin{aligned}\mathbb{Z}_{F_1}(y|x) =& \left(1+\omega_0^+(y,x)\right)\mathbb{Z}_{m_D^+}(y|x) + \left(1+\omega_0^+(y,x)\right)\mathbb{Z}_{m^+_{1-D}}\left(\widetilde{q}_0\left(\widetilde{F}_1(y,x),x\right)|x\right) \\ & - \omega_0^+(y,x)\,\mathbb{Z}_{m_D^-}(y|x) - \omega_0^+(y,x)\,\mathbb{Z}_{m^-_{1-D}}\left(\widetilde{q}_0\left(\widetilde{F}_1(y,x),x\right)|x\right), \\ \mathbb{Z}_{F_0}(y|x) =& -\omega_1^-(y,x)\,\mathbb{Z}_{m_D^+}\left(\widetilde{q}_1\left(\widetilde{F}_0(y,x),x\right)|x\right) - \omega_1^-(y,x)\,\mathbb{Z}_{m^+_{1-D}}(y|x) \\ & + \left(1+\omega_1^-(y,x)\right)\mathbb{Z}_{m_D^-}\left(\widetilde{q}_1\left(\widetilde{F}_0(y,x),x\right)|x\right) + \left(1+\omega_1^-(y,x)\right)\mathbb{Z}_{m^-_{1-D}}(y|x)\end{aligned}$$

where

$$\begin{aligned}\omega_0^+(y,x) =& f_{Y|DRX}\left(\widetilde{q}_0\left(\widetilde{F}_1(y,x),x\right),0,r_0^+,x\right)\left(1-p\left(r_0^+,x\right)\right)\Big/\widetilde{f}_0\left(\widetilde{q}_0\left(\widetilde{F}_1(y,x),x\right),x\right), \\ \omega_1^-(y,x) =& f_{Y|DRX}\left(\widetilde{q}_1\left(\widetilde{F}_0(y,x),x\right),1,r_0^-,x\right)p\left(r_0^-,x\right)\Big/\widetilde{f}_1\left(\widetilde{q}_1\left(\widetilde{F}_0(y,x),x\right),x\right).\end{aligned}$$

Let

$$W_1 = \begin{pmatrix} \mathbb{1}\{Y \le y\}D \\ \\ \mathbb{1}\left\{Y \le \widetilde{q}_0\left(\widetilde{F}_1(y,x),x\right)\right\}(1-D) \end{pmatrix}, \qquad W_0 = \begin{pmatrix} \mathbb{1}\left\{Y \le \widetilde{q}_1\left(\widetilde{F}_0(y,x),x\right)\right\}D \\ \\ \mathbb{1}\{Y \le y\}(1-D) \end{pmatrix},$$

$$v_1(y,x) = \Big(1+\omega_0^+(y,x), 1+\omega_0^+(y,x), -\omega_0^+(y,x), -\omega_0^+(y,x)\Big),$$

and

$$v_0(y,x) = \Big(-\omega_1^-(y,x), -\omega_1^-(y,x), 1+\omega_1^-(y,x), 1+\omega_1^-(y,x)\Big).$$

Then, the covariance function $\Sigma^{F_x}(y,\widetilde{y},x)$ is, for $j,k \in \{1,2\}$,

$$\Sigma_{jk}^{F_x}(y,\widetilde{y},x) = v_{2-j}(y,x)\sigma_{jk}(y,\widetilde{y},x)v'_{2-k}(\widetilde{y},x),$$

where

$$\sigma_{jk}(y,\widetilde{y},x) = \begin{bmatrix} \dfrac{\lambda^+ \mathrm{Cov}\left(W_{2-j}(y,x), W_{2-k}(\widetilde{y},x) | R = r_0^+, X = x\right)}{f_{RX}(r_0,x)} & 0 \\ 0 & \dfrac{\lambda^- \mathrm{Cov}\left(W_{2-j}(y,x), W_{2-k}(\widetilde{y},x) | R = r_0^-, X = x\right)}{f_{RX}(r_0,x)} \end{bmatrix}.$$

■

Lemma A.3. Assume $h/h_r = \gamma^2$ with $0 < \gamma < \infty$. Under the same assumptions as Theorem 2, then

$$\sqrt{nh_r}\begin{pmatrix} \widehat{F}_1(y) - F_1(y) \\ \widehat{F}_0(y) - F_0(y) \end{pmatrix} \to^d \begin{pmatrix} \mathbb{Z}_1(y) \\ \mathbb{Z}_0(y) \end{pmatrix}$$

where $\mathbb{Z}_1$ and $\mathbb{Z}_0$ are tight zero-mean Gaussian processes in $\ell^\infty(S_Y)^2$.

Proof. We only prove the first part, and the remaining term can be similarly verified. According to Lemma A.2, we have the following asymptotic representation

$$\widehat{F}_{1|X}(y|x) - F_{1|X}(y|x) = v_1(y,x)\widehat{M}(y,x) = \sum_{k=1}^{4} v_{1k}(y,x)\widehat{M}_k(y,x),$$

where

$$\widehat{M}(y,x) = \begin{pmatrix} \widehat{m}^+\left(\mathbb{1}\{Y<y\}D|x\right) - m^+\left(\mathbb{1}\{Y<y\}D|x\right) \\ \widehat{m}^+\left(\mathbb{1}\left\{Y<\widetilde{q}_0\left(\widetilde{F}_1(y,x),x\right)\right\}(1-D)|x\right) - m^+\left(\mathbb{1}\left\{Y<\widetilde{q}_0\left(\widetilde{F}_1(y,x),x\right)\right\}(1-D)|x\right) \\ \widehat{m}^-\left(\mathbb{1}\{Y<y\}D|x\right) - m^-\left(\mathbb{1}\{Y<y\}D|x\right) \\ \widehat{m}^-\left(\mathbb{1}\left\{Y<\widetilde{q}_0\left(\widetilde{F}_1(y,x),x\right)\right\}(1-D)|x\right) - m^-\left(\mathbb{1}\left\{Y<\widetilde{q}_0\left(\widetilde{F}_1(y,x),x\right)\right\}(1-D)|x\right) \end{pmatrix}.$$

Let $\widetilde{e}_1 = (1,0,0)'$. By definition,

$$\begin{aligned}
\widehat{F}_1(y) =& \widetilde{e}_1'\left(\frac{1}{nh_r}\sum_{i=1}^{n}\widetilde{\nu}_i(r_0)\widetilde{\nu}_i'(r_0)K\left(\frac{R_i-r_0}{h_r}\right)\right)^{-1}\frac{1}{nh_r}\sum_{i=1}^{n}\widehat{F}_{1|X}(y|X_i)\widetilde{\nu}_i(r_0)K\left(\frac{R_i-r_0}{h_r}\right) \\
=& \widetilde{e}_1'\left(\frac{1}{nh_r}\sum_{i=1}^{n}\widetilde{\nu}_i(r_0)\widetilde{\nu}_i'(r_0)K\left(\frac{R_i-r_0}{h_r}\right)\right)^{-1}\frac{1}{nh_r}\sum_{i=1}^{n}F_{1|X}(y|X_i)\widetilde{\nu}_i(r_0)K\left(\frac{R_i-r_0}{h_r}\right) \\
&+\widetilde{e}_1'\left(\frac{1}{nh_r}\sum_{i=1}^{n}\widetilde{\nu}_i(r_0)\widetilde{\nu}_i'(r_0)K\left(\frac{R_i-r_0}{h_r}\right)\right)^{-1}\frac{1}{nh_r}\sum_{i=1}^{n}\left(\widehat{F}_{1|X}(y|X_i) - F_{1|X}(y|X_i)\right)\widetilde{\nu}_i(r_0)K\left(\frac{R_i-r_0}{h_r}\right).
\end{aligned}$$

Applying similar arguments as in Lemma A.1,

$$\begin{aligned}
&\sqrt{nh_r}\left(\widehat{F}_1(y)-F_1(y)\right)\\
=&\widetilde{e}_1'\left(\widetilde{\Gamma} f_R(r_0)+O_p^v\left(h_r^2+\sqrt{\frac{\log n}{nh_r}}\right)\right)^{-1}\frac{1}{\sqrt{nh_r}}\sum_{i=1}^n \widetilde{\epsilon}_i(y)\widetilde{\nu}_i(r_0)K\left(\frac{R_i-r_0}{h_r}\right)\\
+&\widetilde{e}_1'\left(\widetilde{\Gamma} f_R(r_0)+O_p^v\left(h_r^2+\sqrt{\frac{\log n}{nh_r}}\right)\right)^{-1}\frac{1}{\sqrt{nh_r}}\sum_{i=1}^n \left(\widehat{F}_{1|X}(y|X_i)-F_{1|X}(y|X_i)\right)\widetilde{\nu}_i(r_0)K\left(\frac{R_i-r_0}{h_r}\right)\\
+&O_p^v\left(\sqrt{nh_r}h_r^3\right),
\end{aligned}$$

where

$$\widetilde{\Gamma}=\int (1\ u\ u^2)'\cdot K(u)\cdot(1\ u\ u^2)du,$$

and $\widetilde{\epsilon}_i(y)=F_{1|X}(y|X_i)-E\left[F_{1|X}(y|X_i)|R=R_i\right]$. Similarly, we can show that

$$\frac{1}{\sqrt{nh_r}}\sum_{i=1}^n \widetilde{\epsilon}_i(y)\widetilde{\nu}_i(r_0)K\left(\frac{R_i-r_0}{h_r}\right)$$

weakly converges to a tight mean-zero Gaussian process with a covariance function

$$f_R(r_0)\widetilde{\Gamma}_2\mathrm{Cov}\left(F_{1|X}(y|X),F_{1|X}(\widetilde{y}|X)\middle|R=r_0\right),$$

where

$$\widetilde{\Gamma}_2=\int (1\ u\ u^2)'\cdot K^2(u)\cdot(1\ u\ u^2)du.$$

For the second term, plugging in the asymptotic representation of $\widehat{F}_{1|X}(y|X_i)-F_{1|X}(y|X_i)$, we have

$$\begin{aligned}
&\frac{1}{\sqrt{nh_r}}\sum_{i=1}^n \left(\widehat{F}_{1|X}(y|X_i)-F_{1|X}(y|X_i)\right)\widetilde{\nu}_i(r_0)K\left(\frac{R_i-r_0}{h_r}\right)\\
=&\sum_{k=1}^4\frac{1}{\sqrt{nh_r}}\sum_{i=1}^n v_{1k}(y,X_i)\widehat{M}_k(y,X_i)\widetilde{\nu}_i(r_0)K\left(\frac{R_i-r_0}{h_r}\right).
\end{aligned}$$

We only prove the term of $k=1$, and the remaining terms can be similarly proved. Notice that

$$\begin{aligned}
\widehat{M}_1(y,x)=&\widehat{m}^+\left(\mathbb{1}\{Y<y\}D|x\right)-m^+\left(\mathbb{1}\{Y<y\}D|x\right)\\
=&\frac{1}{nhh_x}\sum_{j=1}^n e_1'\left(\Gamma^+\right)^{-1}\frac{\epsilon_j(y)}{f_{RX}(r_0,x)}\nu_j(r_0,x)K\left(\frac{X_j-x}{h_x}\right)K\left(\frac{R_j-r_0}{h}\right)\mathbb{1}\{R_j>r_0\}\\
&+O_p^v\left(\max\{h^3,h_x^3\}\right)
\end{aligned}$$

uniformly in x. Thus,

$$
\begin{aligned}
&\frac{1}{\sqrt{nh_r}}\sum_{i=1}^{n}\left(1+\omega_0^+(y,X_i)\right)\widehat{M}_1(y,X_i)\widetilde{\nu}_i(r_0)K\left(\frac{R_i-r_0}{h_r}\right)\\
=&\frac{1}{\sqrt{nh_r}nhh_x}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(1+\omega_0^+(y,X_i)\right)e_1'\left(\Gamma^+\right)^{-1}\nu_j(r_0,X_i)\frac{\epsilon_j(y)}{f_{RX}(r_0,X_i)}\widetilde{\nu}_i(r_0)\\
&\times K\left(\frac{X_j-X_i}{h_x}\right)K\left(\frac{R_j-r_0}{h}\right)\mathbb{1}\{R_j>r_0\}K\left(\frac{R_i-r_0}{h_r}\right)\\
+&O_p^v\left(\sqrt{nh_r}\max\{h^3,h_x^3\}\right)\\
=&\frac{1}{\sqrt{nh_r}nhh_x}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathcal{H}_1\left(Z_i,Z_j,y\right)+O_p^v\left(\sqrt{nh_r}\max\{h^3,h_x^3\}\right)\\
=&\frac{\sqrt{nh_r}}{n(n-1)}\sum_{i\neq j}\frac{1}{h_rhh_x}\mathcal{H}_1\left(Z_i,Z_j,y\right)+o_p^v\left(\frac{\sqrt{nh_r}}{\sqrt{n}}\right)+O_p^v\left(\sqrt{nh_r}\max\{h^3,h_x^3\}\right).
\end{aligned}
$$

According to Powell, Stock and Stoker (1989), the condition sufficient for U-statistic projection is $E\left\|\frac{1}{h_rhh_x}\mathcal{H}_1\left(Z_i,Z_j,y\right)\right\|^2=o(n)$. Notice that

$$
\begin{aligned}
E\left\|\frac{1}{h_rhh_x}\mathcal{H}_1\left(Z_i,Z_j,y\right)\right\|^2=&E\left[\frac{1}{h_r^2h^2h_x^2}\mathcal{H}_1'\left(Z_i,Z_j,y\right)\mathcal{H}_1\left(Z_i,Z_j,y\right)\right]\\
=&E\left\{E\left[\frac{1}{h_r^2h^2h_x^2}\mathcal{H}_1'\left(Z_i,Z_j,y\right)\mathcal{H}_1\left(Z_i,Z_j,y\right)|Z_j\right]\right\}.
\end{aligned}
$$

For the term in the brace,

$$
\begin{aligned}
&E\left[\frac{1}{h_r^2h^2h_x^2}\mathcal{H}_1'\left(Z_i,Z_j,y\right)\mathcal{H}_1\left(Z_i,Z_j,y\right)|Z_j\right]\\
=&E\left[\frac{1}{h_r^2h^2h_x^2}\left(\left(1+\omega_0^+(y,X_i)\right)e_1'\left(\Gamma^+\right)^{-1}\nu_j(r_0,X_i)\right)^2\frac{\widetilde{\nu}_i'(r_0)\widetilde{\nu}_i(r_0)}{f_{RX}^2(r_0,X_i)}K^2\left(\frac{X_j-X_i}{h_x}\right)K^2\left(\frac{R_i-R_0}{h_r}\right)|Z_j\right]\\
&\times\epsilon_j^2(y)K^2\left(\frac{R_j-r_0}{h}\right)\mathbb{1}\{R_j>r_0\}.
\end{aligned}
$$

Applying a change of variables and a series of Taylor expansions,

$$
\begin{aligned}
&E\left[\frac{1}{h_r^2h^2h_x^2}\left(\left(1+\omega_0^+(y,X_i)\right)e_1'\left(\Gamma^+\right)^{-1}\nu_j(r_0,X_i)\right)^2\frac{\widetilde{\nu}_i'(r_0)\widetilde{\nu}_i(r_0)}{f_{RX}^2(r_0,X_i)}K^2\left(\frac{X_j-X_i}{h_x}\right)K^2\left(\frac{R_i-R_0}{h_r}\right)|Z_j\right]\\
=&\frac{1}{h_rh^2h_x}\int\left[\left(\left(1+\omega_0^+(y,X_j)\right)e_1'\left(\Gamma^+\right)^{-1}\bar{\nu}_j(r_0,u)\right)^2\frac{1+t^2+t^4}{f_{RX}(r_0,X_j)}K^2(u)K^2(t)+O\left(h_x^2+h_r^2\right)\right]dudt
\end{aligned}
$$

with $\bar{\nu}_j(r_0,u)=\Big(1,(R_j-r_0)/h,(R_j-r_0)^2/h^2,u,u^2,(R_j-r_0)u/h\Big)'$. Similarly, applying a change of variable and Taylor expansion, the expectation over Z_j results

$$
\begin{aligned}
&E\left\|\frac{1}{h_rhh_x}\mathcal{H}_1\left(Z_i,Z_j,y\right)\right\|^2\\
=&\frac{1}{h_rhh_x}\Bigg\{\lambda^+f_R(r_0)E\left[\int\left(1+\omega_0^+(y,X)\right)^2\frac{1+t^2+t^4}{f_{RX}(r_0,X)}K^2(t)dt\cdot\left(\mathbb{1}\{Y\leq y\}D-\widetilde{m}_1\left(y,r_0^+,X\right)\right)^2\middle|R=r_0^+\right]\\
&+O\left(h+h_x^2+h_r^2\right)\Bigg\}\\
=&O\left(\frac{1}{h_rhh_x}\right).
\end{aligned}
$$

Consequently, the sufficient condition holds if $nh_r hh_x \to \infty$ as $n \to \infty$. Notice that by definition, $E\left[\mathcal{H}_1\left(Z_i, Z_j, y\right)|Z_i\right] = 0$. Thus, the result following the U-statistic projection is given by

$$\frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h_r hh_x} \mathcal{H}_1\left(Z_i, Z_j, y\right) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_r hh_x} E\left[\mathcal{H}_1\left(Z_j, Z_i, y\right)|Z_i\right] + o_p^v\left(\frac{1}{\sqrt{n}}\right).$$

After the calculation of expectations,

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_r hh_x} E\left[\mathcal{H}_1\left(Z_j, Z_i, y\right)|Z_i\right] \\
=&\frac{1}{nh} \sum_{i=1}^{n} \left\{ \epsilon_i(y) K\left(\frac{R_i - r_0}{h}\right) \mathbb{1}\left\{R_i > r_0\right\} \left(1 + \omega_0^+(y, X_i)\right) \int e_1' \left(\Gamma^+\right)^{-1} \bar{\nu}_i(r_0, u) K(u) du \cdot (1, 0, \mu_2)' \right. \\
&\left. + O_p^v(h_x^2 + h_r^2) \right\} \\
=&\frac{1}{nh_r} \sum_{i=1}^{n} \left\{ \frac{\epsilon_i(y)}{\gamma} K\left(\frac{R_i - r_0}{h}\right) \mathbb{1}\left\{R_i > r_0\right\} \left(1 + \omega_0^+(y, X_i)\right) \int e_1' \left(\Gamma^+\right)^{-1} \bar{\nu}_i(r_0, u) K(u) du \cdot (1, 0, \mu_2)' \right. \\
&\left. + O_p^v(h_x^2 + h_r^2) \right\} \\
=&\frac{1}{nh_r} \sum_{i=1}^{n} \left[\mathcal{Q}_{11}(Z_i, y) K\left(\frac{R_i - r_0}{h}\right) \mathbb{1}\left\{R_i > r_0\right\} + O_p^v(h_x^2 + h_r^2) \right]
\end{aligned}$$

with $\mu_2 = \int t^2 K(t) dt$. Applying similar arguments as the proceeding one, we can show that

$$\frac{1}{\sqrt{nh_r}} \sum_{i=1}^{n} \mathcal{Q}_{11}(Z_i, y) K\left(\frac{R_i - r_0}{h}\right) \mathbb{1}\left\{R_i > r_0\right\}$$

weakly converges to a tight mean-zero Gaussian process. As a consequence,

$$\begin{aligned}
&\frac{1}{\sqrt{nh_r}} \sum_{i=1}^{n} \left(1 + \omega_0^+(y, X_i)\right) \widehat{M}_1(y, X_i) \widetilde{\nu}_i(r_0) K\left(\frac{R_i - r_0}{h_r}\right) \\
=&\frac{1}{\sqrt{nh_r}} \sum_{i=1}^{n} \mathcal{Q}_{11}(Z_i, y) K\left(\frac{R_i - r_0}{h}\right) \mathbb{1}\left\{R_i > r_0\right\} + o_p^v(1).
\end{aligned}$$

Similarly, we can show that

$$\begin{aligned}
&\sum_{k=2}^{4} \frac{1}{\sqrt{nh_r}} \sum_{i=1}^{n} v_{1k}(y, X_i) \widehat{M}_k(y, X_i) \widetilde{\nu}_i(r_0) K\left(\frac{R_i - r_0}{h_r}\right) \\
=&\frac{1}{\sqrt{nh_r}} \sum_{i=1}^{n} \left(\mathcal{Q}_{12}(Z_i, y) \mathbb{1}\left\{R_i > r_0\right\} + \mathcal{Q}_{13}(Z_i, y) \mathbb{1}\left\{R_i < r_0\right\} + \mathcal{Q}_{14}(Z_i, y) \mathbb{1}\left\{R_i < r_0\right\}\right) K\left(\frac{R_i - r_0}{h}\right) + o_p^v(1),
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{Q}_{12}(Z_i, y) &= \frac{\varepsilon_i\left(\widetilde{q}_0\left(\widetilde{F}_1(y, X_i), X_i\right)\right)}{\gamma} \left(1 + \omega_0^+(y, X_i)\right) \int e_1' \left(\Gamma^+\right)^{-1} \bar{\nu}_i(r_0, u) K(u) du \cdot (1, 0, \mu_2)', \\
\mathcal{Q}_{13}(Z_i, y) &= -\frac{\epsilon_i(y)}{\gamma} \omega_0^+(y, X_i) \int e_1' \left(\Gamma^-\right)^{-1} \bar{\nu}_i(r_0, u) K(u) du \cdot (1, 0, \mu_2)', \\
\mathcal{Q}_{14}(Z_i, y) &= -\frac{\varepsilon_i\left(\widetilde{q}_0\left(\widetilde{F}_1(y, X_i), X_i\right)\right)}{\gamma} \omega_0^+(y, X_i) \int e_1' \left(\Gamma^-\right)^{-1} \bar{\nu}_i(r_0, u) K(u) du \cdot (1, 0, \mu_2)',
\end{aligned}$$

in which

$$\varepsilon_i(y) = \mathbb{1}\{Y_i \leq y\}(1-D_i) - E[\mathbb{1}\{Y < y\}(1-D)|R = R_i, X = X_i],$$

and each term weakly converges to a tight mean-zero Gaussian process. By the Cramér-Wold device, the terms converge jointly. Combining the results above, we have

$$\begin{aligned}\sqrt{nh_r}\left(\widehat{F}_1(y) - F_1(y)\right) =& \frac{1}{\sqrt{nh_r}}\sum_{i=1}^{n} \widetilde{e}_1'\left(\widetilde{\Gamma} f_R(r_0)\right)^{-1} \widetilde{\epsilon}_i(y)\widetilde{\nu}_i(r_0) K\left(\frac{R_i - r_0}{h_r}\right) \\ &+ \frac{1}{\sqrt{nh_r}}\sum_{i=1}^{n} \widetilde{e}_1'\left(\widetilde{\Gamma} f_R(r_0)\right)^{-1} \Big(\mathcal{Q}_{11}(Z_i,y)\mathbb{1}\{R_i > r_0\} + \mathcal{Q}_{12}(Z_i,y)\mathbb{1}\{R_i > r_0\} \\ &+ \mathcal{Q}_{13}(Z_i,y)\mathbb{1}\{R_i < r_0\} + \mathcal{Q}_{14}(Z_i,y)\mathbb{1}\{R_i < r_0\}\Big) K\left(\frac{R_i - r_0}{h}\right) + o_p^{\upsilon}(1),\end{aligned}$$

which converges to a tight mean-zero Gaussian process. Similarly,

$$\begin{aligned}\sqrt{nh_r}\left(\widehat{F}_0(y) - F_0(y)\right) =& \frac{1}{\sqrt{nh_r}}\sum_{i=1}^{n} \widetilde{e}_1'\left(\widetilde{\Gamma} f_R(r_0)\right)^{-1} \widetilde{\varepsilon}_i(y)\widetilde{\nu}_i(r_0) K\left(\frac{R_i - r_0}{h_r}\right) \\ &+ \frac{1}{\sqrt{nh_r}}\sum_{i=1}^{n} \widetilde{e}_1'\left(\widetilde{\Gamma} f_R(r_0)\right)^{-1} \Big(\mathcal{Q}_{01}(Z_i,y)\mathbb{1}\{R_i > r_0\} + \mathcal{Q}_{02}(Z_i,y)\mathbb{1}\{R_i > r_0\} \\ &+ \mathcal{Q}_{03}(Z_i,y)\mathbb{1}\{R_i < r_0\} + \mathcal{Q}_{04}(Z_i,y)\mathbb{1}\{R_i < r_0\}\Big) K\left(\frac{R_i - r_0}{h}\right) + o_p^{\upsilon}(1),\end{aligned}$$

which converges to a tight mean-zero Gaussian process, where

$$\begin{aligned}\mathcal{Q}_{01}(Z_i,y) &= -\frac{\epsilon_i\left(\widetilde{q}_1\left(\widetilde{F}_0(y,X_i),X_i\right)\right)}{\gamma}\omega_1^-(y,X_i)\int e_1'\left(\Gamma^+\right)^{-1}\bar{\nu}_i(r_0,u)K(u)du\cdot(1,0,\mu_2)', \\ \mathcal{Q}_{02}(Z_i,y) &= -\frac{\varepsilon_i(y)}{\gamma}\omega_1^-(y,X_i)\int e_1'\left(\Gamma^+\right)^{-1}\bar{\nu}_i(r_0,u)K(u)du\cdot(1,0,\mu_2)', \\ \mathcal{Q}_{03}(Z_i,y) &= \frac{\epsilon_i\left(\widetilde{q}_1\left(\widetilde{F}_0(y,X_i),X_i\right)\right)}{\gamma}\left(1+\omega_1^-(y,X_i)\right)\int e_1'\left(\Gamma^-\right)^{-1}\bar{\nu}_i(r_0,u)K(u)du\cdot(1,0,\mu_2)', \\ \mathcal{Q}_{04}(Z_i,y) &= \frac{\varepsilon_i(y)}{\gamma}\left(1+\omega_1^-(y,X_i)\right)\int e_1'\left(\Gamma^-\right)^{-1}\bar{\nu}_i(r_0,u)K(u)du\cdot(1,0,\mu_2)',\end{aligned}$$

and

$$\widetilde{\varepsilon}_i(y) = F_{0|X}(y|X_i) - E\left[F_{0|X}(y|X_i)|R = R_i\right].$$

Thus, we can conclude that

$$\sqrt{nh_r}\begin{pmatrix}\widehat{F}_1(y) - F_1(y) \\ \widehat{F}_0(y) - F_0(y)\end{pmatrix} \to^d \begin{pmatrix}\mathbb{Z}_1(y) \\ \mathbb{Z}_0(y)\end{pmatrix}$$

where $\mathbb{Z}_1$ and $\mathbb{Z}_0$ are tight zero-mean Gaussian processes in $\ell^\infty(S_Y)^2$. Let $c = (1,1,1,1,1)'$ and

$$\begin{aligned}\widetilde{\lambda} &= \widetilde{e}_1'\widetilde{\Gamma}^{-1}\widetilde{\Gamma}_2\widetilde{\Gamma}^{-1}\widetilde{e}_1, \\ \bar{\lambda} &= \widetilde{e}_1'\widetilde{\Gamma}^{-1}(1,0,\mu_2)'\cdot(1,0,\mu_2)\widetilde{\Gamma}^{-1}\widetilde{e}_1, \\ \Delta^+ &= \int\left[\int e_1'\left(\Gamma^+\right)^{-1}\left(1,t,t^2,u,u^2,tu\right)K(u)du\right]^2 K^2(t)\mathbb{1}\{t>0\}dt, \\ \Delta^- &= \int\left[\int e_1'\left(\Gamma^-\right)^{-1}\left(1,t,t^2,u,u^2,tu\right)K(u)du\right]^2 K^2(t)\mathbb{1}\{t<0\}dt.\end{aligned}$$

$$v_1^+(y,x) = \Big(1+\omega_0^+(y,x), 1+\omega_0^+(y,x)\Big), \quad v_1^-(y,x) = \Big(-\omega_0^+(y,x), -\omega_0^+(y,x)\Big),$$

and

$$v_0^+(y,x) = \Big(-\omega_1^-(y,x), -\omega_1^-(y,x)\Big), \quad v_0^-(y,x) = \Big(1+\omega_1^-(y,x), 1+\omega_1^-(y,x)\Big).$$

where

$$\widetilde{\Gamma}_2 = \int (1\ u\ u^2)' \cdot K^2(u) \cdot (1\ u\ u^2) du.$$

For $j,k \in \{0,1\}$, also define

$$\sigma_{jk}(y,\widetilde{y}) = \frac{1}{f_R(r_0)} \begin{bmatrix} \widetilde{\lambda}\mathrm{Cov}\left(F_{2-j|X}(y,X), F_{2-k|X}(\widetilde{y},X)|R=r_0\right) & 0 \\ 0 & \widetilde{\sigma}_{jk}(y,\widetilde{y}) \end{bmatrix},$$

where

$$\widetilde{\sigma}_{jk}(y,\widetilde{y}) = \frac{\bar{\lambda}}{\gamma^2} \begin{bmatrix} \widetilde{\sigma}_{jk}^+(y,\widetilde{y}) & 0 \\ 0 & \widetilde{\sigma}_{jk}^-(y,\widetilde{y}) \end{bmatrix},$$

with

$$\widetilde{\sigma}_{jk}^+(y,\widetilde{y}) = \Delta^+ E\big[v_{2-j}^{+\prime}(y,X)\mathrm{Cov}\left(W_{2-j|X}(y,X), W_{2-k|X}(\widetilde{y},X)|R,X\right) v_{2-k}^+(y,X)|R=r_0^+\big]$$

and

$$\widetilde{\sigma}_{jk}^-(y,\widetilde{y}) = \Delta^- E\big[v_{2-j}^{-\prime}(y,X)\mathrm{Cov}\left(W_{2-j|X}(y,X), W_{2-k|X}(\widetilde{y},X)|R,X\right) v_{2-k}^-(y,X)|R=r_0^-\big].$$

Then, the covariance function $\Sigma^F(y,\widetilde{y})$ is, for $j,k \in \{1,2\}$,

$$\Sigma_{jk}^F(y,\widetilde{y}) = c'\sigma_{jk}(y,\widetilde{y})\, c.$$

■

Lemma A.4. Under the same assumptions as Theorem 3,

$$\widehat{f}_R(r_0) - f_R(r_0) = o_p^v(1).$$

Proof. This lemma directly holds by the standard nonparametric results, for example, Li and Racine (2007). ■

Lemma A.5. Under the same assumptions as Theorem 3, then for $d \in \{0,1\}$,

$$\sup_{(y,r)\in S_Y\times S_R} |\widehat{p}_d(y,r) - p_d(y,r)| = o_p^v(1).$$

Proof. Notice that

$$\begin{aligned}\widehat{p}_d(y,r) - p_d(y,r) =& \frac{1}{nh_p}\sum_{j=1}^n \Big(F_{d|X}(y|X_j) - p(y,r)\Big) K\left(\frac{R_j-r}{h_p}\right) \Big/ \frac{1}{nh_p}\sum_{j=1}^n K\left(\frac{R_j-r}{h_p}\right) \\ &+ \frac{1}{nh_p}\sum_{j=1}^n \Big(\widehat{F}_{d|X}(y|X_j) - F_{d|X}(y|X_j)\Big) K\left(\frac{R_j-r}{h_p}\right) \Big/ \frac{1}{nh_p}\sum_{j=1}^n K\left(\frac{R_j-r}{h_p}\right).\end{aligned}$$

By a slight abuse of notation, let

$$\widehat{f}_R(r_0) = \frac{1}{nh_p}\sum_{j=1}^n K\left(\frac{R_j-r}{h_p}\right).$$

We first consider the first term on the right-hand side. Note that the denominator $\widehat{f}_R(r) = f_R(r) + o_p^{\upsilon}(1)$ and $f_R(r)$ is bounded away from 0 under Assumption 5.2 and 5.8. Thus, it suffices to show that the numerator converges to 0 uniformly over (y, r). Note that

$$\begin{aligned}
&\left|\frac{1}{nh_p}\sum_{j=1}^{n}\Big(F_{d|X}(y|X_j) - p(y,r)\Big)K\left(\frac{R_j - r}{h_p}\right)\right| \\
\leq &\left|\frac{1}{nh_p}\sum_{j=1}^{n}\Big(F_{d|X}(y|X_j) - p(y,r)\Big)K\left(\frac{R_j - r}{h_p}\right) - E\frac{1}{nh_p}\sum_{j=1}^{n}\Big(F_{d|X}(y|X_j) - p(y,r)\Big)K\left(\frac{R_j - r}{h_p}\right)\right| \\
+ &\left|E\frac{1}{nh_p}\sum_{j=1}^{n}\Big(F_{d|X}(y|X_j) - p(y,r)\Big)K\left(\frac{R_j - r}{h_p}\right)\right|.
\end{aligned}$$

Li and Racine (2007) suggests that the stochastic part converges almost surely at rate $\sqrt{\log n/nh_p}$ uniformly over (y, r). For the deterministic part, a Taylor expansion gives

$$E\frac{1}{nh_p}\sum_{j=1}^{n}\Big(F_{d|X}(y|X_j) - p(y,r)\Big)K\left(\frac{R_j - r}{h_p}\right) = O\left(h_p^2\right)$$

uniformly over (y, r).

For the second term, according to Lemma A.2 and standard nonparametric results,

$$\widehat{F}_{d|X}(y|x) - F_{d|X}(y|x) = o_p^{\upsilon}(1)$$

uniformly over (y, x).[3] As a consequence,

$$\begin{aligned}
&\left|\frac{1}{nh_p}\sum_{j=1}^{n}\Big(\widehat{F}_{d|X}(y|X_j) - F_{d|X}(y|X_j)\Big)K\left(\frac{R_j - r}{h_p}\right)\right| \\
\leq &\frac{1}{nh_p}\sum_{j=1}^{n}K\left(\frac{R_j - r}{h_p}\right) \times \sup_{(y,x)\in S_Y\times S_X}\left|\widehat{F}_{d|X}(y|x) - F_{d|X}(y|x)\right| = o_p^{\upsilon}(1),
\end{aligned}$$

which concludes the proof. ■

Lemma A.6. Under the same assumptions as Theorem 3, for $d \in \{0, 1\}$,

$$\sup_{(y,x)\in S_Y\times S_X}\left|\widehat{f}_{YD|RX}\left(y,d,r_0^+,x\right) - f_{YD|RX}\left(y,d,r_0^+,x\right)\right| = o_p^{\upsilon}(1),$$
$$\sup_{(y,x)\in S_Y\times S_X}\left|\widehat{f}_{YD|RX}\left(y,d,r_0^-,x\right) - f_{YD|RX}\left(y,d,r_0^-,x\right)\right| = o_p^{\upsilon}(1)$$

and

$$\sup_{(y,r,x)\in S_Y\times S_R\times S_X}\left|\widehat{\widetilde{p}}(y,d,r,x)\mathbb{1}\{r > r_0\} - \widetilde{p}(y,d,r,x)\mathbb{1}\{r > r_0\}\right| = o_p^{\upsilon}(1),$$
$$\sup_{(y,r,x)\in S_Y\times S_R\times S_X}\left|\widehat{\widetilde{p}}(y,d,r,x)\mathbb{1}\{r < r_0\} - \widetilde{p}(y,d,r,x)\mathbb{1}\{r < r_0\}\right| = o_p^{\upsilon}(1).$$

[3]Notice that $\sqrt{nhh_x}\left(\widehat{F}_{d|X}(y|x) - F_{d|X}(y|x)\right)$ is not asymptotically Gaussian uniformly over both y and x as for the uniform convergence rate of kernel estimator should be adjusted by $\sqrt{\log n}$ (Li and Racine, 2007) or $\sqrt{\log h_x}$ (Einmahl and Mason, 2000). Meanwhile, the uniform consistency still holds by the standard nonparametric results in Li and Racine (2007).

Proof. We will show the first claim only, and the remaining terms can be similarly proved. Recall that

$$\begin{aligned}\widehat{f}_{YD|RX}\left(y,d,r_0^+,x\right)=&\frac{n^{-1}h_{f_x}^{-3}\sum_{i=1}^n K\left(\frac{Y_i-y}{h_{f_x}}\right)K\left(\frac{R_i-r_0}{h_{f_x}}\right)K\left(\frac{X_i-x}{h_{f_x}}\right)\mathbb{1}\{D_i=d\}\mathbb{1}\{R_i>r_0\}}{n^{-1}h_{f_x}^{-2}\sum_{i=1}^n K\left(\frac{R_i-r_0}{h_{f_x}}\right)K\left(\frac{X_i-x}{h_{f_x}}\right)\mathbb{1}\{R_i>r_0\}}\\ =&\widehat{f}_{YDRX}\left(y,d,r_0^+,x\right)\Big/\widehat{f}_{RX}\left(r_0^+,x\right).\end{aligned}$$

Note that the denominator $\widehat{f}_{RX}\left(r_0^+,x\right)=\frac{1}{2}f_{RX}(r_0,x)$ and $f_{RX}(r_0,x)$ is bounded away from 0 under Assumption 5.2 and 5.8. Thus, for $d\in\{0,1\}$, it suffices to show

$$\sup_{(y,x)\in S_Y\times S_X}\left|\widehat{f}_{YDRX}\left(y,d,r_0^+,x\right)-\frac{1}{2}f_{YDRX}\left(y,d,r_0^+,x\right)\right|=o_p^v(1).$$

Note that

$$\begin{aligned}&\left|\widehat{f}_{YDRX}\left(y,d,r_0^+,x\right)-\frac{1}{2}f_{YDRX}\left(y,d,r_0^+,x\right)\right|\\ \le&\left|\widehat{f}_{YDRX}\left(y,d,r_0^+,x\right)-E\widehat{f}_{YDRX}\left(y,d,r_0^+,x\right)\right|+\left|E\widehat{f}_{YDRX}\left(y,d,r_0^+,x\right)-\frac{1}{2}f_{YDRX}\left(y,d,r_0^+,x\right)\right|.\end{aligned}$$

Again, as per Li and Racine (2007), the stochastic part satisfies

$$\sup_{(y,x)\in S_Y\times S_X}\left|\widehat{f}_{YDRX}\left(y,d,r_0^+,x\right)-E\widehat{f}_{YDRX}\left(y,d,r_0^+,x\right)\right|=O_{a.s.}^v\left(\sqrt{\frac{\log n}{nh_{f_x}^3}}\right).$$

For the deterministic part, a Taylor expansion gives

$$\begin{aligned}E\widehat{f}_{YDRX}\left(y,d,r_0^+,x\right)=&\int K(u)K(v)K(t)\mathbb{1}\{v>0\}f_{YDRX}(y+uh_{f_x},d,r_0+vh_{f_x},x+th_{f_x})dudvdt\\ =&\frac{1}{2}f_{YDRX}\left(y,d,r_0^+,x\right)+O(h_{f_x})\end{aligned}$$

uniformly over (y,x). This concludes the proof. ■

Lemma A.7. Under the same assumptions as Theorem 3, then (i)

$$\sup_{(y,x)\in S_Y\times S_X}\left|\widehat{\widetilde{f}}_1\left(y,x\right)-\widetilde{f}_1\left(y,x\right)\right|=o_p^v(1),$$
$$\sup_{(y,x)\in S_Y\times S_X}\left|\widehat{\widetilde{f}}_0\left(y,x\right)-\widetilde{f}_0\left(y,x\right)\right|=o_p^v(1),$$

(ii)

$$\sup_{(y,x)\in S_Y\times S_X}\left|\widehat{f}_{1|X}\left(y|x\right)-f_{1|X}\left(y|x\right)\right|=o_p^v(1),$$
$$\sup_{(y,x)\in S_Y\times S_X}\left|\widehat{f}_{0|X}\left(y|x\right)-f_{0|X}\left(y|x\right)\right|=o_p^v(1),$$

and (iii)

$$\sup_{(y,x)\in S_Y\times S_X}\left|\widehat{\omega}_0^+\left(y,x\right)-\omega_0^+\left(y,x\right)\right|=o_p^v(1),$$
$$\sup_{(y,x)\in S_Y\times S_X}\left|\widehat{\omega}_1^-\left(y,x\right)-\omega_1^-\left(y,x\right)\right|=o_p^v(1).$$

Proof. The result of part (i) directly holds by Lemma A.6. For part (ii), note that

$$\begin{aligned}\left|\widehat{f}_{1|X}\left(y|x\right)-f_{1|X}\left(y|x\right)\right| \leq &\left|\widehat{f}_{YD|RX}\left(y,1,r_0^+,x\right)-f_{YD|RX}\left(y,1,r_0^+,x\right)\right| \\ &+\left|\widehat{f}_{YD|RX}\left(\widehat{\widetilde{q}}_0\left(\widehat{\widetilde{F}}_1(y,x),x\right),0,r_0^+,x\right)-f_{YD|RX}\left(\widehat{\widetilde{q}}_0\left(\widehat{\widetilde{F}}_1(y,x),x\right),0,r_0^+,x\right)\right| \\ &+\left|f_{YD|RX}\left(\widehat{\widetilde{q}}_0\left(\widehat{\widetilde{F}}_1(y,x)\right),0,r_0^+,x\right)-f_{YD|RX}\left(\widetilde{q}_0\left(\widetilde{F}_1(y,x)\right),0,r_0^+,x\right)\right|.\end{aligned}$$

The first and second terms converge to 0 uniformly over (y,x) by Lemma A.5. Similarly, according to the proof of Lemma A.2 and standard nonparametric results,

$$\widehat{\widetilde{q}}_0\left(\widehat{\widetilde{F}}_1(y,x)\right)-\widetilde{q}_0\left(\widetilde{F}_1(y,x)\right)=o_p^{\upsilon}(1)$$

uniformly over (y,x). The continuous mapping theorem and the continuity of the conditional density function imply

$$f_{YD|RX}\left(\widehat{\widetilde{q}}_0\left(\widehat{\widetilde{F}}_1(y,x)\right),0,r_0^+,x\right)-f_{YD|RX}\left(\widetilde{q}_0\left(\widetilde{F}_1(y,x)\right),0,r_0^+,x\right)=o_p^{\upsilon}(1)$$

uniformly over (y,x). The proof of $\widehat{f}_{0|X}\left(y|x\right)$ is similar and omitted here.

Finally, part (iii) can be similarly proved by part (i)-(ii) and the continuous mapping theorem. ■

Lemma A.8. Under the same assumptions as Theorem 3, then for $d\in\{0,1\}$,

$$\sup_{y\in S_Y}\left|\widehat{f}_d\left(y\right)-f_d\left(y\right)\right|=o_p^{\upsilon}(1).$$

Proof. This argument can be easily verified by applying identical steps in proving Lemma A.5 and results in Lemma A.7. ■

Lemma A.9. Define

$$\mathbb{T}\left(t_1(\tau),t_2(\tau)\right)=\lim_{n\to\infty}\left(\sum_{i=1}^{n}E\left|f_{ni}\big(t_1(\tau)\big)-f_{ni}\big(t_2(\tau)\big)\right|^2\right)^{1/2},$$

where

$$f_{ni}\big(y\big)=\frac{1}{\sqrt{nh}}\widetilde{e}_1'\left(\widetilde{\Gamma}f_R(r_0)\right)^{-1}\mathcal{Q}_1(Z_i,y)K\left(\frac{R_i-r_0}{h}\right)$$

or

$$f_{ni}\big(y\big)=\frac{1}{\sqrt{nh}}\widetilde{e}_1'\left(\widetilde{\Gamma}f_R(r_0)\right)^{-1}\mathcal{Q}_0(Z_i,y)K\left(\frac{R_i-r_0}{h}\right).$$

Under the same assumptions as Theorem 3, then

$$\sup_{\tau\in\mathcal{T}}\left|t_1(\tau)-t_2(\tau)\right|\to_{\upsilon}^{p}0$$

implies

$$\sup_{\tau\in\mathcal{T}}\mathbb{T}\left(t_1(\tau),t_2(\tau)\right)\to_{\upsilon}^{p}0.$$

Proof. We will show the claim of f_{ni} for the first term only. The case of the second term can be similarly proved. We omit the dependence of t_1 and t_2 on τ. By the law of iterated expectations and Taylor expansions, we have

$$
\begin{aligned}
&\mathbb{T}^2(t_1,t_2)\\
=&\lim_{n\to\infty}\frac{1}{nh}\sum_{i=1}^{n}E\left\{E\left[\left(\widetilde{e}_1'\left(\widetilde{\Gamma}f_R(r_0)\right)^{-1}\left(\mathcal{Q}_1(Z_i,t_1)-\mathcal{Q}_1(Z_i,t_2)\right)\right)^2\middle|R_i\right]K^2\left(\frac{R_i-r_0}{h}\right)\right\}\\
=&\frac{1}{f_R(r_0)}\left\{\widetilde{\lambda}E\left[\left(\widetilde{\epsilon}_i(t_1)-\widetilde{\epsilon}_i(t_2)\right)^2\middle|R_i=r_0\right]+\bar{\lambda}\Delta^+E\left[\left(\widetilde{\mathcal{Q}}^+(Z_i,t_1)-\widetilde{\mathcal{Q}}^+(Z_i,t_2)\right)^2\middle|R_i=r_0^+\right]\right.\\
&\left.+\bar{\lambda}\Delta^-E\left[\left(\widetilde{\mathcal{Q}}^-(Z_i,t_1)-\widetilde{\mathcal{Q}}^-(Z_i,t_2)\right)^2\middle|R_i=r_0^-\right]\right\}+o_p^{\upsilon}(1)
\end{aligned}
$$

with

$$
\begin{aligned}
\widetilde{\mathcal{Q}}^+(Z_i,t)&=\left(\epsilon_i(t_1)+\varepsilon_i\left(\widetilde{q}_0\left(\widetilde{F}_1(t_1,X_i),X_i\right)\right)\right)\left(1+\omega_0^+(t_1,X_i)\right),\\
\widetilde{\mathcal{Q}}^-(Z_i,t)&=-\left(\epsilon_i(t_1)+\varepsilon_i\left(\widetilde{q}_0\left(\widetilde{F}_1(t_1,X_i),X_i\right)\right)\right)\omega_0^+(t_1,X_i).
\end{aligned}
$$

It then suffices to show that

$$
\begin{aligned}
E\left[\left(\widetilde{\epsilon}_i(t_1)-\widetilde{\epsilon}_i(t_2)\right)^2\middle|R_i=r_0\right]&\to_{\upsilon}^{p}0,\\
E\left[\left(\widetilde{\mathcal{Q}}^+(Z_i,t_1)-\widetilde{\mathcal{Q}}^+(Z_i,t_2)\right)^2\middle|R_i=r_0^+\right]&\to_{\upsilon}^{p}0,\\
E\left[\left(\widetilde{\mathcal{Q}}^-(Z_i,t_1)-\widetilde{\mathcal{Q}}^-(Z_i,t_2)\right)^2\middle|R_i=r_0^-\right]&\to_{\upsilon}^{p}0
\end{aligned}
$$

uniformly over τ as t_1-t_2 converges to 0 uniformly over τ. For the first argument,

$$
\begin{aligned}
&E\left[\left(\widetilde{\epsilon}_i(t_1)-\widetilde{\epsilon}_i(t_2)\right)^2\middle|R_i=r_0\right]\\
=&E\left[\left(F_{1|X}(t_1|X_i)-F_{1|X}(t_2|X_i)\right)^2\middle|R_i=r_0\right]-\left\{E\left[F_{1|X}(t_1|X_i)-F_{1|X}(t_2|X_i)\middle|R_i=r_0\right]\right\}^2\to_{\upsilon}^{p}0
\end{aligned}
$$

uniformly by the uniform consistency of t_1-t_2 and the continuity of conditional potential density function. The uniform convergence in probability of the remaining terms can be concluded similarly. ■

Section B: Additional Simulation Results: Without Monotonicity

In this section, we conduct a small-scale Monte Carlo simulation to evaluate the performance of the proposed estimator when the monotonicity condition is violated (or defiers exist). We consider the following DGP:

$$
\begin{cases}
Y_1=R+\omega U_1,\\
Y_0=R+U_0,\\
D=\mathbb{1}\{R>0\}\cdot\mathbb{1}\{V_1>0\}+\mathbb{1}\{R\le 0\}\cdot\mathbb{1}\{V_0>1\},
\end{cases}
$$

where R is independent of (U_1, U_0, V_1, V_0), and follows the standard normal distribution. By simple calculation, we know that the compliers satisfy $\{V_1 > 0, V_0 \leq 1\}$ and the defiers satisfy $\{V_1 \leq 0, V_0 > 1\}$, thus violating the monotonicity condition. We generate the vector of unobserved errors (U_1, U_0, V_1, V_0) using

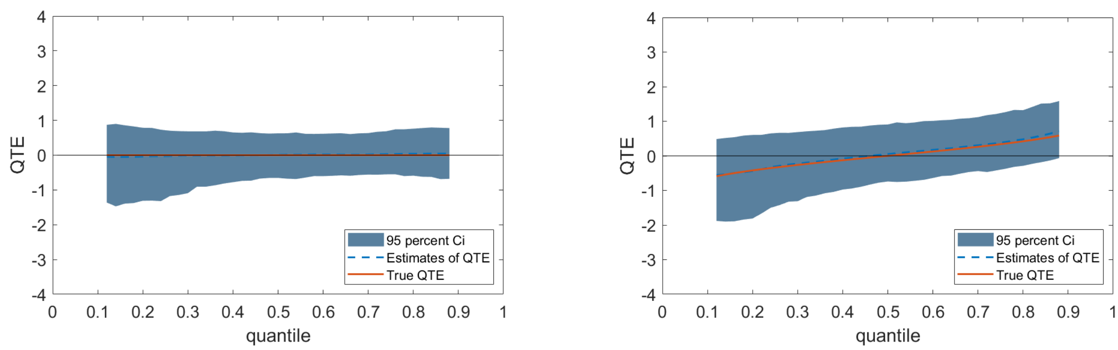
$$\begin{pmatrix} U_1 \\ U_0 \\ V_1 \\ V_0 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 & 0.5 \\ 0.5 & 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 0.5 & 1 \end{pmatrix}\right).$$

Notice that $U_1|(V_1, V_0) \sim^d U_0|(V_1, V_0)$ and $V_1 \sim^d V_0$. Then, we give the determinant of the Jacobian matrix $\Pi'(y_1, y_0)$ by

$$\begin{aligned}
\det\left(\Pi'(y_1, y_0)\right) =& f_{Y|DR}\left(y_1, 1, r_0^+\right) p\left(r_0^+\right) f_{Y|DR}\left(y_0, 0, r_0^-\right)\left(1 - p\left(r_0^-\right)\right) \\
& - f_{Y|DR}\left(y_0, 0, r_0^+\right)\left(1 - p\left(r_0^+\right)\right) f_{Y|DR}\left(y_1, 1, r_0^-\right) p\left(r_0^-\right) \\
=& \frac{1}{4}\Bigg[f_{U_1|V_1}\left(\frac{y_1}{\omega}|V_1 > 0\right) f_{U_1|V_1}\left(y_0|V_1 \leq 1\right) P\{V_1 \leq 1\} \\
& - f_{U_1|V_1}\left(\frac{y_1}{\omega}|V_1 > 1\right) f_{U_1|V_1}\left(y_0|V_1 \leq 0\right) P\{V_1 > 1\}\Bigg] > 0
\end{aligned}$$

for any (y_1, y_0) in its support. Hence, the Jacobian matrix is of full rank.

We set $\omega \in \{1, 1.5\}$ and conduct the simulations based on 2,500 multiplier bootstrap replications with sample size $n = 5,000$. Throughout, we use a uniform kernel.



(a) $\omega = 1$ (b) $\omega = 1.5$

Figure B.1: QTE estimates and 95 percent confidence intervals.

Figure B.1 depicts the QTE estimates for each design and the 95% confidence interval. The bias is small enough to neglect in each case, although the monotonicity condition is violated.

Section C: Bandwidth Selection

We adopt an approach similar to CHS to choose a set of bandwidths for estimated QTEs without covariates in Section 6 based on MSE optimality. We derive the optimal bandwidths for the local quadratic estimation by running the local cubic regression. In practice, the kernel function that we use is the uniform kernel. Let $r(u) = (1, u, u^2)'$ and $e_1 = (1, 0, 0)'$. Recall that

$$\Gamma^+ = \int r(u) \cdot K(u) \cdot r(u)' \mathbb{1}\{u > 0\} du, \quad \Gamma^- = \int r(u) \cdot K(u) \cdot r(u)' \mathbb{1}\{u < 0\} du,$$

$$\Gamma_2^+ = \int r(u) \cdot K^2(u) \cdot r(u)' \mathbb{1}\{u > 0\} du, \quad \Gamma_2^- = \int r(u) \cdot K^2(u) \cdot r(u)' \mathbb{1}\{u < 0\} du.$$

Additionally, define

$$\Lambda_{2,3}^+ = \int u^3 \cdot r(u) \cdot K(u) \mathbb{1}\{u > 0\} du, \quad \Lambda_{2,3}^- = \int u^3 \cdot r(u) \cdot K(u) \mathbb{1}\{u < 0\} du,$$

and

$$\Gamma = \Gamma^+ + \Gamma^-, \quad \Gamma_2 = \Gamma_2^+ + \Gamma_2^-, \quad \Lambda_{2,3} = \Lambda_{2,3}^+ + \Lambda_{2,3}^-.$$

Then, the bandwidth selection strategy consists of the following procedures:

Step 1. Determine the bandwidths to estimate the (conditional) kernel density $f_R(r_0)$ and $f_{YD|R}(y, d, r_0^{\pm})$ by Silverman's rule of thumb

$$\begin{aligned} h_{f_R} &= 1.06 \widehat{\sigma}_R n^{-1/5}, \\ h_{f_1} &= 1.06 \widehat{\sigma}_Y n^{-1/6}, \\ h_{f_2} &= 1.06 \widehat{\sigma}_R n^{-1/6}, \end{aligned}$$

where $\widehat{\sigma}_Y$ and $\widehat{\sigma}_R$ denote the standard deviation of the sample $\{Y_i\}_{i=1}^n$ and $\{R_i\}_{i=1}^n$, respectively. Moreover, h_{f_1} and h_{f_2} represent the bandwidths for Y and R in estimating $f_{YD|R}(y, d, r_0^{\pm})$, respectively. Then, the preliminary bandwidths for $E\left[W|R = r_0^{\pm}\right]$ is

$$h^0 = \left(\frac{1}{6} \frac{C_0'}{C_0^2} \right)^{1/5} n^{-1/5},$$

where the constant terms

$$\begin{aligned} C_0 &= e_1' \left[\frac{(\Gamma^+)^{-1} \Lambda_{2,3}^+}{3!} \bar{\mu}_+^{(3)} - \frac{(\Gamma^-)^{-1} \Lambda_{2,3}^-}{3!} \bar{\mu}_-^{(3)} \right], \\ C_0' &= e_1' \left[\bar{\sigma}_+^2 (\Gamma^+)^{-1} \Gamma_2^+ (\Gamma^+)^{-1} + \bar{\sigma}_-^2 (\Gamma^-)^{-1} \Gamma_2^- (\Gamma^-)^{-1} \right] e_1 \Big/ \widehat{f}_R(r_0) \end{aligned}$$

where we can estimate $\bar{\mu}_{\pm}^{(3)}$ via global cubic regression, and $\bar{\sigma}_{\pm}^2$ is the sample variance of $\bar{\mu}_{\pm}^{(3)}$. That is, we estimate $\bar{\mu}_+^{(3)}$ is estimated as the value of d that solves

$$\arg \min_{a,b,c,d} \sum_{i=1}^n \mathbb{1}\{R_i > r_0\} \left(W_i - a - b(R_i - r_0) - \frac{c}{2!}(R_i - r_0)^2 - \frac{d}{3!}(R_i - r_0)^3 \right)^2$$

and estimate $\bar{\mu}_-^{(3)}$ as the value of d that solves

$$\arg \min_{a,b,c,d} \sum_{i=1}^n \mathbb{1}\{R_i < r_0\} \left(W_i - a - b(R_i - r_0) - \frac{c}{2!}(R_i - r_0)^2 - \frac{d}{3!}(R_i - r_0)^3 \right)^2.$$

Step 2. Using the preliminary bandwidth h^0, estimate $\check{\alpha}_+' = \left(\check{\mu}_+(r_0), \check{\mu}_+^{(1)}(r_0), \check{\mu}_+^{(2)}(r_0), \check{\mu}_+^{(3)}(r_0) \right)'$ as the value of α_+' that solves

$$\arg \min_{\alpha_+} \sum_{i=1}^n \mathbb{1}\{R_i > r_0\} \left(W_i - \left(1, (R_i - r_0), \frac{(R_i - r_0)^2}{2!}, \frac{(R_i - r_0)^3}{3!} \right)' \alpha_+ \right)^2 K\left(\frac{R_i - r_0}{h^0} \right),$$

and estimate $\check{\alpha}'_- = \left(\check{\mu}_-(r_0), \check{\mu}_-^{(1)}(r_0), \check{\mu}_-^{(2)}(r_0), \check{\mu}_-^{(3)}(r_0)\right)'$ as the value of α'_- that solves

$$\arg\min_{\alpha_-} \sum_{i=1}^{n} \mathbb{1}\left\{R_i < r_0\right\} \left(W_i - \left(1, (R_i - r_0), \frac{(R_i - r_0)^2}{2!}, \frac{(R_i - r_0)^3}{3!}\right)' \alpha_-\right)^2 K\left(\frac{R_i - r_0}{h^0}\right).$$

Then, compute the covariance estimates by

$$\check{\sigma}_+^2(r_0) = \sum_{i=1}^{n} (W_i - \check{\mu}(r_0))^2 K\left(\frac{R_i - r_0}{h^0}\right) \mathbb{1}\left\{R_i > r_0\right\} \Big/ \sum_{i=1}^{n} K\left(\frac{R_i - r_0}{h^0}\right) \mathbb{1}\left\{R_i > r_0\right\}$$

$$\check{\sigma}_-^2(r_0) = \sum_{i=1}^{n} (W_i - \check{\mu}(r_0))^2 K\left(\frac{R_i - r_0}{h^0}\right) \mathbb{1}\left\{R_i < r_0\right\} \Big/ \sum_{i=1}^{n} K\left(\frac{R_i - r_0}{h^0}\right) \mathbb{1}\left\{R_i < r_0\right\}.$$

Step 3. Finally, we can derive the bandwidths by

$$h^{MSE} = \left(\frac{1}{6}\frac{\hat{C}'}{\hat{C}^2}\right)^{1/7} n^{-1/7},$$

where the constant terms

$$\hat{C} = e_1' \left[\frac{\left(\Gamma^+\right)^{-1} \Lambda_{2,3}^+}{3!} \check{\mu}_+^{(3)}(r_0) - \frac{\left(\Gamma^-\right)^{-1} \Lambda_{2,3}^-}{3!} \check{\mu}_-^{(3)}(r_0)\right],$$

$$\hat{C}' = e_1' \left[\check{\sigma}_+^2(r_0) \left(\Gamma^+\right)^{-1} \Gamma_2^+ \left(\Gamma^+\right)^{-1} + \check{\sigma}_-^2(r_0) \left(\Gamma^-\right)^{-1} \Gamma_2^- \left(\Gamma^-\right)^{-1}\right] e_1 \Big/ \widehat{f}_R\left(r_0\right).$$

Following Calonico, Cattaneo, and Farrell (2016), we apply the rule-of-thumb (ROT) bandwidth algorithm for coverage error:

$$h^{ROT} = h^{MSE} n^{-2/35} = \left(\frac{1}{6}\frac{\hat{C}'}{\hat{C}^2}\right)^{1/7} n^{-1/5}.$$