SUPPLEMENT TO "ON TESTING CONDITIONAL QUALITATIVE TREATMENT EFFECTS"

By Chengchun Shi, Rui Song and Wenbin Lu

North Carolina State University

9. More on the forward selection algorithm.

9.1. *Statistical properties.* In this section, we investigate the statistical properties of the forward selection algorithm introduced in Section 7. For completeness, we restate the algorithm below.

1. Set $\widehat{D} = \emptyset$. In the first step, for each variable i, define the set $W_i = \{i\}$ and calculate the p-value p_i for each test statistic $\widetilde{T}^{W_i,\widehat{D}}$ as described in Section 5. Stop if $\min_i p_i > \alpha$. Include the variable that gives the smallest p-value in the set $\widehat{D}$, i.e,
$$\widehat{D} \leftarrow \{\arg\min_i p_i\}.$$
2. In the second step, for each variable $i \notin \widehat{D}$, define $W_i = \widehat{D} \cup \{i\}$ and calculate the p-value p_i for each test statistic $\widetilde{T}^{W_i,B}$. Stop if $\min_i p_i > \alpha$. Include the variable that gives the smallest p-value,
$$\widehat{D} \leftarrow \widehat{D} \cup \{\arg\min_i p_i\}.$$
3. Continue the second step until it stops. Output $\widehat{D}$.

Below, we establish the "sure screening property" and selection consistency of the proposed algorithm. Recall that $I = \{1, \ldots, p\}$. For any $D \subseteq I$, recall that

$$V^D = \mathrm{E}\{Y^*(d_{opt}^D)\},$$

is the value function under the treatment regime d_{opt}^D where d_{opt}^D is defined in Section 2.2. Let $V_0 = V^I$. Define

$$\mathcal{D}_0 = \{D \subseteq W_0 : V^D = V_0\} \text{ and } D_0 = \arg\min_{D \subseteq \mathcal{D}_0} |D|,$$

where $|D|$ denotes the cardinality of D.

We assume the set D_0 is nonempty and unique. The objective of our variable selection algorithm is to identify D_0. An algorithm is said to possess the "sure screening property" if its output $\widehat{D}$ satisfies

$$\Pr\left(D_0 \subseteq \widehat{D}\right) \to 1.$$

An algorithm is said to be selection consistent if its output $\widehat{D}$ satisfies

$$\Pr\left(\widehat{D} = D_0\right) \to 1.$$

Let $\bar{\Phi}(z) = \Pr(\mathbb{Z}_0 > z)$ where $\mathbb{Z}_0$ is a standard normal random variable. For any set D, D^c denotes the complement of D. We assume the following conditions.

(C1.) Assume all the covariates are continuous and the support Ω takes the form $\Omega = \prod_{j=1}^{p} \Omega^{\{j\}}$ where each $\Omega^{\{j\}}$ is a bounded open set in $\mathbb{R}$.
(C2.) For any set $D \subsetneq I$, and $j_0 \in I \cap D^c$, assume (A1)-(A8) hold with $W = D \cup \{j_0\}$ and $B = D$.
(C3.) For any set $D \subsetneq D_0$, there exists some $j_0 \in D_0 \cap D^c$ such that

$$V^D < V^{D \cup \{j_0\}}.$$

THEOREM 9.1. *Assume (C1), (C2) and (C3) hold. Assume the threshold α in the forward selection algorithm satisfies that $\alpha \to 0$ and $\alpha \geq \bar{\Phi}(c_n)$ for some sequence $c_n \ll \sqrt{n}$. Then the output $\widehat{D}$ of the forward selection algorithm satisfies*

$$Pr\left(D_0 \subseteq \widehat{D}\right) \to 1.$$

By Theorem 3.6 and 3.8, Condition (C1) and (C2) imply that the test statistic $\widetilde{T}^{D \cup \{j_0\}, D}$ is consistent for any $D \subsetneq I$ and $j_0 \in I \cap D^c$. Theorem 9.1 proves the "sure screening property" of the proposed algorithm. To show its selection consistency, we need the following conditions.

(C4.) Assume there exists some continuous functions $\varphi_{D_0}(\cdot)$ and $\varphi_{D_0^c}(\cdot)$ such that for all x, we have

$$\tau_0(x) = \varphi_{D_0}(x_{D_0})\varphi_{D_0^c}(x_{D_0^c}),$$

where x_{D_0} and $x_{D_0^c}$ are the sub-vectors of x, formed by elements in D_0 and D_0^c.
(C5.) Assume X^{D_0} is independent of $X^{D_0^c}$.

LEMMA 9.1. *Assume (C1), (C2), (C4) and (C5) hold. Assume the threshold α in the forward selection algorithm satisfies that $\alpha \to 0$ and $\alpha \geq \bar{\Phi}(c_n)$ for some sequence $c_n \ll \sqrt{n}$. Then the output $\widehat{D}$ of the forward selection algorithm satisfies*

$$Pr\left(\widehat{D} \subseteq D_0\right) \to 1.$$

THEOREM 9.2. *Assume (C1)-(C5) hold. Assume the threshold α in the forward selection algorithm satisfies that $\alpha \to 0$ and $\alpha \geq \bar{\Phi}(c_n)$ for some sequence $c_n \ll \sqrt{n}$. Then the output $\widehat{D}$ of the forward selection algorithm satisfies*

$$Pr\left(\widehat{D} = D_0\right) \to 1.$$

Under Condition (C4), we can show $\varphi_{D_0^c}(x_{D_0^c})$ is nonnegative for all $x_{D_0^c}$. Condition (C5) assumes that the set of variables that are qualitatively interacted with the treatment are independent of the others. Lemma 9.1 implies that the proposed forward selection algorithm won't choose overfitted models. Theorem 9.2 is directly implied by Lemma 9.1 and Theorem 9.1.

9.2. *Numerical studies.* In this section, we examine the numerical performance of the forward selection algorithm via Monte Carlo simulations. Simulated data are generated from

$$Y = 1 - \frac{X^{(1)} - X^{(2)}}{2} + A\tau(X) + e.$$

As in Section 6, we set $A \sim \text{Binom}(1, 0.5)$, $e \sim N(0, 0.5^2)$, $X^{(1)}, X^{(2)}, \ldots, X^{(p)} \sim \text{Unif}[-2, 2]$. Besides, A, e, $X^{(1)}, \ldots, X^{(p)}$ are mutually independent. Set

$$\tau(x) = \varphi(x_{(1)}) \frac{(x_{(2)} + x_{(3)})^2}{2}.$$

By Theorem 2.2, variables $X^{(2)}, \ldots, X^{(p)}$ doesn't have CQTE given $X^{(1)}$.

We consider two scenarios. In the first scenario, we set

$$\varphi(z) = z^2 - \delta,$$

for some $\delta \geq 0$. In the second scenario, we set

$$\varphi_2(z) = \left\{ \begin{array}{cc} z, & 0 \leq z \leq 2, \\ 0, & \delta - 2 \leq z < 0, \\ 2 + z - \delta, & -2 \leq z < \delta - 2, \end{array} \right.$$

TABLE 3
Variable selection results (standard errors in parenthesis).

			FS		SAS	
	VD	n	TP	FN	TP	FN
S1	0	300	97.7% (0.6%)	0.0% (0.0%)	25.8% (1.8%)	0.0% (0.0%)
		600	97.3% (0.7%)	0.0% (0.0%)	61.3% (2.0%)	0.0% (0.0%)
	0.1	300	48.0% (2.0%)	47.3% (2.0%)	0.0% (0.0%)	10.5% (1.3%)
		600	74.8% (1.8%)	22.0% (1.7%)	0.0% (0.0%)	37.2% (2.0%)
	0.2	300	80.2% (1.6%)	9.3% (1.2%)	0.0% (0.0%)	4.5% (0.8%)
		600	97.5% (0.6%)	1.2% (0.4%)	0.0% (0.0%)	16.7% (1.5%)
S2	0	300	94.3% (0.9%)	0.0% (0.0%)	0.0% (0.0%)	0.0% (0.0%)
		600	95.2% (0.9%)	0.0% (0.0%)	0.0% (0.0%)	0.0% (0.0%)
	0.1	300	74.5% (1.8%)	24.7% (1.8%)	8.5% (1.1%)	0.0% (0.0%)
		600	95.3% (0.9%)	4.5% (0.8%)	6.3% (1.0%)	0.0% (0.0%)
	0.2	300	97.5% (0.6%)	1.5% (0.5%)	20.0% (1.6%)	0.0% (0.0%)
		600	99.8% (0.2%)	0.0% (0.0%)	21.7% (1.7%)	0.0% (0.0%)

for some $\delta \geq 0$. In both scenarios, we can show that $D_0 = \{1\}$ if $\delta > 0$ and $D_0 = \emptyset$ if $\delta = 0$. Moreover, it can be shown that the difference of the value function under the optimal treatment regime and the optimal fixed treatment regime

$$\mathrm{VD} = \mathrm{E}\tau(X)[d_{opt}(X) - I\{\mathrm{E}\tau(X) \geq 0\}]$$

for Scenario 1 and 2 are equal to $4\delta^{3/2}/9$ and $\delta^2/6$, for all $0 \leq \delta \leq 4/3$.

For each scenario, we further consider three settings by setting $VD = 0, 0.1$ and 0.2. Hence, we have $D_0 = \emptyset$ in the first setting and $D_0 = \{1\}$ in other settings. We set $p = 8$, and consider two sample sizes, $n = 300$ and $n = 600$.

The forward selection algorithm is implemented based on the proposed doubly-robust test statistic. The propensity score function is estimated via penalized logistic regression, while the conditional mean functions are estimated via penalized linear regression, with SCAD penalty function (Fan and Li, 2001). These penalized regressions are implemented by the R package `ncvreg` and the tuning parameters are selected via 10-folded cross-validation. The test statistics are constructed as discussed in Section 5. As in Section 6, we choose the bandwidth $h = 6n^{-2/7}$ when there's only one continuous variable in the kernel estimation. Otherwise, we set $h = 2\sqrt{3}n^{-1/7}$. In (5.1) and (5.2), we set $\eta_n = n^{-2/7}$, $C_1 = 3$ and $C_2 = 1$. The threshold α in the algorithm is set to be $\bar{\Phi}(n^{1/6}/2)$.

In Table 3, we report the model selection results of the proposed forward selection algorithm (FS) and the sequential advantage selection (SAS, Fan et al., 2016). Specifically, we report the percentage of selecting the true models (TP) and the false negative (FN) rate (the percentage of important variables that are missed). Results are aggregated over 600 simulations.

From Table 3, it can be seen that the proposed algorithm achieves better model selection results in terms of TP. Except for the first setting of Scenario 1, TPs of SAS are less than 22% in all other cases. Besides, SAS never correctly identify D_0 in the last two settings of Scenario 1. On the contrary, except for the setting where $n = 300$, VD $= 0$, TPs of the proposed algorithm are higher than 74% in all other cases. Moreover, when VD > 0, TP of the proposed algorithm increases as the sample size or VD increases. This demonstrates the model selection consistency of the forward selection algorithm.

When VD $= 0$, we have $D_0 = \emptyset$. As a result, FNs of both algorithms are equal to 0. When VD > 0, we note that SAS has relatively small FNs, which implies that SAS tends to select overfitted models. This is in line with our real data analysis, where SAS selects almost all baseline variables.

10. Choice of tuning parameters. We've shown in Theorem 3.14 that as long as h_W and h_B satisfy Condition (A6) and η_n satisfies Condition (A8), the asymptotic power function of our test $\widetilde{T}_n^{W,B} > z_\alpha$ equals

$$\Pr(\widetilde{T}_n^{W,B} > z_\alpha) = 1 - \Phi\left(z_\alpha - \frac{1}{2\tilde{\sigma}} \int_{x_W \in F_0} |\delta_0^W(x_W)| f^W(x_W) dx_W\right),$$

under the local alternative defined in Section 3.3. Notice that the above asymptotic power function is independent of the tuning parameters. This means our test achieves the same asymptotic power function for any h_W, h_B satisfy (A6) and any η_n satisfies (A8).

In our simulations studies, we set the smoothing parameters as $h_W = c_W n^{-1/7}$, $h_B = c_B n^{-2/7}$ with $c_W = 2\sqrt{3}$ and $c_B = 6$ in Scenario 1-4, and $h_W = c_W n^{-2/7}$ with $c_W = 6$ in Scenario 5. In (5.1) and (5.2), we set $\eta_n = n^{-2/7}$, $C_1 = 3$ and $C_2 = 1$. In this section, we examine the performance of our test under other choices of tuning parameters. Specifically, in Scenario 1-4, we consider the following four choices of c_W and c_B: (A) $c_W = 3$, $c_B = 5$; (B) $c_W = 4$, $c_B = 7$; (C) $c_W = 2\sqrt{5}$, $c_B = 8$; (D) $c_W = 5$, $c_B = 9$. In Scenario 5, we also consider the following four choices of c_W: (A) $c_W = 5$; (B) $c_W = 7$; (C) $c_W = 8$; (D) $c_W = 9$. For each choice of c_W and c_B, we adjust the constants C_1 and C_2 such that they satisfy:

$$\frac{C_1}{3} = \sqrt{\frac{(2\sqrt{3})^{p_W}}{c_W^{p_W}}} \quad \text{and} \quad C_2 = \sqrt{\frac{(6)^{p_B}}{c_B^{p_B}}}. \tag{10.1}$$

Table 4
Simulation results.

	n	VD = 0, α level 0.05	VD = 0, α level 0.1	VD = 4%, α level 0.05	VD = 4%, α level 0.1	VD = 8%, α level 0.05	VD = 8%, α level 0.1	VD = 12%, α level 0.05	VD = 12%, α level 0.1
(A)									
Scenario 1	300	3.0%	5.7%	28.8%	36.8%	57.8%	67.0%	80.2%	87.2%
	600	3.0%	5.3%	36.8%	45.5%	74.2%	82.3%	94.7%	97.2%
Scenario 2	300	8.2%	12.8%	29.3%	39.5%	66.8%	74.3%	89.5%	93.2%
	600	4.8%	9.2%	32.8%	45.3%	83.3%	90.3%	99.3%	99.8%
Scenario 3	300	5.0%	9.3%	39.7%	50.2%	72.0%	78.3%	94.2%	96.5%
	600	5.5%	9.3%	54.2%	64.5%	91.3%	94.0%	99.7%	99.7%
Scenario 4	300	7.5%	12.0%	41.0%	50.7%	84.2%	88.2%	96.3%	97.3%
	600	5.7%	9.5%	55.5%	65.0%	96.7%	98.0%	99.8%	100.0%
Scenario 5	300	5.8%	10.8%	25.2%	34.7%	70.3%	80.2%	93.8%	96.5%
	600	4.7%	8.3%	42.7%	55.7%	93.7%	97.0%	99.8%	99.8%
(B)									
Scenario 1	300	2.7%	5.2%	28.8%	36.3%	56.2%	65.0%	82.2%	86.2%
	600	2.3%	4.0%	39.1%	49.2%	76.8%	82.5%	94.7%	97.7%
Scenario 2	300	7.2%	10.8%	24.5%	36.2%	60.7%	71.0%	88.3%	92.5%
	600	4.8%	7.5%	31.0%	41.2%	82.5%	89.5%	99.3%	99.8%
Scenario 3	300	4.3%	7.5%	34.7%	43.0%	77.3%	83.5%	94.7%	97.0%
	600	3.0%	5.3%	52.2%	63.0%	92.3%	95.0%	99.3%	99.5%
Scenario 4	300	7.2%	11.2%	38.5%	47.2%	78.3%	85.0%	95.8%	97.0%
	600	5.7%	10.2%	56.2%	66.5%	95.3%	97.5%	100.0%	100.0%
Scenario 5	300	5.2%	9.7%	29.3%	40.5%	68.0%	76.3%	94.0%	96.8%
	600	4.3%	8.3%	44.5%	55.5%	94.5%	97.7%	99.7%	99.8%
(C)									
Scenario 1	300	2.8%	4.7%	28.5%	35.2%	56.5%	65.8%	82.8%	87.3%
	600	2.5%	3.8%	40.7%	49.7%	78.7%	83.7%	95.0%	97.3%
Scenario 2	300	6.0%	10.8%	22.8%	32.8%	59.7%	70.3%	88.2%	92.3%
	600	4.0%	7.7%	29.7%	40.7%	82.0%	89.7%	99.5%	99.7%
Scenario 3	300	3.5%	5.3%	39.5%	48.7%	73.0%	79.7%	93.2%	95.7%
	600	3.2%	5.3%	53.7%	62.2%	92.7%	96.2%	99.3%	99.8%
Scenario 4	300	5.7%	8.8%	33.7%	44.2%	76.3%	82.0%	96.8%	97.7%
	600	6.0%	8.5%	56.8%	66.2%	96.0%	98.2%	99.8%	100.0%
Scenario 5	300	5.2%	9.7%	29.3%	40.5%	68.0%	76.3%	94.0%	96.8%
	600	4.7%	8.2%	45.3%	57.0%	95.7%	97.8%	99.8%	99.8%
(D)									
Scenario 1	300	2.0%	4.3%	27.7%	35.2%	58.8%	67.7%	83.5%	89.5%
	600	2.5%	3.7%	42.3%	49.5%	80.7%	85.2%	96.3%	98.5%
Scenario 2	300	7.2%	11.5%	23.3%	31.0%	57.8%	67.3%	88.3%	92.3%
	600	4.0%	8.0%	29.3%	39.8%	81.3%	89.7%	99.5%	99.7%
Scenario 3	300	2.0%	4.3%	40.0%	48.8%	78.8%	84.7%	96.0%	97.8%
	600	3.2%	5.3%	57.0%	65.8%	93.8%	95.8%	99.8%	100.0%
Scenario 4	300	4.5%	8.0%	30.3%	40.7%	78.8%	85.7%	96.3%	97.5%
	600	5.8%	8.7%	53.7%	62.5%	96.5%	98.7%	99.7%	99.8%
Scenario 5	300	5.2%	9.7%	29.3%	40.5%	68.0%	76.3%	94.0%	96.8%
	600	4.7%	8.2%	45.2%	58.0%	96.2%	98.2%	99.8%	100.0%

To ensure the consistency of our test, we require

$$C_1\eta_n \gg \sqrt{\frac{nh_W^{p_W}}{\log n}} \quad \text{and} \quad C_2\eta_n \gg \sqrt{\frac{nh_B^{p_B}}{\log n}}.$$

The constraints in (10.1) guarantee that the ratios $\sqrt{nh_W^{p_W}/\log(n)}/(C_1\eta_n)$ and $\sqrt{nh_B^{p_B}/\log n}/(C_2\eta_n)$ are the same as those in Section 6. Similar to Section 6, for each scenario, we further consider four settings by setting $\mathrm{VD} = 0, 0.04, 0.08$ and 0.12. In Table 4, we report the proportions of rejecting H_0 of our proposed test. It can be seen that for each choice of c_W and c_B, the rejection probabilities are very close to those in Section 6.

11. Fully nonparametric extensions. In this section, we propose a fully nonparametric procedure based on some nonparametric estimators of the propensity score and the conditional mean functions.

Let $\pi(x) = \Pr(A = 1|X = x)$, $\Phi_0(x) = \mathrm{E}(Y|X = x, A = 0)$ and $\Phi_1(x) = \mathrm{E}(Y|X = x, A = 1)$. We estimate $\pi(\cdot)$, $\Phi_0(\cdot)$ and $\Phi_1(\cdot)$ via some nonparametric methods. Denoted by $\hat{\pi}(\cdot)$, $\hat{\Phi}_0(\cdot)$ and $\hat{\Phi}_1(\cdot)$ the corresponding estimators. Let $\hat{\pi}_i = \hat{\pi}(X_i)$, $\hat{\Phi}_{0i} = \hat{\Phi}_0(X_i)$ and $\hat{\Phi}_{1i} = \hat{\Phi}_1(X_i)$. Define

$$\begin{aligned}
\tau_{n,NP}^W(x_W) &= \frac{1}{n}\sum_{i=1}^n \left[\left\{\frac{A_i}{\hat{\pi}_i}Y_i - \left(\frac{A_i}{\hat{\pi}_i} - 1\right)\hat{\Phi}_{1i}\right\}\right. \\
&\quad \left. - \left\{\frac{1-A_i}{1-\hat{\pi}_i}Y_i - \left(\frac{1-A_i}{1-\hat{\pi}_i} - 1\right)\hat{\Phi}_{0i}\right\}\right] K_{h_W}^W(x_W - X_i^W), \\
\tau_{n,NP}^B(x_B) &= \frac{1}{n}\sum_{i=1}^n \left[\left\{\frac{A_i}{\hat{\pi}_i}Y_i - \left(\frac{A_i}{\hat{\pi}_i} - 1\right)\hat{\Phi}_{1i}\right\}\right. \\
&\quad \left. - \left\{\frac{1-A_i}{1-\hat{\pi}_i}Y_i - \left(\frac{1-A_i}{1-\hat{\pi}_i} - 1\right)\hat{\Phi}_{0i}\right\}\right] K_{h_B}^B(x_B - X_i^B).
\end{aligned}$$

Let $d_{n,NP}^W(x_W) = I\{\tau_{n,NP}^W(x_W) \geq 0\}$, $d_{n,NP}^B(x_B) = I\{\tau_{n,NP}^B(x_B) \geq 0\}$. Consider

$$\widetilde{S}_{n,NP}^{W,B} = \int_{x_W \in \Omega^W} \tau_{n,NP}^W(x_W)\{d_{n,NP}^W(x_W) - d_{n,NP}^B(x_{W,B})\} I(x_W \notin \hat{E}_{NP}) dx_W,$$

where

$$\hat{E}_{NP} = \left\{x_W : \left|\frac{\tau_{n,NP}^W(x_W)}{\hat{f}^W(x_W)}\right| \leq \eta_n, \left|\frac{\tau_{n,NP}^B(x_{W,B})}{\hat{f}^B(x_{W,B})}\right| \leq \eta_n\right\}.$$

For any set F, define

$$
\begin{aligned}
\hat{a}_{n,NP}(F) &= \frac{1}{\sqrt{2\pi(h_W)^{p_W}}}\int_{x_W\in F_0}\sqrt{\hat{\mu}^W_{n,NP}(x_W)}dx_W,\\
\hat{\sigma}^2_{n,NP}(F) &= \int_{x_W\in F}\int_{t\in[-1,1]^{p_W}}\hat{\mu}^W_{n,NP}(x_W)\mathrm{cov}(\max\{\sqrt{1-\rho^2(t)}\mathbb{Z}_1+\rho(t)\mathbb{Z}_2,0\},\\
&\quad \max\{\mathbb{Z}_2,0\})dx_W dt,
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{\mu}^W_{n,NP}(x_W) = \frac{1}{n(h_W)^{p_W}}\sum_{i=1}^n\Bigg[\left\{\frac{A_i}{\hat{\pi}_i}Y_i-\left(\frac{A_i}{\hat{\pi}_i}-1\right)\hat{\Phi}_{1i}\right\}\\
- \left\{\frac{1-A_i}{1-\hat{\pi}_i}Y_i-\left(\frac{1-A_i}{1-\hat{\pi}_i}-1\right)\hat{\Phi}_{0i}\right\}\Bigg]^2\left\{K^W\left(\frac{x_W-X_i^W}{h_W}\right)\right\}^2.
\end{aligned}
$$

We estimate the asymptotic mean and variance of $\sqrt{n}\widetilde{S}^{W,B}_{n,NP}$ by $\hat{a}_{n,NP}(\hat{F}_{NP})$ and $\hat{\sigma}^2_{n,NP}(\hat{F}_{NP})$, respectively, with

$$
\hat{F}_{NP}=\{x_W:|\tau^W_{n,NP}(x_W)/\hat{f}^W(x_W)|\le\eta_n,|\tau^B_{n,NP}(x_{W,B})/\hat{f}^B(x_{W,B})|>\eta_n\}.
$$

Define

$$
\widetilde{T}^{W,B}_{n,NP}=\begin{cases}\{\sqrt{n}\widetilde{S}^{W,B}_{n,NP}-\hat{a}_{n,NP}(\hat{F}_{NP})\}/\hat{\sigma}_{n,NP}(\hat{F}_{NP}), & \text{if } \nu(\hat{F}_{NP})=0,\\ \{\sqrt{n}\widetilde{S}^{W,B}_{n,NP}-\hat{a}_{n,NP}(\Omega^W)\}/\hat{\sigma}_{n,NP}(\Omega^W), & \text{otherwise.}\end{cases}
$$

We reject the null when $\widetilde{T}^{W,B}_{n,NP}>z_\alpha$.

We conduct some simulation studies to examine the finite sample properties of $\widetilde{T}^{W,B}_{n,NP}$. We consider the same simulation settings as in Section 6. The propensity score function and the conditional mean functions are estimated via kernel ridge regression. We use the Gaussian radial basis function kernel. The estimating procedure is implemented by the R package `CVST`. The tuning parameters in the kernel functions are selected via 5-folded cross-validation. When implementing the test, we consider the same choices of h_W, h_B and η_n as in Section 6.

Reported in Table 5 are the proportions of rejecting the null hypothesis over 600 simulations. When compared to $\widetilde{T}^{W,B}_{n,DR}$, $\widetilde{T}^{W,B}_{n,NP}$ is more powerful. For example, in the third setting of Scenario 1, when $n=600$, the rejection probabilities of $\widetilde{T}^{W,B}_{n,NP}$ are approximately 20% greater than $\widetilde{T}^{W,B}_{n,DR}$. However, in Scenario 4, when $n=300$ and VD $=0$, the empirical type I error rates of $\widetilde{T}^{W,B}_{n,NP}$ are slightly larger than the nominal level.

TABLE 5
Simulation results.

	n	VD = 0, α level 0.05	VD = 0, α level 0.1	VD = 4%, α level 0.05	VD = 4%, α level 0.1	VD = 8%, α level 0.05	VD = 8%, α level 0.1	VD = 12%, α level 0.05	VD = 12%, α level 0.1
Scenario 1	300	0.5%	1.0%	22.7%	30.7%	66.0%	75.8%	95.7%	97.8%
	600	0.2%	0.3%	35.5%	49.5%	94.3%	96.2%	99.7%	100.0%
Scenario 2	300	6.3%	11.1%	24.0%	33.7%	67.3%	73.7%	93.5%	96.5%
	600	3.2%	6.3%	31.7%	42.8%	90.2%	94.8%	99.8%	100.0%
Scenario 3	300	2.5%	3.7%	43.8%	52.0%	86.2%	91.3%	99.0%	99.2%
	600	1.3%	2.0%	63.5%	72.8%	100.0%	100%	100.0%	100.0%
Scenario 4	300	7.8%	11.5%	41.5%	48.7%	82.8%	88.8%	98.0%	98.3%
	600	3.8%	8.5%	58.8%	70.2%	98.3%	98.8%	100.0%	100.0%
Scenario 5	300	5.5%	8.3%	40.0%	49.0%	87.8%	94.0%	99.7%	100.0%
	600	3.2%	5.7%	61.5%	71.7%	99.5%	99.7%	100.0%	100.0%

12. Extensions to L_p-type functionals. For any $q \geq 1$, let $\phi_q(z) = \mathrm{sgn}(z)|z|^q$. Note that $\phi_q(\cdot)$ is monotonically increasing and satisfies $\phi_q(0) = 0$. Consider the following test statistic

$$\widetilde{S}_{n,q}^{W,B} = \int_{x_W \in \Omega^W} \phi_q\{\tau_n^W(x_W)\}\{d_n^W(x_W) - d_n^B(x_{W,B})\} I(x_W \notin \hat{E}) dx_W,$$

where $\hat{E}$ is defined in Section 3.1. When $q = 1$, $\widetilde{S}_{n,q}^{W,B}$ corresponds to $\widetilde{S}_n^{W,B}$ defined in Section 3.1.

For any set $F \subseteq \Omega^W$, define

$$\begin{aligned}
\hat{a}_{n,q}(F) &= \frac{1}{2\sqrt{(h_W)^{p_W}}} \int_{x_W \in F} \{\hat{\mu}_n^W(x_W)\}^{q/2} dx_W \mathrm{E}|\mathbb{Z}_1^q|, \\
\hat{\sigma}_{n,q}^2(F) &= \int_{\substack{x_W \in F \\ t \in [-1,1]^{p_W}}} \hat{\mu}_n^W(x_W) \mathrm{cov}(\max\{\sqrt{1-\rho^2(t)}\mathbb{Z}_1 + \rho(t)\mathbb{Z}_2, 0\}^q, \max\{\mathbb{Z}_2, 0\}^q) \\
&\quad \times dx_W dt,
\end{aligned}$$

where $\hat{\mu}_n^W(x_W)$ is defined in Section 3.2.

Define the test statistic

$$\widetilde{T}_{n,q}^{W,B} = \begin{cases} \{n^{q/2} h^{p(q-1)/2} \widetilde{S}_{n,q}^{W,B} - \hat{a}_{n,q}(\hat{F})\}/\hat{\sigma}_{n,q}(\hat{F}), & \text{if } \nu(\hat{F}) \neq 0, \\ \{n^{q/2} h^{p(q-1)/2} \widetilde{S}_{n,q}^{W,B} - \hat{a}_{n,q}(\Omega^W)\}/\hat{\sigma}_{n,q}(\Omega^W), & \text{otherwise,} \end{cases}$$

where $\hat{F}$ is defined in Section 3.2. We reject H_0 if $\widetilde{T}_{n,q}^{W,B} > z_\alpha$.

In the following, we establish the statistical properties of $\widetilde{T}_{n,q}^{W,B}$. We first introduce some following conditions.

(A3*.) Assume $\mathrm{E}\exp(t|Y|) < \infty$ for some $t > 0$, and $\sup_{x_W \in \Omega^W} \mathrm{E}(Y^{\max(r_0 q, 4)} | X^W = x_W, A = a) < \infty$ for $a = 0, 1$ and some $r_0 > 2$.
(A8*.) Assume η_n satisfies $\eta_n^{2\xi_0} \gg \log^{\xi_0+q}(n)/\{n(h_W)^{p_W}\}^{\xi_0}$, $n^q h^{p(q-1)} \eta_n^{2\xi_0+2q} \to 0$.

Compared to (A3), when $q \geq 2$, (A3*) assumes a stronger condition on the moments of Y conditional on X and A. When $q = 1$, Condition (A8*) reduces to (A8).

THEOREM 12.1. *Assume Conditions (A1), (A2), (A3*), (A4)-(A7), (A8*) hold. Then, under H_0, we have*

$$\lim_n Pr(\widetilde{T}_{n,q}^{W,B} > z_\alpha) \leq \alpha,$$

for $0 < \alpha \leq 0.5$, where the equality holds when $\nu(F_0) > 0$.

To establish the local power property of $\widetilde{T}_{n,q}^{W,B} > z_\alpha$, consider the following sequence of local alternatives:

$$H_a : \tau_{n,0}^W(x_W) = \tau_0^W(x_W) + n^{-1/2}\delta_0^W(x_W),$$

for some continuous functions τ_0^W and δ_0^W on Ω^W. As in Section 3.3, we assume for any fixed $x_B \in \Omega^B$,

$$\tau_0^W(x_B, x_C) \leq 0 \text{ for any } x_C \in \Omega^C, \text{ or } \tau_0^W(x_B, x_C) \geq 0 \text{ for any } x_C \in \Omega^C,$$

and

$$\delta_0^W(x_B, x_C) \leq 0 \text{ for any } x_C \in \Omega^C, \text{ or } \delta_0^W(x_B, x_C) \geq 0 \text{ for any } x_C \in \Omega^C.$$

In addition,

$$\delta_0^W(x_B, x_C)\tau_0^W(x_B, x_C) \leq 0, \qquad \forall x_B \in \Omega^B, x_C \in \Omega^C.$$

Let

$$\tilde{\sigma}_q^2 = \int_{\substack{x_W \in F_0 \\ t \in [-1,1]^{p_W}}} \mu^W(x_W) \mathrm{cov}(\max\{\sqrt{1-\rho^2(t)}\mathbb{Z}_1 + \rho(t)\mathbb{Z}_2, 0\}^q, \max\{\mathbb{Z}_2, 0\}^q) dx_W dt,$$

where $\mu^W(x_W)$ is defined in Section 3.2, we have the following results.

THEOREM 12.2. *Assume Conditions (A1), (A2), (A3*), (A4)-(A7), (A8*) hold. Assume δ_0^W is bounded on Ω^W. Then, under H_a with*

$$\int_{x_W \in F_0} |\delta_0^W(x_W)| f^W(x_W) dx_W > 0,$$

we have

$$
\begin{aligned}
& \lim_n Pr(\widetilde{T}^{W,B}_{n,q} > z_\alpha) \\
= \;& 1 - \Phi\left(z_\alpha - \int_{x_W \in \dot{F}_0} \frac{2^{(q-3)/2} q \Gamma(q/2)}{\sqrt{\pi}\tilde{\sigma}_q} \{\mu^W(x_W)\}^{(q-1)/2} \delta_0^W(x_W) f^W(x_W) dx_W \right),
\end{aligned}
$$

where $\Gamma(z) = \int_0^\infty x^{z-1}\exp(-x)dx$.

Theorem 12.2 shows that $\widetilde{T}^{W,B}_{n,q}$ has non-negligible powers against some $n^{-1/2}$ local alternatives. For different q, the asymptotic pwoer function increases as

$$
\int_{x \in \dot{F}_0} \frac{2^{(q-3)/2} q \Gamma(q/2)}{\sqrt{\pi}\tilde{\sigma}_q} \{\mu^W(x_W)\}^{(q-1)/2} \delta_0^W(x_W) f^W(x_W) d\nu(x_W)
$$

increases. Since $\Gamma(1/2) = \sqrt{\pi}$, when $q = 1$, the asymptotic power function corresponds to the one given in Theorem 3.14.

13. Supremum-type functionals. In this section, we introduce a test statistic based on the supremum-type functional defined in (2.6). We assume all covariates are bounded and continuous. The test statistic is constructed by the kernel estimators $\tau_n^W(x_W)$ and $\tau_n^B(x_B)$ defined in Section 3, with many different bandwidth values. We write

$$
\begin{aligned}
\tau_n^W(x_W, h_W) &= \frac{1}{n}\sum_{i=1}^n \left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i} \right) Y_i K^W_{h_W}(x_W - X_i^W), \\
\tau_n^B(x_B, h_B) &= \frac{1}{n}\sum_{i=1}^n \left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i} \right) Y_i K^B_{h_B}(x_B - X_i^B),
\end{aligned}
$$

in order to emphasize dependence on the bandwidth. Similarly we set $d_n^W(x_W, h_W) = I\{\tau_n^W(x_W, h_W) > 0\}$, $d_n^B(x_B, h_B) = I\{\tau_n^B(x_B, h_B) > 0\}$,

$$
\begin{aligned}
\hat{f}^W(x_W, h_W) = \frac{1}{n}\sum_{i=1}^n K^W_{h_W}(x_W - X_i^W), \quad & \hat{f}^B(x_B, h_B) = \frac{1}{n}\sum_{i=1}^n K^B_{h_B}(x_B - X_i^B), \\
\hat{\mu}_n^W(x_W, h_W) = \frac{1}{n h_W}\sum_{i=1}^n & \left\{ \left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i} \right) Y_i \right\}^2 \left\{ K^W\left(\frac{x_W - X_i^W}{h_W} \right) \right\}^2,
\end{aligned}
$$

and

$$\hat{E}(h_W,h_B)=\left\{x_W\in\Omega^W:\left|\frac{\tau_n^W(x_W,h_W)}{\hat{f}^W(x_W,h_W)}\right|\le\eta_n,\left|\frac{\tau_n^B(x_{W,B},h_B)}{\hat{f}^B(x_{W,B},h_B)}\right|\le\eta_n\right\},$$

$$\hat{F}(h_W,h_B)=\left\{x_W\in\Omega^W:\left|\frac{\tau_n^W(x_W,h_W)}{\hat{f}^W(x_W,h_W)}\right|\le\eta_n,\left|\frac{\tau_n^B(x_{W,B},h_B)}{\hat{f}^B(x_{W,B},h_B)}\right|>\eta_n\right\},$$

for some sequence $\eta_n\to 0$.

Consider the following set of bandwidth values $\mathcal{H}=\{h=h_{\max}t^k:h\ge h_{\min},k=0,1,2,\dots\}$ for some $0<t<1$. We assume both $h_{\max}$ and $h_{\min}$ converge to zero as $n\to\infty$. Define our test statistic

$$\widetilde{T}_n^{W,B}=\max_{\substack{h_W\in\mathcal{H}\\ h_B\in\mathcal{H}}}\max_{\substack{x_W\in\{X_1^{W*},\dots,X_m^{W*}\}\\ x_W\notin\hat{E}(h_W,h_B)}}\frac{\sqrt{nh_W^{p_W}}\tau_n^W(x_W,h_W)}{\sqrt{\hat{\mu}_n^W(x_W,h_W)}}\{d_n^W(x_W,h_W)-d_n^B(x_{W,B},h_B)\},$$

where $X_1^{W*},\dots,X_m^{W*}$ are independently and uniformly generated from Ω. In addition, $X_1^{W*},\dots,X_m^{W*}$ are independent of the data $\{O_i\}_{i=1,\dots,n}$. We let $m\to\infty$ as $n\to\infty$, to guarantee that the test has enough power to detect CQTE.

Let

$$\hat{k}=\begin{cases}\dfrac{1}{|\mathcal{H}|}\displaystyle\sum_{\substack{h_W\in\mathcal{H}\\ h_B\in\mathcal{H}}}\sum_{j=1}^m I\{X_j^{W*}\in\hat{F}(h_W,h_B)\}, & \text{if } \displaystyle\sum_{j=1}^m I\{X_j^{W*}\in\hat{F}(h_W,h_B)\}>0,\\ m|\mathcal{H}|, & \text{otherwise.}\end{cases}$$

For a given significance level α, we reject H_0 when $\widetilde{T}_n^{W,B}>z_{\alpha/(2\hat{k})}$.

For any $h>0$, define the set

$$F_0(h)=\{x_W\in F_0:\exists y_W\notin F_0,\|x_W-y_W\|_\infty\le h/2\}.$$

We assume $\nu(F_0(h))\to 0$, as $h\to 0$. When $F_0=\emptyset$, it is immediate to see that $F_0(h)=\emptyset$ for any $h>0$. To investigate the statistical properties of $\widetilde{T}_n^{W,B}$, we introduce the following conditions.

(D1.) Assume $h_{\max}\to 0$ and $nh_{\min}^{p_W}/\log^3 n\to\infty$.
(D2.) Assume $\nu(\partial F_0)=0$. Assume there exists some constant $\xi_0>0$ such that for any sufficiently small $t>0$,

$$\nu(\{x_W:0<|\tau_0^W(x_W)|\le t\})=O(t^{\xi_0}),\quad \nu(\{x_B:0<|\tau_0^B(x_B)|\le t\})=O(t^{\xi_0}).$$

(D3.) Assume η_n satisfies $\eta_n\to 0$ and $\eta_n\gg\max(\sqrt{\log n}/\sqrt{nh_{\min}^{p_W}},h_{\max}^{-s})$, where s is defined in Assumption (A2).
(D4.) Assume $m\to\infty$, $m=o(\eta_n^{-\xi_0})$ and $m=o(1/\nu(F_0(h_{\max})))$.

THEOREM 13.1. *Assume (A1)-(A5) and (D1)-(D4) hold. Then, under H_0, we have*

$$\lim_n Pr(\widetilde{T}_n^{W,B} > z_\alpha) \leq \alpha,$$

for any $0 < \alpha < 1$. In addition, when $\nu(F_0) = 0$, we have

$$\lim_n Pr(\widetilde{T}_n^{W,B} > z_\alpha) = 0.$$

To investigate the local power property of the test, we consider the following sequence of local alternatives,

$$H_a : \tau_{n,0}^W(x_W) = \tau_0^W(x_W) + a_n \delta_0^W(x_W),$$

for some continuous functions τ_0^W and δ_0^W on Ω^W, and some sequence $\{a_n\}_n$ such that $a_n \to 0$ as $n \to \infty$. Similarly, we assume for any fixed $x_B \in \Omega^B$,

$$\tau_0^W(x_B, x_C) \leq 0 \text{ for any } x_C \in \Omega^C, \text{ or } \tau_0^W(x_B, x_C) \geq 0 \text{ for any } x_C \in \Omega^C,$$

and

$$\delta_0^W(x_B, x_C) \leq 0 \text{ for any } x_C \in \Omega^C, \text{ or } \delta_0^W(x_B, x_C) \geq 0 \text{ for any } x_C \in \Omega^C.$$

In addition,

$$\delta_0^W(x_B, x_C)\tau_0^W(x_B, x_C) \leq 0, \qquad \forall x_B \in \Omega^B, x_C \in \Omega^C.$$

Below, we show that the supremum-type test has nontrivial power against $\sqrt{\log n}/\sqrt{nh_{\max}^p}$-local alternatives.

THEOREM 13.2. *Assume (A1)-(A5) and (D1)-(D4) hold. Assume the function $\delta_0^W(\cdot)$ is uniformly continuous and bounded on Ω^W. Then, under H_a with $\sup_{x_W \in \dot{F}_0} |\delta_0^W(x_W)| > 0$ and $a_n \gg \sqrt{\log n}/\sqrt{nh_{\max}^p}$, then we have*

$$\lim_n Pr(\widetilde{T}_n^{W,B} > z_\alpha) = 1,$$

for any $0 < \alpha < 1$.

14. Properties of the test when all covariates are discrete. In this section, we investigate the theoretical properties of the proposed test in Section 5.1 when all covariates are discrete. Let $f_W(x_W) = \Pr(X^W = x_W)$ and $f_B(x_B) = \Pr(X^B = x_B)$, for any $x_W \in \Omega^W$, $x_B \in \Omega^B$. For simplicity, we assume $|\Omega^W| < \infty$. This implies X^W and X^B can only take a finite number of possible values. For any $x_W \in \Omega^W$, let

$$\mu^W(x_W) = \mathrm{Var}\left\{\left(\frac{A}{\pi(X)} - \frac{1-A}{1-\pi(X)}\right) Y I(X^W = x_W)\right\}.$$

Define

$$F_0 = \{x_W \in \Omega^W : \tau_0^W(x_W) = 0, \tau_0^B(x_{W,B}) \neq 0\},$$

and

$$\hat{c}_\alpha = \begin{cases} \hat{c}_\alpha(\hat{F}) & \text{if } \hat{F} = \emptyset, \\ \hat{c}_\alpha(\Omega) & \text{otherwise,} \end{cases}$$

where $\hat{c}_\alpha(\hat{F})$ and $\hat{c}_\alpha(\Omega)$ are defined in Section 5.1. In the following, we show the proposed test is consistent.

THEOREM 14.1. *Assume (A3), (A4) hold,* $\min_{x_W \in \Omega^W} \mu^W(x_W) > 0$*, and* η_n *satisfies* $\eta_n \gg n^{-1/2}$, $\eta_n \to 0$. *Then, under* H_0*, we have*

$$\lim_n Pr(\sqrt{n}\widetilde{S}_n^{W,B} \geq \hat{c}_\alpha) \leq \alpha,$$

where the equality holds when $Pr(X^W \in F_0) > 0$. *In addition, under* H_1*, we have*

$$\lim_n Pr(\sqrt{n}\widetilde{S}_n^{W,B} \geq \hat{c}_\alpha) \to 1.$$

APPENDIX: PROOFS OF THEOREMS

In Section A.1-A.7, we provide proofs of Theorem 2.2, 3.11, 3.6, 3.8, 3.14, 4.3, 4.5. Proofs of Theorem 12.1, 12.2, 13.1, 13.2, 14.1 are given in Section A.10, A.11, A.12, A.13 and A.14 respectively. Without loss of generality, we assume in the proofs that $W = I$ such that $x = x_W$, $\tau_0 = \tau_0^W$, $f = f^W$, $\mu = \mu^W$ and $\Omega = \Omega^W$. For notational convenience, we omit the superscript (or the subscript) W in $\mathrm{ER}^{W,B}$, $\mathrm{VD}^{W,B}$, $\widetilde{S}_n^{W,B}$, $\widetilde{T}_n^{W,B}$, δ_0^W, h_W, etc., and present them as ER^B, VD^B, $\widetilde{S}_n^B$, $\widetilde{T}_n^B$, δ_0 and h.

Proof of Theorem 9.1 is provided in Section A.8. Proof of Theorem 9.1 is given in Section A.9. Since Theorem 9.2 is implied by Lemma 9.1 and Theorem 9.1, we omit its proof for brevity. Recall that for any $x \in \mathbb{R}^p$, x_D denotes the sub-vector of x formed by elements in D.

A.1 Proof of Theorem 2.2. We only show that (v), (iv), (ii) and (i) are equivalent. Using similar arguments, we can establish the equivalence between (iv) and (iii). We omit this part of the proof for brevity. To prove the equivalence between (i), (ii), (iv) and (v), we show (v)⇒(iv), (iv)⇒(ii), (ii)⇒(i) and (i)⇒(v). The proof is then completed.

(v)⇒(iv). Fix x_B, assume $\tau_0(x_B, x_C) \geq 0$ for all $x_C \in \Omega^C$. Recall that $\tau_0^B(x_B) = \mathrm{E}\{\tau_0(X_0^B, X_0^C)|X_0^B = x_B\}$. If $\tau_0^B(x_B) = 0$, the continuity of τ_0 implies $\tau_0(x_B, x_C) = 0$ for all $x_C \in \Omega^C$. Hence, (iv) automatically holds. When $\tau_0^B(x_B) > 0$, we have

$$I(\tau_0(x) \geq 0) = I(\tau_0^B(x_B) \geq 0), \tag{A.1}$$

for all $x_C \in \Omega^C$ such that $\tau_0(x_B, x_C) > 0$. This shows (iv) holds. The case where $\tau_0(x_B, x_C) \leq 0$ for all $x_C \in \Omega^C$ can be similarly discussed.

(iv)⇒(ii). By (iv), for any $x \in \Omega$, we obtain

$$\tau_0(x)\{I(\tau_0(x) \geq 0) - I(\tau_0^B(x_B) \geq 0)\} = 0,$$

and (ii) thus follows.

(ii)⇒(i). Assume (ii) holds, (i) doesn't hold. Then we can find some x_{B_0}, x_{C_1} and x_{C_2} such that $\tau_0(x_{B_0}, x_{C_1}) > 0$, $\tau_0(x_{B_0}, x_{C_2}) < 0$. Assume $\tau_0^B(x_{B_0}) > 0$. For some sufficiently small $\varepsilon > 0$, the continuity of τ_0, τ_0^B and f suggests there exists some $\delta > 0$, such that $\tau_0^B(x_B) > 0$ for all $\|x_B - x_{B_0}\|_\infty \leq \delta$, and

$$\tau_0(x_B, x_C) \leq -\varepsilon, f(x_B, x_C) \geq \varepsilon, \qquad \forall \|x_B - x_{B_0}\|_\infty \leq \delta, \|x_C - x_{C_2}\|_\infty \leq \delta,$$

where $f(x_B, x_C)$=$f(y)$ with $y = \{(x_B)^T, (x_C)^T\}^T$.

Recall that x_B and x_C are the sub-vectors of x formed by elements in B and C. Thus we have

$$\int_{\substack{\|x_B - x_{B_0}\|_\infty \leq \delta \\ \|x_C - x_{C_2}\|_\infty \leq \delta}} \tau_0(x)\{d_{opt}(x) - d_{opt}^B(x_B)\} f(x) dx \geq \varepsilon^2 \int_{\substack{\|x_B - x_{B_0}\|_\infty \leq \delta \\ \|x_C - x_{C_2}\|_\infty \leq \delta}} dx = \delta^p \varepsilon^2 > 0.$$

This conflicts (ii). When $\tau_0^B(x_{B_0}) < 0$, we can similarly reach a contradiction. Hence, (i) holds.

It remains to consider when $\tau_0^B(x_{B_0}) = 0$. The continuity of τ_0 and f suggests that there exists some $\delta > 0$, such that

$$\begin{aligned} &\tau_0(x_B, x_{C_1}) \geq \varepsilon, f(x_B, x_{C_1}) \geq \varepsilon, \\ &\forall \|x_B - x_{B_0}\|_\infty \leq \delta, \|x_C - x_{C_1}\|_\infty \leq \delta, \end{aligned} \tag{A.2}$$

$$\begin{aligned} &\tau_0(x_B, x_{C_2}) \leq -\varepsilon, f(x_B, x_{C_2}) \geq \varepsilon, \\ &\forall \|x_B - x_{B_0}\|_\infty \leq \delta, \|x_C - x_{C_2}\|_\infty \leq \delta. \end{aligned} \tag{A.3}$$

If $\tau_0^B(x_B) = 0$ for all $\|x_B - x_{B_0}\|_\infty < \delta$. Then by (A.3), we can similarly show

$$\int_{\substack{\|x_B - x_{B_0}\|_\infty \le \delta \\ \|x_C - x_{C_2}\|_\infty \le \delta}} \tau_0(x)\{d_{opt}(x) - d_{opt}^B(x_B)\} f(x) dx > 0.$$

Otherwise, assume there exists some $y_{B_0} \in \Omega^B$ such that $\|y_{B_0} - x_{B_0}\|_\infty < \delta$ and $\tau_0^B(y_{B_0}) < 0$ (The case when $\tau_0^B(y_{B_0}) > 0$ can be similarly considered). The continuity of τ_0^B suggests there exists some $\delta_0 < \delta$ such that for all $x_B \in \Omega^B$ such that $\|x_B - y_{B_0}\|_\infty \le \delta_0$, we have $\|x_{B_0} - x_B\|_\infty < \delta$, and $\tau_0^B(x_B) < 0$. By (A.2), we can show

$$\int_{\substack{\|x_B - y_{B_0}\|_\infty \le \delta_0 \\ \|x_C - x_{C_1}\|_\infty \le \delta_0}} \tau_0(x)\{d_{opt}(x) - d_{opt}^B(x_B)\} f(x) dx \ge \delta_0^p \varepsilon^2 > 0.$$

This implies (ii)$\Rightarrow$(i).

(i)$\Rightarrow$(v). Assume there exists some x_{B_0}, x_{C_1} and x_{C_2} such that $\tau_0(x_{B_0}, x_{C_1}) > 0$, $\tau_0(x_{B_0}, x_{C_2}) < 0$. The continuity of τ_0 implies there exists some $\delta > 0$ such that for some $\varepsilon > 0$, (A.2) and (A.3) holds. Since the density function is strictly positive, we have

$$\Pr(\|X^B - x_{B_0}\|_\infty \le \delta, \|X^C - x_{C_1}\|_\infty \le \delta) > 0,$$
$$\Pr(\|X^B - x_{B_0}\|_\infty \le \delta, \|X^C - x_{C_2}\|_\infty \le \delta) > 0.$$

Besides, (A.2) and (A.3) further suggest

$$\arg\max_a \mathrm{E}(Y^*(a)|X^B = x_B, X^C = x_C) = 1, \forall \|x_B - x_{B_0}\|_\infty \le \delta, \|x_C - x_{C_1}\|_\infty \le \delta,$$
$$\arg\max_a \mathrm{E}(Y^*(a)|X^B = x_B, X^C = x_C) = 0, \forall \|x_B - x_{B_0}\|_\infty \le \delta, \|x_C - x_{C_2}\|_\infty \le \delta.$$

This contradicts (i). The proof is hence completed.

A.2 Proof of Theorem 3.11. By Assumption, we have $\nu(\partial F_0) = 0$. By the continuity of δ_0 and positiveness of functions f, it is immediate to see that (iii) and (iv) are equivalent. The proof is completed if we can show (iii)$\Rightarrow$(ii), (ii)$\Rightarrow$(i) and (i)$\Rightarrow$(iii).

(iii)$\Rightarrow$(ii). Since δ_0 is bounded on Ω. We assume $\sup_{x\in\Omega} |\delta_0(x)| < M$ for some constant $M > 0$. Consider the set

$$F_n = \left\{ x \in \Omega : 0 < |\tau_0(x)| \le \frac{M}{\sqrt{n}} \right\}.$$

Outside the set $F_n \cup \{x \in \Omega : \tau_0(x) = 0\}$, we have $I\{\tau_0(x) \geq 0\} = I\{\tau_{n,0}(x) \geq 0\}$. This implies outside this set, for any fixed x_B, we have

$$\tau_{n,0}(x_B, x_C) \geq 0, \forall x_C, \text{ or } \tau_{n,0}(x_B, x_C) \leq 0, \forall x_C. \tag{A.4}$$

The set $\{x : \tau_0(x) = 0\}$ can be decomposed into $\{x : \tau(x) = 0, \tau_0^B(x_B) = 0\} + \{x : \tau(x) = 0, \tau_0^B(x_B) \neq 0\}$. Observe that $F_0 = \{x : \tau(x) = 0, \tau_0^B(x_B) \neq 0\}$. On the set $\dot{F}_0$, by (iii), we have $\delta_0(x) = 0$ and hence $\tau_{n,0}(x) = 0$. This together with (A.4) implies that outside the set $F_n \cup \partial F_0 \cup \{\tau_0(x) = 0, \tau_0^B(x_B) = 0\}$, for any fixed x_B,

$$\tau_{n,0}(x_B, x_C) \geq 0, \forall x_C, \text{ and } \tau_{n,0}(x_B, x_C) \leq 0, \forall x_C. \tag{A.5}$$

Since the function $\tau_0(x_B, x_C)$ satisfies H_0, for any x_B such that $\tau_0^B(x_B) = 0$, the continuity of τ_0 implies $\tau_0(x_B, x_C) = 0$ for all x_C. Thus, we have

$$\{x : \tau(x) = 0, \tau_0^B(x_B) = 0\} = \{x : \tau_0^B(x_B) = 0, x_C \in \Omega^C\}. \tag{A.6}$$

For any x_B such that $\tau_0^B(x_B) = 0$, it follows from the definition of δ_0 that $\delta_0(x_B, x_C) \geq 0, \forall x_C$ or $\delta_0(x_B, x_C) \leq 0, \forall x_C$. Therefore, for such x_B, we have

$$\tau_{n,0}(x_B, x_C) = n^{-1/2}\delta_0(x_B, x_C) \geq 0, \quad \forall x_C,$$

or

$$\tau_{n,0}(x_B, x_C) = n^{-1/2}\delta_0(x_B, x_C) \leq 0, \quad \forall x_C.$$

This together with (A.4), (A.5) and (A.6) implies that outside the set $F_n \cup \partial F$, we have

$$\tau_{n,0}(x_B, x_C) \geq 0, \ \forall x_C, \text{ and } \tau_{n,0}(x_B, x_C) \leq 0, \ \forall x_C.$$

Besides, we have $\lim_n \nu(F_n \cup \partial F_0) = \lim_n \nu(F_n) = 0$. This verifies (ii).

(ii)⇒(i). Assume (ii) holds. If X^C has QTE conditional on X^B, by definition, there exist some sets $\mathcal{B}, \mathcal{C}_1, \mathcal{C}_2$ and a sequence n_k such that for all $x_B \in \mathcal{B}$, $x_{C_1} \in \mathcal{C}_1$, $x_{C_2} \in \mathcal{C}_2$,

$$\arg\max_{a=0,1} a\tau_{n_k,0}(x_B, x_{C_1}) \neq \arg\max_{a=0,1} a\tau_{n_k,0}(x_B, x_{C_2}).$$

Under $H_1 : \tau_n = \tau_0 + n^{-1/2}\delta_0$ with δ_0 a bounded function, for any fixed x_B, we have either

$$\tau_{n,0}(x_B, x_{C_1}) > 0, \tau_{n,0}(x_B, x_{C_2}) < 0, \quad \forall x_{C_1} \in \mathcal{C}_1, \forall x_{C_2} \in \mathcal{C}_2, \tag{A.7}$$

or

$$\text{(A.8)}\quad \tau_{n,0}(x_B, x_{C_1}) < 0, \tau_{n,0}(x_B, x_{C_2}) > 0, \quad \forall x_{C_1} \in \mathcal{C}_1, \forall x_{C_2} \in \mathcal{C}_2.$$

Let $\varepsilon = \min\{\Pr(X \in \mathcal{B} \times \mathcal{C}_1), \Pr(X \in \mathcal{B} \times \mathcal{C}_2)\}/2$. By definition, we have $\varepsilon > 0$ and $\Pr(X \in \mathcal{B} \times \mathcal{C}_1) > \varepsilon$, $\Pr(X \in \mathcal{B} \times \mathcal{C}_2) > \varepsilon$. Since the density of X is bounded on Ω, it follows from (ii) that there exists some set N_ε with $\Pr(X \in N_\varepsilon) \le \varepsilon$ such that outside N_ε, for sufficiently large n and any x_B,

$$\text{(A.9)}\qquad \tau_{n,0}(x_B, x_C) \ge 0, \forall x_C, \quad \text{or} \quad \tau_{n,0}(x_B, x_C) \le 0, \forall x_C.$$

Since $\Pr(X \in N_\varepsilon) \le \varepsilon$ and $\Pr(X \in \mathcal{B} \times \mathcal{C}_1) > \varepsilon$, $\Pr(X \in \mathcal{B} \times \mathcal{C}_2) > \varepsilon$, we have $(\mathcal{B} \times \mathcal{C}_1) \cap N_\varepsilon^c \ne \emptyset$ and $(\mathcal{B} \times \mathcal{C}_2) \cap N_\varepsilon^c \ne \emptyset$. By (A.7) and (A.8), this means outside N_ε, we can find some x_B, x_{C_1} and x_{C_2} that for sufficiently large n,

$$\tau_{n,0}(x_B, x_{C_1}) > 0 \quad \text{and} \quad \tau_{n,0}(x_B, x_{C_2}) < 0.$$

This contradicts (A.9). Therefore, (i) must hold.

(i)⇒(iii). Assume there exists some $x_0 \in \dot{F}_0$ such that $\tau_0(x_0) = 0, \tau_0^B(x_{0,B}) > 0, \delta_0(x_0) < 0$ where $x_{0,B}$ is the sub-vector of x_0 formed by elements in B. Since $\tau_0^B(x_{0,B}) > 0$, there exists some $x_1 = (x_{0,B}^T, x_{1,C}^T)^T$ such that $\tau_0(x_1) > 0$. By the continuity of τ_0, there exists some $\varepsilon > 0$ and $\delta_1 > 0$ such that for all $\|x - x_1\|_\infty < \delta_1$, we have

$$\text{(A.10)}\qquad \tau_0(x) > \varepsilon, \text{ and } \tau_0^B(x_B) > 0, \text{ and } \delta_0(x) \le 0.$$

Since $x_0 \in \dot{F}_0$, there exists some small $\delta_2 > 0$ such that $x \in \dot{F}_0$ for all $\|x - x_0\|_\infty < \delta_2$. The continuity of δ_0 and τ_0^B suggests that there exists some δ_3 such that $\delta_0(x) < 0$ and $\tau_0^B(x_B)$ for all $\|x - x_0\|_\infty \le \delta_3$. Take $\delta = \min(\delta_1, \delta_2, \delta_3)$, by the definition of F_0, we have for all $\|x - x_0\|_\infty < \delta$,

$$\text{(A.11)}\qquad \tau_0(x) = 0, \text{ and } \tau_0^B(x_B) > 0, \text{ and } \delta_0(x) < 0.$$

Define

$$\begin{aligned}
\mathcal{B} &= \{x_B : \|x_B - x_{0,B}\|_\infty < \delta\}, \\
\mathcal{C}_1 &= \{x_C : \|x_C - x_{0,C}\|_\infty < \delta\}, \\
\mathcal{C}_2 &= \{x_C : \|x_C - x_{1,C}\|_\infty < \delta\},
\end{aligned}$$

where $x_{0,C}$ is the sub-vector of x_0 formed by elements in C.

Note that δ_0 is bounded on Ω. Assume $|\delta_0(x)| \leq M, \forall x$. For sufficiently large n_0 such that $M/\sqrt{n_0} \leq \varepsilon$, we have for all $n \geq n_0$, $x_B \in \mathcal{B}$, $x_C \in \mathcal{C}_2$, $\tau_0(x) > |\delta_0(x)|/\sqrt{n}$ and hence $\tau_{n,0}(x) > 0$. Therefore, for all $x_B \in \mathcal{B}$, and $x_{C_2} \in \mathcal{C}_2$ and $n \geq n_0$, we have

$$\tau_{n,0}(x_B, x_{C_2}) > 0. \tag{A.12}$$

On the other hand, it follows from (A.11) that for all $x_B \in \mathcal{B}$, $x_{C_1} \in \mathcal{C}_1$, we have

$$\tau_{n,0}(x_B, x_{C_1}) < 0. \tag{A.13}$$

Combining (A.12) with (A.13), we have

$$\arg\max_{a=0,1} a\tau_{n,0}(x_B, x_{C_1}) \neq \arg\max_{a=0,1} a\tau_{n,0}(x_B, x_{C_2}).$$

This contradicts (i). When there exists some $x_0 \in \dot{F}_0$ such that $\tau_0(x_0) = 0, \tau_0^B(x_{0,B}) < 0, \delta_0(x_0) > 0$, we can reach a similar contradiction. The proof is thus completed.

A.3 Proof of Theorem 3.6. Consider separately the cases where (i) $\nu(F_0) > 0$; (ii) $\nu(F_0) = 0$ and $\nu(\hat{F}) > 0$; (iii) $\nu(F_0) = 0$ and $\nu(\hat{F}) = 0$. When $\nu(F_0) > 0$, we will show

$$\frac{\sqrt{n}\widetilde{S}_n^B - \hat{a}_n}{\hat{\sigma}_n} \xrightarrow{d} N(0,1),$$

where $\hat{a}_n$ and $\hat{\sigma}_n$ stand for $\hat{a}_n(\hat{F})$ and $\hat{\sigma}_n(\hat{F})$, respectively.

When $\nu(F_0) = 0$ and $\nu(\hat{F}) > 0$, we show with probability going to 1, $\sqrt{n}\widetilde{S}_n^B \leq \hat{a}_n$ and therefore

$$\lim_n \Pr\left(\frac{\sqrt{n}\widetilde{S}_n^B - \hat{a}_n}{\hat{\sigma}_n} \geq z_\alpha\right) \to 0,$$

for any $\alpha \leq 0.5$.

Finally, when $\nu(\hat{F}) = 0$, we show $\Pr(\sqrt{n}\widetilde{S}_n \leq \hat{a}_n(\Omega)) \to 1$. The proof is therefore completed.

Case 1. To show the asymptotic normality of $\widetilde{T}_n^B$, we break the proof into three steps. In the first step, we show $\sqrt{n}\widetilde{S}_n^B = \sqrt{n}\dot{S}_n^B + o_p(1)$, where

$$\dot{S}_n^B = \int_{x:\tau_0(x)=0,\tau_0^B(x_B)\neq 0} \tau_n(x)\{d_n(x) - d_n^B(x_B)\}dx.$$

Next, we show

$$\frac{\sqrt{n}\dot{S}_n^B - \tilde{a}_n}{\tilde{\sigma}} \xrightarrow{d} N(0,1). \tag{A.14}$$

Finally we show $\hat{a}_n(\hat{F})$ and $\hat{\sigma}_n^2(\hat{F})$ are consistent estimators for $\tilde{a}_n$ and $\tilde{\sigma}^2$, respectively. The asymptotic normality of $(\sqrt{n}\dot{S}_n^B - \tilde{a}_n)/\tilde{\sigma}$ therefore follows.

Step 1: Recall that $\widetilde{S}_n^B$ is defined to be

$$\int_{x\in\Omega} \tau_n(x)\{d_n(x) - d_n^B(x_B)\}I(x \notin \hat{E})dx.$$

It follows from Lemma A.2 that Pr-almost surely, we have for any x,

(A.15)

$$|\tau_n(x) - \tau_0(x)f(x)| \le \frac{\gamma\sqrt{\log n}}{\sqrt{nh^p}}, |\tau_n^B(x_B) - \tau_0^B(x_B)f^B(x_B)| \le \frac{\gamma\sqrt{\log n}}{\sqrt{n(h_B)^{p_B}}},$$

for some constant $\gamma > 0$. Under H_0, when $\tau_0(x) \neq 0$, Theorem 2.2 suggests $I(\tau_0(x) \ge 0) = I(\tau_0^B(x_B) \ge 0)$. Therefore, Define

$$E_1 = \left\{x : |\tau_0(x)f(x)| \le \frac{2\gamma\sqrt{\log n}}{\sqrt{nh^p}}\right\}, E_2 = \left\{x : |\tau_0^B(x_B)f^B(x_B)| \le \frac{2\gamma\sqrt{\log n}}{\sqrt{n(h_B)^{p_B}}}\right\}.$$

Under the event defined in (A.15), outside $E_1 \cup E_2$, we have $I(\tau_n(x) \ge 0) = I(\tau_n^B(x_B) \ge 0)$. Therefore, we have Pr-almost surely, that

$$\bar{S}_n^B = \int_{x\in E_1\cup E_2} \tau_n(x)\{d_n(x) - d_n^B(x_B)\}I(x \notin \hat{E})dx. \tag{A.16}$$

Further define

$$\begin{aligned} E_3 &= \left\{x : 0 < |\tau_0(x)f(x)| \le \frac{2\gamma\sqrt{\log n}}{\sqrt{nh^p}}\right\}, \\ E_4 &= \left\{x : 0 < |\tau_0^B(x_B)f^B(x_B)| \le \frac{2\gamma\sqrt{\log n}}{\sqrt{n(h_B)^{p_B}}}\right\}. \end{aligned} \tag{A.17}$$

We now show

$$\sqrt{n}\int_{x\in E_3\cup E_4} \tau_n(x)\{d_n(x) - d_n^B(x_B)\}I(x \notin \hat{E})dx = o_p(1). \tag{A.18}$$

It is immediate to see that (A.18) is smaller than

$$\sqrt{n}\int_{x\in E_3\cup E_4}|\tau_n(x)|dx.$$

Therefore, it suffices to show

$$\sqrt{n}\int_{x\in E_3}|\tau_n(x)|dx = o_p(1), \tag{A.19}$$

and

$$\sqrt{n}\int_{x\in E_4}|\tau_n(x)|dx = o_p(1). \tag{A.20}$$

It follows from (A.15) and the definition of E_3 that for all $x \in E_3$, almost surely we have

$$|\tau_n(x)| \le \frac{3\gamma\sqrt{\log n}}{\sqrt{nh^p}},$$

and hence under Assumption (A7), LHS of (A.19) is bounded by

$$\begin{aligned}\nu(E_3)\frac{3\gamma\sqrt{\log n}}{\sqrt{h^p}} &\le \nu\left\{x : |\tau_0(x)| = O\left(\frac{\sqrt{\log n}}{\sqrt{nh^p}}\right)\right\}\frac{3\gamma\sqrt{\log n}}{\sqrt{h^p}}\\ &= O\left(\frac{(\log n)^{\xi_0/2+1/2}}{n^{\xi_0/2}h^{p\xi_0/2+p/2}}\right).\end{aligned} \tag{A.21}$$

By $\eta_n \gg \sqrt{\log n}/\sqrt{nh^p}$ and $n\eta_n^{2\xi_0+2} \to 0$ in Assumption (A8), we obtain $(\log n)^{\xi_0+1}/(n^{\xi_0}h^{p(\xi_0+1)}) \to 0$. Therefore, RHS of (A.21) is $o(1)$. This verifies (A.19).

As for (A.20), for all $x \in \Omega$, under the event defined in (A.2), we have

$$\begin{aligned}\sqrt{n}\int_{x\in E_4}|\tau_n(x)|dx &\le \sqrt{n}\int_{x\in E_4}|\tau_0(x)f(x)|dx + \sqrt{n}\nu(E_4)\frac{3\gamma\sqrt{\log n}}{\sqrt{nh^p}}\\ &= R_1 + R_2.\end{aligned}$$

Since $\eta_n \gg \sqrt{\log n/nh^p}$ and $h^p \asymp h_B^{p_B}$, we obtain $\eta_n \gg \sqrt{\log n/n(h_B)^{p_B}}$. Under Assumption (A7), we have

$$R_2 = O\left(\frac{(\log n)^{\xi_0/2+1/2}}{n^{\xi_0/2}(h_B)^{p_B\xi_0/2}h^{p/2}}\right) = O(\sqrt{n}\eta_n^{\xi_0+1}) = o(1). \tag{A.22}$$

It suffices to show $R_1 = o_p(1)$.

Under H_0, it follows from (iv) in Theorem 2.2 that for any fixed $x_B \in \Omega^B$, we have $\tau_0(x_B, x_C) \geq 0, \forall x_C$ or $\tau_0(x_B, x_C) \leq 0, \forall x_C$. Therefore,

$$\begin{aligned} \int_{x \in E_4} |\tau_0(x) f(x)| dx &= \int_{x_B \in E_4^B} \int_{x_C \in \Omega^C} |\tau_0(x) f(x)|\, dx_C dx_B \\ &= \int_{x_B \in E_4^B} \left| \int_{x_C \in \Omega^C} \tau_0(x) f(x) dx_C \right| dx_B, \end{aligned} \tag{A.23}$$

where

$$E_4^B = \left\{ x_B : 0 < |\tau_0^B(x_B) f^B(x_B)| \leq \frac{2\gamma\sqrt{\log n}}{\sqrt{n(h_B)^{p_B}}} \right\}.$$

Observe that for any x,

$$\tau_0^B(x_B) f^B(x_B) = \int_{x_C \in \Omega^C} \tau_0(x) f(x) dx_C.$$

This together with (A.23) implies

$$R_1 = \sqrt{n} \int_{x_B \in E_4^B} |\tau_0^B(x_B) f^B(x_B)| dx_B \leq \nu(E_4^B) \frac{2\gamma\sqrt{\log n}}{\sqrt{(h_B)^{p_B}}} = o(1), \tag{A.24}$$

by $\eta_n \gg \sqrt{\log n}/\sqrt{n(h_B)^{p_B}}$, and $n\eta_n^{2\xi_0+2} \to 0$.

Combining (A.22) together with (A.24) gives (A.18). Hence, we have

$$\begin{aligned} \sqrt{n}\bar{S}_n^B &= \sqrt{n}\check{S}_n^B + o_p(1) \\ &= \sqrt{n} \int_{x \in E_5 \cup E_6} \tau_n(x)\{d_n(x) - d_n^B(x_B)\} I(x \notin \hat{E}) dx + o_p(1), \end{aligned}$$

where

$$E_5 = \{x \in \Omega : \tau_0(x) = 0\}, \qquad E_6 = \{x \in \Omega : \tau_0^B(x_B) = 0\}.$$

Under the null, we have $E_6 \subseteq E_5$. Further decompose E_5 into $E_6 + F_0$. Therefore, we have

$$\check{S}_n^B = \int_{x \in E_6 + F_0} \tau_n(x)\{d_n(x) - d_n^B(x_B)\} I(x \notin \hat{E}) dx.$$

We now claim

$$\sqrt{n} \int_{x \in E_6} \tau_n(x)\{d_n(x) - d_n^B(x_B)\} I(x \notin \hat{E}) dx = o_p(1), \tag{A.25}$$

and

$$\text{(A.26)} \qquad \sqrt{n}\int_{x\in F_0}\tau_n(x)\{d_n(x)-d_n^B(x_B)\}I(x\in\hat{E})dx = o_p(1).$$

When (A.25) and (A.26) holds, we have

$$\sqrt{n}\check{S}_n^B = \sqrt{n}\int_{x\in F_0}\tau_n(x)\{d_n(x)-d_n^B(x_B)\}dx + o_p(1).$$

We first show (A.25). For all $x \in E_6$, we have $\tau_0(x) = 0$. On the set defined in (A.15), if follows from the conditions $\eta_n \gg \sqrt{\log n}/\sqrt{nh^p}$, $\eta_n \gg \sqrt{\log n}/\sqrt{n(h_B)^{p_B}}$ that

$$\text{(A.27)} \qquad |\tau_n(x)| \ll \eta_n \text{ and } |\tau_n^B(x_B)| \ll \eta_n, \qquad \forall x \in E_6.$$

Besides, similar to Lemma A.2, we can show almost surely

$$\text{(A.28)} \qquad \sup_x |\hat{f}(x) - f(x)| = O\left(\frac{\sqrt{\log n}}{\sqrt{nh^p}}\right),$$

$$\text{(A.29)} \qquad \sup_{x_B} |\hat{f}^B(x_B) - f^B(x_B)| = O\left(\frac{\sqrt{\log n}}{\sqrt{n(h_B)^{p_B}}}\right).$$

By Assumption (A1), we have $\inf_x f(x) > 0$ and hence $\inf_{x_B} f^B(x_B) > 0$. This together with (A.28) and (A.29) implies with probability going to 1,

$$\text{(A.30)} \qquad \liminf_n \inf_x \hat{f}(x) > 0, \text{ and } \liminf_n \inf_{x_B} \hat{f}^B(x_B) > 0.$$

Combining (A.27) with (A.30), we have with probability going to 1, that

$$\text{(A.31)} \qquad \left|\frac{\tau_n(x)}{\hat{f}(x)}\right| \ll \eta_n, \text{ and } \left|\frac{\tau_n^B(x_B)}{\hat{f}^B(x_B)}\right| \ll \eta_n, \ \forall x \in E_6.$$

By definition,

$$\hat{E} = \left\{x : \left|\frac{\tau_n(x)}{\hat{f}(x)}\right| \le \eta_n, \left|\frac{\tau_n^B(x_B)}{\hat{f}^B(x_B)}\right| \le \eta_n\right\}.$$

Under the event defined in (A.31), LHS of (A.25) is equal to 0.

It follows from (A.15), (A.28) and (A.30) that

$$\text{(A.32)} \sup_x \left|\frac{\tau_n(x)}{\hat{f}(x)} - \tau_0(x)\right| = \sup_x \left|\frac{\tau_n(x) - \tau_0(x)\hat{f}(x)}{\hat{f}(x)}\right|$$
$$\le \sup_x \left|\frac{\tau_n(x) - \tau_0(x)f(x)}{\hat{f}(x)}\right| + \sup_x \left|\frac{\tau_0(x)}{\hat{f}(x)}\{f(x) - \hat{f}(x)\}\right| = O\left(\frac{\sqrt{\log n}}{\sqrt{nh^p}}\right).$$

Similarly we can show

$$\sup_{x_B}\left|\frac{\tau_n^B(x_B)}{\hat{f}^B(x_B)}-\tau_0^B(x_B)\right| = O\left(\frac{\sqrt{\log n}}{\sqrt{n(h_B)^{p_B}}}\right). \tag{A.33}$$

By (A.33) and $\eta_n \gg \sqrt{\log n}/\sqrt{n(h_B)^{p_B}}$, for all x such that $|\tau_n^B(x_B)/\hat{f}^B(x_B)| \le \eta_n$, we have

$$|\tau_0^B(x_B)| \le 2\eta_n.$$

This suggests with probability going to 1, LHS of (A.26) is smaller than

$$\sqrt{n}\int_{|\tau_0^B(x_B)|\le 2\eta_n} \tau_n(x)\{d_n(x)-d_n^B(x_B)\}dx. \tag{A.34}$$

Using similar arguments as in (A.20), we can show (A.34) is $o_p(1)$. This proves (A.26).

It remains to show

$$\int_{x\in F_0} \tau_n(x)\{I(\tau_n(x)\ge 0)-I(\tau_n^B(x_B)\ge 0)\}dx$$

is equivalent to

$$\dot{S}_n^B = \int_{x\in F_0} \tau_n(x)\{I(\tau_n(x)\ge 0)-I(\tau_0^B(x_B)\ge 0)\}dx.$$

To see this, using similar arguments in (A.20), we can show

$$\begin{aligned} &\sqrt{n}\int_{x\in F_0} \tau_n(x)\{I(\tau_n(x)\ge 0)-I(\tau_n^B(x_B)\ge 0)\}dx \\ = {}& \sqrt{n}\int_{x\in F_0\cap E_4^c} \tau_n(x)\{I(\tau_n(x)\ge 0)-I(\tau_n^B(x_B)\ge 0)\}dx + o_p(1). \end{aligned} \tag{A.35}$$

On the set $F_0 \cap E_4^c$, it follows from (A.15) that almost surely we have $I(\tau_0^B(x_B)\ge 0) = I(\tau_n^B(x_B)\ge 0)$ and hence

$$\begin{aligned} &\int_{x\in F_0\cap E_4^c} \tau_n(x)\{I(\tau_n(x)\ge 0)-I(\tau_n^B(x_B)\ge 0)\}dx \\ = {}& \int_{x\in F_0\cap E_4^c} \tau_n(x)\{I(\tau_n(x)\ge 0)-I(\tau_0^B(x_B)\ge 0)\}dx. \end{aligned} \tag{A.36}$$

Using similar arguments in (A.20), we can show the above expression is equivalent to

$$\dot{S}_n^B = \int_{x\in F_0} \tau_n(x)\{I(\tau_n(x)\ge 0)-I(\tau_0^B(x_B)\ge 0)\}dx. \tag{A.37}$$

Combining (A.37) together with (A.35) and (A.36), we've completed the proof for the first step.

Step 2: In this step, we show the asymptotic normality of $(\dot{S}_n^B - \tilde{a}_n)/\tilde{\sigma}$. We begin by defining a Poissonized version of $\dot{S}_n^B$ as

$$\dot{S}_N^B = \int_{x \in F_0} \tau_N(x)\{I(\tau_N(x) \geq 0) - I(\tau_0^B(x_B) \geq 0)\}dx,$$

where

$$\tau_N(x) = \frac{1}{n}\sum_{i=1}^{N}\left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i}\right)Y_i K_h(x - X_i), \tag{A.38}$$

N is a Poisson variable with mean n, independent of (X_i, A_i, Y_i) for $i = 1, 2, \ldots$.

We first show

$$\lim_n n\mathrm{Var}(\dot{S}_N^B) = \tilde{\sigma}^2. \tag{A.39}$$

By definition, we have

$$\begin{aligned}\mathrm{Var}(\dot{S}_N^B) = \int_{x \in F_0}\int_{y \in F_0} &\mathrm{cov}(\tau_N(x)\{I(\tau_N(x) \geq 0) - I(\tau_0^B(x_B) \geq 0)\},\\ &\tau_N(y)\{I(\tau_N(y) \geq 0) - I(\tau_0^B(y^B) \geq 0)\})dxd\nu(y).\end{aligned}$$

By Assumption (A7), we have $\nu(\partial F_0) = 0$. Therefore,

$$\begin{aligned}\mathrm{Var}(\dot{S}_N^B) = \int_{x \in \dot{F}_0}\int_{y \in \dot{F}_0} &\mathrm{cov}(\tau_N(x)\{I(\tau_N(x) \geq 0) - I(\tau_0^B(x_B) \geq 0)\},\\ &\tau_N(y)\{I(\tau_N(y) \geq 0) - I(\tau_0^B(y^B) \geq 0)\})dxd\nu(y),\end{aligned}$$

where $\dot{F}_0$ stands for the interior of F_0.

By Assumption (A2), $K(x) = \prod_{j=1}^{p} K_j(x_j)$ with each K_j that has bounded support over $[-1/2, 1/2]$. This implies we can represent τ_N as

$$\tau_N(x) = \frac{1}{n}\sum_{i=1}^{N}\left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i}\right)Y_i K_h(x - X_i) I(\|x - X_i\|_\infty \leq h),$$

It follows from Lemma A.1 that for any x, y such that $\|x - y\|_\infty > h$, we have $\tau_N(x) \perp\!\!\!\perp \tau_N(y)$, and hence

$$\begin{aligned}&\mathrm{cov}\left(\tau_N(x)\{I(\tau_N(x) \geq 0) - I(\tau_0^B(x_B) \geq 0)\},\right.\\ &\quad\left.\tau_N(y)\{I(\tau_N(y) \geq 0) - I(\tau_0^B(y_B) \geq 0)\}\right) = 0.\end{aligned} \tag{A.40}$$

It follows from (A.40) that

$$n\mathrm{Var}(\dot{S}_N^B) = nh^p \int_{x\in\dot{F}_0}\int_{t\in[-1,1]^p} \mathrm{cov}(\tau_N(x)\{I(\tau_N(x)\geq 0) - d_{opt}^B(x_B)\}, \tau_N(x+ht)$$
$$\{I(\tau_N(x+ht)\geq 0) - d_{opt}^B(x_B+ht_B)\})I(x+ht\in F_0)dxdt.$$

By Lemma A.4, we obtain

$$\sqrt{nh^p}\begin{pmatrix} \tau_N(x)-\tau_0(x)f(x) \\ \tau_N(x+ht)-\tau_0(x+ht)f(x+ht)\end{pmatrix} \stackrel{d}{\to} \begin{pmatrix} \sqrt{\mu(x)}(\sqrt{1-\rho^2(t)}\mathbb{Z}_1+\rho(t)\mathbb{Z}_2) \\ \sqrt{\mu(x)}\mathbb{Z}_1 \end{pmatrix}, \tag{A.41}$$

for independent bivariate normal variables $\mathbb{Z}_1$ and $\mathbb{Z}_2$.

For any fixed $x\in\dot{F}_0$ and sufficiently large n, we have $x+ht\in F_0$ and hence $\tau_0(x+ht)=0$. Besides,

$$\begin{aligned} &nh^p\mathrm{cov}(\tau_N(x)\{I(\tau_N(x)\geq 0)-d_{opt}^B(x_B)\}, \\ &\tau_N(x+ht)\{I(\tau_N(x+ht)\geq 0)-d_{opt}^B(x_B+ht_B)\}) \leq nh^p\sup_{x\in\Omega}\mathrm{E}\tau_N^2(x)<\infty. \end{aligned}\tag{A.42}$$

The last inequality follows from the boundedness of μ on Ω. Function τ_0^B is continuous. By (A.41), (A.42) and dominated convergence theorem, we have

$$\begin{aligned} &nh^p\mathrm{cov}(\tau_N(x)\{I(\tau_N(x)\geq 0)-d_{opt}^B(x_B)\}, \\ &\tau_N(x+ht)\{I(\tau_N(x+ht)\geq 0)-d_{opt}^B(x_B+ht_B)\}) \\ &\to \mu(x)\mathrm{cov}((\sqrt{1-\rho^2(t)}\mathbb{Z}_1+\rho(t)\mathbb{Z}_2)\{I(\sqrt{1-\rho^2(t)}\mathbb{Z}_1+\rho(t)\mathbb{Z}_2\geq 0) \\ &-I(\tau_0^B(x_B)\geq 0)\}, \mathbb{Z}_1\{I(\mathbb{Z}_1\geq 0)-d_{opt}^B(x_B)\}). \end{aligned}$$

The symmetry of $\mathbb{Z}_1$ and $\mathbb{Z}_2$ suggest that the expression on the last two lines are the same regardless of whether $\tau_0^B(x_B)\geq 0$ or $\tau_0^B(x_B)<0$. Hence, we have

$$\begin{aligned} &nh^p\mathrm{cov}(\tau_N(x)\{I(\tau_N(x)\geq 0)-I(\tau_0^B(x_B)\geq 0)\}, \\ &\tau_N(x+ht)\{I(\tau_N(x+ht)\geq 0)-I(\tau_0^B(x_B+ht_B)\geq 0)\}) \\ &\to \mu(x)\mathrm{cov}(\max\{\sqrt{1-\rho^2(t)}\mathbb{Z}_1+\rho(t)\mathbb{Z}_2, 0\}, \max(\mathbb{Z}_1, 0)). \end{aligned}\tag{A.43}$$

Besides, for fixed $x\in F_0$, we have

$$I(x+ht)\to 1. \tag{A.44}$$

It follows from (A.42) that

$$n\mathrm{Var}(\dot{S}_N^B) = O\left(nh^p \sup_{x\in\Omega} \mathrm{E}\tau_N^2(x)\right) < \infty.$$

Using dominated convergence theorem, it follows from (A.43) and (A.44) that (A.39) holds.

We now show $\tilde{a}_n$ is the asymptotic mean for $\sqrt{n}\dot{S}_n^B$, that is,

$$\mathrm{E}\sqrt{n}\dot{S}_n^B = \tilde{a}_n + o(1). \tag{A.45}$$

By definition, we have

$$\sqrt{n}\mathrm{E}\dot{S}_n^B = \sqrt{n}\int_{x\in\dot{F}_0} \mathrm{E}\tau_N(x)\{I(\tau_N(x)\geq 0) - I(\tau_0^B(x_B)\geq 0)\}dx.$$

Observe that on the set $\dot{F}_0$, we have $\mathrm{E}\tau_N(x) = O(h^s)$, almost surely. By Assumption (A6), we obtain

$$\sup_{x\in\dot{F}_0} |\sqrt{n}\mathrm{E}\tau_N(x)| = O(\sup_{x\in\dot{F}_0} |\sqrt{n}\mathrm{E}\tau_N(x)|) = O(\sqrt{n}h^s) = o(1).$$

This further suggests

$$\sqrt{n}\mathrm{E}\dot{S}_n^B = \sqrt{n}\int_{x\in\dot{F}_0} \mathrm{E}\max(\tau_N(x), 0)dx + o(1). \tag{A.46}$$

Using similar arguments in proofs of Lemma 2.3 of Giné et al. (2003) and Lemma B.8 of Chang et al. (2015), we can show

$$\begin{aligned}
&\left|\mathrm{E}\max\left(\frac{\tau_N(x) - \mathrm{E}\tau_N(x) + \mathrm{E}\tau_N(x)}{\sqrt{\mathrm{Var}(\tau_N(x))}}, 0\right) - \mathrm{E}\max\left(\mathbb{Z}_1 + \frac{\mathrm{E}\tau_N(x)}{\sqrt{\mathrm{Var}(\tau_N(x))}}, 0\right)\right| \\
&\leq O\left\{\frac{1}{\sqrt{nh^{3p}}}\mathrm{E}\left|\left(\frac{A}{\pi(X)} - \frac{1-A}{1-\pi(X)}\right)YK\left(\frac{x-X}{h}\right)\right|^3\right\} = O\left(\frac{1}{\sqrt{nh^p}}\right),
\end{aligned} \tag{A.47}$$

where the last equality is due to Lemma A.5.

By Lemma A.3, we have $\sup_x |\mu_n(x) - \mu(x)| = o(1)$. It follows from Assumption (A5) that $\sup_x |\mu(x)| = O(1)$. Therefore, we have $\sup_x |\mu_n(x)| = O(1)$. Combining this with (A.47) yields

$$\begin{aligned}
&\sqrt{n}\left|\mathrm{E}\max(\tau_N(x), 0) - \mathrm{E}\left\{\max\left(\sqrt{\mathrm{Var}(\tau_N(x))}\mathbb{Z}_1 + \mathrm{E}\tau_N(x), 0\right)\right\}\right| \\
&= O\left(\frac{1}{\sqrt{nh^p}}\right).
\end{aligned} \tag{A.48}$$

By Assumption (A6), we obtain RHS of (A.48) is $o(1)$.

Besides, standard bias calculation suggests

$$\mathrm{E}\tau_N(x) = O(h^s), \qquad \forall x \in \dot{E}_7.$$

Together with Assumption (A6), this further implies

$$\begin{aligned} & \sqrt{n}\left|\mathrm{E}\max\left(\sqrt{\mathrm{Var}(\tau_N(x))}Z_1 + \mathrm{E}\tau_N(x), 0\right) - \mathrm{E}\max\left(\sqrt{\mathrm{Var}(\tau_N(x))}\mathbb{Z}_1, 0\right)\right| \\ = \;& O(\sqrt{n}h^s) = o(1), \end{aligned}$$

Hence, we obtain

$$\sqrt{n}\left|\mathrm{E}\max(\tau_N(x), 0) - \mathrm{E}\max\left(\sqrt{\mathrm{Var}(\tau_N(x))}\mathbb{Z}_1, 0\right)\right| = o(1). \tag{A.49}$$

With some calculations, we can show

$$\begin{aligned} n\mathrm{Var}(\tau_N(x)) &= \mathrm{Var}\left[\frac{1}{h^p}\sum_{i=1}^{\eta}\left\{\left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i}\right)Y_i K\left(\frac{x-X_i}{h}\right)\right\}\right] \\ &= \frac{1}{h^{2p}}\mathrm{E}\left\{\left(\frac{A_0}{\pi_0} - \frac{1-A_0}{1-\pi_0}\right)^2 Y_0^2 K^2\left(\frac{x-X_0}{h}\right)\right\} = \frac{\mu_n(x)}{h^p}. \end{aligned}$$

This together with (A.48) and (A.49) gives

$$\sqrt{n}\int_{x\in\dot{F}_0}\mathrm{E}\max(\tau_N(x), 0)dx = \frac{1}{\sqrt{h^p}}\int_{x\in\dot{F}_0}\sqrt{\mu_n(x)}dx. \tag{A.50}$$

Combining (A.50) together with (A.46) proves (A.45).

It remains to show the asymptotic normality of $(\dot{S}_N^B - \mathrm{E}\dot{S}_N^B)/\sqrt{\mathrm{Var}(\dot{S}_N^B)}$. Define

$$N_1 = \sum_{i=1}^{N} I(X_i \in F_0 + hU),$$
$$N_2 = \sum_{i=1}^{N} I(X_i \notin F_0 + hU),$$

where $U = [-1,1]^p$, and $\alpha_n = \Pr(X_0 \in F_0 + hU)$. Since h is monotonically decreasing, α_n is monotonically decreasing and bounded. Thus, it converges to some α.

We further define

$$U_n = \frac{N_1 - n\alpha_n}{\sqrt{n}} \text{ and } V_n = \frac{N_2 - n(1-\alpha_n)}{\sqrt{n}}.$$

Using similar arguments in the proof of Theorem B.2 in Chang et al. (2015), we can show (i) the random vector $(\dot{S}_N^B, U_n)$ is asymptotically normal; (ii) $\dot{S}_N^B$ and U_n are asymptotically uncorrelated. Besides, Lemma A.1 implies that $(\dot{S}_N^B, U_n)$ is independent of V_n. Since $\nu(F_0) > 0$, we have $\alpha > 0$. By Condition (A7), we have $\nu(\Omega^W \cap F_0^c) > 0$. This means $\alpha < 1$. Hence, it follows from Lemma A.6 that $\dot{S}_n^B$ is asymptotically normally distributed. This completes the proof for the second step.

Step 3. Finally, we show $\hat{a}_n = \tilde{a}_n + o_p(1)$, $\hat{\sigma}_n = \tilde{\sigma} + o_p(1)$. By definition, we have

$$\begin{aligned} |\hat{a}_n - \tilde{a}_n| &\leq \frac{1}{\sqrt{2\pi h^p}} \int_{x \in \hat{F} \Delta F_0} \sqrt{\mu_n(x)} dx \\ + \quad & \frac{1}{\sqrt{2\pi h^p}} \int_{x \in \hat{F}} |\sqrt{\mu_n(x)} - \sqrt{\hat{\mu}_n(x)}| dx \equiv I_1 + I_2, \end{aligned}$$

where Δ stands for the difference of two sets.

Boundedness of $\mu(\cdot)$ in Assumption (A5) implies that $\mu_n(\cdot)$ is bounded on Ω. Thus we have

$$I_1 = O\left(\frac{1}{\sqrt{h^p}} \nu(\hat{F} \Delta F)\right) = O\left(\frac{1}{\sqrt{h^p}} \nu(\hat{F}_1^c \Delta F_1)\right) + O\left(\frac{1}{\sqrt{h^p}} \nu(\hat{F}_2 \Delta F_2^c)\right), \tag{A.51}$$

where $F_1 = \{x \in \Omega : \tau_0(x) = 0\}$, $F_2 = \{x \in \Omega : \tau_0^B(x_B) \neq 0\}$, $\hat{F}_1 = \{x \in \Omega : |\tau_n(x)/\hat{f}(x)| \leq \eta_n\}$ and $\hat{F}_2 = \{x \in \Omega : |\tau_n^B(x_B)/\hat{f}^B(x_B)| > \eta_n\}$.

In the following, we show with probability going to 1, $\nu(\hat{F}_1^c \Delta F_1) \to o(\sqrt{h^p})$. Similarly we have $\nu(\hat{F}_2^c \Delta F_2) \to o(\sqrt{h^p})$ and hence (A.51) holds.

Define $\tilde{F}_1 = \{x \in \Omega : |\tau_0(x)| \leq 2\eta_n\}$. Observe that

$$\begin{aligned} \nu(\hat{F}_1^c \Delta F_1) &= \nu(\hat{F}_1^c \cap F_1) + \nu(F_1^c \cap \hat{F}_1) \\ &\leq \nu(\hat{F}_1^c \cap F_1) + \nu(\tilde{F}_1^c \cap \hat{F}_1) + \nu(F_1^c \cap \tilde{F}_1). \end{aligned} \tag{A.52}$$

Note that $\eta_n \gg \sqrt{\log n}/\sqrt{nh^p}$. Under the event defined in (A.32), we have $\nu(\hat{F}_1^c \cap F_1) = 0$ and $\nu(\tilde{F}_1^c \cap \hat{F}_1) = 0$. Besides, it follows from Assumption (A7) that $\nu(F_1^c \cap \tilde{F}_1) = O(\eta_n^{\xi_1})$. This together with (A.52) implies $\nu(\hat{F}_1^c \Delta F_1) = O(\eta_n^{\xi_1})$.

By Assumption (A8), we have $n\eta_n^{2\xi_1+2} \to 0$. Since $\eta_n \gg \sqrt{\log n}/\sqrt{nh^p}$, this suggests

$$n\eta_n^{2\xi_1} \frac{\log n}{nh^p} = \frac{\eta^{2\xi_1} \log n}{h^p} \to 0,$$

and hence $\eta_n^{\xi_1}/\sqrt{h^p} \to 0$. This proves $\nu(\hat{F}_1^c \Delta F_1) = o(\sqrt{h^p})$.

Now we show $I_2 = o_p(1)$. Since

$$I_2 = \frac{1}{\sqrt{2\pi h^p}} \int_{x\in\hat{F}} |\sqrt{\mu_n(x)} - \sqrt{\hat{\mu}_n(x)}| dx, \tag{A.53}$$

similar to Lemma A.2, with probability 1, we can show

$$\sup_x |\hat{\mu}_n(x) - \mu_n(x)| = O\left(\sqrt{\log n}/\sqrt{nh^p}\right), \tag{A.54}$$

for all sufficiently large n. By Lemma A.3 and the condition $\inf_{x\in\Omega}\mu(x) > 0$ in (A5), we have $\liminf_n \inf_{x\in\Omega}\mu_n(x) > 0$. Under the event defined in (A.54), we have $\liminf_n \inf_{x\in\Omega}\hat{\mu}_n(x) > 0$. Therefore,

$$|\sqrt{\hat{\mu}_n(x)} - \sqrt{\mu_n(x)}| = \frac{|\hat{\mu}_n(x) - \mu_n(x)|}{|\sqrt{\hat{\mu}_n(x)} + \sqrt{\mu_n(x)}|} = O\left(\frac{\sqrt{\log n}}{\sqrt{nh^p}}\right),$$

Since $n \gg \log n/h^{2p}$, in view of (A.53), we have $I_2 = o_p(1)$. Therefore, we've shown $\hat{a}_n = \tilde{a}_n + o_p(1)$. Similarly, we can show $\hat{\sigma}_n = \tilde{\sigma} + o_p(1)$. This completes the proof.

Case 2: Define

$$\mathcal{A}_1 = \left\{\inf_{x\in\Omega}\hat{\mu}_n(x) \geq \frac{c_2}{4} \text{ and } \sup_{x\in\Omega}\hat{\mu}_n(x) \leq 4c_3\right\}.$$

By (A.54), Lemma A.3 and Assumption (A5), we can show that there exist some positive constants c_2 and c_3 such that

$$\Pr(\mathcal{A}_1) \to 1. \tag{A.55}$$

Let $\mathcal{A}_2$ be the intersection of events defined in (A.15), (A.32) and (A.33). On the set $\mathcal{A}_2$, it follows from the conditions $\eta_n \gg \sqrt{\log n}/\sqrt{nh^p}$, $\eta_n \gg \sqrt{\log n}/\sqrt{n(h_B)^{p_B}}$ that for some arbitrary small ϵ and sufficiently large n, we have

$$\begin{aligned} \tilde{F} &= \left\{x : |\tau_0(x)| \leq \frac{\eta_n}{1+\epsilon}, |\tau_0^B(x_B)| \geq \frac{\eta_n}{1-\epsilon}\right\} \\ \subseteq \quad \hat{F} &= \left\{x : \left|\frac{\tau_n(x)}{\hat{f}(x)}\right| \leq \eta_n, \left|\frac{\tau_n^B(x_B)}{\hat{f}^B(x_B)}\right| > \eta_n\right\}. \end{aligned}$$

On the set $\mathcal{A}_1 \cup \mathcal{A}_2$, we have

$$\hat{a}_n = \frac{1}{\sqrt{2\pi h^p}} \int_{x\in\hat{F}} \sqrt{\hat{\mu}_n(x)} dx \geq \frac{1}{\sqrt{2\pi h^p}} \int_{x\in\hat{F}} \frac{\sqrt{c_2}}{2} dx \geq \frac{\sqrt{c_2}}{4\sqrt{\pi h^p}}\nu(\hat{F}) \geq \frac{\sqrt{c_2}}{4\sqrt{\pi h^p}}\nu(\tilde{F}).$$

Since $\nu(F_0) = 0$, we have $\nu(\tilde{F}) = \nu(\check{F})$, where

$$\check{F} = \left\{ x : 0 < |\tau_0(x)| \le \frac{\eta_n}{1+\epsilon}, |\tau_0^B(x_B)| \ge \frac{\eta_n}{1-\epsilon} \right\}.$$

By Assumption (A7), we have for sufficiently small ϵ that

$$\nu(\check{F}) \ge \bar{c}_0 \left(\frac{\eta_n}{1+\epsilon} \right)^{\xi_0}. \tag{A.56}$$

On the other hand, by (A.16), we have almost surely that

$$\sqrt{n}\widetilde{S}_n^B = \sqrt{n} \int_{x \in E_1 \cup E_2} \tau_n(x)\{d_n(x) - d_n^B(x_B)\} I(x \notin \hat{E}) dx. \tag{A.57}$$

We decompose RHS of (A.57) into

$$\begin{aligned} &\sqrt{n} \int_{x \in E_1} \tau_n(x)\{d_n(x) - d_n^B(x_B)\} I(x \notin \hat{E}) dx \\ +&\sqrt{n} \int_{x \in E_2 \cap E_1^c} \tau_n(x)\{d_n(x) - d_n^B(x_B)\} I(x \notin \hat{E}) dx \triangleq J_1 + J_2. \end{aligned}$$

Since $\nu(F_0) = 0$, we have

$$\begin{aligned} J_1 &\le \sqrt{n} \int_{x \in E_3} \tau_n(x)\{d_n(x) - d_n^B(x_B)\} I(x \notin \hat{E}) dx \\ &\le \sqrt{n} \int_{x \in E_3} |\tau_n(x)| dx, \end{aligned} \tag{A.58}$$

where the definition of E_3 is given in (A.17).

Consider J_2. On the set E_1^c, we have $\tau_0(x) \ne 0$. Under the null, for any $\tau_0(x) \ne 0$, we have $\tau_0^B(x_B) \ne 0$. This implies

$$\begin{aligned} J_2 &\le \sqrt{n} \int_{x \in E_2 \cap E_1^c} |\tau_n(x)| dx \le \sqrt{n} \int_{x \in E_4 \cap E_1^c} |\tau_n(x)| dx \\ &\le \sqrt{n} \int_{x \in E_4} |\tau_n(x)| dx, \end{aligned} \tag{A.59}$$

where the definition of E_4 is given in (A.17).

It follows from (A.57)-(A.59) that

$$\sqrt{n}\widetilde{S}_n^B \le \sqrt{n} \int_{x \in E_3 \cup E_4} |\tau_n(x)| dx. \tag{A.60}$$

By (A.19) and (A.20), on the set $\mathcal{A}_2$, we have

$$\int_{x \in E_3 \cup E_4} |\tau_n(x)| dx = O\left(\frac{(\log n)^{\xi_0/2+1/2}}{n^{\xi_0/2} h^{\xi_0/2+p/2}} \right)$$
$$+ \quad O\left(\frac{(\log n)^{\xi_0/2+1/2}}{n^{\xi_0/2} (h_B)^{p_B \xi_0/2} h^{p/2}} \right) = O\left(\frac{(\log n)^{\xi_0/2+1/2}}{n^{\xi_0/2} (h)^{p\xi_0/2+p/2}} \right),$$

where the last equality is due to the condition $h^p \asymp h_B^{p_B}$ in (A6). This together with (A.60) gives

$$\sqrt{n}\widetilde{S}_n^B = O\left(\frac{(\log n)^{\xi_0/2+1/2}}{n^{\xi_0/2} (h)^{p\xi_0/2+p/2}} \right). \tag{A.61}$$

It follows from Assumption (A8) that RHS of (A.61) is smaller than RHS of (A.56), for sufficiently large n. Therefore, we have

$$\Pr(\sqrt{n}\widetilde{S}_n^B \le \hat{a}_n) \ge \Pr(\mathcal{A}_1 \cup \mathcal{A}_2) \to 1. \tag{A.62}$$

This completes the proof.

Case 3: By definition, we have $\hat{a}_n(\Omega) \ge \hat{a}_n(\hat{F})$. Similar to (A.62), we can show

$$\Pr(\sqrt{n}\widetilde{S}_n^B \le \hat{a}_n(\Omega)) \ge \Pr(\sqrt{n}\widetilde{S}_n^B \le \hat{a}_n) \to 1.$$

The proof is hence completed.

A.4 Proof of Theorem 3.8. Under the alternative, according to the definition of CQTE, there must exists some $x_{B_0} \in \Omega^B$, $x_{C_1} \neq x_{C_2} \in \Omega^C$ such that $\tau_0(x_{B_0}, x_{C_1}) > 0$ and $\tau_0(x_{B_0}, x_{C_2}) < 0$. It follows from the continuity of τ_0 that there exists some positive constants ε and δ, such that for all $\|x_B - x_{B_0}\|_\infty \le \delta$ and $\|x_C - x_{C_1}\|_\infty \le \delta$, that

$$\min \tau_0(x_B, x_C) f(x_B, x_C) \ge \varepsilon, \tag{A.63}$$

and for all $\|x_B - x_{B_0}\|_\infty \le \delta$, $\|x_C - x_{C_2}\|_\infty \le \delta$,

$$\max \tau_0(x_B, x_C) f(x_B, x_C) \le -\varepsilon. \tag{A.64}$$

Since $\sqrt{\log n}/\sqrt{nh^p} \to 0$, $\sqrt{\log n}/\sqrt{n(h_B)^{p_B}} \to 0$, under the event defined in (A.15), it follows from (A.63) and (A.64) that for sufficiently large n,

$$\min \tau_n(x_B, x_C) \ge \frac{\varepsilon}{2}, \quad \forall \|x_B - x_{B_0}\|_\infty \le \delta, \|x_C - x_{C_1}\|_\infty \le \delta, \tag{A.65}$$

and that

$$\text{(A.66)} \quad \max \tau_n(x_B, x_C) \leq -\frac{\varepsilon}{2}, \quad \forall \|x_B - x_{B_0}\|_\infty \leq \delta, \|x_C - x_{C_2}\|_\infty \leq \delta.$$

This implies for any fixed x_B such that $\|x_B - x_{B_0}\|_\infty \leq \delta$, we have

$$\text{(A.67)} \quad I(\tau_n(x_B, x_{C_1}) \geq 0) \neq d_n^B(x_B) \,\text{or}\, I(\tau_n(x_B, x_{C_2}) \geq 0) \neq d_n^B(x_B).$$

Define $E_0 = \{x : \|x_B - x_{0,B}\|_\infty \leq \delta, \|x_C - x_{C_1}\|_\infty \leq \delta \text{ or } \|x_C - x_{C_2}\|_\infty \leq \delta\}$. Combining (A.65), (A.66) with (A.67), on the set (A.32), we have

$$\begin{aligned} \text{(A.68)} \quad & \sqrt{n}\int_{x\in E_0} \tau_n(x)\{I(\tau_n(x) \geq 0) - I(\tau_n^B(x_B) \geq 0)\}I(x \notin \hat{E})dx \\ = \; & \sqrt{n}\int_{x\in E_0} \tau_n(x)\{I(\tau_n(x) \geq 0) - I(\tau_n^B(x_B) \geq 0)\}dx \geq \sqrt{n}\frac{\varepsilon}{2}\delta^p, \end{aligned}$$

where the equality is due to that $|\tau_n(x)| \geq \varepsilon/2 \gg \eta_n$, for sufficiently large n. Observe that

$$\sqrt{n}\widetilde{S}_n^B \geq \sqrt{n}\int_{x\in E_0} \tau_n(x)\{d_n(x) - d_n^B(x_B)\}I(x \notin \hat{E})dx.$$

This together with (A.68) implies

$$\text{(A.69)} \quad \sqrt{n}\widetilde{S}_n^B \geq c\sqrt{n},$$

for some constant $c > 0$.

On the other hand, note that $\hat{a}_n(\hat{F}) \leq \hat{a}_n(\Omega)$ and $\hat{\sigma}_n^2(\hat{F}) \leq \hat{\sigma}_n^2(\Omega)$. Using the same arguments in proving (A.55), we can show that the following event holds with probability tending to 1,

$$\text{(A.70)} \quad \sup_x \hat{\mu}_n(x) = O(1).$$

Under the event defined in (A.70), we have

$$\text{(A.71)} \quad \hat{a}(\Omega) = \frac{1}{\sqrt{2\pi h^p}}\int_{x\in\Omega} \sqrt{\hat{\mu}_n(x)}dx = O\left(\frac{1}{\sqrt{h^p}}\right) \ll \sqrt{n},$$

and

$$\begin{aligned} \hat{\sigma}_n^2(\Omega) = \; & \int_{x^W\in\Omega}\int_{t\in[-1,1]^p} \hat{\mu}_n(x^W)\mathrm{cov}(\max\{\sqrt{1-\rho^2(t)}\mathbb{Z}_1 + \rho(t)\mathbb{Z}_2, 0\}, \\ \text{(A.72)} \quad & \max\{\mathbb{Z}_2, 0\})d\nu(x^W)dt = O(1). \end{aligned}$$

Combining (A.69) with (A.71) and (A.72), for any α, we have

$$\Pr\left(\frac{\sqrt{n}\widetilde{S}_n^B - \hat{a}_n(\hat{F})}{\hat{\sigma}_n(\hat{F})} \geq z_\alpha\right) \leq \Pr\left(\frac{\sqrt{n}\widetilde{S}_n^B - \hat{a}_n(\Omega)}{\hat{\sigma}_n(\Omega)} \geq z_\alpha\right) \to 1.$$

The proof is hence completed.

A.5 Proof of Theorem 3.14. Define

$$\hat{\tau}(x) = \frac{1}{nh^p}\sum_{i=1}^n \left\{ \left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i} \right) Y_i - n^{-1/2}\delta_0(X_i) \right\} K_h(x - X_i),$$

$$\hat{\tau}^B(x_B) = \frac{1}{nh_B^{p_B}}\sum_{i=1}^n \left\{ \left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i} \right) Y_i - n^{-1/2}\delta_0(X_i) \right\} K_{h_B}^B(x_B - X_i^B).$$

Similar to Lemma A.2, under the local alternative H_a, we can show for sufficiently large n, almost surely,

(A.73)

$$|\hat{\tau}(x) - \tau_0(x)f(x)| = O\left(\frac{\sqrt{\log n}}{\sqrt{nh^p}} \right), |\hat{\tau}^B(x_B) - \tau_0^B(x_B)f^B(x_B)| = O\left(\frac{\sqrt{\log n}}{\sqrt{nh_B^{p_B}}} \right),$$

for all $x \in \Omega$. Since δ_0 is bounded on Ω, there exists some constant $\bar{C}$ such that

$$\begin{aligned} \sup_{x\in\Omega} |\hat{\tau}(x) - \tau_n(x)| &\leq \sup_x \frac{1}{nh^p}\sum_{i=1}^n |n^{-1/2}\delta_0(X_i)| \left| K\left(\frac{x - X_i}{h} \right) \right| \\ &\leq \bar{C}\frac{1}{n^{3/2}h^p}\sum_{i=1}^n \left| K\left(\frac{x - X_i}{h} \right) \right|. \end{aligned} \tag{A.74}$$

Using similar arguments in the proof of Lemma A.2, we can show for sufficiently large n,

$$\sup_{x\in\Omega} \frac{1}{nh^p}\sum_{i=1}^n \left| K\left(\frac{x - X_i}{h} \right) \right| = O(1) + O\left(\frac{\sqrt{\log n}}{\sqrt{nh^p}} \right), \tag{A.75}$$

almost surely. This together with (A.74) implies that almost surely,

$$\sup_{x\in\Omega} |\hat{\tau}(x) - \tau_n(x)| = O\left(\frac{1}{\sqrt{n}} \right), \tag{A.76}$$

for sufficiently large n.

Besides, the boundedness of δ_0 implies

$$|\tau_{n,0}(x) - \tau_0(x)| = O\left(\frac{1}{\sqrt{n}} \right), |\tau_{n,0}^B(x_B) - \tau_0^B(x_B)| = O\left(\frac{1}{\sqrt{n}} \right). \tag{A.77}$$

Combining (A.76) with (A.77) together, we have almost surely that

$$|\tau_n(x) - \tau_0(x)f(x)| = O\left(\frac{\sqrt{\log n}}{\sqrt{nh^p}} \right), |\tau_n^B(x_B) - \tau_0^B(x_B)f^B(x_B)| = O\left(\frac{\sqrt{\log n}}{\sqrt{n(h_B)^{p_B}}} \right),$$

for sufficiently large n.

Therefore, using similar arguments in the proof of Theorem 3.6, we can show that under H_a, $\sqrt{n}\widetilde{S}_n^B$ is equivalent to $\sqrt{n}\dot{S}_n^B$ defined in (A.37). The asymptotic normality of

$$\frac{\sqrt{n}\dot{S}_n^B - \sqrt{n}\mathrm{E}\dot{S}_N^B}{\tilde{\sigma}}$$

can be similarly established as in Theorem 3.6.

We now show

$$\mathrm{E}\sqrt{n}\dot{S}_N^B - \tilde{a}_n \to \frac{1}{2}\int_{x\in F_0} |\delta_0(x)| f(x) dx. \tag{A.78}$$

Standard bias calculation suggests $\mathrm{E}\tau_N(x) = \tau_{n,0}(x)f(x) + O(h^s)$. On the set F_0, we have $\tau_{n,0}(x) = n^{-1/2}\delta_0(x)$. Since $\sqrt{n}h^s \to 0$ according to Assumption (A6), we obtain

$$\sup_{x\in\Omega} |\sqrt{n}\mathrm{E}\tau_N(x) d_{opt}^B(x_B) - \delta_0(x)f(x)d_{opt}^B(x_B)| = o(1). \tag{A.79}$$

Similar to (A.48), we can show under the local alternative,

$$\begin{aligned}\sqrt{n}\left|\mathrm{E}\max(\tau_N(x),0) - \mathrm{E}\left\{\max\left(\sqrt{\mathrm{Var}(\tau_N(x))}\mathbb{Z}_1 + \mathrm{E}\tau_N(x), 0\right)\right\}\right| \\ = O\left(\frac{1}{\sqrt{n}h^p}\right).\end{aligned} \tag{A.80}$$

Since $\mathrm{E}\tau_N(x) = n^{-1/2}\delta_0(x)f(x) + o(1/\sqrt{n}), \forall x \in F_0$ and $\mathrm{Var}(\tau_N(x)) = \mu_n(x)/(nh^p)$, it follows from (A.80) and the condition $nh^{2p} \to 0$ that

$$\begin{aligned}\sqrt{n}\left|\mathrm{E}\max(\tau_N(x),0) - \mathrm{E}\left\{\max\left(\sqrt{\mu_n(x)/(nh^p)}\mathbb{Z}_1 + \frac{\delta_0(x)f(x)}{\sqrt{n}}, 0\right)\right\}\right| \\ = o(1).\end{aligned} \tag{A.81}$$

Let $\kappa(x) = \Pr\{\mathbb{Z}_1 > \max(-\delta_0(x)f(x)h^{p/2}/\sqrt{\mu_n(x)}, 0)\}$. By Lemma A.9, we have

$$\begin{aligned}&\left|\mathrm{E}\left\{\max\left(\sqrt{\frac{\mu_n(x)}{h^p}}\mathbb{Z}_1 + \delta_0(x)f(x), 0\right)\right\} - \mathrm{E}\left\{\max\left(\sqrt{\frac{\mu_n(x)}{h^p}}\mathbb{Z}_1, 0\right)\right\} - \delta_0(x)f(x)\kappa(x)\right| \\ &\le \sqrt{\frac{\mu_n(x)}{h^p}}\left|\mathrm{E}\left\{\max\left(\mathbb{Z}_1 + \frac{\delta_0(x)f(x)h^{p/2}}{\sqrt{\mu_n(x)}}, 0\right)\right\} - \mathrm{E}\left\{\max\left(\mathbb{Z}_1, 0\right)\right\} - \frac{\delta_0(x)f(x)\kappa(x)h^{p/2}}{\sqrt{\mu_n(x)}}\right| \\ &\qquad\le c^*\sqrt{\frac{\mu_n(x)}{h^p}}\left|\frac{\delta_0(x)f(x)\kappa(x)h^{p/2}}{\sqrt{\mu_n(x)}}\right|^2 = \frac{c^*\delta_0^2(x)f^2(x)\kappa(x)h^{p/2}}{\sqrt{\mu_n(x)}},\end{aligned}$$

for some constant $c^* > 0$. By Lemma A.3, we have $\mu_n(x) \to \mu(x)$. Since $\inf_{x\in\Omega} \mu(x) > 0$, we have $\lim\inf_n \inf_{x\in\Omega} \mu_n(x) > 0$. This implies $\{\inf_n \mu_n(x)\}/h^p \to \infty$. Since $\delta_0(\cdot)$, $f(\cdot)$, $\kappa(\cdot)$ are bounded, we have

$$\sup_x \left| \mathrm{E}\left\{ \max\left(\sqrt{\frac{\mu_n(x)}{h^p}}\mathbb{Z}_1 + \delta_0(x)f(x), 0 \right) \right\} \right.$$
$$\left. - \quad \mathrm{E}\left\{ \max\left(\sqrt{\frac{\mu_n(x)}{h^p}}\mathbb{Z}_1, 0 \right) \right\} - \delta_0(x)f(x)\kappa(x) \right| = o(1).$$

Notice that $\kappa(x) \to \Pr(\mathbb{Z}_1 > 0) = 1/2$. Therefore,

$$\sup_x \left| \mathrm{E}\left\{ \max\left(\sqrt{\frac{\mu_n(x)}{h^p}}\mathbb{Z}_1 + \delta_0(x)f(x), 0 \right) \right\} \right. \tag{A.82}$$
$$\left. - \quad \mathrm{E}\left\{ \max\left(\sqrt{\frac{\mu_n(x)}{h^p}}\mathbb{Z}_1, 0 \right) \right\} - \frac{\delta_0(x)f(x)}{2} \right| = o(1).$$

This together with (A.81) gives

$$\left| \sqrt{n}\mathrm{E}\max(\tau_N(x), 0) - \mathrm{E}\left\{ \max\left(\sqrt{\mu_n(x)/h^p}\mathbb{Z}_1, 0 \right) \right\} - \frac{\delta_0(x)f(x)}{2} \right| = o(1).$$

It follows from dominated convergence theorem that

$$\mathrm{E}\sqrt{n}\dot{S}_n^B - \tilde{a}_n \to \int_{x\in F_0} \delta_0(x) \left(\frac{1}{2} - I(\tau_0^B(x_B) \geq 0) \right) f(x)dx. \tag{A.83}$$

By assumption, we have $\delta_0(x) \geq 0$ if $\tau_0^B(x_B) \leq 0$, and $\delta_0(x) \leq 0$ if $\tau_0^B(x_B) \geq 0$. Therefore,

$$\int_{x\in F_0} \delta_0(x) \left(\frac{1}{2} - I(\tau_0^B(x_B) \geq 0) \right) f(x)dx = \frac{1}{2}\int_{x\in F_0} |\delta_0(x)| f(x)dx. \tag{A.84}$$

Combining (A.83) together with (A.84), we obtain

$$\mathrm{E}\sqrt{n}\dot{S}_n^B - \tilde{a}_n \to \frac{1}{2}\int_{x\in F_0} |\delta_0(x)| f(x)dx.$$

Similarly, under the local alternative, we can show $\hat{a}_n$ and $\hat{\sigma}_n$ are consistent estimators for $\tilde{a}_n$ and $\tilde{\sigma}$, respectively. The proof is thus completed.

A.6 Proof of Theorem 4.3. Define $\mu_{n,DR}(x)$ to be

$$\frac{1}{h^p}\mathrm{E}\left[\left\{\left(\frac{A}{\pi(X,\alpha_0)}-\frac{1-A}{1-\pi(X,\alpha_0)}\right)Y-\left(\frac{A}{\pi(X,\alpha_0)}-1\right)\Phi_1(X,\theta_0)\right.\right.$$
$$\left.\left.+\left(\frac{1-A}{1-\pi(X,\alpha_0)}-1\right)\Phi_0(X,\zeta_0)\right\}^2|X=x\right]\left\{K\left(\frac{(x-X)}{h}\right)\right\}^2.$$

Let $\tilde{a}_{n,DR}=\int_{x\in F}\sqrt{\mu_{n,DR}(x)}dx/(\sqrt{2\pi h^p})$ and

$$\tilde{\sigma}^2_{DR}=\int_{x\in F}\int_{t\in[-1,1]^p}\mu_{DR}(x)\mathrm{cov}(\max\{\sqrt{1-\rho^2(t)}\mathbb{Z}_1+\rho(t)\mathbb{Z}_2,0\},\max\{\mathbb{Z}_2,0\})dxdt.$$

$\tilde{a}_{n,DR}$ and $\tilde{\sigma}^2_{DR}$ correspond to the asymptotic bias and variance of $\widetilde{S}^B_{n,DR}$ under H_0.

We only show the asymptotic distribution under H_0 for brevity. We first argue the following holds:

$$\sup_{x\in\Omega}|\tau_{n,DR}(x)-\tilde{\tau}_{n,DR}(x)|=o_p\left(\frac{1}{\sqrt{n}}\right). \tag{A.85}$$

where

$$\tilde{\tau}_{n,DR}(x)=\frac{1}{nh^p}\sum_{i=1}^n\left[\left\{\frac{A_i}{\pi(X_i,\alpha_0)}Y_i-\left(\frac{A_i}{\pi(X_i,\alpha_0)}-1\right)\Phi_1(X_i,\zeta_0)\right\}\right.$$
$$-\left.\left\{\frac{1-A_i}{1-\pi(X_i,\alpha_0)}Y_i-\left(\frac{1-A_i}{1-\pi(X_i,\alpha_0)}-1\right)\Phi_0(X_i,\theta_0)\right\}\right]K_h(x-X_i)$$
$$-\quad v_1(x)f(x)\frac{1}{n}\sum_{i=1}^n\xi_1(O_i)-v_2(x)f(x)\frac{1}{n}\sum_{i=1}^n\xi_2(O_i)+v_3(x)f(x)\frac{1}{n}\sum_{i=1}^n\xi_3(O_i),$$

and

$$v_1(x)=\mathrm{E}\left\{\left(\frac{A}{\pi_0^2}\{Y-\Phi_1(X,\zeta_0)\}+\frac{1-A}{(1-\pi_0)^2}\{Y-\Phi_0(X,\theta_0)\}\right)\left(\frac{\partial\pi_0}{\partial\alpha_0}\right)|X=x\right\},$$
$$v_2(x)=\mathrm{E}\left\{\left(\frac{A}{\pi_0}-1\right)\frac{\partial\Phi_1(X,\zeta)}{\partial\zeta_0}|X=x\right\},$$
$$v_3(x)=\mathrm{E}\left\{\left(\frac{1-A}{1-\pi_0}-1\right)\frac{\partial\Phi_0(X,\theta)}{\partial\theta_0}|X=x\right\}.$$

where π_0 is a shorthand for $\pi(X,\alpha_0)$.

Let $\pi_i=\pi(X_i,\alpha_0)$, $\Phi_{1,i}=\Phi_1(X_i,\zeta_0)$ and $\Phi_{0,i}=\Phi_0(X_i,\theta_0)$. With some calculations, we can show

$$\sup_{x\in\Omega}|\tau_{n,DR}(x)-\tilde{\tau}_{n,DR}(x)|\le I_1+I_2+I_3,$$

where

$$
\begin{aligned}
I_1 &= \sup_{x\in\Omega} \frac{1}{nh^p}\left|\sum_{i=1}^n \left\{\left(\frac{A_i(1-A_i)}{\pi_i\{1-\pi_i\}} - \frac{A_i(1-A_i)}{\hat{\pi}_i(1-\hat{\pi}_i)}\right) Y_i - \left(\frac{A_i}{\pi_i} - \frac{A_i}{\hat{\pi}_i}\right)\Phi_{1,i}\right.\right. \\
&+ \left.\left.\left(\frac{1-A_i}{1-\pi_i} - \frac{1-A_i}{1-\hat{\pi}_i}\right)\Phi_{0,i}\right\} K\left(\frac{x-X_i}{h}\right) - f(x)v_1^T(x)\sum_{i=1}^n \xi_1(O_i)\right|, \\
I_2 &= \sup_{x\in\Omega} \frac{1}{nh^p}\left|\sum_{i=1}^n \left\{\left(\frac{1-A_i}{1-\pi_i} - 1\right)\{\Phi_{0,i} - \Phi_0(X_i,\hat{\theta})\} - \left(\frac{A_i}{\pi_i}\right)\{\Phi_{1,i} - \Phi_1(X_i,\hat{\zeta})\}\right\}\right. \\
&\qquad \left. K_h(x-X_i) - f(x)v_2^T(x)\sum_{i=1}^n \xi_2(O_i) + f(x)v_3^T(x)\sum_{i=1}^n \xi_3(O_i)\right|, \\
I_3 &= \sup_{x\in\Omega} \frac{1}{nh^p}\left|\sum_{i=1}^n \left\{\left(\frac{1-A_i}{1-\pi_i} - \frac{1-A_i}{1-\hat{\pi}_i}\right)\{\Phi_{0,i} - \Phi_0(X_i,\hat{\theta})\}\right.\right. \\
&- \left.\left.\left(\frac{A_i}{\pi_i} - \frac{A_i}{\hat{\pi}_i}\right)\{\Phi_{1,i} - \Phi_1(X_i,\hat{\zeta})\}\ \right\} K\left(\frac{x-X_i}{h}\right)\right|.
\end{aligned}
$$

A second order Taylor expansion around α_0 yields that $I_1 \le \eta_1 + \eta_2 + \eta_3$, where η_1 is the linear first order term,

$$
\begin{aligned}
\eta_1 &= \sup_{x\in\Omega} \frac{1}{nh^p}\left|\sum_{i=1}^n \left\{\left(\frac{A_i}{\pi_i^2} + \frac{1-A_i}{(1-\pi_i)^2}\right) Y_i - \frac{A_i}{\pi_i^2}\Phi_{1,i} - \frac{1-A_i}{(1-\pi_i)^2}\Phi_{0,i}\right\}\right. \\
&\qquad \left.\left(\frac{\partial\pi}{\partial\alpha_0}\right)^T (\alpha_0 - \hat{\alpha}) K\left(\frac{x-X_i}{h}\right) - f(x)v_1^T(x)(\hat{\alpha} - \alpha_0)\right|,
\end{aligned}
$$

η_2 the second order remainder term,

$$
\begin{aligned}
\eta_2 &= \sup_{x\in\Omega} \frac{1}{nh^p}\left|\sum_{i=1}^n \left\{\left(\frac{2A_i}{(\pi_i^*)^3} - \frac{2(1-A_i)}{(1-\pi_i^*)^3}\right) Y_i - \left(\frac{2A_i}{(\pi_i^*)^3}\Phi_{1,i} - \frac{1-A_i}{(1-\pi_i^*)^3}\Phi_{0,i}\right)\right\}\right. \\
&\qquad \left.\left\{\left(\frac{\partial\pi}{\partial\alpha^*}\right)^T (\alpha_0 - \hat{\alpha})\right\}^2 K\left(\frac{x-X_i}{h}\right)\right| + \sup_{x\in\Omega} \frac{1}{nh^p}\left|\sum_{i=1}^n \left\{\left(\frac{A_i}{(\pi_i^*)^2} + \frac{1-A_i}{(1-\pi_i^*)^2}\right) Y_i\right.\right. \\
&- \left.\left.\frac{A_i}{(\pi_i^*)^2}\Phi_{1,i} - \frac{1-A_i}{(1-\pi_i^*)^2}\Phi_{0,i}\right\}(\alpha_0 - \hat{\alpha})^T \left(\frac{\partial^2\pi}{\partial\alpha^*\partial(\alpha^*)^T}\right)(\alpha_0 - \hat{\alpha}) K\left(\frac{x-X_i}{h}\right)\right|,
\end{aligned}
$$

and

$$
\eta_3 = \sup_{x\in\Omega}\left| f(x)v_1^T(x)(\hat{\alpha} - \alpha_0) - f(x)v_1^T(x)\frac{1}{n}\sum_i \xi_1(O_i)\right|,
$$

where π_i^* is a shorthand for $\pi(X_i, \alpha^*)$ with some vector α^* lying on the line segment jointing α_0 and $\hat{\alpha}$. By Assumptions (A4'), (A9) and (A10), we obtain

$$\sup_x |v_1(x)| < \infty. \tag{A.86}$$

Since $\sup_x f(x) < \infty$, it follows from Assumption (A13) that

$$\eta_3 = o_p(1/\sqrt{n}). \tag{A.87}$$

Using similar arguments in the proof of Lemma A.2, we can show that the following event holds almost surely for any sufficiently large n,

$$\max_{i \in \{1,\dots,n\}} |Y_i| \le c_n, \tag{A.88}$$

with $c_n = 3\log(n)/t$ for some constant $t > 0$ defined in (A3).

On the set $\hat{\alpha} = \alpha_0 + O(n^{-1/2})$, Assumption (A11) implies

$$\left\| \frac{\partial \pi}{\partial \alpha^*} \right\|_2 \text{ and } \left\| \frac{\partial^2 \pi}{\partial \alpha^* \partial (\alpha^*)^T} \right\|_2 \tag{A.89}$$

are uniformly bounded. Under Assumptions (A4') and (A13), combining (A.89) together with (A.88) suggests

$$\eta_2 = \frac{c_n}{n} O_p\left(\frac{1}{nh^p} \sum_{i=1}^n \sup_{x \in \Omega} \left| K\left(\frac{x - X_i}{h} \right) \right| \right). \tag{A.90}$$

It follows from (A.75) that almost surely for all sufficiently large n,

$$\frac{1}{nh^p} \sup_{x \in \Omega} \sum_{i=1}^n \left| K\left(\frac{x - X_i}{h} \right) \right| = O(1).$$

This together with (A.90) suggests

$$\eta_2 = O_p\left(\frac{c_n}{n} \right) = o_p\left(\frac{1}{\sqrt{n}} \right). \tag{A.91}$$

Now we show $\eta_1 = o_p(1/\sqrt{n})$. Since $\Phi_0(x, \theta_0)$, $\Phi_1(x, \zeta_0)$ and $\partial \pi(x, \alpha)/\partial \alpha_0$ are uniformly bounded for $x \in \Omega$, using similar arguments in the proof of Lemma A.2, we can show that

$$I_0 \triangleq \frac{1}{nh^p} \sup_{x \in \Omega} \left\| \sum_{i=1}^n \left\{ \left(\frac{A_i}{\pi_i^2} + \frac{1 - A_i}{(1 - \pi_i)^2} \right) Y_i - \frac{A_i}{\pi_i^2} \Phi_1(X_i, \zeta_0) \right.\right.$$
$$\left.\left. - \frac{1 - A_i}{(1 - \pi_i)^2} \Phi_0(X_i, \theta_0) \right\} \left(\frac{\partial \pi_i}{\partial \alpha_0} \right) K\left(\frac{x - X_i}{h} \right) - v_1(x) f(x) \right\|_2 = o(1),$$

almost surely for all sufficiently large n. This together with Assumption (A10) gives

$$\eta_1 \leq I_0 \|\hat{\alpha} - \alpha_0\|_2 = o_p(1/\sqrt{n}).$$

Combining this together with (A.91) and (A.87) implies $I_1 = o_p(1/\sqrt{n})$.

Similarly, we can show $I_2 = o_p(1/\sqrt{n})$, $I_3 = o_p(1/\sqrt{n})$. Thus, we obtain

$$\sup_{x\in\Omega} |\tau_{n,DR}(x) - \tilde{\tau}_{n,DR}(x)| = o_p\left(\frac{1}{\sqrt{n}}\right).$$

This verifies (A.85).

Using similar arguments in the proof of Lemma A.2, we can show the following event holds with probability 1 for all sufficiently large n,

$$\sup_{x\in\Omega} |\tilde{\tau}_{n,DR}(x) - \tau_0(x)f(x)| = O\left(\frac{\sqrt{\log n}}{\sqrt{nh^p}}\right).$$

This together with (A.85) implies

$$\sup_{x\in\Omega} |\tau_{n,DR}(x) - \tau_0(x)f(x)| = O\left(\frac{\sqrt{\log n}}{\sqrt{nh^p}}\right). \tag{A.92}$$

Similarly, we can show

$$\sup_{x\in\Omega} |\tau^B_{n,DR}(x) - \tau^B_0(x)f^B(x_B)| = O\left(\frac{\sqrt{\log n}}{\sqrt{nh_B^{p_B}}}\right). \tag{A.93}$$

By (A.92) and (A.93), using similar arguments in the proof of Theorem 3.6, we can show that $\sqrt{n}\widetilde{S}^B_{n,DR}$ is equivalent to

$$\sqrt{n}\int_{x\in F_0} \tau_{n,DR}(x)\{I(\tau_{n,DR}(x) \geq 0) - d^B_{opt}(x_B)\}dx. \tag{A.94}$$

By (A.85), we can show (A.94) is equivalent to $\sqrt{n}\dot{S}^B_{n,DR}$, where

$$\dot{S}^B_{n,DR} = \int_{x\in F_0} \tilde{\tau}_{n,DR}(x)\{I(\tilde{\tau}_{n,DR}(x) \geq 0) - d^B_{opt}(x_B)\}dx. \tag{A.95}$$

By Assumption (A11) and (A.86), we have $\sup_{x\in\Omega} \mathrm{E}|f(x)v_1(x)^T \sum_{i=1}^n \xi_1(O_i)|^2 =$

$O(n)$. Therefore,

$$
\begin{aligned}
&\text{(A.96)}\\
&\sup_{x\in\Omega}\left|\mathrm{E}|\tilde{\tau}_{n,DR}(x)|^2 - \mathrm{E}|\tilde{\tau}_{n,DR}(x) + f(x)v_1(x)^T\sum_i \xi_1(O_i)/n|^2\right|\\
\le\ & \frac{1}{n^2}\mathrm{E}\left|f(x)v_1(x)^T\sum_{i=1}^n \xi_1(O_i)\right|^2 + \left|\frac{2}{nh^p}\mathrm{E}\xi_1(O_0)\left\{\frac{A_0}{\pi_0}Y_0 - \left(\frac{A_0}{\pi_0}-1\right)\Phi_1(X_0,\zeta_0)\right\}\right.\\
-\ & \left.\left\{\frac{1-A_0}{1-\pi_0}Y_0 - \left(\frac{1-A_0}{1-\pi_0}-1\right)\Phi_0(X_0,\theta_0)\right\}K\left(\frac{x-X_0}{h}\right)\right| = O\left(\frac{1}{n}\right).
\end{aligned}
$$

Define

$$
\begin{aligned}
\check{\tau}_{n,DR}(x) &= \tilde{\tau}_{n,DR}(x) + \frac{1}{n}f(x)v_1(x)^T\sum_i \xi_1(O_i) + \frac{1}{n}f(x)v_2(x)^T\sum_i \xi_2(O_i)\\
&- \frac{1}{n}f(x)v_3(x)^T\sum_i \xi_3(O_i) = \frac{1}{nh^p}\sum_{i=1}^n\left[\left\{\frac{A_i}{\pi_i}Y_i - \left(\frac{A_i}{\pi_i}-1\right)\Phi_{1,i}\right\}\right.\\
&- \left.\left\{\frac{1-A_i}{1-\pi_i}Y_i - \left(\frac{1-A_i}{1-\pi_i}-1\right)\Phi_{0,i}\right\}\right]K\left(\frac{x-X_i}{h}\right)
\end{aligned}
$$

Similar to (A.96), we can show

$$
\sup_{x\in\Omega}\left|\mathrm{E}|\tilde{\tau}_{n,DR}(x)|^2 - \mathrm{E}|\check{\tau}_{n,DR}(x)|^2\right| = O\left(\frac{1}{n}\right). \tag{A.97}
$$

With some calculations, we have

$$
\mathrm{E}\{\sqrt{nh^p}\check{\tau}_{n,DR}(x)\}^2 = \mu_{n,DR}(x). \tag{A.98}
$$

Similar to Lemma A.3, we can show

$$
\sup_{x\in\Omega}|\mu_{n,DR}(x) - \mu_n(x)| = o(1). \tag{A.99}
$$

Let $\tilde{\tau}_{N,DR}(x)$ and $\check{\tau}_{N,DR}(x)$ be Poissonized versions of $\tilde{\tau}_{n,DR}(x)$ and $\check{\tau}_{n,DR}(x)$, respectively. It follows from (A.97), (A.98) and (A.99) that

$$
\begin{aligned}
\mathrm{Var}\left(\sqrt{nh^p}\tilde{\tau}_{N,DR}(x)\right) &= \mathrm{Var}\left(\sqrt{nh^p}\check{\tau}_{N,DR}(x)\right) + O(h^p)\\
&= \mu_{n,DR}(x) + O(h^p) \to \mu_{DR}(x).
\end{aligned} \tag{A.100}
$$

Therefore, using similar arguments in the proof of Theorem 3.6, we can show $\tilde{\sigma}^2_{DR}$ is the asymptotic variance for $\sqrt{n}\dot{S}^B_{N,DR}$.

For the asymptotic mean, note that for all $x \in F_0$, we have

$$\sup_{x\in F} |\mathrm{E}\tilde{\tau}_{N,DR}(x)| = \sup_{x\in\Omega} \left|\mathrm{E}\tau_0(X_0)K\left(\frac{x-X_0}{h}\right)\right| = O(h^s) = o\left(\frac{1}{\sqrt{n}}\right).$$

This suggests

$$\mathrm{E}\sqrt{n}\int_{x\in F_0} \tilde{\tau}_{N,DR}(x)I(\tau_0^B(x_B)\geq 0)dx = o(1). \tag{A.101}$$

Similar to (A.49), under Assumption (A13) we can show

$$\sup_{x\in F_0}\sqrt{n}\left|\mathrm{E}\max(\tilde{\tau}_{N,DR}(x),0) - \mathrm{E}\max\left(\sqrt{\mathrm{Var}(\tilde{\tau}_{N,DR}(x))}\mathbb{Z}_0, 0\right)\right| = o(1), \tag{A.102}$$

for a standard normal random variable $\mathbb{Z}_0$. Combining (A.102) together with (A.100) suggests

$$\mathrm{E}\sqrt{n}\int_{x\in F_0}\max(\tilde{\tau}_{N,DR}(x),0)dx = \frac{1}{\sqrt{h}}\int_{x\in F_0}\sqrt{\mu_n(x)}\mathrm{E}\max(\mathbb{Z}_0,0)dx.$$

This together with (A.101) shows $\tilde{a}_{n,DR}$ is the asymptotic mean for $\sqrt{n}\dot{S}^B_{N,DR}$.

Asymptotic normality of

$$\frac{\sqrt{n}\dot{S}^B_{N,DR} - \tilde{a}_{n,DR}}{\tilde{\sigma}_{DR}}$$

can be proven using similar arguments as in Theorem 3.6. It follows from Lemma A.6 that $(\sqrt{n}\dot{S}^B_{n,DR} - \tilde{a}_{n,DR})/\tilde{\sigma}_{DR}$ is asymptotically normally distributed.

Under Assumptions (A9), (A10) and (A11), $\hat{\mu}_{n,DR}(x)$ is a consistent estimator for $\mu_{n,DR}(x)$. Similar to the proof of Theorem 3.6, we can show $\hat{a}_{n,DR}$ and $\hat{\sigma}_{n,DR}$ are consistent estimators for $\tilde{a}_{n,DR}$ and $\tilde{\sigma}_{DR}$, respectively. This completes the proof.

A.7 Proof of Theorem 4.5. Similar to the proof for Theorem 4.3, under H_a, we can show $\sqrt{n}\widetilde{S}^B_{n,DR}$ is equivalent to $\sqrt{n}\dot{S}^B_{n,DR}$, where the definition of $\dot{S}^B_{n,DR}$ is given in (A.95). Using similar arguments in the proof of Theorem 3.6, we can show

$$\frac{\sqrt{n}\dot{S}^B_{n,DR} - \mathrm{E}\sqrt{n}\dot{S}^B_{n,DR}}{\tilde{\sigma}_{DR}} \xrightarrow{d} N(0,1).$$

Besides, similar to (A.78), we can show

$$\mathrm{E}\sqrt{n}\dot{S}^B_{n,DR} - \tilde{a}_{n,DR} \to \frac{1}{2}\int_{x\in F_0}|\delta_0(x)|f(x)dx.$$

The proof is thus completed.

A.8 Proof of Theorem 9.1. For any $D \subseteq I$, define the set

$$\mathcal{D}_{-} = \{D : D_0 \not\subseteq D\},$$

it suffices to show

$$\Pr(\widehat{D} \in \mathcal{D}_{-}) \to 0.$$

The dimension of the covariates p is fixed. Hence, the number of elements in the set $\mathcal{D}_{-}$ is bounded. By Bonferroni's inequality, it suffices to show that

$$\Pr(\widehat{D} = D) \to 0, \tag{A.103}$$

for any set $D \in \mathcal{D}_{-}$.

Consider an arbitrary set $D_{-} \in \mathcal{D}_{-}$. By definition, there exists some $j_0 \notin D_{-}$ such that $j_0 \in D_0$. For such j_0, by Condition (C3), we have

$$V^{D_{-}} < V^{D_{-} \cup \{j_0\}}. \tag{A.104}$$

Assumption (A3) implies that the response Y has sub-exponential tail. Hence, we have $\mathrm{E}|Y|^2 < \infty$. By Jensen's inequality, we have $\mathrm{E}\{\mathrm{E}(Y|A = 1, X)\}^2 < \infty$ and $\mathrm{E}\{\mathrm{E}(Y|A = 0, X)\}^2 < \infty$. It follows from Cauchy-Schwarz inequality that

$$\begin{aligned} \mathrm{E}\{\tau_0(X)\}^2 &= \mathrm{E}\{\mathrm{E}(Y|A = 1, X) - \mathrm{E}(Y|A = 0, X)\}^2 \\ &\leq 2\mathrm{E}\{\mathrm{E}(Y|A = 1, X)\}^2 + 2\mathrm{E}\{\mathrm{E}(Y|A = 0, X)\}^2 < \infty. \end{aligned} \tag{A.105}$$

By Jensen's inequality, this implies that

$$\mathrm{E}\{\tau_0^{D_{-} \cup \{j_0\}}(X^{D_{-} \cup \{j_0\}})\}^2 < \infty.$$

Besides, by Condition (C2) and (A1), $\tau_0^{D_{-} \cup \{j_0\}}$ and $\tau_0^{D_{-}}$ are continuous. The conditions in Theorem 2.2 are satisfied for $W = D_{-} \cup \{j_0\}$ and $B = D_{-}$. By (A.104), Theorem 2.2 implies that $X^{(j_0)}$ has CQTE given $X^{D_{-}}$. The threshold α in the forward selection algorithm satisfies $\alpha \geq \bar{\Phi}(c_n)$ for some $c_n \ll n^{1/2}$. As a result, we have $z_\alpha \leq c_n \ll n^{1/2}$. Using similar arguments in the proof of Theorem 3.8, we can show

$$\Pr\left(\widetilde{T}_n^{D_{-} \cup \{j_0\}, D_{-}} > z_\alpha\right) \to 1. \tag{A.106}$$

Under the event $\widehat{D} = D_{-}$, the algorithm stops after obtaining $X^{D_{-}}$. According to the forward selection procedure, we have

$$\widetilde{T}_n^{D_{-} \cup \{j\}, D_{-}} \leq z_\alpha,$$

for all $j \in D_-^c$. This implies that

$$\Pr\left(\widehat{D} = D_-\right) \le \Pr\left(\widetilde{T}_n^{D_- \cup \{j_0\}, D_-} \le z_\alpha\right).$$

In view of (A.106), we have

$$\Pr\left(\widehat{D} = D_-\right) \to 0.$$

This proves (A.103). The proof is hence completed.

A.9 Proof of Lemma 9.1. By definition, we have $\mathrm{VD}^{W_0, D_0^c} = 0$. Assumption (A3) implies that the response Y has sub-exponential tail. By (A.105), we have $\mathrm{E}\tau_0^2(X) < \infty$. Under Condition (C4), $\tau_0^{D_0^c}(\cdot)$ and $\tau_0^{D_0}(\cdot)$ are continuous. Hence, it follows from Theorem 2.2 that for any $\tau_0(x) \neq 0$, we have

$$I\{\varphi_{D_0}(x_{D_0})\varphi_{D_0^c}(x_{D_0^c}) \ge 0\} = I\{\varphi_{D_0}(x_{D_0}) \ge 0\} \tag{A.107}$$

We claim there exists some x_{0,D_0} such that $\varphi_{D_0}(x_{0,D_0}) \neq 0$. Otherwise, we have $\varphi_{D_0}(x_{D_0}) = 0$ for all x_{D_0}. As a result, by Condition (C4), we have $\tau_0(x) = 0$ for all x. This implies $D_0 = \emptyset$, which contradicts the fact that D_0 is not empty. For such x_{0,D_0}, it follows from (A.107) that

$$I\{\varphi_{D_0}(x_{0,D_0})\varphi_{D_0^c}(x_{D_0^c}) \ge 0\} = I\{\varphi_{D_0}(x_{0,D_0}) \ge 0\},$$

for all $x_{D_0^c}$ such that $\phi_{D_0^c}(x_{D_0^c}) \neq 0$. This gives

$$\varphi_{D_0^c}(x_{D_0^c}) \ge 0, \qquad \forall x_{D_0^c}. \tag{A.108}$$

Consider an arbitrary set $D \subsetneq \mathcal{I}$ such that $D \subseteq D_0$ and any $j_0 \in D_0^c$. By Condition (C4), X^{D_0} and $X^{D_0^c}$ are independent. Hence, $X^{D_0 \cap D^c}$ and $X^{D_0^c}$ are conditionally independent given X^D. Similarly, $X^{(j_0)}$ and X^{D_0} are conditionally independent given $X^{D_0^c - \{j_0\}}$. Therefore, we have for any $x_D \in \Omega^D$ and any $x_{(j_0)} \in \Omega^{\{j_0\}}$,

$$\begin{aligned}
& \mathrm{E}\left(\tau(X)|X^D = x_D, X^{(j_0)} = x_{(j_0)}\right) \\
= \; & \mathrm{E}\left(\varphi_{D_0}(X^{D_0})\varphi_{D_0^c}(X^{D_0^c})|X^D = x_D, X^{(j_0)} = x_{(j_0)}\right) \\
= \; & \mathrm{E}\left(\varphi_{D_0}(X^{D_0})|X^D = x_D\right)\mathrm{E}\left(\varphi_{D_0^c}(X^{D_0^c})|X^{(j_0)} = x_{(j_0)}\right) \\
\equiv \; & \varphi_{D_0}^D(x_D)\varphi_{D_0^c}^{(j_0)}(x_{(j_0)}).
\end{aligned}$$

By (A.108), we have $\varphi_{D_0^c}^{(j_0)}(x_{(j_0)}) \geq 0$ for all $x_{(j_0)}$. Therefore, for any x_D, we have

$$\varphi_{D_0}^{D}(x_D)\varphi_{D_0^c}^{(j_0)}(x_{(j_0)}) \geq 0, \qquad \forall x_{(j_0)},$$

or

$$\varphi_{D_0}^{D}(x_D)\varphi_{D_0^c}^{(j_0)}(x_{(j_0)}) \leq 0, \qquad \forall x_{(j_0)}.$$

By Condition (C2) and (A1), $\tau_D(\cdot)$ and $\tau_{D\cup\{j_0\}}(\cdot)$ are continuous. Therefore, it follows from Theorem 2.2 that $X^{(j_0)}$ doesn't have CQTE given X^D. By Theorem 3.6, we have

$$\limsup_n \Pr\left(\widetilde{T}_n^{D\cup\{j_0\},D} \leq z_{\alpha_0}\right) \leq \alpha_0,$$

for any fixed $0 < \alpha_0 \leq 0.5$.

Since the threshold α in the forward selection algorithm satisfies $\alpha \to 0$, we have

$$\Pr\left(\widetilde{T}_n^{D\cup\{j_0\},D} \leq z_{\alpha}\right) \to 0, \tag{A.109}$$

for any $D \subseteq D_0$ and $j_0 \in D_0^c$.

Let

$$\mathcal{D}_+ = \left\{D \subseteq I : D \cap D_0^c \neq \emptyset\right\},$$

we need to prove

$$\Pr\left(\widehat{D} \in \mathcal{D}_+\right) \to 0.$$

The number of set in $\mathcal{D}_+$ is bounded. By Bonferroni's inequality, it suffices to show that

$$\Pr\left(\widehat{D} = D\right) \to 0, \tag{A.110}$$

for any $D \in \mathcal{D}_+$.

Consider an arbitrary set $D_+ \in \mathcal{D}_+$. According to our forward selection procedure, there exists some j_0 and D such that $D \subseteq D_+ \cap D_0$, $j_0 \in D_+ \cap D_0^c$, and at a certain step, given the existing set of variables X^D, our algorithm include the variable X^{j_0}. Hence, we have

$$\Pr\left(\widehat{D} = D_+\right) \leq \Pr\left(\max_{\substack{D\subseteq D_+\cap D_0 \\ j_0\in D_+\cap D_0^c}} \widetilde{T}_n^{D\cup\{j_0\},D} > z_{\alpha}\right).$$

By (A.110) and the Bonferroni's inequality, there exists some constant $\bar{C} > 0$ such that

$$\begin{aligned}
&\Pr\left(\max_{\substack{D\subseteq D_+\cap D_0\\ j_0\in D_+\cap D_0^c}} \widetilde{T}_n^{D\cup\{j_0\},D} > z_\alpha\right)\\
\leq\ & \bar{C}\max_{\substack{D\subseteq D_+\cap D_0\\ j_0\in D_+\cap D_0^c}} \Pr\left(\widetilde{T}_n^{D\cup\{j_0\},D} > z_\alpha\right)\\
\leq\ & \bar{C}\max_{\substack{D\subseteq D_0\\ j_0\in D_0^c}} \Pr\left(\widetilde{T}_n^{D\cup\{j_0\},D} > z_\alpha\right)\to 0.
\end{aligned}$$

This proves (A.110). The proof is hence completed.

A.10 Proof of Theorem 12.1. It suffices to show

$$\limsup_n \Pr\left(\widetilde{T}_{n,q}^B > z_\alpha\right) \leq \alpha,$$

in the following three cases (i) $\nu(F_0) > 0$; (ii) $\nu(F_0) = 0$ and $\nu(\hat{F}) > 0$; (iii) $\nu(F_0) = 0$ and $\nu(\hat{F}) = 0$. In the following, we will show that when $\nu(F_0) > 0$,

$$\frac{n^{q/2}h^{(q-1)p/2}\widetilde{S}_{n,q}^B - \hat{a}_{n,q}}{\hat{\sigma}_{n,q}} \xrightarrow{d} N(0,1), \tag{A.111}$$

where $\hat{a}_{n,q}$ and $\hat{\sigma}_{n,q}$ stand for $\hat{a}_{n,q}(\hat{F})$ and $\hat{\sigma}_{n,q}(\hat{F})$, respectively. Following the arguments in the proof of Theorem 3.6, we can show that

$$\limsup_n \Pr\left(\widetilde{T}_{n,q}^B > z_\alpha\right) \to 0,$$

in Case (ii) and (iii). The proof is therefore completed.

Proof of (A.111)*:* Similar to the proof of Theorem 3.6, under the given conditions, we can show $n^{q/2}h^{(q-1)p/2}\widetilde{S}_{n,q}^B = n^{q/2}h^{(q-1)p/2}\dot{S}_{n,q}^B + o_p(1)$, $\hat{a}_{n,q} = \tilde{a}_{n,q} + o_p(1)$, $\hat{\sigma}_{n,q} = \tilde{\sigma}_q + o_p(1)$ and $\tilde{\sigma}_q > 0$ where

$$\begin{aligned}
\dot{S}_{n,q}^B &= \int_{x\in F_0}\phi_q\{\tau_n(x)\}\{d_n(x) - d_0^B(x_B)\}dx,\\
\tilde{a}_{n,q} &= \frac{1}{2\sqrt{h^p}}\int_{x\in F_0}\{\mu_n(x)\}^{q/2}dx\mathrm{E}|\mathbb{Z}_1^q|,\\
\tilde{\sigma}_q^2 &= \int_{\substack{x\in F_0\\ t\in[-1,1]^p}}\mu(x)\mathrm{cov}(\max\{\sqrt{1-\rho^2(t)}\mathbb{Z}_1+\rho(t)\mathbb{Z}_2, 0\}^q, \max\{\mathbb{Z}_2,0\}^q)\\
&\quad\times dxdt.
\end{aligned}$$

Hence, it suffices to prove the asymptotic normality of

$$\frac{n^{q/2}h^{(q-1)p/2}\dot{S}^{B}_{n,q}-\tilde{a}_{n,q}}{\tilde{\sigma}_q}.$$

Consider the Poissonized version

$$\dot{S}^{B}_{N,q}=\int_{x\in F_0}\phi_q\{\tau_N(x)\}[I\{\tau_N(x)\geq 0\}-d_0^B(x_B)]dx,$$

where N is a Poisson random variable with mean $\mathrm{E}N=n$ independent of the data, $\tau_N(\cdot)$ is defined in (A.38). Similar to (A.39), we can show

$$n^{q}h^{p(q-1)}\mathrm{Var}(\dot{S}^{B}_{N,q})=\tilde{\sigma}_q^2+o(1). \tag{A.112}$$

In the following, we prove

$$n^{q/2}h^{p(q-1)/2}\mathrm{E}\dot{S}^{B}_{N,q}=\tilde{a}_{n,q}+o(1). \tag{A.113}$$

By definition, it suffices to show

$$\sup_{x\in\dot{F}_0}\left|n^{q/2}h^{p(q-1)/2}\mathrm{E}\psi_q\{\tau_N(x)\}[I\{\tau_N(x)\geq 0\}-d_0^B(x_B)]-\frac{\{\mu_n(x)\}^{q/2}\mathrm{E}|\mathbb{Z}_1^q|}{2\sqrt{h^p}}\right|=o(1). \tag{A.114}$$

In the following, we focus on proving

$$\sup_{x\in\dot{F}_0}\left|n^{q/2}h^{p(q-1)/2}\mathrm{E}\max\{\tau_N(x),0\}^q-\frac{\{\mu_n(x)\}^{q/2}}{\sqrt{h^p}}\mathrm{E}\max\{\mathbb{Z},0\}^q\right|=o(1). \tag{A.115}$$

Similarly, we can show

$$\sup_{x\in\dot{F}_0}\left|n^{q/2}h^{p(q-1)/2}\mathrm{E}\min\{-\tau_N(x),0\}^q-\frac{\{\mu_n(x)\}^{q/2}}{\sqrt{h^p}}\mathrm{E}\min\{-\mathbb{Z},0\}^q\right|=o(1).$$

This together with (A.115) gives

$$\sup_{x\in\dot{F}_0}\left|n^{q/2}h^{p(q-1)/2}\mathrm{E}\psi_q\{-\tau_N(x)\}\right|=o(1),$$

and hence

$$\sup_{x\in\dot{F}_0}\left|n^{q/2}h^{p(q-1)/2}\mathrm{E}\psi_q\{-\tau_N(x)\}d_0^B(x_B)\right|=o(1).$$

Combining this together with (A.115) yields (A.114). Assertion (A.113) thus follows. It remains to prove (A.115).

Note that $\tau_N(x)$ can be represented as i.i.d summation of

$$\frac{1}{n}\sum_{i=1}^{\eta}\left(\frac{A_i}{\pi_i}-\frac{1-A_i}{1-\pi_i}\right)Y_iK_h(x-X_i),$$

where η is a Poisson random variable with $\mathrm{E}\eta=1$, independent of data. Define

$$W_0=\frac{1}{\sqrt{h^p\mu_n(x)}}\left\{\sum_{i=1}^{\eta}\left(\frac{A_i}{\pi_i}-\frac{1-A_i}{1-\pi_i}\right)Y_iK\left(\frac{x-X_i}{h}\right)-\mathrm{E}\tau_0(X)K\left(\frac{x-X}{h}\right)\right\},$$

we have $\mathrm{E}W_0=0, \mathrm{Var}(W_0)=1$. Let $W_1,\dots,W_n$ be i.i.d copies of W_0, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_i\stackrel{d}{=}\frac{\sqrt{nh^p}}{\sqrt{\mu_n(x)}}\{\tau_N(x)-\mathrm{E}\tau_n(x)\},$$

and hence

$$\begin{aligned}\max\{\tau_N(x),0\}^q &\stackrel{d}{=} \max\left(\frac{\sqrt{\mu_n(x)}}{\sqrt{nh^p}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_i+\mathrm{E}\tau_n(x),0\right)^q\\ &= \frac{\{\mu_n(x)\}^{q/2}}{(nh^p)^{q/2}}\max\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_i+\frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}},0\right)^q\\ &= \frac{\{\mu_n(x)\}^{q/2}}{(nh^p)^{q/2}}\psi_q\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_i,\frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}}\right), \qquad \text{(A.116)}\end{aligned}$$

where ψ_q is defined in Lemma A.7.

By Condition (A5) and Lemma A.3, we have $\liminf_n \inf_{x\in\Omega}\mu_n(x)>0$. By (A3*), using similar arguments in the proof of Lemma A.5, we can show

$$\mathrm{E}|W_1|^3=O\left(\frac{1}{h^{p/2}}\right),$$

and

$$\mathrm{E}|W_1|^r=O\left(\frac{1}{h^{(r-2)p/2}}\right).$$

It follows from Lemma A.7 that for any $x \in \Omega$,

$$\begin{aligned}
&\left| \mathrm{E}\psi_q \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_i, \frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right) - \mathrm{E}\psi_q \left(\mathbb{Z}_0, \frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right) \right| \\
\le \quad & c^* \left(\left| \frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right|^{q-1} + 1 \right) \left(\frac{1}{\sqrt{nh^p}} + \frac{1}{(nh^p)^{(r-2)/2}} \right) \\
+ \quad & \frac{c^*}{(nh^p)^{q/2}} + \frac{c^*}{\sqrt{nh^p}} \left(1 + \left| \frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right|^{q-1} \right),
\end{aligned}$$

for some constant $c^* > 0$. By Condition (A6), we have $nh^p \to \infty$. Since $r \ge 3$, $q \ge 1$, we have $(nh^p)^{(r-2)/2}, (nh^p)^{q/2} \gg \sqrt{nh^p}$. Hence, for sufficiently large n, we have

$$\begin{aligned}
&\left| \mathrm{E}\psi_q \left(\frac{1}{\sqrt{n}} W_i, \frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right) - \mathrm{E}\psi_q \left(\mathbb{Z}_0, \frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right) \right| \\
\le \quad & \frac{2c^*}{\sqrt{nh^p}} \left(\left| \frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right|^{q-1} + 1 \right) + \frac{c^*}{\sqrt{nh^p}} + \frac{c^*}{\sqrt{nh^p}} \\
& \left(1 + \left| \frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right|^{q-1} \right) \le \frac{4c^*}{\sqrt{nh^p}} \left(\left| \frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right|^{q-1} + 1 \right).
\end{aligned} \tag{A.117}$$

Moreover, it follows from Lemma A.8 that for any $x \in \Omega$,

$$\begin{aligned}
&\left| \mathrm{E}\psi_q \left(\mathbb{Z}_0, \frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right) - \mathrm{E}\psi_q(\mathbb{Z}_0, 0) \right| \\
&\le c^{**} \left| \frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right| + c^{**} \left| \frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right|^{q},
\end{aligned}$$

for some constant $c^{**} > 0$. Combining this together with (A.117), we obtain that for any $x \in \Omega$,

$$\begin{aligned}
&\left| \mathrm{E}\psi_q \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_i, \frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right) - \mathrm{E}\psi_q(\mathbb{Z}_0, 0) \right| \\
\le \quad & \frac{c^{***}}{\sqrt{nh^p}} \left(\left| \frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right|^{q-1} + 1 \right) + c^{***} \left| \frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right| + c^{***} \left| \frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right|^{q},
\end{aligned}$$

for some constant $c^{***} > 0$. By (A.116), this implies that for any $x \in \Omega$,

$$\begin{aligned} & \left| n^{q/2} h^{p(q-1)/2} \mathrm{E} \max\{\tau_N(x), 0\}^q - \frac{\{\mu_n(x)\}^{q/2}}{h^{p/2}} \mathrm{E} \max(\mathbb{Z}_0, 0)^q \right| \\ \leq \;& \frac{c^{***}}{\sqrt{nh^p}} \left(|\sqrt{nh^p} \mathrm{E}\tau_n(x)|^{q-1} \mu_n(x)^{1/2} + \mu_n(x)^{q/2} \right) \\ + \;& c^{***} |\sqrt{n} \mathrm{E}\tau_n(x)| \mu_n^{(q-1)/2}(x) + c^{***} n^{q/2} h^{p(q-1)/2} |\mathrm{E}\tau_n(x)|^q. \end{aligned}$$

For any $x \in \dot{F}_0$, we have $\mathrm{E}\tau_n(x) = O(h^s)$. Under Condition (A5), $\mu(x)$ is bounded. Since $\sup_x |\mu_n(x) - \mu(x)| = o(1)$, $\mu_n(x)$ is also bounded. Hence, we have

$$\begin{aligned} & \sup_{x \in \dot{F}_0} \left| n^{q/2} p^{(q-1)/2} \mathrm{E} \max\{\tau_N(x), 0\}^q - \frac{\{\mu_n(x)\}^{q/2}}{h^{p/2}} \mathrm{E} \max(\mathbb{Z}_0, 0)^q \right| \\ = \;& O\left(\frac{1}{\sqrt{nh^p}} \right) + O(\sqrt{n} h^s) + O(n^{q/2} h^{p(q-1)/2} h^{sq}) + O(n^{(q-2)/2} h^{p(q-3)/2 + s(q-1)}). \end{aligned}$$

By (A6), we have $\sqrt{n} h^p \to \infty$, $\sqrt{n} h^s \to 0$. This implies $s > p$. Since $h \to 0$, we have

$$n^{q/2} h^{p(q-1)/2} h^{sq} \ll n^{q/2} h^{sq} \to 0.$$

Now we claim

$$n^{(q-2)/2} h^{p(q-3)/2 + s(q-1)} \to 0. \tag{A.118}$$

This implies

$$\sup_{x \in \dot{F}_0} \left| n^{q/2} p^{(q-1)/2} \mathrm{E} \max\{\tau_N(x), 0\}^q - \frac{\{\mu_n(x)\}^{q/2}}{h^{p/2}} \mathrm{E} \max(\mathbb{Z}_0, 0)^q \right| \to 0.$$

It remains to show (A.118) holds for any $q \geq 1$. When $1 \leq q \leq 2$, we have

$$n^{(q-2)/2} h^{p(q-3)/2 + s(q-1)} = \frac{h^{(s-p/2)(q-1)}}{n^{(2-q)/2} h^{p(2-q)}}.$$

The numerator converges to 0 since $s > p/2$. The denominator is either equal to 1 or diverges to ∞, by (A6). This shows (A.118) holds for $1 \leq q \leq 2$. When $q > 2$, by (A6), we have $nh^{2s} \to 0$ and hence $n^{(q-2)/2} h^{s(q-2)} \to 0$. Besides, since $s > p$, we have $h^{s+p(q-3)/2} \to 0$ for all $q > 2$. This implies (A.118) holds for any $q > 2$ as well. This proves (A.113).

Define $U = [-1.1]^p$, $\alpha_n = \Pr(X \in F_0 + hU)$ and

$$U_n = \frac{\sum_{i=1}^{N} I(X_i \in F_0 + hU) - n\alpha_n}{\sqrt{n}}, \ V_n = \frac{\sum_{i=1}^{N} I(X_i \notin F_0 + hU) - n(1-\alpha_n)}{\sqrt{n}}.$$

Note that we have $\alpha_n \to \alpha = \Pr(X \in F_0)$. Since $\nu(F_0) > 0$, $\nu(\Omega - F_0) > 0$, we have $0 < \alpha < 1$. By Lemma A.1, U_n and $\dot{S}_{N,q}$ are independent of V_n.

Assume for now, we can show

$$\text{(A.119)} \qquad \frac{n^{q/2}h^{q(p-1)/2}(\dot{S}^B_{N,q} - \mathrm{E}\dot{S}^B_{N,q})}{\tilde{\sigma}_q} \xrightarrow{d} N(0,1),$$

and

$$\text{(A.120)} \qquad \operatorname{cov}(n^{q/2}h^{q(p-1)/2}\dot{S}^B_{N,q}, U_n) \to 0.$$

By (A.113), we have

$$\frac{n^{q/2}h^{q(p-1)/2}\dot{S}^B_{N,q} - \tilde{a}_{n,q}}{\tilde{\sigma}_q} \xrightarrow{d} N(0,1).$$

This together with (A.120) gives

$$(n^{q/2}h^{q(p-1)/2}\dot{S}^B_{N,q} - \tilde{a}_{n,q}, U_n) \xrightarrow{d} (\tilde{\sigma}\mathbb{Z}_1, \alpha\mathbb{Z}_2),$$

for standard normal random variables $\mathbb{Z}_1$ and $\mathbb{Z}_2$. Then, the asymptotic normality of

$$\frac{n^{q/2}h^{q(p-1)/2}\dot{S}^B_{n,q} - \tilde{a}_{n,q}}{\tilde{\sigma}_q}$$

follows by an application of Lemma A.6.

It remains to show (A.119) and (A.120). However, these results can be proven following the arguments in the proof of Theorem 1 in Lee et al. (2013). We omit the details for brevity.

A.11 Proof of Theorem 12.2. Similar to the proof of Theorem 12.1, we can show that $\widetilde{S}^B_{n,q} = \dot{S}^B_{n,q} + o_p(1)$,

$$\frac{n^{q/2}h^{(q-1)p/2}(\dot{S}^B_{n,q} - \mathrm{E}\dot{S}^B_{N,q})}{\hat{\sigma}_{n,q}} \to N(0,1),$$

$\hat{a}_{n,q} = \tilde{a}_{n,q} = o(1)$, and $\hat{\sigma}_{n,q} = \tilde{\sigma}_{n,q} + o(1)$. It suffices to show that

$$\begin{aligned} & n^{q/2} h^{(q-1)p/2} \mathrm{E}\dot{S}_{N,q}^{B} \\ = \quad & \tilde{a}_{n,q} + \int_{x \in \dot{F}_0} \frac{2^{(q-3)/2} q \Gamma(q/2)}{\sqrt{\pi}} \mu_n^{(q-1)/2}(x) \delta_0(x) f(x) dx + o(1). \end{aligned}$$

Similar to (A.117), we can show

$$\begin{aligned} & \sup_{x \in \dot{F}_0} \left| \mathrm{E}\psi_q \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_i, \frac{\sqrt{nh^p} \mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right) - \mathrm{E}\psi_q \left(\mathbb{Z}_0, \frac{\sqrt{nh^p} \mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right) \right| \\ \leq \quad & \frac{c^*}{\sqrt{nh^p}} \sup_{x \in \dot{F}_0} \left(\left| \frac{\sqrt{nh^p} \mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right|^{q-1} + 1 \right), \end{aligned}$$

for some constant $c^* > 0$. Standard bias calculation given $\mathrm{E}\tau_n(x) = \tau_0(x) f(x) + n^{-1/2} \delta_0(x) f(x) + O(h^s)$, where the big-$O$ term is uniform in x. Since $\tau_0(x) = 0$ for $x \in \dot{F}_0$, we have

$$\sup_{x \in \dot{F}_0} |\mathrm{E}\tau_n(x) - n^{-1/2} \delta_0(x) f(x)| = O(h^s). \tag{A.121}$$

Since $\delta_0(\cdot)$ is bounded, and $\sqrt{n} h^s \to 0$, it follows from (A.121) that

$$\sup_{x \in \dot{F}_0} |\mathrm{E}\tau_n(x)| \leq n^{-1/2} \sup_{x} |\delta_0(x) f(x)| + O(h^s) = O(n^{-1/2}). \tag{A.122}$$

By Lemma A.3, we have $\sup_x |\mu_n(x) - \mu(x)| = o(1)$ and $\inf_x \mu(x) > 0$. This implies $\liminf_n \inf_x \mu_n(x) > 0$. Combining this together with (A.122), we have

$$\sup_{x \in \dot{F}_0} \left| \frac{\sqrt{nh^p} \mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right| = O(h^p) = o(1). \tag{A.123}$$

Since $q \geq 1$, we obtain

$$\sup_{x \in \dot{F}_0} \left| \mathrm{E}\psi_q \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_i, \frac{\sqrt{nh^p} \mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right) - \mathrm{E}\psi_q \left(\mathbb{Z}_0, \frac{\sqrt{nh^p} \mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}} \right) \right| \leq \frac{c^{**}}{\sqrt{nh^p}}, \tag{A.124}$$

for some constant $c^{**} > 0$.

Moreover, it follows from Lemma A.8 that

$$\begin{aligned}
&\sup_{x\in\dot{F}_0}\left|\mathrm{E}\psi_q\left(\mathbb{Z}_0, \frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}}\right) - \mathrm{E}\psi_q\left(\mathbb{Z}_0, \frac{h^{p/2}\delta_0(x)f(x)}{\sqrt{\mu_n(x)}}\right)\right| \\
\leq\ & c^{***}\left|\frac{\sqrt{nh^p}\{\mathrm{E}\tau_n(x) - n^{-1/2}\delta_0(x)f(x)\}}{\sqrt{\mu_n(x)}}\right|\left(1 + \left|\frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}}\right|^{q-1}\right) \\
+\ & c^{***}\left|\frac{\sqrt{nh^p}\{\mathrm{E}\tau_n(x) - n^{-1/2}\delta_0(x)f(x)\}}{\sqrt{\mu_n(x)}}\right|^{q},
\end{aligned}$$

for some constant $c^{***} > 0$. By (A.121) and (A.123), we have

$$\sup_{x\in\dot{F}_0}\left|\mathrm{E}\psi_q\left(\mathbb{Z}_0, \frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}}\right) - \mathrm{E}\psi_q\left(\mathbb{Z}_0, \frac{h^{p/2}\delta_0(x)f(x)}{\sqrt{\mu_n(x)}}\right)\right| \leq c^{****}n^{1/2}h^{p/2}h^s,$$

for some constant $c^{****} > 0$. This together with (A.124) gives

$$\begin{aligned}
&\sup_{x\in\dot{F}_0}\left|\mathrm{E}\psi_q\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_i, \frac{\sqrt{nh^p}\mathrm{E}\tau_n(x)}{\sqrt{\mu_n(x)}}\right) - \mathrm{E}\psi_q\left(\mathbb{Z}_0, \frac{h^{p/2}\delta_0(x)f(x)}{\sqrt{\mu_n(x)}}\right)\right| \\
\leq\ & c^{*****}\left(n^{1/2}h^{p/2}h^s + \frac{1}{\sqrt{nh^p}}\right),
\end{aligned}$$

where $c^{*****} = \max(c^{****}, c^{**})$. By (A.116), we have

$$\begin{aligned}
&\sup_{x\in\dot{F}_0}\left|n^{q/2}h^{p(q-1)/2}\mathrm{E}\max\{\tau_N(x), 0\}^q - \frac{\mu_n^{q/2}(x)}{h^{p/2}}\mathrm{E}\psi_q\left(\mathbb{Z}_0, \frac{h^{p/2}\delta_0(x)f(x)}{\sqrt{\mu_n(x)}}\right)\right| \\
\leq\ & c^{*****}\mu_n^{q/2}(x)\left(n^{1/2}h^s + \frac{1}{\sqrt{nh^{2p}}}\right).
\end{aligned}$$

By (A5) and Lemma A.3, function $\mu_n(\cdot)$ is bounded. Under the given conditions, we have $n^{1/2}h^s \to 0$, $nh^{2p} \to \infty$. Hence, we have

(A.125)

$$\sup_{x\in\dot{F}_0}\left|n^{q/2}h^{p(q-1)/2}\mathrm{E}\max\{\tau_N(x), 0\}^q - \frac{\mu_n^{q/2}(x)}{h^{p/2}}\mathrm{E}\psi_q\left(\mathbb{Z}_0, \frac{h^{p/2}\delta_0(x)f(x)}{\sqrt{\mu_n(x)}}\right)\right| = o(1).$$

By Lemma A.9, following the arguments in proving (A.82), we can show

$$\begin{aligned}
\sup_{x\in\dot{F}_0}\Bigg|&\frac{\mu_n^{q/2}(x)}{h^{p/2}}\mathrm{E}\psi_q\left(\mathbb{Z}_0, \frac{h^{p/2}\delta_0(x)f(x)}{\sqrt{\mu_n(x)}}\right) - \frac{\mu_n^{q/2}(x)}{h^{p/2}}\mathrm{E}\psi_q\left(\mathbb{Z}_0, 0\right) \\
&-q\mu_n^{(q-1)/2}(x)\delta_0(x)f(x)\mathrm{E}\mathbb{Z}_0^{q-1}I(\mathbb{Z}_0 > 0)\Bigg| = o(1).
\end{aligned}$$

Since

$$\mathrm{E}\mathbb{Z}_0^{q-1} I(\mathbb{Z}_0 > 0) = \frac{\mathrm{E}|\mathbb{Z}_0|^{q-1}}{2} = \frac{2^{(q-3)/2}\Gamma(q/2)}{\sqrt{\pi}},$$

we have

$$\sup_{x\in \dot{F}_0} \left| \frac{\mu_n^{q/2}(x)}{h^{p/2}} \mathrm{E}\psi_q \left(\mathbb{Z}_0, \frac{h^{p/2}\delta_0(x) f(x)}{\sqrt{\mu_n(x)}}\right) - \frac{\mu_n^{q/2}(x)}{h^{p/2}} \mathrm{E}\psi_q \left(\mathbb{Z}_0, 0\right) \right.$$
$$\left. - \frac{2^{(q-3)/2} q\Gamma(q/2)}{\sqrt{\pi}} \mu_n^{(q-1)/2}(x)\delta_0(x) f(x) \right| = o(1).$$

This together with (A.125) gives

(A.126)
$$\sup_{x\in \dot{F}_0} \left| n^{q/2} h^{p(q-1)/2} \mathrm{E}\max\{\tau_N(x), 0\}^q - \frac{\mu_n^{q/2}(x)}{h^{p/2}} \mathrm{E}\{\max(\mathbb{Z}_0, 0)\}^q \right.$$
$$\left. - \frac{2^{(q-3)/2} q\Gamma(q/2)}{\sqrt{\pi}} \mu_n^{(q-1)/2}(x)\delta_0(x) f(x) \right| = o(1).$$

Similarly, we can show

$$\sup_{x\in \dot{F}_0} \left| n^{q/2} h^{p(q-1)/2} \mathrm{E}\{-\min(\tau_N(x), 0)\}^q - \frac{\mu_n^{q/2}(x)}{h^{p/2}} \mathrm{E}\{-\min(-\mathbb{Z}_0, 0)\}^q \right.$$
$$\left. + \frac{2^{(q-3)/2} q\Gamma(q/2)}{\sqrt{\pi}} \mu_n^{(q-1)/2}(x)\delta_0(x) f(x) \right| = o(1),$$

and hence

$$\sup_{x\in \dot{F}_0} \left| n^{q/2} h^{p(q-1)/2} \mathrm{E}\phi_q\{\tau_N(x)\} - \frac{2^{(q-1)/2} q\Gamma(q/2)}{\sqrt{\pi}} \mu_n^{(q-1)/2}(x)\delta_0(x) f(x) \right| = o(1).$$

Combining this together with (A.126) gives

$$\sup_{x\in \dot{F}_0} \left| n^{q/2} h^{p(q-1)/2} \mathrm{E}\phi_q\{\tau_N(x)\}[I\{\tau_N(x) \ge 0\} - d_0^B(x_B)] - \frac{\mu_n^{q/2}(x)}{2h^{p/2}} \mathrm{E}|\mathbb{Z}_0|^q \right.$$
$$\left. - \frac{2^{(q-1)/2} q\Gamma(q/2)}{\sqrt{\pi}} \mu_n^{(q-1)/2}(x)\delta_0(x) f(x) \left(\frac{1}{2} - d_0^B(x_B)\right) \right| = o(1).$$

By dominated convergence theorem, we have

$$\begin{aligned}
&\mathrm{E}\sqrt{n}\dot{S}_{n,q}^{B}-\tilde{a}_{n,q}\\
\to\;&\int_{x\in\dot{F}_0}\frac{2^{(q-1)/2}q\Gamma(q/2)}{\sqrt{\pi}}\mu_n^{(q-1)/2}(x)\delta_0(x)f(x)\left(\frac{1}{2}-I(\tau_0^B(x_B)\geq 0)\right)dx.
\end{aligned}$$

By assumption, we have $\delta_0(x)\geq 0$ if $\tau_0^B(x_B)\leq 0$, and $\delta_0(x)\leq 0$ if $\tau_0^B(x_B)\geq 0$. Therefore,

$$\delta_0(x)\left(\frac{1}{2}-I(\tau_0^B(x_B)\geq 0)\right)=\frac{1}{2}|\delta_0(x)|,\qquad \forall x\in\dot{F}_0.$$

As a result, we obtain

$$\mathrm{E}\sqrt{n}\dot{S}_{n,q}^{B}-\tilde{a}_{n,q}\to\int_{x\in\dot{F}_0}\frac{2^{(q-3)/2}q\Gamma(q/2)}{\sqrt{\pi}}\mu_n^{(q-1)/2}(x)\delta_0(x)f(x)dx.$$

The proof is hence completed.

A.12 Proof of Theorem 13.1. For ease of notation, we write X_1^{W*}, $\cdots$, X_m^{W*} as $X_1^*,\ldots,X_m^*$. Let k_h denote the number of elements in $\mathcal{H}$. By definition, we have $h_{\max}t^{k_h}\geq h_{\min}$ and hence $k_h\leq\log(h_{\max}/h_{\min})/\log(t^{-1})$. In Condition (D1), we assume $h_{\max}\to 0$ and $h_{\min}\geq\{(\log n)^3/n\}^{1/p}$. Since $0<t<1$, this implies

$$k_h\leq\frac{\log h_{\min}^{-1}}{\log t^{-1}}\leq\frac{p^{-1}\log\{n/(\log n)^3\}}{\log t^{-1}}\leq\frac{\log n}{\log t^{-1}}. \tag{A.127}$$

Consider the event

$$\begin{aligned}
\mathcal{A}_n(h)&=\left\{\sup_{x\in\Omega}|\tau_n(x,h)-\tau_0(x)f(x)|\leq\gamma\frac{\sqrt{\log n}}{\sqrt{nh^p}}+\gamma h^s\right\},\\
\mathcal{A}_n^B(h_B)&=\left\{\sup_{x_B\in\Omega^B}|\tau_n^B(x_B)-\tau_0^B(x_B)f^B(x_B)|\leq\frac{\gamma\sqrt{\log n}}{\sqrt{n(h_B)^{p_B}}}+\gamma h_B^s\right\}.
\end{aligned}$$

Similar to the proof of Lemma A.2, we can show that there exists some $\gamma>0$ such that for any h satisfies $h_{\max}\leq h\leq h_{\min}$, we have

$$\Pr\left(\{\mathcal{A}_n(h)\cap\mathcal{A}_n^B(h_B)\}^c\right)=O(n^{-2}).$$

By (A.127) and Bonferroni's inequality, we obtain

$$\begin{aligned}
&\sum_{n=1}^{\infty} \Pr\left(\bigcup_{\substack{h\in\mathcal{H}\\ h_B\in\mathcal{H}}} \{\mathcal{A}_n(h)\cap \mathcal{A}_n^B(h_B)\}^c \right)\\
\le\;& \sum_{n=1}^{\infty} \sum_{\substack{h\in\mathcal{H}\\ h_B\in\mathcal{H}}} \Pr\left(\{\mathcal{A}_n(h)\cap \mathcal{A}_n^B(h_B)\}^c\right) = O\left(\sum_{n=1}^{\infty} n^{-2}\log^2 n\right) < \infty.
\end{aligned}$$

Hence, it follows from Borel-Cantelli Lemma that

$$\Pr\left(\limsup_n \bigcap_{\substack{h\in\mathcal{H}\\ h_B\in\mathcal{H}}} \{\mathcal{A}_n(h)\cap \mathcal{A}_n^B(h_B)\} \right) = 1.$$

Similarly, we can show

$$\Pr\left(\limsup_n \bigcap_{\substack{h\in\mathcal{H}\\ h_B\in\mathcal{H}}} \{\mathcal{A}_n^*(h)\cap \mathcal{A}_n^{B*}(h_B)\} \right) = 1,$$

where

$$\begin{aligned}
\mathcal{A}_n^*(h) &= \left\{ \sup_{x\in\Omega} |\hat{f}(x,h) - f(x)| \le \gamma\frac{\sqrt{\log n}}{\sqrt{nh^p}} + \gamma h^s \right\},\\
\mathcal{A}_n^{B*}(h_B) &= \left\{ \sup_{x_B\in\Omega^B} |\hat{f}^B(x_B,h_B) - f^B(x_B)| \le \frac{\gamma\sqrt{\log n}}{\sqrt{n(h_B)^{p_B}}} + \gamma h_B^s \right\}.
\end{aligned}$$

This implies that with probability 1, there exists some integer $K>0$, such that the following event holds

$$\bigcap_{n\ge K} \bigcap_{\substack{h\in\mathcal{H}\\ h_B\in\mathcal{H}}} \{\mathcal{A}_n(h)\cap \mathcal{A}_n^B(h_B)\cap \mathcal{A}_n^*(h)\cap \mathcal{A}_n^{B*}(h_B)\}. \tag{A.128}$$

Define the set

$$\begin{aligned}
E_3 &= \left\{ x\in\Omega : 0 < |\tau_0(x)f(x)| \le \frac{2\gamma\sqrt{\log n}}{\sqrt{nh_{\min}^p}} + 2\gamma h_{\max}^s \right\},\\
E_4 &= \left\{ x\in\Omega : 0 < |\tau_0^B(x_B)f^B(x_B)| \le \frac{2\gamma\sqrt{\log n}}{\sqrt{nh_{\min}^{p_B}}} + 2\gamma h_{\max}^s \right\}.
\end{aligned}$$

It follows from Condition (D2), (D3) and Cauchy-Schwarz inequality that

$$\nu(E_3) = O\left\{\left(\frac{\sqrt{\log n}}{\sqrt{nh_{\min}^p}} + h_{\max}^s\right)^{\xi_0}\right\} = O\left(\frac{(\log n)^{\xi_0/2}}{(nh_{\min}^p)^{\xi_0/2}} + h_{\max}^{s\xi_0}\right) \ll \eta_n^{\xi_0},$$

$$\nu(E_4) \le \nu(\Omega^C) \times O\left\{\left(\frac{\sqrt{\log n}}{\sqrt{nh_{\min}^p}} + h_{\max}^s\right)^{\xi_0}\right\} = O\left(\frac{(\log n)^{\xi_0/2}}{(nh_{\min}^p)^{\xi_0/2}} + h_{\max}^{s\xi_0}\right) \ll \eta_n^{\xi_0}.$$

Since $X_1^*, \ldots, X_m^*$ are uniformly distributed on Ω, we have for any $1 \le j \le m$,

$$\Pr(X_j^* \in E_3 \cup E_4) \le \frac{\nu(E_3 \cup E_4)}{\nu(\Omega)} \le \frac{\nu(E_3) + \nu(E_4)}{\nu(\Omega)} \le \eta_n^{\xi_0},$$

for sufficiently large n. By Bonferroni's inequality, we obtain

$$\Pr\left(\bigcup_{j=1}^m \left\{X_j^* \in E_3 \cup E_4\right\}^c\right) \ge 1 - m\eta_n^{\xi_0} \to 1. \tag{A.129}$$

Similarly, we can show

$$\Pr\left(\bigcup_{j=1}^m \left\{X_j^* \in F_0(h_{\max})\right\}^c\right) \ge 1 - m\nu(F_0(h_{\max})) \to 1. \tag{A.130}$$

Let $E_5 = \{x \in \Omega : \tau_0(x) = 0\}$, $E_6 = \{x \in \Omega : \tau_0^B(x_B) = 0\}$. Notice that under the event defined in (A.128), we have for any $x \in E_3^c \cap E_4^c \cap E_5^c \cap E_6^c$, $h, h_B \in \mathcal{H}$, $n \ge K$,

$$d_n(x, h) = I(\tau_0(x) \ge 0) \quad \text{and} \quad d_n^B(x_B, h_B) = I(\tau_0^B(x_B) \ge 0).$$

Under H_0, this implies $d_n(x, h) = d_n^B(x_B, h_B)$ for any $x \in E_3^c \cap E_4^c \cap E_5^c \cap E_6^c$, $h, h_B \in \mathcal{H}$, $n \ge K$. Hence,

$$\widetilde{T}_n^B = \max_{\substack{h\in\mathcal{H} \\ h_B\in\mathcal{H}}} \max_{\substack{x\in\{X_1^*,\ldots,X_m^*\} \\ x\notin\hat{E}(h,h_B) \\ x\in E_3\cup E_4\cup E_5\cup E_6}} \frac{\sqrt{nh^p}\tau_n(x,h)}{\sqrt{\hat{\mu}_n(x,h)}}\{d_n(x,h) - d_n^B(x_B,h_B)\}.$$

Besides, under the event defined in (A.129), we have

$$\widetilde{T}_n^B = \max_{\substack{h\in\mathcal{H} \\ h_B\in\mathcal{H}}} \max_{\substack{x\in\{X_1^*,\ldots,X_m^*\} \\ x\notin\hat{E}(h,h_B) \\ x\in E_5\cup E_6}} \frac{\sqrt{nh^p}\tau_n(x,h)}{\sqrt{\hat{\mu}_n(x,h)}}\{d_n(x,h) - d_n^B(x_B,h_B)\}.$$

Under the null, we have $E_6 = \{x \in \Omega : \tau_0(x) = 0, \tau_0^B(x_B) = 0\}$ and $E_5 = E_6 + F_0$. In addition, under the event defined in (A.128), we have for any $x \in E_6$, $h, h_B \in \mathcal{H}$ and sufficiently large n,

$$|\tau_n(x,h)/\hat{f}(x,h)| \le \eta_n \quad \text{and} \quad |\tau_n^B(x_B,h_B)/\hat{f}^B(x_B,h_B)| \le \eta_n, \tag{A.131}$$

and hence

$$x \in \hat{E}(h, h_B). \tag{A.132}$$

Therefore, under the event defined in (A.128) and (A.129), we have

$$\widetilde{T}_n^B = \max_{\substack{h\in\mathcal{H}\\ h_B\in\mathcal{H}}} \max_{\substack{x\in\{X_1^*,\dots,X_m^*\}\\ x\notin\hat{E}(h,h_B)\\ x\in F_0}} \frac{\sqrt{nh^p}\tau_n(x,h)}{\sqrt{\hat{\mu}_n(x,h)}}\{d_n(x,h) - d_n^B(x_B,h_B)\}. \tag{A.133}$$

Besides, under the event defined in (A.130), we can similarly show

$$\widetilde{T}_n^B = \max_{\substack{h\in\mathcal{H}\\ h_B\in\mathcal{H}}} \max_{\substack{x\in\{X_1^*,\dots,X_m^*\}\\ x\notin\hat{E}(h,h_B)\\ x\in F_0\cap F_0^c(h_{\max})}} \frac{\sqrt{nh^p}\tau_n(x,h)}{\sqrt{\hat{\mu}_n(x,h)}}\{d_n(x,h) - d_n^B(x_B,h_B)\}.$$

Therefore, we have under the events defined in (A.128)-(A.130),

$$\widetilde{T}_n^B \le \max_{h\in\mathcal{H}} \max_{\substack{x\in\{X_1^*,\dots,X_m^*\}\\ x\in F_0\cap F_0^c(h_{\max})}} \left| \frac{\sqrt{nh^p}\tau_n(x,h)}{\sqrt{\hat{\mu}_n(x,h)}} \right|. \tag{A.134}$$

For any $x \in F_0 \cap F_0^c(h_{\max})$, $h \in \mathcal{H}$, we have $\tau_0(y) = 0$ for any $\|y-x\|_\infty \le h/2$. As are result, we have $K\{(x-X_i)/h\}\tau_0(X_i) = 0$, $\forall i = 1,\dots,n$ and hence $\mathrm{E}\tau_n(x,h) = 0$, $\hat{\mu}_n(x,h)$ equals

$$\underbrace{\frac{1}{nh^p}\sum_{i=1}^n \left\{ \left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i}\right) Y_i K\left(\frac{x-X_i}{h}\right) - \mathrm{E}\left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i}\right) Y_i K\left(\frac{x-X_i}{h}\right) \right\}^2}_{\mu_n^*(x,h)}.$$

This together with (A.128)-(A.130), (A.134) yields

$$\widetilde{T}_n^B \le \max_{h\in\mathcal{H}} \max_{\substack{x\in\{X_1^*,\dots,X_m^*\}\\ x\in F_0}} \left| \frac{\sqrt{nh^p}\{\tau_n(x,h) - \mathrm{E}\tau_n(x,h)\}}{\sqrt{\mu_n^*(x,h)}} \right|, \tag{A.135}$$

with probability tending to 1.

When $\nu(F_0) = 0$, we have $X_1^* \notin F_0, X_2^* \notin F_0, \ldots, X_m^* \notin F_0$ and $\widetilde{T}_n^B = 0$, almost surely. By (A.135), this implies $\Pr(\widetilde{T}_n^B = 0) \to 1$. Hence, it suffices to consider the case where $\nu(F_0) > 0$.

For any $x \in \Omega$, we have

$$\begin{aligned} \text{(A.136)} \qquad & \sum_{i=1}^{n} \operatorname{Var}\left\{ \frac{1}{\sqrt{nh^p}} \left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i} \right) Y_i K\left(\frac{x - X_i}{h} \right) \right\} \\ = \quad & \mu_n(x) - h^p \left\{ \frac{1}{h^p} \mathrm{E}\tau(X) K\left(\frac{x-X}{h} \right) \right\}^2 . \end{aligned}$$

Since $h \in \mathcal{H}$ and $h_{\max} \to 0$, standard bias calculation yields

$$\sup_{x \in \Omega, h \in \mathcal{H}} \left| \frac{1}{h^p} \mathrm{E}\tau(X) K\left(\frac{x - X}{h} \right) - \tau_0(x) \right| = o(1).$$

Besides, it follows from Lemma A.3 that

$$\text{(A.137)} \qquad \sup_{x \in \Omega, h \in \mathcal{H}} |\mu_n(x) - \mu(x)| \to 0.$$

By Assumption (A5), (A.136) and (A.137), we have

$$c_2 \le \sum_{i=1}^{n} \operatorname{Var}\left\{ \frac{1}{\sqrt{nh^p}} \left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i} \right) Y_i K\left(\frac{x - X_i}{h} \right) \right\} \le c_3, \qquad \forall h \in \mathcal{H},$$

for some constants $0 < c_2 \le c_3 < 0$.

For any $i \in \{1, \ldots, n\}$, it follows from Jensen's inequality and Cauchy-Schwarz inequality that

$$\begin{aligned} & \mathrm{E}\left| \frac{1}{\sqrt{nh^p}} \left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i} \right) Y_i K\left(\frac{x - X_i}{h} \right) - \mathrm{E} \frac{1}{\sqrt{nh^p}} \left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i} \right) Y_i K\left(\frac{x - X_i}{h} \right) \right|^3 \\ \le & \ \mathrm{E}\left| \frac{1}{\sqrt{nh^p}} \left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i} \right) Y_i K\left(\frac{x - X_j}{h} \right) - \frac{1}{\sqrt{nh^p}} \left(\frac{A_j}{\pi_j} - \frac{1-A_j}{1-\pi_j} \right) Y_j K\left(\frac{x - X_j}{h} \right) \right|^3 \\ & \le 8\mathrm{E}\left| \frac{1}{\sqrt{nh^p}} \left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i} \right) Y_i K\left(\frac{x - X_j}{h} \right) \right|^3 , \end{aligned}$$

for some $j \neq i$ and $j \in \{1, \ldots, n\}$. In addition, it follows from Lemma A.5 that

$$\sum_{i=1}^{n} \mathrm{E}\left| \frac{1}{\sqrt{nh^p}} \left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i} \right) Y_i K\left(\frac{x - X_j}{h} \right) \right|^3 = O\left(\frac{1}{\sqrt{nh^p}} \right).$$

Therefore,

$$\max_{h\in\mathcal{H}}\sum_{i=1}^{n}\mathrm{E}\left|\frac{1}{\sqrt{nh^p}}\left(\frac{A_i}{\pi_i}-\frac{1-A_i}{1-\pi_i}\right)Y_iK\left(\frac{x-X_j}{h}\right)\right|^3=O\left(\frac{1}{\sqrt{nh_{\min}^p}}\right).$$

As a result, it follows from Theorem 7.4 in de la Peña et al. (2009) that for any $x\in\Omega$, $h\in\mathcal{H}$ and $0<t\ll n^{1/6}h_{\min}^{p/6}$,

$$\Pr\left(\frac{\sqrt{nh^p}\{\tau_n(x,h)-\mathrm{E}\tau_n(x,h)\}}{\sqrt{\mu_n^*(x,h)}}\geq t\right)\leq\{1-\Phi(t)\}\left(1+O(1)\frac{(1+t)^3}{\sqrt{nh_{\min}^p}}\right).$$

Similarly, we have

$$\Pr\left(\frac{\sqrt{nh^p}\{\tau_n(x,h)-\mathrm{E}\tau_n(x,h)\}}{\sqrt{\mu_n^*(x,h)}}\leq -t\right)\leq\{1-\Phi(t)\}\left(1+O(1)\frac{(1+t)^3}{\sqrt{nh_{\min}^p}}\right),$$

and hence

$$\Pr\left(\left|\frac{\sqrt{nh^p}\{\tau_n(x,h)-\mathrm{E}\tau_n(x,h)\}}{\sqrt{\mu_n^*(x,h)}}\right|\geq t\right)\leq 2\{1-\Phi(t)\}\left(1+O(1)\frac{(1+t)^3}{\sqrt{nh_{\min}^p}}\right).$$

Let $x_1^*,\dots,x_m^*$ be a realization of $X_1^*,\dots,X_m^*$. By Bonferroni's inequality, we have for any $0<t\ll n^{1/6}h_{\min}^{p/6}$,

$$\begin{aligned}&\Pr\left(\max_{\substack{x\in\{x_1^*,\dots,x_m^*\},x\in F_0\\ h\in\mathcal{H}}}\left|\frac{\sqrt{nh^p}\{\tau_n(X_j,h)-\mathrm{E}\tau_n(X_j,h)\}}{\sqrt{\mu_n^*(X_j,h)}}\right|\geq t\right)\\ &\leq 2k_h\sum_j I(x_j^*\notin F_0)\{1-\Phi(t)\}\left(1+O(1)\frac{(1+t)^3}{\sqrt{nh_{\min}^p}}\right).\end{aligned}\tag{A.138}$$

Under the events defined in (A.128)-(A.130), we have $\widetilde{T}_n^B=0$ when $\sum_j I(x_j^*\in F_0)=0$. Hence, the null will be rejected with probability tending to 0. Therefore, it suffices to consider the case where $\sum_j I(x_j^*\in F_0)>0$. Let $t_0=z_{\alpha/\{2k_h\sum_j I(x_j^*\in F_0)\}}$. Since $k_h\to\infty$ as $n\to\infty$, we have $1<z_{\alpha/2k_h}\leq t_0$ for any $0<\alpha<1$. By definition,

$$\frac{1}{\sqrt{2\pi}}\int_{t_0}^{\infty}\exp\left(-\frac{x^2}{2}\right)dx=\frac{\alpha}{2k_h\sum_{j=1}^m I(x_j^*\in F_0)}.$$

Therefore, we have

$$\int_{t_0}^{\infty}x\exp\left(-\frac{x^2}{2}\right)dx\geq\frac{\alpha}{2k_h\sum_{j=1}^m I(x_j^*\in F_0)},$$

and hence

$$\exp\left(-\frac{t_0^2}{2}\right) \ge \frac{\alpha}{2k_h \sum_{j=1}^m I(x_j^* \in F_0)}.$$

This further implies that

$$t_0 \le \left(2\log[\{2k_h \sum_{j=1}^m I(x_j^* \in F_0)\}/\alpha]\right)^{1/2}. \tag{A.139}$$

Since $m = o(\eta_n^{-\xi_0})$ and $\eta_n \gg \sqrt{\log n}/\sqrt{nh_{\min}^p}$, we have $m = O(n^{l_*})$ for some $l_* > 0$. As a result, $t_0 = O(\sqrt{\log n})$. It follows from Condition (D1) that $t_0^3 \ll \sqrt{nh_{\min}^{pW}}$. By (A.138), we have

$$\Pr\left(\max_{\substack{x\in\{x_1^*,\dots,x_m^*\},x\in F_0\\ h\in\mathcal{H}}}\left|\frac{\sqrt{nh^p}\{\tau_n(X_j,h)-\mathrm{E}\tau_n(X_j,h)\}}{\sqrt{\mu_n^*(X_j,h)}}\right| \ge t_0\right)$$
$$\le 2k_h\sum_j I(x_j^* \notin F_0)\{1-\Phi(t_0)\} + o(1) = \alpha + o(1). \tag{A.140}$$

Let $\hat{t} = z_{\alpha/(2\hat{k})}$. Using similar arguments in (A.131)-(A.133), we can show with probability tending to 1,

$$I(X_j^* \in \hat{F}(h_W, h_B)) = I(X_j^* \in F_0),\ \forall j = 1,\dots,m, h_W, h_B \in \mathcal{H}.$$

This implies

$$\Pr(t_0 = \hat{t}) \to 1,$$

when $\sum_{j=1}^m I(x_j^* \in F_0) = 0$. Combining this together with (A.140) yields

$$\begin{aligned}
&\Pr\left(\max_{\substack{x\in\{x_1^*,\dots,x_m^*\},x\in F_0\\ h\in\mathcal{H}}}\left|\frac{\sqrt{nh^p}\{\tau_n(X_j,h)-\mathrm{E}\tau_n(X_j,h)\}}{\sqrt{\mu_n^*(X_j,h)}}\right| \ge \hat{t}\right) \\
= \ &\Pr\left(\max_{\substack{x\in\{x_1^*,\dots,x_m^*\},x\in F_0\\ h\in\mathcal{H}}}\left|\frac{\sqrt{nh^p}\{\tau_n(X_j,h)-\mathrm{E}\tau_n(X_j,h)\}}{\sqrt{\mu_n^*(X_j,h)}}\right| \ge t_0\right) \\
+ \ &\Pr(\hat{t} \ne t_0) \le \alpha + o(1).
\end{aligned} \tag{A.141}$$

Notice that the little-o term is independent of $\{x_1^*,\dots,x_m^*\}$, the inequality in (A.141) also holds for any realizations of $X_1^*,\dots,X_m^*$. In view of (A.135) and (A.141), we've shown

$$\Pr(\widetilde{T}_n^B \ge \hat{t}) \le \alpha + o(1).$$

The proof is hence completed.

A.13 Proof of Theorem 13.2. For ease of notation, we write X_1^{W*}, $\cdots$, X_m^{W*} as $X_1^*, \dots, X_m^*$. Let $\rho_0 = \sup_{x\in \dot{F}_0} |\delta_0(x)|$. Since $\rho_0 > 0$ and $\delta_0(\cdot)$ is continuous, there exists some $x_0 \in \dot{F}_0$ such that $|\delta_0(x_0)| \geq \rho_0/2$. Without loss of generality, assume $\delta_0(x_0) \geq \rho_0/2$. Hence, we have $\tau_0^B(x_{0,B}) \leq -\rho_*$ for some $\rho_* > 0$. By the continuity of δ_0 and τ_0^B, there exists some $\varepsilon_0 > 0$ such that for any $\|x - x_0\|_\infty \leq \varepsilon_0$, we have $x \in \dot{F}_0$, $\delta_0(x) \geq \rho_0/4$ and $\tau_0^B(x_B) \leq -\rho_*/2$. Define the set

$$F_* = \left\{x \in F_0 : \|x - x_0\|_\infty \leq \frac{\varepsilon_0}{2}\right\}.$$

Apparently, we have $\nu(F_*) > 0$. Since $m \to \infty$, the following event happens with probability tending to 1,

$$X_{j_0}^* \in F_*, \quad \text{for some } 1 \leq j_0 \leq m. \tag{A.142}$$

Let $x_1^*, \dots, x_m^*$ be a realization of $X_1^*, \dots, X_m^*$. Since $h_{\max} \to 0$, we have $x \in F_0$ for any x such that $\|x - x_{j_0}^*\| \leq h_{\max}/2$. Besides, it follows from Assumption (A1) that $\bar{\rho} \equiv \inf_{x\in\Omega} f(x) > 0$. This implies $\inf_{x_B\in\Omega^B} f^B(x_B) \geq \bar{\rho}\nu(\Omega^C)$. Since $f(\cdot)$ has uniformly bounded derivatives, $f(\cdot)$ is uniformly continuous. By assumption, $\delta_0(\cdot)$ is also uniformly continuous. using similar arguments in the proof of Lemma A.3, we can show

$$\begin{aligned}
\mathrm{E}\tau_n(x_{j_0}^*, h_{\max}) &= \frac{1}{h_{\max}^p} \mathrm{E}K\left(\frac{x_{j_0}^* - X}{h_{\max}}\right)\{\tau_0(X) + a_n\delta_0(X)\} \\
&= \frac{a_n}{h_{\max}^p}\int_{\|x - x_{j_0}^*\|_\infty \leq h_{\max}/2} K\left(\frac{x_{j_0}^* - x}{h_{\max}}\right)\delta_0(x)f(x)dx \\
&= a_n \int_{\|z\|\leq h_{\max}/2} K(z)\delta_0(x_{j_0}^* - zh_{\max})f(x_{j_0}^* - zh_{\max})dx \\
&= a_n\delta_0(x_{j_0}^*)f(x_{j_0}^*) + o(a_n) \geq \frac{a_n\rho_0\bar{\rho}}{4} + o(a_n).
\end{aligned} \tag{A.143}$$

Since $a_n \to 0$ and δ_0 is bounded, we can similarly show

$$\begin{aligned}
&\mathrm{E}\tau_n^B(x_{j_0,B}^*, h_{\max}) = \frac{1}{h_{\max}^p}\mathrm{E}K^B\left(\frac{x_{j_0,B}^* - X_B}{h_{\max}}\right)[\tau_0^B(X_B) + a_n\mathrm{E}\{\delta_0(X)|X_B\}] \\
&= \frac{1}{h_{\max}^p}\int_{\|x_B - x_{j_0,B}^*\|_\infty \leq h_{\max}/2} K^B\left(\frac{x_{j_0,B}^* - x_B}{h_{\max}}\right)\tau_0^B(x_B)f^B(x_B)dx_B + o(1) \\
&\leq -\frac{\rho_*\bar{\rho}\nu(\Omega^C)}{2} + o(1) \leq -\frac{\rho_*\bar{\rho}\nu(\Omega^C)}{4},
\end{aligned} \tag{A.144}$$

for sufficiently large n.

Similar to Lemma A.2, we can show that the following events happens with probability 1,

$$\sup_{x\in\Omega}|\tau_n(x,h_{\max})-\mathrm{E}\tau_n(x,h_{\max})|\leq\frac{\gamma\sqrt{\log n}}{\sqrt{nh_{\max}^p}},\quad\forall n\geq K,\tag{A.145}$$

$$\sup_{x_B\in\Omega^B}\left|\tau_n^B(x_B,h_{\max})-\mathrm{E}\tau_n^B(x_B,h_{\max})\right|\leq\frac{\gamma\sqrt{\log n}}{\sqrt{nh_{\max}^{p_B}}},\quad\forall n\geq K,\tag{A.146}$$

for some constants $\gamma,K>0$.

Under the events defined in (A.142), (A.145) and (A.146), it follows from (A.143), (A.144) and the condition $a_n\gg\sqrt{\log n}/\sqrt{nh_{\max}^{p_B}}$ that for sufficiently large n,

$$\tau_n(x_{j_0}^*)\geq\frac{a_n\rho_0\bar{\rho}}{8},\quad\text{and}\quad\tau_n^B(x_{j_0,B}^*)<0,\tag{A.147}$$

with probability tending to 1. Similarly, we can show the following event happens with probability tending to 1,

$$|\tau_n^B(x_{j_0,B}^*)|>\eta_n.\tag{A.148}$$

In addition, it follows from (A.55) that

$$\hat{\mu}_n(x_{j_0}^*,h_{\max})\leq 4c_3,$$

for some constant $c_3>0$. This together with (A.147) and (A.148) yields

$$\begin{aligned}\widetilde{T}_n^B&\geq\frac{\sqrt{nh_{\max}^p}\tau_n(x_{j_0}^*,h_{\max})}{\sqrt{\hat{\mu}_n(x_{j_0}^*,h_{\max})}}\{d_n(x_{j_0}^*,h_{\max})-d_n^B(x_{j_0,B}^*,h_{\max})\}I\{x_{j_0}^*\notin\hat{E}(h_{\max},h_{\max})\}\\&\geq\frac{\sqrt{nh_{\max}^p}\tau_n(x_{j_0}^*,h_{\max})}{\sqrt{\hat{\mu}_n(x_{j_0}^*,h_{\max})}}\geq\frac{\sqrt{nh_{\max}^p}a_n\rho_0\bar{\rho}}{16\sqrt{c_3}},\end{aligned}\tag{A.149}$$

with probability tending to 1.

Similar to (A.139), we can show $z_{\alpha/(2\hat{k})}\leq z_{\alpha/(2k_h|\mathcal{H}|^2)}=O(\sqrt{\log n})$. Since $a_n\gg\sqrt{\log n}/\sqrt{nh_{\max}^p}$, it follows from (A.149) that $\Pr(\widetilde{T}_{n,B}\geq z_{\alpha/(2\hat{k})})\to 1$. The proof is hence completed.

A.14 Proof of Theorem 14.1. We first show the Type-I error rates of the test can be well controlled under H_0. Since all covariates are discrete, we have

$$\inf_{x\in\Omega}f(x)>0\quad\text{and}\quad\inf_{x_B\in\Omega^B}f^B(x_B)>0.\tag{A.150}$$

Under the given conditions, using Chebyshev's inequality, we can show

$$\begin{aligned}
&\Pr(|\tau_n(x) - \tau_0(x)| > c_n n^{-1/2}) \to 0, \quad \Pr(|\tau_n^B(x_B) - \tau_0(x_B)| > c_n n^{-1/2}) \to 0,\\
&\Pr(|f(x) - \hat{f}(x)| > c_n n^{-1/2}) \to 0, \quad \Pr(|f^B(x_B) - \hat{f}^B(x_B)| > c_n n^{-1/2}) \to 0,\\
&\Pr(|\hat{\mu}(x) - \mu(x)| > c_n n^{-1/2}) \to 0.
\end{aligned}$$

for any $x \in \Omega$, $x_B \in \Omega^B$, and any sequence $c_n \to \infty$.

Since $|\Omega| < \infty$, we have $|\Omega^B| < \infty$. It follows from Bonferroni's inequality that

$$\Pr(\mathcal{A}_1^c \cup \mathcal{A}_2^c \cup \mathcal{A}_3^c) \to 0, \tag{A.151}$$

where

$$\begin{aligned}
\mathcal{A}_1 &= \left\{\max_{x\in\Omega} |\tau_n(x) - \tau_0(x)| \le c_n n^{-1/2}\right\} \cup \left\{\max_{x_B\in\Omega^B} |\tau_n^B(x_B) - \tau_0^B(x_B)| \le c_n n^{-1/2}\right\},\\
\mathcal{A}_2 &= \left\{\max_{x\in\Omega} |\hat{f}(x) - f(x)| \le c_n n^{-1/2}\right\} \cup \left\{\max_{x_B\in\Omega^B} |f^B(x_B) - \hat{f}^B(x_B)| \le c_n n^{-1/2}\right\},\\
\mathcal{A}_3 &= \left\{\max_{x\in\Omega} |\hat{\mu}(x) - \mu(x)| \le c_n n^{-1/2}\right\}.
\end{aligned}$$

Since η_n satisfies $\eta_n \gg n^{-1/2}$ and $\eta_n \to 0$, we can find a sequence c_n that satisfies $c_n \to \infty$, $\eta_n \gg c_n n^{-1/2}$, $c_n \ll n^{1/2}$. By (A.150), for such a choice of $\{c_n\}_n$, we have on the set $\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3$,

$$\hat{E} = E_0 \quad \text{and} \quad \hat{F} = F_0, \tag{A.152}$$

$$\begin{aligned}
&I(\tau_n(x) \ge 0) = I(\tau_0(x) \ge 0), \quad \forall x \in \Omega \ \text{ with } \ \tau_0(x) \ne 0,\\
&I(\tau_n^B(x_B) \ge 0) = I(\tau_0^B(x_B) \ge 0), \quad \forall x_B \in \Omega^B \ \text{ with } \ \tau_0^B(x_B) \ne 0.
\end{aligned} \tag{A.153}$$

for sufficiently large n, where $E_0 = \{x \in \Omega : \tau_0(x) = 0, \tau_0^B(x^B) = 0\}$. In addition, similar to Theorem 2.2, we can show that under the null,

$$\tau_0(x) \ge 0, \ \text{ if } \ \tau_0^B(x_B) \ge 0 \quad \text{and} \quad \tau_0(x) \le 0, \ \text{ if } \ \tau_0^B(x_B) \le 0.$$

This implies

$$I(\tau_n(x) \ge 0) = I(\tau_n^B(x_B) \ge 0), \text{if } \ \tau_0(x) \ne 0, \tau_0^B(x_B) \ne 0. \tag{A.154}$$

Since $\tau_0^B(x_B) = \mathrm{E}\{\tau_0(X)|X^B = x_B\}$, under H_0, we have

$$\tau_0(x) = 0, \ \text{ if } \ \tau_0^B(x_B) = 0.$$

This means that under H_0,

$$\{x \in \Omega : \tau_0(x) \neq 0, \tau_0^B(x_B) = 0\} = \emptyset. \tag{A.155}$$

Therefore, we have on the set $\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3$,

$$\begin{aligned}
\widetilde{S}_n^B &= \sum_{x \notin E_0} \tau_n(x)\{I(\tau_n(x) \geq 0) - I(\tau_n^B(x_B) \geq 0)\} \\
&= \sum_{\{x \notin E_0 : \tau_0(x) \neq 0, \tau_0^B(x_B) \neq 0\}} \tau_n(x)\{I(\tau_n(x) \geq 0) - I(\tau_n^B(x_B) \geq 0)\} \\
&+ \sum_{\{x \notin E_0 : \tau_0(x) \neq 0, \tau_0^B(x_B) = 0\}} \tau_n(x)\{I(\tau_n(x) \geq 0) - I(\tau_n^B(x_B) \geq 0)\} \\
&+ \sum_{\{x \notin E_0 : \tau_0(x) = 0, \tau_0^B(x_B) \neq 0\}} \tau_n(x)\{I(\tau_n(x) \geq 0) - I(\tau_n^B(x_B) \geq 0)\} \\
&= \sum_{\{x \notin E_0 : \tau_0(x) = 0, \tau_0^B(x_B) \neq 0\}} \tau_n(x)\{I(\tau_n(x) \geq 0) - I(\tau_n^B(x_B) \geq 0)\} \\
&= \sum_{x \in F_0} \tau_n(x)\{I(\tau_n(x) \geq 0) - I(\tau_0^B(x_B) \geq 0)\},
\end{aligned}$$

where the third equality follows by (A.154) and (A.155), and the last equality is due to (A.153).

Consider the case where $F_0 = \emptyset$. It follows from (A.152) that $F_0 = \emptyset$ on the set $\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3$ and hence $\sqrt{n}\widetilde{S}_n^B = 0$. Therefore, it follows from (A.151) that

$$\Pr(\widetilde{S}_n^B > \hat{c}_\alpha(\Omega)) \leq \Pr(\mathcal{A}_1^c \cup \mathcal{A}_2^c \cup \mathcal{A}_3^c) \to 0.$$

Now consider the case where $F_0 \neq \emptyset$. On the set $\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3$, we have $\hat{F} \neq \emptyset$ and hence $\hat{c}_\alpha = \hat{c}_\alpha(\hat{F})$. By (A.151), we have

$$\left| \sqrt{n}\widetilde{S}_n^B - \sqrt{n} \sum_{x \in F_0} \tau_n(x)\{I(\tau_n(x) \geq 0) - I(\tau_0^B(x_B) \geq 0)\} \right| \xrightarrow{P} 0. \tag{A.156}$$

By the definition of F_0, we have $\mathrm{E}\tau_n(x) = \tau_0(x) = 0, \forall x \in F_0$. In addition, we have for any $x, y \in F_0$, $x \neq y$,

$$n\mathrm{cov}\{\tau_n(x), \tau_n(y)\} = \mathrm{E}\left(\frac{A}{\pi(X)} - \frac{1-A}{1-\pi(X)}\right)^2 Y^2 I(X = x) I(X = y) - \tau_0(x)\tau_0(y) = 0.$$

It follows from central limit theorem that

$$\{\sqrt{n}\tau_n(x)\}_{x \in F_0} \xrightarrow{d} \{\sqrt{\mu(x)}\mathbb{Z}_x\}_{x \in F_0},$$

where $\{\mathbb{Z}_x\}_{x\in F_0}$ are independent standard normal variables.

By the continuous mapping theorem, we have

$$\begin{aligned}\sqrt{n}\sum_{x\in F_0}\tau_n(x)\{I(\tau_n(x)\geq 0)-I(\tau_0^B(x_B)\geq 0)\}\\ \stackrel{d}{\to}\sum_{x\in F_0}\sqrt{\mu(x)}\left(\max(\mathbb{Z}_x,0)-\mathbb{Z}_x I(\tau_0^B(x_B)\geq 0)\right).\end{aligned}\tag{A.157}$$

Notice that for any $x\in F_0$,

$$\max(\mathbb{Z}_x,0)-\mathbb{Z}_x I(\tau_0^B(x_B)\geq 0)\stackrel{d}{=}\mathbb{Z}_x.$$

In addition, $\mathbb{Z}_x$'s are independent. Therefore,

$$\sum_{x\in F_0\sqrt{\mu(x)}}\left(\max(\mathbb{Z}_x,0)-\mathbb{Z}_x I(\tau_0^B(x_B)\geq 0)\right)\stackrel{d}{=}\sum_{x\in F_0}\sqrt{\mu(x)}\max(\mathbb{Z}_x,0).$$

Combining this together with (A.157) implies that

$$\sqrt{n}\sum_{x\in F_0}\tau_n(x)\{I(\tau_n(x)\geq 0)-I(\tau_0^B(x_B)\geq 0)\}\stackrel{d}{\to}\sum_{x\in F_0}\sqrt{\mu(x)}\max(\mathbb{Z}_x,0).$$

By (A.156) and Slutsky's theorem, we obtain

$$\sqrt{n}\widetilde{S}_n^B\stackrel{d}{\to}\sum_{x\in F_0}\sqrt{\mu(x)}\max(\mathbb{Z}_x,0).\tag{A.158}$$

Under the given conditions, we have $\min_{x\in F_0}\sqrt{\mu(x)}>0$. When $\mathcal{A}_3$ holds, we have

$$\hat{c}_\alpha\to c_\alpha.\tag{A.159}$$

Thus, for any $\varepsilon>0$, it follows from (A.158) and (A.151) that

$$\Pr(\sqrt{n}\widetilde{S}_n^B>\hat{c}_\alpha)\leq\Pr(\mathcal{A}_3^c)+\Pr(\sqrt{n}\widetilde{S}_n^B>c_\alpha-\varepsilon)=o(1)+G_0(c_\alpha-\varepsilon).$$

Since ε can be arbitrarily chosen, we have

$$\limsup_n\Pr(\sqrt{n}\widetilde{S}_n^B>\hat{c}_\alpha)\leq G_0(c_\alpha)=\alpha.$$

Similarly, we can show

$$\liminf_n\Pr(\sqrt{n}\widetilde{S}_n^B>\hat{c}_\alpha)\geq\alpha.$$

This implies that $\lim_n \Pr(\sqrt{n}\widetilde{S}_n^B > \hat{c}_\alpha) = \alpha$, when $\nu(F_0) \neq 0$. Thus, the Type-I error rates can be well controlled under H_0.

Under H_1, there exists some $y \in \Omega$ such that $\tau_0(y) > 0$, $\tau_0^B(y_B) \leq 0$ or $\tau_0(y) < 0$, $\tau_0^B(y_B) \geq 0$. For brevity, we only consider the case where $\tau_0(y) > 0$, $\tau_0^B(y_B) \leq 0$. Further assume $\tau_0^B(y_B) < 0$. Under the event defined in $\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3$, we have

$$\sqrt{n}\widetilde{S}_n^B \geq \sqrt{n}\tau_n(y)\{I(\tau_n(y) \geq 0) - I(\tau_n^B(y_B) \geq 0)\} \geq \sqrt{n}|\tau_n(y)| \geq \frac{\sqrt{n}\tau_0(y)}{2},$$

for sufficiently large n. This together with (A.159) and (A.151) gives

$$\Pr\left(\sqrt{n}\widetilde{S}_n^B > \hat{c}_\alpha\right) \geq \Pr\left(\left\{\sqrt{n}\widetilde{S}_n^B > \hat{c}_\alpha\right\} \cap \mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3\right)$$
$$\geq \Pr\left(\left\{\sqrt{n}\tau_0(y)/2 \geq c_\alpha + \varepsilon\right\} \cap \mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3\right) \geq 1 - \Pr(\mathcal{A}_1^c \cup \mathcal{A}_2^c \cup \mathcal{A}_3^c) \to 1,$$

for some sufficiently small $\varepsilon > 0$.

Now assume $\tau_0^B(y_B) = 0$. Since $\tau(y) > 0$ and $\tau_0^B(y_B) = \mathrm{E}\{\tau_0(X)|X^B = y_B\}$, there exists some $y^* \in \Omega$ such that $\tau(y^*) < 0$ and $y_B^* = y_B$. Therefore,

$$\sqrt{n}\widetilde{S}_n^B \geq \sqrt{n}\tau_n(y^*)[I(\tau_n(y^*) \geq 0) - I\{\tau_n^B(y_B) \geq 0\}] \geq \sqrt{n}|\tau_n(y^*)| \geq \frac{\sqrt{n}\tau_0(y^*)}{2}.$$

As a result, we can similarly show that the $\Pr(\sqrt{n}\widetilde{S}_n^B > \hat{c}_\alpha) \to 1$. The proof is hence completed.

A.15 Technical lemmas.

Lemma A.1. *Assume $X_j, j = 1, \ldots,$ are i.i.d random variables distributed according as X_0. Let η be a Poisson random variable with mean λ, independent of $\{X_j, j = 1, \ldots\}$. Then for any disjoint sets $\mathcal{C}$, $\mathcal{D}$, and measurable functions g and f, we have*

$$I_1 = \sum_{j=1}^{\eta} g(X_i) I(X_i \in \mathcal{C}) \perp\!\!\!\perp I_2 = \sum_{j=1}^{\eta} f(X_j) I(X_i \in \mathcal{D}).$$

Proof: With some calculation, we can show the characteristic function of $t_1 I_1 + t_2 I_2$ is

(A.160)
$$\mathrm{E}\exp(it_1 I_1 + it_2 I_2) = \exp\left\{\lambda\left(\mathrm{E}e^{it_1 g(X_0)I(X_0\in\mathcal{C}) + it_2 f(X_0)I(X_0\in\mathcal{D})} - 1\right)\right\}$$
$$= \exp\left[\lambda\left\{\mathrm{E}\left(e^{it_1 g(X_0)} - 1\right) I(X_0 \in \mathcal{C}) + \mathrm{E}\left(e^{it_2 f(X_0)} - 1\right) I(X_0 \in \mathcal{D})\right\}\right],$$

where the last equality is due to that $\mathcal{C}$ and $\mathcal{D}$ are disjoint. Similarly we can show

$$\begin{aligned}
\mathrm{E}\exp(it_1 I_1) &= \exp\left\{\lambda\left(\mathrm{E}e^{it_1 g(X_0)I(X_0\in\mathcal{C})}-1\right)\right\}\\
&= \exp\left[\lambda\left\{\mathrm{E}\left(e^{it_1 g(X_0)}-1\right)I(X_0\in\mathcal{C})\right\}\right],\\
\mathrm{E}\exp(it_2 I_2) &= \exp\left\{\lambda\left(\mathrm{E}e^{it_2 f(X_0)I(X_0\in\mathcal{D})}-1\right)\right\}\\
&= \exp\left[\lambda\left\{\mathrm{E}\left(e^{it_2 f(X_0)}-1\right)I(X_0\in\mathcal{D})\right\}\right].
\end{aligned}$$

Combining these together with (A.160), we obtain

$$\mathrm{E}\exp(it_1 I_1 + it_2 I_2) = \mathrm{E}\exp(it_1 I_1)\mathrm{E}\exp(it_2 I_2).$$

The independence between I_1 and I_2 hence follows.

LEMMA A.2. *Under conditions in Theorem 3.6 or 3.14, we have almost surely, for all $x \in \Omega$ and sufficiently large n,*

$$|\tau_n(x) - \tau_0(x)f(x)| = O\left(\frac{\sqrt{\log n}}{\sqrt{nh^p}}\right),$$
$$|\tau_n^B(x_B) - \tau_0^B(x_B)f^B(x_B)| = O\left(\frac{\sqrt{\log n}}{\sqrt{n(h_B)^{p_B}}}\right).$$

Proof: For brevity, we only show for sufficiently large n,

$$|\tau_n(x) - \tau_0(x)f(x)| = O\left(\frac{\sqrt{\log n}}{\sqrt{nh^p}}\right),$$

almost surely.

By Assumption (A1), f and τ_0 have bounded derivatives of order s. Standard bias calculation in kernel estimation yields

$$|\mathrm{E}\tau_n(x) - \tau_0(x)f(x)| = O(h^s).$$

Assumptions (A6) implies $h^s \ll \sqrt{\log n}/\sqrt{nh^p}$, and it suffices to show almost surely,

$$|\tau_n(x) - \mathrm{E}\tau_n(x)| = O\left(\frac{\sqrt{\log n}}{\sqrt{nh^p}}\right). \tag{A.161}$$

We begin by introducing a truncated version of $\tau_n(x)$. Let

$$\tilde{\tau}_n(x) = \frac{1}{n}\sum_{i=1}^n K_h(x - X_i)\left(\frac{A_i}{\pi_i} - \frac{1-A_i}{1-\pi_i}\right)Y_i I(|Y_i| \le c_n),$$

where $c_n = 3\log(n)/t$ and the constant t is defined in (A3).

Observe that $\sum_{n=2} 1/(n^2) < \infty$. It follows from Markov's inequality that

$$\Pr(\max_i |Y_i| \geq c_n) \leq n\Pr(|Y_1| \geq c_n) \leq n\mathrm{E}\frac{\exp(t|Y_1|)}{\exp(3t\log(n)/t)} \leq O\left(\frac{1}{n^2}\right),$$

and hence

$$\sum_{n=1}^{\infty} \Pr(|Y_i| \leq c_n) < \infty. \tag{A.162}$$

It follows from (A.162) and Borel-Cantelli Lemma that with probability 1, there exists some K such that for all $n \geq K$, $|Y_n| \leq c_n$. Observe that $c_n \geq c_K$ for all $n \geq K$. Hence, with probability 1, for sufficiently large n, we have $|Y_i| \leq c_n$ for all $i \leq n$. This suggests for all x and sufficiently large n, almost surely, we have $\tau_n(x) = \tilde{\tau}_n(x)$.

Besides, by Assumption (A4) and Cauchy-Swartz inequality, we obtain

$$\begin{aligned} |\mathrm{E}\tau_n(x) - \mathrm{E}\tilde{\tau}_n(x)| &\leq \frac{C}{h^p}\mathrm{E}\,|K_h(x - X_0)|\,|Y_0| I(|Y_0| > c_n) \\ &\leq \frac{Cc_n}{h^p}\sqrt{\mathrm{E}^2\,|K_h(x - X_0)|}\sqrt{\Pr(|Y_0| > c_n)}. \end{aligned} \tag{A.163}$$

for some constant $C > 0$. Besides,

$$\frac{1}{h^p}\mathrm{E}^2\,|K_h(x - X_0)| = \int_y |K(y)|^2 f(x - yh)dy = O(1), \tag{A.164}$$

by Assumptions (A1) and (A2). Moreover, it follows from (A3) that

$$\sqrt{\Pr(|Y_0| > c_n)} = O(n^{-3/2}). \tag{A.165}$$

Combining (A.163)-(A.165) together, we obtain

$$|\mathrm{E}\tau_n(x) - \mathrm{E}\tilde{\tau}_n(x)| = O\left(\frac{\sqrt{\log n}}{n}\right)\frac{\sqrt{\log n}}{\sqrt{nh^p}} \ll \frac{\sqrt{\log n}}{\sqrt{nh^p}}.$$

Therefore, it suffices to show

$$|\tilde{\tau}_n(x) - \mathrm{E}\tilde{\tau}_n(x)| = O\left(\frac{\sqrt{\log n}}{\sqrt{nh^p}}\right).$$

By Assumption (A2), $K = \prod_{j=1}^p K_j$ with each K_j of bounded variation, and is bounded. Lemma 22 in Nolan and Pollard (1987) that the class of functions $\kappa_j = \{K_j(a^T(x - X_0)/h^{1/p}) : x \in \mathbb{R}^p\}$ belongs to the Euclidean class for a constant envelope. It also belongs to VC class. By definition, there exists some constant A_j and V_j such that for all $0 < \varepsilon \le 1$,

$$N\left(\varepsilon\sqrt{QM_j^2}, \kappa_j, L_2(Q)\right) \le A_j \varepsilon^{-V_j}, \tag{A.166}$$

for some constant function M_j and all probability measures Q such that $0 < QM_j^2 < \infty$. See the definition of the covering number N in van der Vaart and Wellner (1996).

Let $A = \prod_j A_j$, $V = \sum_j V_j$, for all $0 < \varepsilon \le 1$, it follows from (A.166) that

$$N\left(\varepsilon \prod_j \sqrt{QM_j^2}, \kappa, L_2(Q)\right) \le A\varepsilon^{-V},$$

where κ denotes the class of functions $\{K((x - X_0)/h) : x \in \mathbb{R}^p\}$. Note that M_j is a constant envelope, we have

$$\prod_j \sqrt{QM_j^2} = \sqrt{Q \prod_j M_j^2},$$

and hence κ belongs to the VC class with constant envelope $\prod_j M_j$. It follows by Lemma 2.6.18 in van der Vaart and Wellner (1996) that the class of functions

$$\mathcal{F}_n = \left\{K\left(\frac{x - X_0}{h}\right)\left(\frac{A_0}{\pi_0} - \frac{1 - A_0}{1 - \pi_0}\right) Y_0 I(|Y_0| \le c_n) : x \in \mathbb{R}^p\right\}$$

belongs to the VC class. We can take the envelope function as $M_n = M|Y_0|I(Y_0 < c_n) + 1$ for some constant $M > 0$.

With some calculation, it follows from Assumption (A3) that

$$\sup_{x\in\Omega} n\mathrm{Var}\{\tilde{\tau}_n(x)\} = O\left(\frac{1}{h^p}\right). \tag{A.167}$$

For any $\gamma > 0$, take $\varepsilon_n = \sqrt{\log n}/\sqrt{nh^p}$, it follows by (A.167) that

$$\mathrm{Var}\left(\frac{\tilde{\tau}_n(x)}{\gamma\varepsilon_n}\right) = O\left(\frac{1}{\log n}\right) \ll \frac{1}{2}.$$

Therefore, by the symmetrization inequality (Lemma 2.3.7, van der Vaart and Wellner, 1996), we obtain that for all $\gamma > 0$,

$$\begin{aligned} &\Pr\left(\sup_{x\in\Omega} |\tilde{\tau}_n(x)| > 4\gamma\varepsilon_n\right) \\ \le\ & 4\Pr\left(\sup_{x\in\Omega}\left|\frac{1}{nh^p}\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)\left(\frac{A_i}{\pi_i} - \frac{1 - A_i}{1 - \pi_i}\right) Y_i I(|Y_i| \le c_n) e_i\right| > 2\gamma\varepsilon_n\right), \end{aligned} \tag{A.168}$$

for i.i.d Rademacher random variables e_i. Since $\mathcal{F}_n$ belongs to the VC class, we have

$$N(\delta\sqrt{\mathbb{P}_n M_n^2}, \mathcal{F}_n, L_2(\mathbb{P}_n)) \leq A_0 \delta^{-V_0}, \tag{A.169}$$

for some constant A_0 and $V_0 > 0$ and any $\delta > 0$. Take $\delta = h^p \varepsilon_n / \sqrt{\mathbb{P}_n M_n^2}$, it follows by (A.169) that

$$N(\delta\sqrt{\mathbb{P}_n M_n^2}, \mathcal{F}_n, L_2(\mathbb{P}_n)) \leq A_0 \left(\frac{h^p \varepsilon_n}{\sqrt{\mathbb{P}_n M_n^2}} \right)^{-V_0} = O(n^{l_0}),$$

for some constant $l_0 > 0$, by Assumption A6.

Hoeffding's inequality implies RHS of (A.168) can be bounded by

$$2N(\delta\sqrt{\mathbb{P}_n M_n^2}, \mathcal{F}_n, L_2(\mathbb{P}_n)) \mathrm{E} \exp\left(-\frac{n\gamma^2 h^{2p} \varepsilon_n^2}{\max_j \mathbb{P}_n g_j^2} \right), \tag{A.170}$$

where the maximum runs over all $N(\delta\sqrt{\mathbb{P}_n M_n^2}, \mathcal{F}_n, L_2(\mathbb{P}_n))$ functions $\{g_j\}$ in $\mathcal{F}_n$. Note that

$$\begin{aligned}
\mathrm{E} \exp\left(-\frac{n\gamma^2 h^{2p} \varepsilon_n^2}{\max_j \mathbb{P}_n g_j^2} \right) &= \mathrm{E} \exp\left(-\frac{n\gamma^2 h^{2p} \varepsilon_n^2}{\max_j \mathbb{P}_n g_j^2} \right) I(\max_j \mathbb{P}_n g_j^2 \leq \gamma^2 h^p / (l_0 + 2)) \\
&+ \mathrm{E} \exp\left(-\frac{n\gamma^2 h^{2p} \varepsilon_n^2}{\max_j \mathbb{P}_n g_j^2} \right) I(\max_j \mathbb{P}_n g_j^2 > \gamma^2 h^p / (l_0 + 2)) \\
&\leq n^{-l_0 - 2} + \Pr\left(\max_j \mathbb{P}_n g_j^2 > \frac{\gamma^2 h^p}{l_0 + 2} \right). \qquad \text{(A.171)}
\end{aligned}$$

Since each g_j is bounded by $O(c_n)$ by Assumptions (A2) and (A4). It follows from Bernstein's inequality (Lemma 2.2.9, van der Vaart and Wellner, 1996) that for any $t > 0$, and $j = 1, \ldots, N(\delta\sqrt{\mathbb{P}_n M_n^2}, \mathcal{F}_n, L_2(\mathbb{P}_n))$,

$$\Pr(\mathbb{P}_n g_j^2 - \mathrm{E} g_j^2 > t) \leq \exp\left(-\frac{\bar{c} n t^2}{2v_j + c_n^2 t / 3} \right), \tag{A.172}$$

for some constant $\bar{c} > 0$, where $v_j = \mathrm{Var}(g_j^2) \leq \mathrm{E} g_j^4 = O(c_n^4 h^p)$. By (A.167), we have $\mathrm{E} g_j^2 = O(h^p)$. We can choose γ such that $\gamma^2 h^p / (l_0 + 2) - \mathrm{E} g_j^2 \geq \gamma_0 h^p$ for some $\gamma_0 > 0$. Set $t = \gamma_0 h$ in (A.172), RHS is smaller than

$$\exp\left(-\frac{\bar{c} n \gamma_0^2 h^{2p}}{2v_j + c_n^2 \gamma_0 h^p / 3} \right). \tag{A.173}$$

For sufficiently large n, we have $v_j \leq c_n^5 \gamma_0 h^p/3$, and hence (A.173) is smaller than

$$\exp\left(-\frac{\bar{c} n \gamma_0^2 h^{2p}}{c_n^5 \gamma_0 h^p}\right) \leq \exp(-\gamma_0 \log n),$$

where the last inequality is due to that $nh^p \gg \log n^6$ since we assume $nh^p \gg h^{-p}$ and $h^{-1} \gg n^{1/(2s)}$ in (A6). Take sufficiently large γ, it follows from (A.170) and (A.171) that (A.168) is bounded by $O(n^{l_0})O(n^{-l_0-2}) = O(n^{-2})$. The almost sure convergence follows by an application of Borel-Cantelli Lemma.

LEMMA A.3. *Assume $h \to 0$ as $n \to \infty$. Under Assumption (A1), (A2) and (A5), we have*

$$\sup_{x\in\Omega} |\mu_n(x) - \mu(x)| = o(1).$$

Proof: Recall that

$$\mu_n(x) = \frac{1}{h^p K_*(0)} \mathrm{E}\left[\frac{\mu(X)}{f(X)}\left\{K\left(\frac{x-X}{h}\right)\right\}^2\right] = \frac{1}{h^p K_*(0)} \int_{y\in\Omega} \mu(y) K^2\left(\frac{x-y}{h}\right) dx.$$

By Assumption (A2), $K(z) = 0$ for any z such that $\|z\|_\infty > 1/2$. Using change of variables, we have

$$\mu_n(x) = \frac{1}{K_*(0)} \int_{\|z\|_\infty \leq 1/2} \mu(x - zh) K^2(z) dz.$$

By Assumption (A5), $\mu(\cdot)$ is uniform continuous. Notice that $K_*(0) = \int_{\|z\|_\infty \leq 1/2} K^2(z)dz$. Therefore, we have

$$\begin{aligned}\sup_{x\in\Omega} |\mu_n(x) - \mu(x)| &\leq \frac{1}{K_*(0)} \sup_{x\in\Omega} \int_{\|z\|_\infty \leq 1/2} |\mu(x-zh) - \mu(x)| K^2(z) dz \\ &\leq \sup_{x\in\Omega, \|z\|_\infty \leq 1/2} |\mu(x-zh) - \mu(x)| \to 0.\end{aligned}$$

The proof is hence completed.

LEMMA A.4. *Under conditions in Theorem 3.6 and 3.14, for any $x \in \Omega$, and $t \in [-1,1]^p$ such that $x + ht \in \Omega$, we have*

$$\begin{aligned}&\sqrt{nh^p}\{(\tau_N(x) - \tau_0(x)f(x))/\sqrt{\mu(x)}, (\tau_N(x+th) - \\ &\quad \tau_0(x+th)f(x+th))/\sqrt{\mu(x)}\}^T \xrightarrow{d} N(0, \Sigma(t)),\end{aligned}$$

where

$$\Sigma(t) = \begin{pmatrix} 1 & \rho(t) \\ \rho(t) & 1 \end{pmatrix}.$$

Proof: By Assumption (A6), we have $\sqrt{nh^p}(\mathrm{E}\tau_N(x) - \tau_0(x)f(x)) = o(1)$. With some calculations, we can show $\mathrm{Var}\{\sqrt{nh^p}\tau_N(x)\} = \mathrm{E}nh^p\tau_n^2(x) = \mu_n(x)$. By Lemma A.3, we have $\mu_n(x) \to \mu(x)$. Similarly, we can show $nh^p\mathrm{cov}\{\tau_N(x), \tau_N(x+th)\} \to \mu(x)\rho(t)$ and $nh^p\mathrm{Var}\{\tau_N(x+th)\} \to \mu(x)$. Since Poisson distribution is self-decomposable, the asymptotic normality follows by standard central limit theorem.

LEMMA A.5. *Under Assumption (A1), (A2), (A3) and (A4), we have*

$$\frac{1}{h^p}E\left|\left(\frac{A}{\pi(X)} - \frac{1-A}{1-\pi(X)}\right)^3 Y^3K^3\left(\frac{x-X}{h}\right)\right| = O(1),$$

where the big-O term is uniform in x and h.

Proof: By Assumption (A3), we have

$$\left|\frac{A}{\pi(X)} - \frac{1-A}{1-\pi(X)}\right| = \left|\frac{A\{1-\pi(X)\} - (1-A)\pi(X)}{\pi(X)\{1-\pi(X)\}}\right| \le \frac{1}{c_0(1-c_1)}.$$

Therefore,

$$\begin{aligned} &\frac{1}{h^p}\mathrm{E}\left|\left(\frac{A}{\pi(X)} - \frac{1-A}{1-\pi(X)}\right)^3 Y^3K^3\left(\frac{x-X}{h}\right)\right| \\ \le\ & \frac{1}{c_0^3(1-c_1)^3h^p}\mathrm{E}|Y^3|\left|K^3\left(\frac{x-X}{h}\right)\right|. \end{aligned} \tag{A.174}$$

Besides, it follows from Assumption (A4) and Hölder's inequality that

$$\sup_{X,A}\mathrm{E}(|Y^3||X,A) \le \sup_{X,A}\{\mathrm{E}(Y^4|X,A)\}^{3/4} \le c_*,$$

for some constant $c_* > 0$. This together with (A.174) yields

$$\frac{1}{h^p}\mathrm{E}\left|\left(\frac{A}{\pi(X)} - \frac{1-A}{1-\pi(X)}\right)^3 Y^3K^3\left(\frac{x-X}{h}\right)\right| \le \frac{c_*}{c_0^3(1-c_1)^3h^p}\mathrm{E}\left|K^3\left(\frac{x-X}{h}\right)\right|.$$

In addition, similar to (A.164), we can show

$$\frac{1}{h^p}\mathrm{E}\left|K^3\left(\frac{x-X}{h}\right)\right| = O(1),$$

under Condition (A1) and (A2). The proof is hence completed.

LEMMA A.6. *Let $N_{1,n}$ and $N_{2,n}$ be independent Poisson random variables with $N_{1,n}$ being Poisson$(n\alpha_n)$ and $N_{2,n}$ being Poisson$(n(1-\alpha_n))$ where $\alpha_n \in (0,1)$. Denote $N_n = N_{1n} + N_{2n}$ and set*

$$U_n = \frac{N_{1,n} - n\alpha_n}{\sqrt{n}} \quad \textit{and} \quad V_n = \frac{N_{2,n} - n(1-\alpha_n)}{\sqrt{n}}.$$

Let $\{S_n\}_{n=1}^{\infty}$ be a sequence of random variables such that:

(i) for each $n \geq 1$, the random vector (S_n, U_n) is independent of V_n,
(ii) for some $\sigma^2 < \infty$, $S_n \xrightarrow{d} \sigma Z$,
(iii) $\alpha_n \to \alpha \in (0,1)$.

Then, for all x,

$$Pr\{S_n \leq x | N_n = n\} \to Pr(\sigma Z \leq x).$$

Proof: Lemma A.6 is similar to Lemma 2.2 in Mason and Polonik (2009) and Lemma 2.4 in Giné et al. (2003). It can be proven using similar arguments in these references.

LEMMA A.7. *Define function $\psi_q(z, z_0) = \max\{(z+z_0)^q, 0\}$ for any $q \geq 1$. Suppose W_i is a set of i.i.d random variables such that $EW_i = 0$, $EW_i^2 = 1$ and $E|W_i|^r < \infty$ for $r = \max(q, 3)$. Then, we have*

$$\begin{aligned} & \left| E\psi_q\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n W_i, z_0\right) - E\psi_q(\mathbb{Z}_0, z_0)\right| \\ \leq \quad & c_1(|z_0|^{q-1}+1)\left(\frac{1}{\sqrt{n}}E|W_1|^3 + \frac{1}{n^{(r-2)/2}}E|W_1|^r\right) \\ + \quad & \frac{c_2}{n^{q/2}}(E|W_1|^3)^q + \frac{c_3}{\sqrt{n}}E|W_1|^3(E|\mathbb{Z}|^{q-1} + |z_0|^{q-1}), \end{aligned}$$

where c_1, c_2 are positive constants that depend only on q, and $\mathbb{Z}_0$ stands for a standard normal random variable.

Proof: For any $a, b \in \mathbb{R}$ and any $0 \leq q_0 < 1$, it follows from Minkowski's inequality that

$$|a+b|^{q_0} \leq |a|^{q_0} + |b|^{q_0}.$$

For $q_0 \geq 1$, we have

$$|a+b|^{q_0} \leq 2^{q_0-1}(|a|^{q_0} + |b|^{q_0}).$$

Therefore, there exists some constant C_{q_0} such that

$$|a+b|^{q_0} \le C_{q_0}(|a|^{q_0} + |b|^{q_0}), \tag{A.175}$$

for all $a, b \in \mathbb{R}$, $q_0 \ge 0$.

Therefore, we have

$$\begin{aligned} |\psi_q(z, z_0) - \psi_q(0, z_0)| \le |(z+z_0)^q - z_0^q| &= \left| \int_0^z q(z_0 + t)^{q-1} dt \right| \\ \le\ & q \int_0^{|z|} (|z_0| + |t|)^{q-1} dt \le q C_{q-1} \int_0^{|z|} (|z_0|^{q-1} + |t|^{q-1}) dt \\ =\ & q C_{q-1} |z_0|^{q-1} |z| + C_{q-1} |z|^q, \end{aligned} \tag{A.176}$$

where the second inequality on the second line is due to (A.175).

For any $z \in \mathbb{R}$, we have $|z|^q \le 1 + |z|^r \min(1, |z|)$ and $|z| \le 1 + |z|^r \min(1, |z|)$. Combining this with (A.176), we have

$$\frac{|\psi_q(z, z_0) - \psi_q(0, z_0)|}{1 + |z|^r \min(1, |z|)} \le q C_{q-1} (|z_0|^{q-1} + 1).$$

For any z^*, similar to (A.176), we can show

$$\begin{aligned} |\psi_q(z, z_0) - \psi_q(z^*, z_0)| &\le \left| \int_0^{z - z^*} q(z^* + z_0 + t)^{q-1} dt \right| \\ \le\ & q C_{q-1} |z - z^*| |z_0 + z^*|^{q-1} + C_{q-1} |z - z^*|^q \\ \le\ & q C_{q-1} C_{q-2} |z - z^*| (|z_0|^{q-1} + |z^*|^{q-1}) + C_{q-1} |z - z^*|^q. \end{aligned}$$

This means for any given z^* and any z such that $|z - z^*| \le \varepsilon$, we have

$$|\psi_q(z, z_0) - \psi_q(z^*, z_0)| \le \varepsilon q C_{q-1} C_{q-2} (|z_0|^{q-1} + |z^*|^{q-1}) + \varepsilon^q C_{q-1}.$$

The assertion now follows by Lemma A2 in Lee et al. (2013).

LEMMA A.8. *Let $\psi_q(z, z_0) = \max\{(z + z_0)^q, 0\}$ for any $q \ge 1$. Let $\mathbb{Z}_0$ be a standard normal random variable. For any $z_0, z_1 \in \mathbb{R}$, we have*

$$|E\psi_q(\mathbb{Z}_0, z_0) - E\psi_q(\mathbb{Z}_0, z_1)| \le c_4(|z_0 - z_1| + |z_0 - z_1||z_0|^{q-1} + |z_0 - z_1|^q),$$

where c_4 is a positive constant that depends on q only.

Proof: Similar to (A.176), we have

$$\begin{aligned} |\psi_q(z, z_0) - \psi_q(z, z_1)| \le\ & q C_{q-1} |z + z_0|^{q-1} |z_0 - z_1| + C_{q-1} |z_0 - z_1|^q \\ & q C_{q-1}^2 (|z|^{q-1} + |z_0|^{q-1}) |z_0 - z_1| + C_{q-1} |z_0 - z_1|^q. \end{aligned}$$

Therefore,

$$
\begin{aligned}
&|\mathrm{E}\psi_q(\mathbb{Z}_0, z_0) - \mathrm{E}\psi_q(\mathbb{Z}_0, z_1)| \le \mathrm{E}\,|\psi_q(\mathbb{Z}_0, z_0) - \psi_q(\mathbb{Z}_0, z_1)| \\
\le\ & qC_{q-1}^2 \mathrm{E}|\mathbb{Z}_0|^{q-1}|z_0 - z_1| + qC_{q-1}^2|z_0|^{q-1}|z_0 - z_1| + C_{q-1}|z_0 - z_1|^q.
\end{aligned}
$$

Set $c_4 = \max(qC_{q-1}^2 \mathrm{E}|\mathbb{Z}_0|^{q-1}, qC_{q-1}^2, C_{q-1})$, the assertion thus follows.

LEMMA A.9. *Let $\mathbb{Z}_0$ be a standard normal random variable. For any $q \ge 1$ and $z \in \mathbb{R}$, we have*

$$
\begin{aligned}
&\left| E\{\max(\mathbb{Z}_0 + z, 0)\}^q - E\{\max(\mathbb{Z}_0, 0)\}^q - qzE\mathbb{Z}_0^{q-1} I\{\mathbb{Z}_0 > \max(-z, 0)\} \right| \\
\le\ & c_5|z|^{q+1} + c_5\{|z|^2 + |z|^q\} I(q > 1),
\end{aligned}
$$

where c_5 is a positive constant that depends on q only.

Proof: It is immediate to see that the assertion holds when $z = 0$. Consider the case where $z > 0$. It suffices to show

$$
\begin{aligned}
&|\mathrm{E}[\{\max(\mathbb{Z}_0 + z, 0)\}^q - \{\max(\mathbb{Z}_0, 0)\}^q] I(-z \le \mathbb{Z}_0 \le 0)| \\
\le\ & c_5|z|^{q+1},
\end{aligned} \tag{A.177}
$$

and

$$
\begin{aligned}
&\left| \mathrm{E}[\{\max(\mathbb{Z}_0 + z, 0)\}^q - \{\max(\mathbb{Z}_0, 0)\}^q] I(\mathbb{Z}_0 > 0) - qz\mathrm{E}\mathbb{Z}_0^{q-1} I(\mathbb{Z}_0 > 0) \right| \\
\le\ & c_5\{|z|^2 + |z|^q\} I(q > 1).
\end{aligned} \tag{A.178}
$$

We first prove (A.177). The density function of a standard normal random variable is bounded from above by 1. Hence, we have

$$
\begin{aligned}
&|\mathrm{E}[\{\max(\mathbb{Z}_0 + z, 0)\}^q - \{\max(\mathbb{Z}_0, 0)\}^q] I(-z \le \mathbb{Z}_0 \le 0)| \\
\le\ & \int_{-z}^{0} (z_0 + z)^q dz = \frac{z^{q+1}}{q+1} \le z^{q+1}.
\end{aligned} \tag{A.179}
$$

Note that assertion (A.178) holds trivially for $q = 1$. It suffices to consider the case where $q > 1$. For any $z_0 > 0$, we have

$$
(z_0 + z)^q - z_0^q = qzz_0^{q-1} + q(q-1)\int_0^z \int_0^t (z_0 + t^*)^{q-2} dt^* dt.
$$

If $q > 2$, it follows from (A.175) that

$$\begin{aligned}|(z_0+z)^q - z_0^q - qzz_0^{q-1}| &\leq q(q-1)C_{q-2}\int_0^z \int_0^t (|z_0|^{q-2} + |t^*|^{q-2})dt^*dt \\ &\leq \frac{q(q-1)C_{q-2}|z_0|^{q-2}z^2}{2} + C_{q-2}|z|^q.\end{aligned}$$

If $q = 2$, it is immediate to see that

$$|(z_0+z)^q - z_0^q - qzz_0^{q-1}| = z^2.$$

If $1 < q < 2$, we have

(A.180)

$$|(z_0+z)^q - z_0^q - qzz_0^{q-1}| \leq q(q-1)\int_0^z \int_0^t z_0^{q-2}dt^*dt = \frac{q(q-1)z^2}{2}|z_0|^{q-2}.$$

To summarize, we've shown for any $q > 1$,

$$|(z_0+z)^q - z_0^q - qzz_0^{q-1}| \leq c^*(z^2|z_0|^{q-2} + |z|^q),$$

where c^* is some constant depends on q only.

Therefore, we have

$$\begin{aligned}& |\mathrm{E}\{(\mathbb{Z}_0+z)^q - \mathbb{Z}_0^q - qz\mathbb{Z}_0^{q-1}\}I(\mathbb{Z}_0>0)| \\ \leq\ & \mathrm{E}|(\mathbb{Z}_0+z)^q - \mathbb{Z}_0^q - qz\mathbb{Z}_0^{q-1}|I(\mathbb{Z}_0>0) \\ \leq\ & c^*\{z^2\mathrm{E}|\mathbb{Z}_0^{q-2}|I(\mathbb{Z}_0>0) + |z|^q\}.\end{aligned}$$

When $q > 1$, the expectation $\mathrm{E}|\mathbb{Z}_0^{q-2}|I(\mathbb{Z}_0>0)$ is finite. This proves (A.178).

Consider the case where $z < 0$. Similar to (A.177), we can show

$$\begin{aligned}& |\mathrm{E}[\{\max(\mathbb{Z}_0+z,0)\}^q - \{\max(\mathbb{Z}_0,0)\}^q]I(0 \leq \mathbb{Z}_0 \leq -z)| \\ \leq\ & c_5|z|^{q+1}.\end{aligned}$$

It suffices to show

$$\begin{aligned}\text{(A.181)}\quad & \left|\mathrm{E}[\{\max(\mathbb{Z}_0+z,0)\}^q - \{\max(\mathbb{Z}_0,0)\}^q - qz\mathrm{E}\mathbb{Z}_0^{q-1}]I(\mathbb{Z}_0>-z)\right| \\ \leq\ & c_5\{|z|^2 + |z|^q\}I(q>1).\end{aligned}$$

Assertion (A.181) can be similarly proven when $q = 1$ or $q \geq 2$. Consider the case where $1 < q < 2$. Similar to (A.180), we have for any $z_0 + z > 0$,

$$\begin{aligned}|(z_0+z)^q - z_0^q - qzz_0^{q-1}| &\leq q(q-1)\int_0^{-z}\int_0^t (z_0+z)^{q-2}dt^*dt \\ \text{(A.182)}\qquad &= \frac{q(q-1)z^2}{2}|z_0+z|^{q-2}.\end{aligned}$$

Moreover, note that the normal density function is bounded by 1. Hence, we have

$$
\begin{aligned}
& \mathrm{E}|\mathbb{Z}_0+z|^{q-2}I(\mathbb{Z}_0>-z) = \frac{1}{\sqrt{2\pi}}\int_{-z}^{\infty}(z_0+z)^{q-2}\exp\left(-\frac{z_0^2}{2}\right)dz_0 \\
= \quad & \frac{1}{\sqrt{2\pi}}\int_{-z}^{1-z}(z_0+z)^{q-2}\exp\left(-\frac{z_0^2}{2}\right)dz_0+\frac{1}{\sqrt{2\pi}}\int_{1-z}^{\infty}(z_0+z)^{q-2}\exp\left(-\frac{z_0^2}{2}\right)dz_0 \\
\leq \quad & \int_{-z}^{1-z}(z_0+z)^{q-2}dz_0+\frac{1}{\sqrt{2\pi}}\int_{1-z}^{\infty}\exp\left(-\frac{z_0^2}{2}\right)dz_0 \leq \int_0^1 z_0^{q-2}dz_0 \\
+ \quad & \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{z_0^2}{2}\right)dz_0 = \frac{1}{q-1}+1=\frac{q}{q-1}.
\end{aligned}
$$

Combining this together with (A.182) implies that

$$
\begin{aligned}
& \left|\mathrm{E}[\{\max(\mathbb{Z}_0+z,0)\}^q-\{\max(\mathbb{Z}_0,0)\}^q-qz\mathrm{E}\mathbb{Z}_0^{q-1}]I(\mathbb{Z}_0>-z)\right| \\
\leq \quad & \frac{q(q-1)z^2}{2}\frac{q}{q-1}=\frac{q^2z^2}{2}.
\end{aligned}
$$

This yields (A.181). The proof is hence completed.

REFERENCES

Chang, M., S. Lee, and Y.-J. Whang (2015). Nonparametric tests of conditional treatment effects with an application to single-sex schooling on academic achievements. *Econom. J. 18*(3), 307–346.

de la Peña, V. H., T. L. Lai, and Q.-M. Shao (2009). *Self-normalized processes.* Probability and its Applications (New York). Springer-Verlag, Berlin. Limit theory and statistical applications.

Fan, A., W. Lu, and R. Song (2016). Sequential advantage selection for optimal treatment regime. *Ann. Appl. Stat. 10*(1), 32–53.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc. 96*(456), 1348–1360.

Giné, E., D. M. Mason, and A. Y. Zaitsev (2003). The L_1-norm density estimator process. *Ann. Probab. 31*(2), 719–768.

Lee, S., K. Song, and Y.-J. Whang (2013). Testing functional inequalities. *J. Econometrics 172*(1), 14–32.

Mason, D. M. and W. Polonik (2009). Asymptotic normality of plug-in level set estimates. *Ann. Appl. Probab. 19*(3), 1108–1142.

Nolan, D. and D. Pollard (1987). U-processes: rates of convergence. *Ann. Statist. 15*(2), 780–799.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes.* Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.

Chengchun Shi
Department of Statistics,
North Carolina State University,
Raleigh NC, U.S.A.
E-mail: cshi4@ncsu.edu

Wenbin Lu
Department of Statistics,
North Carolina State University,
Raleigh NC, U.S.A.
E-mail: wlu4@ncsu.edu

Rui Song
Department of Statistics,
North Carolina State University,
Raleigh NC, U.S.A.
E-mail: rsong@ncsu.edu