Online Supplement to “Unconditional Quantile Regression with High-Dimensional Data”

Yuya Sasaki* Takuya Ura† Yichong Zhang‡

Abstract

This article collects the supplemental contents to the main paper. Section A collects the key results for the theorems in Section 3. Section B shows the double robustness of moment condition used in the paper. Section C considers the estimation and inference of UQPE with general machine learning estimators and cross-fitting. Section D collects the proofs of theoretical results in the main paper. Section E collects the proofs of theoretical results in Section C. Section F provides additional simulation results.

*Y. Sasaki: Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235-1819. Email: yuya.sasaki@vanderbilt.edu

†T. Ura: Department of Economics, University of California, Davis, One Shields Avenue, Davis, CA 95616. Email: takura@ucdavis.edu

‡Y. Zhang: School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903, Singapore. Email: yczhang@smu.edu.sg

A Theoretical Results for Section 3

Theorem A.1. *If Assumptions 1 and 2 hold, then for $j = 0, 1$ and any $e > 0$, there exists a class of functions denoted as $\mathcal{G}^{(j)}$ such that $\{\hat{m}_j(x, q) : q \in \mathcal{Q}^\delta\} \subset \mathcal{G}^{(j)}$ with probability greater than $1 - e$ and*

$$\sup_Q N(\mathcal{G}^{(j)}, \|\cdot\|_{Q,2}, \varepsilon \|G^{(j)}\|_{Q,2}) \leq C \left(\frac{p_b}{\varepsilon}\right)^{cs_b} \text{ for every } \varepsilon \in (0, 1], \tag{A.1}$$

where C, c are positive constants, $N(\cdot)$ is the covering number, $G^{(j)}$ is the envelope for $\mathcal{G}^{(j)}$, $\|\cdot\|_{Q,2}$ is the L^2 norm for a probability measure Q, and the supremum is taken over all finitely discrete probability measures. In addition, we have

$$\sup_{q \in \mathcal{Q}^\delta} \int |\hat{m}_j(x, q) - m_j(x, q)|^2 \, dF_X(x) = O_P\left(\frac{s_b \log(p_b)}{N}\right), \tag{A.2}$$

$$\sup_{q \in \mathcal{Q}^\delta, x \in \mathcal{X}} |\hat{m}_j(x, q) - m_j(x, q)| = O_P\left(\zeta_N s_b \sqrt{\frac{\log(p_b)}{N}}\right). \tag{A.3}$$

Theorem A.2. *If Assumptions 1 and 3 holds, then for any $e > 0$, there exists a class of functions denoted as $\mathcal{G}^\omega$ such that $\hat{\omega}(x) \in G^\omega$ with probability greater than $1 - e$ and*

$$\sup_Q N(\mathcal{G}^\omega, \|\cdot\|_{Q,2}, \varepsilon \|G^\omega\|_{Q,2}) \leq C \left(\frac{p_h}{\varepsilon}\right)^{cs_h} \text{ for every } \varepsilon \in (0, 1], \tag{A.4}$$

where G^ω is the envelope of $\mathcal{G}^{(\omega)}$. In addition, we have, for all $c > 0$,

$$\int (\hat{\omega}(x) - \omega(x))^2 \, dF_X(x) = o_P(N^c s_h \log(p_h)/N) \tag{A.5}$$

and

$$\sup_{x \in \mathcal{X}} |\hat{\omega}(x) - \omega(x)| = o_P(1). \tag{A.6}$$

Theorem A.3. *If Assumptions 1–4 hold, then*

$$\sup_{\tau \in \Upsilon} \left| \hat{\theta}(\tau) - \theta(\tau) - \frac{1}{N} \sum_{i=1}^N \mathrm{IF}_i^\theta(\tau) \right| = o_P(N^{-1/2}),$$

$$\sup_{\tau \in \Upsilon} \left| \hat{\theta}^*(\tau) - \hat{\theta}(\tau) - \frac{1}{N} \sum_{i=1}^N \eta_i \mathrm{IF}_i^\theta(\tau) \right| = o_P(N^{-1/2}).$$

Theorem A.4. *If Assumptions 1–3, 5, 6 hold, then*

$$\widehat{UQPE}(\tau) - UQPE(\tau) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{IF}_i(\tau) + \frac{\theta(\tau) f_Y^{(2)}(q_\tau)(\int u^2 K_1(u)du) h_1^2}{2 f_Y^2(q_\tau)} + R(\tau)$$
$$\widehat{UQPE}^*(\tau) - \widehat{UQPE}(\tau) = \frac{1}{N}\sum_{i=1}^{N} \eta_i \cdot \mathrm{IF}_i(\tau) + R^*(\tau),$$

where the residuals satisfy $\sup_{\tau\in\Upsilon}\max\{|R(\tau)|, |R^*(\tau)|\} = o_P((\log(N)Nh_1)^{-1/2})$.

B Double Robustness

The double robustness of (3) follows from Chernozhukov, Escanciano, Ichimura, Newey, and Robins (2021a, Theorem 3). In this section, for the sake of completeness, we demonstrate that (3) is doubly robust.

Lemma B.1 (Double Robustness). *Suppose Assumption 1 holds. If*

(i) $\int |\tilde{m}_1(x, q_\tau)| dF_X(x)$, $\int |\tilde{\omega}(x) 1\{y \le q_\tau\}| dF_{Y,X}(y, x)$, $\int |\tilde{\omega}(x) m_0(x, q_\tau)| dF_X(x)$, *and* $\int |\omega(x)\tilde{m}_0(x, q_\tau)| dF_X(x)$ *are finite;*

(ii) for every x_{-1} *in the support of* X_{-1}*, the mappings* $x_1 \mapsto (m_0(x, q_\tau) - \tilde{m}_0(x, q_\tau))$ *and* $x_1 \mapsto f_{X_1|X_{-1}=x_{-1}}(x_1)$ *are continuously differentiable with*

$$(m_0(x, q_\tau) - \tilde{m}_0(x, q_\tau)) f_{X_1|X_{-1}=x_{-1}}(x_1) \to 0$$

as $x_1 \to \pm\infty$*; and*

(iii) $\int \tilde{m}_1(x, q_\tau) dF_X(x) = \int \frac{\partial}{\partial x_1}\tilde{m}_0(x, q_\tau) dF_X(x)$*;*

then (4) and (5) hold.

In this lemma, conditions (i) and (ii) are regularity conditions for the nuisance parameter values. Condition (iii) is satisfied if $\tilde{m}_1(x, q_\tau) = \frac{\partial}{\partial x_1}\tilde{m}_0(x, q_\tau)$. It is reasonable since $\tilde{m}_0(x, q_\tau)$ is a value for $m_0(x, q_\tau)$ and $\tilde{m}_1(x, q_\tau)$ is a value for $m_1(x, q_\tau) = \frac{\partial}{\partial x_1} m_0(x, q_\tau)$.

Proof. Note that (4) follows from

$$\int (\tilde{m}_1(x, q_\tau) - \omega(x)(1\{y \le q_\tau\} - \tilde{m}_0(x, q_\tau)))\, dF_{Y,X}(y, x)$$
$$= \int \tilde{m}_1(x, q_\tau) dF_X(x) - \iint (m_0(x, q_\tau) - \tilde{m}_0(x, q_\tau)) \left(\frac{\partial}{\partial x_1} f_{X_1|X_{-1}=x_{-1}}(x_1)\right) dx_1 dF_{X_{-1}}(x_{-1})$$

$$
\begin{aligned}
&= \int \tilde{m}_1(x, q_\tau) dF_X(x) + \iint \left(m_1(x, q_\tau) - \left(\frac{\partial}{\partial x_1} \tilde{m}_0(x, q_\tau) \right) \right) \left(f_{X_1 | X_{-1} = x_{-1}}(x_1) \right) dx_1 dF_{X_{-1}}(x_{-1}) \\
&= \int m_1(x, q_\tau) dF_X(x) \\
&= \theta(\tau),
\end{aligned}
$$

where the first equality follows from Fubini's theorem, and the second equality follows from integration by parts. Next, (5) follows from

$$
\begin{aligned}
&\int \left(m_1(x, q_\tau) - \tilde{\omega}(x)(1\{y \le q_\tau\} - m_0(x, q_\tau)) \right) dF_{Y,X}(y, x) \\
&= \int m_1(x, q_\tau) dF_X(x) - \iint \tilde{\omega}(x)(m_0(x, q_\tau) - m_0(x, q_\tau)) f_{X_1 | X_{-1} = x_{-1}} dx_1 dF_{X_{-1}}(x_{-1}) \\
&= \int m_1(x, q_\tau) dF_X(x) \\
&= \theta(\tau).
\end{aligned}
$$

This completes a proof of the lemma. □

C Estimation and Inference of UQPE with General Preliminary Machine Learning Estimators

C.1 Estimation and Inference Procedures with Cross-Fitting

Based on the moment condition (3), we propose to estimate $\theta(\tau)$ by a cross-fitting approach (Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins, 2018, Definition 3.2). We split the sample of size N into a random partition $\{I_1, \cdots, I_L\}$ of approximately equal size. For simplicity, let $|I_l| = n$ for every l so that $N = nL$. In this section, we assume that, for every index $l \in \{1, ..., L\}$ of fold, we can construct an estimator $(\hat{\omega}_l(x), \hat{m}_{0,l}(x, q), \hat{m}_{1,l}(x, q))$ by using all the observations except those in I_l. Letting $\hat{q}_\tau$ be the full sample τ-th empirical quantile of Y, we estimate $\theta(\tau)$ by

$$
\hat{\theta}_{cf}(\tau) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{n} \sum_{i \in I_l} \left(\hat{m}_{1,l}(X_i, \hat{q}_\tau) - \hat{\omega}_l(X_i)(1\{Y_i \le \hat{q}_\tau\} - \hat{m}_{0,l}(X_i, \hat{q}_\tau)) \right). \tag{C.1}
$$

With this estimator for $\theta(\tau)$, our proposed estimator for $UQPE(\tau)$ is

$$\widehat{UQPE}_{cf}(\tau) = -\frac{\hat{\theta}_{cf}(\tau)}{\hat{f}_Y(\hat{q}_\tau)},$$

where $\hat{f}_Y(y)$ is defined in Section 2.2.

For an inference about $UQPE(\tau)$, we propose the multiplier bootstrap without recalculating the preliminary estimators $(\hat{\omega}_l(x), \hat{m}_{0,l}(x,q), \hat{m}_{1,l}(x,q))$ in each bootstrap iteration. Using independent standard normal random variables $\{\eta_i\}_{i=1}^N$ that are independent of the data, we compute the bootstrap estimator $\widehat{UQPE}^*_{cf}(\tau)$ in the following steps. We construct the bootstrap estimator $(\hat{q}^*_\tau, \hat{f}^*_Y)$ for (q_τ, f_Y) in the same way as in Section 2.3. The bootstrap estimator for $\theta(\tau)$ is

$$\hat{\theta}^*_{cf}(\tau) = \frac{1}{L}\sum_{l=1}^{L} \frac{1}{\sum_{i\in I_l}(\eta_i+1)} \sum_{i\in I_l} (\eta_i+1)\left(\hat{m}_{1,l}(X_i, \hat{q}^*_\tau) - \hat{\omega}_l(X_i)(1\{Y_i \leq \hat{q}^*_\tau\} - \hat{m}_{0,l}(X_i, \hat{q}^*_\tau))\right).$$

With these components, the bootstrap estimator $\widehat{UQPE}^*_{cf}(\tau)$ is given by

$$\widehat{UQPE}^*_{cf}(\tau) = -\frac{\hat{\theta}^*_{cf}(\tau)}{\hat{f}^*_Y(\hat{q}^*_\tau)}.$$

We can use the above multiplier bootstrap method to conduct various types of inference. For example, we can construct a confidence band $CB^\theta_{\Upsilon,cf}$ for $\theta(\tau)$ over Υ by computing $\hat{\theta}_{cf}(\tau) \pm \hat{\sigma}^\theta(\tau) c^\theta_{\Upsilon,cf}(1-\alpha)$ for $\tau \in \Upsilon$, where $\hat{\sigma}^\theta(\tau)$ is an estimator of the standard error of $\hat{\theta}_{cf}(\tau)$ and $c^\theta_{\Upsilon,cf}(1-\alpha)$ is the $(1-\alpha)$ quantile of $\sup_{\tau\in\Upsilon}\left|(\hat{\theta}^*_{cf}(\tau) - \hat{\theta}_{cf}(\tau))/\hat{\sigma}^\theta(\tau)\right|$ conditional on the data. Also, we can construct a confidence band $CB_{\Upsilon,cf}$ for $UQPE$ over Υ by computing $\widehat{UQPE}_{cf}(\tau) \pm \hat{\sigma}(\tau) c_{\Upsilon,cf}(1-\alpha)$ for $\tau \in \Upsilon$, where $\hat{\sigma}(\tau)$ is an estimator of the standard error of $\widehat{UQPE}_{cf}(\tau)$ and $c_{\Upsilon,cf}(1-\alpha)$ is the $(1-\alpha)$ quantile of $\sup_{\tau\in\Upsilon}\left|(\widehat{UQPE}^*_{cf}(\tau) - \widehat{UQPE}_{cf}(\tau))/\hat{\sigma}(\tau)\right|$ conditional on the data.

C.2 Asymptotic Theory

In this section, we investigate the asymptotic properties of the estimators $(\widehat{UQPE}_{cf}(\tau), \hat{\theta}_{cf}(\tau))$ and the bootstrap estimators $(\widehat{UQPE}^*_{cf}(\tau), \hat{\theta}^*_{cf}(\tau))$ introduced in the previous section.

Assumption C.1. *For every index $l \in \{1, ..., L\}$ of folds, there exist sequences ν_N, A_N, π_N*

such that the following conditions hold with probability approaching one:

$$\sup_{Q} N(\{\hat{m}_{j,l}(x,q) : q \in \mathcal{Q}^{\delta}\}, \|\cdot\|_{Q,2}, \varepsilon\|G_l^{(j)}\|_{Q,2}) \lesssim \left(\frac{A_N}{\varepsilon}\right)^{\nu_N} \quad \text{for every } \varepsilon \in (0,1], \tag{C.2}$$

$$\sup_{q\in\mathcal{Q}^{\delta}} \int |\hat{m}_{1,l}(x,q) - m_1(x,q)|^2 \, dF_X(x) = O_P(\pi_N^2), \tag{C.3}$$

$$\int |\hat{\omega}_l(x) - \omega(x)|^2 \, dF_X(x) = O_P(\pi_N^2), \tag{C.4}$$

$$\sup_{q\in\mathcal{Q}^{\delta}} \int |\hat{\omega}_l(x)\hat{m}_{0,l}(x,q) - \omega(x)m_0(x,q)|^2 \, dF_X(x) = O_P(\pi_N^2), \tag{C.5}$$

$$\int \left(\sup_{q\in\mathcal{Q}^{\delta}} |\hat{m}_{1,l}(x,q)|\right)^{2+d} dF_X(x) = O_P(1), \tag{C.6}$$

$$\int \left(\sup_{q\in\mathcal{Q}^{\delta}} |\hat{\omega}_l(x)(1+|\hat{m}_{0,l}(x,q)|)|\right)^{2+d} dF_X(x) = O_P(1), \tag{C.7}$$

$$\sup_{q\in\mathcal{Q}^{\delta}} \left|\int \left(\hat{m}_{1,l}(x,q) - \frac{\partial}{\partial x_1}\hat{m}_{0,l}(x,q)\right) dF_X(x)\right| = O_P(\tilde{\pi}_N^2), \tag{C.8}$$

$$\sup_{q\in\mathcal{Q}^{\delta}} \left|\int (\hat{\omega}_l(x) - \omega(x))(\hat{m}_{0,l}(x,q) - m_0(x,q)) dF_X(x)\right| = O_P(\tilde{\pi}_N^2), \tag{C.9}$$

where, in (A.1)*,* $N(\cdot)$ *is the covering number,* $G_l^{(j)}$ *is the envelope for* $\{\hat{m}_j(x,q) : q \in \mathcal{Q}^{\delta}\}$*, and the supremum in* (C.2) *is taken over all finitely discrete probability measures.*

Several comments are in order. First, this assumption consists of a list of high-level conditions that should be satisfied by the preliminary estimator $(\hat{\omega}_l(x), \hat{m}_{0,l}(x,q), \hat{m}_{1,l}(x,q))$. While we state these high-level conditions here for the sake of accommodating a general class of preliminary estimators, the preliminary estimators considered in the paper satisfy these requirements. Second, (A.1) is the entropy condition for the classes of functions $\{\hat{m}_{j,l}(x,q) : q \in \mathcal{Q}^{\delta}\}$. We require this condition because (1) we want to derive the linear expansion for $\hat{\theta}(\tau)$ that is uniform in τ and (2) $\hat{m}_{j,l}(x,\hat{q}_\tau)$ has the estimated $\hat{q}_\tau$ inside for $j = 0, 1$. We can directly verify (A.1) for general machine learning estimators via a kernel convolution technique in Section C.3. Third, it is worth mentioning that term (C.8) is zero if we construct $\hat{m}_1(x,q)$ by $\hat{m}_1(x,q) = \frac{\partial}{\partial x_1}\hat{m}_0(x,q)$.

Theorem C.1. *If Assumptions 1 and C.1 hold,* $\pi_N^2\nu_N\log(A_N/\pi_N) = o(1)$*,* $\nu_N^2\log^2(A_N/\pi_N) = o(N^{\frac{d}{2+d}})$*, and* $\tilde{\pi}_N = o(N^{-1/4})$*, then* $\sup_{\tau\in\Upsilon}\left|\hat{\theta}_{cf}(\tau) - \theta(\tau) - \frac{1}{N}\sum_{i=1}^N \mathrm{IF}_i^{\theta}(\tau)\right| = o_P(N^{-1/2})$*, and* $\sup_{\tau\in\Upsilon}\left|\hat{\theta}_{cf}^*(\tau) - \hat{\theta}_{cf}(\tau) - \frac{1}{N}\sum_{i=1}^N \eta_i\mathrm{IF}_i^{\theta}(\tau)\right| = o_P(N^{-1/2})$.

Corollary C.1. *Suppose assumptions in Theorem C.1 hold and* $\sup_{\tau\in\Upsilon}|\sqrt{N}\hat{\sigma}^{\theta}(\tau) - Var(\mathrm{IF}_i^{\theta}(\tau))| = o_P(1)$*. Then* $\mathbb{P}\left(\{\theta(\tau) : \tau \in \Upsilon\} \in CB_{\Upsilon,cf}^{\theta}\right) \to 1 - \alpha$.

Theorem C.2. *If Assumptions 1, 5, and C.1 hold,* $\pi_N^2 \nu_N \log(N)\log(A_N/\pi_N)h_1 = o(1)$, $\nu_N^2 \log(N)\log^2(A_N/\pi_N)h_1 = o(N^{\frac{d}{2+d}})$, *and* $\tilde{\pi}_N = o((\log(N)Nh_1)^{-1/4})$, *then* $\widehat{UQPE}_{cf}(\tau) - UQPE(\tau) = \frac{1}{N}\sum_{i=1}^N \mathrm{IF}_i(\tau) + \frac{\theta(\tau) f_Y^{(2)}(q_\tau)(\int u^2 K_1(u)du)h_1^2}{2f_Y^2(q_\tau)} + R(\tau)$ *and* $\widehat{UQPE}^*_{cf}(\tau) - \widehat{UQPE}_{cf}(\tau) = \frac{1}{N}\sum_{i=1}^N \eta_i \cdot \mathrm{IF}_i(\tau) + R^*(\tau)$, *where* $\sup_{\tau\in\Upsilon}\max\{|R(\tau)|, |R^*(\tau)|\} = o_P((\log(N)Nh_1)^{-1/2})$.

The following corollary summarizes the validity for the bootstrap inference.

Corollary C.2. *Suppose assumptions in Theorem C.2 hold and* $\sqrt{Nh_1} = o(h_1^{-2})$. *If* $h_1 Var(\mathrm{IF}_i(\tau))$ *is bounded away from zero and* $\sup_{\tau\in\Upsilon}\left|\sqrt{Nh_1}\hat{\sigma}(\tau) - \sqrt{h_1 Var(\mathrm{IF}_i(\tau))}\right| = o_P(\log^{-1/2}(N))$, *then* $\mathbb{P}(\{UQPE(\tau) : \tau \in \Upsilon\} \in CB_{\Upsilon,cf}) \to 1-\alpha$.

C.3 Kernel Smoothing General Machine Learning Estimators

In this section, we propose a kernel convolution method to smooth general machine learning estimators $\hat{m}_{j,l}(x,q)$ over q. This convolution benefits the theoretical arguments for the uniform consistency over q because the resulting convolution is Lipschitz continuous, as shown in the proof of Theorem C.3. Chernozhukov, Fernández-Val, and Kowalski (2015) introduce the kernel convolution as a theoretical device in their proof. The key advantage is they do not need to implement kernel convolution in practice, and thus, avoid the choice of the tuning parameter. On the other hand, we apply the kernel convolution technique in a difference context and require implementing it on the original machine learning estimator. For a generic machine learning estimator $\hat{m}_0(x,q)$, the entropy of the class of functions $\{\hat{m}_{j,l}(x,q) : q \in \mathcal{Q}^\delta\}$ for $j = 0,1$ and $l \in \{1,\ldots,L\}$ is usually unknown. This kernel convolution method provides one way to introduce smoothness to $\hat{m}_{j,l}(x,q)$ over q and thus reduces the entropy of $\{\hat{m}_{j,l}(x,q) : q \in \mathcal{Q}^\delta\}$.

Assumption C.2.

1. *$m_0(x,q)$ and $m_1(x,q)$ are $2k$-th order differentiable with respect to q, and all the derivatives are bounded uniformly over x.*

2. *$K_2(\cdot)$ is a symmetric function with bounded support, $\int K_2(u)du = 1$, $\int u^j K_2(u)du = 0$ for $j = 1,\cdots,2k-1$, $sup_u|K_2(u)| < \infty$ and $\int u^{2k}|K_2(u)|du < \infty$. $h_2 = c_2 N^{\frac{-1}{2(2k+1)}}$ for some positive constant c_2.*

We use the higher-order kernel to fully exploit the smoothness of $m_0(x,q)$ and reduce the bias caused by the kernel convolution method. We further assume that the errors of the initial machine learning estimators $\{\breve{m}_{j,l}(x,q)\}_{j=0,1,l\in\{1,\ldots,L\}}$ and $\{\hat{\omega}_l(x)\}_{l\in\{1,\ldots,L\}}$ satisfy the following conditions.

Assumption C.3. *For every subsample index $l \in \{1, ..., L\}$, there exists a vanishing sequence ρ_N such that*

$$\sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta} |\breve{m}_{j,l}(X,q) - m_j(X,q)| = O_P(h_2\rho_N),\ \ j = 0, 1, \tag{C.10}$$

$$\sup_{q\in\mathcal{Q}^\delta} \int |\breve{m}_0(x,q) - m_1(x,q)|^2\, dF_X(x) = O_P(h_2^2\rho_N^2), \tag{C.11}$$

$$\int |\hat{\omega}(x) - \omega(x)|^2\, dF_X(x) = O_P(h_2^2\rho_N^2), \tag{C.12}$$

$$\int \left(\sup_{q\in\mathcal{Q}^\delta} |\breve{m}_0(x,q)|\right)^{2+d} dF_X(x) = O_P(1), \tag{C.13}$$

$$\int \left(\sup_{q\in\mathcal{Q}^\delta} |\hat{\omega}(x)(1 + |\breve{m}_0(x,q)|)|\right)^{2+d} dF_X(x) = O_P(1), \tag{C.14}$$

$$\sup_{q\in\mathcal{Q}^\delta} \left|\int \left(\breve{m}_0(x,q) - \frac{\partial}{\partial x_1}\breve{m}_0(x,q)\right) dF_X(x)\right| = O_P(\tilde{\pi}_{N,1}^2), \tag{C.15}$$

$$\sup_{q\in\mathcal{Q}^\delta} \left|\int (\hat{\omega}(x) - \omega(x))(\breve{m}_0(x,q) - m_0(x,q))dF_X(x)\right| = O_P(\tilde{\pi}_{N,1}^2). \tag{C.16}$$

Deriving these error bounds for various machine learning methods is beyond the scope of our paper. Partial results are available in the literature. For example, the L_2 bounds for the random forest method and deep neural networks have already been established in Wager and Athey (2018), Schmidt-Hieber (2020), and Farrell, Liang, and Misra (2021)), respectively.

Our final first-stage estimator of ($m_0(x,q)$, $m_1(x,q)$, $\omega(x)$) is ($\hat{m}_0(x,q)$, $\hat{m}_0(x,q)$, $\hat{\omega}(x)$), where $\hat{m}_0(x,q) = \int \frac{\breve{m}_0(x,t)}{h_2} K_2\left(\frac{q-t}{h_2}\right) dt$ for $j = 0, 1$. The next theorem shows that the high-level conditions in Assumption C.1 hold for $(\hat{m}_0(x,q), \hat{m}_0(x,q), \hat{\omega}(x))$.

Theorem C.3. *Suppose Assumptions 1, C.2 and C.3 hold, then $(\hat{m}_0(x,q), \hat{m}_0(x,q), \hat{\omega}(x))$ satisfy Assumption C.1 with $\nu_N = 1$, $A_N = 1$, and $\pi_N = h_2\rho_N + h_2^{2k}$, and $\tilde{\pi}_N^2 = \tilde{\pi}_{N,1}^2 + o(N^{-1/2})$.*

D Proof of the Results in the Main Text

D.1 Proof of (3)

Lemma D.1. *Equation* (3) *holds under Assumption 1.*

Proof. This statement follows from

$$\begin{aligned}
&\mathbb{E}[m_1(X,q_\tau)-\theta(\tau)-\omega(X)(1\{Y\le q_\tau\}-m_0(X,q_\tau))]\\
&=-\int \omega(x)(1\{y\le q_\tau\}-m_0(x,q_\tau))dF_{Y,X}(y,x)\\
&=-\int \omega(x)(\int 1\{y\le q_\tau\}dF_{Y|X=x}(y)-m_0(x,q_\tau))dF_X(x)\\
&=0,
\end{aligned}$$

where the first equality follows from the definition of $\theta(\tau)$, the second equality comes from the law of iterated expectations, and the last equality follows from the definition of $m_0(x,q)$. □

D.2 Proof of Theorem A.1

In the proof of Theorem A.1, we use the notations

$$\mathcal{G}^{(0)}=\{\Lambda(b(X)^\top\beta):\beta\in\mathbb{R}^p,\|\beta\|_0\le Ms_b\},$$

and

$$\mathcal{G}^{(1)}=\left\{\Lambda(b(X)^\top\beta)(1-\Lambda(b(X)^\top\beta))\frac{\partial}{\partial x_1}b(X)^\top\beta:\beta\in\mathbb{R}^p,\|\beta\|_0\le Ms_b,\sup_{x\in\mathcal{X}}\left|\frac{\partial}{\partial x_1}b(x)^\top\beta\right|\le M\right\},$$

where M is a sufficiently large constant.

Lemma D.2. *Under the assumptions in Theorem A.1,*

$$(i)\quad \sup_{q\in\mathcal{Q}^\delta}\|\hat\beta_q-\beta_q\|_1=O_P\left(\sqrt{\frac{s_b^2\log(p_b)}{N}}\right),$$

$$(ii)\quad \sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta}|\hat m_0(x,q)-m_0(x,q)|=O_P\left(\sqrt{\frac{\zeta_N^2 s_b^2\log(p_b)}{N}}\right),$$

$$(iii)\quad \sup_{q\in\mathcal{Q}^\delta}\left(\int(\hat m_0(x,q)-m_0(x,q))^2dF_X(x)\right)^{1/2}=O_P\left(\sqrt{\frac{s_b\log(p_b)}{N}}\right),$$

$$(vi)\quad \sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta}|\hat m_1(x,q)-m_1(x,q)|=O_P\left(\sqrt{\frac{\zeta_N^2 s_b^2\log(p_b)}{N}}\right),$$

$$(v)\quad \sup_{q\in\mathcal{Q}^\delta}\left(\int(\hat m_1(x,q)-m_1(x,q))^2dF_X(x)\right)^{1/2}=O_P\left(\sqrt{\frac{s_b\log(p_b)}{N}}\right),$$

where, in each of the five the statements, the norm on the left hand side is with respect to X *and the stochastic convergence O_P in the right hand side is with respect to the randomness of the estimators.*

Proof. The first three results have been established by Belloni, Chernozhukov, Fernández-Val, and Hansen (2017). For the fourth result, we have

$$\begin{aligned}
&|\hat{m}_1(X,q) - m_1(X,q)| \\
\leq & \left| \Lambda(b(X)^\top \hat{\beta}_q)(1-\Lambda(b(X)^\top \hat{\beta}_q)) \frac{\partial}{\partial x_1} b(X)^\top \hat{\beta}_q - \Lambda(b(X)^\top \beta_q)(1-\Lambda(b(X)^\top \beta_q)) \frac{\partial}{\partial x_1} b(X)^\top \beta_q \right| \\
& + \left| \frac{\partial}{\partial x_1}(m_0(X,q) - \Lambda(b(X)^\top \beta_q)) \right| \\
\leq & \left| \frac{\partial}{\partial x_1} b(X)^\top (\hat{\beta}_q - \beta_q) \right| + \left| \Lambda(b(X)^\top \hat{\beta}_q) - \Lambda(b(X)^\top \beta_q) \right| + \left| \frac{\partial}{\partial x_1}(m_0(X,q) - \Lambda(b(X)^\top \beta_q)) \right| \\
\leq & \sup_{x\in\mathcal{X}} \left| \frac{\partial}{\partial x_1} b(x) \right| \|\hat{\beta}_q - \beta_q\|_1 + |\hat{m}_0(X,q) - m_0(X,q)| + \left| \frac{\partial}{\partial x_1}(m_0(X,q) - \Lambda(b(X)^\top \beta_q)) \right| \\
& + |(m_0(X,q) - \Lambda(b(X)^\top \beta_q))|,
\end{aligned}$$

where the first inequality is due to the triangle inequality and Assumption 2, and the second inequality is due to the facts that $\Lambda(\cdot)(1-\Lambda(\cdot))$ is bounded, $f(u) = u(1-u)$ is Lipschitz-1 continuous in u, and $\sup_{x\in\mathcal{X}, q\in\mathcal{Q}^\delta} |\frac{\partial}{\partial x_1} b(x)^\top \beta_q| < \bar{c}$. Taking $\sup_{q\in\mathcal{Q}^\delta, x\in\mathcal{X}}$ on both sides, we have $\sup_{x\in\mathcal{X}, q\in\mathcal{Q}^\delta} |\hat{m}_1(x,q) - m_1(x,q)| = O_P\left(\sqrt{\zeta_N^2 s_b^2 \log(p_b)/N}\right)$. Similarly, by Assumption 2.3,

$$\sup_{q\in\mathcal{Q}^\delta} \left(\int (\hat{m}_1(x,q) - m_1(x,q))^2 \, dF_X(x) \right)^{1/2} = O_P\left(\sqrt{\frac{s_b \log(p_h \vee N)}{N}} \right).$$

This complete a proof of the lemma. □

We note that Belloni et al. (2017) have shown $\sup_{q\in\mathcal{Q}^\delta} \|\hat{\beta}_q\|_0 = O_P(s_b)$. In addition, by Lemma D.2, we have $\sup_{x\in\mathcal{X}, q\in\mathcal{Q}^\delta} \left| \frac{\partial}{\partial x_1} b(x)^\top (\hat{\beta}_q - \beta_q) \right| \leq \zeta_N \sup_{q\in\mathcal{Q}^\delta} \|\hat{\beta}_q - \beta_q\|_1 = o_P(1)$. This implies, with probability approaching one, $\sup_{x\in\mathcal{X}, q\in\mathcal{Q}^\delta} \left| \frac{\partial}{\partial x_1} b(X)^\top \hat{\beta}_q \right| = O(1)$. Then (A.1) directly follows from the argument in the proof of Belloni et al. (2017, Theorem 5.1). Lemma D.2 verifies (A.2) and (A.3) in Theorem A.1. This concludes the proof.

D.3 Proof of Theorem A.2

Define $\varepsilon_N = \sqrt{\log(p_h)/N}$. We let ℓ_N be a sequence that diverges to ∞ but $\ell_N = o(N^c)$ for any constant $c > 0$. Define $\mathcal{J}_0 = \text{Supp}(\bar{\rho})$ and $\hat{\mathcal{J}} = \text{Supp}(\hat{\rho})$. We denote $\rho^* = \arg\min_\rho \int (\omega(x) - h(x)^\top \rho)^2 dF_X(x) + 2\varepsilon_N \sum_{j\in\mathcal{J}_0^c} |\rho_j|$. For a generic $p \times 1$ vector ρ, let $\rho_{\mathcal{J}}$ be the $p \times 1$ vector such that its jth element is ρ_j if $j \in \mathcal{J}$ and 0 otherwise.

First, we are going to show (A.6). It is sufficient to show

$$\|\hat{\rho} - \overline{\rho}\|_1 = O_P(\ell_N \varepsilon_N^{(2\xi-1)/(1+2\xi)}), \tag{D.1}$$

because, given (D.1), we have

$$\sup_{x\in\mathcal{X}} |\hat{\omega}(x) - \omega(x)| \le \sup_{x\in\mathcal{X}} |h(x)| \|\hat{\rho} - \overline{\rho}\|_1 + \sup_{x\in\mathcal{X}} |h(x)^\top \overline{\rho} - \omega(x)| = o_P(1).$$

The proof of Chernozhukov, Newey, and Singh (2021b, Lemma A6) shows

$$\|\hat{\rho} - \rho^*\|_1 \le C\sqrt{s_h}\|\hat{\rho} - \rho^*\|_2 = O_P(\ell_N \varepsilon_N^{(2\xi-1)/(1+2\xi)}) \tag{D.2}$$

under Assumption 3.3. Therefore, in order to prove (D.1), it suffices to show

$$\|\overline{\rho} - \rho^*\|_1 = O_P(\ell_N \varepsilon_N^{(2\xi-1)/(1+2\xi)}). \tag{D.3}$$

By Chernozhukov et al. (2021b, Equation (B.2)), with $\varepsilon_N = \sqrt{\log(p_h)/N}$, we have

$$(\underline{\rho} - \rho^*)^\top G(\underline{\rho} - \rho^*) + 2\varepsilon_N \sum_{j\in\mathcal{J}_0^c} |\rho_j^*| \le (\underline{\rho} - \overline{\rho})^\top G(\underline{\rho} - \overline{\rho}) \le C_1 \varepsilon_N^{4\xi/(2\xi+1)},$$

where $\underline{\rho}$ is the coefficient of a linear projection of $\omega(X)$ on $h(X)$ such that $\mathbb{E}h(X)(\omega(X) - h(X)^\top \underline{\rho}) = 0$. Given $\overline{\rho}_j = 0$ for $j \in \mathcal{J}_0^c$ and the definition of $\underline{\rho}$, we have

$$\|(\rho^* - \overline{\rho})_{\mathcal{J}_0^c}\|_1 = \sum_{j\in\mathcal{J}_0^c} |\rho_j^*| \le C_1 \varepsilon_N^{(2\xi-1)/(2\xi+1)} \text{ and}$$

$$\int (\omega(x) - h(x)^\top \rho^*)^2 dF_X(x) = (\underline{\rho} - \rho^*)^\top G(\underline{\rho} - \rho^*) \le C_1 \varepsilon_N^{4\xi/(2\xi+1)}.$$

By Bickel, Ritov, and Tsybakov (2009, Lemma 4.1), Assumption 3.4 implies there exist constants κ and c such that

$$\inf_{\rho\neq 0, \|\rho_{\mathcal{J}_0^c}\|_1 \le \kappa \|\rho_{\mathcal{J}_0}\|_1} \frac{\rho^\top G \rho}{\|\rho_{\mathcal{J}_0}\|_2^2} \ge c > 0. \tag{D.4}$$

It implies $\|\rho^* - \overline{\rho}\|_1 \le C\varepsilon_N^{(2\xi-1)/(2\xi+1)}$.[1] Therefore, we have (D.3).

[1] If $\kappa\|(\rho^* - \overline{\rho})_{\mathcal{J}_0}\|_1 \le C_1\varepsilon_N^{(2\xi-1)/(2\xi+1)}$, then $\|\rho^* - \overline{\rho}\|_1 \le \|(\rho^* - \overline{\rho})_{\mathcal{J}_0}\|_1 + \|(\rho^* - \overline{\rho})_{\mathcal{J}_0^c}\|_1 \le C\varepsilon_N^{(2\xi-1)/(2\xi+1)}$. On the other hand, If $\kappa\|(\rho^* - \overline{\rho})_{\mathcal{J}_0}\|_1 > C_1\varepsilon_N^{(2\xi-1)/(2\xi+1)} \ge \|(\rho^* - \overline{\rho})_{\mathcal{J}_0^c}\|_1$, then we have $\|(\rho^* - \overline{\rho})_{\mathcal{J}_0}\|_1^2 \le |\mathcal{J}_0|\|(\rho^* - \overline{\rho})_{\mathcal{J}_0}\|_2^2 \le \frac{1}{c}s_h(\rho^* - \overline{\rho})^\top G(\rho^* - \overline{\rho}) = \frac{1}{c}s_h \int (h(x)(\rho^* - \overline{\rho}))^2 dF_X(x) \le \frac{1}{c}2s_h \left(\int (h(x)\rho^* - \omega(x))^2 dF_X(x) + C(s_h)^{-2\xi}\right) \le C\varepsilon_N^{(2\xi-1)/(2\xi+1)}$, where the second inequality is by (D.4),

Second, we want to show

$$\|\hat{\rho}\|_0 = O_P(s_h). \tag{D.5}$$

Let e be an arbitrary positive number. There exists a constant $\kappa > 0$ such that $\mathbb{P}(\lambda_{\max}(\hat{G}) > \kappa) \leq e/2$, where $\lambda_{\max}(\hat{G})$ is the largest eigenvalue for $\hat{G}$. By the first order condition, we have

$$\begin{aligned}
\lambda_R \|\hat{\rho}\|_0^{1/2} &= \left\| \left\{ -\hat{M} + \hat{G}\hat{\rho} \right\}_{\hat{\mathcal{J}}} \right\|_2 \\
&\leq \|\{\hat{M} - \hat{G}\rho^*\}_{\hat{\mathcal{J}}}\|_2 + \| \left\{ \hat{G}(\hat{\rho} - \rho^*) \right\}_{\hat{\mathcal{J}}} \|_2 \\
&\leq \|\hat{\rho}\|_0^{1/2} \|\hat{M} - \hat{G}\rho^*\|_\infty + \sup_{\|a\|_0 \leq \|\hat{\rho}\|_0, \|a\|_2 = 1} a^\top \hat{G}(\hat{\rho} - \rho^*) \\
&\leq \|\hat{\rho}\|_0^{1/2} \lambda_R / \ell_N + \sup_{\|a\|_0 \leq \|\hat{\rho}\|_0, \|a\|_2 = 1} a^\top \hat{G}(\hat{\rho} - \rho^*),
\end{aligned}$$

where the last equality holds because Chernozhukov et al. (2021b) show $\|\hat{G}\rho^* - \hat{M}\|_\infty = O_P(\varepsilon_N)$ in the proof of their Lemma A5. For the second term on the right-hand side of the above display, there exists a large constant $C > 0$ such that, with probability greater than $1 - e/2$,

$$\begin{aligned}
|a^\top \hat{G}(\hat{\rho} - \rho^*)| &\leq \left(\frac{1}{N} \sum_{i=1}^N (h(X_i)^\top a)^2 \right)^{1/2} \left(\frac{1}{N} \sum_{i=1}^N (h(X_i)^\top (\hat{\rho} - \rho^*))^2 \right)^{1/2} \\
&\leq \left(\frac{1}{N} \sum_{i=1}^N (h(X_i)^\top a)^2 \right)^{1/2} |(\hat{\rho} - \rho^*)^\top \hat{G}(\hat{\rho} - \rho^*)| \\
&\leq \left(\frac{1}{N} \sum_{i=1}^N (h(X_i)^\top a)^2 \right)^{1/2} \left(\|\hat{\rho} - \rho^*\|_1 \|\hat{G}(\hat{\rho} - \rho^*)\|_\infty \right)^{1/2} \\
&\leq C \left(\frac{1}{N} \sum_{i=1}^N (h(X_i)^\top a)^2 \right)^{1/2} (s_h \lambda_R)^{1/2} \lambda_R^{1/2},
\end{aligned}$$

where the last inequality holds due to (D.2) and the fact that $\|\hat{G}(\hat{\rho} - \rho^*)\|_\infty = O_P(\lambda_R)$.[2]

---

the third inequality is due to Assumption 3.3, and the last inequality is by $s_h = C\varepsilon_N^{-2/(2\xi+1)}$. This again implies $\|\rho^* - \overline{\rho}\|_1 \leq \|(\rho^* - \overline{\rho})_{\mathcal{J}_0}\|_1 + \|(\rho^* - \overline{\rho})_{\mathcal{J}_0^c}\|_1 \leq C\varepsilon_N^{(2\xi-1)/(2\xi+1)}$.

[2] We have $\|\hat{G}(\hat{\rho} - \rho^*)\|_\infty \leq \|\hat{G}\hat{\rho} - \hat{M}\|_\infty + \|\hat{M} - M\|_\infty + \|G\rho^* - M\|_\infty + \|(\hat{G} - G)\rho^*\|_\infty = O_P(\lambda_R)$, where we use the facts that by the first order condition for Lasso regressions, $\|\hat{G}\hat{\rho} - \hat{M}\|_\infty = O(\lambda_R)$ and $\|G\rho^* - M\|_\infty = O(\varepsilon_N)$, $\|\hat{M} - M\|_\infty = O_P(\varepsilon_N)$ by Assumption 3, and $\|(\hat{G} - G)\rho^*\|_\infty = O_P(\varepsilon_N)$ by Chernozhukov et al. (2021b, Lemma A4).

Therefore, there exists a large constant $C > 0$ such that

$$\mathbb{P}\left(\|\hat{\rho}\|_0 \leq Cs_h \sup_{\|a\|_0 \leq \|\hat{\rho}\|_0, \|a\|_2 = 1} \frac{1}{N} \sum_{i=1}^N h(X_i)^\top a \right) \geq 1 - e/2. \tag{D.6}$$

If $\|\hat{\rho}\|_0 > 3C\overline{c}s_h$ for the constant C in (D.6) where $\overline{c}$ is defined in Assumption 3.4, then

$$\begin{aligned} Cs_h \sup_{\|a\|_0 \leq \|\hat{\rho}\|_0, \|a\|_2 = 1} \frac{1}{N} \sum_{i=1}^N h(X_i)^\top a &\leq Cs_h \left\lceil \frac{\|\hat{\rho}\|_0}{3C\overline{c}s_h} \right\rceil \sup_{\|a\|_0 \leq 3C\overline{c}s_h, \|a\|_2 = 1} \frac{1}{N} \sum_{i=1}^N h(X_i)^\top a \\ &\leq 0.5\|\hat{\rho}\|_0 < \|\hat{\rho}\|_0, \end{aligned}$$

where the first inequality is due to Belloni and Chernozhukov (2011, Lemma 13) under $\|\hat{\rho}\|_0/(3C\overline{c}s_h) > 1$, and the second inequality uses $\overline{c} \geq \sup_{\|a\|_0 \leq 3C\overline{c}s_h, \|a\|_2 = 1} \frac{1}{N} \sum_{i=1}^N h(X_i)^\top a$ provided that

$$\sup_{\rho \neq 0, \|\rho\|_0 \leq m_N} \frac{\rho^\top \hat{G} \rho}{\|\rho\|_2^2} \leq \overline{c}. \tag{D.7}$$

Therefore, by Assumption 3.4, we have

$$\mathbb{P}(\|\hat{\rho}\|_0 > 3C\overline{c}s_h) \leq \mathbb{P}\left(\|\hat{\rho}\|_0 > Cs_h \sup_{\|a\|_0 \leq \|\hat{\rho}\|_0, \|a\|_2 = 1} \frac{1}{N} \sum_{i=1}^N h(X_i)^\top a \right) + \mathbb{P}(\text{(D.7) is false}) \leq e.$$

Third, we are going to show (A.4). By (D.1) and (D.5), for any $e > 0$, we can find M and c such that $\hat{\omega}(x) \in \mathcal{G}^\omega$ with probability greater than $1 - e$, where

$$\mathcal{G}^\omega = \{h(X)^\top \rho : \rho \in \mathbb{R}^{p_h}, \|\rho\|_0 \leq Ms_h, \|\rho - \overline{\rho}\|_1 \leq MN^c(\log(p_h)/N)^{\frac{\xi - 1/2}{1 + 2\xi}}\}.$$

Then (A.4) directly follows the argument in the proof of Belloni et al. (2017, Theorem 5.1).

Last, we are going to show (A.5). Assumption 3 implies Assumptions 1–6 in Chernozhukov et al. (2021b), where their α_0, ρ_*, $\tilde{\rho}$, and $\hat{\rho}$ in our context are $\omega(x)$, ρ^*, $\overline{\rho}$, and $\hat{\rho}$, respectively.[3] Then (A.5) holds due to Chernozhukov et al. (2021b, Theorem 1) and the fact that $N^c\left(\log(p_h)/N\right)^{2\xi/(1+2\xi)} = O(N^c s_h \log(p_h)/N)$ for any constant $c > 0$.

[3]In particular, Assumption 3.4 implies Assumption 3 in Chernozhukov et al. (2021b) by Bickel et al. (2009, Lemma 4.1).

D.4 Proof of Theorem A.3

For a proof of this theorem, we let $\mathbb{P}_N f$ and $\mathbb{P}f$ denote $\frac{1}{N}\sum_{i=1}^N f(Z_i)$ and $\mathbb{E}f$, respectively. We write $a_N \lesssim b_N$ for two positive sequences a_N and b_N if there exists a constant independent of n such that $a_N \leq cb_N$. The constant c may vary in different contexts. For any estimator $\hat{\theta}$, we follow the empirical processes literature and denote $\mathbb{E}f(X,\hat{\theta})$ as $\mathbb{E}f(X,\theta)$ evaluated at $\theta = \hat{\theta}$.

The proof of Theorem A.3 is divided into three sections. In Section D.4.1, we prove three technical lemmas that will be used later. In Section D.4.2, we derive the linear expansion for $\hat{\theta}(\tau)$. In Section D.4.3, we derive the linear expansion for $\hat{\theta}^*(\tau)$.

D.4.1 Useful Lemmas

Define $\phi_i(q) = m_1(X_i,q) - \omega(X_i)(1\{Y_i \leq q\} - m_0(X_i,q)) - \theta(\tau)$ and $\hat{\phi}_i(q) = \hat{m}_0(X_i,q) - \hat{\omega}(X_i)(1\{Y_i \leq q\} - \hat{m}_0(X_i,q)) - \theta(\tau)$.

Lemma D.3. *If Assumptions 1–3 hold, then* $\sup_{\tau\in\Upsilon}|\mathbb{P}(\hat{\phi}_i(\hat{q}_\tau) - \phi_i(\hat{q}_\tau))| = O_P\left(\frac{N^c\sqrt{s_b s_h \log(p_b)\log(p_h)}}{N}\right)$ *and* $\sup_{\tau\in\Upsilon}|\mathbb{P}(\hat{\phi}_i(\hat{q}^*_\tau) - \phi_i(\hat{q}^*_\tau))| = O_P\left(\frac{N^c\sqrt{s_b s_h \log(p_b)\log(p_h)}}{N}\right)$.

Proof. We focus on the first result and the second one can be proved in the same manner. Using the law of iterated expectations and $m_0(x,q) = \int 1\{y \leq q\}dF_{Y|X=x}(y)$, we have

$$\begin{aligned}
&\int (m_1(x,q) - \omega(x)(1\{y \leq q\} - m_0(x,q)))\, dF_{Y,X}(y,x) \\
&- \int (\hat{m}_0(x,q) - \hat{\omega}(x)(1\{y \leq q\} - \hat{m}_0(x,q)))\, dF_{Y,X}(y,x) \\
&= \int (\hat{m}_1(x,q) - m_1(x,q))\, dF_X(x) + \int \omega(x)(\hat{m}_0(x,q) - m_0(x,q)) dF_X(x) \\
&\quad + \int (\hat{\omega}(x) - \omega(x))(\hat{m}_0(x,q) - m_0(x,q)) dF_X(x).
\end{aligned}$$

The integration by parts implies

$$\begin{aligned}
&\int \omega(x)(\hat{m}_0(x,q) - m_0(x,q)) f_{X_1|X_{-1}=x_{-1}}(x_1)dx_1 \\
&= -\int \left(\frac{\partial}{\partial x_1}\hat{m}_0(x,q) - \frac{\partial}{\partial x_1}m_0(x,q)\right) f_{X_1|X_{-1}=x_{-1}}(x_1)dx_1,
\end{aligned}$$

where $(\hat{m}_0(x,q) - m_0(x,q)) f_{X_1|X_{-1}=x_{-1}}(x_1)$ disappears on the boundary of x_1. Then

$$\int (m_1(x,q) - \omega(x)(1\{y \leq q\} - m_0(x,q)))\, dF_{Y,X}(y,x)$$

$$
\begin{aligned}
&- \int (\hat{m}_1(x,q) - \hat{\omega}(x)(1\{y \le q\} - \hat{m}_0(x,q)))\, dF_{Y,X}(y,x) \\
&= \int \left(\hat{m}_1(x,q) - \frac{\partial}{\partial x_1} \hat{m}_0(x,q) \right) dF_X(x) + \int (\hat{\omega}(x) - \omega(x))(\hat{m}_0(x,q) - m_0(x,q)) dF_X(x).
\end{aligned}
$$

Because $\sup_{\tau \in \Upsilon} |\hat{q}_\tau - q_\tau| = o_P(N^{-1/2})$, we have, with probability approaching one,

$$
\begin{aligned}
|\mathbb{P}(\hat{\phi}_i(\hat{q}_\tau) - \phi_i(\hat{q}_\tau))| &\le \sup_{q \in \mathcal{Q}^\delta} \left| \int \left(\hat{m}_1(x,q) - \frac{\partial}{\partial x_1} \hat{m}_0(x,q) \right) dF_X(x) \right| \\
&\quad + \sup_{q \in \mathcal{Q}^\delta} \left| \int (\hat{\omega}(x) - \omega(x))(\hat{m}_0(x,q) - m_0(x,q)) dF_X(x) \right| \\
&\le \left(\int (\hat{\omega}(x) - \omega(x))^2 dF_X(x) \right)^{1/2} \sup_{q \in \mathcal{Q}^\delta} \left(\int (\hat{m}_0(x,q) - m_0(x,q))^2 dF_X(x) \right)^{1/2} \\
&= O_P\left(\frac{N^c \sqrt{s_b s_h \log(p_b) \log(p_h)}}{N} \right),
\end{aligned}
$$

where the second inequality holds due to the fact that $\hat{m}_1(x,q) = \frac{\partial}{\partial x_1} \hat{m}_0(x,q)$ and the last equality holds due to Theorems A.1 and A.2. □

Lemma D.4. *Let $\tilde{\eta}_i = 1$ for every $i = 1, \ldots, N$ or $\tilde{\eta}_i = 1 + \eta_i$ for every $i = 1, \ldots, N$. If Assumptions 1–3 hold, then*

$$
\begin{aligned}
&\sup_{q \in \mathcal{Q}^\delta} |(\mathbb{P}_N - \mathbb{P}) \tilde{\eta}_i (\hat{\phi}_i(q) - \phi_i(q)) \\
&= O_P\left(\pi_N (\sqrt{s_h \log(p_h)/N} + \sqrt{s_b \log(p_b)/N}) + N^{-(1+d)/(2+d)} (s_h \log(p_h) + s_b \log(p_b)) \right).
\end{aligned}
$$

Proof. Define $\mathcal{M}(M)$ the set of $(\tilde{m}_1(x,q), \tilde{m}_0(x,q), \tilde{\omega}(x))$ which satisfies

$$
\begin{aligned}
&\{\tilde{m}_j(x,q) : q \in \mathcal{Q}^\delta\} \subset \mathcal{G}^{(j)},\ j = 0, 1, \\
&\sup_{q \in \mathcal{Q}^\delta} \int |\tilde{m}_j(x,q) - m_j(x,q)|^2 \, dF_X(x) \le M s_b \log(p_b)/N,\ j = 0, 1, \\
&\sup_{q \in \mathcal{Q}^\delta, x \in \mathcal{X}} |\tilde{m}_j(x,q) - m_j(x,q)| \le M \zeta_N s_b \sqrt{\log(p_b)/N},\ j = 0, 1, \\
&\int |\tilde{\omega}(x) - \omega(x)|^2 \, dF_X(x) \le M N^{2c} s_h \log(p_h)/N.
\end{aligned}
$$

For such $(\tilde{m}_1(x,q), \tilde{m}_0(x,q), \tilde{\omega}(x))$, we have

$$
\sup_{q \in \mathcal{Q}^\delta} \int |\tilde{\omega}(x) \tilde{m}_0(x,q) - \omega(x) m_0(x,q)|^2 \, dF_X(x)
$$

$$
\begin{aligned}
&\le \int (\tilde{\omega}(x)-\omega(x))^2 dF_X(x) + \sup_{q\in\mathcal{Q}^\delta} \int \omega^2(x)(\tilde{m}_0(x,q)-m_0(x,q))^2 dF_X(x)\\
&\le MN^{2c} s_h \log(p_h)/N\\
&+ \sup_{q\in\mathcal{Q}^\delta} \left(\int \omega^{2+d}(x) dF_X(x)\right)^{2/(2+d)} \left(\int (\tilde{m}_0(x,q)-m_0(x,q))^{4/d+2} dF_X(x)\right)^{d/(2+d)}\\
&\le MN^{2c} s_h \log(p_h)/N + M\Bigg(\sup_{q\in\mathcal{Q}^\delta, x\in\mathcal{X}} |\tilde{m}_0(x,q)-m_0(x,q)|^{4/d}\\
&\times \sup_{q\in\mathcal{Q}^\delta} \left(\int \omega^{2+d}(x) dF_X(x)\right)^{2/(2+d)} \int (\tilde{m}_0(x,q)-m_0(x,q))^2 dF_X(x)\Bigg)^{d/(2+d)}\\
&\le MN^{2c} s_h \log(p_h)/N + M(\zeta_N^{4/(2+d)} s_b^{(4+d)/(2+d)}) \log(p_b)/N \equiv \pi_N^2.
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
&\sup_{q\in\mathcal{Q}^\delta} \int |\tilde{\omega}(x)\tilde{m}_0(x,q)-\omega(x)m_0(x,q)|^2 dF_X(x) \le M\pi_N^2,\\
&\sup_{q\in\mathcal{Q}^\delta} |\tilde{\omega}(x)\tilde{m}_0(x,q)-\omega(x)m_0(x,q)| \le M,\\
&\int \left(\sup_{q\in\mathcal{Q}^\delta} |\tilde{m}_0(x,q)| + \sup_{q\in\mathcal{Q}^\delta} |m_1(x,q)|\right)^{2+d} dF_X(x) \le M, \quad \text{and}\\
&\int \left(\sup_{q\in\mathcal{Q}^\delta} |\tilde{\omega}_l(x)(1+|\tilde{m}_0(x,q)|)| + \sup_{q\in\mathcal{Q}^\delta} |\omega(x)(1+m_0(x,q))|\right)^{2+d} dF_X(x) \le M.
\end{aligned}
$$

Define

$$
F(X_i) = 4|\tilde{\eta}_i|(|\omega(x)|+M) + |\tilde{\eta}_i|(\sup_{q\in\mathcal{Q}^\delta} |m_1(x,q)|+M)
$$

and

$$
\mathcal{F} = \left\{ \begin{array}{c} \tilde{\eta}_i(\tilde{m}_1(X_i,q) - \tilde{\omega}(X_i)(1\{Y_i\le q\} - \tilde{m}_0(X_i,q)))\\ -\tilde{\eta}_i(m_1(X_i,q) - \omega(X_i)(1\{Y_i\le q\} - m_0(X_i,q))) \end{array} : (\tilde{m}_1(x,q), \tilde{m}_0(x,q), \tilde{\omega}(x)) \in \mathcal{M}(M) \right\}.
$$

By Theorems A.1 and A.2, we have

$$
\begin{aligned}
&\sup_Q N(\mathcal{F}, \|\cdot\|_{Q,2}, \varepsilon\|F\|_{Q,2}) \le C\left(\frac{p_b}{\varepsilon}\right)^{cs_b}\left(\frac{p_h}{\varepsilon}\right)^{cs_h},\\
&\sup_{f\in\mathcal{F}} \mathbb{E}f^2 \le C \sup_{q\in\mathcal{Q}^\delta} \mathbb{E}(\tilde{m}_1(x,q)-m_1(x,q))^2\\
&+ C\mathbb{E}(\tilde{\omega}(x)-\omega(x))^2 + C\sup_{q\in\mathcal{Q}^\delta} \mathbb{E}(\tilde{m}_0(x,q)\tilde{\omega}(x) - m_0(x,q)\omega(x))^2 \le M\pi_N^2, \text{ and}\\
&\mathbb{E}F^{2+d} < \infty.
\end{aligned}
$$

By Chernozhukov, Chetverikov, and Kato (2014b, Corollary 5.1), we have

$$\mathbb{P}\sup_{f\in\mathcal{F}}|(\mathbb{P}_N-\mathbb{P})f| \lesssim \sqrt{\frac{\pi_N^2 s_b}{N}\log\left(\frac{p_b\mathbb{E}[F(X)^2]^{1/2}}{\pi_N}\right)} + \sqrt{\frac{\pi_N^2 s_h}{N}\log\left(\frac{p_h\mathbb{E}[F(X)^2]^{1/2}}{\pi_N}\right)}$$
$$+ \frac{s_h\mathbb{E}[(\max_i F(X_i))^2]^{1/2}}{N}\log\left(\frac{p_h\mathbb{E}[F(X)^2]^{1/2}}{\pi_N}\right) + \frac{s_b\mathbb{E}[(\max_i F(X_i))^2]^{1/2}}{N}\log\left(\frac{p_b\mathbb{E}[F(X)^2]^{1/2}}{\pi_N}\right)$$
$$\lesssim \pi_N(\sqrt{s_h\log(p_h)/N}+\sqrt{s_b\log(p_b)/N}) + N^{-(1+d)/(2+d)}(s_h\log(p_h)+s_b\log(p_b)),$$

where the last inequality is due to the fact that $\mathbb{E}[F(X)^2]^{1/2} = O(N^{1/(2+d)})$. In addition, Theorems A.1 and A.2 show that, for any $e > 0$, we can find a sufficiently large constant $M > 0$ such that $(\hat{m}_0, \hat{m}_0, \hat{\omega}) \in \mathcal{M}(M)$ occurs with probability greater than $1 - e$. This further implies that $\hat{\phi}_i(q) \in \mathcal{F}$ with probability greater than $1 - e$, and thus,

$$\sup_{q\in\mathcal{Q}^\delta}|(\mathbb{P}_N-\mathbb{P})\tilde{\eta}_i(\hat{\phi}_i(q)-\phi_i(q))$$
$$= O_P\left(\pi_N(\sqrt{s_h\log(p_h)/N}+\sqrt{s_b\log(p_b)/N}) + N^{-(1+d)/(2+d)}(s_h\log(p_h)+s_b\log(p_b))\right).$$

This leads to the desired result. □

Lemma D.5. *If Assumptions 1–3 hold, then* $\sup_{\tau\in\Upsilon}|(\mathbb{P}_N-\mathbb{P})(\phi_i(\hat{q}_\tau)-\phi_i(q_\tau))| = o_P(N^{-1/2})$ *and* $\sup_{\tau\in\Upsilon}|(\mathbb{P}_N-\mathbb{P})(\eta_i+1)(\phi_i(\hat{q}^*_\tau)-\phi_i(q_\tau))| = o_P(N^{-1/2})$.

Proof. We know that $\sup_{\tau\in\Upsilon}|\hat{q}_\tau - q_\tau| = O_P(N^{-1/2})$ and $\sup_{\tau\in\Upsilon}|\hat{q}^*_\tau - q_\tau| = O_P(N^{-1/2})$. (See Section D.6.2 for more detail.) These conditions imply that, for any $\varepsilon > 0$, there exists a constant $M > 0$ such that

$$\mathbb{P}\left(\sup_{\tau\in\Upsilon}|\hat{q}^*(\tau)-q_\tau| \le MN^{-1/2}, \sup_{\tau\in\Upsilon}|\hat{q}_\tau-q_\tau| \le MN^{-1/2}\right) \ge 1-\varepsilon.$$

Next, we show

$$\sup_{|v|\le M,\tau\in\Upsilon}|(\mathbb{P}_N-\mathbb{P})\tilde{\eta}_i(\phi_i(q_\tau+vN^{-1/2})-\phi_i(q_\tau))| = o_P(N^{-1/2}).$$

Let $\mathcal{F} = \{\tilde{\eta}_i\left(\phi_i(q_\tau+vN^{-1/2})-\phi_i(q_\tau)\right) : |v|\le M, \tau\in\Upsilon\}$ with envelope

$$F(X_i) = |\tilde{\eta}_i|\sup_{q\in\mathcal{Q}^\delta}|m_1(x,q)| + |\tilde{\eta}_i|\sup_{q\in\mathcal{Q}^\delta}|\omega(x)(1+m_0(x,q))|.$$

Note that $\mathcal{F}$ is nested in $\{\tilde{\eta}_i(\phi_i(q_1)-\phi_i(q_2)) : q_1, q_2 \in \mathbb{R}\}$. Because $m_j(x,q)$ is Lipschitz continuous in q and $\{1\{Y \le q\} : q \in \mathbb{R}\}$ is a VC class with VC index 2, we have $J(v) = \int_0^v \sqrt{1+\log(\sup_Q N(\mathcal{F}, \|\cdot\|_{Q,2}, \varepsilon\|F\|_{Q,2}))}d\varepsilon \lesssim v\sqrt{\log(a/v)}$ for some constant $a > 0$.

Last,

$$\sup_{f\in\mathcal{F}}\mathbb{P}f^2 \le \mathbb{P}\sup_{\tau\in\Upsilon,|v|\le M}\Big\{\big|m_1(X,q_\tau+vN^{-1/2})-m_1(X,q_\tau)\big|$$
$$+|\omega(X)|\left(\big|m_0(X,q_\tau+vN^{-1/2})-m_0(X,q_\tau)\big|+\big|1\{Y\le q_\tau\}-1\{Y\le q_\tau+vN^{-1/2}\}\big|\right)\Big\}^2$$
$$\lesssim N^{-1/2}.$$

By Chernozhukov et al. (2014b, Corollary 5.1), we have

$$\mathbb{P}\sup_{|v|\le M,\tau\in\Upsilon}|(\mathbb{P}_N-\mathbb{P})(\phi_i(q_\tau+vN^{-1/2})-\phi_i(q_\tau))|$$
$$=\mathbb{P}|\sup_{f\in\mathcal{F}}(\mathbb{P}_N-\mathbb{P})f|$$
$$\lesssim\sqrt{\frac{1}{N^{3/2}}\log\left(a\mathbb{E}[F(X)^2]^{1/2}N\right)}+\frac{\mathbb{E}[(\max_i F(X_i))^2]^{1/2}}{N}\log\left(a\mathbb{E}[F(X)^2]^{1/2}N\right)=o(N^{-1/2}).$$

Therefore the statement of this lemma holds. □

D.4.2 Linear Expansion for $\hat{\theta}(\tau)$

Taking $\tilde{\eta}_i=1$ and by (6) and Lemmas D.3, D.4, and D.5, we have

$$\begin{aligned}\hat{\theta}(\tau)-\theta(\tau)&=\mathbb{P}_N\hat{\phi}_i(\hat{q}_\tau)\\&=(\mathbb{P}_N-\mathbb{P})\phi_i(q_\tau)+\mathbb{P}\phi_i(\hat{q}_\tau)+(\mathbb{P}_N-\mathbb{P})(\hat{\phi}_i(\hat{q}_\tau)-\phi_i(\hat{q}_\tau))\\&\quad+\mathbb{P}(\hat{\phi}_i(\hat{q}_\tau)-\phi_i(\hat{q}_\tau))+(\mathbb{P}_N-\mathbb{P})(\phi_i(\hat{q}_\tau)-\phi_i(q_\tau)).\end{aligned}$$

Rearranging the above equation and the extra condition in Theorem A.3, we have

$$\hat{\theta}(\tau)-\theta(\tau)=(\mathbb{P}_N-\mathbb{P})(\phi_i(q_\tau))+\mathbb{P}\phi_{i,l}(\hat{q}_\tau)+R_N(\tau)$$

where

$$\sup_{\tau\in\Upsilon}|R_N(\tau)|=O_P\left(\pi_N(\sqrt{s_h\log(p_h)/N}+\sqrt{s_b\log(p_b)/N})+N^{-(1+d)/(2+d)}(s_h\log(p_h)+s_b\log(p_b))\right)$$
$$+O_P\left(\frac{N^c\sqrt{s_bs_h\log(p_b)\log(p_h)}}{N}\right)+o_P(N^{-1/2}).$$

By $\mathbb{E}m_1(X, q_\tau) = 0$ and the usual delta method,

$$\mathbb{P}\phi_i(\hat{q}_\tau) = (\mathbb{E}m_1(X, \hat{q}_\tau) - \mathbb{E}m_1(X, q_\tau)) = \frac{\frac{\partial}{\partial q}\mathbb{E}m_1(X, q_\tau)}{f_Y(q_\tau)} \frac{1}{N}\sum_{i=1}^{N}(\tau - 1\{Y_i \leq q_\tau\}) + o_P(N^{-1/2})$$

where the $o_P(N^{-1/2})$ term holds uniformly over $\tau \in \Upsilon$. Then

$$\hat{\theta}(\tau) - \theta(\tau) = \mathbb{P}_N \eta_i \mathrm{IF}_i^\theta(\tau) + R_\theta(\tau),$$

where

$$\sup_{\tau\in\Upsilon} |R_\theta(\tau)| = O_P\left(\pi_N(\sqrt{s_h \log(p_h)/N} + \sqrt{s_b \log(p_b)/N}) + N^{-(1+d)/(2+d)}(s_h \log(p_h) + s_b \log(p_b))\right) + O_P\left(\frac{N^c\sqrt{s_b s_h \log(p_b)\log(p_h)}}{N}\right) + o_P(N^{-1/2}). \tag{D.8}$$

the desired result holds because under the condition in Theorem A.3, $\sup_{\tau\in\Upsilon}|R_\theta(\tau)| = o_P(N^{-1/2})$.

D.4.3 Linear Expansion for $\hat{\theta}^*(\tau)$

By Lemmas D.3, D.4, and D.5 with $\tilde{\eta}_i = 1 + \eta_i$, we have

$$\begin{aligned}
\hat{\theta}^*(\tau) - \theta(\tau) &= \frac{1}{\sum_{i\in[N]} 1 + \eta_i} \sum_{i\in[N]} (1+\eta_i)\hat{\phi}_i(\hat{q}^*_\tau) \\
&= \frac{N}{\sum_{i\in[N]} 1 + \eta_i}\Big((\mathbb{P}_N - \mathbb{P})(1+\eta_i)\phi_i(q_\tau) + \mathbb{P}(1+\eta_i)\phi_i(\hat{q}^*_\tau) \\
&\quad + (\mathbb{P}_N - \mathbb{P})(1+\eta_i)(\hat{\phi}_i(\hat{q}^*_\tau) - \phi_i(\hat{q}^*_\tau)) + \mathbb{P}(1+\eta_i)(\hat{\phi}_i(\hat{q}^*_\tau) - \phi_i(\hat{q}^*_\tau)) \\
&\quad + (\mathbb{P}_N - \mathbb{P})(1+\eta_i)(\phi_i(\hat{q}^*_\tau) - \phi_i(q_\tau))\Big) \\
&= \frac{N}{\sum_{i\in[N]} 1 + \eta_i}\Big((\mathbb{P}_N - \mathbb{P})(1+\eta_i)\phi_i(q_\tau) + \mathbb{P}(1+\eta_i)\phi_i(\hat{q}^*_\tau)\Big) + R^*_N(\tau),
\end{aligned}$$

where

$$\sup_{\tau\in\Upsilon} |R^*_N(\tau)| = O_P\left(\pi_N(\sqrt{s_h \log(p_h)/N} + \sqrt{s_b \log(p_b)/N}) + N^{-(1+d)/(2+d)}(s_h \log(p_h) + s_b \log(p_b))\right) + O_P\left(\frac{N^c\sqrt{s_b s_h \log(p_b)\log(p_h)}}{N}\right) + o_P(N^{-1/2}).$$

In addition,

$$\begin{aligned}\mathbb{P}(1+\eta_i)\phi_i(\hat{q}^*_\tau) &= (\mathbb{E}m_1(X,\hat{q}^*_\tau) - \mathbb{E}m_1(X,q^*_\tau)) \\ &= \frac{\frac{\partial}{\partial q}\mathbb{E}m_1(X,q_\tau)}{f_Y(q_\tau)}\frac{1}{N}\sum_{i=1}^N (1+\eta_i)(\tau - 1\{Y_i \le q_\tau\}) + o_P(N^{-1/2}),\end{aligned}$$

where the $o_P(N^{-1/2})$ term holds uniformly over $\tau \in \Upsilon$. Therefore, we have

$$\hat{\theta}^*(\tau) - \hat{\theta}(\tau) = \mathbb{P}_N\eta_i\mathrm{IF}^\theta_i(\tau) + R^*_N(\tau),$$

where

$$\begin{aligned}\sup_{\tau\in\Upsilon}|R^*_\theta(\tau)| &= O_P\left(\pi_N(\sqrt{s_h\log(p_h)/N} + \sqrt{s_b\log(p_b)/N}) + N^{-(1+d)/(2+d)}(s_h\log(p_h) + s_b\log(p_b))\right) \\ &\quad + O_P\left(\frac{N^c\sqrt{s_bs_h\log(p_b)\log(p_h)}}{N}\right) + o_P(N^{-1/2}). \qquad \text{(D.9)}\end{aligned}$$

the desired result holds because under the condition in Theorem A.3, $\sup_{\tau\in\Upsilon}|R_\theta(\tau)| = o_P(N^{-1/2})$.

D.5 Proof of Corollary 1

The desired result holds due to the linear expansions in Theorem A.3 and the fact that $\{\mathrm{IF}^\theta_i(\tau) : \tau \in \Upsilon\}$ is Donsker.

D.6 Proof of Theorem A.4

D.6.1 Linear Expansion for $\widehat{UQPE}(\tau)$

Note that $\hat{f}_Y(\hat{q}_\tau) - f_Y(q_\tau) = A_1(\tau) + A_2(\tau) + A_3(\tau)$, where $A_1(\tau) \equiv (\mathbb{P}_N - \mathbb{P})\frac{1}{h_1}K_1\left(\frac{Y_i-\hat{q}_\tau}{h_1}\right)$, $A_2(\tau) \equiv \mathbb{P}\frac{1}{h_1}K_1\left(\frac{Y_i-\hat{q}_\tau}{h_1}\right) - f_Y(\hat{q}_\tau)$ and $A_3(\tau) \equiv f_Y(\hat{q}_\tau) - f_Y(q_\tau)$. Below we will analyze $A_1(\tau)$, $A_2(\tau)$, and $A_3(\tau)$, and then derive the linear expansion of $\widehat{UQPE}(\tau)$.

First, we will analyze $A_1(\tau)$. Let $R_1(\tau) = A_1(\tau) - (\mathbb{P}_N - \mathbb{P})\frac{1}{h_1}K_1\left(\frac{Y_i-q_\tau}{h_1}\right)$. Because $\sup_{\tau\in\Upsilon}|\hat{q}_\tau - q_\tau| = O_P(N^{-1/2})$. For any $\varepsilon > 0$, there exists a constant $M > 0$ such that, with probability greater than $1-\varepsilon$,

$$\sup_{\tau\in\Upsilon}|R_1(\tau)| \le \sup_{q\in\mathcal{Q}^\delta, |v|\le M}\left|(\mathbb{P}_N - \mathbb{P})\left(\frac{1}{h_1}K_1\left(\frac{Y_i - q - v/\sqrt{N}}{h_1}\right) - \frac{1}{h_1}K_1\left(\frac{Y_i-q}{h_1}\right)\right)\right|.$$

In the following, we aim to bound $\sup_{q\in\mathcal{Q}^\delta,|v|\le M}\left|(\mathbb{P}_N-\mathbb{P})\frac{\tilde\eta_i}{h_1}\left(K_1\left(\frac{Y_i-q-v/\sqrt{N}}{h_1}\right)-K_1\left(\frac{Y_i-q}{h_1}\right)\right)\right|$. Consider the class of functions $\mathcal{F}=\left\{\frac{\tilde\eta_i}{h_1}\left(K_1\left(\frac{y-q-v/\sqrt{N}}{h_1}\right)-K_1\left(\frac{y-q}{h_1}\right)\right):q\in\mathcal{Q}^\delta,|v|\le M\right\}$ with an envelope function $F_i=C|\tilde\eta_i|/h$ for some constant $C>0$ such that $(\mathbb{E}[(\max_i F_i)^2])^{1/2}\lesssim\sqrt{\log(N)}$. We note that $\mathcal{F}$ is a VC-class with a fixed VC index and

$$\sup_{f\in\mathcal{F}}\mathbb{P}f^2=\sup_{q\in\mathcal{Q}^\delta,|v|\le M}\int\left(K_1(u-\frac{v}{\sqrt{N}h_1})-K_1(u)\right)^2 f_Y(q+h_1u)du\lesssim 1/(Nh_1^2).$$

Therefore, Chernozhukov et al. (2014b, Corollary 5.1) implies

$$\mathbb{E}\sup_{q\in\mathcal{Q}^\delta,|v|\le M}\left|(\mathbb{P}_N-\mathbb{P})\left(\frac{\tilde\eta_i}{h_1}\left(K_1\left(\frac{Y_i-q-v/\sqrt{N}}{h_1}\right)-K_1\left(\frac{Y_i-q}{h_1}\right)\right)\right)\right|\lesssim\frac{\sqrt{\log(N)}}{Nh_1}+\frac{\log(N)^{3/2}}{Nh_1},$$

and thus, $\sup_{\tau\in\Upsilon}|R_1(\tau)|=o_P(N^{-1/2})$.

Second, we will analyze $A_2(\tau)$. Let $R_2(\tau)=A_2(\tau)-\frac{f_Y^{(2)}(q_\tau)(\int u^2K_1(u)du)}{2}h_1^2$. By the Taylor expansion, we have

$$\begin{aligned}\sup_{\tau\in\Upsilon}|R_2(\tau)|\le&\sup_{\tau\in\Upsilon}\left|\int(f_Y(\hat q_\tau+uh_1)-f_Y(\hat q_\tau))K_1(u)du-\frac{f_Y^{(2)}(q_\tau)(\int u^2K_1(u)du)}{2}h_1^2\right|\\ \lesssim&\sup_{\tau\in\Upsilon}\frac{|f_Y^{(2)}(q_\tau)-f_Y^{(2)}(\tilde q_\tau)|(\int u^2K_1(u)du)}{2}h_1^2,\end{aligned}$$

where $\tilde q_\tau$ is between $\hat q_\tau$ and $\hat q_\tau+h_1$ such that $\sup_{\tau\in\Upsilon}|\tilde q_\tau-q_\tau|\le\sup_{\tau\in\Upsilon}|\tilde q_\tau-\hat q_\tau|+\sup_{\tau\in\Upsilon}|\hat q_\tau-q_\tau|=O_P(h_1+N^{-1/2})$. Therefore, $\sup_{\tau\in\Upsilon}|R_2(\tau)|=O_P(h_1^3+h_1N^{-1/2})=o_P(N^{-1/2})$.

Third, we will analyze $A_3(\tau)$. By the delta method, we have

$$A_3(\tau)=f_Y^{(1)}(q_\tau)(\hat q_\tau-q_\tau)+R_3'(\tau)=\frac{f_Y^{(1)}(q_\tau)}{f_Y(q_\tau)}\left(\frac{1}{N}\sum_{i=1}^N(\tau-\mathbf{1}\{Y_i\le q_\tau\})\right)+R_3(\tau),$$

where $\sup_{\tau\in\Upsilon}|R_3'(\tau)|+\sup_{\tau\in\Upsilon}|R_3(\tau)|=o_P(N^{-1/2})$.

Last, we will derive the linear expansion of $\widehat{UQPE}(\tau)$. Combining the analyses of $A_1(\tau)$, $A_2(\tau)$, and $A_3(\tau)$, we have

$$\hat f_Y(\hat q_\tau)-f_Y(q_\tau)=(\mathbb{P}_N-\mathbb{P})\frac{1}{h_1}K_1\left(\frac{Y_i-q_\tau}{h_1}\right)+\frac{f_Y^{(2)}(q_\tau)(\int u^2K_1(u)du)}{2}h_1^2+R_4(\tau),\quad\text{(D.10)}$$

where $\sup_{\tau\in\Upsilon}|R_4(\tau)| = O_P(N^{-1/2})$. By (D.8) and the condition in Theorem A.4, we have

$$\sup_{\tau\in\Upsilon}\left|\hat{\theta}(\tau)-\theta(\tau)\right| = o_P((\log(N)Nh_1)^{-1/2}).$$

Therefore, we have

$$\sup_{\tau\in\Upsilon}|\hat{f}_Y(\hat{q}_\tau)-f_Y(q_\tau)| = O_P(\log^{1/2}(N)(Nh_1)^{-1/2}+h_1^2),$$

$$\sup_{\tau\in\Upsilon}\left|\frac{\hat{\theta}(\tau)-\theta(\tau)}{f_Y(q_\tau)}\right| = o_P((\log(N)Nh_1)^{-1/2}),$$

$$\sup_{\tau\in\Upsilon}\left|\frac{(\hat{\theta}(\tau)-\theta(\tau))(\hat{f}_Y(\hat{q}_\tau)-f_Y(q_\tau))}{\hat{f}_Y(\hat{q}_\tau)f_Y(q_\tau)}\right| = o_P((\log(N)Nh_1)^{-1/2}),$$

and

$$\sup_{\tau\in\Upsilon}\left|\frac{\theta(\tau)(\hat{f}_Y(\hat{q}_\tau)-f_Y(q_\tau))^2}{f_Y^2(q_\tau)\hat{f}_Y(\hat{q}_\tau)}\right| = O_P(\log(N)(Nh_1)^{-1}+h_1^4) = o_P((\log(N)Nh_1)^{-1/2}).$$

Therefore

$$\begin{aligned}\widehat{UQPE}(\tau)-UQPE(\tau) =& -\frac{\hat{\theta}(\tau)-\theta(\tau)}{f_Y(q_\tau)}+\frac{\theta(\tau)(\hat{f}_Y(\hat{q}_\tau)-f_Y(q_\tau))}{f_Y^2(q_\tau)}\\ &+\frac{(\hat{\theta}(\tau)-\theta(\tau))(\hat{f}_Y(\hat{q}_\tau)-f_Y(q_\tau))}{\hat{f}_Y(\hat{q}_\tau)f_Y(q_\tau)}-\frac{\theta(\tau)(\hat{f}_Y(\hat{q}_\tau)-f_Y(q_\tau))^2}{f_Y^2(q_\tau)\hat{f}_Y(\hat{q}_\tau)}\\ =&\frac{1}{N}\sum_{i=1}^N \mathrm{IF}_i(\tau)+\frac{\theta(\tau)f_Y^{(2)}(q_\tau)(\int u^2K_1(u)du)h_1^2}{2f_Y^2(q_\tau)}+R(\tau), \end{aligned} \tag{D.11}$$

where $\sup_{\tau\in\Upsilon}|R(\tau)| = o_P((\log(N)Nh_1)^{-1/2})$.

D.6.2 Linear Expansion for $\widehat{UQPE}^*(\tau)$

First, we will derive the linear expansion of $\hat{q}_\tau^*$. Note that $\hat{q}_\tau^*$ is the optimizer of the objective function $\sum_{i=1}^N \rho_\tau(Y_i-q) - q\sum_{i=1}^N \eta_i(\tau-\mathbf{1}\{Y_i\le\hat{q}_\tau\})$. Define the local parameter as $\hat{u} = \sqrt{N}(\hat{q}_\tau^*-q_\tau)$. Then

$$\hat{u} = \arg\min_u \sum_{i=1}^N \rho_\tau(Y_i-q_\tau-uN^{-1/2}) - uN^{-1/2}\sum_{i=1}^N \eta_i(\tau-\mathbf{1}\{Y_i\le\hat{q}_\tau\})$$

Note that $u \mapsto \sum_{i=1}^N \rho_\tau(Y_i - q_\tau - uN^{-1/2}) - uN^{-1/2}\sum_{i=1}^N \eta_i(\tau - \mathbf{1}\{Y_i \le \hat{q}_\tau\})$ is convex in u for any $\tau \in \Upsilon$. By the Knight's identity, we can show that

$$\left(\sum_{i=1}^N \rho_\tau(Y_i - q_\tau - uN^{-1/2}) - uN^{-1/2}\sum_{i=1}^N \eta_i(\tau - \mathbf{1}\{Y_i \le \hat{q}_\tau\})\right) - \left(-\frac{u}{\sqrt{N}}\sum_{i=1}^N (\eta_i+1)(\tau - \mathbf{1}\{Y_i \le q_\tau\}) + \frac{f_Y(q_\tau)u^2}{2}\right)$$

is $o_P(1)$ pointwise in u. By the convexity lemma (Pollard, 1991), we have

$$\hat{q}^*_\tau - q_\tau = \frac{1}{Nf_Y(q_\tau)}\sum_{i=1}^N (\eta_i+1)(\tau - \mathbf{1}\{Y_i \le q_\tau\}) + R^*_1(\tau), \tag{D.12}$$

where $\sup_{\tau\in\Upsilon}|R^*_1(\tau)| = o_P(N^{-1/2})$.

Second, we will derive the linear expansion of $\hat{f}^*_Y(\hat{q}^*_\tau)$. Let $\hat{N} = \sum_{i=1}^N(\eta_i+1)$. Note that $\hat{f}^*_Y(\hat{q}^*_\tau) - f_Y(q_\tau) = \frac{N}{\hat{N}}(\mathbb{P}_N - \mathbb{P})\frac{(1+\eta_i)}{h_1}K_1\left(\frac{Y_i - \hat{q}^*_\tau}{h_1}\right) + \frac{N}{\hat{N}}\left(\mathbb{P}\frac{1}{h_1}K_1\left(\frac{Y_i - \hat{q}^*_\tau}{h_1}\right) - f_Y(\hat{q}^*_\tau)\right) + \frac{N}{\hat{N}}\left(f_Y(\hat{q}^*_\tau) - f_Y(q_\tau)\right)$. Following the same argument in the proof in Section D.6.1 and the fact that $\left|\frac{N}{\hat{N}} - 1\right| = O_P(N^{-1/2})$, we have

$$\frac{N}{\hat{N}}(\mathbb{P}_N - \mathbb{P})\frac{(1+\eta_i)}{h_1}K_1\left(\frac{Y_i - \hat{q}^*_\tau}{h_1}\right) = (\mathbb{P}_N - \mathbb{P})\frac{(1+\eta_i)}{h_1}K_1\left(\frac{Y_i - q_\tau}{h_1}\right) + R^*_1(\tau),$$

$$\frac{N}{\hat{N}}\left(\mathbb{P}\frac{1}{h_1}K_1\left(\frac{Y_i - \hat{q}^*_\tau}{h_1}\right) - f_Y(\hat{q}^*_\tau)\right) = \frac{f^{(2)}_Y(q_\tau)(\int u^2K_1(u)du)h_1^2}{2} + R^*_2(\tau),$$

and

$$\begin{aligned}\frac{N}{\hat{N}}\left(f_Y(\hat{q}^*_\tau) - f_Y(q_\tau)\right) &= f^{(1)}_Y(q_\tau)(\hat{q}^*_\tau - q_\tau) + R^*_3(\tau) \\ &= \frac{f^{(1)}_Y(q_\tau)}{f_Y(q_\tau)}\left(\frac{1}{N}\sum_{i=1}^N(\eta_i+1)(\tau - \mathbf{1}\{Y_i \le q_\tau\})\right) + R^*_4(\tau),\end{aligned}$$

where $\sup_{\tau\in\Upsilon, j=1,\cdots,4}|R^*_j(\tau)| = o_P(N^{-1/2})$. This implies

$$\hat{f}^*_Y(\hat{q}^*_\tau) - f_Y(q_\tau) = (\mathbb{P}_N - \mathbb{P})(\eta_i+1)\frac{1}{h_1}K_1\left(\frac{Y_i - q_\tau}{h_1}\right) + \frac{f^{(2)}_Y(q_\tau)(\int u^2K_1(u)du)}{2}h_1^2 + R^*_5(\tau),$$

where $\sup_{\tau\in\Upsilon}|R^*_5(\tau)| = O_P(N^{-1/2})$.

Last, we will derive the linear expansion of $\widehat{UQPE}^*(\tau)$. By (D.8), (E.4), and the condition in Theorem A.4, we have

$$\sup_{\tau\in\Upsilon}\left|\hat{\theta}^*(\tau)-\theta(\tau)\right| = o_P((\log(N)Nh_1)^{-1/2}).$$

Therefore,

$$\begin{aligned}\widehat{UQPE}^*(\tau)-UQPE(\tau) =& -\frac{\hat{\theta}^*(\tau)-\theta(\tau)}{f_Y(q_\tau)}+\frac{\theta(\tau)(\hat{f}_Y^*(\hat{q}_\tau^*)-f_Y(q_\tau))}{f_Y^2(q_\tau)}\\ &+\frac{(\hat{\theta}^*(\tau)-\theta(\tau))(\hat{f}_Y^*(\hat{q}_\tau^*)-f_Y(q_\tau))}{\hat{f}_Y^*(\hat{q}_\tau^*)f_Y(q_\tau)}-\frac{\theta(\tau)(\hat{f}_Y^*(\hat{q}_\tau^*)-f_Y(q_\tau))^2}{f_Y^2(q_\tau)\hat{f}_Y^*(\hat{q}_\tau^*)}\\ =&\frac{1}{N}\sum_{i=1}^N(1+\eta_i)\mathrm{IF}_i(\tau)+\frac{\theta(\tau)f_Y^{(2)}(q_\tau)(\int u^2K_1(u)du)h_1^2}{2f_Y^2(q_\tau)}+R_6^*(\tau)\end{aligned} \tag{D.13}$$

where $\sup_{\tau\in\Upsilon}|R_6^*(\tau)| = o_P((\log(N)Nh_1)^{-1/2})$. Taking difference between (D.11) and (D.13), we have

$$\widehat{UQPE}^*(\tau)-\widehat{UQPE}(\tau) = \frac{1}{N}\sum_{i=1}^N\eta_i\mathrm{IF}_i(\tau)+R^*(\tau),$$

where $\sup_{\tau\in\Upsilon}|R^*(\tau)| = o_P((\log(N)Nh_1)^{-1/2})$.

D.7 Proof of Corollary 2

Corollary 2 is a direct consequence of Chernozhukov, Chetverikov, and Kato (2014a, Corollary 3.1). In order to apply Chernozhukov et al. (2014a, Corollary 3.1), we need to verify Conditions H1–H4. Our Theorem A.4 shows that

$$\widehat{UQPE}(\tau)-UQPE(\tau) = \frac{1}{N}\sum_{i=1}^N\frac{\theta(\tau)}{f_Y^2(q_\tau)h_1}K_1\left(\frac{Y_i-q_\tau}{h_1}\right)+R(\tau),$$

and

$$\widehat{UQPE}^*(\tau)-\widehat{UQPE}(\tau) = \frac{1}{N}\sum_{i=1}^N\eta_i\frac{\theta(\tau)}{f_Y^2(q_\tau)h_1}K_1\left(\frac{Y_i-q_\tau}{h_1}\right)+R^*(\tau),$$

where $\sup_{\tau\in\Upsilon}|R(\tau)| = o_P((Nh_1\log(N))^{-1/2})$ and $\sup_{\tau\in\Upsilon}|R^*(\tau)| = o_P((Nh_1\log(N))^{-1/2})$. Therefore, the original and multiplier bootstrap estimators can be approximated by local empirical processes with a kernel function and the approximation errors are $o_P((\log(N))^{-1/2})$ uniformly over $\tau\in\Upsilon$. Following Chernozhukov et al. (2014b, Proposition 3.2 and Remark 3.2), the approximation errors are asymptotically negligible. Focusing on the local empirical process part, Conditions H1–H4 can be verified by Chernozhukov et al. (2014a, Theorem 3.2). Specifically, Condition VC in Chernozhukov et al. (2014a) holds where, in their notations, a_n and v_n are constants, $b_n = h_1^{-1/2}$, $K_n = \log(N)$, σ_n^2 is bounded, and $\log^4(N)/Nh_1 = o(N^{-c})$ for some constant $c > 0$ as we assume $h_1 = cN^{-H}$ for $H < 1/4$.

E Proofs of Results in Section C

E.1 Proof of Theorem C.1

For a proof of this theorem, we let $\mathbb{P}_N f$, $\mathbb{P}_{n,l} f$, $\mathbb{P}_l f$, and $\mathbb{P} f$ denote $\frac{1}{N}\sum_{i=1}^N f(Z_i)$, $\frac{1}{n}\sum_{i\in I_l} f(Z_i)$, $\mathbb{E}(f(Z_i)|\{Z_j\}_{j\in I_l^c})$, and $\mathbb{E}f$, respectively. We write $a_N \lesssim b_N$ for two positive sequences a_N and b_N if there exists a constant independent of n such that $a_N \leq cb_N$. The constant c may vary in different contexts. For any estimator $\hat{\theta}$, we follow the empirical processes literature and denote $\mathbb{E}f(X,\hat{\theta})$ as $\mathbb{E}f(X,\theta)$ evaluated at $\theta = \hat{\theta}$.

The proof of Theorem C.1 is divided into three sections. In Section E.1.1, we prove several technical lemmas that will be used later. In Section E.1.2, we derive the linear expansion of $\widehat{\theta}_{cf}(\tau)$. In Section E.1.3, we derive the linear expansion of $\hat{\theta}^*_{cf}(\tau)$.

E.1.1 Useful Lemmas

Define $\phi_i(q) = m_1(X_i,q) - \omega(X_i)(1\{Y_i \leq q\} - m_0(X_i,q)) - \theta(\tau)$ and $\hat{\phi}_{i,l}(q) = \hat{m}_{1,l}(X_i,q) - \hat{\omega}_l(X_i)(1\{Y_i \leq q\} - \hat{m}_{0,l}(X_i,q)) - \theta(\tau)$.

Lemma E.1. *Under the Assumptions 1 and C.1,* $\frac{1}{L}\sum_{l=1}^L \mathbb{P}_l(\hat{\phi}_{i,l}(\hat{q}_\tau) - \phi_i(\hat{q}_\tau)) = o_P(\tilde{\pi}_N^2)$ *for any estimator* $(\hat{\omega}_l(x), \hat{m}_{0,l}(x,q), \hat{m}_{1,l}(x,q))$ *of* $(\omega(x), m_0(x,q), m_1(x,q))$ *and any quantile index* $\tau \in \Upsilon$.

Proof. Using the law of iterated expectations and $m_0(x,q) = \int 1\{y \leq q\} dF_{Y|X=x}(y)$, we have

$$
\begin{aligned}
&\int \left(m_1(x,q) - \omega(x)(1\{y \leq q\} - m_0(x,q))\right) dF_{Y,X}(y,x) \\
&- \int \left(\hat{m}_{1,l}(x,q) - \hat{\omega}_l(x)(1\{y \leq q\} - \hat{m}_{0,l}(x,q))\right) dF_{Y,X}(y,x) \\
&= \int \left(\hat{m}_{1,l}(x,q) - m_1(x,q)\right) dF_X(x) + \int \omega(x)(\hat{m}_{0,l}(x,q) - m_0(x,q)) dF_X(x) \\
&\quad + \int (\hat{\omega}_l(x) - \omega(x))(\hat{m}_{0,l}(x,q) - m_0(x,q)) dF_X(x).
\end{aligned}
$$

The integration by parts implies

$$\begin{aligned}
&\int \omega(x)(\hat{m}_{0,l}(x,q) - m_0(x,q)) f_{X_1|X_{-1}=x_{-1}}(x_1) dx_1 \\
&= -\int \left(\frac{\partial}{\partial x_1}\hat{m}_{0,l}(x,q) - \frac{\partial}{\partial x_1} m_0(x,q) \right) f_{X_1|X_{-1}=x_{-1}}(x_1) dx_1,
\end{aligned}$$

where $(\hat{m}_{0,l}(x,q) - m_0(x,q)) f_{X_1|X_{-1}=x_{-1}}(x_1)$ disappears on the boundary of x_1. Then

$$\begin{aligned}
&\int (m_1(x,q) - \omega(x)(1\{y \le q\} - m_0(x,q)))\, dF_{Y,X}(y,x) \\
&- \int (\hat{m}_{1,l}(x,q) - \hat{\omega}_l(x)(1\{y \le q\} - \hat{m}_{0,l}(x,q)))\, dF_{Y,X}(y,x) \\
&= \int \left(\hat{m}_{1,l}(x,q) - \frac{\partial}{\partial x_1}\hat{m}_{0,l}(x,q) \right) dF_X(x) + \int (\hat{\omega}_l(x) - \omega(x))(\hat{m}_{0,l}(x,q) - m_0(x,q)) dF_X(x).
\end{aligned}$$

Because $\sup_{\tau\in\Upsilon} |\hat{q}_\tau - q_\tau| = o_P(N^{-1/2})$, we have, with probability approaching one,

$$\begin{aligned}
|\mathbb{P}_l(\hat{\phi}_{i,l}(\hat{q}_\tau) - \phi_i(\hat{q}_\tau))| \le& \sup_{q\in\mathcal{Q}^\delta} \left| \int \left(\hat{m}_{1,l}(x,q) - \frac{\partial}{\partial x_1}\hat{m}_{0,l}(x,q) \right) dF_X(x) \right| \\
&+ \sup_{q\in\mathcal{Q}^\delta} \left| \int (\hat{\omega}_l(x) - \omega(x))(\hat{m}_{0,l}(x,q) - m_0(x,q)) dF_X(x) \right| \\
=& o_P(\tilde{\pi}_N^2),
\end{aligned}$$

where the last equality holds due to (C.8) and (C.9). □

Lemma E.2. *Let $\tilde{\eta}_i = 1$ for every $i = 1, \ldots, N$ or if $\tilde{\eta}_i = 1 + \eta_i$ for every $i = 1, \ldots, N$. If the assumptions in Theorem A.4 hold, then*

$$\begin{aligned}
&\sup_{l\in\{1,\ldots,L\}, q\in\mathcal{Q}^\delta} |(\mathbb{P}_{n,l} - \mathbb{P}_l)\tilde{\eta}_i(\hat{\phi}_{i,l}(q) - \phi_i(q))| \\
&= O_P(\pi_N \nu_N^{1/2} N^{-1/2} \log^{1/2}(A_N/\pi_N) + \nu_N N^{-(1+d)/(2+d)} \log(A_N/\pi_N)).
\end{aligned}$$

Proof. Define $\mathcal{M}_l(M)$ the set of $(\tilde{m}_1(x,q), \tilde{m}_0(x,q), \tilde{\omega}(x))$ which satisfies

$$\begin{aligned}
&\{\tilde{m}_j(x,q) : q \in \mathcal{Q}^\delta\} \subset \{\hat{m}_j(x,q) : q \in \mathcal{Q}^\delta\},\ j = 0, 1 \\
&\sup_{q\in\mathcal{Q}^\delta} \int |\tilde{m}_1(x,q) - m_1(x,q)|^2\, dF_X(x) \le M\pi_N, \\
&\int |\tilde{\omega}(x) - \omega(x)|^2\, dF_X(x) \le M\pi_N,
\end{aligned}$$

$$\sup_{q\in\mathcal{Q}^\delta}\int |\tilde{\omega}(x)\tilde{m}_0(x,q)-\omega(x)m_0(x,q)|^2\,dF_X(x)\le M\pi_N,$$

$$\int\left(\sup_{q\in\mathcal{Q}^\delta}|\tilde{m}_1(x,q)|+\sup_{q\in\mathcal{Q}^\delta}|m_1(x,q)|\right)^{2+d}dF_X(x)\le M,\quad\text{and}$$

$$\int\left(\sup_{q\in\mathcal{Q}^\delta}|\tilde{\omega}_l(x)(1+|\tilde{m}_0(x,q)|)|+\sup_{q\in\mathcal{Q}^\delta}|\omega(x)(1+m_{0,l}(x,q))|\right)^{2+d}dF_X(x)\le M.$$

Define

$$F_l(X_i)=|\tilde{\eta}_i|\sup_{q\in\mathcal{Q}^\delta}|\hat{\omega}_l(x)(1+|\hat{m}_{0,l}(x,q)|)|$$
$$+|\tilde{\eta}_i|\sup_{q\in\mathcal{Q}^\delta}|\omega(x)(1+m_{0,l}(x,q))|+|\tilde{\eta}_i|\sup_{q\in\mathcal{Q}^\delta}|\hat{m}_{1,l}(x,q)|+|\tilde{\eta}_i|\sup_{q\in\mathcal{Q}^\delta}|m_1(x,q)|,$$

and

$$\mathcal{F}_l=\left\{\begin{array}{c}\tilde{\eta}_i(\hat{m}_{1,l}(X_i,q)-\hat{\omega}_l(X_i)(1\{Y_i\le q\}-\hat{m}_{0,l}(X_i,q)))\\-\tilde{\eta}_i(m_1(X_i,q)-\omega(X_i)(1\{Y_i\le q\}-m_0(X_i,q)))\end{array}:q\in\mathcal{Q}^\delta\right\}.$$

By Assumption C.1, for any $\delta>0$, we can find a sufficiently large constant $M>0$ such that $(\hat{m}_{1,l},\hat{m}_{0,l},\hat{\omega}_l)\in\mathcal{M}_l(M)$ occurs with probability greater than $1-\delta$. Conditional on $\{(\hat{m}_{1,l},\hat{m}_{0,l},\hat{\omega}_l)\in\mathcal{M}_l(M)\}$ and $\{X_i,Y_i\}_{i\in I_l^c}$, we can treat $\hat{m}_{1,l},\hat{m}_{0,l},\hat{\omega}_l$ as fixed, and $\mathbb{P}_lF_l^{2+d}<\infty$. In addition, by Van der Vaart and Wellner (1996, Theorem 2.7.11) and the fact that $\sup_Q N(\{\hat{m}_j(x,q):q\in\mathcal{Q}^\delta\},\|\cdot\|_{Q,2},\varepsilon\|G_l^{(j)}\|_{Q,2})\lesssim\left(\frac{A_N}{\varepsilon}\right)^{\nu_N}$, we have $\sup_Q N(\mathcal{F}_l,\|\cdot\|_{Q,2},\varepsilon\|F_l\|_{Q,2})\lesssim\left(\frac{A_N}{\varepsilon}\right)^{\nu_N}$. Furthermore, note that

$$\sup_{f\in\mathcal{F}_l}\mathbb{P}_lf^2\le\sup_{q\in\mathcal{Q}^\delta}\mathbb{P}_l\left(|\hat{m}_{1,l}(X,q)-m_1(X,q)|+|\hat{\omega}_l(X)-\omega(X)|+|\omega(X)m_0(X,q)-\hat{\omega}_l(X)\hat{m}_{0,l}(X,q)|\right)^2$$
$$\lesssim\pi_N^2.$$

By Chernozhukov et al. (2014b, Corollary 5.1), we have

$$\mathbb{P}_l\sup_{q\in\mathcal{Q}^\delta}|(\mathbb{P}_{n,l}-\mathbb{P}_l)(\hat{\phi}_{i,l}(q)-\phi_i(q))|\le\mathbb{P}_l|\sup_{f\in\mathcal{F}_l}(\mathbb{P}_{n,l}-\mathbb{P}_l)f|$$
$$\lesssim\sqrt{\frac{\pi_N^2\nu_N}{N}\log\left(\frac{A_N\left(\int F_l(x)^2dF_X(x)\right)^{1/2}}{\pi_N}\right)}$$
$$+\frac{\nu_N\mathbb{E}[(\max_i F_l(X_i))^2]^{1/2}}{N}\log\left(\frac{A_N\left(\int F_l(x)^2dF_X(x)\right)^{1/2}}{\pi_N}\right).$$

Because $\mathbb{E}[F_l(X)^{2+d}]<\infty$, we have $\mathbb{E}[(\max_i F_l(X_i))^2]^{1/2}=O(N^{1/(2+d)})$ on $\{(\hat{m}_{1,l},\hat{m}_{0,l},\hat{\omega}_l)\in$

$\mathcal{M}(\varepsilon, M)\}$.[4] By letting n be sufficiently large, we have

$$\mathbb{P}_l \sup_{q\in\mathcal{Q}^\delta} |(\mathbb{P}_{n,l}-\mathbb{P}_l)(\hat{\phi}_{i,l}(q)-\phi_i(q))| \lesssim \pi_N \nu_N^{1/2} N^{-1/2}\log^{1/2}(A_N/\pi_N) + \nu_N N^{-(1+d)/(2+d)}\log(A_N/\pi_N).$$

This leads to the desired result. □

Lemma E.3. *Under the assumptions in Theorem A.4,* $\sup_{l\in\{1,\ldots,L\},\tau\in\Upsilon}|(\mathbb{P}_{n,l}-\mathbb{P}_l)(\phi_i(\hat{q}_\tau)-\phi_i(q_\tau))| = o_P(N^{-1/2})$ *and* $\sup_{l\in\{1,\ldots,L\},\tau\in\Upsilon}|(\mathbb{P}_{n,l}-\mathbb{P}_l)(\eta_i+1)(\phi_i(\hat{q}^*_\tau)-\phi_i(q_\tau))| = o_P(N^{-1/2})$.

The proof of this lemma is similar to that of Lemma D.5, and thus, is omitted.

E.1.2 Linear Expansion for $\hat{\theta}_{cf}(\tau)$

Taking $\tilde{\eta}_i = 1$ and by (C.1), Lemmas E.1, E.2, and E.3, we have

$$\begin{aligned}
\hat{\theta}_{cf}(\tau)-\theta(\tau) &= \frac{1}{L}\sum_{l=1}^{L}\mathbb{P}_{n,l}\hat{\phi}_{i,l}(\hat{q}_\tau) \\
&= \frac{1}{L}\sum_{l=1}^{L}(\mathbb{P}_{n,l}-\mathbb{P}_l)\phi_i(q_\tau) + \frac{1}{L}\sum_{l=1}^{L}\mathbb{P}_l\phi_i(\hat{q}_\tau) + \frac{1}{L}\sum_{l=1}^{L}(\mathbb{P}_{n,l}-\mathbb{P}_l)(\hat{\phi}_{i,l}(\hat{q}_\tau)-\phi_i(\hat{q}_\tau)) \\
&+ \frac{1}{L}\sum_{l=1}^{L}\mathbb{P}_l(\hat{\phi}_{i,l}(\hat{q}_\tau)-\phi_i(\hat{q}_\tau)) + \frac{1}{L}\sum_{l=1}^{L}(\mathbb{P}_{n,l}-\mathbb{P}_l)(\phi_i(\hat{q}_\tau)-\phi_i(q_\tau)).
\end{aligned}$$

In addition, by $\mathbb{E}m_1(X,q_\tau)=0$ and the usual delta method,

$$\mathbb{P}_l\phi_i(\hat{q}_\tau) = (\mathbb{E}m_1(X,\hat{q}_\tau)-\mathbb{E}m_1(X,q_\tau)) = \frac{\frac{\partial}{\partial q}\mathbb{E}m_1(X,q_\tau)}{f_Y(q_\tau)}\frac{1}{N}\sum_{i=1}^{N}(\tau-1\{Y_i\le q_\tau\}) + o_P(N^{-1/2})$$

where the $o_P(N^{-1/2})$ term holds uniformly over $l=1,\ldots,L$ and $\tau\in\Upsilon$. Therefore, we have,

$$\begin{aligned}
&\sup_{\tau\in\Upsilon}\left|\hat{\theta}(\tau)-\theta(\tau)-\frac{1}{N}\sum_{i=1}^{N}\left(\phi_i(q_\tau)+\frac{\frac{\partial}{\partial q}\mathbb{E}m_1(X,q_\tau)}{f_Y(q_\tau)}(\tau-1\{Y_i\le q_\tau\})\right)\right| \\
&= O_P(\pi_N\nu_N^{1/2}N^{-1/2}\log^{1/2}(A_N/\pi_N)+\nu_N N^{-(1+d)/(2+d)}\log(A_N/\pi_N)+\tilde{\pi}_N^2)+o_P(N^{-1/2}).
\end{aligned} \tag{E.1}$$

[4] If $\{X_i\}$ is sequence of i.i.d. nonnegative random variables with $\mathbb{E}X_i^{2+d}\le M$, then $(\mathbb{E}(\max_{i=1,\ldots,N}X_i)^2)^{1/2} \lesssim N^{\frac{1}{2+d}}$. It is shown as follows. Note that $\mathbb{E}(\max_{i=1,\ldots,N}X_i)^2 = 2\int_0^\infty x\mathbb{P}(\max_{i=1,\ldots,N}X_i > x)dx = 2\int_0^{\alpha_N} x\mathbb{P}(\max_{i=1,\ldots,N}X_i > x)dx + 2\int_{\alpha_N}^\infty x\mathbb{P}(\max_{i=1,\ldots,N}X_i > x)dx \le \alpha_N^2 + 2N\int_{\alpha_N}^\infty \frac{\mathbb{E}X^{2+d}}{X^{1+d}}dx \le \alpha_N^2 + \frac{2MN}{\delta\alpha_N^\delta}$. We can obtain the desired result by taking $\alpha_N = N^{\frac{1}{2+d}}$.

The RHS of the above display is $o_P(N^{-1/2})$ because under the condition in Theorem C.1, $\tilde{\pi}_N^2 = o(N^{-1/2})$ and

$$\pi_N \nu_N^{1/2} N^{-1/2} \log^{1/2}(A_N/\pi_N) + \nu_N N^{-(1+d)/(2+d)} \log(A_N/\pi_N) = o(N^{-1/2}).$$

This leads to the desired result.

E.1.3 Linear Expansion for $\hat{\theta}^*_{cf}(\tau)$

Let $\hat{n}_l = \sum_{i \in I_l}(\eta_i + 1)$. Then

$$\begin{aligned}
\hat{\theta}^*_{cf}(\tau) - \theta(\tau) =& \frac{1}{L}\sum_{l=1}^{L} \frac{n}{\hat{n}_l} \mathbb{P}_{n,l}(\eta_i + 1)\hat{\phi}_{i,l}(\hat{q}^*_\tau) \\
=& \frac{1}{L}\sum_{l=1}^{L} \frac{n}{\hat{n}_l}(\mathbb{P}_{n,l} - \mathbb{P}_l)(\eta_i + 1)\phi_i(\hat{q}^*_\tau) + \frac{1}{L}\sum_{l=1}^{L} \frac{n}{\hat{n}_l}\mathbb{P}_l\hat{\phi}_{i,l}(\hat{q}^*_\tau) + R^*_1(\tau) \\
=& \frac{1}{L}\sum_{l=1}^{L} \frac{n}{\hat{n}_l}(\mathbb{P}_{n,l} - \mathbb{P}_l)(\eta_i + 1)\phi_i(\hat{q}^*_\tau) + \frac{1}{L}\sum_{l=1}^{L} \frac{n}{\hat{n}_l}\mathbb{P}_l\phi_i(\hat{q}^*_\tau) + R^*_2(\tau) \\
=& \frac{1}{L}\sum_{l=1}^{L} \frac{n}{\hat{n}_l}(\mathbb{P}_{n,l} - \mathbb{P}_l)(\eta_i + 1)\phi_i(q_\tau) + \frac{1}{L}\sum_{l=1}^{L} \frac{n}{\hat{n}_l}\mathbb{P}_l\phi_i(\hat{q}^*_\tau) + R^*_3(\tau) \\
=& (\mathbb{P}_N - \mathbb{P})(\eta_i + 1)\phi_i(q_\tau) + \frac{1}{L}\sum_{l=1}^{L} \frac{n}{\hat{n}_l}\mathbb{P}_l\phi_i(\hat{q}^*_\tau) + R^*_4(\tau),
\end{aligned} \tag{E.2}$$

where

$$\sup_{\tau \in \Upsilon} |R_j(\tau)| = O_P(\pi_N \nu_N^{1/2} N^{-1/2} \log^{1/2}(A_N/\pi_N) + \nu_N N^{-(1+d)/(2+d)} \log(A_N/\pi_N) + \tilde{\pi}_N^2) + o_P(N^{-1/2})$$

for $j = 1, \cdots, 4$, the second equality is due to Lemma D.4 and $\mathbb{P}_l\eta_i\hat{\phi}_{i,l}(\hat{q}^*_\tau) = (\mathbb{P}_l\eta_i)(\mathbb{P}_l\hat{\phi}_{i,l}(\hat{q}^*_\tau)) = 0$, the third equality is due to Lemma E.1, the fourth equality is due to Lemma E.3 and the fact that $\sup_{\tau \in \Upsilon} |\hat{q}^*_\tau - q_\tau| = O_P(N^{-1/2})$, and the fifth equality holds because $\sup_{\tau \in \Upsilon} |(\mathbb{P}_{n,l} - \mathbb{P}_l)(\eta_i + 1)\phi_i(q_\tau)| = O_P(N^{-1/2})$ and $\hat{n}_l/n = 1 + o_P(1)$. For the second term on the RHS of (E.2), we have

$$\begin{aligned}
\frac{1}{L}\sum_{l=1}^{L} \frac{n}{\hat{n}_l}\mathbb{P}_l\phi_i(\hat{q}^*_\tau) =& \left(\frac{1}{L}\sum_{l=1}^{L} \frac{n}{\hat{n}_l}\right)(\mathbb{E}m_1(X, \hat{q}^*_\tau) - \mathbb{E}m_1(X, q_\tau)) \\
=& \frac{\frac{\partial}{\partial q}\mathbb{E}m_1(X, q_\tau)}{f_Y(q_\tau)}\left(\sum_{i=1}^{N} \frac{(\eta_i + 1)}{N}(\tau - 1\{Y_i \leq q_\tau\})\right) + o_P(N^{-1/2}),
\end{aligned} \tag{E.3}$$

where the last equality is due to the delta method and (D.12). Combining (E.2) and (E.3), we have

$$\begin{aligned}\hat{\theta}^*_{cf}(\tau) - \theta(\tau) =& \frac{1}{N}\sum_{i=1}^{N}(\eta_i + 1)\Bigg(m_1(X_i, q_\tau) - \theta(\tau) - \omega(X_i)(1\{Y_i \le q_\tau\} - m_0(X_i, q_\tau)) \\ &+ \frac{\frac{\partial}{\partial q}\mathbb{E}m_1(X, q_\tau)}{f_Y(q_\tau)}(\tau - 1\{Y_i \le q_\tau\})\Bigg) + \tilde{R}^*_N(\tau), \end{aligned} \tag{E.4}$$

where $\sup_{\tau\in\Upsilon}|\tilde{R}^*_N(\tau)| = O_P(\pi_N\nu_N^{1/2}N^{-1/2}\log^{1/2}(A_N/\pi_N) + \nu_N N^{-(1+d)/(2+d)}\log(A_N/\pi_N) + \tilde{\pi}_N^2) + o_P(N^{-1/2})$. Taking the difference between (E.1) and (E.4), we have

$$\begin{aligned}\hat{\theta}^*_{cf}(\tau) - \hat{\theta}_{cf}(\tau) =& \frac{1}{N}\sum_{i=1}^{N}\eta_i\Bigg(m_1(X_i, q_\tau) - \theta(\tau) - \omega(X_i)(1\{Y_i \le q_\tau\} - m_0(X_i, q_\tau)) \\ &+ \frac{\frac{\partial}{\partial q}\mathbb{E}m_1(X, q_\tau)}{f_Y(q_\tau)}(\tau - 1\{Y_i \le q_\tau\})\Bigg) + R^*_N(\tau), \end{aligned}$$

where $\sup_{\tau\in\Upsilon}|R^*_N(\tau)| = O_P(\pi_N\nu_N^{1/2}N^{-1/2}\log^{1/2}(A_N/\pi_N) + \nu_N N^{-(1+d)/(2+d)}\log(A_N/\pi_N) + \tilde{\pi}_N^2)+o_P(N^{-1/2})$. Due to the condition in Theorem C.1, we have $\sup_{\tau\in\Upsilon}|R^*_N(\tau)| = o_P(N^{-1/2})$, which is the desired result.

E.2 Proof of Theorem C.2

E.2.1 Linear Expansion for $\widehat{UQPE}(\tau)$

By Theorem A.3 and the condition in Theorem A.4, we have

$$\sup_{\tau\in\Upsilon}\left|\hat{\theta}_{cf}(\tau) - \theta(\tau)\right| = o_P((\log(N)Nh_1)^{-1/2}).$$

Based on (D.10), we have

$$\sup_{\tau\in\Upsilon}|\hat{f}_Y(\hat{q}_\tau) - f_Y(q_\tau)| = O_P(\log^{1/2}(N)(Nh_1)^{-1/2} + h_1^2),$$

$$\sup_{\tau\in\Upsilon}\left|\frac{(\hat{\theta}_{cf}(\tau) - \theta(\tau))(\hat{f}_Y(\hat{q}_\tau) - f_Y(q_\tau))}{\hat{f}_Y(\hat{q}_\tau)f_Y(q_\tau)}\right| = o_P((\log(N)Nh_1)^{-1/2}),$$

and

$$\sup_{\tau\in\Upsilon}\left|\frac{\theta(\tau)(\hat{f}_Y(\hat{q}_\tau) - f_Y(q_\tau))^2}{f_Y^2(q_\tau)\hat{f}_Y(\hat{q}_\tau)}\right| = O_P(\log(N)(Nh_1)^{-1} + h_1^4) = o_P((\log(N)Nh_1)^{-1/2}).$$

Therefore

$$\begin{aligned}\widehat{UQPE}_{cf}(\tau) - UQPE(\tau) =& -\frac{\hat{\theta}_{cf}(\tau)-\theta(\tau)}{f_Y(q_\tau)} + \frac{\theta(\tau)(\hat{f}_Y(\hat{q}_\tau)-f_Y(q_\tau))}{f_Y^2(q_\tau)} \\ &+ \frac{(\hat{\theta}_{cf}(\tau)-\theta(\tau))(\hat{f}_Y(\hat{q}_\tau)-f_Y(q_\tau))}{\hat{f}_Y(\hat{q}_\tau)f_Y(q_\tau)} - \frac{\theta(\tau)(\hat{f}_Y(\hat{q}_\tau)-f_Y(q_\tau))^2}{f_Y^2(q_\tau)\hat{f}_Y(\hat{q}_\tau)} \\ =& \frac{1}{N}\sum_{i=1}^N \mathrm{IF}_i(\tau) + \frac{\theta(\tau)f_Y^{(2)}(q_\tau)(\int u^2K_1(u)du)h_1^2}{2f_Y^2(q_\tau)} + R(\tau), \qquad \text{(E.5)}\end{aligned}$$

where $\sup_{\tau\in\Upsilon}|R(\tau)| = o_P((\log(N)Nh_1)^{-1/2})$.

E.2.2 Linear Expansion for $\widehat{UQPE}^*_{cf}(\tau)$

Recall that

$$\hat{f}_Y^*(\hat{q}_\tau^*) - f_Y(q_\tau) = (\mathbb{P}_N - \mathbb{P})(\eta_i+1)\frac{1}{h_1}K_1\left(\frac{Y_i-q_\tau}{h_1}\right) + \frac{f_Y^{(2)}(q_\tau)(\int u^2K_1(u)du)}{2}h_1^2 + R_5^*(\tau),$$

where $\sup_{\tau\in\Upsilon}|R_5^*(\tau)| = O_P(N^{-1/2})$. Then

$$\begin{aligned}\widehat{UQPE}^*_{cf}(\tau) - UQPE(\tau) =& -\frac{\hat{\theta}^*_{cf}(\tau)-\theta(\tau)}{f_Y(q_\tau)} + \frac{\theta(\tau)(\hat{f}^*_Y(\hat{q}^*_\tau)-f_Y(q_\tau))}{f_Y^2(q_\tau)} \\ &+ \frac{(\hat{\theta}^*_{cf}(\tau)-\theta(\tau))(\hat{f}^*_Y(\hat{q}^*_\tau)-f_Y(q_\tau))}{\hat{f}^*_Y(\hat{q}^*_\tau)f_Y(q_\tau)} - \frac{\theta(\tau)(\hat{f}^*_Y(\hat{q}^*_\tau)-f_Y(q_\tau))^2}{f_Y^2(q_\tau)\hat{f}^*_Y(\hat{q}^*_\tau)} \\ =& \frac{1}{N}\sum_{i=1}^N (1+\eta_i)\mathrm{IF}_i(\tau) + \frac{\theta(\tau)f_Y^{(2)}(q_\tau)(\int u^2K_1(u)du)h_1^2}{2f_Y^2(q_\tau)} + R_6^*(\tau)\end{aligned} \qquad \text{(E.6)}$$

where $\sup_{\tau\in\Upsilon}|R_6^*(\tau)| = o_P((\log(N)Nh_1)^{-1/2})$. Taking difference between (E.5) and (E.6), we have $\widehat{UQPE}^*(\tau) - \widehat{UQPE}(\tau) = \frac{1}{N}\sum_{i=1}^N \eta_i\mathrm{IF}_i(\tau) + R^*(\tau)$, where $\sup_{\tau\in\Upsilon}|R^*(\tau)| = o_P((\log(N)Nh_1)^{-1/2})$.

E.3 Proof of Theorem C.3

We will show (A.1)–(C.9) in Assumption C.1. First, we will show (A.1). To verify the first condition in Assumption C.1, we note that

$$\begin{aligned}&\sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta}\left|\frac{\partial}{\partial q}\hat{m}_{j,l}(x,q)\right| \\ =& \sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta}\left|\int \frac{\breve{m}_{j,l}(x,t)}{h_2^2}K_2^{(1)}\left(\frac{t-q}{h_2}\right)dt\right|\end{aligned}$$

$$
\begin{aligned}
&\leq \sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta}\left|\int \frac{m_j(x,t)}{h_2} dK_2\left(\frac{t-q}{h_2}\right)\right| + \sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta} \frac{|\breve{m}_{j,l}(x,q)-m_j(x,q)|}{h_2}\int d|K_2(u)| \\
&\leq \sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta}\left|\int \frac{\frac{\partial}{\partial q}m_j(x,t)}{h_2} K_2\left(\frac{t-q}{h_2}\right)dt\right| + \sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta} \frac{|\breve{m}_{j,l}(x,q)-m_j(x,q)|}{h_2}\int d|K_2(u)| \\
&\leq \sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta} |\frac{\partial}{\partial q}m_j(x,q)|\int |K_2(u)|du + \sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta} \frac{|\breve{m}_{j,l}(x,q)-m_j(x,q)|}{h_2}\int d|K_2(u)| \\
&<\infty,
\end{aligned}
$$

where the first inequality is due to the triangle inequality, the second equality is due to the integration by parts and the fact that the kernel function $K_2(\cdot)$ vanishes at the boundary, and the last inequality is due to the facts that $\sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta}|\frac{\partial}{\partial q}m_j(x,q)|$ is bounded and that $\sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta}\frac{|\breve{m}_{j,l}(x,q)-m_j(x,q)|}{h_2} = O_P(\rho_N) = o_P(1)$. Given the derivative $\frac{\partial}{\partial q}\hat{m}_{j,l}(x,q)$ is uniformly bounded with probability approaching one, there exists a constant M such that $|\hat{m}_{j,l}(x,q_1)-\hat{m}_{j,l}(x,q_2)| \leq M|q_1-q_2|$. The class of Lipschitz continuous functions is a VC-class with a fixed VC-index. This implies $\mu_N = A_N = 1$.

Second, (A.3) follows from

$$
\begin{aligned}
&\sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta}|\breve{m}_{j,l}(x,q)-\hat{m}_{j,l}(x,q)| \\
&= \sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta}\left|\int \frac{\breve{m}_{j,l}(x,t)-\breve{m}_{j,l}(x,q)}{h_2}K_2\left(\frac{t-q}{h_2}\right)dt\right| \\
&\leq 2\sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta}\int \frac{\sup_{t\in\mathcal{Q}^\delta}|\breve{m}_{j,l}(x,t)-m_j(x,t)|}{h_2}\left|K_2\left(\frac{t-q}{h_2}\right)\right|dt \\
&+ \sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta}\left|\int \frac{m_j(x,t)-m_j(x,q)}{h_2}K_2\left(\frac{t-q}{h_2}\right)dt\right| \\
&\leq 2\sup_{x\in\mathcal{X},t\in\mathcal{Q}^\delta}|\breve{m}_{j,l}(x,t)-m_j(x,t)|\int |K_2(u)|du \\
&+ \sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta}|\frac{\partial^{2k}}{\partial q^{2k}}m_j(x,q)|h_2^{2k}\int u^{2k}|K_2(u)|du \\
&= O_P(h_2\rho_N + h_2^{2k}),
\end{aligned}
$$

where the last inequality holds because of (C.10) and the fact that $\sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta}|\frac{\partial^{2k}}{\partial q^{2k}}m_j(x,q)| < \infty$. Therefore,

$$
\begin{aligned}
&\sup_{q\in\mathcal{Q}^\delta}\int |\hat{m}_{1,l}(x,q)-m_1(x,q)|^2 dF_X(x) \\
&\lesssim \sup_{q\in\mathcal{Q}^\delta}\int |\hat{m}_{1,l}(x,q)-\breve{m}_1(x,q)|^2 dF_X(x) + \sup_{q\in\mathcal{Q}^\delta}\int |\breve{m}_{1,l}(x,q)-m_1(x,q)|^2 dF_X(x)
\end{aligned}
$$

$$=O_P(\rho_N^2 h_2^2 + h_2^{4k}).$$

Third, (C.4) is the same as (C.12).

Fourth, we will show (C.5). Note that

$$|\hat{\omega}_l(x)\hat{m}_{0,l}(x,q) - \omega(x)m_0(x,q)| \leq |\hat{\omega}_l(x) - \omega(x)||\hat{m}_{0,l}(x,q)| + |\omega(x)(\hat{m}_{0,l}(x,q) - m_0(x,q))|.$$

Then

$$\begin{aligned}
&\sup_{q\in\mathcal{Q}^\delta} \int |\hat{\omega}_l(x)\hat{m}_{0,l}(x,q) - \omega(x)m_0(x,q)|^2 dF_X(x) \\
\lesssim &\sup_{q\in\mathcal{Q}^\delta} \int (\hat{\omega}_l(x) - \omega(x))^2 \hat{m}_{0,l}^2(x,q) dF_X(x) + \int \omega^2(x)(\hat{m}_{0,l}(x,q) - m_0(x,q))^2 dF_X(x) \\
\lesssim &\int (\hat{\omega}_l(x) - \omega(x))^2 dF_X(x) + \int \omega^2(x) dF_X(x) \sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta} |\hat{m}_{0,l}(x,q) - m_0(x,q)|^2 \\
= &O_P(\rho_N^2 h_2^2 + h_2^{4k}),
\end{aligned}$$

where the last equality holds due to the fact that

$$\begin{aligned}
&\sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta} |\hat{m}_{0,l}(X,q) - m_0(X,q)| \\
&\leq \sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta} |\hat{m}_{0,l}(X,q) - \breve{m}_{0,l}(X,q)| + \sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta} |\breve{m}_{0,l}(X,q) - m_0(X,q)| \\
&= O_P(h_2\rho_N + h_2^{2k}).
\end{aligned}$$

Fifth, (C.6) holds because $\sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta} |m_j(x,q)|$ is bounded for $j = 0, 1$ and $\sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta} |\breve{m}_{1,l}(x,q) - \hat{m}_{1,l}(x,q)| = o_P(1)$.

Sixth, (C.7) follows from $\sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta} |m_0(x,q)| \leq 1$, $\sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta} |\breve{m}_{0,l}(x,q) - \hat{m}_{0,l}(x,q)| = o_P(1)$, and $\mathbb{E}|\omega(X)|^{2+d} < \infty$.

Seventh, (C.8) holds because

$$\begin{aligned}
&\sup_{q\in\mathcal{Q}^\delta} \left| \int \left(\hat{m}_{1,l}(x,q) - \frac{\partial}{\partial x_1} \hat{m}_{0,l}(x,q) \right) dF_X(x) \right| \\
&\leq \sup_{q\in\mathcal{Q}^\delta} \int \frac{1}{h_2} \sup_{t\in\mathcal{Q}^\delta} \left| \int \left(\breve{m}_{1,l}(x,t) - \frac{\partial}{\partial x_1} \breve{m}_{0,l}(x,t) \right) dF_X(x) \right| \left| K_2\left(\frac{q-t}{h_2}\right) \right| dt \\
&= O_P(\tilde{\pi}_{N,1}^2).
\end{aligned}$$

Eighth, we will show (C.9). Note that

$$
\begin{aligned}
&\left|\int (\hat{\omega}_l(x)-\omega(x))(\hat{m}_{0,l}(x,q)-m_0(x,q))dF_X(x)\right| \\
\leq &\left|\int (\hat{\omega}_l(x)-\omega(x))(\breve{m}_{0,l}(x,q)-m_0(x,q))dF_X(x)\right| \\
&+\int \frac{\left|\int (\hat{\omega}_l(x)-\omega(x))(\breve{m}_{0,l}(x,q)-m_0(x,q))dF_X(x)\right|}{h_2}\left|K_2\left(\frac{t-q}{h_2}\right)\right|dt \\
&+\int \frac{\left|\int (\hat{\omega}_l(x)-\omega(x))(\breve{m}_{0,l}(x,t)-m_0(x,t))dF_X(x)\right|}{h_2}\left|K_2\left(\frac{t-q}{h_2}\right)\right|dt \\
&+\int |\hat{\omega}_l(x)-\omega(x)|\left|\int \frac{m_0(x,t)-m_0(x,q)}{h_2}K_2\left(\frac{t-q}{h_2}\right)dt\right|dF_X(x). \qquad \text{(E.7)}
\end{aligned}
$$

By Assumption C.3, we have

$$
\sup_{q\in\mathcal{Q}^\delta}\left|\int (\hat{\omega}_l(x)-\omega(x))(\breve{m}_{0,l}(x,q)-m_0(x,q))dF_X(x)\right| = O_P(\tilde{\pi}_{N,1}^2).
$$

For the second term on the RHS of (E.7), we have, by (C.16),

$$
\sup_{q\in\mathcal{Q}^\delta}\int \frac{\left|\int (\hat{\omega}_l(x)-\omega(x))(\breve{m}_{0,l}(x,t)-m_0(x,t))dF_X(x)\right|}{h_2}K_2\left(\frac{t-q}{h_2}\right)dt = O_P(\tilde{\pi}_{N,1}^2).
$$

Similarly, we can show the third term is $O_P(\tilde{\pi}_{N,1}^2)$ uniformly over $q\in\mathcal{Q}^\delta$ as well. For the fourth term on the RHS of (E.7), we have

$$
\begin{aligned}
&\int |\hat{\omega}_l(x)-\omega(x)|\left|\int \frac{m_0(x,t)-m_0(x,q)}{h_2}K_2\left(\frac{t-q}{h_2}\right)dt\right|dF_X(x) \\
\leq &\int |\hat{\omega}_l(x)-\omega(x)|\left|\frac{\partial^{2k}}{\partial q^{2k}}m_0(x,\tilde{q})\right|h^{2k}dF_X(x)\int u^{2k}|K_2(u)|du = o_P(N^{-1/2}),
\end{aligned}
$$

where we use the fact that $\sup_{x\in\mathcal{X},q\in\mathcal{Q}^\delta}|\frac{\partial^{2k}}{\partial q^{2k}}m_0(x,q)|<\infty$, $h_2^{2k}=O(N^{\frac{-k}{2k+1}})$, and

$$
\int |\hat{\omega}_l(x)-\omega(x)|\,dF_X(x)\leq\left(\int (\hat{\omega}_l(x)-\omega(x))^2\,dF_X(x)\right)^{1/2} = O_P(h_2\rho_N)=o_P(N^{\frac{-1}{2(2k+1)}}).
$$

F Additional Simulation Studies

In this section, we present additional Monte Carlo simulations to those presented in Section 4 in the main text.

DGP	N	p	$\binom{p}{2}$	τ	True UQPE	Estimates Mean	Bias	RMSE	95% Cover Point	Unif.
1 (i)	500	15	105	0.20	1.00	1.03	0.03	0.16	0.962	0.966
				0.40	1.00	1.02	0.02	0.13	0.956	
				0.60	1.00	1.02	0.02	0.14	0.936	
				0.80	1.00	1.02	0.02	0.17	0.958	
2 (i)	500	15	105	0.20	1.12	1.14	0.02	0.17	0.970	0.964
				0.40	1.03	1.05	0.03	0.13	0.956	
				0.60	0.95	0.97	0.02	0.13	0.948	
				0.80	0.88	0.90	0.02	0.15	0.952	
3 (i)	500	15	105	0.20	1.14	1.17	0.02	0.17	0.966	0.958
				0.40	1.05	1.07	0.02	0.13	0.958	
				0.60	0.97	0.99	0.02	0.13	0.952	
				0.80	0.90	0.92	0.02	0.16	0.944	

Table 8: Monte Carlo simulation results with the interaction terms as well as the powers of x_{-1} in $h(x_{-1})$. The true UQPE is numerically computed. The 95% coverage is uniform over the set $[0.20, 0.80]$.

F.1 Interaction Terms

The dictionaries $b(x)$ and $h(x)$ employed for the simulations presented in the main text include the powers of x, but do not include interactions among the coordinates of x. In this section, we present simulation analysis when $b(x)$ and $h(x)$ include the interactions among x_{-1} as well as the powers of x. We follow the same data generating design as in Section 4 in the main text. While we use $(N, p) = (500, 100)$ in Section 4, we use $(N, p) = (500, 15)$ in the current section. This choice is made because $p = 15$ entails $\binom{p}{2} = 105$ interactions, and the dimensions are therefore comparable with those considered in Section 4 in the main text.

Table 8 summarizes simulation results with the interaction terms of x_{-1} as well as the powers of x included in $b(x)$ and $h(x)$. In comparison with the baseline case without the interaction terms, the results are very similar in terms of the bias, RMSE, and the 95% coverage accuracy.

F.2 Kernel Convolution

We directly use the lasso preliminary estimator (cf. Section 2.2) in the baseline simulation studies presented in Section 4 in the main text. In this appendix section, we present additional simulation analysis based on further applying the kernel convolution method (cf. Section C.3) to the preliminary lasso estimator. We continue to use the same data generating design as in the baseline design presented in Section 5 in the main text for the purpose

DGP	N	p	h_2	τ	True UQPE	Estimates Mean	Bias	RMSE	95% Cover Point	Unif.
1 (i)	500	100	(i)	0.20	1.00	1.03	0.03	0.16	0.950	0.954
				0.40	1.00	1.02	0.02	0.13	0.948	
				0.60	1.00	1.03	0.02	0.14	0.944	
				0.80	1.00	1.00	0.00	0.16	0.950	
	500	100	(ii)	0.20	1.00	1.04	0.04	0.16	0.946	0.956
				0.40	1.00	1.02	0.02	0.13	0.948	
				0.60	1.00	1.02	0.02	0.14	0.948	
				0.80	1.00	1.01	0.01	0.15	0.950	
2 (i)	500	100	(i)	0.20	1.12	1.15	0.03	0.18	0.950	0.956
				0.40	1.03	1.05	0.02	0.13	0.946	
				0.60	0.96	0.98	0.02	0.13	0.952	
				0.80	0.87	0.88	0.01	0.14	0.942	
	500	100	(ii)	0.20	1.12	1.16	0.04	0.18	0.956	0.950
				0.40	1.03	1.05	0.02	0.13	0.954	
				0.60	0.95	0.98	0.03	0.13	0.954	
				0.80	0.87	0.89	0.02	0.14	0.946	
3 (i)	500	100	(i)	0.20	1.15	1.17	0.03	0.18	0.952	0.952
				0.40	1.04	1.06	0.02	0.13	0.946	
				0.60	0.97	1.00	0.03	0.13	0.950	
				0.80	0.91	0.91	0.00	0.14	0.950	
	500	100	(ii)	0.20	1.15	1.18	0.04	0.18	0.954	0.950
				0.40	1.04	1.06	0.02	0.13	0.954	
				0.60	0.97	1.00	0.03	0.13	0.950	
				0.80	0.90	0.92	0.01	0.14	0.954	

Table 9: Monte Carlo simulation results with the kernel convolution. The true UQPE is numerically computed. The 95% coverage is uniform over the set $[0.20, 0.80]$.

of comparisons.

Table 9 summarizes simulation results based on the kernel convolution method applied to the lasso preliminary estimator, with the tuning parameter value given by two alternative rules, (i) $h_2 = 0.1N^{-1/6}$ and (ii) $h_2 = 0.2N^{-1/6}$. Observe that the results are overall very similar to those presented in Table 1 in Section 5 in the main text, in terms of the magnitudes of the bias and RMSE as well as the 95% coverage. This finding suggests that, when the lasso preliminary estimator is used, there do not seem substantially additional benefits of using the kernel convolution method. This is reasonable because of the sufficiently low complexity of the function space of the lasso estimates.

DGP	N	p	τ	True UQPE	Estimates Mean	Bias	RMSE	95% Cover Point	Unif.
1 (i)	500	100	0.20	1.00	1.03	0.03	0.19	0.972	0.986
			0.40	1.00	1.02	0.02	0.15	0.968	
			0.60	1.00	1.02	0.02	0.17	0.970	
			0.80	1.00	0.99	-0.01	0.19	0.956	
2 (i)	500	100	0.20	1.12	1.14	0.02	0.21	0.968	0.984
			0.40	1.03	1.04	0.02	0.16	0.968	
			0.60	0.95	0.98	0.03	0.16	0.976	
			0.80	0.88	0.87	-0.01	0.17	0.952	
3 (i)	500	100	0.20	1.15	1.17	0.02	0.21	0.966	0.986
			0.40	1.04	1.06	0.02	0.16	0.964	
			0.60	0.97	1.00	0.02	0.16	0.972	
			0.80	0.90	0.90	0.00	0.16	0.954	

Table 10: Monte Carlo simulation results with the second-order kernel. The true UQPE is numerically computed. The 95% coverage is uniform over the set $[0.20, 0.80]$.

F.3 Higher-Order Kernel

In Section 4 in the main text, we use a second-order kernel (second-order Epanechnikov kernel) with the undersmoothed rule-of-thumb optimal choice h_1 of the bandwidth to satisfy the assumption for valid inference. Another option is to use the rule-of-thumb optimal choice $h_1 = 1.06\sigma(Y)N^{-1/5}$ without an undersmoothing while using a higher-order-kernel function (e.g., fourth-order Epanechnikov kernel) instead of a second-order kernel function. With this approach, the optimal rate $h_1 \propto N^{-1/5}$ with respect to a second-order kernel is effectively undersmoothing with respect to the fourth-order kernel. This may have a practical advantage in that a researcher can directly use the choice rule $h_1 = 1.06\sigma(Y)N^{-1/5}$ without an *ad hoc* undersmoothing. A disadvantage, on the other hand, is that we require a higher-order of smoothness of the density function. In this section, we demonstrate through simulations that this alternative approach works as well.

Table 10 summarizes simulation results based on the rule-of-thumb optimal choice $h_1 = 1.06\sigma(Y)N^{-1/5}$ along with the fourth-order Epanechnikov kernel function. These simulation results overall indicate accurate estimates as the baseline results presented in Section 4 in the main text albeit slight over-coverages. That said, we emphasize once again that this approach works at the expense of more smoothness assumption.

F.4 Testing $UQPE(\tau) = 0, \forall \tau \in \Upsilon$

While we have thus far studied the finite sample performance of $\widehat{UQPE}(\tau)$ for general purposes of estimation and inference for $UQPE(\tau)$, we now focus on the finite sample performance of $\hat{\theta}(\tau)$. Recall that $\hat{\theta}(\tau)$ and its asymptotic properties are useful for testing $\theta(\tau) = 0, \forall\tau \in \Upsilon$, which in turn is equivalent to the hypothesis $UQPE(\tau) = 0, \forall\tau \in \Upsilon$.

The basic simulation design carries over from Section 4, but the current design differs in the following two points. First, the function $g(\cdot)$ is now defined by $g(x) = 0$ and we refer to it as DGP 0. This design conforms with the null hypothesis $UQPE(\tau) = 0, \forall\tau \in \Upsilon$. Second, for the purpose of evaluating the rate of convergence, we vary the sample size $N \in \{250, 500\}$. Table 11 summarizes the simulation results for $\hat{\theta}$. Observe the the biases are small and the coverage frequencies are close to the nominal probabilities. Furthermore, the convergence rate is consistent with the theoretical prediction of the root-N rate.

(i) The Most Sparse Design – θ

				True	Estimates			95% Cover	
DGP	N	p	τ	θ	Mean	Bias	RMSE	Point	Unif.
0 (i)	250	100	0.20	0.00	0.00	0.00	0.03	0.952	0.938
			0.40	0.00	0.00	0.00	0.04	0.930	
			0.60	0.00	0.00	0.00	0.03	0.938	
			0.80	0.00	0.00	0.00	0.03	0.944	
0 (i)	500	100	0.20	0.00	0.00	0.00	0.02	0.936	0.940
			0.40	0.00	0.00	0.00	0.02	0.930	
			0.60	0.00	-0.01	-0.01	0.03	0.934	
			0.80	0.00	0.00	0.00	0.02	0.932	

(ii) The Second Most Sparse Design – θ

				True	Estimates			95% Cover	
DGP	N	p	τ	θ	Mean	Bias	RMSE	Point	Unif.
0 (ii)	250	100	0.20	0.00	-0.01	-0.01	0.03	0.946	0.940
			0.40	0.00	-0.01	-0.01	0.04	0.916	
			0.60	0.00	-0.01	-0.01	0.04	0.924	
			0.80	0.00	-0.01	-0.01	0.03	0.936	
0 (ii)	500	100	0.20	0.00	0.00	0.00	0.02	0.918	0.918
			0.40	0.00	-0.01	-0.01	0.02	0.924	
			0.60	0.00	-0.01	-0.01	0.02	0.902	
			0.80	0.00	-0.01	-0.01	0.02	0.914	

Table 11: Monte Carlo simulation results for θ under the sparsity designs (i) and (ii). The 95% coverage is uniform over the set $[0.20, 0.80]$.

References

Belloni, A. and V. Chernozhukov (2011): "ℓ_1-penalized quantile regression in high-dimensional sparse models," *The Annals of Statistics*, 39, 82–130.

Belloni, A., V. Chernozhukov, I. Fernández-Val, and C. Hansen (2017): "Program Evaluation with High-dimensional Data," *Econometrica*, 85, 233–298.

Bickel, P. J., Y. Ritov, and A. B. Tsybakov (2009): "Simultaneous Analysis of Lasso and Dantzig Selector," *The Annals of Statistics*, 37, 1705–1732.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018): "Double/debiased machine learning for treatment and structural parameters," *Econometrics Journal*, 21, C1–C68.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014a): "Anti-concentration and Honest, Adaptive Confidence Bands," *The Annals of Statistics*, 42, 1787–1818.

——— (2014b): "Gaussian Approximation of Suprema of Empirical Processes," *The Annals of Statistics*, 42, 1564–1597.

Chernozhukov, V., J. C. Escanciano, H. Ichimura, W. K. Newey, and J. M. Robins (2021a): "Locally Robust Semiparametric Estimation," *Econometrica, forthcoming.*

Chernozhukov, V., I. Fernández-Val, and A. E. Kowalski (2015): "Quantile regression with censoring and endogeneity," *Journal of Econometrics*, 186, 201–221.

Chernozhukov, V., W. K. Newey, and R. Singh (2021b): "Automatic debiased machine learning of causal and structural effects," *Econometrica, forthcoming.*

Farrell, M. H., T. Liang, and S. Misra (2021): "Deep Neural Networks for Estimation and Inference," *Econometrica*, 89, 181–213.

Pollard, D. (1991): "Asymptotics for Least Absolute Deviation Regression Estimators," *Econometric Theory*, 7, 186–199.

Schmidt-Hieber, J. (2020): "Nonparametric regression using deep neural networks with ReLU activation function," *Annals of Statistics*, 48, 1875–1897.

Van der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer.

Wager, S. and S. Athey (2018): "Estimation and inference of heterogeneous treatment effects using random forests," *Journal of the American Statistical Association*, 113, 1228–1242.