Online Supplement to "A Simple Nonparametric Approach for Estimation and Inference of Conditional Quantile Functions"

Zheng Fang[1], Qi Li*[1], and Karen X. Yan[2]

[1]Department of Economics, Texas A&M University
[2]School of Economics, Georgia Institute of Technology

October 29, 2021

This supplement contains:

(i) proofs of Lemmas A.1–Lemma A.10 (Section S1)

(ii) proof of Proposition 5.1 (Section S2)

(iii) additional simulation and empirical results (Section S3).

S1. Proofs of Lemmas A.1–A.10

Proof of Lemma A.1

Proof: (i) and (ii) follow directly from Masry (1996).

(iii) First note that

$$\hat{\sigma}^2_{h_2}(x) = \frac{(nh_2)^{-1}\sum_{i=1}^n [Y_i - \hat{m}_{h_1}(X_i)]^2 K((X_i - x)/h_2)}{\hat{f}_{h_2}(x)}, \tag{S1.1}$$

where

$$\hat{f}_{h_2}(x) = (nh_2)^{-1}\sum_{i=1}^n K((X_i - x)/h_2),$$

and

$$\hat{m}_{h_1}(X_i) = \frac{(nh_1)^{-1}\sum_{j=1}^n Y_i K((X_i - x)/h_1)}{\hat{f}_{h_1}(x)} .$$

*The corresponding author: Qi Li, email: qi-li@tamu.edu (Zheng Fang, email: zfang@tamu.edu; Karen X. Yan, email: xyan306@gatech.edu)

Replacing Y_i in Equation (S1.1) by $m_0(X_i) + \sigma_0(X_i)\epsilon_i$, we get

$$\begin{aligned}\hat{\sigma}^2_{h_2}(x) - \sigma^2_0(x) &= \frac{(nh_2)^{-1}\sum_{i=1}^n [m_0(X_i) - \hat{m}_{h_1}(X_i)]^2 K((X_i - x)/h_2)}{\hat{f}_{h_2}(x)} \\ &\quad - 2\frac{(nh_2)^{-1}\sum_{i=1}^n [\hat{m}_{h_1}(X_i) - m_0(X_i)]\sigma_0(X_i)\epsilon_i K((X_i - x)/h_2)}{\hat{f}_{h_2}(x)} \\ &\quad + \frac{(nh_2)^{-1}\sum_{i=1}^n [\sigma^2_0(X_i)\epsilon^2_i - \sigma^2_0(x)] K((X_i - x)/h_2)}{\hat{f}_{h_2}(x)} \\ &\equiv I_{1n} - 2I_{2n} + I_{3n}, \end{aligned} \tag{S1.2}$$

where the definitions of I_{jn}, $j = 1, 2, 3$, should be apparent.

For $i = 1, \ldots, n$, define

$$\Delta_{n,m,h_1}(X_i) = \hat{m}_{h_1}(X_i) - m_0(X_i) \ .$$

Using Lemma A.1 (i),

$$max_{1 \le i \le n} |\Delta_{n,m,h_1}(X_i)\mathbf{1}_{i,n}| \le \sup_{x \in \mathcal{X}_n} |m_0(x) - \hat{m}_{h_1}(x)| = O_p\left(h_1^2 + \sqrt{\frac{\log n}{nh_1}}\right), \tag{S1.3}$$

where $\mathbf{1}_{i,n} = \mathbf{1}(X_i \in \mathcal{X}_n)$ is the indicator function, and $\mathcal{X}_n$ is a trimmed set. We trimmed off the boundary region of $\mathcal{X}$ to avoid the slow convergence rate at the boundary.

Therefore,

$$\begin{aligned} I_{1n} &= \frac{(nh_2)^{-1}\sum_{i=1}^n [m_0(X_i) - \hat{m}_{h_1}(X_i)]^2 K((X_i - x)/h_2)}{\hat{f}_{h_2}(x)} \\ &\le max_{1 \le i \le n} |\Delta_{n,m,h_1}(X_i)|^2 \mathbf{1}_{i,n} + max_{1 \le i \le n} |\Delta_{n,m,h_1}(X_i)|^2 (1 - \mathbf{1}_{i,n}) \\ &= O_p\left(h_1^4 + \frac{\log n}{nh_1}\right). \end{aligned} \tag{S1.4}$$

The term I_{3n} is the leading term that contributes to the variance of the conditional quantile estimator, and by the results from Masry (1996), uniformly in $x \in \mathcal{X}_n$,

$$I_{3n} = O_p\biggl(h_2^2 + \sqrt{\frac{\log(n)}{nh_2}}\biggr). \tag{S1.5}$$

$$
\begin{aligned}
I_{2n} &= \frac{1}{\hat{f}_{h_2}(x)} \frac{1}{n^2 h_1 h_2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{[Y_j - m_0(X_i)]\sigma_0(X_i)\epsilon_i K((X_j - X_i)/h_1) K((X_i - x)/h_2)}{\hat{f}_{h_1}(X_i)} \\
&= \frac{K(0)}{\hat{f}_{h_1}(x)} \frac{1}{n^2 h_1 h_2} \sum_{i=1}^{n} \frac{\sigma_0^2(X_i)\epsilon_i^2 K((X_i - x)/h_2)}{\hat{f}_{h_2}(X_i)} \\
&+ \quad \frac{1}{\hat{f}_{h_2}(x)} \frac{1}{n^2 h_1 h_2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{[Y_j - m_0(X_i)]\sigma_0(X_i)\epsilon_i K((X_j - X_i)/h_1) K((X_i - x)/h_2)}{\hat{f}_{h_1}(X_i)} \\
&\equiv I_{2n,1} + I_{2n,2}, \qquad \text{(S1.6)}
\end{aligned}
$$

where the first term $I_{2n,1}$ is a partial sum and it is straightforward to see that

$$
I_{2n,1} = o_p((nh_1)^{-1}) \qquad \text{(S1.7)}
$$

uniformly in $x \in \mathcal{X}_n$.

For the second term $I_{2n,2}$, using the identity that

$$
\frac{1}{\hat{f}_{h_1}(X_i)} = \frac{1}{f(X_i)} + \frac{f(X_i) - \hat{f}_{h_1}(X_i)}{f(X_i)\hat{f}_{h_1}(X_i)}, \qquad \text{(S1.8)}
$$

we obtain

$$
\begin{aligned}
I_{2n,2} &= \frac{1}{\hat{f}_{h_2}(x)} \frac{1}{n^2 h_1 h_2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{[Y_j - m_0(X_i)]\sigma_0(X_i)\epsilon_i K((X_j - X_i/h_1) K((X_i - x)/h_2)}{f(X_i)} \frac{f(X_i)}{\hat{f}_{h_1}(X_i)} \\
&\equiv \frac{1}{\hat{f}_{h_2}(x)} \frac{n(n-1)}{n^2} \left[A_{n,x} + B_{n,x}\right], \qquad \text{(S1.9)}
\end{aligned}
$$

where

$$
\begin{aligned}
A_{n,x} &= \frac{1}{n(n-1)h_1 h_2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{[Y_j - m_0(X_i)]\sigma_0(X_i)\epsilon_i K((X_j - X_i)/h_1) K((X_i - x)/h_2)}{f(X_i)}, \\
B_{n,x} &= \frac{1}{n(n-1)h_1 h_2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{[Y_j - m_0(X_i)][f(X_i) - \hat{f}_{h_1}(X_i)]\sigma_0(X_i)\epsilon_i K((X_j - X_i)/h_1) K((X_i - x)/h_2)}{f(X_i)}.
\end{aligned}
$$

We first evaluate $A_{n,x}$. Let $Z_i = (X_i, \epsilon_i)$ and $\mathcal{G}_n$ the class of functions $g_{n,x} : \mathcal{Z}^2 \to \mathbf{R}$ defined by

$$
g_{n,x}(z_1, z_2) = \frac{1}{h_1 h_2} \frac{[m_0(x_2) + \sigma_0(x_2)\epsilon_2 - m_0(x_1)]\sigma_0(x_1)\epsilon_1 K((x_2 - x_1)/h_1) K((x_1 - x)/h_2)}{f(x_1)},
$$

where $z_1 = (x_1, \epsilon_1)$ and $z_2 = (x_2, \epsilon_2)$.

Let P_X and P_ϵ be the marginal laws of X and ϵ respectively, and $P \equiv P_X \times P_\epsilon$. The term $A_{n,x}$ is simply the U-statistic $U_n^2 g_{n,x}$ for each fixed x, and the kernel $g_{n,x}$ is

P^2-centered though not symmetric. Define

$$\mathcal{F}_{1n} = \left\{ K\left(\frac{X_i - x}{h_2}\right) : x \in \mathcal{X}_n \right\} ,$$

and

$$\mathcal{F}_{2n} = \left\{ \frac{1}{h_1 h_2} \frac{[m_0(X_j) + \sigma_0(X_j)\epsilon_j - m_0(X_i)]\sigma_0(X_i)\epsilon_i K((X_j - X_i)/h_1)}{f(X_i)} \right\} .$$

Since $K(\cdot)$ is of bounded variation (because it is differentiable and has bounded derivative), it follows from Ghosal et al. (2000, p.1073) that, for any $\epsilon > 0$,

$$\sup_Q N(\epsilon, \mathcal{F}_{1n}, \|\cdot\|_{Q,2}) \lesssim \epsilon^{-4}, \tag{S1.10}$$

where the supremum is taken over the set of all probability measures. In turn, we have by Lemma A.1 in Ghosal et al. (2000) and $\mathcal{G}_n = \mathcal{F}_{1n}\mathcal{F}_{2n}$ (defined by pointwise product) that

$$\sup_Q N(\epsilon \|G_n\|_{Q,2}, \mathcal{G}_n, \|\cdot\|_{Q,2}) \lesssim \epsilon^{-4}, \tag{S1.11}$$

where G_n is the envelope function for the class of function $\mathcal{G}_n$. Because the kernel function is bounded: $|K(\cdot)| \leq C$ for some finite positive constant C. The envelop function for the class of function $\mathcal{F}_{1n}$ is a positive constant C, hence, $\mathcal{G}_n = \mathcal{F}_{1n}\mathcal{F}_{2n}$ is bounded by

$$G_n(z_i, z_j) \equiv \frac{C}{h_1 h_2} \left| \frac{[m_0(X_j) + \sigma_0(X_j)\epsilon_j - m_0(X_i)]\sigma_0(X_i)\epsilon_i K((X_j - X_i)/h_1)}{f(X_i)} \right| . \tag{S1.12}$$

Therefore, $\mathcal{G}_n$ is Euclidean (Donsker) for each n (Pollard, 1990; Sherman, 1994). Moreover, by simple calculations and exploiting the boundedness of m_0, σ_0 and f, differentiability of f, and the independence between X and ϵ, using Equation (S1.12), we obtain

$$\|G_n\|^2_{P^2,2} \lesssim (h_1 h_2)^{-2} E[K^2((X_j - X_i)/h_1)] \lesssim h_1^{-1} h_2^{-2}. \tag{S1.13}$$

By the Hoeffding decomposition (Sherman, 1994, p.449) and $P^2 g_{n,x} = 0$, we have

$$A_{n,x} \equiv U_n^2 g_{n,x} = \sum_{j=1}^{2} U_n^j \pi_j g_{n,x}, \tag{S1.14}$$

where $\pi_j g_{n,x}$ are the Hoeffding projections of $g_{n,x}$.

Following the Main Corollary and the proof of Corollary 4 in Sherman (1994) and using Equation (S1.13), we obtain, for $j = 1, 2$,

$$E[\sup_{x\in\mathcal{X}_n} |n^{j/2} U_n^j \pi_j g_{n,x}|] \lesssim \|G_n\|_{P^2,2}^{\alpha} \lesssim h_1^{-\alpha/2} h_2^{-\alpha}, \tag{S1.15}$$

where α can be chosen to be in $(0, \frac{1}{3})$.

If we assume $h_{1n} \asymp h_{2n}$ or more generally $h_1^{1-\alpha}/h_2^{2\alpha} \to 0$, then we have that

$$\sup_{x\in\mathcal{X}_n} |A_{n,x}| \equiv \sup_{x\in\mathcal{X}_n} |U_n^2 g_{n,x}| = O_p\left(h_1^{(1-\alpha)/2} h_2^{-\alpha} (nh_1)^{-1/2}\right) = O_p\left(n^{-\frac{1}{2}} h_2^{-\frac{3\alpha}{2}}\right). \tag{S1.16}$$

Turning to the term $B_{n,x}$, using Lemma A.1 (i), (ii), we have

$$\begin{aligned} |B_{n,x}| &\leq \left\{\sup_{x\in\mathcal{X}_n} |\hat{f}_{h_1}(x) - f(x)|\right\}\left\{\sup_{x\in\mathcal{X}_n} |\hat{m}_{h_1}(x) - m_0(x)|\right\} \\ &\quad \sup_{x\in\mathcal{X}_n}\left[\frac{1}{nh_2}\sum_{i=1}^{n} |\sigma_0(X_i)\epsilon_i K((X_i - x)/h_2)|\right] \\ &= O_p\left(h_1^4 + \frac{log(n)}{nh_1}\right). \end{aligned} \tag{S1.17}$$

Equations (S1.9), (S1.16) and (S1.17) imply that

$$I_{2n,2} = O_p\left(n^{-\frac{1}{2}} h_2^{-\frac{3\alpha}{2}}\right) . \tag{S1.18}$$

Combining Equations (S1.6), (S1.7) and (S1.18) yields

$$I_{2n} = O_p\left(n^{-\frac{1}{2}} h_2^{-\frac{3\alpha}{2}}\right) = o_p\left(h_1^2 + \sqrt{\frac{log(n)}{nh_1}}\right). \tag{S1.19}$$

Equations (S1.1), (S1.4), (S1.5), and (S1.19) together imply that

$$\sup_{x\in\mathcal{X}_n} |\hat{\sigma}_{h_2}^2(x) - \sigma_0^2(x)| = O_p\left(h_1^2 + h_2^2 + \sqrt{\frac{\log(n)}{nh_1}} + \sqrt{\frac{\log(n)}{nh_2}}\right). \tag{S1.20}$$

By Assumption 3.3 (ii) and Equation (S1.20), we have that

$$\begin{aligned} \sup_{x\in\mathcal{X}_n} |\hat{\sigma}_{h_2}(x) - \sigma_0(x)| &= \sup_{x\in\mathcal{X}_n} \frac{|\hat{\sigma}_{h_2}^2(x) - \sigma_0^2(x)|}{\hat{\sigma}_{h_2}(x) + \sigma_0(x)} \\ &\leq \frac{\sup_{x\in\mathcal{X}_n} |\hat{\sigma}_{h_2}^2(x) - \sigma_0^2(x)|}{2\inf_{x\in\mathcal{X}_n} \sigma_0(x)} \\ &= O_p\left(h_1^2 + h_2^2 + \sqrt{\frac{\log(n)}{nh_1}} + \sqrt{\frac{\log(n)}{nh_2}}\right). \end{aligned} \tag{S1.21}$$

Proof of Lemma A.2

Proof of (i): Note that

$$\begin{aligned}\hat{m}_{b_1}(x) - m_0(x) &= \frac{(nb_1)^{-1}\sum_{i=1}^n Y_i K((X_i - x)/b_1)}{(nb_1)^{-1}\sum_{i=1}^n K((X_i - x)/b_1)} - m_0(x) \\ &= \frac{(nb_1)^{-1}\sum_{i=1}^n [Y_i - m_0(x)] K((X_i - x)/b_1)}{\hat{f}(x)} \\ &= \frac{M_n}{f(x)} + M_n\left[\frac{1}{\hat{f}_{b_1}(x)} - \frac{1}{f(x)}\right], \end{aligned} \quad \text{(S1.22)}$$

where

$$\begin{aligned} M_n &= \frac{1}{nb_1}\sum_{i=1}^n [Y_i - m_0(x)] K\left(\frac{X_i - x}{b_1}\right) \\ &= \frac{1}{nb_1}\sum_{i=1}^n [m_0(X_i) - m_0(x)] K\left(\frac{X_i - x}{b_1}\right) + \frac{1}{nb_1}\sum_{i=1}^n \sigma_0(X_i)\epsilon_i K\left(\frac{X_i - x}{b_1}\right) \\ &\equiv M_{n,1} + M_{n,2}, \end{aligned} \quad \text{(S1.23)}$$

where the definitions of $M_{n,1}$ and $M_{n,2}$ should be obvious.

It is easy to see that

$$\begin{aligned} E\left[\frac{M_{n,1}}{f(x)}\right] &= \frac{1}{2} b_1^2 \mu_2 \left[2\dot{m}_0(x)\dot{f}(x)/f(x) + \ddot{m}_0(x)\right] + O(b_1^3) \\ &\equiv b_1^2 B_1(x) + O(b_1^3), \end{aligned} \quad \text{(S1.24)}$$

where $\mu_2 = \int u^2 K(u) du$, $B_1(x) = \frac{1}{2}\mu_2\left[2\dot{m}_0(x)\dot{f}(x)/f(x) + \ddot{m}_0(x)\right]$ and

$$Var\left[\frac{M_{n,1}}{f(x)}\right] = O\left(\frac{b_1}{n}\right) = o\left(\frac{1}{n}\right). \quad \text{(S1.25)}$$

Therefore, (S1.24) and (S1.25) imply that

$$\frac{M_{n,1}}{f(x)} = \frac{1}{2} b_1^2 \mu_2 \left[2\dot{m}_0(x)\dot{f}(x)/f(x) + \ddot{m}_0(x)\right] + o_p\left(\frac{1}{\sqrt{n}}\right), \quad \text{(S1.26)}$$

where in the second equality we use the fact that $b_1 = o(n^{-1/5})$.

Combining Equations (S1.23) and (S1.26), we have that

$$\frac{M_n}{f(x)} = b_1^2 B_1(x) + \frac{1}{nb_1}\sum_{i=1}^n \sigma_0(X_i)\epsilon_i K\left(\frac{X_i - x}{b_1}\right) + o_p\left(\frac{1}{\sqrt{n}}\right). \quad \text{(S1.27)}$$

Note that by Lemma A.1 (ii), uniformly in $x \in \mathcal{X}_n$,

$$M_n\left[\frac{1}{\hat{f}_{b_1}(x)} - \frac{1}{f(x)}\right] \quad = \quad O_p\left(b_1^3 + \frac{\log(n)}{nb_1}\right) = o_p\left(\frac{1}{\sqrt{n}}\right). \tag{S1.28}$$

Combining (S1.22), (S1.27) and (S1.28), we derive Lemma A.2 (i).

Proof of Lemma A.2 (ii): Following the same decomposition as in Equation (S1.2) of Lemma A.1 (iii),

$$\begin{aligned}\hat{\sigma}_{b_2}^2(x) - \sigma_0^2(x) &= \frac{(nb_2)^{-1}\sum_{i=1}^n[\sigma_0^2(X_i)\epsilon_i^2 - \sigma_0^2(x)]K((X_i - x)/b_2)}{\hat{f}_{b_2}(x)} + O_p\left(n^{-\frac{1}{2}}b_2^{-\frac{3\alpha}{2}}\right) \\ &\equiv I_{3n} + O_p\left(n^{-\frac{1}{2}}b_2^{-\frac{3\alpha}{2}}\right),\end{aligned} \tag{S1.29}$$

where $\alpha \in (0, \frac{1}{3})$, I_{3n} is defined similar as in Lemma A.1 (iii) and should be apparent.

Next we examine the term I_{3n}. Using the identity $\frac{1}{\hat{f}_{b_2}(x)} = \frac{1}{f(x)} + [\frac{1}{\hat{f}_{b_2}(x)} - \frac{1}{f(x)}]$,

$$\begin{aligned}I_{3n} &= \frac{1}{nb_2\hat{f}_{b_2}(x)}\sum_{i=1}^n[\sigma_0^2(X_i)\epsilon_i^2 - \sigma_0^2(x)]K\left(\frac{X_i - x}{b_2}\right) \\ &= \frac{A_n}{f(x)} + A_n\left[\frac{1}{\hat{f}_{b_2}(x)} - \frac{1}{f(x)}\right],\end{aligned} \tag{S1.30}$$

where

$$\begin{aligned}A_n &= \frac{1}{nb_2}\sum_{i=1}^n[\sigma_0^2(X_i)\epsilon_i^2 - \sigma_0^2(x)]K\left(\frac{X_i - x}{b_2}\right) \\ &= \frac{1}{nb_2}\sum_{i=1}^n\sigma_0^2(X_i)(\epsilon_i^2 - 1)K\left(\frac{X_i - x}{b_2}\right) + \frac{1}{nb_2}\sum_{i=1}^n[\sigma_0^2(X_i) - \sigma_0^2(x)]K\left(\frac{X_i - x}{b_2}\right) \\ &= A_{n,1} + A_{n,2},\end{aligned} \tag{S1.31}$$

where $A_{n,1}$ and $A_{n,2}$ should be apparent.

It is straightforward to see that

$$E\left[\frac{A_{n,2}}{f(x)}\right] = \frac{1}{2}b_2^2\mu_2\left[2\dot{\delta}_2(x)\dot{f}(x)/f(x) + \ddot{\delta}_2(x)\right] + O(b_2^3), \tag{S1.32}$$

where $\mu_2 = \int u^2K(u)du$, $\delta_2(x) = \sigma_0^2(x)$, and

$$var\left[\frac{A_{n,2}}{f(x)}\right] = O\left(\frac{b_2}{n}\right) = o\left(\frac{1}{n}\right). \tag{S1.33}$$

Equations (S1.32) and (S1.33) imply that

$$\frac{A_{n,2}}{f(x)} = \frac{1}{2}b_2^2\mu_2\Big[2\dot{\delta}_2(x)\dot{f}(x)/f(x)+\ddot{\delta}_2(x)\Big]+o_p\Big(\frac{1}{\sqrt{n}}\Big), \qquad \text{(S1.34)}$$

where in the second equality we used the fact that $b_2 = o(n^{-1/5})$.

By Lemma A.1 (i), (ii), uniformly in $x \in \mathcal{X}_n$,

$$A_n\Big[\frac{1}{\hat{f}_{b_2}(x)}-\frac{1}{f(x)}\Big] = O_p\Big(b_2^3+\frac{log(n)}{nb_2}\Big) = o_p\Big(\frac{1}{\sqrt{n}}\Big). \qquad \text{(S1.35)}$$

Equations (S1.31) and (S1.34) imply that

$$\frac{A_n}{f(x)} = \frac{1}{nb_2 f(x)}\sum_{i=1}^n \sigma_0^2(X_i)(\epsilon_i^2-1)K\Big(\frac{X_i-x}{b_2}\Big)+\frac{1}{2}b_2^2\mu_2\Big[2\dot{\delta}_2(x)\dot{f}(x)/f(x)+\ddot{\delta}_2(x)\Big]+o_p\Big(\frac{1}{\sqrt{n}}\Big). \qquad \text{(S1.36)}$$

Combining (S1.29), (S1.35) and (S1.36), we have that

$$\begin{aligned}\hat{\sigma}_{b_2}^2(x)-\sigma_0^2(x) &= \frac{b_2^2\mu_2}{2}\{2\dot{\delta}_2(x)\dot{f}(x)/f(x)+\ddot{\delta}_2(x)\} \\ &\quad +\frac{1}{nb_2}\sum_{i=1}^n \sigma_0^2(X_i)(\epsilon_i^2-1)K\Big(\frac{X_i-x}{b_2}\Big)+o_p\Big(\frac{1}{\sqrt{nb_2}}\Big). \qquad \text{(S1.37)}\end{aligned}$$

Using the notation $A_n \approx B_n$ meaning that $A_n = B_n(1+o_p(1))$, we have that

$$\begin{aligned}\hat{\sigma}_{b_2}(x)-\sigma_0(x) &= \frac{\hat{\sigma}_{b_2}^2(x)-\sigma_0^2(x)}{\hat{\sigma}_{b_2}(x)+\sigma_0(x)} \\ &\approx \frac{\hat{\sigma}_{b_2}^2(x)-\sigma_0^2(x)}{2\sigma_0(x)} \\ &= \frac{b_2^2\mu_2}{4\sigma_0(x)}\{2\dot{\delta}_2(x)\dot{f}(x)/f(x)+\ddot{\delta}_2(x)\} \\ &\quad +\frac{1}{2nb_2 f(x)\sigma_0(x)}\sum_{i=1}^n \sigma_0^2(X_i)(\epsilon_i^2-1)K\Big(\frac{X_i-x}{b_2}\Big)+o_p\Big(\frac{1}{\sqrt{nb_2}}\Big) \\ &= b_2^2 B_2(x)+\frac{1}{2nb_2 f(x)\sigma_0(x)}\sum_{i=1}^n \sigma_0^2(X_i)(\epsilon_i^2-1)K\Big(\frac{X_i-x}{b_2}\Big)+o_p\Big(\frac{1}{\sqrt{nb_2}}\Big), \qquad \text{(S1.38)}\end{aligned}$$

where $B_2(x) = \frac{\mu_2}{4\sigma_0(x)}\big\{2\dot{\delta}_2(x)\dot{f}(x)/f(x)+\ddot{\delta}_2(x)\big\}$.

Proof of Lemma A.3 (i):

Our proof is similar to the proof of Proposition 1 of Akritas and Van Keilegom (2001). Define $m_f(x) = m_0(x)f(x)$, $\widehat{m}_f(x) = \widehat{m}_0(x)\widehat{f}(x) = (nh)^{-1}\sum_{i=1}^n Y_i K_{x,i}$, where

$K_{x,i} = K((x - X_i)/h)$. From $m_0(x) = m_f(x)/f(x)$ and $\widehat{m}_h(x) = \widehat{m}_f(x)/\widehat{f}(x)$, using a similar decomposition as in the proof of Proposition 1 in Akritas and Van Keilegom (2001), we obtain the following decomposition of $\dot{\widehat{m}}_h(x) - \dot{m}_0(x)$:

$$\begin{aligned}
\dot{\widehat{m}}_h(x) - \dot{m}_0(x) &= \frac{\dot{\widehat{m}}_f(x) - \dot{m}_f(x)}{\widehat{f}(x)} - \frac{\dot{m}_f(x)[\widehat{f}(x) - f(x)]}{f(x)\widehat{f}(x)} \\
&\quad - \frac{\widehat{m}_f(x)[\dot{\widehat{f}}(x) - \dot{f}(x)]}{\widehat{f}^2(x)} + \frac{m_f(x)\dot{f}(x)[\widehat{f}^2(x) - f^2(x)]}{f^2(x)\widehat{f}^2(x)} \\
&\quad - \frac{[\widehat{m}_f(x) - m_f(x)]\dot{f}(x)}{f^2(x)}. \qquad \text{(S1.39)}
\end{aligned}$$

We will only prove that the first term is $O_p(c_n)$ uniformly in $x \in \mathcal{X}_n$ as the proofs for other terms are similar. Because $\sup_x |1/\hat{f}(x)| = O_p(1)$ ($\sup_x \equiv \sup_{x \in \mathcal{X}_n}$), we only need to consider the numerator of the first term. Let $\mathbb{X} = \{X_i\}_{i=1}^n$, we write

$$\begin{aligned}
\dot{\widehat{m}}_f(x) - \dot{m}_f(x) &= \{\dot{\widehat{m}}_f(x) - E[\dot{\widehat{m}}_f(x)|\mathbb{X}]\} + \{E[\dot{\widehat{m}}_f(x)|\mathbb{X}] - E[\dot{\widehat{m}}_f(x)]\} + \{E[\dot{\widehat{m}}_f(x)] - \dot{m}_f(x)\} \\
&\equiv \gamma_1(x) + \gamma_2(x) + \gamma_3(x).
\end{aligned}$$

Let $\dot{K}_{x,i} = [dK(v)/dv]|_{v=(x-X_i)/h}$, we have

$$\begin{aligned}
\gamma_1 &= \frac{1}{nh^2}\sum_{i=1}^n (Y_i - m_0(X_i))\dot{K}_{x,i} \\
&= \frac{1}{nh^2}\sum_{i=1}^n \sigma_0(X_i)\epsilon_i \dot{K}_{x,i} \\
&= O_p(c_n)
\end{aligned}$$

uniformly over $x \in \mathcal{X}_n$ by Theorem 2 of Einmahl and Mason (2005), where c_n is defined in Equation (A.1). Similarly, one can show that $\gamma_2 = O_p(c_n)$ and $\gamma_3 = O_p(c_n)$, uniformly in $x \in \mathcal{X}_n$.

Proof of Lemma A.3 (ii): We first introduce some short-hand notation. Let $m_i = m_0(X_i)$, $\sigma_i = \sigma_0(X_i)$, $\sigma_x = \sigma_0(x)$, $\widehat{m}_i = \widehat{m}_h(X_i)$, $\widehat{\sigma}_i^2 = \widehat{\sigma}_h^2(X_i)$, $\widehat{\sigma}_x^2 = \widehat{\sigma}_0^2(x)$, $\widehat{\sigma}_x = \sqrt{\widehat{\sigma}_x^2}$, $u_i = Y_i - m_i = \sigma_i\epsilon_i$, $\widehat{u}_i = Y_i - \widehat{m}_i$, $f_i = f(X_i)$, $\widehat{f}_i = \widehat{f}(X_i)$. From $\widehat{\sigma}_x = \sqrt{\widehat{\sigma}_x^2} = \sqrt{\widehat{A}_x/\widehat{f}_x}$, we get

$$\begin{aligned}
\dot{\widehat{\sigma}}_x &= \frac{1}{2\widehat{\sigma}_x}\dot{\widehat{\sigma}}_x^2 \\
&= \frac{1}{2\sigma_x}\dot{\widehat{\sigma}}_x^2 + \frac{1}{2}\left(\frac{1}{\widehat{\sigma}_x} - \frac{1}{\sigma_x}\right)\dot{\widehat{\sigma}}_x^2 \\
&= \frac{1}{2\sigma_x}\dot{\widehat{\sigma}}_x^2 + O_p(c_n),
\end{aligned}$$

because $\sup_x |\widehat{\sigma}_x^{-1} - \sigma_x^{-1}| \le \sup_x |(\widehat{\sigma}_x - \sigma_x)|/\inf_x(\hat{\sigma}_x\sigma_x) = O_p(c_n)$ by Lemmas A.1 and

A.2, and it is easy to show that $\sup_x |\dot{\widehat{\sigma}}_x^2| = O_p(1)$.

Define $A_x = \sigma_x^2 f_x$, $\widehat{A}_x = \widehat{\sigma}_x^2 \widehat{f}_x$, from $\widehat{u}_i = Y_i - \widehat{m}_i = (m_i - \widehat{m}_i) + u_i$, we obtain $\widehat{u}_i^2 = u_i^2 + 2u_i(m_i - \widehat{m}_i) + (m_i - \widehat{m}_i)^2$. Then

$$
\begin{aligned}
\widehat{A}_x &= \frac{1}{nh}\sum_{i=1}^{n} \widehat{u}_i^2 K_{x,i} \\
&= \frac{1}{nh}\sum_{i=1}^{n} [u_i^2 + 2u_i(m_i - \widehat{m}_i) + (m_i - \widehat{m}_i)^2] K_{x,i} \\
&= \widehat{A}_{1,x} + 2\widehat{A}_{2,x} + \widehat{A}_{3,x}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\dot{\widehat{A}}_x &= \frac{1}{nh}\sum_{i=1}^{n} \widehat{u}_i^2 \dot{K}_{x,i} \\
&= \frac{1}{nh}\sum_{i=1}^{n} [u_i^2 + 2u_i(m_i - \widehat{m}_i) + (m_i - \widehat{m}_i)^2] \dot{K}_{x,i} \\
&= \dot{\widehat{A}}_{1,x} + 2\dot{\widehat{A}}_{2,x} + \dot{\widehat{A}}_{3,x}.
\end{aligned}
$$

By Lemma A.1 that $\sup_x |\widehat{m}_h(x) - m_0(x)| = O_p(h_1^2 + (log(n)/(nh_1))^{1/2})$, it is easy to show that $\dot{\widehat{A}}_{2,x} = O_p(c_n)$ and $\dot{\widehat{A}}_{3,x} = O_p(c_n)$ uniformly in $x \in \mathcal{X}_n$.

Therefore, we only need to consider $\widehat{A}_{1,x}$. We will use the notation: $A_n \approx B_n$ to stand for $A_n = B_n + O_p(c_n)$, and $A_n \asymp B_n$ to mean that A_n and B_n have exactly the same probability order. Using a similar decomposition as in (S1.39), we have:

$$
\begin{aligned}
\dot{\widehat{\sigma}}_x - \dot{\sigma}_x &\approx \frac{1}{2\sigma_x}\left[\dot{\widehat{\sigma}}_x^2 - \dot{\sigma}_x^2\right] \\
&\asymp \dot{\widehat{\sigma}}_x^2 - \dot{\sigma}_x^2 \\
&= \frac{\dot{\widehat{A}}_{1,x} - \dot{A}_{1,x}}{\widehat{f}_x} - \frac{\dot{A}_{1,x}[\widehat{f}_x - f_x]}{f_x \widehat{f}_x} - \frac{\widehat{A}_{1,x}[\dot{\widehat{f}}_x - \dot{f}_x]}{\widehat{f}_x^2} \\
&\quad + \frac{A_{1,x}\dot{f}(x)[\widehat{f}_x^2 - f_x^2]}{f_x^2 \widehat{f}_x^2} - \frac{[\widehat{A}_{1,x} - A_{1,x}]\dot{f}_x}{f_x^2}.
\end{aligned}
$$

We consider the first term of the above decomposition as other terms can be similarly analyzed. Because $\sup_x |1/\hat{f}_x| = O_p(1)$, we only need to consider the numerator.

$$
\begin{aligned}
\dot{\widehat{A}}_{1,x} - \dot{A}_{1,x} &= \{\dot{\widehat{A}}_{1,x} - E[\dot{A}_{1,x}|\mathbb{X}]\} + \{E[\dot{A}_{1,x}|\mathbb{X}] - E[\dot{A}_{1,x}]\} + \{E[\dot{A}_{1,x}] - \dot{A}_{1,x}\} \\
&= \eta_1(x) + \eta_2(x) + \eta_3(x),
\end{aligned}
$$

We only analyze $\eta_1(x)$ as the analysis for other two terms are similar.

$$\eta_1(x) \quad = \quad \frac{1}{nh}\sum_i [u_i^2 - \sigma_i^2]\dot{K}_{x,i} = O_p(c_n)$$

by Theorem 2 of Finmahl and Mason's (2005), where c_n is defined in Equation (A.1).

Proof of Lemma A.4

By using the results of Lemma A.3 and a similar argument as in the proof of Proposition 5 of Akritas and Van Keilegom (2001), one can prove Lemma A.4. Therefore, the proof is omitted.

Proof of Lemma A.5 (i)

By the definition of bootstrap construction and using $Y_i^* = \hat{m}_{h_1}(X_i) + \hat{\sigma}_{h_2}(X_i)\epsilon_i^*$, we have

$$\hat{m}^*_{h_1^*}(x) - \hat{m}_{h_1}(x) = \frac{\sum_{i=1}^n \hat{\sigma}_{h_2}(X_i)\epsilon_i^* K(\frac{X_i - x}{h_1^*})}{\sum_{i=1}^n K(\frac{X_i - x}{h_1^*})} + \left[\frac{\sum_{i=1}^n \hat{m}_{h_1}(X_i) K(\frac{X_i - x}{h_1^*})}{\sum_{i=1}^n K(\frac{X_i - x}{h_1^*})} - \hat{m}_{h_1}(x)\right]. \tag{S1.40}$$

Below we deal with the two terms on the right side of Equation (S1.40) separately. For $i = 1, \ldots, n$, define

$$\Delta_{n,\sigma,h_2}(X_i) = \hat{\sigma}_{h_2}(X_i) - \sigma_0(X_i) \ .$$

First, note that by Lemma A.2 (ii), uniformly in $x \in \mathcal{X}_n$,

$$\max_{1 \le i \le n} |\Delta_{n,\sigma,h_2}(X_i)\mathbf{1}_{i,n}| \le \sup_{x \in \mathcal{X}_n} |\hat{\sigma}_{h_2}(x) - \sigma_0(x)| = O_{P_Z}\left(h_2^2 + \sqrt{\frac{\log(n)}{nh_2}}\right), \tag{S1.41}$$

where $\mathbf{1}_{i,n} = \mathbf{1}(X_i \in \mathcal{X}_n)$ is the indicator function, and $\mathcal{X}_n$ is a trimmed set. We trimmed off the boundary region of $\mathcal{X}$ to avoid the slow convergence rate at the boundary.

Using $\hat{\sigma}_{h_2}(X_i) = \sigma_0(X_i) + \Delta_{n,\sigma,h_2}(X_i)$, the first term on the right side of Equation

(S1.40) can be decomposed into two terms:

$$\frac{\sum_{i=1}^n \hat{\sigma}_{h_2}(X_i)\epsilon_i^* K(\frac{X_i-x}{h_1^*})}{\sum_{i=1}^n K(\frac{X_i-x}{h_1^*})} = \underbrace{\frac{\sum_{i=1}^n \sigma_0(X_i)\epsilon_i^* K(\frac{X_i-x}{h_1^*})}{\sum_{i=1}^n K(\frac{X_i-x}{h_1^*})}}_{A_n} + \underbrace{\frac{\sum_{i=1}^n \Delta_{n,\sigma,h_2}(X_i)\epsilon_i^* K(\frac{X_i-x}{h_1^*})}{\sum_{i=1}^n K(\frac{X_i-x}{h_1^*})}}_{B_n}. \tag{S1.42}$$

For the term B_n, we have: in P_Z probability,

$$\begin{aligned}
\sup_{x\in\mathcal{X}_n} |B_n| &\leq \sup_{x\in\mathcal{X}_n} \frac{\sum_{i=1}^n \left|\Delta_{n,\sigma,h_2}(X_i)\epsilon_i^*\right| K(\frac{X_i-x}{h_1^*})}{\sum_{i=1}^n K(\frac{X_i-x}{h_1^*})} \\
&\leq \max_{1\leq i\leq n} |\Delta_{n,\sigma,h_2}(X_i)| \sup_{x\in\mathcal{X}_n} \frac{\sum_{i=1}^n \left|\epsilon_i^*\right| K(\frac{X_i-x}{h_1^*})}{\sum_{i=1}^n K(\frac{X_i-x}{h_1^*})} \\
&= \max_{1\leq i\leq n} |\Delta_{n,\sigma,h_2}(X_i)| \mathbf{1}_{i,n} \sup_{x\in\mathcal{X}_n} \frac{\sum_{i=1}^n \left|\epsilon_i^*\right| K(\frac{X_i-x}{h_1^*})}{\sum_{i=1}^n K(\frac{X_i-x}{h_1^*})} \\
&\quad + \max_{1\leq i\leq n} |\Delta_{n,\sigma,h_2}(X_i)| (1-\mathbf{1}_{i,n}) \sup_{x\in\mathcal{X}_n} \frac{\sum_{i=1}^n \left|\epsilon_i^*\right| K(\frac{X_i-x}{h_1^*})}{\sum_{i=1}^n K(\frac{X_i-x}{h_1^*})} \\
&= O_{P_W}\left(h_2^2 + \sqrt{\frac{\log(n)}{nh_2}}\right) = o_{P_W}\left(\frac{1}{\sqrt{nh_2^*}}\right).
\end{aligned} \tag{S1.43}$$

Turning to the A_n term, it is easy to show that

$$\begin{aligned}
A_n &= \frac{1}{nh_1^*} \frac{\sum_{i=1}^n \sigma_0(X_i)\epsilon_i^* K(\frac{X_i-x}{h_1^*})}{f(x)} \frac{f(x)}{\hat{f}_{h_1^*}(x)} \\
&= \frac{1}{nh_1^*} \frac{\sum_{i=1}^n \sigma_0(X_i)\epsilon_i^* K(\frac{X_i-x}{h_1^*})}{f(x)} + \frac{1}{nh_1^*} \frac{\sum_{i=1}^n \sigma_0(X_i)\epsilon_i^* K(\frac{X_i-x}{h_1^*})}{f(x)} \left[\frac{f(x)}{\hat{f}_{h_1^*}(x)} - 1\right].
\end{aligned} \tag{S1.44}$$

By the same argument as in Masry (1996), we have that: in P_Z probability,

$$\frac{1}{nh_1^*} \frac{\sum_{i=1}^n \sigma_0(X_i)\epsilon_i^* K\left(\frac{X_i-x}{h_1^*}\right)}{f(x)} = O_{P_W}\left(\sqrt{\frac{\log(n)}{nh_1^*}}\right). \tag{S1.45}$$

Also, by Lemma A.1, uniformly in $x \in \mathcal{X}_n$,

$$\frac{f(x)}{\hat{f}_{h_1^*}(x)} - 1 = O_{P_Z}\left(h_1^{*2} + \sqrt{\frac{\log(n)}{nh_1^*}}\right). \tag{S1.46}$$

Combining results (S1.44), (S1.45) and (S1.46), together with Lemma 3-(73) in Cheng and Huang (2010), we have: uniformly in $x \in \mathcal{X}_n$ and in P_Z probability,

$$A_n = \frac{1}{nh_1^*}\frac{\sum_{i=1}^n \sigma_0(X_i)\epsilon_i^* K(\frac{X_i-x}{h_1^*})}{f(x)} + o_{P_W}\left(\frac{1}{\sqrt{nh_1^*}}\right). \tag{S1.47}$$

Equations (S1.42), (S1.43) and (S1.47) imply that

$$\frac{\sum_{i=1}^n \hat{\sigma}_{h_2}(X_i)\epsilon_i^* K(\frac{X_i-x}{h_1^*})}{\sum_{i=1}^n K(\frac{X_i-x}{h_1^*})} = \frac{1}{nh_1^*}\frac{\sum_{i=1}^n \sigma_0(X_i)\epsilon_i^* K(\frac{X_i-x}{h_1^*})}{f(x)} + o_{P_W}\left(\frac{1}{\sqrt{nh_1^*}} + \frac{1}{\sqrt{nh_2^*}}\right). \tag{S1.48}$$

Finally, let us look at the second term on the right side of (S1.40). We rewrite this term as

$$\frac{\sum_{i=1}^n \hat{m}_{h_1}(X_i) K(\frac{X_i-x}{h_1^*})}{\sum_{i=1}^n K(\frac{X_i-x}{h_1^*})} - \hat{m}_{h_1}(x) = \frac{\frac{1}{nh_1^*}\sum_{i=1}^n [\hat{m}_{h_1}(X_i) - \hat{m}_{h_1}(x)] K\left(\frac{X_i-x}{h_1^*}\right)}{\hat{f}_{h_1^*}(x)}. \tag{S1.49}$$

The numerator of the right side of (S1.49) can be further rewritten as

$$\begin{aligned} &\frac{1}{nh_1^*}\sum_{i=1}^n [\hat{m}_{h_1}(X_i) - \hat{m}_{h_1}(x)] K\left(\frac{X_i-x}{h_1^*}\right) \\ = \;& \underbrace{\frac{1}{nh_1^*}\sum_{i=1}^n [\hat{m}_{h_1}(X_i) - m_0(X_i)] K\left(\frac{X_i-x}{h_1^*}\right)}_{C_n(x)} + \underbrace{\frac{1}{nh_1^*}\sum_{i=1}^n [m_0(X_i) - m_0(x)] K\left(\frac{X_i-x}{h_1^*}\right)}_{D_n(x)} \\ & \underbrace{-\frac{1}{nh_1^*}\sum_{i=1}^n [\hat{m}_{h_1}(x) - m_0(x)] K\left(\frac{X_i-x}{h_1^*}\right)}_{E_n(x)}. \end{aligned} \tag{S1.50}$$

Following the similar proof as in Lemma A.2 (i), uniformly in $x \in \mathcal{X}_n$,

$$\frac{D_n(x)}{\hat{f}_{h_1^*}(x)} = h_1^{*2} B_1(x) + o_{P_Z}\left(\frac{1}{\sqrt{n}}\right), \tag{S1.51}$$

where $B_1(x) = \frac{1}{2}\mu_2\Big[2\dot{m}_0(x)\dot{f}(x)/f(x) + \ddot{m}_0(x)\Big]$, and $\mu_2 = \int u^2 K(u) du$.

Next, we examine the $E_n(x)$ term. By Lemma A.2 (i), uniformly in $x \in \mathcal{X}_n$,

$$\hat{m}_{h_1}(x) - m_0(x) = h_1^2 B_1(x) + \frac{1}{nh_1}\sum_{i=1}^n \sigma_0(X_i)\epsilon_i K\left(\frac{X_i-x}{h_1}\right) + o_{P_Z}\left(\frac{1}{\sqrt{n}}\right). \tag{S1.52}$$

By Lemma A.1, uniformly in $x \in \mathcal{X}_n$,

$$\frac{1}{nh_1^*}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_1^*}\right) = f(x) + o_{P_Z}\left(h_1^{*2} + \sqrt{\frac{\log(n)}{nh_1^*}}\right). \tag{S1.53}$$

Combining (S1.46), (S1.52), (S1.53) and Assumption 4.1, we have uniformly in $x \in \mathcal{X}_n$,

$$\begin{aligned}
\frac{E_n(x)}{\hat{f}_{h_1^*}(x)} &= \frac{E_n}{f(x)} + E_n\left[\frac{f(x)}{\hat{f}_{h_1^*}(x)} - 1\right] \\
&= h_1^2 B_1(x) + \frac{1}{nh_1}\sum_{i=1}^{n} \sigma_0(X_i)\epsilon_i K\left(\frac{X_i - x}{h_1}\right) + o_{P_Z}\left(h_1^{*2}h_1^2 + h_1^{*2}\sqrt{\frac{\log(n)}{n^2 h_1^*}}\right) \\
&= h_1^2 B_1(x) + \frac{1}{nh_1}\sum_{i=1}^{n} \sigma_0(X_i)\epsilon_i K\left(\frac{X_i - x}{h_1}\right) + o_{P_Z}\left(\frac{1}{\sqrt{nh_1^*}}\right) \\
&= o_{P_Z}\left(\frac{1}{\sqrt{nh_1^*}}\right).
\end{aligned} \tag{S1.54}$$

Finally, we examine $C_n(x)$. We have:

$$\begin{aligned}
C_n(x) &= \frac{1}{nh_1^*}\sum_{i=1}^{n}\left[\frac{\frac{1}{nh_1}\sum_{j=1}^{n} Y_j K\left(\frac{X_j - X_i}{h_1}\right)}{\frac{1}{nh_1}\sum_{j=1}^{n} K\left(\frac{X_j - X_i}{h_1}\right)} - m_0(X_i)\right] K\left(\frac{X_i - x}{h_1^*}\right) \\
&= \frac{1}{nh_1^*}\sum_{i=1}^{n}\frac{\frac{1}{nh_1}\sum_{j=1}^{n}[Y_j - m_0(X_i)] K\left(\frac{X_j - X_i}{h_1}\right)}{\frac{1}{nh_1}\sum_{j=1}^{n} K\left(\frac{X_j - X_i}{h_1}\right)} K\left(\frac{X_i - x}{h_1^*}\right) \\
&= \frac{1}{n^2 h_1^* h_1}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{[Y_j - m_0(X_i)] K\left(\frac{X_j - X_i}{h_1}\right)}{\hat{f}_{h_1}(X_i)} K\left(\frac{X_i - x}{h_1^*}\right) \\
&= \frac{1}{n^2 h_1^* h_1}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{[m_0(X_j) - m_0(X_i)] K\left(\frac{X_j - X_i}{h_1}\right)}{\hat{f}_{h_1}(X_i)} K\left(\frac{X_i - x}{h_1^*}\right) \\
&\quad + \frac{1}{n^2 h_1^* h_1}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\sigma_0(X_j)\epsilon_j K\left(\frac{X_j - X_i}{h_1}\right)}{\hat{f}_{h_1}(X_i)} K\left(\frac{X_i - x}{h_1^*}\right) \\
&= C_{n,1}(x) + C_{n,2}(x).
\end{aligned} \tag{S1.55}$$

It is easy to see that replacing the $\frac{1}{\hat{f}_{h_1}(X_i)}$ by $\frac{1}{f(X_i)}$ in $C_{n,j}$ gives the leading term of $C_{n,j}$, for $j = 1, 2$. Using the U-statistics and following the similar procedure as in Lemma A.2 (i), it is easy to show that uniformly in $x \in \mathcal{X}_n$,

$$\frac{C_{n,1}(x)}{\hat{f}_{h_1^*}(x)} = h_1^2 B_1(x) + o_{P_Z}\left(\frac{\sqrt{\log(n)}}{\sqrt{nh_1}}\right) = h_1^2 B_1(x) + o_{P_Z}\left(\frac{1}{\sqrt{nh_1^*}}\right), \tag{S1.56}$$

and

$$
\begin{aligned}
\frac{C_{n,2}(x)}{\hat{f}_{h_1^*}(x)} &= \frac{1}{nh_1}\sum_{i=1}^{n}\frac{\sigma_0(X_i)\epsilon_i}{f(x)}\int K\Big(\frac{h_1^*u+x-X_i}{h_1}\Big)K(u)du + o_{P_Z}\Big(\frac{1}{\sqrt{nh_1}}\Big) \\
&= o_{P_Z}\Big(\frac{1}{\sqrt{nh_1^*}}\Big). \qquad \text{(S1.57)}
\end{aligned}
$$

Equations (S1.55), (S1.56), (S1.57) and Assumption 4.1 imply that

$$
\frac{C_n(x)}{\hat{f}_{h_1^*}(x)} = o_{P_Z}\Big(\frac{1}{\sqrt{nh_1^*}}\Big). \qquad \text{(S1.58)}
$$

Combining the results (S1.49), (S1.50), (S1.51), (S1.54), (S1.58) and Assumption 4.1, we have that

$$
\frac{\sum_{i=1}^n \hat{m}_{h_1}(X_i)K(\frac{X_i-x}{h_1^*})}{\sum_{i=1}^n K(\frac{X_i-x}{h_1^*})} - \hat{m}_{h_1}(x) = o_{P_Z}\Big(\frac{1}{\sqrt{nh_1^*}}\Big). \qquad \text{(S1.59)}
$$

Equations (S1.40), (S1.48), and (S1.59) imply the Lemma A.5 (i). ■

Proof of Lemma A.5 (ii): By simple algebra we have that

$$
\begin{aligned}
&\hat{\sigma}^{*2}_{h_2^*}(x) - \hat{\sigma}^2_{h_2}(x) \\
=\ & \frac{\sum_{i=1}^n [Y_i^* - \hat{m}^*_{h_1^*}(X_i)]^2 K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} - \frac{\sum_{i=1}^n [Y_i - \hat{m}_{h_1}(X_i)]^2 K\big(\frac{X_i-x}{h_2}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2}\big)} \\
=\ & \frac{\sum_{i=1}^n [Y_i^* - \hat{m}^*_{h_1^*}(X_i)]^2 K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} - \frac{\sum_{i=1}^n [Y_i - \hat{m}_{h_1}(X_i)]^2 K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} \\
&+ \frac{\sum_{i=1}^n [Y_i - \hat{m}_{h_1}(X_i)]^2 K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} - \frac{\sum_{i=1}^n [Y_i - \hat{m}_{h_1}(X_i)]^2 K\big(\frac{X_i-x}{h_2}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2}\big)} \\
=\ & \Big[\frac{\sum_{i=1}^n [Y_i^* - \hat{m}^*_{h_1^*}(X_i)]^2 K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} - \frac{\sum_{i=1}^n [Y_i - \hat{m}_{h_1}(X_i)]^2 K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)}\Big] \\
&+ \Big[\frac{\sum_{i=1}^n [Y_i - \hat{m}_{h_1}(X_i)]^2 K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} - \sigma_0^2(x)\Big] - \Big[\frac{\sum_{i=1}^n [Y_i - \hat{m}_{h_1}(X_i)]^2 K\big(\frac{X_i-x}{h_2}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2}\big)} - \sigma_0^2(x)\Big] \\
=\ & D_1 + D_2 - D_3, \qquad \text{(S1.60)}
\end{aligned}
$$

where the definitions of $D_i, i = 1, 2, 3$ should be apparent.

By Lemma A.2 (ii), we have that

$$
D_2 = h_2^{*2}B_2(x) + \frac{1}{nh_2^* f(x)}\sum_{i=1}^n \sigma_0^2(X_i)(\epsilon_i^2 - 1)K\Big(\frac{X_i-x}{h_2^*}\Big) + o_{P_Z}\Big(\frac{1}{\sqrt{nh_2^*}}\Big), \qquad \text{(S1.61)}
$$

and

$$D_3 = h_2^2 B_2(x) + \frac{1}{nh_2 f(x)} \sum_{i=1}^{n} \sigma_0^2(X_i)(\epsilon_i^2 - 1) K\left(\frac{X_i - x}{h_2}\right) + o_{P_Z}\left(\frac{1}{\sqrt{nh_2}}\right), \quad \text{(S1.62)}$$

where $B_2(x) = \frac{1}{2}\mu_2\left[\frac{2\dot{\delta}_2(x)\dot{f}(x)}{f(x)} + \ddot{\delta}_2(x)\right]$, and $\delta_2(x) = \sigma_0^2(x)$.

Note that since $h_2^* = o_p(h_2)$, we have that

$$D_3 = o_{P_Z}\left(\frac{1}{\sqrt{nh_2^*}}\right). \quad \text{(S1.63)}$$

Let us now examine D_1.

$$\begin{aligned}
D_1 &= \frac{\sum_{i=1}^n [Y_i^* - \hat{m}^*_{h_1^*}(X_i)]^2 K\left(\frac{X_i-x}{h_2^*}\right)}{\sum_{i=1}^n K\left(\frac{X_i-x}{h_2^*}\right)} - \frac{\sum_{i=1}^n [Y_i - \hat{m}_{h_1}(X_i)]^2 K\left(\frac{X_i-x}{h_2^*}\right)}{\sum_{i=1}^n K\left(\frac{X_i-x}{h_2^*}\right)} \\
&= \frac{\sum_{i=1}^n [\hat{m}_{h_1}(X_i) + \hat{\sigma}_{h_2}(X_i)\epsilon_i^* - \hat{m}^*_{h_1^*}(X_i)]^2 K\left(\frac{X_i-x}{h_2^*}\right)}{\sum_{i=1}^n K\left(\frac{X_i-x}{h_2^*}\right)} \\
&\quad - \frac{\sum_{i=1}^n [m_0(X_i) + \sigma_0(X_i)\epsilon_i - \hat{m}_{h_1}(X_i)]^2 K\left(\frac{X_i-x}{h_2^*}\right)}{\sum_{i=1}^n K\left(\frac{X_i-x}{h_2^*}\right)} \\
&= \left\{\frac{\sum_{i=1}^n [\hat{m}_{h_1}(X_i) - \hat{m}^*_{h_1^*}(X_i)]^2 K\left(\frac{X_i-x}{h_2^*}\right)}{\sum_{i=1}^n K\left(\frac{X_i-x}{h_2^*}\right)} - \frac{\sum_{i=1}^n [m_0(X_i) - \hat{m}_{h_1}(X_i)]^2 K\left(\frac{X_i-x}{h_2^*}\right)}{\sum_{i=1}^n K\left(\frac{X_i-x}{h_2^*}\right)}\right\} \\
&\quad + \left\{\frac{\sum_{i=1}^n [\hat{\sigma}^2_{h_2}(X_i)\epsilon_i^{*2} - \sigma_0^2(X_i)\epsilon_i^2] K\left(\frac{X_i-x}{h_2^*}\right)}{\sum_{i=1}^n K\left(\frac{X_i-x}{h_2^*}\right)}\right\} \\
&\quad + 2\left\{\frac{\sum_{i=1}^n [\hat{m}_{h_1}(X_i)\hat{\sigma}_{h_2}(X_i)\epsilon_i^* - m_0(X_i)\sigma_0(X_i)\epsilon_i] K\left(\frac{X_i-x}{h_2^*}\right)}{\sum_{i=1}^n K\left(\frac{X_i-x}{h_2^*}\right)}\right\} \\
&\quad + 2\left\{\frac{\sum_{i=1}^n [\sigma_0(X_i)\hat{m}_{h_1}(X_i)\epsilon_i - \hat{\sigma}_{h_2}(X_i)\hat{m}^*_{h_1^*}(X_i)\epsilon_i^*] K\left(\frac{X_i-x}{h_2^*}\right)}{\sum_{i=1}^n K\left(\frac{X_i-x}{h_2^*}\right)}\right\} \\
&\equiv D_{1,1} + D_{1,2} + 2D_{1,3} + 2D_{1,4},
\end{aligned}$$

where the definitions of $C_i, i = 1, 2, 3, 4$ are obvious.

For $i = 1, \cdots, n$, define

$$\Delta_{n,m,h_1}(X_i) = \hat{m}_{h_1}(X_i) - m_0(X_i),$$

$$\Delta_{n,\hat{m},h_1^*}(X_i) = \hat{m}_{h_1^*}(X_i) - \hat{m}_{h_1}(X_i).$$

Then by Lemma A.1, uniformly in $x \in \mathcal{X}_n$,

$$\max_{1\le i\le n} |\Delta^2_{n,m,h_1}(X_i)\mathbf{1}_{i,n}| \le \big\{\sup_{x\in\mathcal{X}_n} |\hat{m}_{h_1}(x) - m_0(x)|\big\}^2 = O_{P_Z}\left(h_1^4 + \frac{\log(n)}{nh_1}\right). \quad \text{(S1.64)}$$

By Lemma A.5 (i), uniformly in $x \in \mathcal{X}_n$,

$$\max_{1\le i\le n} |\Delta^2_{n,\hat{m},h_1^*}(X_i)\mathbf{1}_{i,n}| \le \big\{\sup_{x\in\mathcal{X}_n} |\hat{m}^*_{h_1^*}(x) - \hat{m}_{h_1}(x)|\big\}^2 = O_{P_W}\left(h_1^{*^4} + \frac{\log(n)}{nh_1^*}\right). \quad \text{(S1.65)}$$

Equations (S1.64) and (S1.65) imply that

$$\begin{aligned} D_{1,1} &= \frac{\sum_{i=1}^n [\hat{m}_{h_1}(X_i) - \hat{m}^*_{h_1^*}(X_i)]^2 K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} - \frac{\sum_{i=1}^n [m_0(X_i) - \hat{m}_{h_1}(X_i)]^2 K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} \\ &\le O_{P_Z}\left(h_1^4 + \frac{\log(n)}{nh_1}\right) + O_{P_W}\left(h_1^{*^4} + \frac{\log(n)}{nh_1^*}\right) \\ &= o_{P_W}\left(\frac{1}{\sqrt{nh_1^*}}\right). \end{aligned} \quad \text{(S1.66)}$$

Next we examine the term $D_{1,2}$.

$$\begin{aligned} D_{1,2} &= \frac{\sum_{i=1}^n [\hat{\sigma}^2_{h_2}(X_i)\epsilon_i^{*2} - \sigma_0^2(X_i)\epsilon_i^2] K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} \\ &= \frac{\sum_{i=1}^n [\hat{\sigma}^2_{h_2}(X_i)\epsilon_i^{*2} - \sigma_0^2(X_i)\epsilon_i^{*2}] K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} + \frac{\sum_{i=1}^n \sigma_0^2(X_i)\epsilon_i^{*2} K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} \\ &\quad - \frac{\sum_{i=1}^n \sigma_0^2(X_i)\epsilon_i^2 K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} \\ &= \frac{\sum_{i=1}^n [\hat{\sigma}^2_{h_2}(X_i) - \sigma_0^2(X_i)]\epsilon_i^{*2} K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} + \frac{\sum_{i=1}^n \sigma_0^2(X_i)(\epsilon_i^{*2}-1) K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} \\ &\quad - \frac{\sum_{i=1}^n \sigma_0^2(X_i)(\epsilon_i^2-1) K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} \\ &= D_{1,2,1} + D_{1,2,2} + D_{1,2,3}, \end{aligned} \quad \text{(S1.67)}$$

where the definitions of $D_{1,3,i}$, $i = 1, 2, 3$ are obvious.

For $i = 1, \cdots, n$, define

$$\Delta_{n,\sigma^2,h_2}(X_i) = \hat{\sigma}^2_{h_2}(X_i) - \sigma_0^2(X_i).$$

Then by Lemma A.1,

$$\max_{1\le i\le n}|\Delta_{n,\sigma^2,h_2}(X_i)\mathbf{1}_{i,n}| \le sup_{x\in\mathcal{X}_n}|\hat{\sigma}^2_{h_2}(x)-\sigma^2_0(x)| = O_{P_Z}\left(h_1^2+h_2^2+\sqrt{\frac{\log(n)}{nh_1}}+\sqrt{\frac{\log(n)}{nh_2}}\right).$$

Using a similar argument leading to Equation (S1.43) in Lemma A.5 (i), one can show that, uniformly in $x \in \mathcal{X}_n$ and in P_Z probability,

$$D_{1,2,1} \quad = \quad O_{P_W}\left(h_2^2+\sqrt{\frac{\log(n)}{nh_2}}\right) = o_{P_W}\left(\frac{1}{\sqrt{nh_2^*}}\right). \qquad \text{(S1.68)}$$

Equations (S1.67) and (S1.68) imply that

$$D_{1,2} \quad = \quad \frac{\sum_{i=1}^n \sigma_0^2(X_i)(\epsilon_i^{*2}-1)K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} - \frac{\sum_{i=1}^n \sigma_0^2(X_i)(\epsilon_i^{2}-1)K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} + o_{P_W}\left(\frac{1}{\sqrt{nh_2^*}}\right). \qquad \text{(S1.69)}$$

Next we examine $D_{1,3}$.

$$\begin{aligned} D_{1,3} \quad &= \quad \frac{\sum_{i=1}^n[\hat{m}_{h_1}(X_i)\hat{\sigma}_{h_2}(X_i)\epsilon_i^* - \hat{m}_{h_1}(X_i)\sigma_0(X_i)\epsilon_i^*]K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} \\ &\quad + \frac{\sum_{i=1}^n[\hat{m}_{h_1}(X_i)\sigma_0(X_i)\epsilon_i^* - m_0(X_i)\sigma_0(X_i)\epsilon_i^*]K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} \\ &\quad + \frac{\sum_{i=1}^n[m_0(X_i)\sigma_0(X_i)\epsilon_i^* - m_0(X_i)\sigma_0(X_i)\epsilon_i]K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} \\ &= \quad D_{1,3,1}+D_{1,3,2}+D_{1,3,3}, \end{aligned}$$

where the terms $D_{1,2,i}$, $i=1,2,3$ are obvious.

Note that

$$\begin{aligned} D_{1,3,1} \quad &= \quad \frac{\sum_{i=1}^n[\hat{m}_{h_1}(X_i)\hat{\sigma}_{h_2}(X_i)\epsilon_i^* - \hat{m}_{h_1}(X_i)\sigma_0(X_i)\epsilon_i^*]K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} \\ &= \quad \frac{\sum_{i=1}^n \hat{m}_{h_1}(X_i)[\hat{\sigma}_{h_2}(X_i)-\sigma_0(X_i)]\epsilon_i^* K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} \\ &= \quad \frac{\sum_{i=1}^n [\hat{m}_{h_1}(X_i)-m_0(X_i)][\hat{\sigma}_{h_2}(X_i)-\sigma_0(X_i)]\epsilon_i^* K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} \\ &\quad + \frac{\sum_{i=1}^n m_0(X_i)[\hat{\sigma}_{h_2}(X_i)-\sigma_0(X_i)]\epsilon_i^* K\big(\frac{X_i-x}{h_2^*}\big)}{\sum_{i=1}^n K\big(\frac{X_i-x}{h_2^*}\big)} \\ &= \quad D_{1,3,1,1}+D_{1,3,1,2}, \end{aligned}$$

where the definitions of $C_{1,3,1,1}$ and $C_{1,3,1,2}$ are obvious.

Using a similar argument leading to Equation (S1.43) in Lemma A.5 (i), one can show that, uniformly in $x \in \mathcal{X}_n$ and in P_Z probability,

$$D_{1,3,1,1} = o_{P_W}\left(\frac{1}{\sqrt{nh_2^*}}\right), \qquad D_{1,3,1,2} = o_{P_W}\left(\frac{1}{\sqrt{nh_2^*}}\right), \qquad D_{1,3,2} = o_{P_W}\left(\frac{1}{\sqrt{nh_2^*}}\right).$$

Therefore,

$$D_{1,3} = \frac{\sum_{i=1}^n [m_0(X_i)\sigma_0(X_i)\epsilon_i^* - m_0(X_i)\sigma_0(X_i)\epsilon_i] K\left(\frac{X_i - x}{h_2^*}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h_2^*}\right)} + o_{P_W}\left(\frac{1}{\sqrt{nh_2^*}}\right). \quad \text{(S1.70)}$$

Finally we examine the term $D_{1,4}$.

$$\begin{aligned} D_{1,4} &= \frac{\sum_{i=1}^n [\sigma_0(X_i)\hat{m}_{h_1}(X_i)\epsilon_i - \hat{\sigma}_{h_2}(X_i)\hat{m}^*_{h_1^*}(X_i)\epsilon_i^*] K\left(\frac{X_i - x}{h_2^*}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h_2^*}\right)} \\ &= \left\{\frac{\sum_{i=1}^n \sigma_0(X_i)[\hat{m}_{h_1}(X_i) - m_0(X_i)]\epsilon_i K\left(\frac{X_i - x}{h_2^*}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h_2^*}\right)}\right\} \\ &\quad + \left\{\frac{\sum_{i=1}^n m_0(X_i)[\sigma_0(X_i) - \hat{\sigma}_{h_2}(X_i)]\epsilon_i^* K\left(\frac{X_i - x}{h_2^*}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h_2^*}\right)}\right\} \\ &\quad - \left\{\frac{\sum_{i=1}^n [m_0(X_i)\sigma_0(X_i)\epsilon_i^* - m_0(X_i)\sigma_0(X_i)\epsilon_i] K\left(\frac{X_i - x}{h_2^*}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h_2^*}\right)}\right\} \\ &\equiv D_{1,4,1} + D_{1,4,2} + D_{1,4,3}, \end{aligned} \quad \text{(S1.71)}$$

where the second equality holds by adding/subtracting same terms, and the definitions of $D_{1,4,i}, i = 1, 2, 3$ should be apparent.

Using a similar argument leading to Equation (S1.43) in Lemma A.5 (i), one can show that, uniformly in $x \in \mathcal{X}_n$ and in P_Z probability,

$$D_{1,4,1} = o_{P_W}\left(\frac{1}{\sqrt{nh_2^*}}\right), \qquad D_{1,4,2} = o_{P_W}\left(\frac{1}{\sqrt{nh_2^*}}\right). \quad \text{(S1.72)}$$

Equations (S1.71) and (S1.72) imply that

$$D_{1,4} = -\frac{\sum_{i=1}^n [m_0(X_i)\sigma_0(X_i)\epsilon_i^* - m_0(X_i)\sigma_0(X_i)\epsilon_i] K\left(\frac{X_i - x}{h_2^*}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h_2^*}\right)} + o_{P_W}\left(\frac{1}{\sqrt{nh_2^*}}\right). \quad \text{(S1.73)}$$

Combining Equations (S1.64), (S1.66), (S1.67), (S1.70) and (S1.71), we have that

$$\begin{aligned} D_1 &= \frac{\sum_{i=1}^n \sigma_0^2(X_i)(\epsilon_i^{*2}-1)K\left(\frac{X_i-x}{h_2^*}\right)}{\sum_{i=1}^n K\left(\frac{X_i-x}{h_2^*}\right)} - \frac{\sum_{i=1}^n \sigma_0^2(X_i)(\epsilon_i^{2}-1)K\left(\frac{X_i-x}{h_2^*}\right)}{\sum_{i=1}^n K\left(\frac{X_i-x}{h_2^*}\right)} \\ &\quad + o_{P_W}\left(\frac{1}{\sqrt{nh_1^*}}+\frac{1}{\sqrt{nh_2^*}}\right). \end{aligned} \tag{S1.74}$$

Results (S1.60), (S1.61), (S1.63), (S1.74), and Assumption 4.1 imply that

$$\hat{\sigma}_{h_2^*}^{*2}(x) - \hat{\sigma}_{h_2}^2(x) = \frac{1}{nh_2^* f(x)} \sum_{i=1}^n \sigma_0^2(X_i)(\epsilon_i^{*2}-1)K\left(\frac{X_i-x}{h_2^*}\right) + o_{P_W}\left(\frac{1}{\sqrt{nh_1^*}}+\frac{1}{\sqrt{nh_2^*}}\right).$$

This completes the proof of Lemma A.5 (ii). ■

Proof of Lemma A.6

The proof of Lemma A.6 is similar to the proof of Lemma A.3 and is omitted.

Proof of Lemma A.7

The proof of Lemma A.7 is similar to the proof of Lemma A.4, i.e., using the results of Lemma A.6 and the similar arguments as in the proof of Proposition 5 in Akritas and Van Keilegom (2001). The proof is omitted.

Proof of Lemma A.8

Proof: Our proof follows closely the proof of Lemma 1 in Akritas and Van Keilegom (2001), though the conclusion is more general. Fix $\epsilon \in (0,1)$ and $r \geq 1$. Without loss of generality, we may assume that $N_{[]}(\epsilon, \mathcal{G}_m, L^r(P_X))$ and $N_{[]}(\epsilon, \mathcal{G}_s, L^r(P_X))$ are both finite. Let $\{[a_i^l, a_i^u]\}_{i=1}^{m_a}$ and $\{[b_j^l, b_j^u]\}_{j=1}^{m_b}$ be $L^r(P_X)$-brackets of size ϵ. This construction readily yields: for any pair t and $\eta = (a,b)$, there exist some i and j such that

$$\begin{aligned} 1\{y \leq a_i^l(x) + b_j^l(x)t\} \leq 1\{y \leq a(x) + b(x)t\} \leq 1\{y \leq a_i^u(x) + b_j^u(x)t\} \text{ for all } t \geq 0\ , \\ 1\{y \leq a_i^l(x) + b_j^u(x)t\} \leq 1\{y \leq a(x) + b(x)t\} \leq 1\{y \leq a_i^u(x) + b_j^l(x)t\} \text{ for all } t < 0. \end{aligned}$$

Consider $t \geq 0$ for the moment. For each $i = 1, \ldots, m_a$ and $j = 1, \ldots, m_b$, define

$$F_{ij}^{l+}(t) = P(Y \leq a_i^l(X) + b_j^l(X)t) \text{ and } F_{ij}^{u+}(t) = P(Y \leq a_i^u(X) + b_j^u(X)t)\ .$$

Since both F_{ij}^{l+} and F_{ij}^{u+} are (potentially improper) distribution functions on $\mathbf{R}$, we may form a partition $0 = t_{ij,1}^{l+} < t_{ij,2}^{l+} < \cdots < t_{ij,m_l^+}^{l+} = \infty$ of $\bar{\mathbf{R}}_+$ into segments of F_{ij}^{l+}-size less than ϵ, and another partition $0 = t^{u+}_{ij,1} < t^{u+}_{ij,2} < \cdots < t^{u+}_{ij,m^+_u} = \infty$ of $\bar{\mathbf{R}}_+$ into segments of F^{u+}_{ij}-size less than ϵ. Similarly, for $t < 0$, define for each i and j:

$$F^{l-}_{ij}(t) = P(Y \leq a^l_i(X) + b^u_j(X)t) \text{ and } F^{u-}_{ij}(t) = P(Y \leq a^u_i(X) + b^l_j(X)t) \ .$$

Let $-\infty = t^{l-}_{ij,1} < t^{l-}_{ij,2} < \cdots < t^{l-}_{ij,m^-_l} = 0$ be a partition of $[-\infty, 0]$ into segments of F^{l-}_{ij}-size less than ϵ, and $-\infty = t^{u-}_{ij,1} < t^{u-}_{ij,2} < \cdots < t^{u-}_{ij,m^-_u} = 0$ another partition of $[-\infty, 0]$ into segments of F^{u-}_{ij}-size less than ϵ.

Now if $t \geq 0$, let t^{l+}_{ij,k^*} be the largest value in the first partition of $\bar{\mathbf{R}}_+$ that is less than or equal to t, and t^{u+}_{ij,ℓ^*} be the smallest value in the second partition of $\bar{\mathbf{R}}_+$ that is larger than or equal to t. If $t < 0$, let t^{l-}_{ij,p^*} be the largest value in the first partition of $[-\infty, 0]$ that is less than or equal to t, and t^{u-}_{ij,q^*} be the smallest value in the second partition of $[-\infty, 0]$ that is larger than or equal to t. Then for any pair t and $\eta = (a, b)$, there exist i, j, k^*, ℓ^*, p^* and q^* such that

$$1\{y \leq a^l_i(x) + b^l_j(x)t^{l+}_{ij,k^*}\} \leq 1\{y \leq a(x) + b(x)t\} \leq 1\{y \leq a^u_i(x) + b^u_j(x)t^{u+}_{ij,\ell^*}\} \text{ for all } t \geq 0 \ ,$$
$$1\{y \leq a^l_i(x) + b^u_j(x)t^{l-}_{ij,p^*}\} \leq 1\{y \leq a(x) + b(x)t\} \leq 1\{y \leq a^u_i(x) + b^l_j(x)t^{u-}_{ij,q^*}\} \text{ for all } t < 0.$$

We have thus constructed a collection of brackets that cover $\mathcal{F}$.

We next proceed to calculate the $L^r(P_Z)$-size of the above brackets. By Assumptions 3.1 (ii), (iii), 3.3 (ii), and 3.4 (i), the support of conditional distribution $F_{Y|X}(\cdot\,; x)$ for each $x \in \mathcal{X}$ is convex. Moreover, by the independence between X and ϵ, we have

$$F_{Y|X}(y; x) \equiv F_{Y|X=x}(y) = F_\epsilon\left(\frac{y - m_0(x)}{\sigma_0(x)}\right). \tag{S1.75}$$

It follows by Assumption 3.4 (ii), (iii) that $F_{Y|X}(\cdot\,; x)$ is differentiable for each x with uniformly continuous derivatives. Hence, by Lemma 3.11 in Aliprantis and Border (2006), we may assume that $\mathcal{Y}$ is open so that $f_{Y|X=x}(y)$ exists for all $y \in \mathcal{Y}$ and all $x \in \mathcal{X}$. Consider the brackets with $t \geq 0$ first. Then

$$\begin{aligned}
&\|1\{y \leq a^u_i(x) + b^u_j(x)t^{u+}_{ij,\ell^*}\} - 1\{y \leq a^l_i(x) + b^l_j(x)t^{l+}_{ij,k^*}\}\|^r_{P_Z,r} \\
&\quad = F^{u+}_{ij}(t^{u+}_{ij,\ell^*}) - F^{l+}_{ij}(t^{l+}_{ij,k^*}) \leq F^{u+}_{ij}(t) - F^{l+}_{ij}(t) + 2\epsilon \\
&\quad = \int [F_{Y|X=x}(a^u_i(x) + b^u_j(x)t) - F_{Y|X=x}(a^l_i(x) + b^l_j(x)t)]F_X(dx) + 2\epsilon \\
&\quad = \int f_{Y|X=x}(\tilde{a}_i(x) + \tilde{b}_j(x)t)[(a^u_i(x) - a^l_i(x)) + (b^u_j(x) - b^l_j(x))t]F_X(dx) + 2\epsilon \\
&\quad = \int f_{Y|X=x}(\tilde{a}_i(x) + \tilde{b}_j(x)t)(a^u_i(x) - a^l_i(x))F_X(dx) + 2\epsilon \\
&\qquad + \int f_{Y|X=x}(\tilde{a}_i(x) + \tilde{b}_j(x)t)(\tilde{a}_i(x) + \tilde{b}_j(x)t)\tilde{b}_j(x)^{-1}(b^u_j(x) - b^l_j(x))F_X(dx) \\
&\qquad - \int f_{Y|X=x}(\tilde{a}_i(x) + \tilde{b}_j(x)t)\tilde{a}_i(x)\tilde{b}_j(x)^{-1}(b^u_j(x) - b^l_j(x))F_X(dx) + 2\epsilon, \quad \text{(S1.76)}
\end{aligned}$$

where the third equality follows from the mean value theorem so that (i) $\tilde{a}_i(x)$ is between $a_i^u(x)$ and $a_i^l(x)$ and (ii) $\tilde{b}_i(x)$ is between $b_j^u(x)$ and $b_j^l(x)$. By Assumption 3.4 (iii), (iv), f_ϵ is bounded. This, together with Assumptions 3.1 (ii), (iii) and 3.3 (ii) give us

$$\begin{aligned}\sup_{x\in\mathcal{X},y\in\mathcal{Y}}|f_{Y|X=x}(y)| &= \sup_{x\in\mathcal{X},y\in\mathcal{Y}}\left|\frac{1}{\sigma_0(x)}f_\epsilon\left(\frac{y-m_0(x)}{\sigma_0(x)}\right)\right| \\ &\leq \frac{1}{\inf_{x\in\mathcal{X}_n}\sigma_0(x)}\sup_{x\in\mathcal{X},y\in\mathcal{Y}}f_\epsilon\left(\frac{y-m_0(x)}{\sigma_0(x)}\right)<\infty, \qquad \text{(S1.77)}\end{aligned}$$

and

$$\begin{aligned}\sup_{x\in\mathcal{X},y\in\mathcal{Y}}|yf_{Y|X=x}(y)| &= \sup_{x\in\mathcal{X},y\in\mathcal{Y}}\left|\frac{y}{\sigma_0(x)}f_\epsilon\left(\frac{y-m_0(x)}{\sigma_0(x)}\right)\right| \\ &= \sup_{x\in\mathcal{X},y\in\mathcal{Y}}\left|\frac{y-m_0(x)}{\sigma_0(x)}f_\epsilon\left(\frac{y-m_0(x)}{\sigma_0(x)}\right)\right| + \sup_{x\in\mathcal{X},y\in\mathcal{Y}}\left|\frac{m_0(x)}{\sigma_0(x)}f_\epsilon\left(\frac{y-m_0(x)}{\sigma_0(x)}\right)\right| \\ &< \infty, \qquad \text{(S1.78)}\end{aligned}$$

where for last step we exploited Assumption 3.4 (iv), m_0 being bounded above, σ_0 being bounded below, and f_ϵ being bounded. Hence, by results (S1.76), (S1.77) and (S1.78),

$$\begin{aligned}&\|1\{y\leq a_i^u(x)+b_j^u(x)t_{ij,\ell^*}^{u+}\}-1\{y\leq a_i^l(x)+b_j^l(x)t_{ij,k^*}^{l+}\}\|_{P,r}^r \\ &\qquad\lesssim \|a_i^u-a_i^l\|_{P,1}+\|b_j^u-b_j^l\|_{P,1}+2\epsilon \\ &\qquad\leq \|a_i^u-a_i^l\|_{P,r}+\|b_j^u-b_j^l\|_{P,r}+2\epsilon\leq 4\epsilon.\end{aligned}$$

On the other hand, if $t<0$, we have by analogous arguments that

$$\|1\{y\leq a_i^u(x)+b_j^l(x)t_{ij,q^*}^{u-}\}-1\{y\leq a_i^l(x)+b_j^u(x)t_{ij,p^*}^{l-}\}\|_{P_Z,r}^r\lesssim 4\epsilon.$$

We thus conclude that the $L^r(P_Z)$-size of the brackets is equal to some universal positive constant times ϵ^β. By construction, we may choose $m_a=N_{[]}(\epsilon,\mathcal{G}_m,L^r(P_X))$, $m_b=N_{[]}(\epsilon,\mathcal{G}_s,L^r(P_X))$, $m_l^+=m_u^+\leq 2\epsilon^{-1}$, and $m_l^-=m_u^-\leq 2\epsilon^{-1}$. Lemma A.8 then follows by a change of variables. ■

Proof of Lemma A.9

Proof: By Lemmas A.2 (i), A.2 (ii), A.5 (i) and A.5 (ii), we know that the leading term of $\bar{\mathbb{G}}_n^*(x)$ is given by $\tilde{\mathbb{G}}_n^*(x)$ which is defined as:

$$\tilde{\mathbb{G}}_n^*(x)\equiv\begin{bmatrix}\frac{1}{\sqrt{nh_1^*}}\sum_{i=1}^n\frac{\sigma_0(X_i)\epsilon_i^*}{f(x)}K(\frac{X_i-x}{h_1^*}) \\ \frac{1}{\sqrt{nh_2^*}}\sum_{i=1}^n\frac{\sigma_0^2(X_i)(\epsilon_i^{*2}-1)}{2f(x)\sigma_0(x)}K(\frac{X_i-x}{h_2^*})\end{bmatrix}\equiv\sum_{i=1}^n Z_{ni}^*.$$

Since $\{\epsilon_i^*\}_{i=1}^n$ is i.i.d. with mean zero and unit variance conditional on the data, by

Lemma A.5, we have for each fixed x belonging to the interior of $\mathcal{X}$ and almost every $\{Y_i, X_i\}_{i=1}^n$,

$$
\begin{aligned}
&\mathrm{Var}_W(\tilde{\mathbb{G}}_n^*(x)) \\
&= E_W \begin{bmatrix} \frac{1}{nh_1^*}\sum_{i=1}^n \frac{\sigma_0^2(X_i)}{f^2(x)} K^2(\frac{X_i-x}{h_1^*})\epsilon_i^{*2} & \frac{1}{n\sqrt{h_1^*h_2^*}}\sum_{i=1}^n \frac{\sigma_0^3(X_i)}{2f^2(x)\sigma_0(x)} K(\frac{X_i-x}{h_1^*})K(\frac{X_i-x}{h_2^*})\epsilon_i^{*3} \\ \frac{1}{n\sqrt{h_1^*h_2^*}}\sum_{i=1}^n \frac{\sigma_0^3(X_i)}{2f^2(x)\sigma_0(x)} K(\frac{X_i-x}{h_1^*})K(\frac{X_i-x}{h_2^*})\epsilon_i^{*3} & \frac{1}{nh_2^*}\sum_{i=1}^n \frac{\sigma_0^4(X_i)}{4f^2(x)\sigma_0^2(x)} K^2(\frac{X_i-x}{h_1^*})(\epsilon_i^{*2}-1)^2 \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{nh_1^*}\sum_{i=1}^n \frac{\sigma_0^2(X_i)}{f^2(x)} K^2(\frac{X_i-x}{h_1^*}) & \frac{1}{n\sqrt{h_1^*h_2^*}}\sum_{i=1}^n \frac{\sigma_0^3(X_i)}{2f^2(x)\sigma_0(x)} K(\frac{X_i-x}{h_1^*})K(\frac{X_i-x}{h_2^*})\mu_3 \\ \frac{1}{n\sqrt{h_1^*h_2^*}}\sum_{i=1}^n \frac{\sigma_0^3(X_i)}{2f^2(x)\sigma_0(x)} K(\frac{X_i-x}{h_1^*})K(\frac{X_i-x}{h_2^*})\mu_3 & \frac{1}{nh_2^*}\sum_{i=1}^n \frac{\sigma_0^4(X_i)}{4f^2(x)\sigma_0^2(x)} K^2(\frac{X_i-x}{h_1^*})(\hat{\mu}_4-1) \end{bmatrix} \\
&\equiv \begin{bmatrix} V_{1n} & V_{2n} \\ V_{2n} & V_{3n} \end{bmatrix},
\end{aligned}
$$

where we used the fact that $E_W(\epsilon_i^{*2}) = 1$, $E_W(\epsilon_i^*) = 0$, and $\hat{\mu}_j = E_W(\epsilon_i^{*j}) = \frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i^j$, $j = 3, 4$, and the implied definitions of $V_{i,n}$, $i = 1, 2, 3$.

It is easy to see that, uniformly in $x \in \mathcal{X}_n$,

$$
V_{1n} = \frac{\sigma_0^2(x)}{f(x)}\nu_0 + O_{P_Z}\left(h_1^{*2} + \sqrt{\frac{\log(n)}{nh_1^*}}\right), \tag{S1.79}
$$

where $\nu_0 = \int K^2(u)du$. Also, uniformly in $x \in \mathcal{X}_n$,

$$
V_{2n} = \frac{\sqrt{c}\sigma_0^2(x)\mu_3\nu_{1,2}}{2f(x)} + O_{P_Z}\left(h_1^{*2} + \sqrt{\frac{\log(n)}{nh_1^*}}\right), \tag{S1.80}
$$

where $c = \lim_{n\to\infty} h_1^*/h_2^*$, $\nu_{1,2} = \int K(u)K(cu)du$, and $\mu_3 = E(\epsilon_i^3)$.

Similarly, uniformly in $x \in \mathcal{X}_n$,

$$
V_{3n} = \frac{\sigma_0^2(x)\nu_0}{4f(x)}(\mu_4 - 1) + O_{P_Z}\left(h_1^{*2} + \sqrt{\frac{\log(n)}{nh_1^*}}\right), \tag{S1.81}
$$

where $\mu_4 = E(\epsilon_i^4)$.

Equations (S1.79), (S1.80) and (S1.81) together imply that uniformly in $x \in \mathcal{X}_n$,

$$
\mathrm{Var}_W(\mathbb{G}_n^*(x)) \xrightarrow{p} \frac{\sigma_0^2(x)}{f(x)} \begin{bmatrix} \nu_0 & \frac{1}{2}\sqrt{c}\mu_3\nu_{1,2} \\ \frac{1}{2}\sqrt{c}\mu_3\nu_{1,2} & \frac{1}{4}\nu_0(\mu_4-1) \end{bmatrix}.
$$

This, together with the Lyapunov central limit theorem, allows us to conclude that for each $x \in \mathcal{X}_n$ and with probability approaching one, $\tilde{\mathbb{G}}_n^*(x) \xrightarrow{L} \bar{\mathbb{G}}(x)$. Hence, $\bar{\mathbb{G}}_n^*(x) \xrightarrow{L} \bar{\mathbb{G}}(x)$ because $\tilde{\mathbb{G}}_n^*(x)$ is the leading term of $\bar{\mathbb{G}}_n^*(x)$. Since the bounded Lipschitz metric metrizes weak convergence, Lemma A.9 then follows. ∎

Proof of Lemma A.10

Proof of (i): Note that

$$\begin{aligned}
\hat{\epsilon}_i &= \frac{m_0(X_i) + \sigma_0(X_i)\epsilon_i - \hat{m}_{h_1}(X_i)}{\hat{\sigma}_{h_2}(X_i)} \\
&= \frac{m_0(X_i) - \hat{m}_{h_1}(X_i)}{\hat{\sigma}_{h_2}(X_i)} + \frac{\sigma_0(X_i)\epsilon_i}{\hat{\sigma}_{h_2}(X_i)} \\
&= \frac{m_0(X_i) - \hat{m}_{h_1}(X_i)}{\hat{\sigma}_{h_2}(X_i)} + \frac{\sigma_0(X_i) - \hat{\sigma}_{h_2}(X_i)}{\hat{\sigma}_{h_2}(X_i)}\epsilon_i + \epsilon_i. \qquad \text{(S1.82)}
\end{aligned}$$

Using a similar argument used in the proofs of Lemma A.5 (i) and A.5 (ii),

$$\begin{aligned}
\max_{1\le i\le n} \left|m_0(X_i) - \hat{m}_{h_1}(X_i)\right| &\le \sup_{x\in\mathcal{X}} |m(x) - \hat{m}_{h_1}(x)| \\
&= O_p\left(h_1 + \frac{\sqrt{\log(n)}}{\sqrt{nh_1}}\right), \qquad \text{(S1.83)}
\end{aligned}$$

where the bias is $O(h_1)$ rather than $O(h_1^2)$ because x may take values at the boundary of $\mathcal{X}$. Furthermore, we have that

$$\begin{aligned}
\max_{1\le i\le n} \left|\hat{\sigma}_{h_2}(X_i) - \sigma_0(X_i)\right| &\le \sup_{x\in\mathcal{X}} |\hat{\sigma}_{h_2}(x) - \sigma_0(x)| \\
&= O_p\left(h_1 + h_2 + \frac{\sqrt{\log(n)}}{\sqrt{nh_1}} + \frac{\sqrt{\log(n)}}{\sqrt{nh_2}}\right). \qquad \text{(S1.84)}
\end{aligned}$$

Plugging (S1.83) and (S1.84) into (S1.82) yields

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i - \frac{1}{n}\sum_{i=1}^n \epsilon_i &= \frac{1}{n}\sum_{i=1}^n \frac{m_0(X_i) - \hat{m}_{h_1}(X_i)}{\hat{\sigma}_{h_2}(X_i)} + \frac{1}{n}\sum_{i=1}^n \frac{\sigma_0(X_i) - \hat{\sigma}_{h_2}(X_i)}{\hat{\sigma}_{h_2}(X_i)}\epsilon_i \\
&= G_1 + G_2, \qquad \text{(S1.85)}
\end{aligned}$$

where the definitions of G_i, $i = 1, 2$ should be obvious.

$$\begin{aligned}
|G_1| &\lesssim \frac{1}{n}\sum_{i=1}^n \frac{\max_{1\le i\le n} |m_0(X_i) - \hat{m}_{h_1}(X_i)|}{\min_{1\le i\le n} \sigma_0(X_i)} \\
&\le \frac{1}{C}\max_{1\le i\le n} |m_0(X_i) - \hat{m}_{h_1}(X_i)| = o_p(1), \qquad \text{(S1.86)}
\end{aligned}$$

where in the second inequality we use the fact that $\min_{1\le i\le n} \sigma_0(X_i) \ge \inf_{x\in\mathcal{X}} \sigma_0^2(x) \equiv C > 0$ by Assumption 3.3 (ii), and in the last equality we use Equation (S1.83).

$$
\begin{aligned}
|G_2| &\lesssim \frac{1}{n}\sum_{i=1}^{n}\frac{\max_{1\le i\le n}|\sigma_0(X_i)-\hat{\sigma}_{h_2}(X_i)||\epsilon_i|}{\min_{1\le i\le n}\sigma_0(X_i)} \\
&\le \frac{1}{C}\sup_{x\in\mathcal{X}}|\sigma_0(x)-\hat{\sigma}_{h_2}(x)|\frac{1}{n}\sum_{i=1}^{n}|\epsilon_i| = o_p(1), \qquad \text{(S1.87)}
\end{aligned}
$$

where in the last equality we use Equation (S1.84), and the fact that $\frac{1}{n}\sum_{i=1}^{n}|\epsilon_i| \to E|\epsilon| = O(1)$. By a law of large numbers argument,

$$
\frac{1}{n}\sum_{i=1}^{n}\epsilon_i = E(\epsilon_i) + o_p(1). \qquad \text{(S1.88)}
$$

Equations (S1.85), (S1.86), (S1.87) and (S1.88) imply that

$$
\frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i = E(\epsilon_i) + o_p(1) = o_p(1). \qquad \text{(S1.89)}
$$

Using a similar argument leading to (S1.89), it is easy to show that $\frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i^2 \xrightarrow{p} E(\epsilon_i^2) = 1$ and $\frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i^3 \xrightarrow{p} E(\epsilon_i^3)$. The proof for (ii) is similar and is thus omitted. ■

S2. Proof of Proposition 5.1

In this Subsection we prove Proposition 5.1. First, we introduce some notation.

Definition 1. Let χ denote an $n\times 2$ matrix with the i^{th} row given by $(1,(X_i-x))$, $W_h = diag\big(K\big(\frac{X_i-x}{h}\big)\big)$, $S_h = \begin{bmatrix} s_0^h & s_1^h \\ s_1^h & s_2^h \end{bmatrix}$, where

$$
s_j^h(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{X_i-x}{h}\right)(X_i-x)^j, \qquad \text{(S2.1)}
$$

$$
\tilde{s}_j^h(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{X_i-x}{h}\right)(X_i-x)^j\sigma_0(X_i)\epsilon_i \qquad \text{(S2.2)}
$$

for $j\in\{0,1,\cdots,4\}$.

Definition 2. Define $\int_R K(u)u^q\,\mathrm{d}u = \mu_q \in [0,\infty)$, for $q\in\{0,1,\ldots,4\}$; $\int_R k^2(u)u^r\,\mathrm{d}u = \nu_r \in (0,\infty)$, for $r\in\{0,1,2\}$.

Definition 3. Define $\delta_1(x) = m_0^2(x)$, $\delta_2(x) = \sigma_0^2(x)$, and

$$
\mathbf{a}_1(x) = \begin{bmatrix} m_0(x) \\ \dot{m}_0(x) \end{bmatrix}, \quad \mathbf{a}_2(x) = \begin{bmatrix} \sigma_0(x) \\ \dot{\sigma}_0(x) \end{bmatrix}, \quad \mathbf{a}_3(x) = \begin{bmatrix} \delta_2(x) \\ \dot{\delta}_2(x) \end{bmatrix}.
$$

Let $\hat{m}_{h_1,LL}(x)$ and $\hat{\sigma}_{h_2,LL}(x)$ denote the local linear estimators for $m_0(x)$ and $\sigma_0(x)$ defined in Step 2 and Step 3 below, we replace (h_1, h_2) by undersmoothed bandwidth (b_1, b_2) to obtain $\hat{\epsilon}_{i,LL} = (Y_i - \hat{m}_{b_1,LL}(X_i))/\hat{\sigma}_{b_2,LL}(X_i)$. Inverting the empirical cumulative distribution (computed using $\{\hat{\epsilon}_{i,LL}\}_{i=1}^n$) leads to $\hat{Q}_{\epsilon,LL}(\tau)$. Using similar arguments as in the proof of Proposition 3.1, we can show that, uniformly in $\tau \in \mathcal{T}$,

$$\hat{Q}_{\epsilon,LL}(\tau) - Q_\epsilon(\tau) = O_p(b_1^2 + b_2^2 + n^{-1/2}). \tag{S2.3}$$

With the help of (S2.3), our proof of Proposition 5.1 consists of three steps.

Step 1: Decompose $\hat{\mathbf{a}}(x)$.

Let $\hat{\mathbf{a}}_1(x)$ and $\hat{\mathbf{a}}_2(x)$ be the local linear estimators of $\mathbf{a}_1(x)$ and $\mathbf{a}_2(x)$ defined in Steps 2 and 3 below, $\hat{Q}_{\epsilon,LL}(\tau)$ be defined as described above, then we have

$$\begin{aligned} \hat{\mathbf{a}}_\tau(x) &\stackrel{def}{=} \hat{\mathbf{a}}_1(x) + \hat{\mathbf{a}}_2(x)\hat{Q}_\epsilon(\tau) \\ &= \hat{\mathbf{a}}_1(x) + \hat{\mathbf{a}}_2(x)Q_\epsilon(\tau) + \hat{\mathbf{a}}_2(x)\big[\hat{Q}_{\epsilon,LL}(\tau) - Q_\epsilon(\tau)\big] \\ &= \hat{\mathbf{a}}_1(x) + \hat{\mathbf{a}}_2(x)Q_\epsilon(\tau) + (\text{s.o.}), \end{aligned} \tag{S2.4}$$

where the notation (s.o.) in $A_n(\tau) = B_n(\tau) + (s.o.)$ stands for terms having smaller probability order than $B_n(\tau)$.

Step 2: Property of $\hat{\mathbf{a}}_1(x)$.

The local linear estimator $\hat{\mathbf{a}}_1(x) = (\hat{a}_{1,1}(x), \hat{a}_{1,2}(x))^T = (\hat{m}_{h_1,LL}(x), \dot{\hat{m}}_{h_1,LL}(x))^T$ minimizes the following weighted squares objective function, i.e.,

$$\hat{\mathbf{a}}_1(x) = argmin_{a_1(x)=(a_{1,1}(x),a_{1,2}(x))^T} \sum_{i=1}^n \{Y_i - a_{1,1}(x) - a_{1,2}(x)(X_i - x)\}^2 k\bigg(\frac{X_i - x}{h_1}\bigg). \tag{S2.5}$$

Solving (S2.5) yields

$$\begin{aligned} \hat{\mathbf{a}}_1(x) &= (\chi^T W_{h_1}\chi)^{-1}\chi^T W_{h_1} Y \\ &= \bigg[\frac{1}{nh_1}\sum_{i=1}^n k\big((X_i - x)/h_1\big)\begin{pmatrix} 1 & X_i - x \\ X_i - x & (X_i - x)^2\end{pmatrix}\bigg]^{-1}\bigg\{\frac{1}{nh_1}\sum_{i=1}^n k\big((X_i - x)/h_1\big)\begin{pmatrix} 1 \\ X_i - x\end{pmatrix} Y_i\bigg\}. \end{aligned}$$

Plugging $Y_i = m_0(X_i) + \sigma_0(X_i)\epsilon_i$ into $\hat{\mathbf{a}}_1(x)$, we obtain

$$\hat{\mathbf{a}}_1(x) = \mathbf{a}_1(x) + S_{h_1}^{-1}\bigg[\frac{1}{2}\ddot{m}_0(x)c_{1n}^{h_1} + \frac{1}{6}(d^3 m_0(x)/dx^3)c_{2n}^{h_1}\bigg] + S_{h_1}^{-1}\tilde{c}_{3n}^{h_1} + (\text{s.o.}), \tag{S2.6}$$

where the expression of S_{h_1} is given in Definition 1 above, and

$$c_{1n}^h = \begin{bmatrix} s_2^h \\ s_3^h \end{bmatrix}, \quad c_{2n}^h = \begin{bmatrix} s_3^h \\ s_4^h \end{bmatrix}, \quad \tilde{c}_{3n}^h = \begin{bmatrix} \tilde{s}_0^h \\ \tilde{s}_1^h \end{bmatrix}, \tag{S2.7}$$

the expressions of s_j^h and $\tilde{s}_j^h$, $j \in \{1, \cdots, 4\}$ are given in (S2.1) and (S2.2).

It is straightforward to show that

$$s_j^{h_1}(x) = h_1^j \big[f(x)\mu_j + h_1 \dot{f}(x)\mu_{j+1} + O_p(h_1^2 + (nh_1)^{-1/2})\big].$$

Hence, we can further write

$$S_{h_1} = D_{h_1}\big[f(x)\bar{S}_1 + h_1\dot{f}(x)\bar{S}_2 + O_p(h_1^2 + (nh_1)^{-1/2})\big]D_{h_1}, \tag{S2.8}$$

$$c_{1n}^{h_1} = h_1^2 D_{h_1}\big[f(x)\bar{c}_{1,1} + h_1\dot{f}(x)\bar{c}_{1,2} + O_p(h_1^2 + (nh_1)^{-1/2})\big], \tag{S2.9}$$

$$c_{2n}^{h_1} = h_1^3 D_{h_1}\{f(x)\bar{c}_{1,2} + O_p(h_1^2 + (nh_1)^{-1/2})\}, \tag{S2.10}$$

where

$$D_{h_1} = \begin{bmatrix} 1 & 0 \\ 0 & h_1 \end{bmatrix}, \bar{S}_1 = \begin{bmatrix} 1 & 0 \\ 0 & \mu_2 \end{bmatrix}, \bar{S}_2 = \begin{bmatrix} 0 & \mu_2 \\ \mu_2 & 0 \end{bmatrix}, \bar{c}_{1,1} = \begin{bmatrix} \mu_2 \\ 0 \end{bmatrix}, \bar{c}_{1,2} = \begin{bmatrix} 0 \\ \mu_4 \end{bmatrix}.$$

From Equation (S2.8), and using the fact that $\{A+hB+o(h)\}^{-1} = A^{-1} - hA^{-1}BA^{-1} + o(h)$, we have

$$(S_{h_1})^{-1} = D_{h_1}^{-1}\left\{\frac{1}{f(x)}\bar{S}_1^{-1} - h_1\frac{\dot{f}(x)}{f^2(x)}\bar{S}_1^{-1}\bar{S}_2\bar{S}_1^{-1} + O_p(h_1^2 + (nh_1)^{-1/2})\right\}D_{h_1}^{-1}. \tag{S2.11}$$

Combining (S2.6), (S2.9), (S2.10) and (S2.11), we have that

$$\hat{\mathbf{a}}_1(x) = \mathbf{a}_1(x) + h_1^2\mathbf{B}_{1,LL}(x) + S_{h_1}^{-1}\tilde{c}_{3n}^{h_1} + (\text{s.o.}), \tag{S2.12}$$

where

$$\mathbf{B}_{1,LL}(x) = \begin{bmatrix} \frac{1}{2}\mu_2\ddot{m}_0(x) \\ \frac{\mu_4-\mu_2^2}{2\mu_2}\frac{\dot{f}(x)}{f(x)}\ddot{m}_0(x) + \frac{\mu_4}{6\mu_2}d^3m_0(x)/dx^3 \end{bmatrix}.$$

Step 3: Property of $\hat{\mathbf{a}}_2(x)$.

The LL estimator $\hat{\mathbf{a}}_3(x) = \begin{bmatrix} \hat{\delta}_{2,h_2,LL} \\ \dot{\hat{\delta}}_{2,h_2,LL} \end{bmatrix}$, which estimates $\mathbf{a}_3(x) = \begin{bmatrix} \delta_2(x) \\ \dot{\delta}_2(x) \end{bmatrix}$, is the minimizer of the following objective function:

$$\hat{\mathbf{a}}_3(x) = argmin_{a_{3,1}(x),a_{3,2}(x)}\sum_{i=1}^n\{\hat{u}_i^2 - a_{3,1}(x) - a_{3,2}(x)(X_i - x)\}^2 k\left(\frac{X_i - x}{h_2}\right),$$

where $\hat{u}_i = Y_i - \hat{m}_{h_1,LL}(X_i)$.

It is straightforward to show that

$$\begin{aligned}\hat{\mathbf{a}}_3(x) &= (\chi^T W_{h_2}\chi)^{-1}\chi^T W_{h_2}\hat{U}^2 \\ &= S_{h_2}^{-1}\left\{\frac{1}{nh_2}\sum_{i=1}^n k\big((X_i-x)/h_2\big)\begin{pmatrix}1\\ X_i-x\end{pmatrix}\hat{u}_i^2\right\},\end{aligned} \tag{S2.13}$$

where $\hat{U}^2 = (\hat{u}_1^2, \cdots, \hat{u}_n^2)^T$.

Noting that

$$\begin{aligned}\hat{u}_i^2 &= \big[Y_i - \hat{m}_{h_1,LL}(X_i)\big]^2 \\ &= \Big\{\big[m_0(X_i) - \hat{m}_{h_1,LL}(X_i)\big] + \sigma_0(X_i)\epsilon_i\Big\}^2 \\ &= \sigma_0^2(X_i)\epsilon_i^2 + \big[m_0(X_i) - \hat{m}_{h_1,LL}(X_i)\big]^2 + 2\sigma_0(X_i)\epsilon_i\big[m_0(X_i) - \hat{m}_{h_1,LL}(X_i)\big].\end{aligned} \tag{S2.14}$$

From Step 2, the bias of $\hat{m}_{h_1,LL}(X_i)$ is of order $O(h_1^2)$, however, its contribution to $\hat{a}_3(x)$ is $o(h_1^2)$. Using a similar argument in Fan and Yao (1998),[1] we have, uniformly for $x \in \mathcal{X}$, that

$$\hat{\mathbf{a}}_3(x) = S_{h_2}^{-1}\left\{\frac{1}{nh_2}\sum_{i=1}^n k\big((X_i-x)/h_2\big)\begin{pmatrix}1\\ X_i-x\end{pmatrix}\sigma_0^2(X_i)\epsilon_i^2\right\} + (\text{s.o.}), \tag{S2.15}$$

where the contribution of second and third terms in $\hat{u}_i$ as given in (S2.14) is of small order.

Using a similar leading-bias-term calculation as in Step 2, we obtain

$$\begin{aligned}\hat{\mathbf{a}}_3(x) &= \begin{bmatrix}\delta_2(x)\\ \dot{\delta}_2(x)\end{bmatrix} + h_2^2\mathbf{B}_{\delta_2,LL}(x) \\ &\quad + S_{h_2}^{-1}\left\{\frac{1}{nh_2}\sum_{i=1}^n k\big((X_i-x)/h_2\big)\begin{pmatrix}1\\ X_i-x\end{pmatrix}\sigma_0^2(X_i)\big[\epsilon_i^2-1\big]\right\} + (\text{s.o.}),\end{aligned} \tag{S2.16}$$

where

$$\mathbf{B}_{\delta_2,LL}(x) = \begin{bmatrix}\frac{1}{2}\mu_2\ddot{\delta}_2(x)\\ \frac{\mu_4-\mu_2^2}{2\mu_2}\frac{\dot{f}(x)}{f(x)}\ddot{\delta}_2(x) + \frac{\mu_4}{6\mu_2}d^3\delta_2(x)/dx^3\end{bmatrix}.$$

Using a similar argument as in the proof of Lemma A.1, we can show that

$$\sup_{x\in\mathcal{X}_n}|\hat{\sigma}_{h_2,LL}(x) - \sigma_0(x)| = O_p\left(h_1^2 + h_2^2 + \sqrt{\frac{log(n)}{nh_1}} + \sqrt{\frac{log(n)}{nh_2}}\right). \tag{S2.17}$$

[1]As pointed out in Fan and Yao (1998), using $\hat{u}_i = Y_i - \hat{m}_{h_1,LL}(X_i)$ is more efficient than using $\hat{u}_i = Y_i - \hat{m}_{h_1,LL}(x)$, where the former leads to smaller bias than the latter.

We estimate $\sigma_0(x)$ by $\hat{\sigma}_{h_2,LL}(x) = \sqrt{\hat{\delta}_{2,h_2,LL}}$, and $\dot{\sigma}_0(x)$ by $\dot{\hat{\sigma}}_{h_2,LL}(x) = \frac{\dot{\hat{\delta}}_{2,h_2,LL}}{2\hat{\sigma}_{h_2,LL}(x)}$. Noting that

$$\begin{aligned}
\hat{\delta}_{2,h_2,LL} - \delta_2(x) &\equiv \big[\hat{\sigma}_{h_2,LL}(x)\big]^2 - \sigma_0^2(x) \\
&= \big[\hat{\sigma}_{h_2,LL}(x) + \sigma_0(x)\big]\big[\hat{\sigma}_{h_2,LL}(x) - \sigma_0(x)\big] \\
&= 2\sigma_0(x)\big[\hat{\sigma}_{h_2,LL}(x) - \sigma_0(x)\big] + \text{(s.o.)}, \qquad \text{(S2.18)}
\end{aligned}$$

where the last equality follows from (S2.17). Hence, we have that

$$\hat{\sigma}_{h_2,LL}(x) - \sigma_0(x) = \frac{\hat{\delta}_{2,h_2,LL} - \delta_2(x)}{2\sigma_0(x)} + \text{(s.o.)}. \qquad \text{(S2.19)}$$

Similarly,

$$\begin{aligned}
\dot{\hat{\delta}}_{2,h_2,LL} - \dot{\delta}_2(x) &\equiv 2\hat{\sigma}_{h_2,LL}(x)\dot{\hat{\sigma}}_{h_2,LL}(x) - 2\sigma_0(x)\dot{\sigma}_0(x) \\
&= 2\hat{\sigma}_{h_2,LL}(x)\dot{\hat{\sigma}}_{h_2,LL}(x) - 2\hat{\sigma}_{h_2,LL}(x)\dot{\sigma}_0(x) + 2\hat{\sigma}_{h_2,LL}(x)\dot{\sigma}_0(x) - 2\sigma_0(x)\dot{\sigma}_0(x) \\
&= 2\hat{\sigma}_{h_2,LL}(x)\big[\dot{\hat{\sigma}}_{h_2,LL}(x) - \dot{\sigma}_0(x)\big] + 2\dot{\sigma}_0(x)\big[\hat{\sigma}_{h_2,LL}(x) - \sigma_0(x)\big] \\
&= 2\sigma_0(x)\big[\dot{\hat{\sigma}}_{h_2,LL}(x) - \dot{\sigma}_0(x)\big] + 2\dot{\sigma}_0(x)\big[\hat{\sigma}_{h_2,LL}(x) - \sigma_0(x)\big] + \text{(s.o.)}. \quad \text{(S2.20)}
\end{aligned}$$

Therefore,

$$\dot{\hat{\sigma}}_{h_2,LL}(x) - \dot{\sigma}_0(x) = \frac{\dot{\hat{\delta}}_{2,h_2,LL} - \dot{\delta}_2(x)}{2\sigma_0(x)} - \frac{\dot{\sigma}_0(x)}{\sigma_0(x)}\big[\hat{\sigma}_{h_2,LL}(x) - \sigma_0(x)\big] + \text{(s.o.)}. \qquad \text{(S2.21)}$$

Plugging (S2.19) into (S2.21), we have that

$$\dot{\hat{\sigma}}_{h_2,LL}(x) - \dot{\sigma}_0(x) = \frac{\dot{\hat{\delta}}_{2,h_2,LL} - \dot{\delta}_2(x)}{2\sigma_0(x)} - \frac{\dot{\sigma}_0(x)}{2\sigma_0^2(x)}\big[\hat{\delta}_{2,h_2,LL} - \delta_2(x)\big] + \text{(s.o.)}. \qquad \text{(S2.22)}$$

Combining (S2.16), (S2.19) and (S2.22), yields

$$
\begin{aligned}
\hat{\mathbf{a}}_2(x) - \mathbf{a}_2(x) &\equiv \begin{bmatrix} \hat{\sigma}_{h_2,LL}(x) - \sigma_0(x) \\ \dot{\hat{\sigma}}_{h_2,LL}(x) - \dot{\sigma}_0(x) \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{2\sigma_0(x)} & 0 \\ -\frac{\dot{\sigma}_0(x)}{2\sigma_0^2(x)} & \frac{1}{2\sigma_0(x)} \end{bmatrix} \begin{bmatrix} \hat{\delta}_{2,h_2,LL} - \delta_2(x) \\ \dot{\hat{\delta}}_{2,h_2,LL} - \dot{\delta}_2(x) \end{bmatrix} + (\text{s.o.}) \\
&\equiv \mathbf{H}\big[\hat{\mathbf{a}}_3(x) - \mathbf{a}_3(x)\big] + (\text{s.o.}) \\
&= h_2^2 \mathbf{H}\mathbf{B}_{\delta_2,LL}(x) + \mathbf{H}S_{h_2}^{-1}\left\{ \frac{1}{nh_2}\sum_{i=1}^n k\big((X_i - x)/h_2\big) \begin{pmatrix} 1 \\ X_i - x \end{pmatrix} \sigma_0^2(X_i)\big[\epsilon_i^2 - 1\big] \right\} + (\text{s.o.}),
\end{aligned}
\tag{S2.23}
$$

where $\mathbf{H} \equiv \begin{bmatrix} \frac{1}{2\sigma_0(x)} & 0 \\ -\frac{\dot{\sigma}_0(x)}{2\sigma_0^2(x)} & \frac{1}{2\sigma_0(x)} \end{bmatrix}$.

Combining (S2.4), (S2.12), (S2.23), gives

$$
\begin{aligned}
\hat{\mathbf{a}}_\tau(x) - \mathbf{a}(x) &= h_1^2 \mathbf{B}_{1,LL}(x) + h_2^2 Q_\epsilon(\tau)\mathbf{H}\mathbf{B}_{\delta_2,LL}(x) + S_{h_1}^{-1}\tilde{c}_{3n}^{h_1} \\
&\quad + \mathbf{H}Q_\epsilon(\tau)S_{h_2}^{-1}\left\{ \frac{1}{nh_2}\sum_{i=1}^n k\big((X_i - x)/h_2\big) \begin{pmatrix} 1 \\ X_i - x \end{pmatrix} \sigma_0^2(X_i)\big[\epsilon_i^2 - 1\big] \right\} + (\text{s.o.}).
\end{aligned}
$$

Therefore, by the Lyapunov central limit theorem, we have

$$
\sqrt{nh_1}\begin{bmatrix} 1 & 0 \\ 0 & h_1 \end{bmatrix} \big(\hat{\mathbf{a}}_\tau(x) - \mathbf{a}_\tau(x) - h_1^2\mathbf{B}_{1,LL}(x) - h_2^2\mathbf{B}_{2,LL}(x)\big) \xrightarrow{d} N(0, \mathbf{V}_{\tau,LL}), \tag{S2.24}
$$

where

$$
\mathbf{B}_{1,LL}(x) = \begin{bmatrix} \frac{\mu_2 \ddot{m}_0(x)}{2} \\ 0 \end{bmatrix}, \quad \mathbf{B}_{2,LL}(x) \begin{bmatrix} \frac{\mu_2 Q_\epsilon(\tau)\ddot{\delta}_2(x)}{4\sigma_0(x)} \\ 0 \end{bmatrix}, \quad \mathbf{V}_{\tau,LL} = \begin{bmatrix} V_{\tau,LL,1} & 0 \\ 0 & V_{\tau,LL,2} \end{bmatrix},
$$

$$
V_{\tau,LL,1} = V_\tau = \frac{\sigma_0^2(x)}{f(x)}\left\{ \nu_0 + \frac{c\nu_0 Q_\epsilon^2(\tau)}{4}\big[E(\epsilon^4) - 1\big] + cQ_\epsilon(\tau)E(\epsilon^3)\int K(u)K(cu)du \right\},
$$

$$
V_{\tau,LL,2} = \frac{\sigma_0^2(x)}{4\mu_2^2 f(x)}\left\{ 4\nu_2 + c^3\nu_2 Q_\epsilon^2(\tau)\big[E(\epsilon^4) - 1\big] + 4c^3 Q_\epsilon(\tau)E(\epsilon^3)\int u^2K(u)K(cu)du \right\},
$$

$$
c = \lim_{n\to\infty} h_1/h_2, \quad \mu_2 = \int u^2K(u)du, \quad \nu_j = \int u^jK^2(u)du \text{ for } j = 0, 2.
$$

S3. Additional Simulation and Empirical Results

Table S3.1 reports estimation MSE of our proposed estimator with the Epanechnikov Kernel. In Section 6 of the paper, we compute uniform confidence intervals over $\tau \in [0.1, 0.9]$. In Tables S3.2 and S3.3, we report uniform confidence intervals over $\tau \in [0.2, 0.8]$

for our proposed estimator using the residual bootstrap method with Gaussian kernel function. Table S3.4 presents estimated uniform confidence intervals over $\tau \in [0.2, 0.8]$ with the Epanechnikov Kernel replacing the Gaussian kernel. Next, we replace the residual based bootstrap method by the nonparametric (pair-wise) resampling bootstrap method and the score bootstrap method. The resulting uniform confidence intervals are provided in Tables S3.5–S3.8. These results are self-explanatory. The results are similar to the case of using the Gaussian kernel. We also report the computation time comparison between the score bootstrap and the residual bootstrap in Table S3.9. Finally, Figures S3.1 and S3.2 plot the estimated conditional quantile functions of our proposed method and the check function method using wage, instead of log(wage) as the dependent variable. The general pattern is somewhat similar to the case with logwage as the dependent variable. For example, the maximum wages for higher quantile curves appear at a much later age than the lower quantile curves. We observe from Figure S3.1 that for the lower quantile, the estimated quantile function takes negative values. However, this can be avoided by simply imposing a non-negativity condition on the estimated quantile function.

Table S3.1: Mean MSE ($\times$ 100) of our method, Epanechnikov Kernel

Error Distr.	Sample Size	Quantile Index						
		$\tau = 0.1$	$\tau = 0.15$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.85$	$\tau = 0.9$
$N(0,1)$	100	1.9956	1.7358	1.4820	1.3045	1.5772	1.9495	2.2805
	200	1.1200	0.9754	0.8155	0.7278	0.8817	1.0793	1.2533
	400	0.6230	0.5432	0.4518	0.4023	0.4795	0.5798	0.6706
$exp(1)$	100	0.9844	0.9755	0.9857	1.1228	1.9846	3.0838	4.2756
	200	0.5827	0.5804	0.5603	0.6165	1.0769	1.8531	2.6036
	400	0.3133	0.3093	0.3006	0.3217	0.5963	1.0369	1.4993
$\chi^2(5)$	100	1.2089	1.1778	1.0773	1.1676	1.8641	2.7197	3.5384
	200	0.6802	0.6585	0.5885	0.6420	1.0599	1.5877	2.0694
	400	0.3651	0.3477	0.3341	0.3641	0.6023	0.9038	1.1927

Table S3.2: The 95% Coverage Ratio (CR) of Uniform Confidence Interval over $\tau \in [0.2, 0.8]$, Gaussian Kernel, Residual Bootstrap

Error Distri.	Bandwidth	Eval. Points				
		-2/3	-1/3	0	1/3	2/3
Panel A: $n = 100$						
$N(0,1)$	Optimal	0.8380	0.8940	0.8710	0.9350	0.8840
	Undersmoothed	0.8680	0.9060	0.8920	0.9310	0.8900
$\chi^2(5)$	Optimal	0.8360	0.9350	0.8550	0.9460	0.8510
	Undersmoothed	0.8560	0.9330	0.8580	0.9310	0.8810
$exp(1)$	Optimal	0.8050	0.9230	0.8640	0.9370	0.8670
	Undersmoothed	0.8500	0.9210	0.8910	0.9500	0.8850
Panel B: $n = 200$						
$N(0,1)$	Optimal	0.9270	0.9130	0.9260	0.9740	0.9570
	Undersmoothed	0.9150	0.9310	0.9170	0.9510	0.9330
$\chi^2(5)$	Optimal	0.8790	0.9300	0.8760	0.9450	0.9090
	Undersmoothed	0.9200	0.9400	0.8940	0.9140	0.9020
$exp(1)$	Optimal	0.9320	0.9040	0.9340	0.9400	0.9200
	Undersmoothed	0.9180	0.9580	0.9280	0.9020	0.8840
Panel C: $n = 400$						
$N(0,1)$	Optimal	0.8950	0.9170	0.8830	0.9220	0.9270
	Undersmoothed	0.9360	0.9440	0.9240	0.9460	0.9500
$\chi^2(5)$	Optimal	0.8880	0.9460	0.8840	0.9620	0.9270
	Undersmoothed	0.9390	0.9710	0.9120	0.9590	0.9490
$exp(1)$	Optimal	0.9030	0.9510	0.9050	0.9680	0.9400
	Undersmoothed	0.9410	0.9610	0.9160	0.9700	0.9500

Table S3.3: The 90% Coverage Ratio (CR) of Uniform Confidence Interval over $\tau \in [0.2, 0.8]$, Gaussian Kernel, Residual Bootstrap

Error Distri.	Bandwidth	Eval. Points				
		-2/3	-1/3	0	1/3	2/3
Panel A: $n = 100$						
$N(0,1)$	Optimal	0.7780	0.8360	0.8000	0.8900	0.8050
	Undersmoothed	0.7800	0.8390	0.8150	0.8690	0.8240
$\chi^2(5)$	Optimal	0.7670	0.8760	0.7950	0.8910	0.7840
	Undersmoothed	0.7790	0.8590	0.7810	0.8900	0.7840
$exp(1)$	Optimal	0.7140	0.8620	0.7990	0.8970	0.7790
	Undersmoothed	0.7810	0.8410	0.8150	0.8690	0.8270
Panel B: $n = 200$						
$N(0,1)$	Optimal	0.8320	0.8470	0.8310	0.8870	0.8560
	Undersmoothed	0.8630	0.8820	0.8570	0.8950	0.8770
$\chi^2(5)$	Optimal	0.8090	0.8810	0.8210	0.9010	0.8460
	Undersmoothed	0.8250	0.8960	0.8430	0.8950	0.8650
$exp(1)$	Optimal	0.7930	0.8830	0.8280	0.9160	0.8480
	Undersmoothed	0.8510	0.8830	0.8400	0.9160	0.8680
Panel C: $n = 400$						
$N(0,1)$	Optimal	0.8330	0.8540	0.8200	0.8730	0.8600
	Undersmoothed	0.8840	0.8840	0.8610	0.9020	0.8990
$\chi^2(5)$	Optimal	0.8260	0.8960	0.8290	0.9200	0.8770
	Undersmoothed	0.8930	0.9190	0.8480	0.9080	0.8960
$exp(1)$	Optimal	0.8190	0.8850	0.8510	0.9180	0.8820
	Undersmoothed	0.8900	0.9100	0.8620	0.9150	0.9140

Table S3.4: The 95% and 90% Coverage Ratios (CRs) of Uniform Confidence Interval over $\tau \in [0.2, 0.8]$, Epanechnikov Kernel with Optimal Bandwidth

Error Distri.	CR	Eval. Points				
		-2/3	-1/3	0	1/3	2/3
Panel A: $n = 100$						
$N(0,1)$	95%	0.8420	0.8960	0.8490	0.9330	0.8950
	90%	0.7790	0.8300	0.7860	0.8880	0.8390
$\chi^2(5)$	95%	0.8120	0.9230	0.8590	0.9350	0.9010
	90%	0.7420	0.8640	0.7880	0.8890	0.8450
$exp(1)$	95%	0.8020	0.9300	0.8550	0.9290	0.8820
	90%	0.7080	0.8800	0.7720	0.8670	0.7870
Panel B: $n = 200$						
$N(0,1)$	95%	0.8830	0.8900	0.8810	0.9390	0.9130
	90%	0.8110	0.8400	0.8120	0.8810	0.8610
$\chi^2(5)$	95%	0.8630	0.9540	0.8720	0.9500	0.9250
	90%	0.7790	0.8930	0.7980	0.9040	0.8640
$exp(1)$	95%	0.8710	0.9390	0.8650	0.9570	0.9220
	90%	0.7920	0.8760	0.8120	0.9120	0.8420
Panel C: $n = 400$						
$N(0,1)$	95%	0.8930	0.9140	0.8940	0.9470	0.9170
	90%	0.8290	0.8420	0.8210	0.8820	0.8520
$\chi^2(5)$	95%	0.9060	0.9470	0.8880	0.9470	0.9480
	90%	0.8300	0.8930	0.8230	0.8900	0.8930
$exp(1)$	95%	0.8920	0.9450	0.9180	0.9550	0.9510
	90%	0.8290	0.8820	0.8640	0.9090	0.8750

Table S3.5: The 95% Coverage Ratio (CR) of Uniform Confidence Interval over $\tau \in [0.2, 0.8]$, Gaussian Kernel, Resampling Bootstrap

Error Distri.	Bandwidth	Eval. Points				
		-2/3	-1/3	0	1/3	2/3
Panel A: $n = 100$						
$N(0,1)$	Optimal	0.8980	0.8360	0.8980	0.9380	0.8940
	Undersmoothed	0.9120	0.8880	0.9320	0.9440	0.8700
$\chi^2(5)$	Optimal	0.9320	0.8440	0.9140	0.9260	0.8360
	Undersmoothed	0.9020	0.9060	0.9400	0.8960	0.8480
$exp(1)$	Optimal	0.9020	0.8720	0.9240	0.9160	0.8680
	Undersmoothed	0.8820	0.9160	0.9300	0.8820	0.8560
Panel B: $n = 200$						
$N(0,1)$	Optimal	0.9380	0.8880	0.9360	0.9400	0.9240
	Undersmoothed	0.9420	0.9260	0.9460	0.9360	0.9080
$\chi^2(5)$	Optimal	0.9340	0.8780	0.9160	0.9420	0.9100
	Undersmoothed	0.9200	0.9400	0.8940	0.9140	0.9020
$exp(1)$	Optimal	0.9320	0.9040	0.9340	0.9400	0.9200
	Undersmoothed	0.9180	0.9580	0.9280	0.9020	0.8840
Panel C: $n = 400$						
$N(0,1)$	Optimal	0.9460	0.8480	0.9440	0.9520	0.9160
	Undersmoothed	0.9600	0.9440	0.9580	0.9500	0.9420
$\chi^2(5)$	Optimal	0.9420	0.8780	0.9140	0.9640	0.9260
	Undersmoothed	0.9300	0.9340	0.9420	0.9380	0.9180
$exp(1)$	Optimal	0.9220	0.9120	0.9080	0.9460	0.9240
	Undersmoothed	0.9140	0.9440	0.9000	0.9240	0.9180

Table S3.6: The 90% Coverage Ratio (CR) of Uniform Confidence Interval over $\tau \in [0.2, 0.8]$, Gaussian Kernel, Resampling Bootstrap

Error Distri.	Bandwidth	Eval. Points				
		-2/3	-1/3	0	1/3	2/3
Panel A: $n = 100$						
$N(0,1)$	Optimal	0.8400	0.7600	0.8540	0.8840	0.8540
	Undersmoothed	0.8520	0.8340	0.8940	0.8940	0.8260
$\chi^2(5)$	Optimal	0.8560	0.7500	0.8760	0.8560	0.8040
	Undersmoothed	0.8400	0.8320	0.8880	0.8380	0.7720
$exp(1)$	Optimal	0.8520	0.8040	0.8780	0.8780	0.7980
	Undersmoothed	0.8140	0.8400	0.8480	0.8400	0.8320
Panel B: $n = 200$						
$N(0,1)$	Optimal	0.8940	0.8020	0.8940	0.8980	0.8760
	Undersmoothed	0.8460	0.8530	0.8520	0.8910	0.8640
$\chi^2(5)$	Optimal	0.8800	0.7900	0.8400	0.9100	0.8620
	Undersmoothed	0.8780	0.9080	0.8600	0.8600	0.8560
$exp(1)$	Optimal	0.8980	0.8100	0.8960	0.8800	0.8480
	Undersmoothed	0.8800	0.8980	0.8640	0.8580	0.8340
Panel C: $n = 400$						
$N(0,1)$	Optimal	0.8980	0.7620	0.8840	0.9120	0.8700
	Undersmoothed	0.9260	0.9040	0.9320	0.8960	0.9120
$\chi^2(5)$	Optimal	0.8940	0.7820	0.8640	0.9040	0.8820
	Undersmoothed	0.8920	0.8820	0.8860	0.9060	0.8880
$exp(1)$	Optimal	0.8760	0.8160	0.8600	0.9020	0.8840
	Undersmoothed	0.8600	0.8920	0.8660	0.8960	0.8800

Table S3.7: The 95% Coverage Ratio (CR) of Uniform Confidence Band over $\tau \in [0.2, 0.8]$, Gaussian Kernel, Score Bootstrap

Error Distri.	Bandwidth	Eval. Points				
		-2/3	-1/3	0	1/3	2/3
Panel A: $n = 100$						
$N(0,1)$	Optimal	0.8580	0.8180	0.8110	0.9150	0.8720
	Undersmoothed	0.8910	0.8690	0.8360	0.9310	0.9000
$\chi^2(5)$	Optimal	0.8420	0.8460	0.8350	0.8970	0.8470
	Undersmoothed	0.8440	0.8820	0.8400	0.8820	0.8470
$exp(1)$	Optimal	0.8230	0.8220	0.8490	0.8960	0.8340
	Undersmoothed	0.8130	0.8780	0.8580	0.8800	0.8340
Panel B: $n = 200$						
$N(0,1)$	Optimal	0.8990	0.8400	0.8240	0.9370	0.9320
	Undersmoothed	0.9300	0.8980	0.8850	0.9490	0.9420
$\chi^2(5)$	Optimal	0.8780	0.8640	0.8750	0.9420	0.8990
	Undersmoothed	0.9210	0.9220	0.9090	0.9300	0.9270
$exp(1)$	Optimal	0.8850	0.8710	0.9020	0.9280	0.9040
	Undersmoothed	0.9140	0.9220	0.9030	0.9410	0.9240
Panel C: $n = 400$						
$N(0,1)$	Optimal	0.9030	0.8530	0.8760	0.9450	0.9430
	Undersmoothed	0.9360	0.9230	0.9000	0.9650	0.9720
$\chi^2(5)$	Optimal	0.9260	0.8910	0.8990	0.9420	0.9380
	Undersmoothed	0.9380	0.9480	0.9330	0.9550	0.9350
$exp(1)$	Optimal	0.9170	0.8690	0.9200	0.9440	0.9160
	Undersmoothed	0.9460	0.9510	0.9370	0.9600	0.9520

Table S3.8: The 90% Coverage Ratio (CR) of Uniform Confidence Interval over $\tau \in [0.2, 0.8]$, Gaussian Kernel, Score Bootstrap

Error Distri.	Bandwidth	Eval. Points				
		-2/3	-1/3	0	1/3	2/3
Panel A: $n = 100$						
$N(0,1)$	Optimal	0.7950	0.7280	0.7480	0.8730	0.8320
	Undersmoothed	0.8160	0.7940	0.7680	0.8930	0.8540
$\chi^2(5)$	Optimal	0.7780	0.7600	0.7530	0.8470	0.7760
	Undersmoothed	0.7720	0.8180	0.7610	0.8370	0.7940
$exp(1)$	Optimal	0.7530	0.7320	0.7650	0.8440	0.7570
	Undersmoothed	0.7420	0.8110	0.7960	0.8260	0.7800
Panel B: $n = 200$						
$N(0,1)$	Optimal	0.8390	0.7590	0.7410	0.8870	0.8800
	Undersmoothed	0.8860	0.8350	0.8200	0.9180	0.9060
$\chi^2(5)$	Optimal	0.8190	0.7620	0.7970	0.8940	0.8440
	Undersmoothed	0.8700	0.8660	0.8500	0.8850	0.8810
$exp(1)$	Optimal	0.8210	0.7860	0.8300	0.8780	0.8420
	Undersmoothed	0.8480	0.8620	0.8310	0.8820	0.8710
Panel C: $n = 400$						
$N(0,1)$	Optimal	0.8610	0.7570	0.8060	0.9070	0.8950
	Undersmoothed	0.9020	0.8750	0.8480	0.9230	0.9320
$\chi^2(5)$	Optimal	0.8640	0.7920	0.8210	0.9050	0.8890
	Undersmoothed	0.8890	0.8920	0.8720	0.9220	0.8930
$exp(1)$	Optimal	0.8570	0.7740	0.8520	0.8900	0.8590
	Undersmoothed	0.9010	0.8980	0.8900	0.9140	0.9060

Table S3.9: Computation Time Comparison between Score and Residual Bootstrap

Estimator	Sample Size	Computation Time	Time Ratio in Perc.
Score Bootstrap	100	24.58 secs	29.84%
Residual Bootstrap	100	82.38 secs	
Score Bootstrap	200	41.03 secs	24.99%
Residual Bootstrap	200	164.17 secs	
Score Bootstrap	400	97.08 secs	20.50%
Residual Bootstrap	400	473.65 secs	

Note: The time ratio in perc is calculated as: $\frac{\text{computation time of Score Bootstrap}}{\text{computation time of Residual Bootstrap}} \times 100\%$.

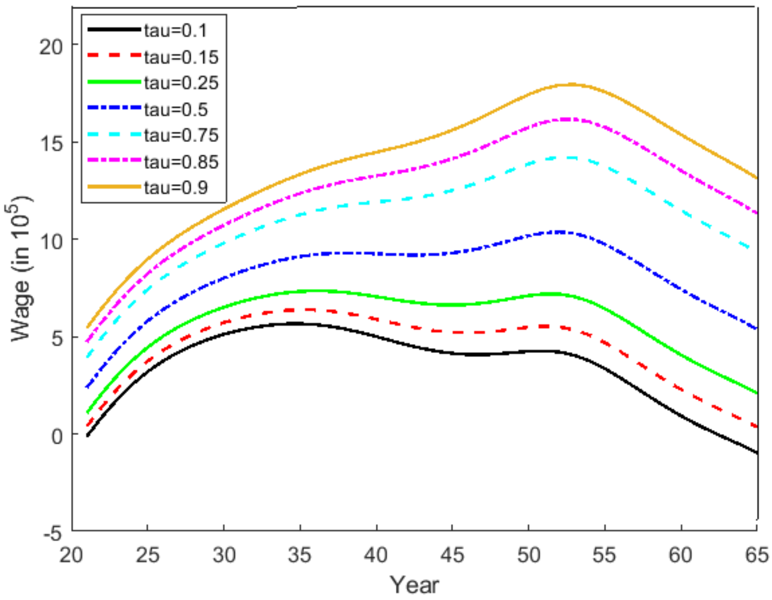

Figure S3.1: Our-method-based Wage Conditional Quantile Curve

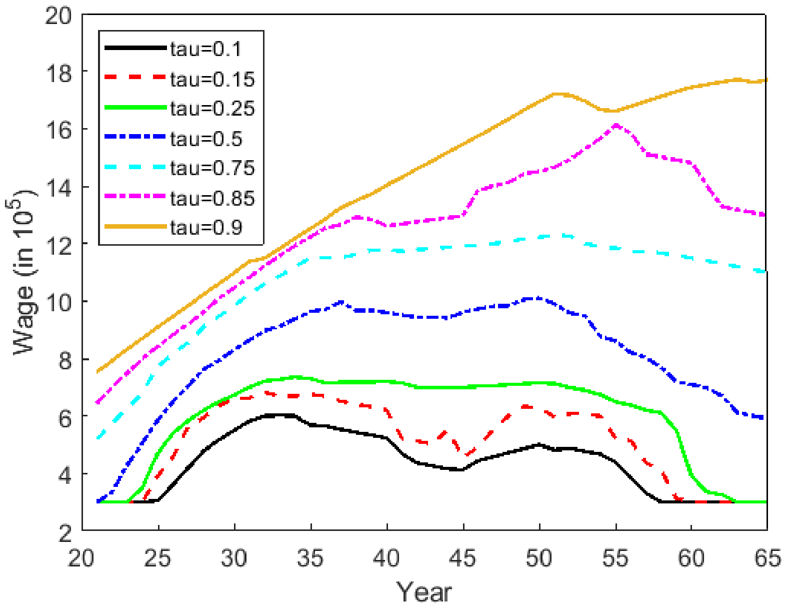

Figure S3.2: Check-function-based Wage Conditional Quantile Curve

References

Akritas, M. G. and Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics*, **28** 549–567.

Aliprantis, C. D. and Border, K. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. 3rd ed. Springer Verlag.

Cheng, G. and Huang, J. (2010). Bootstrap consistency for general semiparametric m-estimation. *The Annals of Statistics*, **38** 2884–2915.

Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *The Annals of Statistics*, **33** 1380–1403.

Ghosal, S., Sen, A. and van der Vaart, A. W. (2000). Testing monotonicity of regression. *The Annals of Statistics*, **28** 1054–1082.

Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis*, **17** 571–599.

Pollard, D. (1990). Empirical processes: Theory and applications. In *NSF-CBMS Regional Conference Series in Probability and Statistics*, vol. 2.

Sherman, R. P. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *The Annals of Statistics*, **22** 439–459.