SUPPLEMENTARY MATERIAL ON LOCAL COMPOSITE QUANTILE REGRESSION SMOOTHING: FLEXIBLE DATA STRUCTURE AND CROSS-VALIDATION*

XIAO HUANG
Kennesaw State University
ZHONGJIAN LIN
Emory University

This online supplement is an appendix to the paper and it includes all the proofs.

1 Propositions and Lemmas

***Proposition* 1.** Minimizing Equation (3) is equivalent to minimizing

$$L(\theta) = \sum_{k=1}^{q} u_k \left(\sum_{i=1}^{n} \frac{K_{ix}\eta_{i,k}^*}{\sqrt{nh^p}} \right) + v^T \left(\sum_{k=1}^{q} \sum_{i=1}^{n} \frac{K_{ix}(X_i^c - x^c)\eta_{i,k}^*}{\sqrt{nh^{p+2}}} \right) + \sum_{k=1}^{q} B_{n,k}(\theta)$$

with respect to θ, where

$$B_{n,k}(\theta) = \sum_{i=1}^{n} K_{ix} \int_0^{\Delta_{i,k}} \Big[1\{\sigma(X_i)(\varepsilon_i - c_k) + d_{i,k} \le z\} - 1\{\sigma(X_i)(\varepsilon_i - c_k) + d_{i,k} \le 0\} \Big] dz. \quad \text{(A.1)}$$

***Proposition* 2.** Under Assumptions 1 to 6, we have $L_n(\theta) = \frac{1}{2}\theta^T S_n \theta + (W_n^*)^T \theta + o_p(1)$.

***Proposition* 3.** Under Assumptions 1 to 6, we have

$$\hat{\theta} + \frac{\sigma(x)}{f(x)} S^{-1} \mathbb{E}(W_n^* | X) \xrightarrow{d} MVN\left(0, \frac{\sigma^2(x)}{f(x)} S^{-1} \Sigma S^{-1}\right). \quad \text{(A.2)}$$

***Lemma* 1.** Under Assumptions 1 to 6, we have

$$\frac{1}{nh^p} \sum_{i=1}^{n} K_{ix}/\sigma(X_i) = \frac{f(x)}{\sigma(x)} + O_p\left(\lambda + h^2 + \frac{1}{\sqrt{nh^p}}\right),$$
$$\frac{1}{nh^{p+1}} \sum_{i=1}^{n} K_{ix}(X_i^c - x^c)/\sigma(X_i) = h\mu_2 \frac{\nabla_{x^c} f(x)}{\sigma(x)} + O_p\left(\lambda h + h^3 + \frac{1}{\sqrt{nh^p}}\right),$$
$$\frac{1}{nh^{p+2}} \sum_{i=1}^{n} K_{ix}(X_i^c - x^c)(X_i^c - x^c)^T/\sigma(X_i) = \mu_2 \frac{f(x)}{\sigma(x)} + O_p\left(\lambda + h^2 + \frac{1}{\sqrt{nh^p}}\right).$$

*E-mail: xhuang3@kennesaw.edu and zhongjian.lin@emory.edu.

***Lemma* 2.** Under Assumptions 1 to 6, we have

$$\frac{1}{nh^p}\sum_{i=1}^{n} K_{ix}^2 = \nu_0 f(x) + O_p\Big(\lambda^2 + h^2 + \frac{1}{\sqrt{nh^p}}\Big),$$
$$\frac{1}{nh^{p+1}}\sum_{i=1}^{n} K_{ix}^2 (X_i^c - x^c) = h\nu_2 \nabla_{x^c} f(x) + O_p\Big(\lambda^2 h + h^3 + \frac{1}{\sqrt{nh^p}}\Big),$$
$$\frac{1}{nh^{p+2}}\sum_{i=1}^{n} K_{ix}^2 (X_i^c - x^c)(X_i^c - x^c)^T = \nu_2 f(x) + O_p\Big(\lambda^2 + h^2 + \frac{1}{\sqrt{nh^p}}\Big).$$

***Lemma* 3.** Under Assumptions 1 to 6, we have

$$Var(W_n|X) \xrightarrow{P} f(x)\Sigma \equiv \Omega.$$

Based on the definition of d_{ix} in Section 2, let $d_{ji} = \sum_{t=1}^{r}\{X_{t,j}^d \neq X_{t,i}^d\}$.
***Lemma* 4.** Under Assumptions 1 to 6, for all $i, j = 1, \cdots, n$ and $k = 1, \cdots, q$, we have

$$E\left(\eta_{j,k}^*|X_i, X_j\right) = \begin{cases} C_{1i,k}(X_j^c - X_i^c) + s.o. & \text{if } d_{ji} = 0, \\ C_{2i,k} + C_{3i,k}(X_j^c - X_i^c) + s.o. & \text{if } d_{ji} = 1, \end{cases}$$

where $C_{1i,k}$, $C_{2i,k}$, and $C_{3i,k}$ are functions of X_i and are defined in the proof of this lemma.
***Lemma* 5.** Under Assumptions 1 to 6, for all $i, j = 1, \cdots, n$, and $k, m = 1, \cdots, q$, we have

$$E\left(\eta_{j,k}^*\eta_{j,m}^*|X_i, X_j\right) = \begin{cases} (1-\tau_m)\tau_k + C_{4i,km}(X_j^c - X_i^c) + s.o. & \text{if } d_{ji} = 0, \\ C_{5i,km} + C_{6i,km}(X_j^c - X_i^c) + s.o. & \text{if } d_{ji} = 1, \end{cases}$$

where $C_{4i,km}$, $C_{5i,km}$, and $C_{6i,km}$ are functions of (X_i, τ_k, τ_m) and are defined in the proof of this lemma.

The following two lemmas use the U-statistics H-decomposition with variable kernels to calculate the expectations of $S_{1,km}$ and $S_{2,k}$. See Appendix B in Racine and Li (2004) for an intuitive explanation of H-decomposition.
***Lemma* 6.** Under Assumptions 1 to 6, for all $k, m = 1, \cdots, q$, we have

$$S_{1,km} = A_{1,km}^* h^4 - A_{2,km}^* h^2\lambda + A_{3,km}^*\lambda^2 + A_{4,km}^*(nh)^{-1}$$
$$+ \tilde{A}_{1,km}^* h^6 + \tilde{A}_{2,km}^* h^4\lambda + \tilde{A}_{3,km}^* h^2\lambda^2 + \tilde{A}_{4,km}^* + s.o.,$$

where the coefficients $A_{1,km}^*$, $A_{2,km}^*$, $A_{3,km}^*$, and $A_{4,km}^*$ are defined in the proof of this lemma.
***Lemma* 7.** Under Assumptions 1 to 6, for all $k = 1, \cdots, q$, we have

$$S_{2,k} = B_{1,k}^*\frac{h^2}{\sqrt{n}} + B_{2,k}^*\frac{\lambda}{\sqrt{n}} + s.o.,$$

where the coefficients $B_{1,k}^*$ and $B_{2,k}^*$ are defined in the proof of this lemma.
***Lemma* 8.** Under Assumptions 1 to 6, the leading term in the $o_p(1)$ term in Equation (A.6) has

order $O_p(h+\lambda+\frac{1}{\sqrt{nh^p}})$. Omitting this term in Equation (8) does not affect the asymptotic results in Theorem 2.

2 Proofs

Proof of Proposition 1. We write $Y_i - a_k - b(X_i^c - x^c) = \sigma(X_i)(\varepsilon_i - c_k) + d_{i,k} - \Delta_{i,k}$ in order to use the identity in Knight (1998), Kai, Li, and Zou (2010). By the identity in Knight (1998), minimizing Equation (3) is equivalent to minimizing

$$
\begin{aligned}
L_n(\theta) &= \sum_{i=1}^n \left\{ K_{ix} \sum_{k=1}^q \left[\rho_{\tau_k}\Big(\sigma(X_i)(\varepsilon_i - c_k) + d_{i,k} - \Delta_{i,k}\Big) - \rho_{\tau_k}\Big(\sigma(X_i)(\varepsilon_i - c_k) + d_{i,k}\Big)\right]\right\} \\
&= \sum_{i=1}^n \left\{ K_{ix} \sum_{k=1}^q \left[\Delta_{i,k}\Big[1\{\sigma(X_i)(\varepsilon_i - c_k) + d_{i,k} \le 0\} - \tau_k\Big] \right.\right. \\
&\qquad \left.\left. + \int_0^{\Delta_{i,k}} \Big[1\{\sigma(X_i)(\varepsilon_i - c_k) + d_{i,k} \le v\} - 1\{\sigma(X_i)(\varepsilon_i - c_k) + d_{i,k} \le 0\}\Big] dv \right]\right\} \\
&= \sum_{i=1}^n \left(K_{ix} \sum_{k=1}^q \left[\left(\frac{u_k}{\sqrt{nh^p}} + \frac{v^T (X_i^c - x^c)}{\sqrt{nh^{p+2}}} \right) \Big[1\{\sigma(X_i)(\varepsilon_i - c_k) + d_{i,k} \le 0\} - \tau_k\Big]\right]\right) \\
&\qquad + \sum_{k=1}^q B_{n,k}(\theta) \\
&= \sum_{k=1}^q u_k \left(\sum_{i=1}^n \frac{K_{ix}\eta_{i,k}^*}{\sqrt{nh^p}} \right) + v^T \left(\sum_{k=1}^q \sum_{i=1}^n \frac{K_{ix}(X_i^c - x^c)\eta_{i,k}^*}{\sqrt{nh^{p+2}}} \right) + \sum_{k=1}^q B_{n,k}(\theta).
\end{aligned}
$$

■

Proof of Proposition 2. Write $L_n(\theta)$ as

$$
L_n(\theta) = \sum_{k=1}^q u_k \left(\sum_{i=1}^n \frac{K_{ix}\eta_{i,k}^*}{\sqrt{nh^p}} \right) + v^T \left(\sum_{k=1}^q \sum_{i=1}^n \frac{K_{ix}(X_i^c - x^c)\eta_{i,k}^*}{\sqrt{nh^{p+2}}} \right) + \sum_{k=1}^q \mathbb{E}_\varepsilon[B_{n,k}(\theta)|X] + \sum_{k=1}^q R_{n,k}(\theta),
$$

where $R_{n,k}(\theta) = B_{n,k}(\theta) - \mathbb{E}_\varepsilon[B_{n,k}(\theta)|X]$. Using $F_\varepsilon(c_k + z) - F_\varepsilon(c_k) = zf(c_k) + o(z)$, we have

$$
\begin{aligned}
\sum_{k=1}^q \mathbb{E}_\varepsilon[B_{n,k}(\theta)|X] &= \sum_{k=1}^q \sum_{i=1}^n K_{ix} \int_0^{\Delta_{i,k}} \mathbb{E}_\varepsilon \left[1\{\varepsilon_i \le c_k + \frac{z - d_{i,k}}{\sigma(X_i)}\} - 1\{\varepsilon_i \le c_k - \frac{d_{i,k}}{\sigma(X_i)}\} \Big| X \right] dz \\
&= \sum_{k=1}^q \sum_{i=1}^n \left(K_{ix} \int_0^{\Delta_{i,k}} \left[\frac{z}{\sigma(X_i)} f_\varepsilon\Big(c_k - \frac{d_{i,k}}{\sigma(X_i)}\Big) + o(z) \right] dz \right) \\
&= \sum_{k=1}^q \sum_{i=1}^n \left[K_{ix} \frac{\Delta_{i,k}^2}{2\sigma(X_i)} f_\varepsilon\Big(c_k - \frac{d_{i,k}}{\sigma(X_i)}\Big) \right] + o_p(1) \\
&= \sum_{k=1}^q \sum_{i=1}^n \left[K_{ix} \frac{\Delta_{i,k}^2}{2\sigma(X_i)} f_\varepsilon(c_k) \right] + o_p(1) = \frac{1}{2}\theta^T S_n \theta + o_p(1). \qquad \text{(A.3)}
\end{aligned}
$$

We now prove that $R_{n,k}(\theta) = o_p(1)$. It is sufficient to show that $Var_\varepsilon[B_{n,k}(\theta)|X] = o_p(1)$. In fact,

$$\begin{aligned}
Var_\varepsilon[B_{n,k}(\theta)|X] &= \sum_{i=1}^{n} Var_\varepsilon\Big[\Big(K_{ix}\int_0^{\Delta_{i,k}}\Big(1\Big\{\varepsilon_i \le c_k - \frac{d_{i,k}}{\sigma(X_i)} + \frac{z}{\sigma(X_i)}\Big\} - 1\Big\{\varepsilon_i \le c_k - \frac{d_{i,k}}{\sigma(X_i)}\Big\}\Big)dz\Big)\Big|X\Big] \\
&\le \sum_{i=1}^{n} \mathbb{E}_\varepsilon\Big[\Big(K_{ix}\int_0^{\Delta_{i,k}}\Big(1\Big\{\varepsilon_i \le c_k - \frac{d_{i,k}}{\sigma(X_i)} + \frac{z}{\sigma(X_i)}\Big\} - 1\Big\{\varepsilon_i \le c_k - \frac{d_{i,k}}{\sigma(X_i)}\Big\}\Big)dz\Big)^2\Big|X\Big] \\
&\le \sum_{i=1}^{n} K_{ix}^2 \int_0^{|\Delta_{i,k}|}\int_0^{|\Delta_{i,k}|}\Big[F\Big(c_k - \frac{d_{i,k}}{\sigma(X_i)} + \frac{|\Delta_{i,k}|}{\sigma(X_i)}\Big) - F\Big(c_k - \frac{d_{i,k}}{\sigma(X_i)}\Big)\Big]dz_1 dz_2 \\
&= o\Big(\sum_{i=1}^{n} K_{ix}^2 \Delta_{i,k}^2\Big) = o_p(1).
\end{aligned}$$

■

Proof of Proposition 3. From Lemma 2, we have

$$S_n \xrightarrow{P} \frac{f(x)}{\sigma(x)} S = \frac{f(x)}{\sigma(x)}\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22}\end{pmatrix}. \tag{A.4}$$

Together with Propositions 1 and 2, we have

$$L_n(\theta) = \frac{1}{2}\frac{f(x)}{\sigma(x)}\theta' S\theta + (W_n^*)'\theta + o_p(1). \tag{A.5}$$

Since the convex function $L_n(\theta) - (W_n^*)'\theta$ converges in probability to the convex function $\frac{1}{2}\frac{f(x)}{\sigma(x)}\theta' S\theta$, it follows from the convexity lemma (Pollard (1991)) that, for any compact set Θ, the quadratic approximation to $L_n(\theta)$ holds uniformly for θ in any compact set, which leads to

$$\hat{\theta} = -\frac{\sigma(x)}{f(x)} S^{-1} W_n^* + o_p(1). \tag{A.6}$$

By the Cramér-Wold theorem, it is easy to see that the central limit theorem for $W_n|X$ holds:

$$\frac{W_n|X - \mathbb{E}[W_n|X]}{\sqrt{Var(W_n|X)}} \xrightarrow{d} MVN(0, I_{q+p}).$$

Note that

$$\begin{aligned}
Cov(\eta_{i,k}, \eta_{i,k'}) &= Cov(1\{\varepsilon_i \le c_k\} - \tau_k, 1\{\varepsilon_i \le c_{k'}\} - \tau_{k'}) = Cov(1\{\varepsilon_i \le c_k\}, 1\{\varepsilon_i \le c_{k'}\}) \\
&= \mathbb{E}[1\{\varepsilon_i \le c_k\} \times 1\{\varepsilon_i \le c_{k'}\}] - \tau_k \cdot \tau_{k'} = \tau_k \wedge \tau_{k'} - \tau_k \cdot \tau_{k'} = \tau_{kk'}, \\
Cov(\eta_{i,k}, \eta_{j,k'}) &= Cov(1\{\varepsilon_i \le c_k\} - \tau_k, 1\{\varepsilon_j \le c_{k'}\} - \tau_{k'}) = Cov(1\{\varepsilon_i \le c_k\}, 1\{\varepsilon_j \le c_{k'}\}) \\
&= \mathbb{E}[1\{\varepsilon_i \le c_k\} \times 1\{\varepsilon_j \le c_{k'}\}] - \tau_k \cdot \tau_{k'} = \tau_k \cdot \tau_{k'} - \tau_k \cdot \tau_{k'} = 0, \text{ if } i \ne j.
\end{aligned}$$

Further, for $k = 1, \cdots, q$,

$$
\begin{aligned}
&\mathbb{E}[w_{1k}|X] = \frac{1}{\sqrt{nh^p}} \sum_{i=1}^{n} \mathbb{E}[K_{ix}\eta_{i,k}|X] = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} K_{ix}\mathbb{E}[\eta_{i,k}|X] = 0,\\
&\mathbb{E}[w_{21}|X] = \frac{1}{\sqrt{nh^{p+2}}} \sum_{k=1}^{q}\sum_{i=1}^{n} \mathbb{E}[K_{ix}(X_i^c - x^c)\eta_{i,k}|X] = \frac{1}{\sqrt{nh^{p+2}}} \sum_{k=1}^{q}\sum_{i=1}^{n} K_{ix}(X_i^c - x^c)\mathbb{E}[\eta_{i,k}|X] = 0,\\
&Var(W_n|X) \xrightarrow{P} f(x)\Sigma.
\end{aligned}
$$

Therefore, $W_n|X \xrightarrow{d} MVN\Big(0, f(x)\Sigma\Big)$. Moreover, we have

$$
\begin{aligned}
Var(w_{1k}^* - w_{1k}|X) &= \frac{1}{nh^p} \sum_{i=1}^{n} K_{ix}^2 Var(\eta_{i,k}^* - \eta_{i,k}|X) \le \frac{1}{nh} \sum_{i=1}^{n} K_{ix}^2 \Big[F\Big(c_k + \frac{|d_{i,k}|}{\sigma(X_i)}\Big) - F(c_k)\Big] = o_p(1),\\
Var(w_{21}^* - w_{21}|X) &= \frac{1}{nh^{p+2}} \sum_{i=1}^{n} K_{ix}^2 (X_i^c - x^c)(X_i^c - x^c)^T Var[\sum_{k=1}^{q}(\eta_{i,k}^* - \eta_{i,k})|X]\\
&\le \frac{1}{nh^{p+2}} \sum_{i=1}^{n} K_{ix}^2 (X_i^c - x^c)(X_i^c - x^c)^T \max_k \Big[F\Big(c_k + \frac{|d_{i,k}|}{\sigma(X_i)}\Big) - F(c_k)\Big] = o_p(1).
\end{aligned}
$$

Thus $Var(W_n^* - W_n|X) = o_p(1)$. By the Slutsky's theorem, conditioning on X, we have $W_n^*|X - \mathbb{E}(W_n^*|X) \xrightarrow{d} MVN\Big(0, f(x)\Sigma\Big)$. Therefore,

$$
\hat{\theta} + \frac{\sigma(x)}{f(x)} S^{-1}\mathbb{E}(W_n^*|X) \xrightarrow{d} MVN\Big(0, \frac{\sigma^2(x)}{f(x)} S^{-1}\Sigma S^{-1}\Big).
$$

■

Proof of Theorem 1. $e_{q\times 1}$ denotes the $q \times 1$ vector of ones. S^{-1} is a diagonal matrix

$$
S^{-1} = \begin{pmatrix}
f(c_1) & 0 & \cdots & 0 & 0\\
0 & f(c_2) & \cdots & 0 & 0\\
\vdots & \vdots & \ddots & \vdots & \vdots\\
0 & 0 & \cdots & f(c_q) & 0\\
0 & 0 & \cdots & 0 & \mu_2 I_p \sum_{k=1}^{q} f(c_k)
\end{pmatrix}^{-1}.
$$

By the definition of θ and u_k, we have $a_k = \frac{u_k}{\sqrt{nh^p}} + g(x) + \sigma(x)c_k$.

$$
\begin{aligned}
\mathbb{E}(\hat{g}(x)|X) &= g(x) + \frac{1}{q}\sum_{k=1}^{q} \sigma(x)c_k + \frac{1}{q\sqrt{nh^p}} \sum_{k=1}^{q} \mathbb{E}(\hat{u}_k|X)\\
&= g(x) + \frac{1}{q}\sum_{k=1}^{q} \sigma(x)c_k - \frac{\sigma(x)}{q\sqrt{nh^p} f(x)} e_{q\times 1}(S^{-1})_{11}\mathbb{E}(W_{1n}^*|X),
\end{aligned}
$$

where W_{1n}^* are the first q elements of W_n^* and $(S^{-1})_{11}$ is the upper-left $q \times q$ block matrix of S^{-1}. Therefore,

$$\begin{aligned} Bias(\hat{g}(x)|X) &= \frac{1}{q}\sum_{k=1}^{q}\sigma(x)c_k - \frac{\sigma(x)}{q\sqrt{nh^p}f(x)}e_{q\times 1}(S^{-1})_{11}\mathbb{E}(W_{1n}^*|X) \\ &= \frac{1}{q}\sum_{k=1}^{q}\sigma(x)c_k - \frac{1}{qnh^p}\frac{\sigma(x)}{f(x)}\sum_{i=1}^{n}K_{ix}\sum_{k=1}^{q}\frac{1}{f(c_k)}\Big[F\Big(c_k - \frac{d_{i,k}}{\sigma(X_i)}\Big) - F(c_k)\Big]. \end{aligned}$$

Since the error is symmetric, $\sum_{k=1}^{q} c_k = 0$. Furthermore, it is easy to check that

$$\begin{aligned} \frac{1}{q}\sum_{k=1}^{q}\frac{1}{f(c_k)}\Big[F\Big(c_k - \frac{d_{i,k}}{\sigma(X_i)}\Big) - F(c_k)\Big] &= \frac{1}{q}\sum_{k=1}^{q}\frac{1}{f(c_k)}f(c_k)\times\Big[-\frac{d_{i,k}}{\sigma(X_i)} + o_p\Big(\frac{d_{i,k}}{\sigma(X_i)}\Big)\Big] \\ &= \frac{1}{q}\sum_{k=1}^{q}\Big[-\frac{c_k(\sigma(X_i)-\sigma(x,z)) + r_i}{\sigma(X_i)}\Big](1+o_p(1)) = -\frac{r_i}{\sigma(X_i)}(1+o_p(1)). \end{aligned}$$

Therefore,

$$\begin{aligned} Bias(\hat{g}(x)|X) &= -\frac{1}{qnh^p}\frac{\sigma(x)}{f(x)}\sum_{i=1}^{n}K_{ix}\sum_{k=1}^{q}\frac{1}{f(c_k)}\Big[F\Big(c_k - \frac{d_{i,k}}{\sigma(X_i)}\Big) - F(c_k)\Big] \\ &= -\frac{1}{nh^p}\frac{\sigma(x)}{f(x)}\sum_{i=1}^{n}K_{ix}\times\Big[-\frac{r_i}{\sigma(X_i)}(1+o_p(1))\Big] \\ &= -\frac{1}{nh^p}\frac{\sigma(x)}{f(x)}\sum_{i=1}^{n}K_{ix}\times\Big[-\frac{g(X_i)-g(x)-\beta(x)(X_i^c - x^c)}{\sigma(X_i)}\Big] \\ &= \Big[h^2\frac{tr[\beta'(x)]\mu_2}{2} + \lambda\sum_{\tilde{x}^d, d_{\tilde{x},x}=1}\frac{\sigma(x)f(x^c,\tilde{x}^d)[g(x^c,\tilde{x}^d)-g(x)]}{f(x)\sigma(x^c,\tilde{x}^d)}\Big](1+o_p(1)), \end{aligned}$$

when using

$$\begin{aligned} &\mathbb{E}\Big[\frac{1}{h^p}K_{ix}r_i/\sigma(X_i)\Big] \approx \frac{1}{h^p}\mathbb{E}[K_{ix}r_i/\sigma(X_i)|d_{ix}=0]\cdot P(d_{ix}=0) + \frac{1}{h^p}\mathbb{E}[K_{ix}r_i/\sigma(X_i)|d_{ix}=1]\cdot P(d_{ix}=1) \\ &= \frac{1}{h^p}\mathbb{E}[W_{ix}r_i/\sigma(X_i)|d_{ix}=0]\cdot P(d_{ix}=0) + \frac{\lambda}{h^p}\mathbb{E}[W_{ix}r_i/\sigma(X_i)|d_{ix}=1]\cdot P(d_{ix}=1) \\ &= \int f(X_i^c,x^d)W\Big(\frac{X_i^c - x^c}{h}\Big)[g(X_i^c,x^d) - g(x) - \beta(x)(X_i^c - x^c)]/\sigma(X_i^c,x^d)dX_i^c/h^p \\ &+ \lambda\sum_{\tilde{x}^d, d_{\tilde{x},x}=1}\int f(X_i^c,\tilde{x}^d)W\Big(\frac{X_i^c - x^c}{h}\Big)[g(X_i^c,\tilde{x}^d) - g(x) - \beta(x)(X_i^c - x^c)]/\sigma(X_i^c,\tilde{x}^d)dX_i^c/h^p \\ &= h^2\frac{f(x)tr[\beta'(x)]\mu_2}{2\sigma(x)} + O(h^4) + \lambda\sum_{\tilde{x}^d, d_{\tilde{x},x}=1}\frac{f(x^c,\tilde{x}^d)[g(x^c,\tilde{x}^d)-g(x)]}{\sigma(x^c,\tilde{x}^d)} + O(\lambda^2) \end{aligned}$$

$$= h^2 \frac{f(x) tr[\beta'(x)]\mu_2}{2\sigma(x)} + \lambda \sum_{\tilde{x}^d, d_{\tilde{x},x}=1} \frac{f(x^c, \tilde{x}^d)[g(x^c, \tilde{x}^d) - g(x)]}{\sigma(x^c, \tilde{x}^d)} + o(h^2 + \lambda).$$

Furthermore, the conditional variance of $\hat{g}(x)$ is

$$\begin{aligned}
Var[\hat{g}(x)|X] &= \frac{1}{nh^p} \frac{\sigma^2(x)}{f(x)} \frac{1}{q^2} e'_{q\times 1} (S^{-1}\Sigma S^{-1})_{11} e_{q\times 1} + o_p(\frac{1}{nh^p}) \\
&= \frac{1}{nh^p} \frac{\sigma^2(x)}{f(x)} \frac{1}{q^2} \sum_{k=1}^{q} \sum_{k'=1}^{q} \frac{\nu_0 \tau_{kk'}}{f(c_k) f(c_{k'})} + o_p(\frac{1}{nh}) \\
&= \frac{1}{nh^p} \frac{\sigma^2(x)}{f(x)} \nu_0 R_1(q) + o_p(\frac{1}{nh^p}).
\end{aligned}$$

■

Proof of Theorem 2. From Lemmas 6 and 7 and the fact that $B^*_{1,k}$ and $B^*_{2,k}$ are both zero mean $O_p(1)$ random variables, we conclude that orders in $S_{1,km}$ dominate those in $S_{2,k}$ and the leading term in cross-validation becomes Equation (16). Define

$$A_1 = q^{-2} \sum_{k=1}^{q} \sum_{m=1}^{q} A^*_{1,km}, \qquad A_2 = q^{-2} \sum_{k=1}^{q} \sum_{m=1}^{q} A^*_{2,km}, \qquad A_3 = q^{-2} \sum_{k=1}^{q} \sum_{m=1}^{q} A^*_{3,km},$$
$$A_4 = q^{-2} \sum_{k=1}^{q} \sum_{m=1}^{q} A^*_{4,km}, \qquad B_1 = 2q^{-1} \sum_{k=1}^{q} B^*_{1,k}, \qquad B_2 = 2q^{-1} \sum_{k=1}^{q} B^*_{2,k}.$$

Hence, Equation (11) can be written as

$$CV(h,\lambda) = A_1 h^4 - A_2 h^2 \lambda + A_3 \lambda^2 + A_4 (nh^p)^{-1} + B_1 \frac{h^2}{\sqrt{n}} + B_2 \frac{\lambda}{\sqrt{n}} + s.o.,$$

and the leading terms of $CV(h,\lambda)$ are collected in CV_0 in Equation (16).

From Equation (16), we have

$$\begin{aligned}
CV_0 &= A_1 h^4 - A_2 h^2 \lambda + A_3 \lambda^2 + A_4 (nh^p)^{-1} \\
&= A_3 \left(\lambda - \frac{A_2}{2A_3} h^2\right)^2 + \left(A_1 - \frac{A_2^2}{4A_3}\right) h^4 + \frac{A_4}{nh^p}.
\end{aligned}$$

CV_0 is minimized when $\lambda_0 = \frac{A_2}{2A_3} h_0^2$ and $h_0^{p+4} = \frac{A_4}{4n(A_1 - A_2^2/(4A_3))}$. Hence,

$$h_0 = c_1 n^{-1/(p+4)}, \;\; \lambda_0 = c_2 n^{-2/(p+4)}, \tag{A.7}$$

where

$$c_1 = \left[\frac{A_4}{A_1 - A_2^2/(4A_3)}\right]^{1/(p+4)}, \qquad c_2 = \left[\frac{A_2 A_4}{2A_3(A_1 - A_2^2/(4A_3))}\right]^{1/(p+4)}.$$

To prove the rate of convergence of $\hat{h}$ and $\hat{\lambda}$, we rewrite

$$\begin{aligned} CV(h,\lambda) &= A_3\left(\lambda - \frac{A_2h^2 - B_2n^{-1/2}}{2A_3}\right)^2 - A_3\left(\frac{A_2h^2 - B_2n^{-1/2}}{2A_3}\right)^2 \\ &\quad + A_1h^4 + A_4(nh^p)^{-1} + B_1h^2/\sqrt{n}, \\ &= A_3\left(\lambda - \frac{A_2h^2 - B_2n^{-1/2}}{2A_3}\right)^2 + \left(A_1 - \frac{A_2^2}{2}\right)h^4 + (A_2B_2 + B_1)n^{-1/2}h^2 + \frac{A_4}{nh^p} - \frac{B_2^2}{2}n^{-1}. \quad \text{(A.8)} \end{aligned}$$

Minimizing $CV(h,\lambda)$ w.r.t. (h,λ) in Equation (A.8) gives

$$\hat{\lambda} = \frac{A_2\hat{h}^2 - B_2n^{-1/2}}{2A_3}, \quad \text{(A.9)}$$

$$4(A_1 - A_2^2/2)\hat{h}^3 + 2(A_2B_2 + B_1)n^{-1/2}\hat{h} - \frac{pA_4}{n\hat{h}^{p+1}} = 0. \quad \text{(A.10)}$$

Let $\hat{h} = h_0 + h_1$, where h_1 is $o(h_0)$ since $\left(\hat{h} - h_0\right)/h_0 = o(1)$ and $CV(h,\lambda) = CV_0(h,\lambda) + o(CV_0)$. Substitute $\hat{h} = h_0 + h_1$, $(h_0 + h_1)^{p+4} = h_0^{p+4} + (p+4)h_0^{p+3}h_1 + s.o.$, and Equation (A.7) into Equation (A.10) to have

$$4\left(A_1 - A_2^2/(4A_3)\right)(p+4)h_0^{p+3}h_1 + 2(A_2B_2/(2A_3) + B_1)n^{-1/2}h_0^{p+2} + s.o. = 0,$$

which gives

$$h_1 = \frac{(A_2B_2/(2A_3) + B_1)n^{-1/2}h_0^3}{2(A_1 - A_2^2/(4A_3))(p+4)h_0^4}. \quad \text{(A.11)}$$

Replacing h_1 with $\hat{h} - h_0$ in Equation (A.11) gives

$$\frac{\hat{h} - h_0}{h_0} = \frac{(A_2B_2/(2A_3) + B_1)}{2(A_1 - A_2^2/(4A_3))(p+4)c_1^2}n^{-p/(2(p+4))} = O_p(n^{-p/(2(p+4))}). \quad \text{(A.12)}$$

For $\hat{\lambda}$, substitute $\hat{h} = h_0 + h_1$ and $\lambda_0 = \frac{A_2}{2A_3}h_0^2$ into Equation (A.9) to obtain

$$\begin{aligned} \hat{\lambda} &= A_2(h_0 + h_1)^2/(2A_3) - n^{-1/2}B_2/(2A_3) \\ &= \lambda_0 + 2h_0h_1A_2/(2A_3) + h_1^2A_2/(2A_3) - h^{-1/2}B_2/(2A_3) \quad \text{(A.13)} \\ &= \lambda_0 + O_p(n^{-1/2}), \quad \text{(A.14)} \end{aligned}$$

where the last line in Equation (A.14) follows $h_0h_1 = O(n^{-1/2})$, which can be verified by multiplying both sides of Equation (A.11) by h_0. ■

Proof of Corollary 1. There are two ways to establish the asymptotic normality results, by stochastic equicontinuity (Hall, Racine, and Li, 2004, Ichimura, 2000) or by Taylor expansion (Li and Racine, 2004, Racine and Li, 2004). Because we consider nonparametric regression, we follow the Taylor expansion proof strategy of Li and Racine (2004), Racine and Li (2004), which mainly deals with nonparametric regression, to show our result. Define $\bar{g}(x)$ in the same manner as $\hat{g}_{\hat{h},\hat{\lambda}}(x)$ but with h_0 and λ_0 replacing $\hat{h}$ and $\hat{\lambda}$. It is shown in the proof of Theorem 2 that

$h_0 = O\left(n^{-1/(p+4)}\right)$ and $\lambda_0 = O\left(n^{-2/(p+4)}\right)$, respectively. Assumption 3 holds with h_0 and λ_0 and therefore, by Theorem 1, we have

$$\sqrt{nh_0^p}\left[\bar{g}(x)-g(x)-h_0^2\frac{tr\left[\beta'(x)\right]\mu_2}{2}-\lambda_0\sum_{\tilde{x}^d,d_{\tilde{x},x}=1}\frac{\sigma(x)f(x^c,\tilde{x}^d)[g(x^c,\tilde{x}^d)-g(x)]}{f(x)\sigma(x^c,\tilde{x}^d)}\right]\xrightarrow{d} N\left(0,\frac{\sigma^2(x)}{f(x)}\nu_0 R_1(q)\right). \tag{A.15}$$

From Theorem 2, we have $\hat{h}^2 = h_0^2\left[1+O_p\left(n^{-p/(p+4)}\right)\right] = h_0^2 + o_p\left(n^{-2/(p+4)}\right)$, $\hat{\lambda} = \lambda_0(1+O_p(n^{-1/2})) = \lambda_0 + o_p\left(n^{-2/(p+4)}\right)$, and $\frac{1}{\hat{h}^p} = \frac{1}{h_0^p} + \frac{1}{h_0^p}O_p\left(\frac{\hat{h}-h_0}{h_0}\right) = \frac{1}{h_0^p}(1+o_p(1))$. From Equation (9), we see

$$\begin{aligned}
\hat{g}_{\hat{h},\hat{\lambda}}(x)-g(x) &= -\frac{\sigma(x)}{n\hat{h}^p q f(x)}\sum_{k=1}^{q}\sum_{i=1}^{n}\frac{\eta_{i,k}^*}{f(c_k)}W\left(\frac{X_i^c-x^c}{\hat{h}}\right)L_{\hat{\lambda}}(X_i^d,x^d),\\
&= -\frac{\sigma(x)}{nh_0^p q f(x)}\sum_{k=1}^{q}\sum_{i=1}^{n}\frac{\eta_{i,k}^*}{f(c_k)}W\left(\frac{X_i^c-x^c}{\hat{h}}\right)L_{\hat{\lambda}}(X_i^d,x^d)+s.o., \qquad (A.16)\\
&= -\frac{\sigma(x)}{nh_0^p q f(x)}\sum_{k=1}^{q}\sum_{i=1}^{n}\frac{\eta_{i,k}^*}{f(c_k)}\cdot L_{\lambda_0}(X_i^d,x^d)\cdot W\left(\frac{X_i^c-x^c}{\hat{h}}\right)+s.o.
\end{aligned}$$

By Taylor expansion, we have $W\left(\frac{X_i^c-x^c}{\hat{h}}\right) = W\left(\frac{X_i^c-x^c}{h_0}\right) + \widetilde{W}\left(\frac{X_i^c-x^c}{h_0}\right)\left(\frac{\hat{h}-h_0}{h_0}\right)+s.o.$, where $\widetilde{W}\left(\frac{X_i^c-x^c}{h_0}\right) \equiv h_0\cdot\frac{\partial\widetilde{W}\left(\frac{X_i^c-x^c}{h_0}\right)}{\partial h}$ and *s.o.* stands for "smaller order term." Therefore, we have

$$\hat{g}_{\hat{h},\hat{\lambda}}(x)-g(x) = \bar{g}(x)-g(x)-\frac{\sigma(x)}{nh_0^p q f(x)}\sum_{k=1}^{q}\sum_{i=1}^{n}\frac{\eta_{i,k}^*}{f(c_k)}\cdot L_{\lambda_0}(X_i^d,x^d)\cdot\widetilde{W}\left(\frac{X_i^c-x^c}{h_0}\right)\left(\frac{\hat{h}-h_0}{h_0}\right)+s.o. \tag{A.17}$$

It is easy to see that $\widetilde{W}(v)$ contains terms of $\partial W(v)/\partial v_k\cdot v_k$, $k=1,\cdots,p$. Because $W(\cdot)$ is a symmetric function by Assumption 4, we know that $\partial W(v)/\partial v_k$ are odd functions and $\partial W(v)/\partial v_k\cdot v_k, k=1,\cdots,p$ are symmetric functions. Thus $\widetilde{W}(v)$ is a symmetric function and can be taken as a second-order kernel function. Define $\widetilde{K}_{h_0,ix}\equiv L_{\lambda_0}(X_i^d,x^d)\cdot\widetilde{W}\left(\frac{X_i^c-x^c}{h_0}\right)$. We can take $\widetilde{K}_{h_0,ix}$ as a kernel function, similar to K_{ix}. With similar argument in the proof of Theorem 1, we can show that $\frac{\sigma(x)}{nh_0^p qf(x)}\sum_{k=1}^q\sum_{i=1}^n\frac{\eta_{i,k}^*}{f(c_k)}\cdot\widetilde{K}_{h_0,ix} = O_p\left(h_0^2+\lambda_0+\frac{1}{\sqrt{nh_0^p}}\right)$. Thus, $\hat{g}_{\hat{h},\hat{\lambda}}(x)-\bar{g}(x) = O_p\left(h_0^2+\lambda_0+\frac{1}{\sqrt{nh_0^p}}\right)\cdot\left(\frac{\hat{h}-h_0}{h_0}\right) = o_p\left(n^{-2/(p+4)}\right)$.

Since

$$\sqrt{nh_0^p}\cdot o_p\left(n^{-2/(p+4)}\right) = o_p\left(n^{1/2-p/2(p+4)-2/(p+4)}\right) = o_p(1),$$

replacing $\bar{g}(x)$, h_0 and λ_0 by $\hat{g}_{\hat{h},\hat{\lambda}}(x)$, $\hat{h}$ and $\hat{\lambda}$ in Equation (A.15) only introduces an $o_p(1)$ term and thus we prove the corollary. ■

Proof of Lemma 1. We derive the expectation and variance for each of the three terms. For the first term, we have

$$
\begin{aligned}
&\mathbb{E}\Big[\frac{1}{nh^p}\sum_{i=1}^{n} K_{ix}/\sigma(X_i)\Big] = \frac{1}{h^p}\mathbb{E}(K_{ix}/\sigma(X_i)) \\
&= \frac{1}{h^p}\Big[\mathbb{E}(K_{ix}/\sigma(X_i)|d_{ix}=0)\cdot P(d_{ix}=0) + \mathbb{E}(K_{ix}/\sigma(X_i)|d_{ix}=1)\cdot P(d_{ix}=1)(1+O(\lambda))\Big] \\
&= \frac{1}{h^p}\Big[\mathbb{E}(W_{ix}/\sigma(X_i)|d_{ix}=0)\cdot P(d_{ix}=0) + \lambda\mathbb{E}(W_{ix}/\sigma(X_i)|d_{ix}=1)\cdot P(d_{ix}=1)\cdot(1+O(\lambda))\Big] \\
&= \int \frac{f(X_i^c, x^d)}{\sigma(X_i^c, x^d)} W\Big(\frac{X_i^c - x^c}{h}\Big) dX_i^c/h^p + \lambda \sum_{\tilde{x}^d, d_{\tilde{x},x}=1} \int \frac{f(X_i^c, \tilde{x}^d)}{\sigma(X_i^c, \tilde{x}^d)} W\Big(\frac{X_i^c - x^c}{h}\Big) dX_i^c/h^p \cdot (1+O(\lambda)) \\
&= \int \frac{f(x^c + hz, x^d)}{\sigma(x^c + hz, x^d)} W(z)dz + \lambda \sum_{\tilde{x}^d, d_{\tilde{x},x}=1} \int \frac{f(x^c + hz, \tilde{x}^d)}{\sigma(x^c + hz, \tilde{x}^d)} W(z)dz \cdot (1+O(\lambda)) = \frac{f(x)}{\sigma(x)} + O(h^2 + \lambda)
\end{aligned}
$$

and

$$
\begin{aligned}
&Var\Big[\frac{1}{nh^p}\sum_{i=1}^{n} K_{ix}/\sigma(X_i)\Big] = \frac{1}{nh^{2p}} Var(K_{ix}/\sigma(X_i)) = \frac{1}{nh^{2p}}\Big[\mathbb{E}(K_{ix}^2/\sigma^2(X_i)) + O(h^{2p})\Big] \\
&\approx \frac{1}{nh^{2p}}\Big[\mathbb{E}(K_{ix}^2/\sigma^2(X_i)|d_{ix}=0)\cdot P(d_{ix}=0) + \mathbb{E}(K_{ix}^2/\sigma^2(X_i)|d_{ix}=1)\cdot P(d_{ix}=1) + O(h^{2p})\Big] \\
&= \frac{1}{nh^{2p}}\Big[\mathbb{E}(W_{ix}^2/\sigma^2(X_i)|d_{ix}=0)\cdot P(d_{ix}=0) + \lambda^2\mathbb{E}(W_{ix}^2/\sigma^2(X_i)|d_{ix}=1)\cdot P(d_{ix}=1) + O(h^{2p})\Big] \\
&= \frac{1}{nh^p}\Big[\int f(X_i^c, x^d) W^2\Big(\frac{X_i^c - x^c}{h}\Big) dX_i^c/h^p + O(\lambda^2 + h^{2p})\Big] = \frac{1}{nh^p}\nu_0\frac{f(x)}{\sigma^2(x)} + o\Big(\frac{1}{nh^p}\Big).
\end{aligned}
$$

Thus

$$
\frac{1}{nh^p}\sum_{i=1}^{n} K_{ix}/\sigma(X_i) = \frac{f(x)}{\sigma(x)} + O_p\Big(\lambda + h^2 + \frac{1}{\sqrt{nh^p}}\Big).
$$

For the second term, we have

$$
\begin{aligned}
&\mathbb{E}\Big[\frac{1}{nh^{p+1}}\sum_{i=1}^{n}\frac{K_{ix}}{\sigma(X_i)}(X_i^c - x^c)\Big] = \frac{1}{h^{p+1}}\mathbb{E}\Big[\frac{K_{ix}}{\sigma(X_i)}(X_i^c - x^c)\Big] \\
&\approx \frac{1}{h^{p+1}}\Big[\mathbb{E}\Big[\frac{K_{ix}}{\sigma(X_i)}(X_i^c - x^c)|d_{ix}=0\Big]\cdot P(d_{ix}=0) + \mathbb{E}\Big[\frac{K_{ix}}{\sigma(X_i)}(X_i^c - x^c)|d_{ix}=1\Big]\cdot P(d_{ix}=1)\Big] \\
&= \frac{1}{h^{p+1}}\Big[\mathbb{E}\Big[\frac{W_{ix}}{\sigma(X_i)}(X_i^c - x^c)|d_{ix}=0\Big]\cdot P(d_{ix}=0) + \lambda\mathbb{E}\Big[\frac{W_{ix}}{\sigma(X_i)}(X_i^c - x^c)|d_{ix}=1\Big]\cdot P(d_{ix}=1)\Big] \\
&= \frac{1}{h}\Big[\int \frac{f(X_i^c, x^d)}{\sigma(X_i^c, x^d)} W\Big(\frac{X_i^c - x^c}{h}\Big)(X_i^c - x^c) dX_i^c/h^p + O(\lambda h^2)\Big] = \frac{1}{h}\int \frac{f(x^c + hz, x^d)}{\sigma(x^c + hz, x^d)} W(z)hzdz + O(\lambda h) \\
&= h\mu_2\frac{\nabla_{x^c} f(x)}{\sigma(x)} + O(\lambda h + h^3)
\end{aligned}
$$

and

$$
\begin{aligned}
&Var\Big[\frac{1}{nh^{p+1}}\sum_{i=1}^{n}\frac{K_{ix}}{\sigma(X_i)}(X_i^c-x^c)\Big]=\frac{1}{nh^{2p+2}}Var[\frac{K_{ix}}{\sigma(X_i)}(X_i^c-x^c)]\\
&=\frac{1}{nh^{2p+2}}\Big[\mathbb{E}\Big[\Big(\frac{K_{ix}}{\sigma(X_i)}\Big)^2(X_i^c-x^c)(X_i^c-x^c)^T\Big]+O(h^{2p+4})\Big]\\
&\approx\frac{1}{nh^{2p+2}}\Big[\mathbb{E}\Big[\Big(\frac{W_{ix}}{\sigma(X_i)}\Big)^2(X_i^c-x^c)(X_i^c-x^c)^T|d_{ix}=0\Big]\cdot P(d_{ix}=0)\\
&\qquad+\lambda^2\mathbb{E}\Big[\Big(\frac{W_{ix}}{\sigma(X_i)}\Big)^2(X_i^c-x^c)(X_i^c-x^c)^T|d_{ix}=1)\cdot P(d_{ix}=1\Big]+O(h^{2p+4})\Big]\\
&=\frac{1}{nh^{p+2}}\Big[\int f(X_i^c,x^d)W^2\Big(\frac{X_i^c-x^c}{h}\Big)(X_i^c-x^c)(X_i^c-x^c)^T dX_i^c/h^p+O(\lambda^2h^2+h^{2p+4})\Big]\\
&=\frac{1}{nh^p}I_p\nu_2\frac{f(x)}{\sigma^2(x)}+o\Big(\frac{1}{nh^p}\Big).
\end{aligned}
$$

Thus

$$
\frac{1}{nh^{p+1}}\sum_{i=1}^{n}K_{ix}(X_i^c-x^c)/\sigma(X_i)=h\mu_2\frac{\nabla_{x^c}f(x)}{\sigma(x)}+O_p\Big(\lambda h+h^3+\frac{1}{\sqrt{nh^p}}\Big).
$$

For the third term, we have

$$
\begin{aligned}
\mathbb{E}\Big[\frac{1}{nh^{p+2}}\sum_{i=1}^{n}\frac{K_{ix}}{\sigma(X_i)}(X_i^c-x^c)(X_i^c-x^c)^T\Big]&=\frac{1}{h^{p+2}}\mathbb{E}[\frac{K_{ix}}{\sigma(X_i)}(X_i^c-x^c)(X_i^c-x^c)^T]\\
&=\frac{1}{h^{p+2}}\Big[\mathbb{E}[\frac{K_{ix}}{\sigma(X_i)}(X_i^c-x^c)(X_i^c-x^c)^T|d_{ix}=0]\cdot P(d_{ix}=0)\\
&\qquad+\mathbb{E}[\frac{K_{ix}}{\sigma(X_i)}(X_i^c-x^c)(X_i^c-x^c)^T|d_{ix}=1]\cdot P(d_{ix}=1)(1+O(\lambda))\Big]\\
&=\frac{1}{h^{p+2}}\Big[\mathbb{E}[\frac{W_{ix}}{\sigma(X_i)}(X_i^c-x^c)(X_i^c-x^c)^T|d_{ix}=0]\cdot P(d_{ix}=0)\\
&\qquad+\lambda\mathbb{E}[\frac{W_{ix}}{\sigma(X_i)}(X_i^c-x^c)(X_i^c-x^c)^T|d_{ix}=1]\cdot P(d_{ix}=1)(1+O(\lambda))\Big]\\
&=\frac{1}{h^2}\int\frac{f(X_i^c,x^d)}{\sigma(X_i^c,x^d)}W\Big(\frac{X_i^c-x^c}{h}\Big)(X_i^c-x^c)(X_i^c-x^c)^T dX_i^c/h^p\\
&\qquad+\frac{1}{h^2}\lambda\sum_{\tilde{x}^d,d_{\tilde{x},x}=1}\int\frac{f(X_i^c,\tilde{x}^d)}{\sigma(X_i^c,\tilde{x}^d)}W\Big(\frac{X_i-x}{h}\Big)(X_i^c-x^c)(X_i^c-x^c)^T dX_i^c/h^p\\
&=I_p\mu_2\frac{f(x)}{\sigma(x)}+O(\lambda+h^2)
\end{aligned}
$$

and

$$
Var\Big[\frac{1}{nh^{p+2}}\sum_{i=1}^{n}\frac{K_{ix}}{\sigma(X_i)}(X_i^c-x^c)(X_i^c-x^c)^T\Big]=\frac{1}{nh^{2p+4}}Var[\frac{K_{ix}}{\sigma(X_i)}(X_i^c-x^c)(X_i^c-x^c)^T]
$$

$$= \frac{1}{nh^{2p+4}}\Big[\mathbb{E}[\Big(\frac{K_{ix}}{\sigma(X_i)}\Big)^2(X_i^c-x^c)(X_i^c-x^c)^T\otimes(X_i^c-x^c)(X_i^c-x^c)^T]+O(h^{2p+4})\Big]$$
$$= \frac{1}{nh^{2p+4}}\Big[\mathbb{E}(\Big(\frac{W_{ix}}{\sigma(X_i)}\Big)^2(X_i^c-x^c)(X_i^c-x^c)^T\otimes(X_i^c-x^c)(X_i^c-x^c)^T|d_{ix}=0)\cdot P(d_{ix}=0)$$
$$+\lambda^2\mathbb{E}(\Big(\frac{W_{ix}}{\sigma(X_i)}\Big)^2(X_i^c-x^c)(X_i^c-x^c)^T\otimes(X_i^c-x^c)(X_i^c-x^c)^T|d_{ix}=1)\cdot P(d_{ix}=1)\Big]+O\Big(\frac{1}{n}\Big)$$
$$= \frac{1}{nh^{p+4}}\Big[\int\frac{f(X_i^c,x^d)}{\sigma^2(X_i^c,x^d)}W^2\Big(\frac{X_i^c-x^c}{h}\Big)(X_i^c-x^c)(X_i^c-x^c)^T\otimes(X_i^c-x^c)(X_i^c-x^c)^T dX_i^c/h^p$$
$$+\lambda^2\sum_{\tilde{x}^d,d_{\tilde{x},x}=1}\int f(X_i^c,\tilde{x}^d)W^2\Big(\frac{X_i^c-x^c}{h}\Big)(X_i^c-x^c)(X_i^c-x^c)^T\otimes(X_i^c-x^c)(X_i^c-x^c)^T dX_i^c/h^p\Big]+O\Big(\frac{1}{n}\Big)$$
$$= \frac{1}{nh^p}\int vv^TK^4(v)dvf(x)\frac{f(x)}{\sigma^2(x)}+o\Big(\frac{1}{nh^p}\Big).$$

Finally, we have

$$\frac{1}{nh^{p+2}}\sum_{i=1}^n K_{ix}(X_i^c-x^c)(X_i^c-x^c)^T = I_p\mu_2\frac{f(x)}{\sigma(x)}+O_p\Big(\lambda+h^2+\frac{1}{\sqrt{nh^p}}\Big).$$

∎

Proof of Lemma 2. To prove the first result, we have

$$\mathbb{E}\Big[\frac{1}{nh^p}\sum_{i=1}^n K_{ix}^2\Big]=\frac{1}{h^p}\mathbb{E}\Big(K_{ix}^2\Big)\approx\frac{1}{h^p}\Big[\mathbb{E}\Big(K_{ix}^2|d_{ix}=0\Big)\cdot P(d_{ix}=0)+\mathbb{E}\Big(K_{ix}^2|d_{ix}=1\Big)\cdot P(d_{ix}=1)\Big]$$
$$=\int f(X_i^c,x^d)W^2\Big(\frac{X_i^c-x^c}{h}\Big)dX_i^c/h^p+\lambda^2\sum_{\tilde{x}^d,d_{\tilde{x},x}=1}\int f(X_i^c,\tilde{x}^d)W^2\Big(\frac{X_i^c-x^c}{h}\Big)dX_i^c/h^p$$
$$=\nu_0 f(x)+O(\lambda^2+h^2)$$

and

$$Var\Big[\frac{1}{nh^p}\sum_{i=1}^n K_{ix}^2\Big]=\frac{1}{nh^{2p}}Var\Big(K_{ix}^2\Big)=\frac{1}{nh^{2p}}[\mathbb{E}(K_{ix}^4)+O(h^{2p})]$$
$$\approx\frac{1}{nh^{2p}}\Big[\mathbb{E}(K_{ix}^4|d_{ix}=0)\cdot P(d_{ix}=0)+\mathbb{E}(K_{ix}^4|d_{ix}=1)\cdot P(d_{ix}=1)+O(h^{2p})\Big]$$
$$=\frac{1}{nh^p}\Big[\int f(X_i^c,x^d)W^4\Big(\frac{X_i^c-x^c}{h}\Big)dX_i^c/h^p+\lambda^4\sum_{\tilde{x}^d,d_{\tilde{x},x}=1}\int f(X_i^c,\tilde{x}^d)W^4\Big(\frac{X_i^c-x^c}{h}\Big)dX_i^c/h^p\Big]+O\Big(\frac{1}{n}\Big)$$
$$=\frac{1}{nh^p}\int K^4(v)dvf(x)+o\Big(\frac{1}{nh^p}\Big).$$

Therefore,

$$\frac{1}{nh^p}\sum_{i=1}^n K_{ix}^2=\nu_0 f(x)+O_p\Big(\lambda^2+h^2+\frac{1}{\sqrt{nh^p}}\Big).$$

Next,

$$
\begin{aligned}
&\mathbb{E}\Big[\frac{1}{nh^{p+1}}\sum_{i=1}^{n}K_{ix}^{2}(X_i^c-x^c)\Big]=\frac{1}{h^{p+1}}\mathbb{E}[K_{ix}^{2}(X_i^c-x^c)]\\
&\approx\frac{1}{h^{p+1}}\Big[\mathbb{E}[W_{ix}^{2}(X_i^c-x^c)|d_{ix}=0]\cdot P(d_{ix}=0)+\lambda^2\mathbb{E}[W_{ix}^{2}(X_i^c-x^c)|d_{ix}=1]\cdot P(d_{ix}=1)\Big]\\
&=\frac{1}{h}\int f(X_i^c,x^d)W^2\Big(\frac{X_i^c-x^c}{h}\Big)(X_i^c-x^c)dX_i^c/h^p\\
&\quad+\frac{1}{h}\lambda^2\sum_{\tilde{x}^d,d_{\tilde{x},x}=1}\int f(X_i^c,\tilde{x}^d)W^2\Big(\frac{X_i^c-x^c}{h}\Big)(X_i^c-x^c)dX_i^c/h^p=h\nu_2\nabla_{x^c}f(x)+O(\lambda^2h+h^3)
\end{aligned}
$$

and

$$
\begin{aligned}
&Var\Big[\frac{1}{nh^{p+1}}\sum_{i=1}^{n}K_{ix}^{2}(X_i^c-x^c)\Big]=\frac{1}{nh^{2p+2}}Var[K_{ix}^{2}(X_i^c-x^c)]\\
&=\frac{1}{nh^{2p+2}}[\mathbb{E}[K_{ix}^{4}(X_i^c-x^c)(X_i^c-x^c)^T]+O(h^{2p+4})]\\
&\approx\frac{1}{nh^{2p+2}}\Big[\mathbb{E}(W_{ix}^{4}(X_i^c-x^c)(X_i^c-x^c)^T|d_{ix}=0)\cdot P(d_{ix}=0)\\
&\quad+\lambda^4\mathbb{E}(W_{ix}^{4}(X_i^c-x^c)(X_i^c-x^c)^T|d_{ix}=1)\cdot P(d_{ix}=1)+O(h^{2p+4})\Big]\\
&=\frac{1}{nh^{p+2}}\Big[\int f(X_i^c,x^d)W^4\Big(\frac{X_i^c-x^c}{h}\Big)(X_i^c-x^c)(X_i^c-x^c)^TdX_i^c/h^p\\
&\quad+\lambda^4\sum_{\tilde{x}^d,d_{\tilde{x},x}=1}\int f(X_i^c,\tilde{x}^d)W^4\Big(\frac{X_i^c-x^c}{h}\Big)(X_i^c-x^c)(X_i^c-x^c)^TdX_i^c/h^p\Big]+O\Big(\frac{h^2}{n}\Big)\\
&=\frac{1}{nh^p}\int vv^TK^4(v)dvf(x)+o\Big(\frac{1}{nh^p}\Big).
\end{aligned}
$$

Therefore, $\frac{1}{nh^{p+1}}\sum_{i=1}^{n}K_{ix}^{2}(X_i^c-x^c)=h\nu_2\nabla_{x^c}f(x)+O_p\Big(\lambda^2h+h^3+\frac{1}{\sqrt{nh^p}}\Big)$. Finally, for the last result, we have

$$
\begin{aligned}
&\mathbb{E}\Big[\frac{1}{nh^{p+2}}\sum_{i=1}^{n}K_{ix}^{2}(X_i^c-x^c)(X_i^c-x^c)^T\Big]=\frac{1}{h^{p+2}}\mathbb{E}[K_{ix}^{2}(X_i^c-x^c)(X_i^c-x^c)^T]\\
&\approx\frac{1}{h^{p+2}}\Big[\mathbb{E}[W_{ix}^{2}(X_i^c-x^c)(X_i^c-x^c)^T|d_{ix}=0]\cdot P(d_{ix}=0)\\
&\quad+\lambda^2\mathbb{E}[W_{ix}^{2}(X_i^c-x^c)(X_i^c-x^c)^T|d_{ix}=1]\cdot P(d_{ix}=1)\Big]\\
&=\frac{1}{h^2}\int f(X_i^c,x^d)W^2\Big(\frac{X_i^c-x^c}{h}\Big)(X_i^c-x^c)(X_i^c-x^c)^TdX_i^c/h^p\\
&\quad+\frac{1}{h^2}\lambda^2\sum_{\tilde{x}^d,d_{\tilde{x},x}=1}\int f(X_i^c,\tilde{x}^d)W^2\Big(\frac{X_i^c-x^c}{h}\Big)(X_i^c-x^c)(X_i^c-x^c)^TdX_i^c/h^p=I_p\nu_2f(x)+O(\lambda^2+h^2)
\end{aligned}
$$

and

$$
\begin{aligned}
&Var\Big[\frac{1}{nh^{p+2}}\sum_{i=1}^{n}K_{ix}^2(X_i^c-x^c)(X_i^c-x^c)^T\Big]=\frac{1}{nh^{2p+4}}Var[K_{ix}^2(X_i^c-x^c)(X_i^c-x^c)^T]\\
&=\frac{1}{nh^{2p+4}}[\mathbb{E}[K_{ix}^4(X_i^c-x^c)(X_i^c-x^c)^T\otimes(X_i^c-x^c)(X_i^c-x^c)^T]+O(h^{2p+4})]\\
&\approx\frac{1}{nh^{2p+4}}\Big[\mathbb{E}(W_{ix}^4(X_i^c-x^c)(X_i^c-x^c)^T\otimes(X_i^c-x^c)(X_i^c-x^c)^T|d_{ix}=0)\cdot P(d_{ix}=0)\\
&\quad+\lambda^4\mathbb{E}(W_{ix}^4(X_i^c-x^c)(X_i^c-x^c)^T\otimes(X_i^c-x^c)(X_i^c-x^c)^T|d_{ix}=1)\cdot P(d_{ix}=1)\Big]+O\Big(\frac{1}{n}\Big)\\
&=\frac{1}{nh^{p+4}}\Big[\int f(X_i^c,x^d)W^4\Big(\frac{X_i^c-x^c}{h}\Big)(X_i^c-x^c)(X_i^c-x^c)^T\otimes(X_i^c-x^c)(X_i^c-x^c)^T dX_i^c/h^p\\
&\quad+\lambda^4\sum_{\tilde{x}^d,d_{\tilde{x},x}=1}\int f(X_i^c,\tilde{x}^d)W^4\Big(\frac{X_i^c-x^c}{h}\Big)(X_i^c-x^c)(X_i^c-x^c)^T\otimes(X_i^c-x^c)(X_i^c-x^c)^T dX_i^c/h^p\Big]+O\Big(\frac{1}{n}\Big)\\
&=\frac{1}{nh^p}\int vv^T\otimes vv^T W^4(v)dv f(x)+o\Big(\frac{1}{nh^p}\Big).
\end{aligned}
$$

Thus $\frac{1}{nh^{p+2}}\sum_{i=1}^n K_{ix}^2(X_i^c-x^c)(X_i^c-x^c)^T=I_p\nu_2 f(x)+O_p\Big(\lambda^2+h^2+\frac{1}{\sqrt{nh^p}}\Big)$. ■

Proof of Lemma 3.

$$
\begin{aligned}
&Var(w_{1k}|X)=\mathbb{E}[w_{1k}^2|X]-0=\frac{1}{nh}\sum_{i=1}^n K_{ix}^2\mathbb{E}[\eta_{i,k}^2|X]\\
&=\frac{1}{nh^p}\sum_{i=1}^n K_{ix}^2\tau_{kk}\xrightarrow{p}\tau_{kk}\nu_0 f(x), k=1,\cdots,q,\\
&Var(w_{21}|X)=\mathbb{E}[w_{21}w_{21}^T|X]-0=\frac{1}{nh^{p+2}}\sum_{i=1}^n\mathbb{E}\Big(K_{ix}^2(X_i^c-x^c)(X_i^c-x^c)^T(\sum_{k=1}^q\eta_{i,k})^2\Big|X\Big)\\
&=\frac{1}{nh^{p+2}}\sum_{i=1}^n K_{ix}^2(X_i^c-x^c)(X_i^c-x^c)^T\sum_{k,k'=1}^q\tau_{kk'}\xrightarrow{p}I_p\nu_2 f(x)\sum_{k,k'=1}^q\tau_{kk'},\\
&Cov(w_{1k},w_{1k'}|X)=\mathbb{E}[w_{1k}\cdot w_{1k'}|X]-0=\frac{1}{nh^p}\sum_{i=1}^n K_{ix}^2\mathbb{E}[\eta_{i,k}\eta_{i,k'}|X]\\
&=\frac{1}{nh^p}\sum_{i=1}^n K_{ix}^2\tau_{kk'}\xrightarrow{p}\tau_{kk'}\nu_0 f(x), k\neq k',\\
&Cov(w_{1k},w_{21}|X)=\mathbb{E}[w_{1k}\cdot w_{21}|X]=\frac{1}{nh^{p+1}}\sum_{i=1}^n K_{ix}^2(X_i^c-x^c)\sum_{k'=1}^q\mathbb{E}[\eta_{i,k}\eta_{i,k'}|X]\\
&=\frac{1}{nh^{p+1}}\sum_{i=1}^n K_{ix}^2(X_i^c-x^c)\sum_{k'=1}^q\tau_{kk'}\xrightarrow{p}0.
\end{aligned}
$$

■

Proof of Lemma 4. Given the definition of $\eta^*_{j,k}$, we have

$$\begin{aligned}\mathbb{E}\Big[\eta^*_{j,k}|X_i, X_j\Big] &= \mathbb{E}\left[I\left(\varepsilon_j \le c_k - \frac{d_{j,k}}{\sigma(X_j)}\right) - \tau_k\right] \\ &= \int_{-\infty}^{c_k - \frac{d_{j,k}}{\sigma(X_j)}} f(v)dv - \tau_k \\ &= F\left(c_k - \frac{d_{j,k}}{\sigma(X_j)}\right) - F(c_k) \\ &= -\frac{d_{j,k}}{\sigma(X_j)} f(c_k) + o(\frac{d_{j,k}}{\sigma(X_j)}).\end{aligned}$$

Next, we analyze the term $-d_{j,k}/\sigma(X_j)$.

When $d_{ji} = 0$, for $d_{j,k}$, we have

$$\begin{aligned}d_{j,k} &= c_k\Big[\sigma(X^c_j, X^d_i) - \sigma(X_i)\Big] \\ &\quad + g(X^c_j, X^d_i) - g(X_i) - \beta(X_i)^T(X^c_j - X^c_i) \\ &= c_k\Big[\sigma'(X_i)^T(X^c_j - X^c_i) + D_1\Big] + D_2,\end{aligned}$$

where D_1 is defined as

$$D_1 = O\Big(\frac{1}{2}(X^c_j - X^c_i)^T \sigma''(X_i)(X^c_j - X^c_i)\Big).$$

Note that we will substitute $\mathbb{E}\big(\eta^*_{j,k}|X_i, X_j\big)$ into $\mathbb{E}\big(h^{-p}W_{ji}L_{ji}\eta^*_{j,k}|X_i\big)$ in Equation (A.23), and the order of D_1 becomes h^2v^2 in integration, which is equivalent to an $O(h^2)$ term if we define $v = (X^c_j - X^c_i)/h$. Similarly, we have $D_2 = O\big(\frac{1}{2}(X^c_j - X^c_i)^T\beta'(X_i)(X^c_j - X^c_i)\big)O\big((X^c_j - X^c_i)^2\big)$, and it will become a smaller order term when combined with the kernel function W. Hereafter, we denote all such terms as *s.o.* for notational simplicity.

For $1/\sigma(X_j)$, we have

$$\frac{1}{\sigma(X_j)} = \frac{1}{\sigma(X^c_j, X^d_i)} = \frac{1}{\sigma(X_i)} - \frac{1}{\sigma^2(X_i)}\sigma'(X_i)(X^c_j - X^c_i) + s.o.$$

In this case, it will be sufficient to keep the first term $\frac{1}{\sigma(X_i)}$ only and we have

$$-\frac{d_{j,k}}{\sigma(X_j)} f(c_k) = C_{1i,k}(X^c_j - X^c_i) + s.o.,$$

where

$$C_{1i,k} = -f(c_k)c_k\sigma'(X_i)^T/\sigma(X_i).$$

When $d_{ij} = 1$, we have

$$\begin{aligned} d_{j,k} &= c_k \left[\sigma(X_i^c, X_j^d) + \sigma'(X_i^c, X_j^d)^T (X_j^c - X_i^c) - \sigma(X_i) + s.o. \right] \\ &\quad + g(X_i^c, X_j^d) + \beta(X_i^c, X_j^d)^T (X_j^c - X_i^c) - g(X_i) - \beta(X_i)^T (X_j^c - X_i^c) \\ &= c_k \left[\sigma(X_i^c, X_j^d) - \sigma(X_i) \right] + g(X_i^c, X_j^d) - g(X_i) + c_k s.o. \\ &\quad + \left[c_k \sigma'(X_i^c, X_j^d)^T + \beta(X_i^c, X_j^d)^T - \beta(X_i)^T \right] (X_j^c - X_i^c). \end{aligned}$$

Next,

$$\begin{aligned} \frac{1}{\sigma(X_j)} &= \frac{1}{\sigma(X_i^c, X_j^d)} - \frac{\sigma'(X_i^c, X_j^d)^T}{\sigma^2(X_i^c, X_j^d)} (X_j^c - X_i^c) + s.o. \\ &= \frac{1}{\sigma(X_i)} + \frac{\sigma(X_i) - \sigma(X_i^c, X_j^d)}{\sigma(X_i^c, X_j^d)\sigma(X_i)} - \frac{\sigma'(X_i^c, X_j^d)^T}{\sigma^2(X_i^c, X_j^d)} (X_j^c - X_i^c) + s.o. \end{aligned} \tag{A.18}$$

It can be shown that the second term in Equation (A.18) is of order λ and the third term in Equation (A.18) is of order h in $\mathbb{E}\left(h^{-p} W_{ji} L_{ji} \eta_{j,k}^* | X_i\right)$, both of which can be omitted so that $1/\sigma(X_j) = 1/\sigma(X_i) + s.o.$ and

$$- \frac{d_{j,k}}{\sigma(X_j)} f(c_k) = C_{2i,k} + C_{3i,k}(X_j^c - X_i^c) + s.o., \tag{A.19}$$

where

$$\begin{aligned} C_{2i,k} &= -f(c_k) \left[c_k \left(\sigma(X_i^c, X_j^d) - \sigma(X_i) \right) + g(X_i^c, X_j^d) - g(X_i) \right] / \sigma(X_i), \\ C_{3i,k} &= -f(c_k) \left[c_k \sigma'(X_i^c, X_j^d)^T + \beta(X_i^c, X_j^d)^T - \beta(X_i)^T \right] / \sigma(X_i). \end{aligned}$$

■

Proof of Lemma 5. In taking expectations of $\eta_{j,k}^* \eta_{j,m}^*$, we assume $c_k - \frac{d_{j,k}}{\sigma(X_j)} \leq c_m - \frac{d_{j,m}}{\sigma(X_j)}$. The results for the case of $c_k - \frac{d_{j,k}}{\sigma(X_j)} \geq c_m - \frac{d_{j,m}}{\sigma(X_j)}$ will be the same.

$$\begin{aligned} \mathbb{E}\left[\eta_{j,k}^* \eta_{j,m}^* | X_i, X_j\right] &= \mathbb{E}\left[1\left(\varepsilon_j \leq c_k - \frac{d_{j,k}}{\sigma(X_j)}\right) 1\left(\varepsilon_j \leq c_m - \frac{d_{j,m}}{\sigma(X_j)}\right) | X_i, X_j \right] \\ &\quad - \tau_m \mathbb{E}\left[1\left(\varepsilon_j \leq c_k - \frac{d_{j,k}}{\sigma(X_j)}\right) | X_i, X_j \right] - \tau_k \mathbb{E}\left[1\left(\varepsilon_j \leq c_m - \frac{d_{j,m}}{\sigma(X_j)}\right) | X_i, X_j \right] + \tau_k \tau_m \\ &= (1 - \tau_m) \mathbb{E}\left[1\left(\varepsilon_j \leq c_k - \frac{d_{j,k}}{\sigma(X_j)}\right) | X_i, X_j \right] - \tau_k \mathbb{E}\left[1\left(\varepsilon_j \leq c_m - \frac{d_{j,m}}{\sigma(X_j)}\right) | X_i, X_j \right] + \tau_k \tau_m. \end{aligned} \tag{A.20}$$

Similar to Lemma 4, we analyze $\mathbb{E}\left[\eta_{j,k}^* \eta_{j,m}^* | X_i, X_j\right]$ under two cases: $d_{ji} = 0$ and $d_{ji} = 1$. Lemma 4 implies

$$E\left(\eta_{j,m}^{*}|X_i,X_j\right)=\begin{cases}C_{1i,m}(X_j^c-X_i^c)+s.o. & \text{if } d_{ji}=0,\\ C_{2i,m}+C_{3i,m}(X_j^c-X_i^c)+s.o. & \text{if } d_{ji}=1,\end{cases} \tag{A.21}$$

where $C_{1i,m}$, $C_{2i,m}$, and $C_{3i,m}$ are defined similarly to $C_{1i,k}$, $C_{2i,k}$, and $C_{3i,k}$ in Lemma 4.

Using both Lemma 4 and Equation (A.21), we obtain the following two results. When $d_{ji}=0$, Equation (A.20) becomes

$$\begin{aligned}\mathbb{E}\left[\eta_{j,k}^{*}\eta_{j,m}^{*}|X_i,X_j\right]&=(1-\tau_m)\left(C_{1i,k}(X_j^c-X_i^c)+\tau_k\right)-\tau_k\left(C_{1i,m}(X_j^c-X_i^c)+\tau_m\right)+\tau_k\tau_m+s.o.\\&=(1-\tau_m)\tau_k+C_{4i,km}(X_j^c-X_i^c)+s.o.,\end{aligned}$$

where

$$C_{4i,km}=(1-\tau_m)C_{1i,k}-\tau_k C_{1i,m}.$$

When $d_{ji}=1$, Equation (A.20) becomes

$$\begin{aligned}\mathbb{E}\left[\eta_{j,k}^{*}\eta_{j,m}^{*}|X_i,X_j\right]&=(1-\tau_m)\left(C_{2i,k}+C_{3i,k}(X_j^c-X_i^c)+\tau_k\right)\\&\quad-\tau_k\left(C_{2i,m}+C_{3i,m}(X_j^c-X_i^c)+\tau_m\right)+\tau_k\tau_m+s.o.\\&=C_{5i,km}+C_{6i,km}(X_j^c-X_i^c)+s.o.,\end{aligned}$$

where

$$\begin{aligned}C_{5i,km}&=(1-\tau_m)\tau_k+(1-\tau_m)C_{2i,k}-\tau_k C_{2i,m}\\C_{6i,km}&=(1-\tau_m)C_{3i,k}-\tau_k C_{3i,m}.\end{aligned}$$

■

Proof of Lemma 6. Rewrite $S_{1,km}$ as

$$\begin{aligned}S_{1,km}&=\frac{1}{n^3h^{2p}}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j\neq l}}^{n}\sum_{l=1}^{n}\gamma_i^2W_{ji}L_{ji}W_{li}L_{li}\eta_{j,k}^{*}\eta_{l,m}^{*}+\frac{1}{n^3h^{2p}}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_i^2W_{ji}^2L_{ji}^2\eta_{j,k}^{*}\eta_{j,m}^{*}\\&=S_{1,km,a}+S_{1,km,b}.\end{aligned}$$

The first term $S_{1,km,a}$ can be written as

$$S_{1,km,a}=\frac{1}{n^3}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j\neq l}}^{n}\sum_{l=1}^{n}H_{1a}(X_i^c,X_j^c,X_l^c),$$

where $H_{1a}(X_i^c,X_j^c,X_l^c)$ is a symmetrized version of $\gamma_i^2W_{ji}L_{ji}W_{li}L_{li}\eta_{j,k}^{*}\eta_{l,m}^{*}$ and is given by

$$
\begin{aligned}
H_{1a}(X_i^c, X_j^c, X_l^c) = \frac{1}{3}(&h^{-2p}\gamma_i^2 W_{ji}L_{ji}W_{li}L_{li}\eta_{j,k}^*\eta_{l,m}^* + h^{-2p}\gamma_j^2 W_{ij}L_{ij}W_{lj}L_{lj}\eta_{i,k}^*\eta_{l,m}^* \\
&+ h^{-2p}\gamma_l^2 W_{jl}L_{jl}W_{il}L_{il}\eta_{i,k}^*\eta_{j,m}^*).
\end{aligned}
$$

We first calculate the expectation of the following term:

$$
\mathbb{E}\Big[H_{1a}(X_i^c, X_j^c, X_l^c)\Big] = \mathbb{E}\Big\{\gamma_i^2 \mathbb{E}\Big[h^{-p}W_{ji}L_{ji}\eta_{j,k}^*|X_i\Big] \cdot \mathbb{E}\Big[h^{-p}W_{li}L_{li}\eta_{l,m}^*|X_i\Big]\Big\}. \tag{A.22}
$$

The calculation of $\mathbb{E}\Big[h^{-1}W_{ji}L_{ji}\eta_{j,k}^*|X_i\Big]$ will suffice.

$$
\begin{aligned}
\mathbb{E}\Big[h^{-p}W_{ji}L_{ji}\eta_{j,k}^*|X_i\Big] &= \mathbb{E}\Big[h^{-p}W_{ji}\eta_{j,k}^*|X_i, d_{ji}=0\Big] \cdot P\Big(d_{ji}=0|X_i\Big) \\
&\quad + \mathbb{E}\Big[h^{-p}W_{ji}\eta_{j,k}^*|X_i, d_{ji}=1\Big] \cdot P\Big(d_{ji}=1|X_i\Big)\lambda \times (1+O(\lambda)) \\
&= C_{1i,k}\int h^{-p}W_{ji}(X_j^c - X_i^c)f(X_j^c, X_i^d)dX_j^c \\
&\quad + \lambda C_{2i,k}\sum_{\tilde{X}^d, d_{\tilde{x}i}=1}\int h^{-p}W_{ji}f(X_j^c|\tilde{X}^d)dX_j^c \times P(d_{\tilde{x}i}=1)\times(1+O(\lambda)) \\
&\quad + \lambda C_{3i,k}\sum_{\tilde{X}^d, d_{\tilde{x}i}=1}\int h^{-p}W_{ji}(X_j^c - X_i^c)f(X_j^c|\tilde{X}^d)dX_j^c \times P(d_{\tilde{x}i}=1)\times(1+O(\lambda)) \\
&= C_{1i,k}\int W(v)hv\Big[f(X_i) + f'(X_i)hv + \frac{1}{2}h^2 v^T f''(X_i)v + o(h^2)\Big]dv \\
&\quad + \lambda C_{2i,k}\sum_{\tilde{X}^d, d_{\tilde{x}i}=1}\int W(v)\Big[f(X_i^c, \tilde{X}^d) + hvf'(X_i^c, \tilde{X}^d) + o(h)\Big]dv \times (1+O(\lambda)) \\
&\quad + \lambda h C_{3i,k}\sum_{\tilde{X}^d, d_{\tilde{x}i}=1}\int vW(v)\Big[f(X_i^c, \tilde{X}^d) + hvf'(X_i^c, \tilde{X}^d) + o(h)\Big]dv \times (1+O(\lambda)) \\
&= A_{1i,k}h^2 + O(h^4) - A_{2i,k}\lambda + O(\lambda h^2),
\end{aligned} \tag{A.23}
$$

where

$$
\begin{aligned}
A_{1i,k} &= C_{1i,k}\mu_2 f(X_i), \\
A_{2i,k} &= -C_{2i,k}\sum_{\tilde{X}^d, d_{\tilde{x}i}=1} f(X_i^c, \tilde{X}_i^d).
\end{aligned}
$$

Similarly,

$$
\mathbb{E}\Big[h^{-p}W_{li}L_{li}\eta_{l,m}^*|X_i\Big] = A_{1i,m}h^2 + O(h^4) - A_{2i,m}\lambda + O(\lambda h^2),
$$

where

$$
\begin{aligned}
A_{1i,m} &= C_{1i,m}\mu_2 f(X_i), \\
A_{2i,m} &= -C_{2i,m}\sum_{\tilde{X}^d, d_{\tilde{x}i}=1} f(X_i^c, \tilde{X}_i^d).
\end{aligned}
$$

Substitute the results for both $\mathbb{E}\left[h^{-p}W_{ji}L_{ji}\eta^*_{j,k}|X_i\right]$ and $\mathbb{E}\left[h^{-p}W_{li}L_{li}\eta^*_{l,m}|X_i\right]$ into Equation (A.22) to get

$$\mathbb{E}\left[H_{1a}(X_i^c, X_j^c, X_l^c)\right] = A^*_{1,km}h^4 - A^*_{2,km}h^2\lambda + A^*_{3,km}\lambda^2 + \tilde{A}^*_{1,km}h^6 + \tilde{A}^*_{2,km}h^4\lambda + \tilde{A}^*_{3,km}h^2\lambda^2 + s.o.,$$

where

$$\begin{aligned} A^*_{1,km} &= \mathbb{E}\left[\gamma_i^2 A_{1i,k}A_{1i,m}\right], \\ A^*_{2,km} &= \mathbb{E}\left[\gamma_i^2\left(A_{1i,k}A_{2i,m} + A_{1i,m}A_{2i,k}\right)\right], \\ A^*_{3,km} &= \mathbb{E}\left[A_{2i,k}A_{2i,m}\right]. \end{aligned}$$

The expressions for $\tilde{A}^*_{1,km}$, $\tilde{A}^*_{2,km}$, and $\tilde{A}^*_{3,km}$ are omitted as they are associated with terms of smaller order and are not used in the proof.

Next, consider $S_{1,km,b}$.

$$S_{1,km,b} = \frac{1}{n}\cdot\frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} H_{1b}(X_i^c, X_j^c),$$

where H_{1b} is a symmetrized version of $h^{-2p}\gamma_i^2 W_{ji}^2 L_{ji}^2\eta^*_{j,k}\eta^*_{j,m}$ and it takes the following form:

$$H_{1b}(X_i^c, X_j^c) = \frac{1}{2}\left(h^{-2p}\gamma_i^2 W_{ji}^2 L_{ji}^2\eta^*_{j,k}\eta^*_{j,m} + h^{-2p}\gamma_j^2 W_{ij}^2 L_{ij}^2\eta^*_{i,k}\eta^*_{i,m}\right).$$

We utilize Lemma 5 to calculate the expectation of $\mathbb{E}\left[H_{1b}(X_i^c, X_j^c)\right]$.

$$\mathbb{E}\left[H_{1b}(X_i^c, X_j^c)\right] = \mathbb{E}\left[\gamma_i^2\mathbb{E}\left[h^{-2p}W_{ji}^2 L_{ji}^2\eta^*_{j,k}\eta^*_{j,m}|X_i\right]\right]$$

and

$$\begin{aligned} \mathbb{E}&\left[h^{-2p}W_{ji}^2L_{ji}^2\eta^*_{j,k}\eta^*_{j,m}|X_i\right] = \mathbb{E}\left[h^{-2p}W_{ji}^2\eta^*_{j,k}\eta^*_{j,m}|X_i, d_{ji}=0\right]\cdot P(d_{ji}=0|X_i) \\ &+ \mathbb{E}\left[h^{-2p}W_{ji}^2\eta^*_{j,k}\eta^*_{j,m}|X_i, d_{ji}=1\right]\cdot P(d_{ji}=1|X_i)\lambda^2\times(1+O(\lambda^2)) \\ &= (1-\tau_m)\tau_k h^{-p}\int h^{-p}W_{ji}^2 f(X_j^c, X_j^d)dX_j^c \\ &+ h^{-p}C_{4i,km}\int h^{-p}W_{ji}^2(X_j^c - X_i^c)f(X_j^c, X_j^d)dX_j^c \\ &+ \lambda^2\sum_{\tilde{X}^d, d_{\tilde{x}i}}\int h^{-p}W_{ji}^2[C_{5i,km} + C_{6i,km}(X_j^c - X_i^c)]f(X_j^c|\tilde{X}^d)dX_j^c\times P(d_{\tilde{x}i}=1)\times(1+O(\lambda^2)) \\ &= (1-\tau_m)\tau_k h^{-p}\int W^2(v)\left[f(X_i) + f'(X_i)hv + O(h^2)\right]dv \\ &+ C_{4i,km}\int W^2(v)v\left[f(X_i) + f'(X_i)hv + O(h^2)\right]dv + O(\lambda^2) \\ &= (1-\tau_m)\tau_k h^{-p}f(X_i)\int W^2(v)dv + s.o. \end{aligned}$$

Hence,

$$\mathbb{E}\Big[H_{1b}(X_i^c, X_j^c)\Big] = A_{4,km}^* h^{-p},$$

where

$$A_{4,km}^* = (1-\tau_m)\tau_k v_0 \mathbb{E}\Big[\gamma_i^2 f(X_i)\Big].$$

Similarly, we can show $\mathbb{E}\Big[H_{1b}(X_i^c, X_j^c)|X_i\Big] = O(h^{-p})$. Hence, by the H-decomposition, we have

$$\begin{aligned} S_{1,km,b} &= n^{-1}\mathbb{E}\Big[H_{1b}(X_i^c, X_j^c)\Big] \\ &\quad + 2n^{-1}\sum_{i=1}^{n}\Big\{\mathbb{E}\Big[H_{1b}(X_i^c, X_j^c)|X_i\Big] - \mathbb{E}\Big[H_{1b}(X_i^c, X_j^c)\Big]\Big\} + s.o. \\ &= A_{4,km}^*(nh^p)^{-1} + n^{-1/2}O((nh^p)^{-1}). \end{aligned}$$

Combining the results for $S_{1,km,a}$ and $S_{1,km,b}$, we have

$$S_{1,km} = A_{1,km}^* h^4 - A_{2,km}^* h^2\lambda + A_{3,km}^*\lambda^2 + A_{4,km}^*(nh^p)^{-1} + \tilde{A}_{1,km}^* h^6 + \tilde{A}_{2,km}^* h^4\lambda + \tilde{A}_{3,km}^* h^2\lambda^2 + s.o.$$

■

Proof of Lemma 7. Write

$$S_{2,k} = n^{-2}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} H_2(X_i^c, X_j^c),$$

where

$$H_2(X_i^c, X_j^c) = \frac{1}{2}\Big(h^{-p}\delta_i W_{ji} L_{ji}\eta_{j,k}^*\varepsilon_i + h^{-p}\delta_j W_{ij} L_{ij}\eta_{i,k}^*\varepsilon_j\Big).$$

Using Equation (A.23) and the independence between ε_i and $\eta_{j,k}^*$, we have

$$\begin{aligned} \mathbb{E}\Big[\delta_i h^{-p} W_{ji} L_{ji}\eta_{j,k}^*\varepsilon_i|X_i\Big] &= \delta_i\varepsilon_i\mathbb{E}\Big[h^{-p} W_{ji} L_{ji}\eta_{j,k}^*|X_i\Big] \\ &= \delta_i\varepsilon_i\Big(A_{1i,k}h^2 - A_{2i,k}\lambda + O(h^4) + O(\lambda h^2)\Big) \\ &= B_{11i,k}h^2 + B_{12i,k}\lambda + O(h^4) + O(\lambda h^2), \end{aligned}$$

where $B_{11i,k} = \delta_i\varepsilon_i A_{1i,k}$ and $B_{12i,k} = \delta_i\varepsilon_i A_{2i,k}$. Using the H-decomposition similar to that for $S_{1,km,b}$ in Lemma 6 and the fact that $\mathbb{E}\Big[H_2(X_i^c, X_j^c)\Big] = 0$, we have

$$S_{2,k} = \frac{2}{n}\sum_{i=1}^{n} B_{11,k}h^2 + \frac{2}{n}\sum_{i=1}^{n} B_{12,k}\lambda + s.o. = B_{1,k}^*\frac{h^2}{\sqrt{n}} + B_{2,k}^*\frac{\lambda}{\sqrt{n}} + s.o.,$$

where

$$B_{1,k}^* = \frac{2}{\sqrt{n}}\sum_{i=1}^{n} B_{11i,k} \text{ and } B_{2,k}^* = \frac{2}{\sqrt{n}}\sum_{i=1}^{n} B_{12i,k},$$

and both $B_{1,k}^*$ and $B_{2,k}^*$ are $O_p(1)$. ■

Proof of Lemma 8. The $o_p(1)$ term in Equation (A.6) results from four different sources throughout the derivation:

1. In the proof of Proposition 2, the use of $F(c_k+z)-F(c_k)=zf(c_k)+o(z)$ gives a smaller order term $o(z)$ that contributes to the $o_p(1)$ term in Equation (A.6).
2. In the proof of Proposition 2, a Taylor series approximation is applied in the second-to-last equality in Equation (A.3),

$$f_\epsilon(c_k-\frac{d_{i,k}}{\sigma(X_i)})=f(c_k)-f'(c_k)\frac{d_{i,k}}{\sigma(X_i)}+o_p(1),$$

   yielding an $o_p(1)$ term.
3. The term $R_{n,k}(\theta)$ adds another $o_p(1)$ term in the proof of Proposition 2.
4. Finally, the use of Equation (A.4) in $L_n(\theta)$ in the proof of Proposition 3 gives another $o_p(1)$ term.

We first analyze the leading term of each of the above four $o_p(1)$ terms and show that, when combined, the $o_p(1)$ term in Equation (A.6) is $O_p(h+\lambda+\frac{1}{\sqrt{nh^p}})$. To simplify notations, assume $p=1$ so that both X_i^c and v are scalars. The result remains the same for the general case when $p>1$. Consider source 1. The second equality in Equation (A.3) can be written as

$$\sum_{k=1}^{q}\mathbb{E}_\varepsilon[B_{n,k}(\theta)|X]=\sum_{k=1}^{q}\sum_{i=1}^{n}\left(K_{ix}\int_0^{\Delta_{i,k}}\left[\frac{z}{\sigma(X_i)}f_\varepsilon\left(c_k-\frac{d_{i,k}}{\sigma(X_i)}\right)+\frac{1}{2}f_\varepsilon'\left(c_k-\frac{d_{i,k}}{\sigma(X_i)}\right)\frac{z^2}{\sigma^2(X_i)}+o(z^2)\right]dz\right),$$

where the leading term of the $o_p(1)$ term is $\sum_{k=1}^{q}\sum_{i=1}^{n}\left(K_{ix}\int_0^{\Delta_{i,k}}\frac{1}{2}f_\varepsilon'\left(c_k-\frac{d_{i,k}}{\sigma(X_i)}\right)\frac{z^2}{\sigma^2(X_i)}dz\right)$. Let $f_{\varepsilon,i}=f_\varepsilon\left(c_k-\frac{d_{i,k}}{\sigma(X_i)}\right)\frac{z^2}{\sigma^2(X_i)}dz\Big)$ and $\tilde{x}_i^c=\frac{X_i^c-x^c}{h}$. A further analysis shows

$$\begin{aligned}
&\sum_{k=1}^{q}\sum_{i=1}^{n}\left(K_{ix}\int_0^{\Delta_{i,k}}\frac{1}{2}f_{\varepsilon,i}'\frac{z^2}{\sigma^2(X_i)}dz\right)\\
&=\frac{1}{6}\sum_{k=1}^{q}\sum_{i=1}^{n}\frac{K_{ix}f_{\varepsilon,i}'}{\sigma^2(X_i)}\Delta_{i,k}^3\\
&=\frac{1}{\sqrt{nh^p}}\frac{1}{6}\sum_{k=1}^{q}\left(u_k,v^T\right)\begin{pmatrix}\frac{1}{nh^p}\sum_{i=1}^{n}\frac{K_{ix}f_{\varepsilon,i}'}{\sigma^2(X_i)}(u_k+v^T\tilde{x}_i^c) & \frac{1}{nh^p}\sum_{i=1}^{n}\frac{K_{ix}f_{\varepsilon,i}'}{\sigma^2(X_i)}(u_k+v^T\tilde{x}_i^c)\tilde{x}_i^{cT}\\ \frac{1}{nh^p}\sum_{i=1}^{n}\frac{K_{ix}f_{\varepsilon,i}'}{\sigma^2(X_i)}(u_k+v^T\tilde{x}_i^c)\tilde{x}_i^c & \frac{1}{nh^p}\sum_{i=1}^{n}\frac{K_{ix}f_{\varepsilon,i}'}{\sigma^2(X_i)}(u_k+v^T\tilde{x}_i^c)\tilde{x}_i^c\tilde{x}_i^{cT}\end{pmatrix}\begin{pmatrix}u_k\\ v\end{pmatrix}\\
&=\frac{1}{\sqrt{nh^p}}\frac{1}{6}\sum_{k=1}^{q}\left(u_k,v^T\right)\begin{pmatrix}O_p(1) & O_p(h)\\ O_p(h) & O_p(1)\end{pmatrix}\begin{pmatrix}u_k\\ v\end{pmatrix}+o_p(1)\\
&=O_p\left(\frac{1}{\sqrt{nh^p}}\right),
\end{aligned}$$

where the probability orders are obtained by a method similar to Lemma 1. Hence the third-to-last equality in Equation (A.3) can be rewritten as

$$\sum_{k=1}^{q}\mathbb{E}_{\varepsilon}[B_{n,k}(\theta)|X] == \sum_{k=1}^{q}\sum_{i=1}^{n}\frac{K_{ix}\Delta_{i,k}^2}{2\sigma(X_i)}f_{\varepsilon,i}+O_p\left(\frac{1}{\sqrt{nh^p}}\right). \tag{A.24}$$

Consider source 2, where an additional $o_p(1)$ term is generated while approximating f_ε in Equation (A.24). Rewrite Equation (A.24) as

$$\begin{aligned}\sum_{k=1}^{q}\mathbb{E}_{\varepsilon}[B_{n,k}(\theta)|X] &= \sum_{k=1}^{q}\sum_{i=1}^{n}\frac{K_{ix}\Delta_{i,k}^2}{2\sigma(X_i)}\left[f_\varepsilon(c_k)-f_\varepsilon'(c_k)\frac{d_{i,k}}{\sigma(X_i)}+\cdots\right]+O_p\left(\frac{1}{\sqrt{nh^p}}\right)\\ &= \sum_{k=1}^{q}\sum_{i=1}^{n}\frac{K_{ix}\Delta_{i,k}^2}{2\sigma(X_i)}\left[f_\varepsilon(c_k)-\frac{f_\varepsilon'(c_k)}{\sigma(X_i)}(c_k(\sigma(X_i)-\sigma(x))+r_i)+\cdots\right]+O_p\left(\frac{1}{\sqrt{nh^p}}\right)\\ &= \sum_{k=1}^{q}\sum_{i=1}^{n}\frac{K_{ix}\Delta_{i,k}^2 f_\varepsilon(c_k)}{2\sigma(X_i)}+\sum_{k=1}^{q}\sum_{i=1}^{n}\frac{K_{ix}\Delta_{i,k}^2 c_k f_\varepsilon'(c_k)}{2\sigma^2(X_i)}\sigma'(x)(X_i^c-x^c)+O_p\left(\frac{1}{\sqrt{nh^p}}\right),\end{aligned}$$

where $\sum_{k=1}^{q}\sum_{i=1}^{n}\frac{K_{ix}\Delta_{i,k}^2 c_k f_\varepsilon'(c_k)}{2\sigma^2(X_i)}\sigma'(x)(X_i^c-x^c)$ is the leading term of the $o_p(1)$ term from source 2 and we can further show that

$$\begin{aligned}&\sum_{k=1}^{q}\sum_{i=1}^{n}\frac{K_{ix}\Delta_{i,k}^2 c_k f_\varepsilon'(c_k)}{2\sigma^2(X_i)}\sigma'(x)(X_i^c-x^c)\\ &=\sum_{k=1}^{q}c_k f_\varepsilon'(c_k)\left(u_k, v^T\right)\\ &\times\begin{pmatrix}\frac{1}{nh^p}\sum_{i=1}^{n}\frac{K_{ix}\sigma'(x)(X_i^c-x^c)}{\sigma^2(X_i)} & \frac{1}{nh^p}\sum_{i=1}^{n}\frac{K_{ix}\sigma'(x)(X_i^c-x^c)}{\sigma^2(X_i)}(X_i^c-x^c)^T\\ \frac{1}{nh^p}\sum_{i=1}^{n}\frac{K_{ix}\sigma'(x)(X_i^c-x^c)}{\sigma^2(X_i)}(X_i^c-x^c) & \frac{1}{nh^p}\sum_{i=1}^{n}\frac{K_{ix}\sigma'(x)(X_i^c-x^c)}{\sigma^2(X_i)}(X_i^c-x^c)(X_i^c-x^c)^T\end{pmatrix}\begin{pmatrix}u_k\\ v\end{pmatrix}\\ &=\sum_{k=1}^{q}c_k f_\varepsilon'(c_k)\left(u_k, v^T\right)\begin{pmatrix}O_p(h^2) & O_p(h)\\ O_p(h) & O_p(h^2)\end{pmatrix}\begin{pmatrix}u_k\\ v\end{pmatrix}=O_p(h),\end{aligned}$$

where the probability orders are obtained by a method similar to that in Lemma 1.

Combined with Equation (A.24), Equation (A.3) can be rewritten as

$$\sum_{k=1}^{q}\mathbb{E}_{\varepsilon}[B_{n,k}(\theta)|X] = \frac{1}{2}\theta^T\left(S_n+O_p(h)\right)\theta+O_p\left(\frac{1}{\sqrt{nh^p}}\right). \tag{A.25}$$

Consider source 3. To explicitly see the probability order of $R_{n,k}(\theta)$, it is helpful to work with $B_{n,k}(\theta)$ first. Assume $\Delta_{i,k}\geq 0$ in Equation (A.1). The symmetric case $\Delta_{i,k}\leq 0$ gives the same result.

$$\begin{aligned}\sum_{k=1}^{q} B_{n,k}(\theta) &= \sum_{k=1}^{q}\sum_{i=1}^{n} K_{ix}\int_0^{\Delta_{i,k}} \text{Prob}\left(c_k - \frac{d_{i,k}}{\sigma(X_i)} \leq \varepsilon_i \leq c_k - \frac{d_{i,k}}{\sigma(X_i)} + \frac{z}{\sigma(X_i)}\right)dz \\ &= \sum_{k=1}^{q}\sum_{i=1}^{n} K_{ix}\int_0^{\Delta_{i,k}} \left[F_\varepsilon\left(c_k - \frac{d_{i,k}}{\sigma(X_i)} + \frac{z}{\sigma(X_i)}\right) - F_\varepsilon\left(c_k - \frac{d_{i,k}}{\sigma(X_i)}\right)\right]dz \\ &= \frac{1}{2}\theta^T S_n \theta + o_p(1),\end{aligned}$$

where the last line follows from the proof in Equation (A.3). We conclude that $\sum_{k=1}^{q} B_{n,k}$ and $\sum_{k=1}^{q} \mathbb{E}_\varepsilon[B_{n,k}(\theta)|X]$ have the same probability order and the leading term of the $o_p(1)$ term in source 3 can also be characterized by $O_p(h)$ and $O_p\left(\frac{1}{\sqrt{nh^p}}\right)$ in Equation (A.25).

Consider source 4. S_n is defined in Section 2. Using Lemma 1, it is easy to see the leading terms of the $o_p(1)$ term in each of the block matrices in Equation (A.4) is $O_p(h^2+\lambda+\frac{1}{\sqrt{nh^p}})$, $O_p(h+\frac{1}{\sqrt{nh^p}})$, and $O_p(h^2+\lambda+\frac{1}{\sqrt{nh^p}})$. Hence, we conclude that the overall leading term is $O_p(h+\lambda+\frac{1}{\sqrt{nh^p}})$.

Given the above analysis for the leading terms from four sources, Equation (A.5) can be written as

$$L_n(\theta) = \frac{1}{2}\frac{f(x)}{\sigma(x)}\theta^T\left(S + O_p(h+\lambda+\frac{1}{\sqrt{nh^p}})\right)\theta + W_n^{*T}\theta + O_p(\frac{1}{\sqrt{nh^p}}), \tag{A.26}$$

where the $O_p(h)$ term from source 2 is combined with the leading term from source 4. Let

$$D = O_p(h+\lambda+\frac{1}{\sqrt{nh^p}}).$$

The first order condition of Equation (A.26) w.r.t. θ gives

$$\hat{\theta}_n = -\frac{\sigma(x)}{f(x)}(S+D)^{-1}W_n^*. \tag{A.27}$$

We note that

$$\begin{aligned}(S+D)^{-1} &= S^{-1} - S^{-1}D\left(I_{q+p} + S^{-1}D\right)^{-1}S^{-1} \\ &= S^{-1} - S^{-1}D\left(I_{q+p} - \left(S^{-1}D\right)^1 + \left(S^{-1}D\right)^2 - \cdots\right)S^{-1} \\ &= S^{-1} - S^{-1}DS^{-1} + o_p(1),\end{aligned}$$

where the second equality follows from the fact that $D \to 0$ as $n \to \infty$ and the sum of the absolute value of each row in $S^{-1}D$ will be less than 1 as $n \to \infty$. Thus the eigenvalues of $S^{-1}D$ are all less than 1 in absolute value and it justifies the expansion in the second equality. This result allows us to write Equation (A.27) as

$$\hat{\theta}_n \approx -\frac{\sigma(x)}{f(x)}S^{-1}W_n^* + \frac{\sigma(x)}{f(x)}S^{-1}DS^{-1}W_n^*. \tag{A.28}$$

Compared with Equation (A.6), Equation (A.28) includes the leading term of the $o_p(1)$ term. Given Equation (A.28), Equation (8) can be written as

$$\begin{pmatrix} \hat{u}_1 \\ \vdots \\ \hat{u}_q \end{pmatrix} = -\frac{\sigma(x)}{f(x)}\left(S^{-1}\right)_{11} W_{1n}^* + \frac{\sigma(x)}{f(x)}\left(S^{-1}DS^{-1}\right)_{11} W_{1n}^*, \tag{A.29}$$

where $\left(S^{-1}DS^{-1}\right)_{11}$ is the upper-left $q \times q$ block matrix in $S^{-1}DS^{-1}$ and it is $O_p(h + \lambda + \frac{1}{\sqrt{nh^p}})$. The remaining analysis essentially repeats that in Section 3 with an extra term on the r.h.s. of Equation (A.29). For example, both $\hat{g}(x)$ in Equation (9) and $\hat{g}_{-i}(X_i)$ in Equation (10) can be updated as

$$\hat{g}(x) = -\frac{1}{nh^p q}\frac{\sigma(x)}{f(x)}\sum_{k=1}^{q}\sum_{i=1}^{n}\frac{K_{ix}\eta_{i,k}^*}{f(c_k)} + \frac{1}{q\sqrt{nh^p}}\frac{\sigma(x)}{f(x)}\iota_q^T\left(S^{-1}DS^{-1}\right)_{11} W_{1n}^* + g(x), \tag{A.30}$$

$$\hat{g}_{-i}(x) = -\frac{1}{nh^p q}\frac{\sigma(x)}{f(x)}\sum_{k=1}^{q}\sum_{\substack{j=1 \\ j\neq i}}^{n}\frac{K_{ji}\eta_{j,k}^*}{f(c_k)} + \frac{1}{q\sqrt{nh^p}}\frac{\sigma(x)}{f(x)}\iota_q^T\left(S^{-1}DS^{-1}\right)_{11} W_{1n,-i}^* + g(x), \tag{A.31}$$

where $W_{1n,-i}$ denotes the variable obtained by excluding X_i. Define

$$G_{1,-i} = -\frac{1}{nh^p q}\frac{\sigma(x)}{f(x)}\sum_{k=1}^{q}\sum_{\substack{j=1 \\ j\neq i}}^{n}\frac{K_{ji}\eta_{j,k}^*}{f(c_k)}, \tag{A.32}$$

$$G_{2,-i} = \frac{1}{q\sqrt{nh^p}}\frac{\sigma(x)}{f(x)}\iota_q^T\left(S^{-1}DS^{-1}\right)_{11} W_{1n,-i}^*. \tag{A.33}$$

Substitute Equation (A.31) into Equation (12) and we have

$$\begin{aligned} CV_1(h,\lambda) &= n^{-1}\sum_{i=1}^{n}(G_{1,-i}+G_{2,-i})^2 + 2n^{-1}\sum_{i=1}^{n}\sigma(X_i)(G_{1,-i}+G_{2,-i}) \\ &= \frac{1}{n}\sum_{i=1}^{n}G_{1,-i}^2 + \frac{2}{n}\sum_{i=1}^{n}G_{1,-i}G_{2,-i} + \frac{1}{n}\sum_{i=1}^{n}G_{2,-i}^2 + \frac{2}{n}\sum_{i=1}^{n}\sigma(X_i)\varepsilon_i G_{1,-i} \\ &\quad + \frac{2}{n}\sum_{i=1}^{n}\sigma(X_i)\varepsilon_i G_{2,-i} \qquad \text{(A.34)} \\ &= G1 + G2 + G3 + G4 + G5, \qquad \text{(A.35)} \end{aligned}$$

where terms $G1$ to $G5$ are defined in the order of the terms in the second equality. Both $G1$ and $G4$ are already analyzed in Equation (12). We will focus on $G2$, $G3$, and $G5$. We will show that the orders of $G2$, $G3$, and $G5$ are smaller than those of $G1$ and $G4$. Hence omitting the $o_p(1)$ term in writing Equation (8) does not affect the cross-validation result.

Consider $G3$. Recall S is $O(1)$ while D is $O_p(h+\lambda+\frac{1}{\sqrt{nh^p}})$ so that $\left(S^{-1}DS^{-1}\right)_{11}$ is also $O_p(h+$

$\lambda + \frac{1}{\sqrt{nh^p}}$).

$$
\begin{aligned}
G3 &= \frac{1}{n}\sum_{i=1}^{n} G_{2,-i}^2 \\
&= \frac{1}{q^2}\frac{1}{n^3h^{2p}}\sum_{i=1}^{n}\gamma_i^2 \iota_q^T \left(S^{-1}DS^{-1}\right)_{11} \\
&\times \begin{pmatrix} \sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{\substack{l=1\\ l\neq i}}^{n} K_{ji}K_{li}\eta_{j,1}^{*}\eta_{l,1}^{*} & \cdots & \sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{\substack{l=1\\ l\neq i}}^{n} K_{ji}K_{li}\eta_{j,1}^{*}\eta_{l,q}^{*} \\ \vdots & \vdots & \vdots \\ \sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{\substack{l=1\\ l\neq i}}^{n} K_{ji}K_{li}\eta_{j,q}^{*}\eta_{l,1}^{*} & \cdots & \sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{\substack{l=1\\ l\neq i}}^{n} K_{ji}K_{li}\eta_{j,q}^{*}\eta_{l,q}^{*} \end{pmatrix} \left(S^{-1}DS^{-1}\right)_{11}^{T} \iota_q. \qquad \text{(A.36)}
\end{aligned}
$$

Proof of the probability orders for terms in Equation (A.36) follows exactly the same steps for $S_{1,km}$ in Equation (15) except for the multiplication of $\left(S^{-1}DS^{-1}\right)_{11}$ and $\left(S^{-1}DS^{-1}\right)_{11}^{T}$. Consequently, $G3$'s contribution to $CV_1(h,\lambda)$ is the result in Lemma 6 multiplied by a factor of $O_p(h+\lambda+\frac{1}{\sqrt{nh^p}})^2$.

Using a similar argument and by noting that $G_{2,-i}$ is equivalent to $G_{1,-i}$ multiplied by an $O_p(h+\lambda+\frac{1}{\sqrt{nh^p}})$ term , we can show $G2$'s contribution to $CV_1(h,\lambda)$ is less than the result in Lemma 6 multiplied by a factor of $O_p(h+\lambda+\frac{1}{\sqrt{nh^p}})$. Similarly, $G5$'s effect on $CV_1(h,\lambda)$ is less than the result in Lemma 7 multiplied by a factor of $O_p(h+\lambda+\frac{1}{\sqrt{nh^p}})$.

To summarize, in writing Equation (8), we omit the $o_p(1)$ in Equation (A.6), the leading term of which is added back in Equation (A.28) and Equation (A.29). The impact of the second term on the r.h.s. of Equation (A.29) on the cross-validation exercise is summarized by $G2$, $G3$, and $G5$ in Equation (A.34). We show that the probability orders of $G1$ and $G4$ dominate those of $G2$, $G3$, and $G5$ in Equation (A.34), which justifies the omission of the $o_p(1)$ term in Equation (8). ■

3 Figures for other distributions

Similar to Figures 1 and 2, we include the results for the Laplace and two mixture normal distributions in Figures 1 to 3. Next, similar to Figures 3 and 4, we attach figures for the Laplace and two mixture normal distributions in Figures 4 to 6.

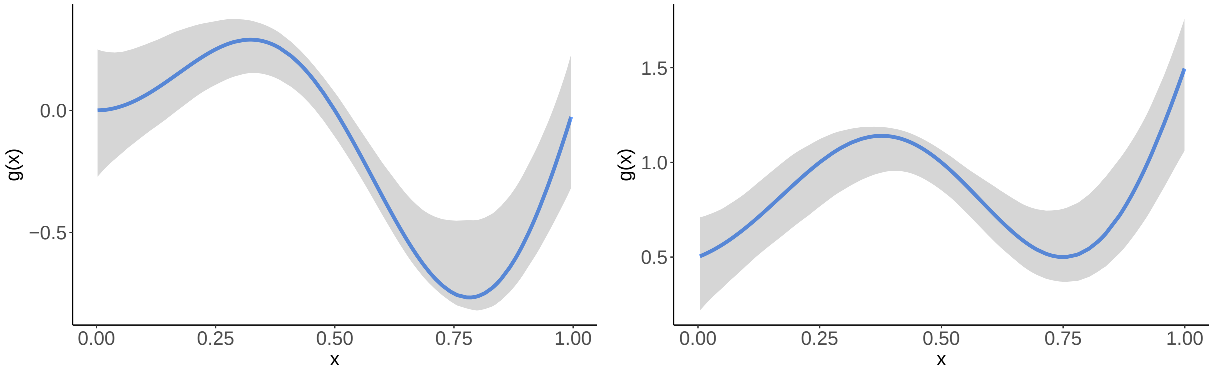

Figure 1: DGP 2 Laplace distribution coverage results. Shaded interval: asymptotic 95% confidence interval, averaged over 1000 replications; solid line: $g(x)$. $X^d = 0$ in the left panel and $X^d = 1$ in the right panel.

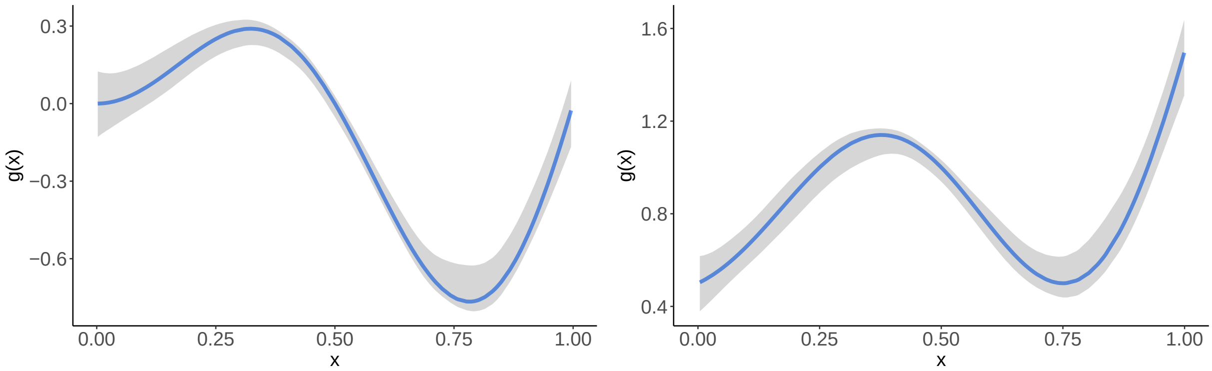

Figure 2: DGP 2 $0.95\mathcal{N}(0,1) + 0.05\mathcal{N}(0,9)$ distribution coverage results. Shaded interval: asymptotic 95% confidence interval, averaged over 1000 replications; solid line: $g(x)$. $X^d = 0$ in the left panel and $X^d = 1$ in the right panel.

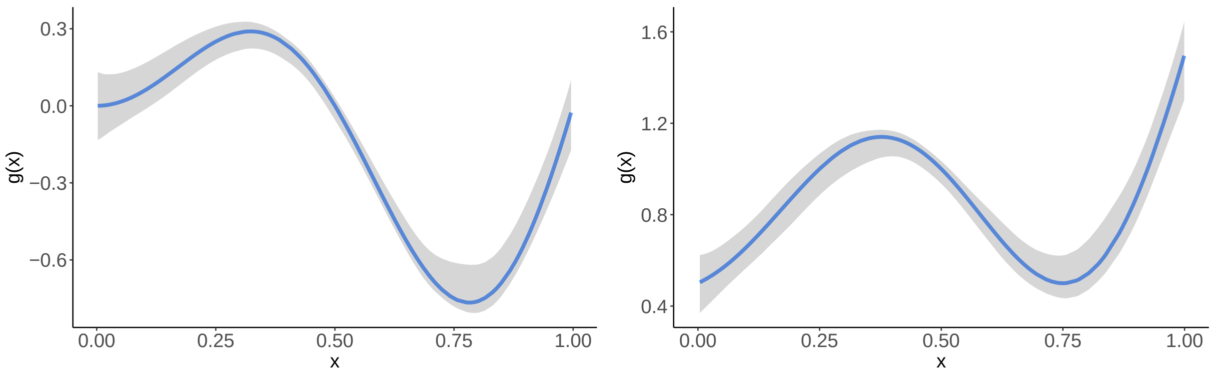

Figure 3: DGP 2 $0.95\mathcal{N}(0,1) + 0.05\mathcal{N}(0,100)$ distribution coverage results. Shaded interval: asymptotic 95% confidence interval, averaged over 1000 replications; solid line: $g(x)$. $X^d = 0$ in the left panel and $X^d = 1$ in the right panel.

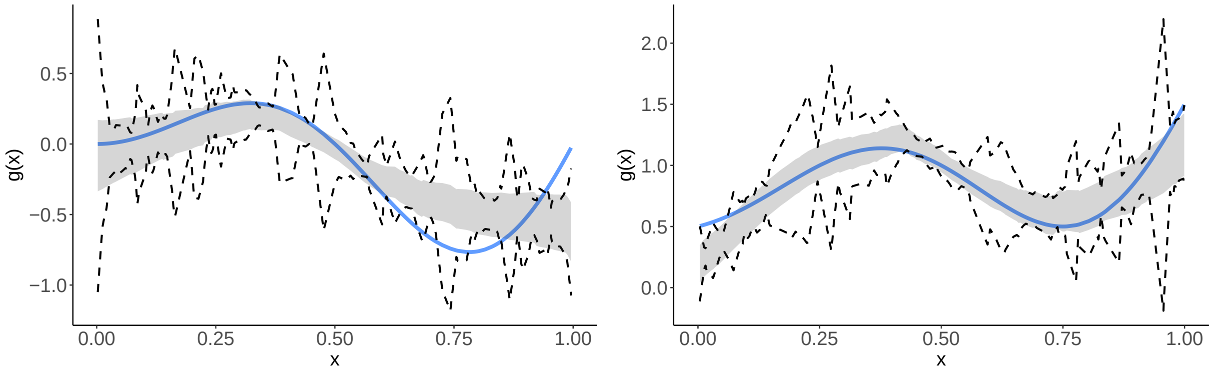

Figure 4: DGP 2 Laplace distribution results. Dashed lines: bootstrap 95% confidence interval; shaded interval: asymptotic 95% confidence interval; solid line: $g(x)$. $X^d = 0$ in the left panel and $X^d = 1$ in the right panel.

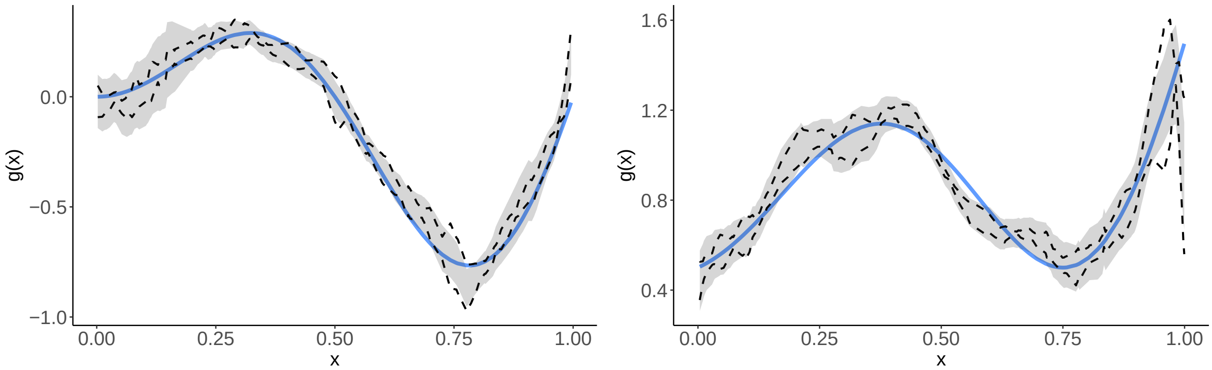

Figure 5: DGP 2 $0.95\mathcal{N}(0,1) + 0.05\mathcal{N}(0,9)$ distribution results. Dashed lines: bootstrap 95% confidence interval; shaded interval: asymptotic 95% confidence interval; solid line: $g(x)$. $X^d = 0$ in the left panel and $X^d = 1$ in the right panel.

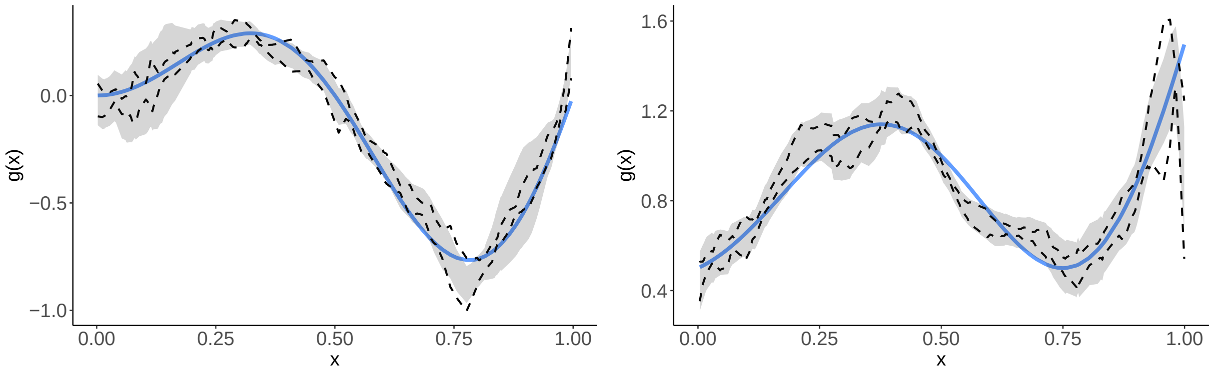

Figure 6: DGP 2 $0.95\mathcal{N}(0,1) + 0.05\mathcal{N}(0,100)$ distribution results. Dashed lines: bootstrap 95% confidence interval; shaded interval: asymptotic 95% confidence interval; solid line: $g(x)$. $X^d = 0$ in the left panel and $X^d = 1$ in the right panel.