Supplementary Material on "Adaptive Nonparametric Regression with Conditional Heteroskedasticity"

Sainan Jin[a], Liangjun Su[a], Zhijie Xiao[b]

[a] *School of Economics, Singapore Management University*

[b] *Department of Economics, Boston College*

This appendix provides proofs for all technical lemmas in the above paper.

C Proofs of the Technical Lemmas

To facilitate the proof, we define an $N \times N$ matrix $M_n(x)$ and $N \times 1$ vectors $\Psi_{s,n}(x)$ $(s = 1, 2)$ as:

$$M_n(x) \equiv \begin{bmatrix} M_{n,0,0}(x) & M_{n,0,1}(x) & \cdots & M_{n,0,p}(x) \\ M_{n,1,0}(x) & M_{n,1,1}(x) & \cdots & M_{n,1,p}(x) \\ \vdots & \vdots & \ddots & \vdots \\ M_{n,p,0}(x) & M_{n,p,1}(x) & \cdots & M_{t,p,p}(x) \end{bmatrix}, \ \Psi_{s,n}(x) \equiv \begin{bmatrix} \Psi_{s,n,0}(x) \\ \Psi_{s,n,1}(x) \\ \vdots \\ \Psi_{s,n,p}(x) \end{bmatrix}, \tag{C.1}$$

where $M_{n,|j|,|k|}(x)$ is an $N_{|j|} \times N_{|k|}$ submatrix with the (l, r) element given by

$$\left[M_{n,|j|,|k|}(x)\right]_{l,r} \equiv \frac{1}{nh^d} \sum_{i=1}^n \left(\frac{X_i - x}{h_1}\right)^{\phi_{|j|}(l)+\phi_{|k|}(r)} K\left(\frac{X_i - x}{h_1}\right),$$

$\Psi_{1,n,|j|}(x)$ is an $N_{|j|} \times 1$ subvector whose r-th element is given by

$$\left[\Psi_{1,n,|j|}(x)\right]_r \equiv \frac{1}{nh^d} \sum_{i=1}^n \left(\frac{X_i - x}{h_1}\right)^{\phi_{|j|}(r)} K\left(\frac{X_i - x}{h_1}\right) Y_i,$$

and $\Psi_{2,n,|j|}(x)$ is an $N_{|j|} \times 1$ subvector whose r-th element is given by

$$\left[\Psi_{2,n,|j|}(x)\right]_r \equiv \frac{1}{nh^d} \sum_{i=1}^n \left(\frac{X_i - x}{h_1}\right)^{\phi_{|j|}(r)} K\left(\frac{X_i - x}{h_1}\right) u_i^2.$$

Define $\widetilde{\Psi}_{2,n}(x)$ analogously as $\Psi_{2,n}(x)$ with u_i^2 being replaced by $\tilde{u}_i^2$, where $\tilde{u}_i \equiv Y_i - \tilde{m}(X_i)$. The p-th order local polynomial estimates of $m(x)$ and $\sigma^2(x)$ are given respectively by

$$\tilde{m}(x) = e_1^\top M_n^{-1}(x)\, \Psi_{1,n}(x) \text{ and } \tilde{\sigma}^2(x) = e_1^\top M_n^{-1}(x)\, \widetilde{\Psi}_{2,n}(x).$$

For $s=1,2$, let

$$U_{s,n}\left(x\right)\equiv\left[\begin{array}{c}U_{s,n,0}(x)\\ U_{s,n,1}(x)\\ \vdots\\ U_{1,n,p}(x)\end{array}\right],\ B_{s,n}\left(x\right)\equiv\left[\begin{array}{c}B_{s,n,0}(x)\\ B_{s,n,1}(x)\\ \vdots\\ B_{s,n,p}(x)\end{array}\right],$$

where $U_{s,n,l}\left(x\right)$ and $B_{s,n,l}\left(x\right)$ are defined analogously as $\Psi_{s,n,l}(x)$ so that $U_{s,n,|j|}\left(x\right)$ and $B_{s,n,|j|}\left(x\right)$ are $N_{|j|}\times 1$ subvectors whose r-th elements are given by

$$\begin{aligned}\left[U_{s,n,|j|}\left(x\right)\right]_r &= \frac{1}{nh_1^d}\sum_{i=1}^{n}\left(\frac{X_i-x}{h_1}\right)^{\phi_{|j|}(r)}K\left(\frac{X_i-x}{h_1}\right)u_{s,i},\\ \left[B_{s,n,|j|}\left(x\right)\right]_r &= \frac{1}{nh_1^d}\sum_{i=1}^{n}\left(\frac{X_i-x}{h_1}\right)^{\phi_{|j|}(r)}K\left(\frac{X_i-x}{h_1}\right)\Delta_{s,i}\left(x\right),\end{aligned}$$

where $u_{1,i}\equiv u_i$, $u_{2,i}\equiv u_i^2-E(u_i^2|X_i)=\sigma^2\left(X_i\right)\left(\varepsilon_i^2-1\right)$, and $\Delta_{s,i}\left(x\right)\equiv\beta_s\left(X_i\right)-\sum_{0\le|\mathbf{j}|\le p}\beta_{s\mathbf{j}}\left(x\right)$ $\times\left(X_i-x\right)^{\mathbf{j}}$. We further define $\tilde{U}_{2,n}\left(x\right)$ analogously as $U_{2,n}\left(x\right)$ but with $u_{2,i}$ being replaced by $\tilde{u}_{2,i}\equiv$ $\tilde{u}_i^2-E\left(u_i^2|X_i\right)$. Then

$$\begin{aligned}\tilde{m}(x)-m\left(x\right) &= e_1^{\top}M_n^{-1}(x)U_{1,n}\left(x\right)+e_1^{\top}M_n^{-1}(x)B_{1,n}\left(x\right),\text{ and}\\ \tilde{\sigma}^2(x)-\sigma^2\left(x\right) &= e_1^{\top}M_n^{-1}(x)\tilde{U}_{2,n}\left(x\right)+e_1^{\top}M_n^{-1}(x)B_{2,n}\left(x\right).\end{aligned}\tag{C.2}$$

By Masry 1996(a), we can readily show that

$$\tilde{m}(x)-m\left(x\right)=e_1^{\top}\left[f_X\left(x\right)M\right]^{-1}\frac{1}{n}\sum_{i=1}^{n}K_{h_1}\left(x-X_i\right)\mathbf{Z}_iu_i+h_1^{p+1}e_1^{\top}M^{-1}B\mathbf{m}^{(p+1)}\left(x\right)+o_p(h_1^{p+1})\tag{C.3}$$

uniformly in x. Furthermore,

$$\sup_{x\in\mathcal{X}}\left|M_n(x)-f_X\left(x\right)M\right|=O_p\left(\upsilon_{0n}\right)\text{ and }\sup_{x\in\mathcal{X}}\left|\tilde{m}(x)-m\left(x\right)\right|=O_p\left(\upsilon_{1n}\right).\tag{C.4}$$

The following lemma studies the asymptotic property of the local polynomial estimator $\tilde{\sigma}^2(x)$ of $\sigma^2(x)$.

Lemma C.1 *Suppose Assumptions A1-A5 hold. Then* $\tilde{\sigma}^2(x)-\sigma^2(x)=e_1^{\top}M_n^{-1}(x)U_{2,n}\left(x\right)+e_1^{\top}M_n^{-1}(x)$ $\times B_{2,n}\left(x\right)+O_p\left((\upsilon_{0n}+\upsilon_{1n})\upsilon_{1n}\right)$ *uniformly in* x.

Proof of Lemma C.1. Let $K^*\left(X_i,x\right)\equiv e_1^{\top}M_n^{-1}(x)K\left(\left(X_i-x\right)/h_1\right)\tilde{\mathbf{Z}}_i$. Then $\tilde{\sigma}^2(x)=(nh_1^d)^{-1}\sum_{i=1}^{n}K^*$ $(X_i,x)\tilde{u}_i^2$. It follows from $M_n^{-1}(x)M_n(x)=I_N$ that

$$\frac{1}{nh_1^d}\sum_{i=1}^{n}K^*\left(X_i,x\right)=\frac{1}{nh_1^d}\sum_{i=1}^{n}e_1^{\top}M_n^{-1}(x)K\left(\left(X_i-x\right)/h_1\right)\tilde{\mathbf{Z}}_i=1,$$

and

$$\frac{1}{nh_1^d}\sum_{i=1}^{n}K^*\left(X_i,x\right)\left(X_i-x\right)^{\mathbf{j}}=\frac{1}{nh_1^d}\sum_{i=1}^{n}e_1^{\top}M_n^{-1}(x)K\left(\left(X_i-x\right)/h_1\right)\tilde{\mathbf{Z}}_i\left(X_i-x\right)^{\mathbf{j}}=0,$$

for $1 \leq |\mathbf{j}| \leq p$. Consequently,

$$\tilde{\sigma}^2(x) - \sigma^2(x) = e_1^\top M_n^{-1}(x)\widetilde{\Psi}_{2,n}(x) = \frac{1}{nh_1^d}\sum_{i=1}^n K^*(X_i, x)\left\{\tilde{u}_i^2 - \overline{\sigma}^2(x, X_i)\right\},$$

where $\overline{\sigma}^2(x, X_i) \equiv \sum_{0\leq|\mathbf{j}|\leq p}\left(D^{(\mathbf{j})}\sigma^2\right)(x)(X_i - x)^{\mathbf{j}}$. Noting that $\tilde{u}_i^2 = [Y_i - \tilde{m}(X_i)]^2 = [\sigma(X_i)\varepsilon_i + m(X_i) - \tilde{m}(X_i)]^2 = \sigma^2(X_i)\varepsilon_i^2 + 2\sigma(X_i)\varepsilon_i[m(X_i) - \tilde{m}(X_i)] + [m(X_i) - \tilde{m}(X_i)]^2$, we have

$$\begin{aligned}
\tilde{\sigma}^2(x) - \sigma^2(x) &= \frac{1}{nh_1^d}\sum_{i=1}^n K^*(X_i, x)\left\{\sigma^2(X_i) - \overline{\sigma}^2(x, X_i)\right\} \\
&\quad + \frac{1}{nh_1^d}\sum_{i=1}^n K^*(X_i, x)\,\sigma^2(X_i)\left(\varepsilon_i^2 - 1\right) \\
&\quad + \frac{2}{nh_1^d}\sum_{i=1}^n K^*(X_i, x)\,\sigma(X_i)\varepsilon_i\left[m(X_i) - \tilde{m}(X_i)\right] \\
&\quad + \frac{1}{nh_1^d}\sum_{i=1}^n K^*(X_i, x)\left[m(X_i) - \tilde{m}(X_i)\right]^2 \\
&\equiv A_1(x) + A_2(x) + 2A_3(x) + A_4(x), \text{ say.}
\end{aligned}$$

Noting that $\Delta_{2,i}(x) = \sigma^2(X_i) - \overline{\sigma}^2(x, X_i)$, we have $A_1(x) = e_1^\top M_n^{-1}(x)B_{2,n}(x)$. In addition $A_2(x) = e_1^\top M_n^{-1}(x)U_{2,n}(x)$ by the definition of $u_{2,i}$, and $\sup_{x\in\mathcal{X}}|A_4(x)| = \upsilon_{1n}^2$ by (C.4). For $A_3(x)$, write $-A_3(x) = A_{31}(x) + A_{32}(x)$, where

$$\begin{aligned}
A_{31}(x) &\equiv \frac{1}{nh_1^d}\sum_{i=1}^n K^*(X_i, x)\,u_i e_1^\top M_n^{-1}(X_i)U_{1,n}(X_i), \text{ and} \\
A_{32}(x) &\equiv \frac{1}{nh_1^d}\sum_{i=1}^n K^*(X_i, x)\,u_i e_1^\top M_n^{-1}(X_i)B_{1,n}(X_i).
\end{aligned}$$

Note that

$$\begin{aligned}
A_{31}(x) &= \frac{1}{nh_1^d}\sum_{i=1}^n K^*(X_i, x)\,u_i e_1^\top\left[Mf_X(X_i)\right]^{-1}U_{1,n}(X_i) \\
&\quad - \frac{1}{nh_1^d}\sum_{i=1}^n K^*(X_i, x)\,u_i e_1^\top\left\{M_n(x)^{-1} - \left[Mf_X(X_i)\right]^{-1}\right\}U_{1,n}(X_i) \\
&\equiv A_{31,1}(x) - A_{31,2}(x), \text{ say.}
\end{aligned}$$

We dispose $A_{31,2}(x)$ first. By (C.4), the facts that $\sup_{x\in\mathcal{X}}\|U_{1,n}(x)\| = O_p(n^{-1/2}h_1^{-d/2}\sqrt{\log n})$ and $\sup_{x\in\mathcal{X}}\frac{1}{nh_1^d}\sum_{i=1}^n|K^*(X_i, x)\,u_i| = O_p(1)$, we have

$$\begin{aligned}
\sup_{x\in\mathcal{X}}|A_{31,2}(x)| &\leq \sup_{x\in\mathcal{X}}\left\|M_n(x)^{-1} - \left[Mf_X(X_i)\right]^{-1}\right\|\sup_{x\in\mathcal{X}}\|U_{1,n}(x)\|\sup_{x\in\mathcal{X}}\frac{1}{nh_1^d}\sum_{i=1}^n|K^*(X_i, x)\,u_i| \\
&= O_p(\upsilon_{0n})O_p(n^{-1/2}h_1^{-d/2}\sqrt{\log n})O_p(1) = O_p(\upsilon_{0n}n^{-1/2}h_1^{-d/2}\sqrt{\log n}).
\end{aligned}$$

Using $U_{1,n}(x) = \frac{1}{nh_1^d}\sum_{j=1}^n K((X_j - x)/h_1)\tilde{\mathbf{Z}}_j u_j$ and $K^*(X_i, x) = e_1^\top M_n^{-1}(x)K((X_i - x)/h_1)\tilde{\mathbf{Z}}_i$, we have

$$
\begin{aligned}
A_{31,1}(x) &= \frac{1}{n^2 h_1^{2d}} e_1^{\top} M_n^{-1}(x) \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\left(X_i - x\right)/h_1\right) \tilde{\mathbf{Z}}_i e_1^{\top} \left[M f_X\left(X_i\right)\right]^{-1} K\left(\left(X_j - X_j\right)/h_1\right) \tilde{\mathbf{Z}}_j u_i u_j \\
&= \frac{1}{n^2 h_1^{2d}} e_1^{\top} M_n^{-1}(x) \sum_{1 \le i \ne j \le n} K\left(\left(X_i - x\right)/h_1\right) \tilde{\mathbf{Z}}_i e_1^{\top} \left[M f_X\left(X_i\right)\right]^{-1} K\left(\left(X_j - X_j\right)/h_1\right) \tilde{\mathbf{Z}}_j u_i u_j \\
&\quad + \frac{1}{n^2 h_1^{2d}} e_1^{\top} M_n^{-1}(x) \sum_{i=1}^{n} K\left(\left(X_i - x\right)/h_1\right) \tilde{\mathbf{Z}}_i e_1^{\top} \left[M f_X\left(X_i\right)\right]^{-1} K(0) \tilde{\mathbf{Z}}_i u_i^2 \\
&\equiv A_{31,1a}(x) + A_{31,1b}(x), \text{ say.}
\end{aligned}
$$

Let $\varsigma_{ij}(x) \equiv \{e_1^{\top} \left[M f_X(x)\right]^{-1} K\left(\left(X_i - x\right)/h_1\right) \tilde{\mathbf{Z}}_i\}\{e_1^{\top} \left[M f_X\left(X_i\right)\right]^{-1} K\left(\left(X_j - X_i\right)/h_1\right) \tilde{\mathbf{Z}}_j\} u_i u_j$. Then by (C.4), $A_{31,1a}(x) = [1 + O_p(\upsilon_{0n})] \bar{A}_{31,1a}(x)$, where

$$
\bar{A}_{31,1a}(x) = \frac{1}{n^2 h_1^{2d}} \sum_{1 \le i \ne j \le n} \varsigma_{ij}(x).
$$

is a second order degenerate U-statistic. We can readily show that $\bar{A}_{31,1a}(x) = O_p\left(n^{-1} h_1^{-d}\right)$ for each x by Chebyshev inequality. By using Bickel's (1975) standard chaining argument, we can show $\sup_{x \in \mathcal{X}} \left|\bar{A}_{31,1a}(x)\right| = O_p\left(n^{-1} h_1^{-d} \log n\right)$. For $A_{31,1b}(x)$, we have

$$
\begin{aligned}
\sup_{x \in \mathcal{X}} \left|A_{31,1b}(x)\right| &\le \frac{1}{n h_1^d} \sup_{x \in \mathcal{X}} \left\|M_n^{-1}(x)\right\| \sup_{x \in \mathcal{X}} \left\| \frac{1}{n h_1^d} \sum_{i=1}^{n} K\left(\left(X_i - x\right)/h_1\right) \tilde{\mathbf{Z}}_i e_1^{\top} \left[M f_X\left(X_i\right)\right]^{-1} K(0) \tilde{\mathbf{Z}}_i u_i^2 \right\| \\
&= O_p\left(n^{-1} h_1^{-d}\right) O_p(1) O_p(1) = O_p\left(n^{-1} h_1^{-d}\right).
\end{aligned}
$$

It follows that $\sup_{x \in \mathcal{X}} \left|A_{31,1}(x)\right| = O_p\left(n^{-1} h_1^{-d} \log n\right)$. Consequently, we have shown that $\sup_{x \in \mathcal{X}} \left|A_{31}(x)\right| = O_p(\upsilon_{0n} n^{-1/2} h_1^{-d/2} \sqrt{\log n})$.

Note that

$$
\begin{aligned}
A_{32}(x) &= \frac{1}{n h_1^d} \sum_{i=1}^{n} K^*\left(X_i, x\right) u_i e_1^{\top} \left[M f_X\left(X_i\right)\right]^{-1} B_{1,n}\left(X_i\right) \\
&\quad - \frac{1}{n h_1^d} \sum_{i=1}^{n} K^*\left(X_i, x\right) u_i e_1^{\top} \left\{M_n(x)^{-1} - \left[M f_X\left(X_i\right)\right]^{-1}\right\} B_{1,n}\left(X_i\right) \\
&\equiv A_{32,1}(x) - A_{32,2}(x), \text{ say.}
\end{aligned}
$$

As in the study of $A_{31}(x)$, using (C.4) and the fact that $\sup_{x \in \mathcal{X}} \left|B_{1,n}(x)\right| = O_p(h_1^{p+1})$ we can readily show that $\sup_{x \in \mathcal{X}} \left|A_{32,2}(x)\right| = O_p(\upsilon_{0n} h_1^{p+1})$ and that $\sup_{x \in \mathcal{X}} \left|A_{32,2}(x)\right| = O_p(n^{-1/2} h_1^{-d/2} \sqrt{\log n} h_1^{p+1})$. Hence $\sup_{x \in \mathcal{X}} \left|A_{32}(x)\right| = O_p(\upsilon_{0n} h_1^{p+1})$. Consequently, $\sup_{x \in \mathcal{X}} \left|A_3(x)\right| = O_p(\upsilon_{0n} \upsilon_{1n})$. This completes the proof. ■

Remark C.1. Using the notation defined in the proof of Lemma C.1, we can also show that $A_1(x) = h_1^{p+1} e_1^{\top} M^{-1} B \boldsymbol{\sigma}^{2(p+1)}(x) + o_p(h_1^{p+1})$, and $\sqrt{n h_1^d} A_2(x) \xrightarrow{d} N(0, \left(\sigma^4(x)/f_X(x)\right) E\left(\varepsilon_1^2 - 1\right)^2 e_1^{\top} M^{-1} \Gamma M^{-1} e_1^{\top})$. By standard results on local polynomial estimators, Lemma A.1 implies

$$
\sup_{x \in \mathcal{X}} \left|\tilde{\sigma}^2(x) - \sigma^2(x)\right| = O_p(\upsilon_{1n}), \tag{C.5}
$$

where υ_{1n} is the rate we can obtain even if the conditional mean function $m(x)$ is known.

Let δ_i and $v_{ri}(x)$ be as defined in Appendix A. To prove Lemmas A.1-A.2, we will frequently use the facts that

$$\delta_i = O_p\left(h^{p+1}\right) \text{ uniformly on the set } \{K_{ix} > 0\}, \tag{C.6}$$

$$v_{ri}(x) = O_p\left(\left(h_1^{p+1} + n^{-1/2} h_1^{-d/2}\right)\left(1 + (h/h_1)^p\right)\right) \text{ on the set } \{K_{ix} > 0\}, \ r = 1, 2, \tag{C.7}$$

$$\max_{\{K_{ix}>0\}} |v_{ri}(x)| = O_p(\upsilon_{2n}), \ r = 1, 2. \tag{C.8}$$

To facilitate the asymptotic analysis, we also define the kernel density and derivative estimator based on the unobserved errors $\{\varepsilon_j\}$:

$$\overline{f}_i(e_i) = \frac{1}{nh_0} \sum_{j \neq i} k_0\left(\frac{e_i - \varepsilon_j}{h_0}\right), \text{ and } \overline{f}_i^{(s)}(e_i) = \frac{1}{nh_0^{1+s}} \sum_{j \neq i} k_0^{(s)}\left(\frac{e_i - \varepsilon_j}{h_0}\right) \text{ for } s = 1, 2, 3.$$

We will need the result in the following lemma which is adopted from Hansen (2008).

Lemma C.2 *Let ε_i, $i = 1, ..., n$, be IID. Assume that (i) the PDF of ε_i, $f(\cdot)$, is uniformly bounded, and the $(p+1)$th derivative of $f^{(s)}(\varepsilon)$ is uniformly continuous; (ii) there exists $q > 0$ such that $\sup_\varepsilon |\varepsilon|^q f(\varepsilon) < \infty$ and $|k_0^{(s)}(e)| \leq C|e|^{-q}$ for $|e|$ large; (iii) $k_0(\cdot)$ is a (p+1)th order kernel and $\int |e|^{p+s+1} |k_0(e)| de < \infty$; (iv) $h_0 \to 0$ and $nh_0^{1+2s}/\log n \to \infty$ as $n \to \infty$. Then*

$$\max_{1 \leq i \leq n} \left| \overline{f}_i^{(s)}(\bar{\varepsilon}_i) - f^{(s)}(\bar{\varepsilon}_i) \right| = O_p(h_0^{p+1} + n^{-1/2} h_0^{-1/2-s} \sqrt{\log n}).$$

Proof of Lemma C.2. The above result is essentially a special case of Theorem 6 in Hansen (2008) who allows for strong mixing processes. For an IID sequence, the parameters β and θ in Hansen (2008) correspond to ∞ and one, respectively. Another noticeable difference is that Hansen considers the usual kernel estimates whereas we consider the leave-one-out kernel estimates here. The difference between these two kernel estimates is uniformly $(nh_0^{1+s})^{-1} k_0^{(s)}(0)$, which is $o(n^{-1/2} h_0^{-1/2-s} \sqrt{\log n})$ under condition (iv) and thus does not contribute to the uniform convergence rate of $\overline{f}_i^{(s)}(\bar{\varepsilon}_i) - f^{(s)}(\bar{\varepsilon}_i)$ to 0. ■

Proof of Lemma A.1. We only prove the lemma with $s = 0$ as the other cases can be treated analogously. Write $\tilde{f}_i(\overrightarrow{\varepsilon}_i) - f(\bar{\varepsilon}_i) = [\overline{f}(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)] + [\tilde{f}_i(\overrightarrow{\varepsilon}_i) - \overline{f}(\bar{\varepsilon}_i)]$. Noting that k_0 is a $(p+1)$-th order kernel with compact support by Assumption A6, the conditions on the kernel in Lemma C.2 are satisfied. One can readily check that the other conditions in that lemma are also satisfied under Assumptions A1, A2, and A7. So we can apply Lemma C.2 to obtain $\max_{1 \leq i \leq n} \left|\overline{f}_i(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)\right| = O_p(h_0^{p+1} + n^{-1/2} h_0^{-1/2} \sqrt{\log n})$. Let

$$\begin{aligned} r_{1ij} \equiv\ & \frac{\bar{\varepsilon}_i \left[\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{1/2} - \varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}\right]}{\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}} - \frac{v_{1i}(x)}{\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}} + \frac{\tilde{m}(X_j) - m(X_j)}{\sigma(X_j)} \\ & + \left[\varepsilon_j + \frac{m(X_j) - \tilde{m}(X_j)}{\sigma(X_j)}\right] \frac{\tilde{\sigma}(X_j) - \sigma(X_j)}{\tilde{\sigma}(X_j)}. \end{aligned} \tag{C.9}$$

Then

$$\overrightarrow{\varepsilon}_i - \tilde{\varepsilon}_j = (\bar{\varepsilon}_i - \varepsilon_j) + r_{1ij}. \tag{C.10}$$

By a first order Taylor expansion with an integral remainder, we have

$$\begin{aligned}
\tilde{f}_i(\overrightarrow{\varepsilon}_i) - \overline{f}(\bar{\varepsilon}_i) &= \frac{1}{nh_0}\sum_{j\neq i}\left[k_0\left(\frac{\overrightarrow{\varepsilon}_i - \tilde{\varepsilon}_j}{h_0}\right) - k_0\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\right] \\
&= \frac{-1}{nh_0^2}\sum_{j\neq i}k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\frac{v_{1i}(x)}{\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}} \\
&\quad + \frac{1}{nh_0^2}\sum_{j\neq i}k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\bar{\varepsilon}_i\left[\varphi(P_i(\boldsymbol{\beta}_2^0))^{1/2} - \varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}\right]\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{-1/2} \\
&\quad + \frac{1}{nh_0^2}\sum_{j\neq i}k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\frac{\tilde{m}(X_j) - m(X_j)}{\sigma(X_j)} \\
&\quad + \frac{1}{nh_0^2}\sum_{j\neq i}k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\left[\varepsilon_j + \frac{m(X_j) - \tilde{m}(X_j)}{\sigma(X_j)}\right]\frac{\tilde{\sigma}(X_j) - \sigma(X_j)}{\tilde{\sigma}(X_j)} \\
&\quad + \frac{1}{nh_0^2}\sum_{j\neq i}\int_0^1\left[k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j + wr_{1ij}}{h_0}\right) - k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\right]dwr_{1ij} \\
&\equiv -B_{1i}(x) + B_{2i}(x) + B_{3i}(x) + B_{4i}(x) + B_{5i}(x), \text{ say.}
\end{aligned} \tag{C.11}$$

We will establish the uniform probability order for $B_{ji}(x)$, $j = 1, 2, ..., 5$, in order.

For $B_{1i}(x)$, we apply Lemma C.2 to obtain that, uniformly in i,

$$\frac{1}{nh_0^2}\sum_{j\neq i}k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right) = f'(\bar{\varepsilon}_i) + O_p\left(n^{-1/2}h_0^{-3/2}\sqrt{\log n} + h_0^{p+1}\right). \tag{C.12}$$

Then by (C.8) and the uniform boundedness of $f'(\varepsilon)$, we have

$$\max_{\{K_{ix}>0\}}|B_{1i}(x)| = O_p(\upsilon_{2n}). \tag{C.13}$$

Similarly, by (C.12), (C.8), and the uniform boundedness of $f'(\varepsilon)\varepsilon$, we have

$$\max_{\{K_{ix}>0\}}|B_{2i}(x)| = O_p(\upsilon_{2n}). \tag{C.14}$$

Expanding $M_n^{-1}(x)$ around its probability limit $[Mf_X(x)]^{-1}$, we have

$$\begin{aligned}
B_{3i}(x) &= \frac{1}{nh_0^2}\sum_{j\neq i}k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\sigma^{-1}(X_j)e_1^\top[Mf_X(X_j)]^{-1}U_{1,n}(X_j) \\
&\quad - \frac{1}{nh_0^2}\sum_{j\neq i}k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\sigma^{-1}(X_j)e_1^\top\alpha_n(X_j)M_n^{-1}(X_j)U_{1,n}(X_j) \\
&\quad + \frac{1}{nh_0^2}\sum_{j\neq i}k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\sigma^{-1}(X_j)e_1^\top[Mf_X(X_j)]^{-1}B_{1,n}(X_j) \\
&\quad - \frac{1}{nh_0^2}\sum_{j\neq i}k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\sigma^{-1}(X_j)e_1^\top\alpha_n(X_j)M_n^{-1}(X_j)B_{1,n}(X_j) \\
&\equiv B_{31i}(x) - B_{32i}(x) + B_{33i}(x) - B_{34i}(x),
\end{aligned}$$

where $\alpha_n(x) \equiv [Mf_X(x)]^{-1}[M_n(x) - Mf_X(x)]$. Write

$$\begin{aligned}
B_{31i}(x) &= \frac{1}{nh_0^2}\sum_{j\neq i} k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\sigma^{-1}(X_j)\, e_1^\top [Mf_X(X_j)]^{-1} U_{1,n}(X_j) \\
&= \frac{1}{nh_0^2}\sum_{j\neq i} E_j\left[k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\right]\sigma^{-1}(X_j)\, e_1^\top [Mf_X(X_j)]^{-1} U_{1,n}(X_j) \\
&\quad + \frac{1}{nh_0^2}\sum_{j\neq i}\left\{k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right) - E_j\left[k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\right]\right\}\sigma^{-1}(X_j)\, e_1^\top [Mf_X(X_j)]^{-1} U_{1,n}(X_j) \\
&\quad \equiv B_{31i,1}(x) + B_{31i,2}(x)\,, \text{ say.}
\end{aligned}$$

For $B_{31i,1}(x)$, we have

$$\begin{aligned}
\max_{1\le i\le n}|B_{31i,1}(x)| &\le \max_{1\le i\le n}\left|\frac{n-1}{nh_0^2}E_j\left[k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\right]\right| \times \sup_{x\in\mathcal{X}}\left\|\sigma^{-1}(x)\, e_1^\top [Mf_X(x)]^{-1}\right\| \sup_{x\in\mathcal{X}}\|U_{1,n}(x)\| \\
&= O_p(1)\, O_p(1)\, O_p\left(n^{-1/2}h_1^{-d/2}\sqrt{\log n}\right) = O_p\left(n^{-1/2}h_1^{-d/2}\sqrt{\log n}\right),
\end{aligned}$$

where we use the facts that $\sup_{x\in\mathcal{X}}\|U_{1,n}(x)\| = O_p(n^{-1/2}h_1^{-d/2}\sqrt{\log n})$ by Masry (1996a), $\max_{1\le i\le n}|h_0^{-2} \times E_j[k_0'((\bar{\varepsilon}_i - \varepsilon_j)/h_0)] - f'(\bar{\varepsilon}_i)| = O(h_0^{p+1})$ by standard bias calculation for kernel estimates and $\max_{1\le i\le n} |f'(\bar{\varepsilon}_i)| \le \sup_\varepsilon |f'(\varepsilon)| \le C < \infty$.

Let $v_j(\bar{\varepsilon}_i) = k_0'((\bar{\varepsilon}_i - \varepsilon_j)/h_0) - E[k_0'((\bar{\varepsilon}_i - \varepsilon_j)/h_0)]$. Then

$$\begin{aligned}
B_{31i,2}(x) &= \frac{1}{n^2h_1^dh_0^2}\sum_{j\neq i}\sum_{l} v_j(\bar{\varepsilon}_i)\,\sigma^{-1}(X_j)\, e_1^\top [Mf_X(X_j)]^{-1}\tilde{\mathbf{Z}}_l K\left(\frac{X_l - X_j}{h_1}\right)u_l \\
&= \frac{1}{n^2h_1^dh_0^2}\sum_{j\neq i}\sum_{l\neq j,i} v_j(\bar{\varepsilon}_i)\,\sigma^{-1}(X_j)\, e_1^\top [Mf_X(X_j)]^{-1}\tilde{\mathbf{Z}}_l K\left(\frac{X_l - X_j}{h_1}\right)u_l \\
&\quad + \frac{1}{n^2h_1^dh_0^2}\sum_{j\neq i} v_j(\bar{\varepsilon}_i)\,\sigma^{-1}(X_j)\, e_1^\top [Mf_X(X_j)]^{-1}\tilde{\mathbf{Z}}_j K(0)\, u_j \\
&\quad + \frac{1}{n^2h_1^dh_0^2}\sum_{j\neq i} v_j(\bar{\varepsilon}_i)\,\sigma^{-1}(X_j)\, e_1^\top [Mf_X(X_j)]^{-1}\tilde{\mathbf{Z}}_i K\left(\frac{X_i - X_j}{h_1}\right)u_i \\
&\equiv B_{31i,2a}(x) + B_{31i,2b}(x) + B_{31i,2c}(x)\,, \text{ say.}
\end{aligned}$$

By construction, $B_{31i,2a}(x)$ is a second order degenerate U-statistic (see, e.g., Lee (1990)) and we can bound it by straightforward moment calculations. Let $\epsilon_n \equiv Cn^{-1/2}h_1^{-d/2}\sqrt{\log n}$ for some $C > 0$. By the Boole and Markov inequalities,

$$P\left(\max_{1\le i\le n}|B_{31i,2a}(x)| \ge \epsilon_n\right) \le \sum_{i=1}^n P(|B_{31i,2a}(x)| \ge \epsilon_n) \le \sum_{i=1}^n \frac{E\left[|B_{31i,2a}(x)|^4\right]}{\epsilon_n^4}.$$

Let $a_{lj} = e_1^\top [Mf_X(X_j)]^{-1}\tilde{\mathbf{Z}}_l K((X_l - X_j)/h_1)$. Note that

$$\begin{aligned}
E\left[|B_{31i,2}(x)|^4\right] &= \frac{1}{(n^2h_1^dh_0^2)^4}\sum_{j_s\neq l_s\neq i \text{ for } s=1,2,3,4} \\
&\quad \times E\left\{a_{l_1j_1}a_{l_2j_2}a_{l_3j_3}a_{l_4j_4}v_{j_1}(\bar{\varepsilon}_i)\,u_{l_1}v_{j_2}(\bar{\varepsilon}_i)\,u_{l_2}v_{j_3}(\bar{\varepsilon}_i)\,u_{l_3}v_{j_4}(\bar{\varepsilon}_i)\,u_{l_4}\right\},
\end{aligned}$$

where the summations are only taken with respect to j and l's. Consider the index set $S \equiv \{j_s, l_s, s = 1, 2, 3, 4\}$. If the number of distinct elements in S is larger than 4, then the expectation in the last expression is zero by the IID condition in Assumption A1. We can readily show that $E\left[|B_{31i,2}(x)|^4\right] = O(n^{-4}h_1^{-2d}h_0^{-6})$. It follows that

$$
\begin{aligned}
P\left(\max_{1\le i\le n}|B_{31i,2a}(x)| \ge \epsilon_n\alpha_{n,0}\right) &\le \frac{nO(n^{-4}h_1^{-2d}h_0^{-6})}{Cn^{-2}h_1^{-2d}(\log n)^2\alpha_{n,0}^4} = \frac{O\left(n^{-1}h_0^{-6}(\log n)^{-2}\right)}{\alpha_n^4} \\
&= O\left(n^{-1}h^{-(2p+1)-d}\right) = O(1).
\end{aligned}
$$

where recall $\alpha_{n,s} = h^{[(2p+d)/4-(s+1)]}(\log n)^{s+1}$. Then $\max_{1\le i\le n}|B_{31i,2a}(x)| = O_p(\alpha_{n,0}n^{-1/2}h_1^{-d/2}\sqrt{\log n})$ by the Markov inequality. Analogously, we can show that $\max_{1\le i\le n}|B_{31i,2c}(x)| = o(n^{-1/2}h_1^{-d/2}\sqrt{\log n})$. For $B_{31i,2b}(x)$, we continue to decompose it as follows

$$
\begin{aligned}
B_{31i,2b}(x) &= \frac{K(0)}{n^2h_1^dh_0^2}\sum_{j\ne i}\sigma^{-1}(X_j)e_1^\top[Mf_X(X_j)]^{-1}\tilde{\mathbf{Z}}_j\{v_j(\bar{\varepsilon}_i)u_j - E_j[v_j(\bar{\varepsilon}_i)u_j]\} \\
&\quad + \frac{K(0)}{n^2h_1^dh_0^2}\sum_{j\ne i}\sigma^{-1}(X_j)e_1^\top[Mf_X(X_j)]^{-1}\tilde{\mathbf{Z}}_jE_j[v_j(\bar{\varepsilon}_i)u_j] \\
&\equiv B_{31i,2b1}(x) + B_{31i,2b2}(x),
\end{aligned}
$$

where E_j denotes expectation with respect to the variable indexed by j. We bound the second term first:

$$
\begin{aligned}
\max_{1\le i\le n}|B_{31i,2b2}(x)| &\le \max_{1\le i\le n}\left|h_0^{-1}E_j[v_j(\bar{\varepsilon}_i)u_j]\right|\frac{K(0)}{n^2h_1^dh_0}\sum_{j=1}^n\sigma^{-1}(X_j)\left|e_1^\top[Mf_X(X_j)]^{-1}\tilde{\mathbf{Z}}_j\right| \\
&= O_p(1)O(n^{-1}h_1^{-d}h_0^{-1}) = O_p(n^{-1}h_1^{-d}h_0^{-1}).
\end{aligned}
$$

By the Boole and Markov inequalities,

$$
\begin{aligned}
P\left(\max_{1\le i\le n}|B_{31i,2b1}(x)| \ge \epsilon_n\right) &\le \sum_{i=1}^n\frac{E\left[|B_{31i,2b1}(x)|^4\right]}{\epsilon_n^4} = \frac{nO(n^{-6}h_1^{-4d}h_0^{-6})}{Cn^{-2}h_1^{-2d}(\log n)^2} \\
&= O(n^{-3}h_1^{-2d}h_0^{-6}(\log n)^{-2}) = o(1),
\end{aligned}
$$

implying that $\max_{1\le i\le n}|B_{31i,2b1}(x)| = o_p(n^{-1/2}h_1^{-d/2}\sqrt{\log n})$. Hence $\max_{1\le i\le n}|B_{31i,2}(x)| = O_p(n^{-1}h_1^{-d}h_0^{-1}) + o_p(n^{-1/2}h_1^{-d/2}\sqrt{\log n})$. Consequently, we have shown that

$$
\max_{1\le i\le n}|B_{31i}(x)| = O_p(n^{-1}h_1^{-d}h_0^{-1}) + (\alpha_{n,0} + o(1))\,O_p(n^{-1/2}h_1^{-d/2}\sqrt{\log n}).
$$

By (C.4), the fact that $\sup_{x\in\mathcal{X}}\|U_{1,n}(x)\| = O_p(n^{-1/2}h_1^{-d/2}\sqrt{\log n})$, and the fact that $\max_{1\le i\le n}\frac{1}{nh_0^2}\sum_{j\ne i}|k_0'((\bar{\varepsilon}_i - \varepsilon_j)/h_0)| = O(h_0^{-1})$, we can readily show that $\max_{1\le i\le n}|B_{32i}(x)| = O_p(\upsilon_{0n}n^{-1/2}h_1^{-d/2}\sqrt{\log n}\,h_0^{-1})$. For the other terms, we have $\max_{1\le i\le n}|B_{33i}(x)| = O_p(h_1^{p+1})$, and $\max_{1\le i\le n}|B_{34i}(x)| = O_p(h_1^{p+1})O_p(\upsilon_{0n})\,O_p(h_0^{-1}) = O_p(\upsilon_{0n}h_1^{p+1}h_0^{-1})$. Consequently,

$$
\max_{1\le i\le n}|B_{3i}(x)| = O_p\left(n^{-1}h_1^{-d}h_0^{-1} + \upsilon_{1n} + \upsilon_{0n}\upsilon_{1n}h_0^{-1} + \alpha_{n,0}n^{-1/2}h_1^{-d/2}\sqrt{\log n}\right). \tag{C.15}
$$

Now write

$$
\begin{aligned}
B_{4i}(x) &= \frac{1}{nh_0^2}\sum_{j\neq i} k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\varepsilon_j\frac{\tilde{\sigma}(X_j)-\sigma(X_j)}{\tilde{\sigma}(X_j)} \\
&\quad +\frac{1}{nh_0^2}\sum_{j\neq i} k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\frac{m(X_j)-\tilde{m}(X_j)}{\sigma(X_j)}\frac{\tilde{\sigma}(X_j)-\sigma(X_j)}{\tilde{\sigma}(X_j)} \\
&\equiv B_{41i}(x)+B_{42i}(x).
\end{aligned}
$$

By (C.4) and Lemma C.1, it is easy to show that $\max_{1\leq i\leq n}|B_{42i}(x)| = O_p\left(\upsilon_{1n}^2 h_0^{-1}\right)$. Using analogous arguments as used in the analysis of $B_{3i}(x)$ and Lemma C.1, we can show that $\max_{1\leq i\leq n}|B_{41i}(x)| = O_p(n^{-1}h_1^{-d}h_0^{-1}+\upsilon_{0n}\upsilon_{1n}h_0^{-1}+h_1^{p+1})$. Consequently,

$$
\max_{1\leq i\leq n}|B_{4i}(x)| = O_p(n^{-1}h_1^{-d}h_0^{-1}+\upsilon_{0n}\upsilon_{1n}h_0^{-1}+h_1^{p+1}). \tag{C.16}
$$

where we use the fact that $\upsilon_{1n}^2 h_0^{-1} = o_p(n^{-1}h_1^{-d}h_0^{-1}+h_1^{p+1})$. As argued by Hansen (2008, pp.740-741), under Assumption A6 there exists an integral function k_0^* such that

$$
\left|k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j+wr_{1ij}}{h_0}\right)-k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right| \leq wh_0^{-1}k_0^*\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)|r_{1ij}|.
$$

It follows that

$$
\begin{aligned}
\max_{1\leq i\leq n}|B_{5i}(x)| &\leq \frac{1}{nh_0^3}\sum_{j\neq i} k_0^*\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)r_{1ij}^2 = \frac{O_p\left(\upsilon_{2n}^2\right)}{nh_0^3}\sum_{j\neq i} k_0^*\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\left(\bar{\varepsilon}_i^2+\varepsilon_j^2\right) \\
&= O_p\left(\upsilon_{2n}^2 h_0^{-2}\right).
\end{aligned} \tag{C.17}
$$

Combining (C.11), (C.13), (C.14), (C.15), (C.16), and (C.17) and using the facts that $n^{-1}h_1^{-d}h_0^{-1} = o\left(\upsilon_{1n}^2 h_0^{-2}\right)$ and that $h_1^{p+1} = o(\upsilon_{2n})$ yield the desired result for $s=0$.

When $s>0$, we can decompose $\tilde{f}_i^{(s)}(\overrightarrow{\varepsilon}_i)-\overline{f}^{(s)}(\bar{\varepsilon}_i)$ as in (C.11) with the corresponding terms denoted as $B_{ri}^{(s)}(x)$ for $r=1,2,...,5$. The probability orders of $B_{1i}^{(s)}(x)$ and $B_{2i}^{(s)}(x)$ are the same as those of $B_{1i}(x)$ and $B_{2i}(x)$, those of $B_{3i}^{(s)}(x)$ and $B_{4i}^{(s)}(x)$ become $O_p(n^{-1}h_1^{-d}h_0^{-1-s}+(\upsilon_{0n}h_0^{-1-s}+\alpha_{n,s})n^{-1/2}h_1^{-d/2}\sqrt{\log n}+h_1^{p+1})$, and the probability order of $B_{5i}^{(s)}(x)$ is $O_p(\upsilon_{2n}^2 h_0^{-2-s})$. Consequently, $\max_{1\leq i\leq n}|\tilde{f}_i^{(s)}(\overrightarrow{\varepsilon}_i)-\overline{f}^{(s)}(\bar{\varepsilon}_i)| = O_p(\upsilon_{2n}+(\upsilon_{0n}h_0^{-1-s}+\alpha_{n,s})n^{-1/2}h_1^{-d/2}\sqrt{\log n}+\upsilon_{2n}^2 h_0^{-2-s})$. ■

Proof of Lemma A.2. The proof is similar to but much simpler than that of Lemma A.1 and thus omitted. ■

Proof of Lemma A.3. The proof is analogous to that of Lemma USSLN in Gozalo and Linton (2000) and thus we only sketch the proof for the $r=1$ case. Let $\mathcal{C}_n = \{q_{1n}(\cdot,\theta):\theta\in\Theta\}$. Under the permissibility and envelope integrability of $\mathcal{C}_n$, the almost sure convergence of $\sup_{\theta\in\Theta}\left|h^{-d}\left[P_n q_{n,1}(Z,\theta)-Pq_{n,1}(Z_i,\theta)\right]\right|$ is equivalent to its convergence in probability. By the boundedness of Θ and measurability of the $q_{n,1}$, the class $\mathcal{C}_n$ is permissible in the sense of Pollard (1984, p196). We now show the envelope integrability of $\mathcal{C}_n$.

By Assumption A1 and the compactness of K, $|\log(f(\varepsilon_i(\boldsymbol{\beta})))| \leq D(Y_i)$ on the set $K_{ix} > 0$. Consequently, we can take the dominance function $\overline{q}_n = D(Y) K((x-X)/h)$. Let $E[D(Y)|X] = \bar{D}(X)$. Assumptions A1 and A3 ensure that

$$P\overline{q}_n = E\left[\bar{D}(X) K((x-X)/h)\right] = h^d \int \bar{D}(x-hu) f(x-hu) K(u)\, du = O\left(h^d\right).$$

The envelope integrability allows us to truncate the functions to a finite range. Let $\alpha_n > 1$ be a sequence of constants such that $\alpha_n \to \infty$ as $n \to \infty$. Define

$$\mathcal{C}^*_{\alpha_n} = \left\{q^*_{\alpha_n} = \alpha_n^{-1} q_{n,1} \mathbf{1}\left\{\overline{q}_n \leq \alpha_n\right\} : q_n \in \mathcal{C}_n\right\}.$$

Let b_n be a non-increasing sequence of positive numbers for which $nh^d b_n^2 \gg \log n$. By analysis similar to that of Gozalo and Linton (2000) and Theorem II.37 of Pollard (1984, p.34), to show that $\sup_{\mathcal{C}_n} |P_n q_{n,1} - P q_{n,1}| = o_p\left(h^d b_n\right)$, it suffices to show

$$\sup_{\mathcal{C}^*_{\alpha_n}} \left|P_n q^*_{\alpha_n} - P q^*_{\alpha_n}\right| = o_p\left(h^d b_n\right), \tag{C.18}$$

which holds provided

$$\sup_{\mathcal{C}^*_{\alpha_n}} \left\{P\left[q^*_{\alpha_n}\right]^2\right\}^{1/2} < h^{d/2} \tag{C.19}$$

and

$$\sup N_1\left(\epsilon, G, \mathcal{C}^*_{\alpha_n}\right) \leq C_1 \epsilon^{-C_2} \text{ for } 0 < \epsilon \leq 1, \tag{C.20}$$

where $N_1\left(\epsilon, G, \mathcal{C}^*_{\alpha_n}\right)$ is the covering number of $\mathcal{C}^*_{\alpha_n}$, i.e., the smallest value J for which there exists functions $g_1, ..., g_J$ such that $\min_{j \leq J} G|q - g_j| \leq \epsilon$ for each $q \in \mathcal{C}^*_{\alpha_n}$, the supremum is taken over all probability measures G, and C_1 and C_2 are positive constants independent of n.

(C.19) holds by construction. For (C.20), we need to show that $\mathcal{C}^*_{\alpha_n}$ is a Euclidean class (Nolan and Pollard, 1987, p.789). Since the functions in $\mathcal{C}^*_{\alpha_n}$, $q^*_{\alpha_n} = \alpha_n^{-1} \log(f(\varepsilon(\boldsymbol{\beta}))) G_b(f(\varepsilon(\boldsymbol{\beta}))) K((x-X)/h)$ $\mathbf{1}\{\overline{q}_n \leq \alpha_n\}$, are composed from the classes of functions

$$\begin{aligned}
\mathcal{C}_1 &= \left\{c_1 \log f\left(\frac{y - P(\boldsymbol{\beta}_1)}{\sqrt{\exp(P(\boldsymbol{\beta}_2))}}\right) : \left(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top\right)^\top \in \mathcal{B},\ c_1 \leq 1\right\}, \\
\mathcal{C}_2 &= \left\{c_2 G_b\left(f\left(\frac{y - P(\boldsymbol{\beta}_1)}{\sqrt{\exp(P(\boldsymbol{\beta}_2))}}\right)\right) : \left(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top\right)^\top \in \mathcal{B},\ c_2 \leq 1\right\}, \\
\mathcal{C}_3 &= \left\{K\left(x^\top c_3 + c_4\right) : c_3 \in \mathbb{R}^d,\ c_4 \in \mathbb{R}\right\}, \text{ and } \mathcal{C}_4 = \left\{\mathbf{1}\left\{c_5 \overline{q}_n \leq 1\right\} : c_5 \in \mathbb{R}\right\},
\end{aligned}$$

it suffices to show that the $\mathcal{C}_j' s$ form Euclidean classes by Nolan and Pollard (1987, pp. 796-797) and Pakes and Pollard (1989, Lemmas 2.14 and 2.15).

First, for $j = 1, 2$, $\{P(\boldsymbol{\beta}_j)\}$ forms a polynomial class of functions and is Euclidean by Lemma 2.12 of Pakes and Pollard (1989). By Example 2.10 of Pakes and Pollard (1989) and the bounded variation assumption on f, the class $\{f\left(\frac{y-m}{s}\right) : m \in \mathbb{R},\, s > 0\}$ is Euclidean for the constant envelope $\sup_{\varepsilon} |f(\varepsilon)|$. It follows from Pakes and Pollard (1989, Lemmas 2.15) that $\mathcal{C}_1$ is also Euclidean. Similarly, $\mathcal{C}_2$ is Euclidean. By Nolan and Pollard (1987, Lemma 22) and the bounded variation of K, $\mathcal{C}_3$ forms a Euclidean class with constant envelope $\sup_x |K(x)|$. Finally, by Pollard (1984, Lemma II.25) and the Euclidean property of $\mathcal{C}_j,\ j = 1, 2, 3$, $\mathcal{C}_4$ is Euclidean. Consequently

$$\sup_{\theta} \left| \frac{1}{nh^d} \sum_{i=1}^{n} q_{1n}(Z_i, \theta) - E q_{1n}(Z_i, \theta) \right| = o_{a.s.}(b_n).$$

Since Pollard's Theorem requires that $b_n \gg n^{-1/2} h^{-d/2} \sqrt{\log n}$, we can take $b_n = n^{-1/2} h^{-d/2} \sqrt{\log n}$ to obtain the desired result. ■

Proof of Lemma A.4. The proof is analogous to that of Newey (1991, Corollary 3.2). We first show $\bar{P}_{n,1}(\theta)$ is equicontinuous. Let $D_{n,i}(S) = \mathbf{1}\{Y_i \notin S\} D(Y_i) K_h(x - X_i)$ for a compact set S on $\mathbb{R}$. By the Hölder inequality and the law of iterated expectations,

$$\begin{aligned} ED_{n,i}(S) &= EE\left[D_{n,i}(S) | X_i\right] \\ &\leq E\left[\{P(Y_i \notin S | X_i)\}^{(\gamma-1)/\gamma} \{E[D^{\gamma}(Y_i) | X_i]\}^{1/\gamma} K_h(x - X_i)\right] \\ &= E\left[\{P(Y_i \notin S | X_i)\}^{(\gamma-1)/\gamma} \left[\bar{D}(X_i)\right]^{1/\gamma} K_h(x - X_i)\right]. \end{aligned} \tag{C.21}$$

Note that

$$E\left[\left[\bar{D}(X_i)\right]^{1/\gamma} K_h(x - X_i)\right] = \int \left[\bar{D}(x - hv)\right]^{1/\gamma} f(x - hv) K(v)\, dv \leq C \int K(v)\, dv. \tag{C.22}$$

Consider $\epsilon, \eta > 0$. By Assumption A2, we can choose S large enough such that $P(Y_i \notin S | X_i)$ is arbitrary small to ensure $ED_{n,i}(S) < \epsilon\eta/4$. Also, $q_n(z, \theta)$ is uniformly continuous on $(\mathcal{X} \times S) \times \Theta$ for each compact set $\mathcal{X} \times S$, implying that for any $\theta \in \Theta$ there exists $\mathcal{N} \equiv \mathcal{N}(\theta)$ such that $\sup_{(z,\theta') \in (\mathcal{X} \times S) \times \mathcal{N}} |p_1(z, \theta') - p_1(z, \theta)| < \epsilon/2$. Consequently

$$\sup_{\theta' \in \mathcal{N}} \left| p_1(Z_i, \theta') - p_1(Z_i, \theta) \right| < \epsilon/2 + 2 \cdot \mathbf{1}\{Y_i \notin S\} D(Y_i) K_h(x - X_i). \tag{C.23}$$

Let $\triangle_n(\epsilon, \eta) = \epsilon/2 + 2\bar{D}_n(S)$, where $\bar{D}_n(S) = n^{-1} \sum_{i=1}^{n} D_{n,i}(S)$. By (C.23) and the triangle inequality

$$\sup_{\theta' \in \mathcal{N}} \left| P_n p_1(Z, \theta') - P_n p_1(Z, \theta) \right| < \triangle_n(\epsilon, \eta).$$

Also,

$$P(\triangle_n(\epsilon, \eta) > \epsilon) = P\left(\bar{D}_n(S) > \epsilon/4\right) \leq \frac{E\left[D_{n,i}(S)\right]}{\epsilon/4} < \eta.$$

Consequently

$$\begin{aligned} \sup_{\theta' \in \mathcal{N}} \left| \bar{P}_{n,1}(\theta') - \bar{P}_{n,1}(\theta) \right| &= \sup_{\theta' \in \mathcal{N}} \left| E\left[P_n p_1(Z, \theta') - P_n p_1(Z, \theta)\right] \right| \\ &\leq E\left[\sup_{\theta' \in \mathcal{N}} \left| P_n p_1(Z, \theta') - P_n p_1(Z, \theta) \right|\right] \leq E\left[\triangle_n(\epsilon, \eta)\right] < \eta. \end{aligned}$$

That is, $\left\{\bar{P}_{n,1}(\theta)\right\}$ is equicontinuous.

Notice that under our assumption on the compactness of $\mathcal{B}$ and the support of K, $P_i(\boldsymbol{\beta}_2)$ is bounded. So the proof for the equicontinuity of $\bar{P}_{n,2}(\theta)$ is simpler than that of $\bar{P}_{n,1}(\theta)$ and thus omitted. ■

Proof of Lemma B.1. We only prove the case $(r,s)=(1,1)$ as the other cases are similar. For notational simplicity, write $T_{1n\mathbf{j}}=T_{1n\mathbf{j}}(1,1)$. By the fact that $\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{-1/2}-\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}=O_p(\upsilon_{2n})$ uniformly in i on the set $\{K_{ix}>0\}$, we can write

$$\begin{aligned}
\tilde{q}_{1,i}(\tilde{\boldsymbol{\beta}})-q_{1,i}(\boldsymbol{\beta}^0) &= \frac{\tilde{f}_i'(\overrightarrow{\varepsilon}_i)}{\tilde{f}_i(\overrightarrow{\varepsilon}_i)}\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{-1/2}-\frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2} \\
&= \left[\frac{\tilde{f}_i'(\overrightarrow{\varepsilon}_i)}{\tilde{f}_i(\overrightarrow{\varepsilon}_i)}-\frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\right]\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{-1/2}+\frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\left[\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{-1/2}-\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}\right] \\
&= \left[\frac{\tilde{f}_i'(\overrightarrow{\varepsilon}_i)}{\tilde{f}_i(\overrightarrow{\varepsilon}_i)}-\frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\right]\left[\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}+O_p(\upsilon_{2n})\right]+\frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}O_p(\upsilon_{2n}). \qquad \text{(C.24)}
\end{aligned}$$

Thus

$$\begin{aligned}
|T_{1n\mathbf{j}}| &\leq \frac{1}{nh^d}\sum_{i=1}^n\left|K_{ix,\mathbf{j}}\tilde{G}_iq_{1,i}(\boldsymbol{\beta}^0)\left[\frac{\tilde{f}_i'(\overrightarrow{\varepsilon}_i)}{\tilde{f}_i(\overrightarrow{\varepsilon}_i)}-\frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\right]\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}\right| \\
&\quad+\frac{O_p(\upsilon_{2n})}{nh^d}\sum_{i=1}^n\left|K_{ix,\mathbf{j}}\tilde{G}_iq_{1,i}(\boldsymbol{\beta}^0)\left[\frac{\tilde{f}_i'(\overrightarrow{\varepsilon}_i)}{\tilde{f}_i(\overrightarrow{\varepsilon}_i)}-\frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\right]\right| \\
&\quad+\frac{O_p(\upsilon_{2n})}{nh^d}\sum_{i=1}^n\left|K_{ix,\mathbf{j}}\tilde{G}_iq_{1,i}(\boldsymbol{\beta}^0)\frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\right|.
\end{aligned}$$

Since the last two terms are of smaller order, it suffices to show the first term (denoted as $|\bar{T}_{1n\mathbf{j}}|$) is $O_p(h^\epsilon)$. By Lemma A.1, the definition of $\tilde{G}_i$, and Assumption A7,

$$\begin{aligned}
&\left|\frac{\tilde{f}_i'(\overrightarrow{\varepsilon}_i)}{\tilde{f}_i(\overrightarrow{\varepsilon}_i)}-\frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\right|\tilde{G}_i=\left|\frac{\tilde{f}_i'(\overrightarrow{\varepsilon}_i)-f'(\bar{\varepsilon}_i)}{\tilde{f}_i(\overrightarrow{\varepsilon}_i)}+\frac{f'(\bar{\varepsilon}_i)\left[f(\bar{\varepsilon}_i)-\tilde{f}_i(\overrightarrow{\varepsilon}_i)\right]}{f(\bar{\varepsilon}_i)\tilde{f}_i(\overrightarrow{\varepsilon}_i)}\right|\tilde{G}_i \\
\leq\ & O_p(b^{-1}\nu_{3n,1})+(f'(\bar{\varepsilon}_i)/f(\bar{\varepsilon}_i))O_p(b^{-1}\nu_{3n,0})=O_p(h^\epsilon)\{1+|f'(\bar{\varepsilon}_i)/f(\bar{\varepsilon}_i)|\}. \qquad \text{(C.25)}
\end{aligned}$$

Therefore $\left|\bar{T}_{1n\mathbf{j}}\right|=\frac{O_p(h^\epsilon)}{nh^d}\sum_{i=1}^n\left|K_{ix,\mathbf{j}}q_{1,i}(\boldsymbol{\beta}^0)\left(1+\frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\right)\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}\right|=O_p(h^\epsilon)$ by Markov inequality and the fact that

$$\begin{aligned}
\frac{1}{h^d}E\left|K_{ix,\mathbf{j}}q_{1,i}(\boldsymbol{\beta}^0)\left(1+\frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\right)\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}\right| &= \frac{1}{h^d}E\left|K_{ix,\mathbf{j}}\frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\left(1+\frac{f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\right)\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1}\right| \\
&= \frac{1}{h^d}E\left|K_{ix,\mathbf{j}}\sigma^{-2}(X_i)\frac{f'(\varepsilon_i)}{f(\varepsilon_i)}\left(1+\frac{f'(\varepsilon_i)}{f(\varepsilon_i)}\right)\right|\{1+o(1)\} \\
&\leq \frac{f_X(x)}{\sigma^2(x)}\int\left|K(u)u^{\mathbf{j}}\right|du\left\{I^{1/2}(f)+I(f)\right\}=O(1),
\end{aligned}$$

where $I(f)\equiv E\left[\psi^2(\varepsilon_i)\right]$ and we use the fact that $\varphi(P_i(\boldsymbol{\beta}_2^0))$ is the p-th order Taylor expansion of $\sigma^2(X_i)$ around x. This completes the proof of the lemma. ■

Proof of Lemma B.2. We only prove the case $(r,s)=(1,1)$ as the other cases are similar. For notational simplicity, write $T_{2n\mathbf{j}}=T_{2n\mathbf{j}}(1,1)$. That is, we will show

$$T_{2n\mathbf{j}}=\frac{1}{nh^d}\sum_{i=1}^{n}K_{ix,\mathbf{j}}\tilde{G}_i\left\{\tilde{q}_{1,i}(\tilde{\boldsymbol{\beta}})-q_1\left(Y_i;P_i\left(\boldsymbol{\beta}_1^0\right),P_i\left(\boldsymbol{\beta}_2^0\right)\right)\right\}^2=O_p\left(h^{\epsilon}\right).$$

By (C.24) and (C.25) in the proof of Lemma B.1, we can write

$$\begin{aligned}
&\left|\tilde{q}_{1,i}(\tilde{\boldsymbol{\beta}})-q_1\left(Y_i;P_i\left(\boldsymbol{\beta}_1^0\right),P_i\left(\boldsymbol{\beta}_2^0\right)\right)\right|^2\tilde{G}_i\\
=&\left[O_p\left(h^{\epsilon}\right)\left(1+\left|\frac{f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\right|\right)^2\left[\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1/2}+O_p(\upsilon_{2n})\right]^2+\left(\frac{f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\right)^2O_p(\upsilon_{2n}^2)\right]\tilde{G}_i\\
\leq&\left(1+\left|\frac{f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\right|\right)^2\left[\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1}+1\right]\tilde{G}_iO_p\left(h^{\epsilon}\right).
\end{aligned}$$

Thus

$$T_{2n\mathbf{j}}\leq\frac{O_p\left(h^{\epsilon}\right)}{nh^d}\sum_{i=1}^{n}\left|K_{ix,\mathbf{j}}\right|\left(1+\left|\frac{f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\right|\right)^2\left[\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1}+1\right]=O_p\left(h^{\epsilon}\right)$$

by Markov inequality and the fact that

$$\begin{aligned}
&\frac{1}{h^d}E\left|K_{ix,\mathbf{j}}\left(1+\left|\frac{f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\right|\right)^2\left[\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1}+1\right]\right|\\
=&\frac{1}{h^d}E\left|K_{ix,\mathbf{j}}\left(1+\left|\frac{f'\left(\bar{\varepsilon}_i\right)}{f\left(\bar{\varepsilon}_i\right)}\right|\right)^2\left[\sigma^{-2}\left(X_i\right)+1+o\left(1\right)\right]\right|\left\{1+o\left(1\right)\right\}\\
\leq&f_X\left(x\right)\left[\sigma^{-2}\left(x\right)+1\right]\int\left|K\left(u\right)u^{\mathbf{j}}\right|du\left[1+I\left(f\right)+2I^{1/2}\left(f\right)\right]\left\{1+o\left(1\right)\right\}=O\left(1\right).
\end{aligned}$$

This completes the proof of the lemma. ■

Proof of Lemma B.3. We only prove the case $(r,s)=(1,1)$ as the other cases are similar. For notational simplicity, write $T_{3n\mathbf{j}}=T_{3n\mathbf{j}}(1,1)$. That is, we will show

$$T_{3n\mathbf{j}}=\frac{1}{nh^d}\sum_{i=1}^{n}K_{ix,\mathbf{j}}\left(1-\tilde{G}_i\right)q_{1,i}\left(\boldsymbol{\beta}^0\right)^2=O_p\left(h^{\epsilon}\right).$$

We decompose $T_{3n\mathbf{j}}$ as follows

$$\begin{aligned}
T_{3n\mathbf{j}}=&\frac{1}{nh^d}\sum_{i=1}^{n}K_{ix,\mathbf{j}}\left[1-G_b\left(f\left(\bar{\varepsilon}_i\right)\right)\right]\psi^2\left(\bar{\varepsilon}_i\right)\\
&+\frac{1}{nh^d}\sum_{i=1}^{n}K_{ix,\mathbf{j}}\left[G_b\left(f\left(\bar{\varepsilon}_i\right)\right)-G_b(\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right))\right]\psi^2\left(\bar{\varepsilon}_i\right)\\
\equiv&T_{3n\mathbf{j},1}+T_{3n\mathbf{j},2},\text{ say.}
\end{aligned}$$

By Lemma A.1,

$$\begin{aligned}
\max_{1\leq i\leq n}\left|\tilde{G}_i-G_i\right|=&\max_{1\leq i\leq n}\left|G_b(\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right))-G_b\left(f\left(\bar{\varepsilon}_i\right)\right)\right|\\
\leq&\frac{C}{b}\max_{1\leq i\leq n}\left|\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right)-f\left(\bar{\varepsilon}_i\right)\right|=b^{-1}O_p\left(\upsilon_{3n,0}\right)=O_p\left(h^{\epsilon}\right).
\end{aligned}\tag{C.26}$$

With this, we can readily obtain $|T_{3n\mathbf{j},2}| \leq O_p(h^{\epsilon}) \frac{1}{nh^d}\sum_{i=1}^n |K_{ix,\mathbf{j}}| \psi^2(\bar{\varepsilon}_i) = O_p(h^{\epsilon})$ by Markov inequality. For $T_{3n\mathbf{j},1}$, we have

$$\begin{aligned} E|T_{3n\mathbf{j},1}| &\leq E\left[\frac{1}{h^d}|K_{ix,\mathbf{j}}|\left[1-G_b(f(\bar{\varepsilon}_i))\right]\psi^2(\bar{\varepsilon}_i)\right] \\ &= \frac{1}{h^d}E\left[|K_{ix,\mathbf{j}}|\right]E\left\{\left[1-G_b(f(\varepsilon_i))\right]\psi^2(\varepsilon_i)\right\}\{1+o(1)\}. \end{aligned}$$

By the Hölder inequality,

$$\begin{aligned} E\left\{\left[1-G_b(f(\varepsilon_i))\right]\psi^2(\varepsilon_i)\right\} &\leq E\left[\psi^2(\varepsilon_i)\ \mathbf{1}\{f(\varepsilon)\leq 2b\}\right] \\ &\leq \left\{E\left[\psi^{2\gamma}(\varepsilon_i)\right]\right\}^{1/\gamma}\left[P(f(\varepsilon_i)\leq 2b)\right]^{(\gamma-1)/\gamma} \\ &\leq C\left[P(f(\varepsilon_i)\leq 2b)\right]^{(\gamma-1)/\gamma} = O\left(b^{(\gamma-1)/(2\gamma)}\right) = O(h^{\epsilon}), \end{aligned}$$

where the last line follows from Lemma 6 of Robinson (1988) and the Markov inequality because by taking $\overline{B} = b^{-1/2}$, we have $P(f(\varepsilon_i)\leq 2b) \leq 2b\int_{|\varepsilon_i|\leq\overline{B}} dz + P\left(|\varepsilon_i| > \overline{B}\right) \leq 2b2b^{-1/2} + E|\varepsilon_i|\, b^{1/2} = O\left(b^{1/2}\right) = O(h^{\epsilon})$. This, in conjunction with the fact that $\frac{1}{h^d}E\left[|K_{ix,\mathbf{j}}|\right] = O(1)$, implies that $T_{3n\mathbf{j},1} = O_p(h^{\epsilon})$ by Markov inequality. Consequently, we have shown that $T_{3n\mathbf{j}} = O_p(h^{\epsilon})$. ■

Proof of Lemma B.4. Let $\tilde{f}_i = \tilde{f}_i(\overrightarrow{\varepsilon}_i)$ and $f_i = f(\bar{\varepsilon}_i)$.Note that $\tilde{f}_i^{-1} = f_i^{-1} - (\tilde{f}_i - f_i)/f_i^2 + R_{2i}$, where $R_{2i} \equiv (\tilde{f}_i - f_i)^2/\{(f_i^2\tilde{f}_i)\}$. First, we expand the trimming function to the second order:

$$G_b(\tilde{f}_i) - G_b(f_i) = g_b(f_i)\left(\tilde{f}_i - f_i\right) + \frac{1}{2}g_b'(f_i^*)\left(\tilde{f}_i - f_i\right)^2, \tag{C.27}$$

where f_i^* is an intermediate value between $\tilde{f}_i$ and f_i. Let $\rho_i(\boldsymbol{\beta}) \equiv \psi(\varepsilon_i(\boldsymbol{\beta}))\,\varepsilon_i(\boldsymbol{\beta}) + 1$, $\bar{\rho}_i \equiv \rho_i\left(\boldsymbol{\beta}^0\right)$, and $\rho_i \equiv \psi(\varepsilon_i)\,\varepsilon_i + 1$. Let $\varphi_i \equiv \varphi'(P_i(\boldsymbol{\beta}_2^0))/\varphi(P_i(\boldsymbol{\beta}_2^0))$. Then we have

$$\begin{aligned} -\mathcal{S}_{1n\mathbf{j}} &= \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^n K_{ix,\mathbf{j}}\varphi_i\left\{\bar{\rho}_i\left[G_b(\tilde{f}_i)-1\right] + \log\left(f(\bar{\varepsilon}_i)\,\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}\right)g_b(f(\bar{\varepsilon}_i))\,f'(\bar{\varepsilon}_i)\,\bar{\varepsilon}_i\right\} \\ &= \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^n K_{ix,\mathbf{j}}\varphi_i\left\{\bar{\rho}_i\left[G_b(f_i)-1\right] + \log\left(f(\bar{\varepsilon}_i)\,\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}\right)g_b(f(\bar{\varepsilon}_i))\,f'(\bar{\varepsilon}_i)\,\bar{\varepsilon}_i\right\} \\ &\quad + \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^n K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_i g_b(f_i)\left(\tilde{f}_i - f_i\right) \\ &\quad + \frac{1}{4\sqrt{nh^d}}\sum_{i=1}^n K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_i g_b'(f_i^*)\left(\tilde{f}_i - f_i\right)^2 \\ &\equiv \mathcal{S}_{1n\mathbf{j},1} + \mathcal{S}_{1n\mathbf{j},2} + \mathcal{S}_{1n\mathbf{j},3},\ \text{say.} \end{aligned}$$

Using a crude bound on the last term, we have $|\mathcal{S}_{1n\mathbf{j},3}| = O_p\left(v_{3n,0}^2 b^{-2}n^{1/2}h^{d/2}\right) = o_p(1)$ by Lemma A.1, the fact that $\sup_s |g_b'(s)| = O(b^{-2})$, and Assumption A7.

To show the first term is $o_p(1)$, write

$$\mathcal{S}_{1n\mathbf{j},1} = \frac{-1}{\sqrt{nh^d}}\sum_{i=1}^n K_{ix,\mathbf{j}}\xi_{1i} + \frac{1}{\sqrt{nh^d}}\sum_{i=1}^n K_{ix,\mathbf{j}}\xi_{2i} \equiv -\mathcal{S}_{1n\mathbf{j},11} + \mathcal{S}_{1n\mathbf{j},12},\ \text{say,}$$

where $\xi_{1i}=\frac{1}{2}\bar{\rho}_i\varphi_i$ and $\xi_{2i}=\frac{1}{2}\left\{\bar{\rho}_iG_b\left(f_i\right)+\log(f\left(\bar{\varepsilon}_i\right)\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{-1/2})g_b\left(f\left(\bar{\varepsilon}_i\right)\right)f'\left(\bar{\varepsilon}_i\right)\bar{\varepsilon}_i\right\}\varphi_i$.

Let $Q_{1n,i}\left(\boldsymbol{\beta}\right)\equiv\log\{f\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)\varphi(P_i\left(\boldsymbol{\beta}_2\right))^{-1/2}\}K_h\left(x-X_i\right)$, $Q_{2n,i}\left(\boldsymbol{\beta}\right)\equiv\log\{f\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)\varphi(P_i\left(\boldsymbol{\beta}_2\right))^{-1/2}\}$ $\times G_b(f\left(\varepsilon_i\left(\boldsymbol{\beta}\right)\right)K_h\left(x-X_i\right)$, and $\varsigma_n\left(\boldsymbol{\beta}\right)\equiv E\left[Q_{2n,i}\left(\boldsymbol{\beta}\right)\right]-E\left[Q_{1n,i}\left(\boldsymbol{\beta}\right)\right]$. Then it is easy to show that (i) $\varsigma_n\left(\boldsymbol{\beta}^0\right)\to 0$, (ii) $\varsigma_n\left(\boldsymbol{\beta}\right)$ is differentiable in a small ϵ_0-neighborhood $N_{\epsilon_0}\left(\boldsymbol{\beta}^0\right)$ of $\boldsymbol{\beta}^0$ with $N_{\epsilon_0}\left(\boldsymbol{\beta}^0\right)\equiv\{\boldsymbol{\beta}:\left\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\right\|\leq\epsilon_0\}$, (iii) $\varsigma_n'\left(\boldsymbol{\beta}\right)$ converges uniformly on $N_{\epsilon_0}\left(\boldsymbol{\beta}^0\right)$. Then by Theorem 7.17 of Rudin (1976) and the fact that $h^{-|\mathbf{j}|}\partial Q_{1n,i}\left(\boldsymbol{\beta}^0\right)/\partial\boldsymbol{\beta}_{2\mathbf{j}}=-h^{-d}\xi_{1i}K_{ix,\mathbf{j}}$ and $h^{-|\mathbf{j}|}\partial Q_{2n,i}\left(\boldsymbol{\beta}^0\right)/\partial\boldsymbol{\beta}_{2\mathbf{j}}=-h^{-d}\xi_{2i}K_{ix,\mathbf{j}}$, we have

$$\begin{aligned}E\left(\mathcal{S}_{1n\mathbf{j},12}\right) &= -\sqrt{n}h^{d/2}h^{-|\mathbf{j}|}E\left[\frac{\partial Q_{2n,i}\left(\boldsymbol{\beta}^0\right)}{\partial\boldsymbol{\beta}_{2\mathbf{j}}}\right]\\ &= -\sqrt{n}h^{d/2}h^{-|\mathbf{j}|}E\left[\frac{\partial Q_{1n,i}\left(\boldsymbol{\beta}^0\right)}{\partial\boldsymbol{\beta}_{2\mathbf{j}}}\right]\{1+o\left(1\right)\}=E\left(\mathcal{S}_{1n\mathbf{j},11}\right)\{1+o\left(1\right)\}.\end{aligned}$$

Consequently, $E\left(\mathcal{S}_{1n\mathbf{j},1}\right)=o\left(1\right)E\left(\mathcal{S}_{1n\mathbf{j},11}\right)=o\left(1\right)$ as $\mathcal{S}_{1n\mathbf{j},11}=n^{1/2}h^{-d/2}E\left[K_{ix,\mathbf{j}}\xi_{1i}\right]=O\left(n^{1/2}h^{d/2}h^{p+1}\right)$ $=O\left(1\right)$. By straightforward calculations and the IID assumption, we can readily show that $\mathrm{Var}(\mathcal{S}_{1n\mathbf{j},1})=o\left(1\right)$. Therefore, $\mathcal{S}_{1n\mathbf{j},1}=o_p\left(1\right)$ by the Chebyshev inequality.

Now, we show that $\mathcal{S}_{1n\mathbf{j},2}=o_p\left(1\right)$. Decompose $\mathcal{S}_{1n\mathbf{j},2}=\mathcal{S}_{1n\mathbf{j},21}+\mathcal{S}_{1n\mathbf{j},22}$, where $\mathcal{S}_{1n\mathbf{j},21}\equiv\frac{1}{2\sqrt{nh^d}}\sum_{i=1}^n K_{ix,\mathbf{j}}\varphi_iq_{2,i}\left(\boldsymbol{\beta}^0\right)g_b\left(f_i\right)\left(\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right)-\overline{f}\left(\bar{\varepsilon}_i\right)\right)$, and $\mathcal{S}_{1n\mathbf{j},22}\equiv\frac{1}{2\sqrt{nh^d}}\sum_{i=1}^n K_{ix,\mathbf{j}}\varphi_iq_{2,i}\left(\boldsymbol{\beta}^0\right)g_b\left(f_i\right)\left(\overline{f}\left(\bar{\varepsilon}_i\right)-f\left(\bar{\varepsilon}_i\right)\right)$. It suffices to show that $\mathcal{S}_{1n\mathbf{j},2s}=o_p\left(1\right)$, $s=1,2$. For $\mathcal{S}_{1n\mathbf{j},21}$, by a Taylor expansion and (C.9)-(C.10), we have

$$\begin{aligned}\mathcal{S}_{1n\mathbf{j},21} &= \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^n K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_ig_b\left(f_i\right)\frac{1}{nh_0}\sum_{j\neq i}\left[k_0\left(\frac{\overrightarrow{\varepsilon}_i-\tilde{\varepsilon}_j}{h_0}\right)-k_0\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right]\\ &= \frac{1}{2n^{3/2}h^{d/2}h_0^2}\sum_{i=1}^n\sum_{j\neq i}K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_ig_b\left(f_i\right)k_0'\left(\frac{\varepsilon_i-\varepsilon_j}{h_0}\right)\left(\overrightarrow{\varepsilon}_i-\tilde{\varepsilon}_j-\bar{\varepsilon}_i+\varepsilon_j\right)+o_p\left(1\right)\\ &= -\frac{1}{2n^{3/2}h^{d/2}h_0^2}\sum_{i=1}^n\sum_{j\neq i}K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_ig_b\left(f_i\right)k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\frac{v_{1i}\left(x\right)}{\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}}\\ &\quad+\frac{1}{2n^{3/2}h^{d/2}h_0^2}\sum_{i=1}^n\sum_{j\neq i}K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_ig_b\left(f_i\right)k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\frac{\tilde{m}\left(X_j\right)-m\left(X_j\right)}{\sigma\left(X_j\right)}\\ &\quad+\frac{1}{2n^{3/2}h^{d/2}h_0^2}\sum_{i=1}^n\sum_{j\neq i}K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_ig_b\left(f_i\right)k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\varepsilon_j\frac{\tilde{\sigma}\left(X_j\right)-\sigma\left(X_j\right)}{\tilde{\sigma}\left(X_j\right)}\\ &\quad+\frac{1}{2n^{3/2}h^{d/2}h_0^2}\sum_{i=1}^n\sum_{j\neq i}K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_ig_b\left(f_i\right)k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\frac{u_i+\delta_i}{\varphi(P_i(\boldsymbol{\beta}_2^0))^{1/2}}\frac{\varphi(P_i(\boldsymbol{\beta}_2^0))^{1/2}-\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}}{\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}}\\ &\quad+o_p\left(1\right)\\ &\equiv -\mathcal{S}_{1n\mathbf{j},211}+\mathcal{S}_{1n\mathbf{j},212}+\mathcal{S}_{1n\mathbf{j},213}+\mathcal{S}_{1n\mathbf{j},214}+o_p\left(1\right).\end{aligned}$$

For the first term, by Lemma A.2 and the fact that $v_{1i}\left(x\right)=O_p\left(\upsilon_{2n}\right)$, $\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))=\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))+O_p\left(\upsilon_{2n}\right)$

uniformly on the set $\{K_{ix}>0\}$, we have

$$\begin{aligned}
|\mathcal{S}_{1n\mathbf{j},211}| &= \left|\frac{1}{2\sqrt{nh^d}}\sum_{i=1}^n K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_i g_b(f_i)\bar{f}_i'(\varepsilon_i)\frac{v_{1i}(x)}{\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}}\right| \\
&= \left|\frac{1}{2\sqrt{nh^d}}\sum_{i=1}^n K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_i g_b(f_i) f'(\varepsilon_i)\frac{v_{1i}(x)}{\varphi(P_i(\boldsymbol{\beta}_2^0))^{1/2}}\right| + o_p(1) \\
&\leq \max_{\{K_{ix}>0\}}|v_{1i}(x)|\frac{1}{2\sqrt{nh^d}}\sum_{i=1}^n\left|K_{ix,\mathbf{j}}\varphi_i\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}\bar{\rho}_i g_b(f_i) f'(\varepsilon_i)\right| + o_p(1).
\end{aligned}$$

The first term in the last expression is $o_p(1)$ if $n^{1/2}h^{-d/2}E\left|K_{ix,\mathbf{j}}\varphi_i\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}\bar{\rho}_i g_b(f_i) f'(\varepsilon_i)\right| = o\left(\upsilon_{2n}^{-1}\right)$ by Markov inequality. Note that

$$\bar{\varepsilon}_i-\varepsilon_i=\{\varepsilon_i[\sigma(X_i)-\varphi(P_i(\boldsymbol{\beta}_2^0))^{1/2}]+\delta_i\}/\varphi(P_i(\boldsymbol{\beta}_2^0))^{1/2}=\varepsilon_i d_i+\overline{\delta}_i. \tag{C.28}$$

where $d_i\equiv\sigma(X_i)\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}-1=O_p\left(h^{p+1}\right)$ and $\overline{\delta}_i\equiv\delta_i\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}=O_p\left(h^{p+1}\right)$ uniformly on the set $\{K_{ix}>0\}$. Then by the triangle inequality,

$$\begin{aligned}
& h^{-d}E\left|K_{ix,\mathbf{j}}\varphi_i\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}\bar{\rho}_i g_b(f_i) f'(\varepsilon_i)\right| \\
= \;& h^{-d}E\left|K_{ix,\mathbf{j}}\varphi_i\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}\left[\bar{\rho}_i g_b(f_i) f'(\varepsilon_i)|X_i\right]\right| \\
= \;& h^{-d}E\left|\frac{K_{ix,\mathbf{j}}\varphi_i\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}}{d_i+1}\int_{b\leq f(\varepsilon)\leq 2b}[\psi(\varepsilon)\varepsilon+1]f'(\varepsilon)g_b(f(\varepsilon))f\left(\frac{\varepsilon-\overline{\delta}_i}{d_i+1}\right)d\varepsilon\right| \\
\leq \;& h^{-d}E\left|\frac{K_{ix,\mathbf{j}}\varphi_i\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}}{d_i+1}\int_{b\leq f(\varepsilon)\leq 2b}\rho(\varepsilon)f'(\varepsilon)g_b(f(\varepsilon))f(\varepsilon)d\varepsilon\right| \\
& +h^{-d}E\left|\frac{K_{ix,\mathbf{j}}\varphi_i\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}}{d_i+1}\int_{b\leq f(\varepsilon)\leq 2b}\rho(\varepsilon)f'(\varepsilon)g_b(f(\varepsilon))\left[f\left(\frac{\varepsilon-\overline{\delta}_i}{d_i+1}\right)-f(\varepsilon)\right]d\varepsilon\right| \\
\equiv \;& S_{n1}+S_{n2},\ \text{say.}
\end{aligned}$$

For S_{n1}, we have

$$\begin{aligned}
S_{n1} &= h^{-d}E\left|K_{ix,\mathbf{j}}\varphi_i\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}(d_i+1)^{-1}\int_{b\leq f(\varepsilon)\leq 2b}\rho(\varepsilon)f'(\varepsilon)g_b(f(\varepsilon))f(\varepsilon)d\varepsilon\right| \\
&\leq \sup_{b\leq f(\varepsilon)\leq 2b}[f(\varepsilon)g_b(f(\varepsilon))]h^{-d}E\left|K_{ix,\mathbf{j}}\varphi_i\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}(d_i+1)^{-1}\int_{b\leq f(\varepsilon)\leq 2b}\rho(\varepsilon)f'(\varepsilon)d\varepsilon\right| \\
&\leq Ch^{-d}E\left|K_{ix,\mathbf{j}}\varphi_i\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}(d_i+1)^{-1}\right|\left|\int_{b\leq f(\varepsilon)\leq 2b}\rho(\varepsilon)f'(\varepsilon)d\varepsilon\right| \\
&\leq Ch^{-d}E\left|K_{ix,\mathbf{j}}\varphi_i\varphi(P_i(\boldsymbol{\beta}_2^0))^{-1/2}(d_i+1)^{-1}\right|\left\{\int_{b\leq f(\varepsilon)\leq 2b}\rho(\varepsilon)^2 f(\varepsilon)d\varepsilon\int_{b\leq f(\varepsilon)\leq 2b}\psi(\varepsilon)f(\varepsilon)d\varepsilon\right\}^{1/2} \\
&= O(h^{\epsilon})
\end{aligned}$$

where the third inequality follows from the Hölder inequality and the independence between X_i and ε_i. By a Taylor expansion, $f\left(\frac{\varepsilon-\overline{\delta}_i}{1+d_i}\right)-f(\varepsilon)\simeq -f'(\varepsilon)\left(\overline{\delta}_i+d_i\varepsilon\right)$. With this, we can readily show that $S_{n2}=O(h^{\epsilon})$. Consequently, $|\mathcal{S}_{1n\mathbf{j},211}|=O_p(\upsilon_{2n}\sqrt{nh^d}h^{\epsilon})=o_p(1)$.

For $\mathcal{S}_{1n\mathbf{j},212}$, using (C.2) we can write

$$\begin{aligned}\mathcal{S}_{1n\mathbf{j},212} &= \frac{1}{2n^{3/2}h^{d/2}h_0^2}\sum_{i=1}^{n}\sum_{j\neq i}K_{ix,\mathbf{j}}\varphi_i\bar{\rho}_i g_b(f_i)\,k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\frac{\tilde{m}(X_j)-m(X_j)}{\sigma(X_j)} \\ &= \frac{1}{2n^{3/2}h^{d/2}h_0^2}\sum_{i=1}^{n}\sum_{j\neq i}\frac{K_{ix,\mathbf{j}}\varphi_i}{\sigma(X_j)}\bar{\rho}_i g_b(f_i)\,k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)e_1^{\top}M_n^{-1}(X_j)\,U_{1,n}(X_j) \\ &\quad -\frac{1}{2n^{3/2}h^{d/2}h_0^2}\sum_{i=1}^{n}\sum_{j\neq i}\frac{K_{ix,\mathbf{j}}\varphi_i}{\sigma(X_j)}\bar{\rho}_i g_b(f_i)\,k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)e_1^{\top}M_n^{-1}(X_j)\,B_{1,n}(X_j) \\ &\equiv \mathcal{S}_{1n\mathbf{j},212a}+\mathcal{S}_{1n\mathbf{j},212b}. \end{aligned} \tag{C.29}$$

Recall $\tilde{\mathbf{Z}}_i$ is defined analogously to $\mathbf{Z}_i$ with h_1 in place of h. So $\mathcal{S}_{1n\mathbf{j},212a}$ can be written as

$$\mathcal{S}_{1n\mathbf{j},212a}=\sum_{i=1}^{n}\sum_{j\neq i}\varsigma_{2n}(\varepsilon_i,\varepsilon_j)+\sum_{i=1}^{n}\sum_{j\neq i}\sum_{l\neq i,l\neq j}\varsigma_{3n}(\varepsilon_i,\varepsilon_j,\varepsilon_l),$$

where $\varsigma_{2n}(\varepsilon_i,\varepsilon_j)=\frac{1}{2n^{5/2}h^{d/2}h_1^dh_0^2}\frac{K_{ix,\mathbf{j}}\varphi_i}{\sigma(X_j)}\bar{\rho}_i g_b(f_i)\,k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)e_1^{\top}M_n^{-1}(X_j)\,\tilde{\mathbf{Z}}_iK\left(\frac{X_j-X_i}{h_1}\right)u_i$ and $\varsigma_{3n}(\varepsilon_i,\varepsilon_j,\varepsilon_l)$ $=\frac{1}{2n^{5/2}h^{d/2}h_1^dh_0^2}\frac{K_{ix,\mathbf{j}}\varphi_i}{2\sigma(X_j)}\bar{\rho}_i g_b(f_i)\,k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)e_1^{\top}M_n^{-1}(X_j)\,\tilde{\mathbf{Z}}_iK\left(\frac{X_j-X_l}{h_1}\right)u_l$. Let $\mathbb{X}\equiv\{X_1,...,X_n\}$. Then $E[\mathcal{S}_{1n\mathbf{j},212a}|\mathbb{X}]=\sum_{i=1}^{n}\sum_{j\neq i}E[\varsigma_{2n}(z_i,z_j)|\mathbb{X}]=O_p\left(n^{-1/2}h^{d/2}b^{-1}\right)=o_p(1)$. For the variance of $\mathcal{S}_{1n\mathbf{j},212a}$, it is easy to show that

$$\begin{aligned}\mathrm{Var}\left[\sum_{i=1}^{n}\sum_{j\neq i}\varsigma_{2n}(\varepsilon_i,\varepsilon_j)|\mathbb{X}\right] &= O\left(n^2\right)E\left[\varsigma_{2n}(\varepsilon_i,\varepsilon_j)^2+\varsigma_{2n}(\varepsilon_i,\varepsilon_j)\varsigma_{2n}(\varepsilon_j,\varepsilon_i)|\mathbb{X}\right] \\ &\quad +O\left(n^3\right)E\left[\varsigma_{2n}(\varepsilon_i,\varepsilon_j)\varsigma_{2n}(\varepsilon_l,\varepsilon_j)+\varsigma_{2n}(\varepsilon_i,\varepsilon_j)\varsigma_{2n}(\varepsilon_i,\varepsilon_i)|\mathbb{X}\right] \\ &= O_p\left(n^{-3}h^{-d-4}b^{-2}\right)+O_p\left(n^{-2}b^{-2}\right)=o_p(1).\end{aligned}$$

Similarly, one can show that $E(\varsigma_{3n}(z_i,z_j,z_l)|\mathbb{X})=0$ and $\mathrm{Var}\left[\sum_{i=1}^{n}\sum_{j\neq i}\sum_{l\neq i,l\neq j}\varsigma_{3n}(z_i,z_j,z_l)|\mathbb{X}\right]=o_p(1)$. Consequently, $\mathcal{S}_{1n\mathbf{j},212a}=o_p(1)$ by the conditional Chebyshev inequality. For $\mathcal{S}_{1n\mathbf{j},212b}$, we have $\mathcal{S}_{1n\mathbf{j},212b}=O_p(n^{1/2}h^{d/2}h_1^{p+1})=o_p(1)$. Thus we have shown that $\mathcal{S}_{1n\mathbf{j},212}=o_p(1)$. By analogous arguments, Lemma A.1, and (C.8), we can show that $\mathcal{S}_{1n\mathbf{j},21s}=o_p(1)$ for $s=3,4$. It follows that $\mathcal{S}_{1n\mathbf{j},21}=o_p(1)$.

For $\mathcal{S}_{1n\mathbf{j},22}$, we make the following decomposition:

$$\mathcal{S}_{1n\mathbf{j},22}=\frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n}K_{ix,\mathbf{j}}\varphi_i q_{2,i}\left(\boldsymbol{\beta}^0\right)g_b(f_i)\left\{\mathcal{V}(\bar{\varepsilon}_i)+\mathcal{B}(\bar{\varepsilon}_i)\right\}\equiv\mathcal{S}_{1n\mathbf{j},221}+\mathcal{S}_{1n\mathbf{j},222},$$

where

$$\mathcal{V}(\bar{\varepsilon}_i) = \frac{1}{nh_0}\sum_{j\neq i}\left\{k_0\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)-E_j\left[k_0\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right]\right\}, \tag{C.30}$$

$$\mathcal{B}(\bar{\varepsilon}_i) = \frac{1}{nh_0}\sum_{j\neq i}E_j\left[k_0\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right]-f(\bar{\varepsilon}_i), \tag{C.31}$$

and E_j indicates expectation with respect to the variable indexed by j. Writing $\mathcal{S}_{1n\mathbf{j},221}$ as a second order degenerate statistic we verify that $E\left[\mathcal{S}_{1n\mathbf{j},221}\right]^2 = o(1)$ and thus $\mathcal{S}_{1n\mathbf{j},221} = o_p(1)$. For $\mathcal{S}_{1n\mathbf{j},222}$, we verify that $\mathcal{S}_{1n\mathbf{j},222} = O_p(n^{1/2}h^{d/2}h_0^{p+1}) = o_p(1)$. Consequently, $\mathcal{S}_{1n\mathbf{j},22} = o_p(1)$. This concludes the proof of the lemma. ■

Proof of Lemma B.5. By a geometric expansion: $\tilde{f}_i = f^{-1} - (\tilde{f}_i - f)/f^2 + (\tilde{f}_i - f)^2/(f^2\tilde{f}_i)$ where $\tilde{f}_i = \tilde{f}_i(\bar{\varepsilon}_i)$, we have

$$
\begin{aligned}
\mathcal{S}_{2n\mathbf{j}} &= \frac{1}{\sqrt{nh^d}}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\left\{\tilde{G}_i\left[\tilde{q}_{2,i}\left(\boldsymbol{\beta}^0\right) - q_{2,i}\left(\boldsymbol{\beta}^0\right)\right]\right\} \\
&= -\frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\varphi_i \frac{\tilde{f}_i'(\bar{\varepsilon}_i) - f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\bar{\varepsilon}_i\tilde{G}_i \\
&\quad + \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\varphi_i \frac{\tilde{f}_i'(\bar{\varepsilon}_i)\left[\tilde{f}_i(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)\right]}{f(\bar{\varepsilon}_i)}\bar{\varepsilon}_i\tilde{G}_i \\
&\quad - \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\varphi_i \frac{\tilde{f}_i'(\bar{\varepsilon}_i)\left[\tilde{f}_i(\bar{\varepsilon}_i) - f(\bar{\varepsilon}_i)\right]^2}{f^2(\bar{\varepsilon}_i)\tilde{f}_i(\bar{\varepsilon}_i)}\bar{\varepsilon}_i\tilde{G}_i \\
&\equiv -\mathcal{S}_{2n\mathbf{j},1} + \mathcal{S}_{2n\mathbf{j},2} - \mathcal{S}_{2n\mathbf{j},3}.
\end{aligned}
$$

where recall $\varphi_i \equiv \varphi'(P_i(\boldsymbol{\beta}_2^0))/\varphi(P_i(\boldsymbol{\beta}_2^0))$. It suffices to show that each of these three terms is $o_p(1)$. For $\mathcal{S}_{2n\mathbf{j},1}$, noticing that $G_b(\tilde{f}_i) - G_b(f_i) = g_b(f_i)(\tilde{f}_i - f_i) + \frac{1}{2}g_b'(f_i^*)(\tilde{f}_i - f_i)^2$, we can apply Lemma A.2 and show that

$$
\begin{aligned}
\mathcal{S}_{2n\mathbf{j},1} &= \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\varphi_i \frac{\tilde{f}_i'(\bar{\varepsilon}_i) - \overline{f}'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\bar{\varepsilon}_i G_b(f_i) \\
&\quad + \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} K_{ix,\mathbf{j}}\varphi_i \frac{\overline{f}'(\bar{\varepsilon}_i) - f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\bar{\varepsilon}_i G_b(f_i) + o_p(1) \\
&\equiv \mathcal{S}_{2n\mathbf{j},11} + \mathcal{S}_{2n\mathbf{j},12} + o_p(1).
\end{aligned}
$$

For the first term, we have

$$
\begin{aligned}
\mathcal{S}_{2n\mathbf{j},11} &= \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} \frac{K_{ix,\mathbf{j}}\varphi_i}{f(\bar{\varepsilon}_i)}\frac{1}{nh_0^2}\sum_{j\neq i}\left[k_0'\left(\frac{\bar{\varepsilon}_i - \tilde{\varepsilon}_j}{h_0}\right) - k_0'\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\right]\bar{\varepsilon}_i G_b(f_i) \\
&= \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} \frac{K_{ix,\mathbf{j}}\varphi_i}{f(\bar{\varepsilon}_i)}\frac{1}{nh_0^3}\sum_{j\neq i} k_0''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)(\tilde{\varepsilon}_j - \varepsilon_j)\bar{\varepsilon}_i G_b(f_i) + o_p(1) \\
&= \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} \frac{K_{ix,\mathbf{j}}\varphi_i}{f(\bar{\varepsilon}_i)}\frac{1}{nh_0^3}\sum_{j\neq i} k_0''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\frac{m(X_j) - \tilde{m}(X_j)}{\sigma(X_j)}\bar{\varepsilon}_i G_b(f_i) \\
&\quad + \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n} \frac{K_{ix,\mathbf{j}}\varphi_i}{f(\bar{\varepsilon}_i)}\frac{1}{nh_0^3}\sum_{j\neq i} k_0''\left(\frac{\bar{\varepsilon}_i - \varepsilon_j}{h_0}\right)\varepsilon_j\frac{\sigma(X_j) - \tilde{\sigma}(X_j)}{\sigma(X_j)}\bar{\varepsilon}_i G_b(f_i) + o_p(1) \\
&\equiv \mathcal{S}_{2n\mathbf{j},111} + \mathcal{S}_{2n\mathbf{j},112} + o_p(1), \text{ say.}
\end{aligned}
$$

Write

$$
\begin{aligned}
\mathcal{S}_{2n\mathbf{j},111} &= \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n}\frac{K_{ix,\mathbf{j}}\varphi_i}{f(\bar{\varepsilon}_i)}\frac{1}{nh_0^3}\sum_{j\neq i}k_0''\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\frac{e_1^\top M_n^{-1}(X_j)\,U_{1,n}(X_j)}{\sigma(X_j)}\bar{\varepsilon}_i G_b(f_i) \\
&\quad + \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n}\frac{K_{ix,\mathbf{j}}\varphi_i}{f(\bar{\varepsilon}_i)}\frac{1}{nh_0^3}\sum_{j\neq i}k_0''\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\frac{e_1^\top M_n^{-1}(X_j)\,B_{1,n}(X_j)}{\sigma(X_j)}\bar{\varepsilon}_i G_b(f_i) \\
&\equiv \mathcal{S}_{2n\mathbf{j},111a}+\mathcal{S}_{2n\mathbf{j},111b}.
\end{aligned}
$$

Writing $\mathcal{S}_{2n\mathbf{j},111a}$ as a third order U-statistic, we can show that $\mathcal{S}_{2n\mathbf{j},111a} = O_p(h^{d/2}) = o_p(1)$ by conditional moment calculations and conditional Chebyshev inequality. For $\mathcal{S}_{2n\mathbf{j},111b}$, we have $\mathcal{S}_{2n\mathbf{j},111b} = O_p(\sqrt{nh^d}h_1^{p+1}) = o_p(1)$. Similarly, we can verify that $\mathcal{S}_{2n\mathbf{j},112} = o_p(1)$. Consequently $\mathcal{S}_{2n\mathbf{j},11} = o_p(1)$. For $\mathcal{S}_{2n\mathbf{j},12}$, we have

$$
\begin{aligned}
\mathcal{S}_{2n\mathbf{j},12} &= \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n}\frac{K_{ix,\mathbf{j}}\varphi_i}{f(\bar{\varepsilon}_i)}\frac{\overline{f}'(\bar{\varepsilon}_i)-f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\bar{\varepsilon}_i G_b(f_i) \\
&= \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n}\frac{K_{ix,\mathbf{j}}\varphi_i}{f(\bar{\varepsilon}_i)}\left\{\frac{1}{nh_0^2}\sum_{j\neq i}\left\{k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)-E_j\left[k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right]\right\}\right\}\bar{\varepsilon}_i G_b(f_i) \\
&\quad + \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n}\frac{K_{ix,\mathbf{j}}\varphi_i}{f(\bar{\varepsilon}_i)}\left\{\frac{1}{nh_0^2}\sum_{j\neq i}E_j\left[k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right]-f'(\bar{\varepsilon}_i)\right\}\bar{\varepsilon}_i G_b(f_i) \\
&= \mathcal{S}_{2n\mathbf{j},121}+\mathcal{S}_{2n\mathbf{j},122},
\end{aligned}
$$

where E_j indicates expectation with respect to the variable indexed by j. Noting $\mathcal{S}_{2n\mathbf{j},121}$ is a second order statistic, it is easy to verify that $E\left[\mathcal{S}_{2n\mathbf{j},121}\right]^2 = O(h^d) = o(1)$, implying that $\mathcal{S}_{2n\mathbf{j},121} = o_p(1)$. For $\mathcal{S}_{2n\mathbf{j},122}$, noticing that

$$
\frac{1}{nh_0^2}\sum_{j\neq i}E_j\left[k_0'\left(\frac{\bar{\varepsilon}_i-\varepsilon_j}{h_0}\right)\right]-f'(\bar{\varepsilon}_i) = h_0^{p+1}f^{(p+2)}(\bar{\varepsilon}_i)\int k_0(u)\,u^{p+1}du,
$$

we can show that $\mathcal{S}_{2n\mathbf{j},122} = O_p(\sqrt{nh^d}h_0^{p+1}) = o_p(1)$. Consequently, $\mathcal{S}_{2n\mathbf{j},12} = o_p(1)$ and $\mathcal{S}_{2n\mathbf{j},1} = o_p(1)$.

For $\mathcal{S}_{2n\mathbf{j},2}$, we can easily show that

$$
\mathcal{S}_{2n\mathbf{j},2} = \frac{1}{2\sqrt{nh^d}}\sum_{i=1}^{n}\frac{K_{ix,\mathbf{j}}\varphi_i f'(\bar{\varepsilon}_i)}{f(\bar{\varepsilon}_i)}\left[\tilde{f}_i(\bar{\varepsilon}_i)-f(\bar{\varepsilon}_i)\right]\bar{\varepsilon}_i G_b(f_i)+o_p(1).
$$

The rest of the proof is similar to that of $\mathcal{S}_{2n\mathbf{j},1}$ and thus omitted. For $\mathcal{S}_{2n\mathbf{j},3}$, by Lemma A.2, $\mathcal{S}_{2n\mathbf{j},3} = O_p(\sqrt{nh^d}b^{-2}\nu_{3n}^2) = o_p(1)$. This concludes the proof of the lemma. ■

Proof of Lemma B.6. Write $\mathcal{S}_{3n\mathbf{j}} = \frac{1}{2}\{\mathcal{S}_{3n\mathbf{j},1} - \mathcal{S}_{3n\mathbf{j},2} + \mathcal{S}_{3n\mathbf{j},3} - \mathcal{S}_{3n\mathbf{j},4}\}$, where

$$\begin{aligned}
\mathcal{S}_{3n\mathbf{j},1} &\equiv \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K_{ix,\mathbf{j}} \Big\{ \log\left(\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right)\right) g_b\left(\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right)\right) \tilde{f}_i'\left(\overrightarrow{\varepsilon}_i\right) \overrightarrow{\varepsilon}_i \varphi_i(\tilde{\boldsymbol{\beta}}_2) \\
&\quad - \log\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right) g_b\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right) \tilde{f}_i'\left(\bar{\varepsilon}_i\right) \bar{\varepsilon}_i \varphi_i \Big\}, \\
\mathcal{S}_{3n\mathbf{j},2} &\equiv \frac{1}{2\sqrt{nh^d}} \sum_{i=1}^n K_{ix,\mathbf{j}} \Big\{ \log\varphi\left(P_i(\tilde{\boldsymbol{\beta}}_2)\right) g_b\left(\tilde{f}_i\left(\overrightarrow{\varepsilon}_i\right)\right) \tilde{f}_i'\left(\overrightarrow{\varepsilon}_i\right) \overrightarrow{\varepsilon}_i \varphi_i(\tilde{\boldsymbol{\beta}}_2) \\
&\quad - \log\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right) g_b\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right) \tilde{f}_i'\left(\bar{\varepsilon}_i\right) \bar{\varepsilon}_i \varphi_i \Big\}, \\
\mathcal{S}_{3n\mathbf{j},3} &\equiv \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K_{ix,\mathbf{j}} \Big\{ \log\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right) g_b\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right) \tilde{f}_i'\left(\bar{\varepsilon}_i\right) - \log\left(f\left(\bar{\varepsilon}_i\right)\right) g_b\left(f\left(\bar{\varepsilon}_i\right)\right) f'\left(\bar{\varepsilon}_i\right) \Big\} \bar{\varepsilon}_i \varphi_i, \\
\mathcal{S}_{3n\mathbf{j},4} &\equiv \frac{1}{2\sqrt{nh^d}} \sum_{i=1}^n K_{ix,\mathbf{j}} \log\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right) \Big\{ g_b\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right) \tilde{f}_i'\left(\bar{\varepsilon}_i\right) - g_b\left(f\left(\bar{\varepsilon}_i\right)\right) f'\left(\bar{\varepsilon}_i\right) \Big\} \bar{\varepsilon}_i \varphi_i.
\end{aligned}$$

where $\varphi_i(\boldsymbol{\beta}_2) \equiv \varphi'(P_i(\boldsymbol{\beta}_2))/\varphi(P_i(\boldsymbol{\beta}_2))$ and $\varphi_i = \varphi_i\left(\boldsymbol{\beta}_2^0\right)$. We will only show that $\mathcal{S}_{3n\mathbf{j},1} = o_p(1)$ since the proofs of $\mathcal{S}_{3n\mathbf{j},s} = o_p(1)$ for $s = 2, 3, 4$ are similar.

For $\mathcal{S}_{3n\mathbf{j},1}$, noticing that $\tilde{v}_{2i}(x) = (\varphi(P_i\left(\boldsymbol{\beta}_2^0\right))^{1/2} - \varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2})/\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}$ and $\tilde{v}_{1i}(x) = v_{1i}(x)$ $/\varphi(P_i(\tilde{\boldsymbol{\beta}}_2))^{1/2}$ are both $O_p(\upsilon_{2n})$ uniformly in i on the set $\{K_{ix} > 0\}$, and $\overrightarrow{\varepsilon}_i - \bar{\varepsilon}_i = \bar{\varepsilon}_i \tilde{v}_{2i}(x) - \tilde{v}_{1i}(x)$, we can show that

$$\begin{aligned}
\mathcal{S}_{3n\mathbf{j},1} &= \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K_{ix,\mathbf{j}} \varphi_i \Big\{ \tilde{\psi}_i\left(\bar{\varepsilon}_i\right) g_b\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right) \tilde{f}_i'\left(\bar{\varepsilon}_i\right) \bar{\varepsilon}_i \left\{\bar{\varepsilon}_i \tilde{v}_{2i}(x) - \tilde{v}_{1i}(x)\right\} \\
&\quad + \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K_{ix,\mathbf{j}} \varphi_i \log\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right) g_b'\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right) \tilde{f}_i'\left(\bar{\varepsilon}_i\right) \bar{\varepsilon}_i \left\{\bar{\varepsilon}_i \tilde{v}_{2i}(x) - \tilde{v}_{1i}(x)\right\} \\
&\quad + \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K_{ix,\mathbf{j}} \varphi_i \log\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right) g_b\left(\tilde{f}_i\left(\bar{\varepsilon}_i\right)\right) \tilde{f}_i''\left(\bar{\varepsilon}_i\right) \bar{\varepsilon}_i \left\{\bar{\varepsilon}_i \tilde{v}_{2i}(x) - \tilde{v}_{1i}(x)\right\} + o_p(1) \\
&\equiv \mathcal{S}_{3n\mathbf{j},11} + \mathcal{S}_{3n\mathbf{j},12} + \mathcal{S}_{3n\mathbf{j},13} + o_p(1).
\end{aligned}$$

By Lemma A.1, we can show

$$\mathcal{S}_{3n\mathbf{j},11} \simeq \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K_{ix,\mathbf{j}} \varphi_i \psi\left(\bar{\varepsilon}_i\right) g_b\left(f\left(\bar{\varepsilon}_i\right)\right) f'\left(\bar{\varepsilon}_i\right) \bar{\varepsilon}_i \left\{\bar{\varepsilon}_i \tilde{v}_{2i}(x) - \tilde{v}_{1i}(x)\right\}, \tag{C.32}$$

$$\mathcal{S}_{3n\mathbf{j},12} \simeq \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K_{ix,\mathbf{j}} \varphi_i \log\left(f\left(\bar{\varepsilon}_i\right)\right) g_b'\left(f\left(\bar{\varepsilon}_i\right)\right) f'\left(\bar{\varepsilon}_i\right) \bar{\varepsilon}_i \left\{\bar{\varepsilon}_i \tilde{v}_{2i}(x) - \tilde{v}_{1i}(x)\right\}, \tag{C.33}$$

$$\mathcal{S}_{3n\mathbf{j},13} \simeq \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n K_{ix,\mathbf{j}} \varphi_i \log\left(f\left(\bar{\varepsilon}_i\right)\right) g_b\left(f\left(\bar{\varepsilon}_i\right)\right) f''\left(\bar{\varepsilon}_i\right) \bar{\varepsilon}_i \left\{\bar{\varepsilon}_i \tilde{v}_{2i}(x) - \tilde{v}_{1i}(x)\right\}. \tag{C.34}$$

The rest of the proof relies on the repeated applications of the dominated convergence arguments. For example, the right hand side of (C.32) is smaller than

$$\begin{aligned}
&\frac{1}{\sqrt{nh^d}} \max_{\{K_{ix}>0\}} \left|\tilde{v}_{2i}(x)\right| \sum_{i=1}^n \left|K_{ix,\mathbf{j}} \varphi_i \psi\left(\bar{\varepsilon}_i\right) g_b\left(f\left(\bar{\varepsilon}_i\right)\right) f'\left(\bar{\varepsilon}_i\right) \bar{\varepsilon}_i^2\right| \\
&+ \frac{1}{\sqrt{nh^d}} \max_{\{K_{ix}>0\}} \left|\tilde{v}_{1i}(x)\right| \sum_{i=1}^n \left|K_{ix,\mathbf{j}} \varphi_i \psi\left(\bar{\varepsilon}_i\right) g_b\left(f\left(\bar{\varepsilon}_i\right)\right) f'\left(\bar{\varepsilon}_i\right) \bar{\varepsilon}_i\right|.
\end{aligned}$$

Noting that

$$
\begin{aligned}
E\left|K_{ix,\mathbf{j}}\varphi_i\psi\left(\bar{\varepsilon}_i\right)g_b\left(f\left(\bar{\varepsilon}_i\right)\right)f'\left(\bar{\varepsilon}_i\right)\bar{\varepsilon}_i^r\right| &= E\left[\left|\frac{K_{ix,\mathbf{j}}\varphi_i}{d_i+1}\int\psi\left(\varepsilon\right)g_b\left(f\left(\varepsilon\right)\right)f'\left(\varepsilon\right)\varepsilon^r\right|f\left(\frac{\varepsilon-\overline{\delta}_i}{d_i+1}\right)d\varepsilon\right] \\
&\leq \sup_{\varepsilon}\left[g_b\left(f\left(\varepsilon\right)\right)f\left(\varepsilon\right)\right]E\left|\frac{K_{ix,\mathbf{j}}\varphi_i}{d_i+1}\right|\int_{b\leq f(\varepsilon)\leq 2b}\frac{f'\left(\varepsilon\right)^2}{f\left(\varepsilon\right)}\left|\varepsilon^r\right|d\varepsilon+O\left(h^{\epsilon}\right) \\
&\leq Cf\left(\varepsilon\right)\int_{b\leq f(\varepsilon)\leq 2b}\left|\frac{f'\left(\varepsilon\right)^2}{f\left(\varepsilon\right)}\varepsilon^r\right|d\varepsilon+O\left(h^{\epsilon}\right) \\
&\leq C\int_{b\leq f(\varepsilon)\leq 2b}\left|\frac{f'\left(\varepsilon\right)^2}{f\left(\varepsilon\right)}\varepsilon^r\right|d\varepsilon+O\left(h^{p+1}\right)=O\left(b^{(\gamma-1)/(2\gamma)}+h^{\epsilon}\right),
\end{aligned}
$$

where the last equality follows from similar argument to the proof of Lemma B.3, we have $\mathcal{S}_{3n\mathbf{j},11} = O_p(\upsilon_{2n}\sqrt{nh^d}(b^{(\gamma-1)/(2\gamma)}+h^{\epsilon})) = o_p\left(1\right)$. Similarly, we can show that $\mathcal{S}_{3n\mathbf{j},1s} = o_p\left(1\right),\ s=2,3.$ ■

Proof of Lemma B.7. Observe that

$$
\begin{aligned}
\mathcal{R}_{1n} &= \frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)\bar{H}^{-1}[\tilde{G}_i\tilde{s}_i\left(\boldsymbol{\beta}^0\right)\tilde{s}_i\left(\boldsymbol{\beta}^0\right)^{\top}-G_is_i\left(\boldsymbol{\beta}^0\right)s_i\left(\boldsymbol{\beta}^0\right)^{\top}]\otimes(\tilde{\mathbf{X}}_i\tilde{\mathbf{X}}_i^{\top})\bar{H}^{-1} \\
&= \frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)\bar{H}^{-1}G_i[\tilde{s}_i\left(\boldsymbol{\beta}^0\right)\tilde{s}_i\left(\boldsymbol{\beta}^0\right)^{\top}-s_i\left(\boldsymbol{\beta}^0\right)s_i\left(\boldsymbol{\beta}^0\right)^{\top}]\otimes(\tilde{\mathbf{X}}_i\tilde{\mathbf{X}}_i^{\top})\bar{H}^{-1} \\
&\quad+\frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)\bar{H}^{-1}\left(\tilde{G}_i-G_i\right)s_i\left(\boldsymbol{\beta}^0\right)s_i\left(\boldsymbol{\beta}^0\right)^{\top}\otimes(\tilde{\mathbf{X}}_i\tilde{\mathbf{X}}_i^{\top})\bar{H}^{-1} \\
&\quad+\frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)\bar{H}^{-1}\left(\tilde{G}_i-G_i\right)[\tilde{s}_i\left(\boldsymbol{\beta}^0\right)\tilde{s}_i\left(\boldsymbol{\beta}^0\right)^{\top}-s_i\left(\boldsymbol{\beta}^0\right)s_i\left(\boldsymbol{\beta}^0\right)^{\top}]\otimes(\tilde{\mathbf{X}}_i\tilde{\mathbf{X}}_i^{\top})\bar{H}^{-1} \\
&\equiv \mathcal{R}_{1n,1}+\mathcal{R}_{1n,2}+\mathcal{R}_{1n,3},\ \text{say.}
\end{aligned}
$$

It suffices to prove the lemma by showing that $\mathcal{R}_{1n,r} = o_p\left(1\right)$ for $r=1,2,3$. We only prove $\mathcal{R}_{1n,1} = o_p\left(1\right)$ and $\mathcal{R}_{1n,2} = o_p\left(1\right)$ as $\mathcal{R}_{1n,3}$ is a smaller order term and can be studied analogously.

First, we show that $\mathcal{R}_{1n,1} = o_p\left(1\right)$. Note that

$$
\tilde{s}_i\left(\boldsymbol{\beta}^0\right)\tilde{s}_i\left(\boldsymbol{\beta}^0\right)^{\top} = \begin{bmatrix} \frac{\tilde{\psi}_i^2(\bar{\varepsilon}_i)}{\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} & \frac{\varphi'(P_i(\boldsymbol{\beta}_2))\tilde{\psi}_i(\bar{\varepsilon}_i)\left[\tilde{\psi}_i(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1\right]}{2\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{3/2}} \\ \frac{\varphi'(P_i(\boldsymbol{\beta}_2))\tilde{\psi}_i(\bar{\varepsilon}_i)\left[\tilde{\psi}_i(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1\right]}{2\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{3/2}} & \frac{\left[\varphi'(P_i(\boldsymbol{\beta}_2))\right]^2\left[\tilde{\psi}_i(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1\right]^2}{4\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^2} \end{bmatrix},
$$

and $s_i\left(\boldsymbol{\beta}^0\right)s_i\left(\boldsymbol{\beta}^0\right)^{\top}$ has a similar expression with $\psi\left(\bar{\varepsilon}_i\right)$ in the place of $\tilde{\psi}_i\left(\bar{\varepsilon}_i\right)$. It follows that

$$
\begin{aligned}
\mathcal{R}_{1n,1} &= \frac{1}{nh^d}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)G_i \\
&\quad\times\begin{bmatrix} \frac{\tilde{\psi}_i^2(\bar{\varepsilon}_i)-\psi^2(\bar{\varepsilon}_i)}{\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} & \frac{\varphi_i\left\{\tilde{\psi}_i(\bar{\varepsilon}_i)\left[\tilde{\psi}_i(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1\right]-\psi(\bar{\varepsilon}_i)[\psi(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1]\right\}}{2\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{1/2}} \\ \frac{\varphi_i\left\{\tilde{\psi}_i(\bar{\varepsilon}_i)\left[\tilde{\psi}_i(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1\right]-\psi(\bar{\varepsilon}_i)[\psi(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1]\right\}}{2\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{1/2}} & \frac{\varphi_i^2\left[\tilde{\psi}_i(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1\right]^2-[\psi(\bar{\varepsilon}_i)\bar{\varepsilon}_i+1]^2}{4} \end{bmatrix}\otimes(\mathbf{Z}_i\mathbf{Z}_i^{\top}) \\
&\equiv \begin{bmatrix} \mathcal{R}_{1n,1,11} & \mathcal{R}_{1n,1,12} \\ \mathcal{R}_{1n,1,21} & \mathcal{R}_{1n,1,22} \end{bmatrix},\ \text{say,}
\end{aligned}
$$

where recall $\varphi_i = \varphi'\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)/\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)$, $\mathcal{R}_{1n,1,21} = \mathcal{R}_{1n,1,12}^{\top}$, and $\mathcal{R}_{1n,1,rs}$, $r,s=1,2$, are all $N \times N$ matrices. We need to show that $\mathcal{R}_{1n,1,11}$, $\mathcal{R}_{1n,1,12}$ and $\mathcal{R}_{1n,1,22}$ are all $o_p(1)$. Noting that

$$\begin{aligned}
\tilde{\psi}_i^2\left(\bar{\varepsilon}_i\right) - \psi^2\left(\bar{\varepsilon}_i\right) &= \frac{\tilde{f}_i'\left(\bar{\varepsilon}_i\right)^2 f\left(\bar{\varepsilon}_i\right)^2 - \tilde{f}_i\left(\bar{\varepsilon}_i\right)^2 f'\left(\bar{\varepsilon}_i\right)^2}{\tilde{f}_i\left(\bar{\varepsilon}_i\right)^2 f\left(\bar{\varepsilon}_i\right)^2} \\
&= \frac{\left[\tilde{f}_i'\left(\bar{\varepsilon}_i\right)^2 - f_i'\left(\bar{\varepsilon}_i\right)^2\right] f\left(\bar{\varepsilon}_i\right)^2 + \left[f\left(\bar{\varepsilon}_i\right)^2 - \tilde{f}_i\left(\bar{\varepsilon}_i\right)^2\right] f_i'\left(\bar{\varepsilon}_i\right)^2}{\tilde{f}_i\left(\bar{\varepsilon}_i\right)^2 f\left(\bar{\varepsilon}_i\right)^2},
\end{aligned}$$

we have

$$\begin{aligned}
\mathcal{R}_{1n,1,11} &= \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) G_i \varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{-1}\left[\tilde{\psi}_i^2\left(\bar{\varepsilon}_i\right) - \psi^2\left(\bar{\varepsilon}_i\right)\right]\mathbf{Z}_i\mathbf{Z}_i^{\top} \\
&= \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) G_i \varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{-1}\tilde{f}_i\left(\bar{\varepsilon}_i\right)^{-2}\left[\tilde{f}_i'\left(\bar{\varepsilon}_i\right)^2 - f_i'\left(\bar{\varepsilon}_i\right)^2\right]\mathbf{Z}_i\mathbf{Z}_i^{\top} \\
&\quad + \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) G_i \varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{-1}\tilde{f}_i\left(\bar{\varepsilon}_i\right)^{-2}\psi\left(\bar{\varepsilon}_i\right)^2\left[\tilde{f}_i\left(\bar{\varepsilon}_i\right)^2 - f_i\left(\bar{\varepsilon}_i\right)^2\right]\mathbf{Z}_i\mathbf{Z}_i^{\top} \\
&\equiv \mathcal{R}_{1n,1,11,a} + \mathcal{R}_{1n,1,11,b} \text{ say.}
\end{aligned}$$

Noting that $G_i\tilde{f}_i\left(\bar{\varepsilon}_i\right)^{-2} = O\left(b^{-2}\right)$, by Lemma A.2, we have

$$\begin{aligned}
\left\|\mathcal{R}_{1n,1,11,a}\right\| &\leq O_p\left(\upsilon_{3n,1}b^{-2}\right)\frac{1}{nh^d}\sum_{i=1}^n\left\|K\left(\frac{x-X_i}{h}\right)G_i\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{-1}\mathbf{Z}_i\mathbf{Z}_i^{\top}\right\| \\
&= O_p\left(\upsilon_{3n,1}b^{-2}\right)O_p(1) = O_p\left(\upsilon_{3n,1}b^{-2}\right) = o_p(1).
\end{aligned}$$

By the same token, $\left|\mathcal{R}_{1n,1,11,b}\right| = o_p(1)$. Thus $\mathcal{R}_{1n,1,11} = o_p(1)$. Analogously, we can show $\mathcal{R}_{1n,1,12} = o_p(1)$ and $\mathcal{R}_{1n,1,22} = o_p(1)$. Hence we have shown that $\mathcal{R}_{1n,1} = o_p(1)$.

Now, we show that $\mathcal{R}_{1n,2} = o_p(1)$. By (C.26) and Markov inequality, we have

$$\begin{aligned}
\left|\mathcal{R}_{1n,2}\right| &\leq O_p\left(h^{\epsilon}\right)\frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)\left\|s_i\left(\boldsymbol{\beta}^0\right)s_i\left(\boldsymbol{\beta}^0\right)^{\top}\otimes\left(\mathbf{Z}_i\mathbf{Z}_i^{\top}\right)\right\| \\
&= O_p\left(h^{\epsilon}\right)O_p(1) = O_p(1).
\end{aligned}$$

This completes the proof of the lemma. ■

Proof of Lemma B.8. Observe that

$$\begin{aligned}
\mathcal{R}_{2n} &= \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)\bar{H}^{-1}\left[\tilde{G}_i\frac{\partial\tilde{s}_i\left(\boldsymbol{\beta}^0\right)}{\partial\boldsymbol{\beta}^{\top}} - G_i\frac{\partial s_i\left(\boldsymbol{\beta}^0\right)}{\partial\boldsymbol{\beta}^{\top}}\right]\otimes\tilde{\mathbf{X}}_i\bar{H}^{-1} \\
&= \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)\bar{H}^{-1}G_i\left\{\left[\frac{\partial\tilde{s}_i\left(\boldsymbol{\beta}^0\right)}{\partial\boldsymbol{\beta}^{\top}} - \frac{\partial s_i\left(\boldsymbol{\beta}^0\right)}{\partial\boldsymbol{\beta}^{\top}}\right]\otimes\tilde{\mathbf{X}}_i\right\}\bar{H}^{-1} \\
&\quad + \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)\bar{H}^{-1}\left(\tilde{G}_i - G_i\right)\left\{\frac{\partial s_i\left(\boldsymbol{\beta}^0\right)}{\partial\boldsymbol{\beta}^{\top}}\otimes\tilde{\mathbf{X}}_i\right\}\bar{H}^{-1} \\
&\quad + \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)\bar{H}^{-1}\left(\tilde{G}_i - G_i\right)\left\{\left[\frac{\partial\tilde{s}_i\left(\boldsymbol{\beta}^0\right)}{\partial\boldsymbol{\beta}^{\top}} - \frac{\partial s_i\left(\boldsymbol{\beta}^0\right)}{\partial\boldsymbol{\beta}^{\top}}\right]\otimes\tilde{\mathbf{X}}_i\right\}\bar{H}^{-1} \\
&\equiv \mathcal{R}_{2n,1} + \mathcal{R}_{2n,2} + \mathcal{R}_{2n,3}, \text{ say.}
\end{aligned}$$

We prove the lemma by showing that $\mathcal{R}_{2n,s} = o_P(1)$ for $s = 1, 2, 3$. We will only show that $\mathcal{R}_{2n,1} = o_P(1)$ as the other two cases can be proved analogously. Recall $c_{i\varphi} = \varphi'\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^2 - \varphi''\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)$ and $\varphi_i \equiv \varphi'\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)/\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)$. Noting that

$$\frac{\partial \tilde{s}_i\left(\boldsymbol{\beta}^0\right)}{\partial \boldsymbol{\beta}^\top} = \begin{pmatrix} \frac{\tilde{\psi}_i'(\bar{\varepsilon}_i)}{\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} & \frac{\varphi_i\left[\tilde{\psi}_i'(\bar{\varepsilon}_i)\bar{\varepsilon}_i + \tilde{\psi}_i(\bar{\varepsilon}_i)\right]}{2\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{1/2}} \\ \frac{\varphi_i\left[\tilde{\psi}_i'(\bar{\varepsilon}_i)\bar{\varepsilon}_i + \tilde{\psi}_i(\bar{\varepsilon}_i)\right]}{2\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{1/2}} & \frac{2c_{i\varphi}\left[\tilde{\psi}_i(\bar{\varepsilon}_i)\bar{\varepsilon}_i + 1\right] + \varphi'\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^2\bar{\varepsilon}_i\left[\tilde{\psi}_i'(\bar{\varepsilon}_i)\bar{\varepsilon}_i + \tilde{\psi}_i(\bar{\varepsilon}_i)\right]}{4\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^2} \end{pmatrix} \otimes \tilde{\mathbf{X}}_i^\top,$$

and $\partial s_i\left(\boldsymbol{\beta}^0\right)/\partial \boldsymbol{\beta}^\top$ has similar expression with $\psi_i(\bar{\varepsilon}_i)$ in the place of $\tilde{\psi}_i(\bar{\varepsilon}_i)$, we have

$$\begin{aligned} \mathcal{R}_{2n,1} &= \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) G_i \\ &\quad \times \begin{pmatrix} \frac{\tilde{\psi}_i'(\bar{\varepsilon}_i) - \psi'(\bar{\varepsilon}_i)}{\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)} & \frac{\varphi_i\left\{\left[\tilde{\psi}_i'(\bar{\varepsilon}_i) - \psi'(\bar{\varepsilon}_i)\right]\bar{\varepsilon}_i + \left[\tilde{\psi}_i(\bar{\varepsilon}_i) - \psi(\bar{\varepsilon}_i)\right]\right\}}{2\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{1/2}} \\ \frac{\varphi_i\left\{\left[\tilde{\psi}_i'(\bar{\varepsilon}_i) - \psi'(\bar{\varepsilon}_i)\right]\bar{\varepsilon}_i + \left[\tilde{\psi}_i(\bar{\varepsilon}_i) - \psi(\bar{\varepsilon}_i)\right]\right\}}{2\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^{1/2}} & \frac{2c_{i\varphi}\left[\tilde{\psi}_i(\bar{\varepsilon}_i) - \psi(\bar{\varepsilon}_i)\right]\bar{\varepsilon}_i}{4\varphi\left(P_i\left(\boldsymbol{\beta}_2^0\right)\right)^2} + \frac{\tilde{d}_i}{4} \end{pmatrix} \otimes \left(\mathbf{Z}_i\mathbf{Z}_i^\top\right) \\ &\equiv \begin{bmatrix} \mathcal{R}_{2n,1,11} & \mathcal{R}_{2n,1,12} \\ \mathcal{R}_{2n,1,12}^\top & \mathcal{R}_{2n,1,22} \end{bmatrix}, \text{ say,} \end{aligned}$$

where $\tilde{d}_i \equiv \varphi_i^2\bar{\varepsilon}_i[\tilde{\psi}_i'(\bar{\varepsilon}_i) - \psi'(\bar{\varepsilon}_i)]\bar{\varepsilon}_i + [\tilde{\psi}_i(\bar{\varepsilon}_i) - \psi(\bar{\varepsilon}_i)]$. As in the analysis of $\mathcal{R}_{1n,1}$, using Lemma A.2, we can readily demonstrate that $\mathcal{R}_{2n,1,11} = o_p(1)$, $\mathcal{R}_{2n,1,12} = o_p(1)$ and $\mathcal{R}_{2n,1,22} = o_p(1)$. It follows that $\mathcal{R}_{2n,1} = o_p(1)$. Similarly, we can show that $\mathcal{R}_{2n,s} = o_P(1)$ for $s = 2, 3$. This completes the proof of the lemma. ■

D Derivative Matrices in the Proof of Proposition 2.1

In this appendix, we give explicit expressions for the elements of some derivative matrices of the log-likelihood function defined in the proof of Proposition 2.1. The elements of the Hessian matrix are

$$\begin{aligned} q_{11}(y;\beta_1,\beta_2) &= \frac{\partial^2 \log f(\varepsilon(\beta))}{\partial\varepsilon^2}\frac{1}{\varphi(\beta_2)}, \\ q_{12}(y;\beta_1,\beta_2) &= \left\{\frac{\partial^2 \log f(\varepsilon(\beta))}{\partial\varepsilon^2}\varepsilon(\beta) + \frac{\partial \log f(\varepsilon(\beta))}{\partial\varepsilon}\right\}\frac{\varphi'(\beta_2)}{2\varphi(\beta_2)^{3/2}}, \\ q_{22}(y;\beta_1,\beta_2) &= \frac{-\varphi''(\beta_2)}{2\varphi(\beta_2)}\left[\frac{\partial \log f(\varepsilon(\beta))}{\partial\varepsilon}\varepsilon(\beta) + 1\right] \\ &\quad + \frac{\varphi'(\beta_2)^2}{4\varphi(\beta_2)^2}\left\{\frac{\partial^2 \log f(\varepsilon(\beta))}{\partial\varepsilon^2}\varepsilon(\beta)^2 + 3\frac{\partial \log f(\varepsilon(\beta))}{\partial\varepsilon}\varepsilon(\beta) + 2\right\}, \end{aligned}$$

and $q_{21}(y;\beta_1,\beta_2) = q_{12}(y;\beta_1,\beta_2)$ by Young's theorem, where, e.g., $\frac{\partial^2 \log f(\varepsilon)}{\partial\varepsilon^2} = \frac{f''(\varepsilon)f(\varepsilon) - f'(\varepsilon)^2}{f^2(\varepsilon)}$ and $\frac{\partial^2 \log f(\varepsilon(\beta))}{\partial\varepsilon^2} \equiv \left.\frac{\partial^2 \log f(\varepsilon)}{\partial\varepsilon^2}\right|_{\varepsilon=\varepsilon(\beta)}$. Note that when we restrict our attention to the case $\varphi(u) = u$ or $\exp(u)$, the above formulae can be greatly simplified.

In addition, in the proof of Proposition 2.1, we also need that $q_{rst}(y;\beta_1,\beta_2) \equiv \frac{\partial^3}{\partial\beta_r\partial\beta_s\partial\beta_t}\log(f(y;\beta_1,\beta_2))$, $r,s,t=1,2$, should be well behaved. Using the expressions

$$\frac{\partial\varepsilon(\beta)}{\partial\beta} = \begin{pmatrix} \frac{\partial\varepsilon(\beta)}{\partial\beta_1} \\ \frac{\partial\varepsilon(\beta)}{\partial\beta_2} \end{pmatrix} = \begin{pmatrix} -\frac{1}{\varphi(\beta_2)^{1/2}} \\ -\frac{\varphi'(\beta_2)}{2\varphi(\beta_2)}\varepsilon(\beta) \end{pmatrix} \text{ and } \frac{\partial^2\log f(\varepsilon)}{\partial\varepsilon^2} = \frac{f''(\varepsilon)f(\varepsilon)-f'(\varepsilon)^2}{f^2(\varepsilon)}$$

and by straightforward calculations, we have

$$\begin{aligned}
q_{111}(y;\beta_1,\beta_2) &= \frac{\partial^3\log f(\varepsilon(\beta))}{\partial\varepsilon^3}\frac{1}{\varphi(\beta_2)}, \\
q_{112}(y;\beta_1,\beta_2) &= \frac{\partial^3\log f(\varepsilon(\beta))}{\partial\varepsilon^3}\frac{\partial\varepsilon(\beta)}{\partial\beta_2}\frac{1}{\varphi(\beta_2)} - \frac{\partial^2\log f(\varepsilon(\beta))}{\partial\varepsilon^2}\frac{\varphi'(\beta_2)}{\varphi(\beta_2)^2}, \\
q_{121}(y;\beta_1,\beta_2) &= \left\{\frac{\partial^3\log f(\varepsilon(\beta))}{\partial\varepsilon^3}\frac{\partial\varepsilon(\beta)}{\partial\beta_1}\varepsilon(\beta) + 2\frac{\partial^2\log f(\varepsilon(\beta))}{\partial\varepsilon^2}\frac{\partial\varepsilon(\beta)}{\partial\beta_1}\right\}\frac{\varphi'(\beta_2)}{2\varphi(\beta_2)^{3/2}} = q_{112}(y;\beta_1,\beta_2), \\
q_{122}(y;\beta_1,\beta_2) &= \left\{\frac{\partial^3\log f(\varepsilon(\beta))}{\partial\varepsilon^3}\frac{\partial\varepsilon(\beta)}{\partial\beta_2}\varepsilon(\beta) + 2\frac{\partial^2\log f(\varepsilon(\beta))}{\partial\varepsilon^2}\frac{\partial\varepsilon(\beta)}{\partial\beta_2}\right\}\frac{\varphi'(\beta_2)}{2\varphi(\beta_2)^{3/2}}, \\
&\quad + \left\{\frac{\partial^2\log f(\varepsilon(\beta))}{\partial\varepsilon^2}\varepsilon(\beta) + \frac{\partial\log f(\varepsilon(\beta))}{\partial\varepsilon}\right\}\frac{\varphi''(\beta_2)\varphi(\beta_2)^{3/2} - \frac{3}{2}\varphi'(\beta_2)^2\varphi(\beta_2)^{1/2}}{2\varphi(\beta_2)^3}, \\
q_{221}(y;\beta_1,\beta_2) &= \frac{-\varphi''(\beta_2)}{2\varphi(\beta_2)}\left[\frac{\partial^2\log f(\varepsilon(\beta))}{\partial\varepsilon^2}\varepsilon(\beta) + \frac{\partial\log f(\varepsilon(\beta))}{\partial\varepsilon}\right]\frac{\partial\varepsilon(\beta)}{\partial\beta_1} + \frac{\varphi'(\beta_2)^2}{4\varphi(\beta_2)^2}\kappa(\beta)\frac{\partial\varepsilon(\beta)}{\partial\beta_1} \\
&= q_{122}(y;\beta_1,\beta_2) \\
q_{222}(y;\beta_1,\beta_2) &= \frac{-\varphi''(\beta_2)}{2\varphi(\beta_2)}\left[\frac{\partial^2\log f(\varepsilon(\beta))}{\partial\varepsilon^2}\varepsilon(\beta) + \frac{\partial\log f(\varepsilon(\beta))}{\partial\varepsilon}\right]\frac{\partial\varepsilon(\beta)}{\partial\beta_2} \\
&\quad - \frac{\varphi'''(\beta_2)\varphi(\beta_2) - \varphi''(\beta_2)\varphi'(\beta_2)}{2\varphi(\beta_2)^2}\left[\frac{\partial\log f(\varepsilon(\beta))}{\partial\varepsilon}\varepsilon(\beta) + 1\right] \\
&\quad + \frac{\varphi'(\beta_2)^2}{4\varphi(\beta_2)^2}\kappa(\beta)\frac{\partial\varepsilon(\beta)}{\partial\beta_2} \\
&\quad + \frac{\varphi'(\beta_2)\varphi''(\beta_2) - \varphi'(\beta_2)^3\varphi(\beta_2)}{2\varphi(\beta_2)^4}\left\{\frac{\partial^2\log f(\varepsilon(\beta))}{\partial\varepsilon^2}\varepsilon(\beta)^2 + 3\frac{\partial\log f(\varepsilon(\beta))}{\partial\varepsilon}\varepsilon(\beta) + 2\right\},
\end{aligned}$$

$q_{211} = q_{121} = q_{112}$, and $q_{212} = q_{122} = q_{221}$ by Young's Theorem, where $\kappa(\beta) \equiv \frac{\partial^3\log f(\varepsilon(\beta))}{\partial\varepsilon^3}\varepsilon(\beta)^2$ $+2\frac{\partial^2\log f(\varepsilon(\beta))}{\partial\varepsilon^2}\varepsilon(\beta) + 3\frac{\partial^2\log f(\varepsilon(\beta))}{\partial\varepsilon^2}\varepsilon(\beta) + 3\frac{\partial\log f(\varepsilon(\beta))}{\partial\varepsilon}$. Note that under our assumptions (X_i has compact support, the parameter space is compact, $\sigma^2(x)$ is bounded away from 0) the terms associated with $\varphi(\cdot)$ or its derivatives are all well behaved when $\varphi(\cdot)$ is evaluated in the neighborhood of $\beta_2^0(x)$.

REFERENCES

Bickel, P. J. (1975) One-step Huber estimates in linear models. *Journal of the American Statistical Association* 70, 428-433.

Gozalo, P. & O. Linton (2000) Local non-linear least squares: using parametric information in nonparametric regression. *Journal of Econometrics* 99, 63-106.

Hansen, B. E. (2008) Uniform convergence rates for kernel regression. *Econometric Theory*, 24, 726-748.

Lee, A. J. (1990) *U-statistics: Theory and Practice.* Marcel Dekker, New York and Basel.

Linton, O., and Z. Xiao (2007) A nonparametric regression estimator that adapts to error distribution of unknown form. *Econometric Theory* 23, 371-413.

Masry, E. (1996a) Multivariate regression estimation: local polynomial fitting for time series. *Stochastic Processes and Their Applications* 65, 81-101.

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

Newey, W. K. (1991) Uniform convergence in probability and stochastic equicontinuity. *Econometrica* 59, 1161-1167.

Nolan, D. and D. Pollard (1987) U-processes: rates of convergence. *Annals of Statistics* 15, 780-799.

Pakes, A. and D. Pollard (1989) Simulation and the asymptotics of optimization estimators. *Econometrica* 57, 1027-1057.

Pollard, D. (1984) *Convergence of Stochastic Processes*. Springer, New York.

Robinson, P. M. (1988) Root-N-consistent semiparametric regression. *Econometrica* 56, 931-954.

Rudin, W. (1976) *Principles of Mathematical Analysis*. 3rd ed., McGraw-Hill, London.