SUPPLEMENTAL MATERIAL FOR "MULTIVARIATE VARYING COEFFICIENT MODEL AND ITS APPLICATION TO NEUROIMAGING DATA"

BY HONGTU ZHU, RUNZE LI AND LINGLONG KONG

University of North Carolina and The Pennsylvania State University

1. Proofs of Theorems.

We introduce some notation. We define

$$
\begin{aligned}
T_{B,j}(h,s) &= \sum_{i=1}^{n}\sum_{m=1}^{M} K_h(s_m - s)[\mathbf{x}_i \otimes \mathbf{z}_h(s_m - s)]\mathbf{x}_i^T B_j(s_m),\\
T_{\eta,j}(h,s) &= \sum_{i=1}^{n}\sum_{m=1}^{M} K_h(s_m - s)[\mathbf{x}_i \otimes \mathbf{z}_h(s_m - s)]\eta_{ij}(s_m),\\
T_{\epsilon,j}(h,s) &= \sum_{i=1}^{n}\sum_{m=1}^{M} K_h(s_m - s)[\mathbf{x}_i \otimes \mathbf{z}_h(s_m - s)]\epsilon_{ij}(s_m),\\
r_u(K;s,h) &= \frac{u_2(K;s,h)^2 - u_1(K;s,h)u_3(K;s,h)}{u_0(K;s,h)u_2(K;s,h) - u_1(K;s,h)^2},
\end{aligned} \tag{1}
$$

where $u_r(K;s,h) = \int_0^1 h^{-r}(u-s)^r K_h(u-s)du$ for $r \geq 0$. Throughout the proofs, C_ks stand for a generic constant, and it may vary from line to line.

Without special saying, we consider the random grid points throughout the paper, but we also discuss several key equations for the fixed grid points. We discuss the key steps in the proof of Theorem 1 and then present on the proof of these key steps as lemmas.

Proof of Theorem 1. Define

$$
\begin{aligned}
&\mathbf{U}_2(K;s,\mathbf{H}) = \mathrm{diag}(r_u(K;s,h_{11}),\ldots,r_u(K;s,h_{1J})),\\
&X_n(s) = \sqrt{n}\{\hat{\mathbf{B}}(s) - E[\hat{\mathbf{B}}(s)|\mathcal{S}]\} = \sqrt{n}\{\mathrm{vec}(\hat{\mathbf{B}}(s) - \mathbf{B}(s) - 0.5\ddot{\mathbf{B}}(s)\mathbf{U}_2(K;s,\mathbf{H})\mathbf{H}^2) + o_p(||\mathbf{H}^2||_2)\},\\
&X_{n,j}(s) = \sqrt{n}\{\hat{B}_j(s) - E[\hat{B}_j(s)|\mathcal{S}]\} = \sqrt{n}\{\hat{B}_j(s) - B_j(s) - 0.5r_u(K;s,h_{1j})h_{1j}^2\ddot{B}_j(s) + o_p(h_{1j}^2)\}.
\end{aligned}
$$

According to the definition of $\mathrm{vec}(\hat{A}_j(s))$, it is easy to see that

$$
\mathrm{vec}(\hat{A}_j(s)) = \Sigma(s,h_{1j})^{-1}[T_{B,j}(h_{1j},s) + T_{\epsilon,j}(h_{1j},s) + T_{\eta,j}(h_{1j},s)], \tag{2}
$$

$$
X_{n,j}(s) = \sqrt{n}[\mathbf{I}_p \otimes (1,0)]\Sigma(s,h_{1j})^{-1}[T_{\epsilon,j}(h_{1j},s) + T_{\eta,j}(h_{1j},s)]. \tag{3}
$$

Since Theorem 1 (ii) is a direct consequence of Theorem 1 (i) and Lemma 4, we focus on Theorem 1 (i). The proof of Theorem 1 (i) consists of two parts.

- Part 1 is to show that $\sqrt{n}\Sigma(s,h_{1j})^{-1}T_{\epsilon,j}(h_{1j},s) = o_p(1)$ holds uniformly for all $s \in [0,1]$ and $j = 1,\ldots,J$.
- Part 2 is to show that $\sqrt{n}\Sigma(s,h_{1j})^{-1}T_{\eta,j}(h_{1j},s)$ converges weakly to a Gaussian process $G(\cdot)$ with mean zero and covariance matrix $\Sigma_{\eta,jj}(s,s')\Omega_X^{-1}$ for each j.

In part 1, we show that

$$
\sqrt{n}[\mathbf{I}_p \otimes (1,0)]\Sigma(s,h_{1j})^{-1}T_{\epsilon,j}(h_{1j},s) = O_p(|\log(h_{1j})|^{1/2}(Mh_{1j})^{-1/2}) = o_p(1). \tag{4}
$$

It follows from Lemma 1 that

$$n^{-1/2}\sum_{i=1}^{n}\mathbf{x}_i\otimes\{M^{-1}\sum_{m=1}^{M}K_{h_{1j}}(s_m-s)\mathbf{z}_{h_{1j}}(s_m-s)\epsilon_{i,j}(s_m)\}=O_p(|\log(h_{1j})|^{1/2}(Mh_{1j})^{-1/2})=o_p(1)$$

hold uniformly for all $s\in[0,1]$. It follows from Lemma 2 that

$$\begin{aligned}(5)\ (nM)^{-1}\Sigma(s,h_{1j}) &= [n^{-1}\sum_{i=1}^{n}\mathbf{x}_i^{\otimes 2}\otimes M^{-1}\sum_{m=1}^{M}K_{h_{1,j}}(s_m-s)\mathbf{z}_{h_{1,j}}(s_m-s)^{\otimes 2}]\\ &= \Omega_X\otimes\Omega_1(h_{1j},s)+O_p(M^{-1/2}h_{1j}^{-1}+n^{-1/2})=\Omega_X\otimes\Omega_1(h_{1j},s)+o_p(1)\end{aligned}$$

hold uniformly for all $s\in[0,1]$. Based on these results, we can finish the proof of (4).

In part 2, we show the weak convergence of $\sqrt{n}[\mathbf{I}_p\otimes(1,0)]\Sigma(s,h_{1j})^{-1}T_{\eta,j}(h_{1j},s)$ for $j=1,\dots,J$. The part 2 consists of two steps. In Step 1, it follows from the standard central limit theorem that for each $s\in[0,1]$,

$$\sqrt{n}[\mathbf{I}_p\otimes(1,0)]\Sigma(s,h_{1j})^{-1}T_{\eta,j}(h_{1j},s)\rightarrow^{L}N(\mathbf{0},\Sigma_{\eta,jj}(s,s)\Omega_X^{-1}),\tag{6}$$

where $\rightarrow^L$ denotes convergence in distribution.

Step 2 is to the asymptotic tightness of $\sqrt{n}[\mathbf{I}_p\otimes(1,0)]\Sigma(s,h_{1j})^{-1}T_{\eta,j}(h_{1j},s)$. We define

$$H_h(s_m-s)=K_h(s_m-s)\mathbf{z}_h(s_m-s),$$

$$\Delta_j(s;\boldsymbol{\eta}_i,h_{1j})=M^{-1}\sum_{m=1}^{M}H_{h_{1j}}(s_m-s)\eta_{ij}(s_m)-\int_0^1 H_{h_{1j}}(u-s)\eta_{ij}(u)\pi(u)du.\tag{7}$$

By using (5) and (7), we can show that $\sqrt{n}\Sigma(s,h_{1j})^{-1}T_{\eta,j}(h_{1j},s)[1+o_p(1)]$ can be approximated by three terms as follows:

$$\begin{aligned}&(8)\quad \sqrt{n}\Sigma(s,h_{1j})^{-1}T_{\eta,j}(h_{1j},s)[1+o_p(1)]=\text{(I)}+\text{(II)}+\text{(III)}\\ =&\quad n^{-1/2}\sum_{i=1}^{n}\Omega_X^{-1}\mathbf{x}_i\otimes\Omega_1(h_{1j},s)^{-1}\Delta_j(s;\boldsymbol{\eta}_i,h_{1j})\\ +&\quad n^{-1/2}\sum_{i=1}^{n}\Omega_X^{-1}\mathbf{x}_i\otimes\Omega_1(h_{1j},s)^{-1}\eta_{ij}(s)\int_{\max(-sh_{1j}^{-1},-1)}^{\min((1-s)h_{1j}^{-1},1)}K(u)(1,u)^T\pi(s+h_{1j}u)du\\ +&\quad n^{-1/2}\sum_{i=1}^{n}\Omega_X^{-1}\mathbf{x}_i\otimes\Omega_1(h_{1j},s)^{-1}\int_{\max(-sh_{1j}^{-1},-1)}^{\min((1-s)h_{1j}^{-1},1)}K(u)\begin{pmatrix}1\\ u\end{pmatrix}[\eta_{ij}(s+h_{1j}u)-\eta_{ij}(s)]\pi(s+h_{1j}u)du\end{aligned}$$

We investigate the three terms on the right hand side of (8) as follows. It follows from Lemma 3 that the first term on the right hand side of (8) converges to zero uniformly.

We prove the asymptotic tightness of (II) as follows. Define

$$\hat{X}_{n,j}(s)=n^{-1/2}\sum_{i=1}^{n}\Omega_X^{-1}\mathbf{x}_i\otimes(1,0)\Omega_1(h_{1j},s)^{-1}\eta_{ij}(s)\int_{\max(-sh_{1j}^{-1},-1)}^{\min((1-s)h_{1j}^{-1},1)}K(u)(1,u)^T\pi(s+h_{1j}u)du.$$

Thus, we only need to prove the asymptotic tightness of $\hat{X}_{n,j}(s)$. The asymptotic tightness of $\hat{X}_{n,j}(s)$ can be proved using the empirical process techniques [13]. It follows from (40) that

$$\begin{aligned}&(1,0)\Omega_1(h_{1j},s)^{-1}\int_{\max(-sh_{1j}^{-1},-1)}^{\min((1-s)h_{1j}^{-1},1)}K(u)(1,u)^T\pi(s+h_{1j}u)du\\ =&\ \frac{u_2(K;s,h_{1j})u_0(K;s,h_{1j})-u_1(K;s,h_{1j})^2+o(h_{1j})}{u_2(K;s,h_{1j})u_0(K;s,h_{1j})-u_1(K;s,h_{1j})^2+o(h_{1j})}=1+o(h_{1j}).\end{aligned}$$

Thus, $\hat{X}_{n,j}(s)$ can be simplified as

$$\hat{X}_{n,j}(s) = [1 + o(h_{1j})]n^{-1/2}\sum_{i=1}^{n}\eta_{ij}(s)\Omega_X^{-1}\mathbf{x}_i.$$

We consider a function class $\mathcal{E}_\eta = \{f(s; \mathbf{x}, \eta_{\cdot,j}) = \Omega_X^{-1}\mathbf{x}\eta_{\cdot,j}(s) : s \in [0, 1]\}$. Due to Assumption (C2), $\mathcal{E}_\eta$ is a $P-$Donsker class.

Finally, we consider the third term (III) on the right hand side of (8). It is easy to see that (III) can be written as

$$\Omega_X^{-1} \otimes \Omega_1(h_{1j}, s)^{-1}\int_{\max(-sh_{1j}^{-1},-1)}^{\min((1-s)h_{1j}^{-1},1)} K(u)[n^{-1/2}\sum_{i=1}^{n}\mathbf{x}_i\{\eta_{ij}(s + h_{1j}u) - \eta_{ij}(s)\}] \otimes \begin{pmatrix} 1 \\ u \end{pmatrix}\pi(s + h_{1j}u)du$$

Using the same argument of proving the second term (II), we can show the asymptotic tightness of $n^{-1/2}\sum_{i=1}^{n}\mathbf{x}_i\eta_{ij}(s)$. Therefore, for any $h_{1j} \to 0$,

$$\sup_{s\in[0,1],|u|\le 1}|n^{-1/2}\sum_{i=1}^{n}\mathbf{x}_i\{\eta_{ij}(s + h_{1j}u) - \eta_{ij}(s)\}| = o_p(1). \tag{9}$$

It follows from Assumptions (C5) and (C7) and (9) that (III) converges to zero uniformly. Therefore, we can finish the proof of Theorem 1.

Lemma 1. Under Assumptions (C1), (C3)-(C5), and (C7), we have that for each j,

$$\sup_{s\in[0,1]} n^{-1/2}h_{1j}|T_{\epsilon,j}(h_{1j}, s)| = O_p(\sqrt{Mh_{1j}|\log h_{1j}|}) = o_p(Mh_{1j}). \tag{10}$$

Proof. Let $F_n(s_m) = n^{-1/2}\sum_{i=1}^{n}\mathbf{x}_i\epsilon_{ij}(s_m)$. Then it follows by the definition of $T_{\epsilon,j}(h_{1j}, s)$ that

$$n^{-1/2}h_{1j}T_{\epsilon,j}(h_{1j}, s) = h_{1j}\sum_{m=1}^{M}K_{h_{1j}}(s_m - s)F_n(s_m) \otimes \mathbf{z}_{h_{1j}}(s_m - s).$$

Let $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and $\tilde{T}_{\epsilon,j}(h_{1j}, s) = \{T'_{\epsilon,j}(h_{1j}, s) - E[T'_{\epsilon,j}(h_{1j}, s)|\mathbf{X}, \mathcal{S}]\}$, where

$$T'_{\epsilon,j}(h_{1j}, s) = \sqrt{n}\sum_{m=1}^{M}K_{h_{1j}}(s_m - s)F_n(s_m)\mathbf{1}(||F_n(s_m)||_2 \le \gamma_M) \otimes \mathbf{z}_{h_{1j}}(s_m - s),$$

in which γ_M is a positive number to be specified below. The proof of Lemma 1 consists of three steps. In Step 1, we show that

$$\sup_{s\in[0,1]} n^{-1/2}h_{1j}||T_{\epsilon,j}(h_{1j}, s) - \tilde{T}_{\epsilon,j}(h_{1j}, s)||_2 = o_p(\sqrt{Mh_{1j}|\log h_{1j}|}). \tag{11}$$

In Step 2, we define an equally-spaced grid $\tilde{\mathbf{S}} = \{\tilde{s}_l = lh_{1j} : l = 0, \ldots, 1h_{1j}^{-1}\}$ and then show that

$$\max_l h_{1j}||n^{-1/2}\tilde{T}_{\epsilon,j}(h_{1j}, \tilde{s}_l)||_2 = O_p(\sqrt{Mh_{1j}|\log h_{1j}|}). \tag{12}$$

In Step 3, we show that

$$\max_l \sup_{s\in[\tilde{s}_{l-1},\tilde{s}_l]} n^{-1/2}h_{1j}||\tilde{T}_{\epsilon,j}(h_{1j}, \tilde{s}_{l-1}) - \tilde{T}_{\epsilon,j}(h_{1j}, s)||_2 = O_p(\sqrt{Mh_{1j}|\log h_{1j}|}). \tag{13}$$

It is easy to see that the proof of Lemma 1 is completed by combing (11)-(13).

We first show (11). It follows from Assumption (C5) and $s_m, s \in [0,1]$ that

$$\begin{aligned} & n^{-1/2} h_{1j} ||T_{\epsilon,j}(h_{1j}, s) - \tilde{T}_{\epsilon,j}(h_{1j}, s)||_2 \\ \leq \;& C_1 \sum_{m=1}^{M} ||F_n(s_m)||_2 \mathbf{1}(||F_n(s_m)||_2 \geq \gamma_M) + C_1 \sum_{m=1}^{M} E[||F_n(s_m)||_2 \mathbf{1}(||F_n(s_m)||_2 \geq \gamma_M)|\mathbf{X}, \mathcal{S}], \end{aligned}$$

for a positive constant C_1. Let $\gamma_M = \delta(M/|\log h_{1j}|)^{1/q_1}$, where δ is a positive scalar. It follows from Assumption (C7) that $(|\log h_{1j}|/M)^{1-2/q_1} \leq h_{1j} \to 0$ and $1 - 2/q_1 > 0$, which yields that $|\log h_{1j}|/M \to 0$ and $\gamma_M \to \infty$. As $\gamma_M \to \infty$, we can show that

$$\max_m E[||F_n(s_m)||_2^{q_1} \mathbf{1}(||F_n(s_m)||_2 \geq \gamma_M)|\mathbf{X}, \mathcal{S}] = o(1). \tag{14}$$

For notational simplicity, we only consider the case $p = 1$, (i.e., $\mathbf{x}_i$ is scalar). For any $c > 0$ with $q_1 + c < 4$, equation (14) is followed from Assumptions (C1), (C3) and (C7) and the partial sum moment inequality [2] as follows:

$$\begin{aligned} & \max_m E[||F_n(s_m)||_2^{q_1} \mathbf{1}(||F_n(s_m)||_2 \geq \gamma_M)|\mathbf{X}, \mathcal{S}] \leq \max_m E[|n^{-1/2} \sum_{i=1}^{n} \mathbf{x}_i \epsilon_{ik}(s_m)|^{q_1+c}|\mathbf{X}, \mathcal{S}]/\gamma_M^c \\ & \leq \max_m n^{-(q_1+c)/2} C(q_1) n^{(q_1+c)/2-1} \sum_{i=1}^{n} |\mathbf{x}_i|^{q_1+c} E[|\epsilon_{ik}(s_m)|^{q_1+c}|]/\gamma_M^c = o(1), \end{aligned}$$

where $C(q_1)$ is a universal constant independent of n. It follows from Assumptions (C1) and (C7) and (14) that

$$\begin{aligned} & \sum_{m=1}^{M} E[||F_n(s_m)||_2 \mathbf{1}(||F_n(s_m)||_2 \geq \gamma_M)|\mathbf{X}, \mathcal{S}] \\ \leq \;& \sum_{m=1}^{M} E[||F_n(s_m)||_2^{q_1} \mathbf{1}(||F_n(s_m)||_2 \geq \gamma_M)|\mathbf{X}, \mathcal{S}]/\gamma_M^{q_1-1} \\ \leq \;& o(1) M^{1/q_1} |\log h_{1j}|^{1-1/q_1} \leq o(\sqrt{M h_{1j} |\log h_{1j}|}). \end{aligned} \tag{15}$$

Furthermore, we have

$$\begin{aligned} & \mathrm{Var}(\sum_{m=1}^{M} ||F_n(s_m)||_2 \mathbf{1}(||F_n(s_m)||_2 \geq \gamma_M)|\mathbf{X}, \mathcal{S}) \\ & = \sum_{m=1}^{M} \mathrm{Var}(||F_n(s_m)||_2 \mathbf{1}(||F_n(s_m)||_2 \geq \gamma_M)|\mathbf{X}, \mathcal{S}) \\ & \leq E\{[\sum_{m=1}^{M} ||F_n(s_m)||_2^2 \mathbf{1}(||F_n(s_m)||_2 \geq \gamma_M)]|\mathbf{X}, \mathcal{S}\} \\ & \leq E[\sum_{m=1}^{M} ||F_n(s_m)||_2^{q_1} \mathbf{1}(||F_n(s_m)||_2 \geq \gamma_M)|\mathbf{X}, \mathcal{S}]/\gamma_M^{q_1-2} \\ & \leq \max_m E[||F_n(s_m)||_2^{q_1} \mathbf{1}(||F_n(s_m)||_2 \geq \gamma_M)|\mathbf{X}, \mathcal{S}] M/\gamma_M^{q_1-2} \\ & \leq \max_m E[||F_n(s_m)||_2^{q_1} \mathbf{1}(||F_n(s_m)||_2 \geq \gamma_M)|\mathbf{X}, \mathcal{S}] M h_{1j} = o(M h_{1j} |\log h_{1j}|). \end{aligned} \tag{16}$$

Therefore, combing equations (15) and (16), we have

$$\sum_{m=1}^{M} ||F_n(s_m)||_2 \mathbf{1}(||F_n(s_m)||_2 \geq \gamma_M) = o_p(\sqrt{Mh_{1j}|\log h_{1j}|}),$$

which yields (11).

We next prove (12). It follows from Assumption (C5) that

$$\begin{aligned} & h_{1j}||K_{h_{1j}}(s_m - s)\{F_n(s_m)\mathbf{1}(||F_n(s_m)||_2 \leq \gamma_M) - \\ & \quad E[F_n(s_m)\mathbf{1}(||F_n(s_m)||_2 \leq \gamma_M)|\mathbf{X}, \mathcal{S}]\} \otimes \mathbf{z}_{h_{1j}}(s_m - s)||_2 \\ \leq \quad & C_2(M/|\log h_{1j}|)^{1/q_1} \leq C_2\sqrt{Mh_{1j}/|\log h_{1j}|}, \end{aligned}$$

where $C_2 = 4\delta \sup_{t\in[-1,1]} |K(t)|$. Furthermore, let $E_{\mathcal{S}}$ denote the expectation to s_m. We have

$$\begin{aligned} & \mathrm{Var}(\sum_{m=1}^{M} h_{1j}K_{h_{1j}}(s_m - s)F_n(s_m)\mathbf{1}(||F_n(s_m)||_2 \leq \gamma_M) \otimes \mathbf{z}_{h_{1j}}(s_m - s)|\mathbf{X}) \\ = & \sum_{m=1}^{M} \mathrm{Var}(h_{1j}K_{h_{1j}}(s_m - s)F_n(s_m)\mathbf{1}(||F_n(s_m)||_2 \leq \gamma_M) \otimes \mathbf{z}_{h_{1j}}(s_m - s)|\mathbf{X}) \\ \leq & \sum_{m=1}^{M} E_{\mathcal{S}}\{h_{1j}^2 K_{h_{1j}}(s_m - s)^2 E[F_n(s_m)^{\otimes 2}\mathbf{1}(||F_n(s_m)||_2 \leq \gamma_M) \otimes \mathbf{z}_{h_{1j}}(s_m - s)^{\otimes 2}|\mathbf{X}, \mathcal{S}]\} \\ \leq & \sum_{m=1}^{M} h_{1j}^2 E_{\mathcal{S}}[K_{h_{1j}}(s_m - s)^2 n^{-1}\sum_{i=1}^{n} \mathbf{x}_i^{\otimes 2}\Sigma_\epsilon(s_m, s_m) \otimes \mathbf{z}_{h_{1j}}(s_m - s)^{\otimes 2}] = O_p(Mh_{1j}). \end{aligned}$$

Therefore, by applying Bernstein's inequality to each component of $h_{1j}n^{-1/2}\tilde{T}_{\epsilon,j}(h_{1j}, \tilde{s}_l)$ [13], we can prove (12). For instance, let $\mathbf{e}_1$ be a $\dim(\tilde{T}_{\epsilon,j}(h_{1j}, \tilde{s}_l)) \times 1$ vector with the first element 1 and zero otherwise; we have

$$\begin{aligned} & P(\max_l |\mathbf{e}_1 h_{1j}n^{-1/2}\tilde{T}_{\epsilon,j}(h_{1j}, \tilde{s}_l)| > t|\mathbf{X}) \\ \leq \quad & C_3(1h_{1j}^{-1} + 1)E[\exp(-\frac{1}{2}\frac{t^2}{v(\mathbf{X}) + tC_3\sqrt{Mh_{1j}/|\log h_{1j}|}/3})|\mathbf{X}], \end{aligned} \tag{17}$$

where $C_3 = O(1)$, t is a positive scalar, and $v(\mathbf{X}) \geq \mathrm{Var}(\mathbf{e}_1 h_{1j}n^{-1/2}\tilde{T}_{\epsilon,j}(h_{1j}, \tilde{s}_l)|\mathbf{X})$ for all l. By setting $t = C_4\sqrt{Mh_{1j}|\log h_{1j}|}$ for large $C_4 > 0$, we can show that the right hand side of (17) is of order $h_{1j}^{C_5}$, where C_5 is a positive scalar. Thus, for sufficiently large $C_4 > 0$, we have

$$P(\max_l |\mathbf{e}_1 h_{1j}n^{-1/2}\tilde{T}_{\epsilon,j}(h_{1j}, \tilde{s}_l)| > C_4\sqrt{Mh_{1j}|\log h_{1j}|}) \to 0 \quad \text{as} \quad h_{1j} \to 0.$$

In Step 3, we focus on the first component of $\mathbf{z}_{h_{1j}}(s_m - s)$. We first consider the following function class:

$$\mathcal{E}_l = \{w_l(S; s) = h_{1j}[K_{h_{1j}}(S - \tilde{s}_l) - K_{h_{1j}}(S - s)]F_n(S)\mathbf{1}(||F_n(S)||_2 \leq \gamma_M) : s \in [\tilde{s}_{l-1}, \tilde{s}_l]\}.$$

It follows from Assumption (C5) and γ_M that $\mathcal{E}_l$ is a pointwise measurable class of functions and $\sup_{s\in[0,1]} |w_l(S; s)| \leq C_6\gamma_M \leq C_7\sqrt{Mh_{1j}/|\log h_{1j}|}$. Let $||\phi||_D = \sup_{z\in D} |\phi(z)|$ for any real valued function ϕ defined on a set D and $\tau_1, \ldots, \tau_M$ be a sequence of independent Rademacher random variables independent of observed data. It follows from an inequality of Talagrand [12, 3] that conditioning on $\mathbf{X}$, we have for suitable finite constants $A_1, A_2 > 0$

$$\begin{aligned}&P\{||\sum_{m=1}^{M}[w_l(s_m;s) - E[w_l(s_m;s)|\mathbf{X}]||_{\mathcal{E}_l} \geq A_1(E[||\sum_{m=1}^{M}\tau_j w_l(s_m;s)||_{\mathcal{E}_l}|\mathbf{X}] + t)|\mathbf{X}\}\\ &\leq 2[\exp(-A_2t^2/(MV_{\mathcal{E}_l}(\mathbf{X}))) + \exp(-A_2t/(C_7\sqrt{Mh_{1j}/|\log h_{1j}|}))],\end{aligned}$$

where $V_{\mathcal{E}_l}(\mathbf{X}) = \sup_{s\in[\tilde{s}_{l-1},\tilde{s}_l]} \mathrm{Var}(w_l(S;s)|\mathbf{X})$. It can be shown that

$$V_{\mathcal{E}_l}(\mathbf{X}) \leq \sup_{s\in[\tilde{s}_{l-1},\tilde{s}_l]} E_S\{h_{1j}^2[K_{h_{1j}}(S-\tilde{s}_l) - K_{h_{1j}}(S-s)]^2 E[F_n(S)^2|\mathbf{X}]\} \leq C_8 h_{1j} n^{-1}\sum_{i=1}^{n}\mathbf{x}_i^{\otimes 2},$$

where C_8 is a positive scalar. By setting $t = C_9\sqrt{Mh_{1j}|\log h_{1j}|}$ for a large $C_9 > 0$, we can show that $A_2t^2/(MV_{\mathcal{E}_l}(\mathbf{X})) = C_{10}|\log h_{1j}|$ and $A_2t/(C_7\sqrt{Mh_{1j}/|\log h_{1j}|}) = C_{11}|\log h_{1j}|$. Moreover, it follows from Assumption (C5) that $\mathcal{E}_l$ is a pointwise measurable Vapnik and Cervonenkis (VC) class [13]. By using Proposition A.1 of [3], we can show that $\max_l E[||\sum_{m=1}^{M}\tau_j w_l(s_m;s)||_{\mathcal{E}_l}|\mathbf{X}] \leq O(\sqrt{Mh_{1j}|\log h_{1j}|})$. This yields (13).

Lemma 2 . Under Assumptions (C1), (C4), (C5) and (C7), we have that for any $r \geq 0$ and j,

$$\sup_{s\in[0,1]}\left|\int K_{h_{1j}}(u-s)\frac{(u-s)^r}{h_{1j}^r}d[\Pi_M(u) - \Pi(u)]\right| = O_p((Mh_{1j})^{-1/2}), \tag{18}$$

$$\sup_{s\in[0,1]}\left|\int K_{h_{1j}}(u-s)\frac{(u-s)^r}{h_{1j}^r}\epsilon_{ij}(u)d\Pi_M(u)\right| = O_p((Mh_{1j})^{-1/2}\sqrt{|\log h_{1j}|}), \tag{19}$$

where $\Pi_M(\cdot)$ is the sampling distribution function based on $\mathcal{S} = \{s_1, \ldots, s_M\}$, and $\Pi(\cdot)$ is the distribution function of s_m.

Proof. Equations (18) can be proved by using the empirical process techniques [13]. Specifically, it follows from Assumptions (C4) and (C5) that

$$\{K\left(\frac{\cdot - s}{h}\right)\frac{(\cdot - s)^r}{h^r} : s \in [0,1]\} \text{ is a } P-\text{Donsker class.}$$

Equation (19) can be proved by using the same arguments of Lemma 1, so we omit the details.

Lemma 3 . Under Assumptions (C2)-(C5), we have

$$\sup_{s\in[0,1]}|n^{-1/2}\sum_{i=1}^{n}\mathbf{x}_i \otimes \Delta_j(s;\boldsymbol{\eta}_i, h_{1j})| = o_p(1). \tag{20}$$

Proof. It follows from the Donsker Theorem [13] that

$$A_{n,\eta}(s) = n^{-1/2}\sum_{i=1}^{n}\mathbf{x}_i\eta_{ij}(s) \Rightarrow G_1(s), \text{ and } M^{1/2}[\Pi_M(s) - \Pi(s)] \Rightarrow G_2(s), \tag{21}$$

$$\sup_{s\in[0,1]}||n^{-1/2}\sum_{i=1}^{n}\mathbf{x}_i\eta_{ij}(s)|| = O_p(1) \text{ and } \sup_{s\in[0,1]}|\Pi_M(s) - \Pi(s)| = O_p(M^{-1/2}), \tag{22}$$

where $G_1(\cdot)$ and $G_2(\cdot)$ are two centered Gaussian processes. Furthermore, we have

$$
\begin{aligned}
(23)\quad & n^{-1/2}\sum_{i=1}^{n}\mathbf{x}_i\otimes\Delta_j(s;\boldsymbol{\eta}_i,h_{1j}) = \text{(I)+(II)+(III)} \\
& = M^{-1}\sum_{m=1}^{M}K_{h_{1j}}(s_m-s)\mathbf{z}_{h_{1j}}(s_m-s)[n^{-1/2}\sum_{i=1}^{n}\mathbf{x}_i\otimes\eta_{ij}(s_m)-n^{-1/2}\sum_{i=1}^{n}\mathbf{x}_i\otimes\eta_{ij}(s)] \\
+ \; & [M^{-1}\sum_{m=1}^{M}K_{h_{1j}}(s_m-s)\mathbf{z}_{h_{1j}}(s_m-s)-\int_0^1 K_{h_{1j}}(u-s)\mathbf{z}_{h_{1j}}(u-s)d\Pi(u)]n^{-1/2}\sum_{i=1}^{n}\mathbf{x}_i\otimes\eta_{ij}(s) \\
+ \; & \int K_{h_{1j}}(u-s)\mathbf{z}_{h_{1j}}(u-s)n^{-1/2}\sum_{i=1}^{n}\mathbf{x}_i\otimes[\eta_{ij}(s)-\eta_{ij}(u)]d\Pi(u).
\end{aligned}
$$

We examine the three terms on the righthand side of (23) as follows:

$$
\begin{aligned}
\text{(I)} \;\leq\; & M^{-1}\sum_{m=1}^{M}K_{h_{1j}}(s_m-s)||\mathbf{z}_{h_{1j}}(s_m-s)||_2|A_{n,\eta}(s_m)-A_{n,\eta}(s)| \\
\leq\; & \sup_{|s-s'|\leq h_{1j}}|A_{n,\eta}(s')-A_{n,\eta}(s)|\sup_s M^{-1}\sum_{m=1}^{M}K_{h_{1j}}(s_m-s)||\mathbf{z}_{h_{1j}}(s_m-s)||_2 = o_p(1), \\
\text{(II)} \;\leq\; & \sup_{s\in[0,1]}|A_{n,\eta}(s)|\sup_{s\in[0,1]}||\int_0^1 K_{h_{1j}}(u-s)\mathbf{z}_{h_{1j}}(u-s)d[\Pi_M(u)-\Pi(u)]||_2 = O_p((Mh_{1j})^{-1/2}), \\
\text{(III)} \;\leq\; & \int K_{h_{1j}}(u-s)||\mathbf{z}_{h_{1j}}(u-s)||_2|A_{n,\eta}(u)-A_{n,\eta}(s)|d\Pi(u) \\
\leq\; & \sup_{|s-s'|\leq h_{1j}}|A_{n,\eta}(s')-A_{n,\eta}(s)|\sup_s\int K_{h_{1j}}(u-s)||\mathbf{z}_{h_{1j}}(u-s)||_2 d\Pi(u) = o_p(1).
\end{aligned}
$$

This finishes the proof of Lemma 3.
Lemma 4 . If Assumptions (C1) and (C3)-(C6) hold, then for any $s\in(0,1)$, we have

$$
\begin{aligned}
(24)\quad \mathrm{E}[\hat{B}_j(s)|\mathcal{S}]-B_j(s) &= 0.5h_{1j}^2u_2(K)\ddot{B}_j(s)[1+O_p(n^{-1/2}+h_{1j}+(Mh_{1j})^{-1/2})] \\
&= 0.5h_{1j}^2u_2(K)\ddot{B}_j(s)[1+o_p(1)], \\
\mathrm{Var}[\hat{B}_j(s)|\mathcal{S}] &= n^{-1}\Sigma_{\eta,jj}(s,s)\Omega_X^{-1}[1+o_p(1)],
\end{aligned}
$$

where $e_n(s)=O_p((Mh_{1j})^{-1/2})$ is defined in (37) with $E[e_n(s)]=0$.

Proof. First, we calculate $\mathrm{E}[\hat{B}_j(s)|\mathcal{S}]$. The $\mathrm{vec}(\hat{A}_j(s))$ can be written as follows:

$$\mathrm{vec}(\hat{A}_j(s)) = \Sigma(s,h_{1j})^{-1}[T_{B,j}(h_{1j},s)+T_{\epsilon,j}(h_{1j},s)+T_{\eta,j}(h_{1j},s)].$$

Because the components of $B_j(s)$ are differentiable in the neighborhood of $|s_m-s|\leq h_{1j}$, it follows from a Taylor's expansion and Assumption (C6) that

$$(25)\qquad \mathbf{x}_i^T B_j(s_m) = \mathbf{x}_i^T A_j(s)\mathbf{z}_{h_{1j}}(s_m-s)+0.5h_{1j}^2\mathbf{x}_i^T\ddot{B}_j(s)\frac{(s_m-s)^2}{h_{1j}^2}+o_P(h_{1j}^2).$$

For $r\geq 0$, it follows from Assumptions (C4) and (C5) that

$$h_{1j}^{-r}|M^{-1}\sum_{m=1}^{M}(s_m-s)^r K_{h_{1j}}(s_m-s)-\int(u-s)^r K_{h_{1j}}(u-s)\pi(u)du| = O_p((Mh_{1j})^{-1/2}).$$

Thus, by substituting (25) into $\hat{B}_j(s)$, we have

$$\hat{B}_j(s) - B_j(s) = [I_p \otimes (1,0)]\Sigma(s,h_{1j})^{-1} \times \{T_{\epsilon,j}(h_{1j},s) + T_{\eta,k}(h_{1j},s) + 0.5h_{1j}^2[\sum_{i=1}^n \mathbf{x}_i^{\otimes 2}\ddot{B}_j(s)] \otimes [\sum_{m=1}^M K_{h_{1j}}(s_m - s)\mathbf{z}_{h_{1j}}(s_m - s)\frac{(s_m - s)^2}{h_{1j}^2}][1 + o_P(1)]\}.$$

It follows from Assumptions (C4)-(C6) that

$$\begin{aligned} \Sigma(s,h_{1j}) &= (nM)(n^{-1}\sum_{i=1}^n \mathbf{x}_i^{\otimes 2}) \otimes [M^{-1}\sum_{m=1}^M K_{h_{1j}}(s_m - s)\mathbf{z}_{h_{1j}}(s_m - s)^{\otimes 2}] \\ &= (nM)[\Omega_X \otimes \Omega_1(h_{1j},s) + O_p(n^{-1/2} + M^{-1/2}h_{1j}^{-1})] \\ &= (nM)[\pi(s)\Omega_X \otimes \mathrm{diag}(1,u_2(K)) + O_p(n^{-1/2} + M^{-1/2}h_{1j}^{-1} + h_{1j})]. \end{aligned} \tag{26}$$

Thus, with some calculation, we have

$$E[\hat{B}_j(s)|\mathcal{S}] - B_j(s) = 0.5h_{1j}^2 u_2(K)\ddot{B}_j(s)[1 + O_p(n^{-1/2} + h_{1j} + (Mh_{1j})^{-1/2})]. \tag{27}$$

Secondly, we calculate $\mathrm{Var}[\hat{B}_j(s)|\mathcal{S}]$. We note that

$$\mathrm{Var}[\hat{B}_j(s)|\mathcal{S}] = E\{\mathrm{Var}[\hat{B}_j(s)|\mathcal{S},\mathbf{X}]|\mathcal{S}\} + \mathrm{Var}\{E[\hat{B}_j(s)|\mathcal{S},\mathbf{X}]|\mathcal{S}\}. \tag{28}$$

Define $\hat{\Omega}_X^{-1} = \sum_{i=1}^n \mathbf{x}_i^{\otimes 2}$, $P_1(s) = [I_p \otimes (1,0)]\Sigma(s,h_{1j})^{-1}$, $P_2(s) = M^{-1}\sum_{m=1}^M K_{h_{1j}}(s_m - s)\mathbf{z}_{h_{1j}}(s_m - s)^{\otimes 2}$ and $P_3(s) = M^{-1}\sum_{m=1}^M K_{h_{1j}}(s_m - s)B_j(s_m) \otimes \mathbf{z}_{h_{1j}}(s_m - s)$. With some calculation, we have

$$\begin{aligned} E[\hat{B}_j(s)|\mathcal{S},\mathbf{X}] &= P_1(s)\sum_{i=1}^n\{\mathbf{x}_i \otimes [\sum_{m=1}^M H_{h_{1j}}(s_m - s)\mathbf{x}_i^T B_j(s_m)]\} \qquad (29) \\ &= [I_p \otimes (1,0)P_2(s)^{-1}]P_3(s), \\ \mathrm{Var}[\hat{B}_j(s)|\mathcal{S},\mathbf{X}] &= P_1(s)\mathrm{Var}[T_{\epsilon,j}(h_{1j},s) + T_{\eta,j}(h_{1j},s)|\mathcal{S},\mathbf{X}]P_1(s)^T. \qquad (30) \end{aligned}$$

Thus, because $P_2(s)$ and $P_3(s)$ solely depend on $\mathcal{S}$, we have

$$\mathrm{Var}\{E[\hat{B}_j(s)|\mathcal{S},\mathbf{X}]|\mathcal{S}\} = 0. \tag{31}$$

We calculate $E\{\mathrm{Var}[\hat{B}_j(s)|\mathcal{S},\mathbf{X}]|\mathcal{S}\}$. Define $A_{\epsilon j}(s) = \sum_{m=1}^M H_{h_{1j}}(s_m - s)H_{h_{1j}}(s_m - s)^T\Sigma_{\epsilon,jj}(s_m,s_m)$ and $A_{\eta j}(s) = \sum_{m,m'=1}^M H_{h_{1j}}(s_m - s)H_{h_{1j}}(s_{m'} - s)^T\Sigma_{\eta,jj}(s_m,s_{m'})$. It is easy to see that

$$\mathrm{Var}[T_{\epsilon,j}(h_{1j},s) + T_{\eta,j}(h_{1j},s)|\mathcal{S},\mathbf{X}] = n(\hat{\Omega}_X - \Omega_X + \Omega_X) \otimes [A_{\epsilon j}(s) + A_{\eta j}(s)].$$

With some calculation, we have

$$\begin{aligned} \Sigma(s,h_{1j}) &= nM[n^{-1}\sum_{i=1}^n(\mathbf{x}_i^{\otimes 2} - \Omega_X)\Omega_X^{-1} \otimes I_2 + I_p \otimes I_2][\Omega_X \otimes P_2(s)], \\ \Sigma(s,h_{1j})^{-1} &= (nM)^{-1}[\Omega_X^{-1} \otimes P_2(s)^{-1}]\{I_{2p} - n^{-1}\sum_{i=1}^n(\mathbf{x}_i^{\otimes 2} - \Omega_X)\Omega_X^{-1} \otimes I_2 \\ &\quad +[n^{-1}\sum_{i=1}^n(\mathbf{x}_i^{\otimes 2} - \Omega_X)\Omega_X^{-1} \otimes I_2]^2 + O_p(n^{-3/2})\}. \end{aligned}$$

Based on the expansion of $\Sigma(s,h_{1j})^{-1}$, we have

$$
\begin{aligned}
(32)\quad & \mathrm{Var}[\hat{B}_j(s)|\mathcal{S},\mathbf{X}] \\
= \; & [I_p\otimes(1,0)][\Omega_X^{-1}\otimes P_2(s)^{-1}]\times\{n^{-1}M^{-2}(\hat{\Omega}_X-\Omega_X+\Omega_X)\otimes[A_{\epsilon j}(s)+A_{\eta j}(s)] \\
& -2(nM)^{-2}\sum_{i=1}^{n}(\mathbf{x}_i^{\otimes 2}-\Omega_X)\otimes[A_{\epsilon j}(s)+A_{\eta j}(s)]+O_p(n^{-2}+(n^2Mh_{1j})^{-1})\}\times \\
& [\Omega_X^{-1}\otimes P_2(s)^{-1}][I_p\otimes(1,0)^T], \\
(33)\quad & E\{\mathrm{Var}[\hat{B}_j(s)|\mathcal{S},\mathbf{X}]|\mathcal{S}\} \\
= \; & O_p(n^{-2}+(n^2Mh_{1j})^{-1})+n^{-1}M^{-2}\Omega_X^{-1}\otimes\{(1,0)P_2(s)^{-1}[A_{\epsilon j}(s)+A_{\eta j}(s)]P_2(s)^{-1}(1,0)^T\}.
\end{aligned}
$$

We approximate $A_{\epsilon j}(s)$ and $A_{\eta j}(s)$ as follows. It follows from Assumption (C1) that

$$
(34)\qquad (1,0)A_{\epsilon j}(s)(1,0)^T\le\sup_{s_m}\Sigma_{\epsilon,jj}(s_m,s_m)\sum_{m=1}^{M}K_{h_{1j}}(s_m-s)^2=O(Mh_{1j}^{-1}).
$$

Moreover, $A_{\eta j}(s)=A_{\eta j}^{(1)}(s)+A_{\eta j}^{(2)}(s)$, where $A_{\eta j}^{(1)}(s)=\sum_{m=1}^{M}H_{h_{1j}}(s_m-s)H_{h_{1j}}(s_m-s)^T\Sigma_{\eta,jj}(s_m,s_m)$ and $A_{\eta j}^{(2)}(s)=\sum_{m\neq m'}^{M}H_{h_{1j}}(s_m-s)H_{h_{1j}}(s_{m'}-s)^T\Sigma_{\eta,jj}(s_m,s_{m'})$. Particularly, $A_{\eta j}^{(2)}(s)/[(M-1)M]$ is a U-statistic [2]. Similar to $A_{\epsilon j}(s)$, it can be shown that

$$
(35)\qquad (1,0)A_{\eta j}(s)^{(1)}(1,0)^T=Mh_{1j}^{-1}[\pi(s)\Sigma_{\eta,jj}(s,s)v_0(K)+O_p((Mh_{1j})^{-1/2}+h_{1j})].
$$

For $A_{\eta}^{(2)}(s)$, we define three 2×2 matrices $U_\eta(s)=(U_{\eta,ll'}(s))=[M(M-1)]^{-1}A_{\eta}^{(2)}(s)$, $\theta(s)=(\theta_{ll'}(s))=E[U_\eta(s)]$, and $P_\eta(v)=(P_{\eta,ll'}(v))=\int H_{h_{1j}}(v-s)H_{h_{1j}}(u-s)^T\Sigma_{\eta,jj}(v,u)\pi(u)du$. By using the Hajek projection, we have

$$
(36)\qquad U_{\eta,ll'}(s)=\theta_{ll'}(s)+\frac{2}{M}\sum_{m=1}^{M}[P_{\eta,ll'}(s_m)-\theta_{ll'}(s)]+\tilde{E}_{n,ll'}(s)\quad\text{for}\quad l,l'=1,2,
$$

in which $2\sum_{m=1}^{M}[P_{\eta,ll'}(s_m)-\theta_{ll'}(s)]/M$ is the projection of $U_{\eta,ll'}(s)-\theta_{ll'}(s)$ onto the set of all statistics of the form $\sum_{m=1}^{M}f_m(s_m)$. Thus, with some calculation, we have

$$
\mathrm{Var}(\tilde{E}_{n,ll'}(s)) \;=\; \mathrm{Var}(U_{\eta,ll'}(s)-\theta_{ll'}(s))-\mathrm{Var}(\frac{2}{M}\sum_{m=1}^{M}[P_{\eta,ll'}(s_m)-\theta_{ll'}(s)])=O((Mh_{1j})^{-2}).
$$

As $h_{1j}\to 0$, it follows from Taylor's expansion that

$$
\theta_{ll'}(s)+O(h_{1j})=\pi(s)^2u_{l-1}(K)u_{l'-1}(K)\Sigma_{\eta,jj}(s,s).
$$

Define

$$
(37)\qquad e_n(s)=2\sum_{m=1}^{M}[P_{\eta,11}(s_m)-\theta_{11}(s)]/M+\tilde{E}_{n,11}(s)=O_p((Mh_{1j})^{-1/2}).
$$

Then, we have

$$
(38)\qquad (1,0)A_{\eta j}(s)^{(2)}(1,0)^T=M^2e_n(s)+M^2\pi(s)^2\Sigma_{\eta,jj}(s,s)[1+o_p(1)].
$$

Substituting (30)-(38) into (28), we can obtain $E\{\mathrm{Var}[\hat{B}_j(s)|\mathcal{S},\mathbf{X}]|\mathcal{S}\}$ and $\mathrm{Var}[\hat{B}_j(s)|\mathcal{S}]$.

Lemma 5. If Assumptions (C1) and (C3)-(C6) hold, then for $s = 0$ or 1, we have

$$\begin{aligned} \mathrm{E}[\hat{B}_j(s)|\mathcal{S}] - B_j(s) &= 0.5h_{1j}^2 r_u(K; s, h_{1j})\ddot{B}_j(s)[1 + o_p(1)], \\ \mathrm{Var}[\hat{B}_j(s)|\mathcal{S}] &= n^{-1}\Sigma_{\eta,jj}(s, s)\Omega_X^{-1}[1 + o_p(1)]. \end{aligned} \tag{39}$$

Proof. Since Lemma 5 follows directly from the proof of Lemma 4, we just highlight how to compute the bias of $\hat{B}_j(s)$ given $\mathcal{S}$. If follows from the change of variable that all elements of $\Omega_1(h, s)$ can be written as

$$\int_0^1 h^{-r}(u - s)^r K_h(u - s)\pi(u)du = \pi(s)u_r(K; s, h) + O(h) \tag{40}$$

for $r = 0, 1,$ and 2. Thus, it follows from the arguments in (26) and (40) that

$$\Sigma(s, h_{1j}) = (nM)[\pi(s)\Omega_X \otimes \begin{pmatrix} u_0(K; s, h_{1j}) & u_1(K; s, h_{1j}) \\ u_1(K; s, h_{1j}) & u_2(K; s, h_{1j}) \end{pmatrix} + O_p(n^{-1/2} + (Mh_{1j})^{-1/2} + h_{1j})].$$

Using arguments similar to (27), we get

$$E[\hat{B}_j(s)|\mathcal{S}] - B_j(s) = 0.5h_{1j}^2 \frac{u_2(K; s, h_{1j})^2 - u_1(K; s, h_{1j})u_3(K; s, h_{1j})}{u_0(K; s, h_{1j})u_2(K; s, h_{1j}) - u_1(K; s, h_{1j})^2}\ddot{B}_j(s)[1 + o_p(1)]. \tag{41}$$

Proof of Theorem 2. Let $\tilde{K}_{M,h}(s) = \tilde{K}_M(s/h)/h$, where $\tilde{K}_M(s)$ is the empirical equivalent kernels for the first-order local polynomial kernel [4]. Thus, we have

$$\begin{aligned} \hat{\eta}_{ij}(s) - \eta_{ij}(s) &= \sum_{m=1}^M \tilde{K}_{M,h_{2j}}(s_m - s)\mathbf{x}_i^T[B_j(s_m) - \hat{B}_j(s_m)] \\ &\quad + \sum_{m=1}^M \tilde{K}_{M,h_{2j}}(s_m - s)[\eta_{ij}(s_m) + \epsilon_{ij}(s_m) - \eta_{ij}(s)]. \end{aligned} \tag{42}$$

It follows from a Taylor's expansion that

$$\sum_{m=1}^M \tilde{K}_{M,h_{2j}}(s_m - s)[\eta_{ij}(s_m) - \eta_{ij}(s)] = 0.5u_2(K)\ddot{\eta}_{ij}(s)h_{2j}^2[1 + o_p(1)],$$

and

$$\begin{aligned} &\sum_{m=1}^M \tilde{K}_{M,h_{2j}}(s_m - s)\mathbf{x}_i^T\{B_j(s_m) - E[\hat{B}_j(s_m)|\mathcal{S}, \boldsymbol{\eta}, \mathbf{X}]\} \\ = &\sum_{m=1}^M \tilde{K}_{M,h_{2j}}(s_m - s)[0.5h_{1j}^2 u_2(K)\mathbf{x}_i^T\ddot{B}_j(s_m) + \mathbf{x}_i^T\Omega_X^{-1}n^{-1}\sum_{i'=1}^n \mathbf{x}_{i'}\eta_{i',j}(s_m)] \\ &\times[1 + O_p(h_{1j} + n^{-1/2} + (Mh_{1j})^{-1/2})] \\ = &[0.5h_{1j}^2 u_2(K)\mathbf{x}_i^T\ddot{B}_j(s) + \mathbf{x}_i^T\Omega_X^{-1}n^{-1}\sum_{i'=1}^n \mathbf{x}_{i'}\eta_{i',j}(s)] \\ &\times[1 + O_p(h_{1j} + h_{2j} + n^{-1/2}) + O_p((Mh_{2j})^{-1/2} + (Mh_{1j})^{-1/2})] \\ = &[0.5h_{1j}^2 u_2(K)\mathbf{x}_i^T\ddot{B}_j(s) + O_p(n^{-1/2})] \\ &\times[1 + O_p(h_{1j} + h_{2j} + n^{-1/2}) + O_p((Mh_{2j})^{-1/2} + (Mh_{1j})^{-1/2})], \end{aligned}$$

which leads to Bias$[\hat{\eta}_{ij}(s)|\mathcal{S},\boldsymbol{\eta},\mathbf{X}]$.

Furthermore, it can be shown that

$$\hat{\eta}_{ij}(s)-E[\hat{\eta}_{ij}(s)|\mathcal{S},\boldsymbol{\eta},\mathbf{X}]=\sum_{m=1}^{M}\tilde{K}_{M,h_{2j}}(s_m-s)\{\epsilon_{ij}(s_m)-\mathbf{x}_i^T[\mathbf{I}_p\otimes(1,0)]\Sigma(s_m,h_{1j})^{-1}T_{\epsilon,j}(h_{1j},s_m)\}$$
$$=\sum_{m=1}^{M}\tilde{K}_{M,h_{2j}}(s_m-s)\{\epsilon_{ij}(s_m)-\mathbf{x}_i^T[\mathbf{I}_p\otimes(1,0)]\Sigma(s_m,h_{1j})^{-1}\sum_{i'=1}^{n}\mathbf{x}_{i'}\otimes\sum_{m'=1}^{M}H_{h_{1j}}(s_{m'}-s_m)\epsilon_{i',j}(s_{m'})\}.$$

With tedious calculations, we have

$$\begin{aligned}&\mathrm{Cov}(\hat{\eta}_{ij}(s)-\eta_{ij}(s),\hat{\eta}_{ij}(t)-\eta_{ij}(t)|\mathcal{S},\boldsymbol{\eta},\mathbf{X})\\=\;&K^*((s-t)/h_{2j})\pi(t)^{-1}O_p((Mh_{2j})^{-1})-\mathbf{x}_i^T\Omega_X^{-1}\mathbf{x}_i(nMh_{1j})^{-1}\pi(s)^{-1}\pi(t)^{-1}O_p(1).\end{aligned}$$

Furthermore, for $i=1,\ldots,n$, after dropping some higher order terms, we have

$$\begin{aligned}&E\{[\hat{\eta}_{ij}(s)-\eta_{ij}(s)]^2|\mathcal{S},\boldsymbol{\eta},\mathbf{X}\}\\&=\{E[\hat{\eta}_{ij}(s)-\eta_{ij}(s)|\mathcal{S},\boldsymbol{\eta},\mathbf{X}]\}^2+\mathrm{Var}[\hat{\eta}_{ij}(s)-\eta_{ij}(s)|\mathcal{S},\boldsymbol{\eta},\mathbf{X}]\\&=[0.5h_{1j}^2u_2(K)\mathbf{x}_i^T\ddot{B}_j(s_m)+0.5u_2(K)\ddot{\eta}_{ij}(s)h_{2j}^2+\mathbf{x}_i^T\Omega_X^{-1}n^{-1}\sum_{i'=1}^{n}\mathbf{x}_{i'}\eta_{i'j}(s_m)]^2[1+o_p(1)]\\&\quad+v_0(K)\pi(s)^{-1}O_p((Mh_{2j})^{-1})-\mathbf{x}_i^T\Omega_X^{-1}\mathbf{x}_i(nMh_{1j})^{-1}\pi(s)^{-1}O_p(1).\end{aligned}$$

This completes the proof of Theorem 2 (a).

It follows from (27) that

$$\begin{aligned}&\sum_{m=1}^{M}\tilde{K}_{M,h_{2j}}(s_m-s)\mathbf{x}_i^T\{B_j(s_m)-E[\hat{B}_j(s_m)|\mathcal{S},\mathbf{X}]\}\\=\;&\sum_{m=1}^{M}\tilde{K}_{M,h_{2j}}(s_m-s)0.5h_{1j}^2u_2(K)\mathbf{x}_i^T\ddot{B}_j(s_m)[1+O_p(h_{1j}+n^{-1/2}+(Mh_{1j})^{-1/2})].\end{aligned}$$

Furthermore, it can be shown that

$$\begin{aligned}&\hat{\eta}_{ij}(s)-\eta_{ij}(s)-E[\hat{\eta}_{ij}(s)|\mathcal{S},\mathbf{X}]\\=\;&\sum_{m=1}^{M}\tilde{K}_{M,h_{2j}}(s_m-s)[\eta_{ij}(s_m)+\epsilon_{ij}(s_m)-\eta_{ij}(s)]\\&-\sum_{m=1}^{M}\tilde{K}_{M,h_{2j}}(s_m-s)\{\mathbf{x}_i^T[\mathbf{I}_p\otimes(1,0)]\Sigma(s_m,h_{1j})^{-1}[T_{\eta,j}(h_{1j},s_m)+T_{\epsilon,j}(h_{1j},s_m)]\}.\end{aligned}$$

With tedious calculations, we have

$$\begin{aligned}&\mathrm{Cov}(\hat{\eta}_{ij}(s)-\eta_{ij}(s),\hat{\eta}_{ij}(t)-\eta_{ij}(t)|\mathcal{S},\mathbf{X})\\=\;&K^*((s-t)/h_{2j})\pi(t)^{-1}O_p((Mh_{2j})^{-1})-\mathbf{x}_i^T\Omega_X^{-1}\mathbf{x}_i(nMh_{1j})^{-1}\pi(s)^{-1}\pi(t)^{-1}O_p(1)\\&+[1+o_p(1)]\{0.25u_2(K)^2h_{1j}^{(2)4}\Sigma_{\eta,jj}^{(2,2)}(s,t)+n^{-1}\mathbf{x}_i^T\Omega_X^{-1}\mathbf{x}_i\Sigma_{\eta,jj}(s,t)\\&-0.5n^{-1}u_2(K)h_{2j}^2\mathbf{x}_i^T\Omega_X^{-1}\mathbf{x}_i[\Sigma_{\eta,jj}^{(2,0)}(s,t)\pi(s)^{-1}+\Sigma_{\eta,jj}^{(0,2)}(s,t)\pi(t)^{-1}]\}.\end{aligned}$$

It follows from (27) that

$$
\begin{aligned}
& E\{[\hat{\eta}_{ij}(s)-\eta_{ij}(s)]^2|\mathcal{S},\mathbf{X}] \\
= \; & \{E[\hat{\eta}_{ij}(s)-\eta_{ij}(s)|\mathcal{S},\mathbf{X}]\}^2+\mathrm{Var}[\hat{\eta}_{ij}(s)-\eta_{ij}(s)|\mathcal{S},\mathbf{X}] \\
= \; & \{0.25h_{1j}^4u_2(K)^2[\mathbf{x}_i^T\ddot{B}_j(s)]^2+0.25u_2(K)^2h_{1j}^{(2)4}\Sigma_{\eta,jj}^{(2,2)}(s,t) \\
& +n^{-1}\mathbf{x}_i^T\Omega_X^{-1}\mathbf{x}_i\Sigma_{\eta,jj}(s,t)+v_0(K)\pi(s)^{-1}O_p((Mh_{2j})^{-1})\}[1+o_p(1)],
\end{aligned}
$$

which leads to Theorem 2 (b). Furthermore, by noting that $E\{[\hat{\eta}_{ij}(s)-\eta_{ij}(s)]^2|\mathcal{S}]=E(E\{[\hat{\eta}_{ij}(s)-\eta_{ij}(s)]^2|\mathcal{S},\mathbf{X}]|\mathcal{S})$, we can easily get Theorem 2 (c) and (d).

We define

$$
\begin{aligned}
\bar{\epsilon}_{ij}(s) &= \sum_{m=1}^M \tilde{K}_{M,h_{2j}}(s_m-s)\epsilon_{ij}(s_m), \\
\Delta\eta_{ij}(s) &= \sum_{m=1}^M \tilde{K}_{M,h_{2j}}(s_m-s)[\eta_{ij}(s_m)-\eta_{ij}(s)], \\
\Delta B_j(s) &= \sum_{m=1}^M \tilde{K}_{M,h_{2j}}(s_m-s)[B_j(s_m)-\hat{B}_j(s_m)], \\
\Delta_{ij}(s) &= \bar{\epsilon}_{ij}(s)+\Delta\eta_{ij}(s)+\mathbf{x}_i^T\Delta B_j(s).
\end{aligned}
$$

Recall from (42) that

$$
\hat{\eta}_{ij}(s)-\eta_{ij}(s)=\Delta_{ij}(s)=\bar{\epsilon}_{ij}(s)+\Delta\eta_{ij}(s)+\mathbf{x}_i^T\Delta B_j(s). \tag{43}
$$

It follows from Lemma 2 and a Taylor's expansion that

$$
\sup_{s\in[0,1]}|\bar{\epsilon}_{ij}(s)|=O_p(\sqrt{\frac{|\log(h_{2j})|}{Mh_{2j}}}) \text{ and } \sup_{s\in[0,1]}|\Delta\eta_{ij}(s)|=O_p(1)\sup_{s\in[0,1]}|\ddot{\eta}_{ij}(s)|h_{1j}^{(2)2}.
$$

Since $\sqrt{n}\{\hat{B}_j(\cdot)-B_j(\cdot)-0.5u_2(K)^2h_{1j}^2\ddot{B}_j(\cdot)[1+o_p(1)]\}$ weakly converges to a Gaussian process in $\ell^\infty([0,1])$ as $n\to\infty$, $\sqrt{n}\{\hat{B}_j(\cdot)-B_j(\cdot)-0.5u_2(K)^2h_{1j}^2\ddot{B}_j(\cdot)[1+o_p(1)]\}$ is asymptotically tight. Thus, we have

$$
\begin{aligned}
& \Delta B_{ij}(s)=-\sum_{m=1}^M \tilde{K}_{M,h_{2j}}(s_j-s)0.5u_2(K)^2h_{1j}^2\ddot{B}_j(s_m)[1+o_p(1)] \\
& +\sum_{m=1}^M \tilde{K}_{M,h_{2j}}(s_j-s)\{0.5u_2(K)^2h_{1j}^2\ddot{B}_j(s_m)[1+o_p(1)]+B_j(s_m)-\hat{B}_j(s_m)\}, \\
& \sup_{s\in[0,1]}||\Delta B_j(s)||=O_p(n^{-1/2})+O_p(h_{1j}^2).
\end{aligned}
$$

Combining these results, we have

$$
\sup_{s\in[0,1]}|\hat{\eta}_{ij}(s)-\eta_{ij}(s)|=O_p(|\log(h_{2j})|^{1/2}(Mh_{2j})^{-1/2}+h_{1j}^{(2)2}+h_{1j}^2+n^{-1/2}).
$$

Lemma 6. Under Assumptions (C1)-(C9), we have

$$\sup_{(s,t)} n^{-1}|\sum_{i=1}^{n}\overline{\epsilon}_{ij}(s)\eta_{ij}(t)| = O_p(n^{-1/2}(\log n)^{1/2}),$$
$$\sup_{(s,t)} n^{-1}|\sum_{i=1}^{n}\overline{\epsilon}_{ij}(s)\Delta\eta_{ij}(t)| = O_p(n^{-1/2}(\log n)^{1/2}),$$
$$\sup_{s} n^{-1}|\sum_{i=1}^{n}\overline{\epsilon}_{ij}(s)\mathbf{x}_i| = O_p(n^{-1/2}(\log n)^{1/2}),$$
$$\sup_{s} n^{-1}|\sum_{i=1}^{n}\Delta\eta_{ij}(s)\mathbf{x}_i| = O_p(n^{-1/2}(\log n)^{1/2}).$$

Proof. For simplicity, we only prove the first result. We define $c^+ = c\mathbf{1}(c \geq 0)$ and $c^- = c\mathbf{1}(c < 0)$ for any scalar c, $\Delta_{n,\epsilon\eta}(s,t) = \sum_{i=1}^{n}\bar{\epsilon}_{ij}(s)\eta_{ij}(t)$, and

$$G_{n,\epsilon\eta}(s,s+v,t) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{M\pi(s)}\sum_{m=1}^{M}\epsilon_{ij}(s_m)\eta_{ij}(t)\mathbf{1}(s \leq s_m \leq s+v),$$
$$G^+_{n,\epsilon\eta}(s,s+v,t) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{M\pi(s)}\sum_{m=1}^{M}[\epsilon_{ij}(s_m)\eta_{ij}(t)]^+\mathbf{1}(s \leq s_m \leq s+v),$$
$$G^-_{n,\epsilon\eta}(s,s+v,t) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{M\pi(s)}\sum_{m=1}^{M}[\epsilon_{ij}(s_m)\eta_{ij}(t)]^-\mathbf{1}(s \leq s_m \leq s+v).$$

It follows from Lemma 2 that for large enough n, there exists a constant $C_0 > 1$ such that

$$\begin{aligned}
&\sup_{(s,t)\in[0,1]^2} n^{-1}|\Delta_{n,\epsilon\eta}(s,t)| \leq C_0 \sup_{(s,t)\in[0,1]^2}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{M\pi(s)}\sum_{m=1}^{M}K_{h_{2j}}(s_m - s)\epsilon_{ij}(s_m)\eta_{ij}(t)\right| \\
&= C_0 \sup_{(s,t)\in[0,1]^2}\left|\int_{-h_{1j}^{(2)}}^{h_{1j}^{(2)}}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{M\pi(s)}\sum_{m=1}^{M}\epsilon_{ij}(s_m)\eta_{ij}(t)\mathbf{1}(v \leq s_m - s \leq h_{2j})dK_{h_{2j}}(v)\right| \\
&\leq C_0 \sup_{(s,t)\in[0,1]^2}\sup_{|v|\leq 2h_{2j}}|G_{n,\epsilon\eta}(s,s+v,t)|/h_{2j}.
\end{aligned}$$

Let $\beta_n = h_{2j}^2 + h_{2j}/M$. By combining Lemma 1 with Lemma 2 of Li and Hsing [11], we can show that

$$\sup_{(s,t)\in[0,1]^2}\sup_{|u|\leq 2h_{2j}}|G_{n,\epsilon\eta}(s,s+v,t)| = O_p(n^{-1/2}\{\beta_n \log n\}^{1/2}). \tag{44}$$

Since $E[G_{n,\epsilon\eta}(s,v,t)] = 0$, we have

$$|G_{n,\epsilon\eta}(s,s+v,t)| \leq |G^+_{n,\epsilon\eta}(s,s+v,t) - E[G^+_{n,\epsilon\eta}(s,s+v,t)]| + |G^-_{n,\epsilon\eta}(s,s+v,t) - E[G^-_{n,\epsilon\eta}(s,s+v,t)]|.$$

From now on, we focus on $V_{n,\epsilon\eta}(s,s+v,t) = |G^+_{n,\epsilon\eta}(s,s+v,t) - E[G^+_{n,\epsilon\eta}(s,s+v,t)]|$. We define an equally-spaced grid $\tilde{\mathbf{S}} = \{\tilde{s}_l = lh_{1j} : l = 0,\ldots,1h_{1j}^{-1}\}$ and $\tilde{s}_l(s)$ to be a grid point that is within h_{1j} of both s and $s+v$. Since $V_{n,\epsilon\eta}(s,s+v,t)$ is upper bounded by $V_{n,\epsilon\eta}(\tilde{s}_l(s),s+v,t) + V_{n,\epsilon\eta}(\tilde{s}_l(s),s,t)$, we have

$$\sup_{(s,t)\in[0,1]^2}\sup_{|v|\leq 2h_{2j}} V_{n,\epsilon\eta}(s,s+v,t) \leq 2\sup_{\tilde{s}_l\in\tilde{\mathbf{S}}}\sup_{|v|\leq 2h_{2j}}\sup_{t\in[0,1]} V_{n,\epsilon\eta}(\tilde{s}_l,\tilde{s}_l+v,t). \tag{45}$$

Let $a_n^{-1/2}\{\beta_n \log n\}^{1/2}$ and $Q_n = \beta_n/a_n$. We define

$$\tilde{V}_{n,\epsilon\eta}(s, s+v, t) = |\tilde{G}^+_{n,\epsilon\eta}(s, s+v, t) - E[\tilde{G}^+_{n,\epsilon\eta}(s, s+v, t)]|,$$

$$\tilde{G}^+_{n,\epsilon\eta}(s, s+v, t) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{M\pi(s)}\sum_{m=1}^{M}[\epsilon_{ij}(s_m)\eta_{ij}(t)]^+\mathbf{1}(s \le s_m \le s+v)\mathbf{1}(|\epsilon_{ij}(s_m)| \le Q_n),$$

$$\hat{G}^+_{n,\epsilon\eta}(s, s+v, t) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{M\pi(s)}\sum_{m=1}^{M}[\epsilon_{ij}(s_m)\eta_{ij}(t)]^+\mathbf{1}(s \le s_m \le s+v)\mathbf{1}(|\epsilon_{ij}(s_m)| > Q_n).$$

Then, we have

$$\begin{aligned}(46)\qquad &\sup_{(s,t)\in[0,1]^2}\sup_{|v|\le 2h_{2j}} V_{n,\epsilon\eta}(s, s+v, t)\\ &\le 2\sup_{\tilde{s}_l\in\tilde{\mathbf{S}}}\sup_{|v|\le 2h_{2j}}\sup_{t\in[0,1]}\{\tilde{V}_{n,\epsilon\eta}(s, s+v, t) + |\hat{G}^+_{n,\epsilon\eta}(s, s+v, t)| + |E[\hat{G}^+_{n,\epsilon\eta}(s, s+v, t)]|\}.\end{aligned}$$

We consider the three terms on the right side of (46). It is obvious to see that

$$\begin{aligned}(47)\qquad a_n^{-1}E[\hat{G}^+_{n,\epsilon\eta}(s, s+v, t)] &\le a_n^{-1}Q_n^{1-q_2}\frac{1}{n}\sum_{i=1}^{n}E[\sup_{t\in[0,1]}|\eta_{ij}(t)|]\frac{1}{M}\sum_{m=1}^{M}E[|\epsilon_{ij}(s_m)|^{q_2}]\\ &= a_n^{-1}Q_n^{1-q_2}E[\sup_{t\in[0,1]}|\eta_{1,k}(t)|]\frac{1}{M}\sum_{m=1}^{M}E[|\epsilon_{1,k}(s_m)|^{q_2}],\end{aligned}$$

which is independent of (s, v, t) and converges almost surely to zero. Similarly, by using Markov's inequality, we have

$$a_n^{-1}\hat{G}^+_{n,\epsilon\eta}(s, s+v, t) \le a_n^{-1}Q_n^{1-q_2}\frac{1}{n}\sum_{i=1}^{n}\sup_{t\in[0,1]}|\eta_{ij}(t)|\frac{1}{M}\sum_{m=1}^{M}|\epsilon_{ij}(s_m)|^{q_2},$$

which converges almost surely to zero.

We consider a further partition of $[0,1]$ in order to bound $\sup_{|v|\le 2h_{2j}}\sup_{t\in[0,1]}\tilde{V}_{n,\epsilon\eta}(s, s+v, t)$ for each fixed $s \in [0,1]$. Following Li and Hsing (2010), let c_n be any positive sequence tending to 0. We define $\omega_n = [Q_n c_n/a_n + 1]$ and $u_r = rc_n/\omega_n$ for $r = -\omega_n, -\omega_n+1, \ldots, \omega_n$. Since $\tilde{G}^+_{n,\epsilon\eta}(s, s+v, t)$ is monotone in $|v|$, we have that for $v \in [u_r, u_{r+1}]$,

$$|\tilde{G}^+_{n,\epsilon\eta}(s, s+v, t) - E[\tilde{G}^+_{n,\epsilon\eta}(s, s+v, t)]| \le \max(\xi_{nr}(t), \xi_{n,r+1}(t)) + E[\tilde{G}^+_{n,\epsilon\eta}(s+u_r, s+u_{r+1}, t)],$$

where $\xi_{nr}(t) = |\tilde{G}^+_{n,\epsilon\eta}(s+u_r, s+u_{r+1}, t) - E[\tilde{G}^+_{n,\epsilon\eta}(s+u_r, s+u_{r+1}, t)]|$. Thus,

$$\sup_{|v|\le 2h_{2j}}\sup_{t\in[0,1]}\tilde{V}_{n,\epsilon\eta}(s, s+v, t) \le \max_{-\omega_n\le r\le\omega_n}\sup_{t\in[0,1]}\xi_{nr}(t) + \max_{-\omega_n\le r\le\omega_n}\sup_{t\in[0,1]}E[\tilde{G}^+_{n,\epsilon\eta}(s+u_r, s+u_{r+1}, t)].$$

For all r, we have

$$E[\tilde{G}^+_{n,\epsilon\eta}(s+u_r, s+u_{r+1}, t)] \le Q_n E[\sup_{t\in[0,1]}|\eta_{1,k}(t)|]\frac{1}{M}\sum_{m=1}^{M}E[\mathbf{1}(s_m \in [s+u_r, s+u_{r+1})] \le C_1 a_n,$$

where C_1 is a given scalar.

For any $B > 0$, we have

$$P\{\sup_{|v|\le 2h_{2j}}\sup_{t\in[0,1]}\tilde{V}_{n,\epsilon\eta}(s, s+v, t) \ge Ba_n\} \le P\{\max_{-\omega_n\le r\le\omega_n}\sup_{t\in[0,1]}\xi_{nr}(t) \ge (B-C_1)a_n\}.$$

For each fixed s, we consider the following function class:

$$\mathcal{X}_j = \{g_j(\boldsymbol{\epsilon}, \boldsymbol{\eta}, t) = \frac{1}{M\pi(s)} \sum_{m=1}^{M} [\epsilon_j(s_m)\eta_j(t)]^{+}\mathbf{1}(s+u_r \leq s_m \leq s+u_{r+1})\mathbf{1}(|\epsilon_j(s_m)| \leq Q_n) : t \in [0,1]\}.$$

It follows from Assumption (C9a) that $\mathcal{X}_j$ is a pointwise measurable class of functions and $||g_j||_{[0,1]} \leq C_2 Q_n$, where $C_2 > 0$ is a given scalar. It follows from Talagrand's inequality [12, 3] that for suitable finite constants $A_1, A_2 > 0$

$$\begin{aligned} &P\{||\sum_{i=1}^{n}[g_j(\boldsymbol{\epsilon}, \boldsymbol{\eta}, t) - E[g_j(\boldsymbol{\epsilon}, \boldsymbol{\eta}, t)]||_{[0,1]} \geq A_1(E[||\sum_{i=1}^{n}\tau_i g_j(\boldsymbol{\epsilon}, \boldsymbol{\eta}, t)||_{\mathcal{X}_j}] + t)\} \\ &\leq 2[\exp(-A_2 t^2/(n\sigma^2_{\mathcal{X}_j})) + \exp(-A_2 t/(C_2 Q_n))], \end{aligned} \tag{48}$$

where $\tau_1, \ldots, \tau_n$ is a sequence of independent Rademacher random variables and

$$\sigma^2_{\mathcal{X}_j} = \sup_{t\in[0,1]} \mathrm{Var}(g_j(\boldsymbol{\epsilon}, \boldsymbol{\eta}, t)) \leq C_3\beta_n,$$

in which C_3 is a positive scalar. It follows from Assumption (C4) that $\mathcal{X}_j$ is a pointwise measurable Vapnik and Cervonenkis (VC) class [13]. By using Proposition A.1 of Einmahl and Mason [3], we can show that for a scalar $A_3 > 0$,

$$E[||\sum_{i=1}^{n}\tau_i g_j(\boldsymbol{\epsilon}, \boldsymbol{\eta}, t)||_{\mathcal{X}_j}] \leq A_3\sqrt{n\beta_n|\log\beta_n|} = A_3 a_n\sqrt{|\log\beta_n|/\log(n)}.$$

By substituting $t = A_4 n a_n$ into (48), we have

$$\begin{aligned} &P\{||\sum_{i=1}^{n}[g_j(\boldsymbol{\epsilon}, \boldsymbol{\eta}, t) - E[g_j(\boldsymbol{\epsilon}, \boldsymbol{\eta}, t)]||_{[0,1]} \geq A_1(A_3 n^{-1}\sqrt{|\log\beta_n|/\log(n)} + A_4) n a_n\} \\ \leq\ & 2[\exp(-A_2 A_4^2 n a_n^2/\beta_n) + \exp(-A_2 A_4 n a_n^2/\beta_n)] = 2n^{-A_2 A_4^2} + 2n^{-A_2 A_4}. \end{aligned}$$

Therefore, by using Bool's inequality, for relatively large $B > 0$, we have

$$P\{\max_{-\omega_n \leq r \leq \omega_n} \sup_{t\in[0,1]} \xi_{nr}(t) \geq (B - C_1)a_n\} \leq n^{-B^*},$$

in which B^* can be chosen to be sufficiently large if we choose a large B, and

$$P\{\sup_{|v|\leq 2h_{2j}} \sup_{t\in[0,1]} \tilde{V}_{n,\epsilon\eta}(s, s+v, t) \geq B a_n\} \leq A_5 \frac{Q_n}{a_n} n^{-B^*} = A_5 n^{-B^*+1}/\log n.$$

Therefore,

$$\sup_{|v|\leq 2h_{2j}} \sup_{t\in[0,1]} \tilde{V}_{n,\epsilon\eta}(s, s+v, t) = O_p(a_n). \tag{49}$$

Hence, (39) follows from combining (44)-(49).

Lemma 7. Under Assumptions (C1)-(C9a), we have

$$\sup_{(s,t)} n^{-1}|\sum_{i=1}^{n}\bar{\epsilon}_{ij}(s)\bar{\epsilon}_{ij}(t)| = O((Mh_{2j})^{-1} + (\log n/n)^{1/2}) = o_p(1).$$

Proof. We define $\Delta_{n,\epsilon\epsilon}(s,t)=\sum_{i=1}^{n}\bar{\epsilon}_{ij}(s)\bar{\epsilon}_{ij}(t)$,

$$\Delta^{(1)}_{n,\epsilon\epsilon}(s,t)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{M^2\pi(s)\pi(t)}\sum_{m=1}^{M}K_{h_{2j}}(s_m-s)K_{h_{2j}}(s_m-t)\epsilon_{ij}(s_m)^2$$

$$\Delta^{(2)}_{n,\epsilon\epsilon}(s,t)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{M^2\pi(s)\pi(t)}\sum_{m\neq m'}K_{h_{2j}}(s_m-s)K_{h_{2j}}(s_{m'}-t)\epsilon_{ij}(s_m)\epsilon_{ij}(s_{j'})$$

$$\begin{aligned}&V_{n,\epsilon\epsilon}(s,v_1,t,v_2)\\ =\ &|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{M^2\pi(s)\pi(t)}\sum_{m,m'=1}^{M}\epsilon_{ij}(s_m)\epsilon_{ij}(s_{m'})\mathbf{1}(s\le s_m\le s+v_1)\mathbf{1}(t\le s_{m'}\le t+v_2)|.\end{aligned}$$

It follows from Lemma 2 that for large enough n, there exists a constant $C_1>1$ such that

$$\begin{aligned}&\sup_{(s,t)\in[0,1]^2}n^{-1}|\Delta_{n,\epsilon\epsilon}(s,t)|\le\\ &C_1\sup_{(s,t)\in[0,1]^2}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{M^2\pi(s)\pi(t)}\sum_{m,m'=1}^{M}K_{h_{2j}}(s_m-s)K_{h_{2j}}(s_{m'}-s)\epsilon_{ij}(s_m)\epsilon_{ij}(s_{m'})\right|\\ &\le C_1\{\sup_{(s,t)\in[0,1]^2}\left|\Delta^{(1)}_{n,\epsilon\epsilon}(s,t)\right|+\sup_{(s,t)\in[0,1]^2}\left|\Delta^{(2)}_{n,\epsilon\epsilon}(s,t)\right|\}.\end{aligned}$$

Similar to the arguments in Lemmas 3 and 4 of Li and Hsing [11], we have

$$\sup_{(s,t)\in[0,1]^2}\left|\Delta^{(2)}_{n,\epsilon\epsilon}(s,t)\right|=O(\sqrt{\log n/n})\ \text{ a.s.}$$

We need to consider $\sup_{(s,t)\in[0,1]^2}\left|\Delta^{(1)}_{n,\epsilon\epsilon}(s,t)\right|$. It follows from assumption C4 and by some calculations that

$$\sup_{(s,t)\in[0,1]^2}\left|\Delta^{(1)}_{n,\epsilon\epsilon}(s,t)\right|=O_p((Mh_{2j})^{-1}+(\log n/n)^{1/2}).$$

This completes the proof.

Proof of Theorem 3. Recall that $\hat{\eta}_{ij}(s)=\eta_{ij}(s)+\Delta_{i,j}(s)$, we have

$$\begin{aligned}n^{-1}\sum_{i=1}^{n}\hat{\eta}_{ij}(s)\hat{\eta}_{ij}(t)\ &=\ n^{-1}\sum_{i=1}^{n}\Delta_{ij}(s)\Delta_{ij}(t)+n^{-1}\sum_{i=1}^{n}\eta_{ij}(s)\Delta_{ij}(t)\\ &\quad+n^{-1}\sum_{i=1}^{n}\Delta_{ij}(s)\eta_{ij}(t)+n^{-1}\sum_{i=1}^{n}\eta_{ij}(s)\eta_{ij}(t).\end{aligned}\tag{50}$$

This proof consists of two steps. The first step is to show that the first three terms on the right hand side of (50) converge to zero uniformly for all $(s,t)\in[0,1]^2$ in probability. The second step is to show the uniform convergence of $n^{-1}\sum_{i=1}^{n}\eta_{ij}(s)\eta_{ij}(t)$ to $\Sigma_\eta(s,t)$ over $(s,t)\in[0,1]^2$ in probability.

We first show that

$$\sup_{(s,t)}n^{-1}|\sum_{i=1}^{n}\Delta_{ij}(s)\eta_{ij}(t)|=O_p(n^{-1/2}+h_{1j}^2+h_{2j}^2+(\log n/n)^{1/2}).\tag{51}$$

Since

$$\sum_{i=1}^{n}\Delta_{ij}(s)\eta_{ij}(t) \le n^{-1}\{|\sum_{i=1}^{n}\overline{\epsilon}_{ij}(s)\eta_{ij}(t)|+|\sum_{i=1}^{n}\Delta\eta_{ij}(s)\eta_{ij}(t)|+|\sum_{i=1}^{n}\mathbf{x}_i^T\Delta B_j(s)\eta_{ij}(t)|\}, \tag{52}$$

it is sufficient to focus on the three terms on the right-hand side of (52). Since

$$|\mathbf{x}_i^T\Delta B_j(s)\eta_{ij}(t)| \le ||\mathbf{x}_i||_2 \sup_{s\in[0,1]}||\Delta B_k(s)||_2 \sup_{t\in[0,1]}|\eta_{ij}(t)|,$$

we have

$$n^{-1}|\sum_{i=1}^{n}\mathbf{x}_i^T\Delta B_j(s)\eta_{ij}(t)| \le \sup_{s\in[0,1]}||\Delta B_k(s)||_2 n^{-1}\sum_{i=1}^{n}||\mathbf{x}_i||_2|\eta_{ij}(t)| = O_p(n^{-1/2}+h_{1j}^2).$$

Similarly, we have

$$n^{-1}|\sum_{i=1}^{n}\Delta\eta_{ij}(s)\eta_{ij}(t)| \le n^{-1}\sum_{i=1}^{n}\sup_{s,t\in[0,1]}|\Delta\eta_{ij}(s)\eta_{ij}(t)| = O_p(h_{1j}^{(2)2}) = o_p(1).$$

It follows from Lemma 6 that $\sup_{(s,t)} n^{-1}\{|\sum_{i=1}^n \overline{\epsilon}_{ij}(s)\eta_{ij}(t)| = O((\log n/n)^{1/2})$. Similarly, we can show that $\sup_{(s,t)} n^{-1}|\sum_{i=1}^n \Delta_{ij}(t)\eta_{ij}(s)| = O_p(n^{-1/2}+h_{1j}^2+h_{2j}^2+(\log n/n)^{1/2})$.

We can show that

$$\sup_{(s,t)}|n^{-1}\sum_{i=1}^{n}[\eta_{ij}(s)\eta_{ij}(t)-\Sigma_{\eta,jj}(s,t)]| = O_p(n^{-1/2}). \tag{53}$$

Note that

$$\begin{aligned}&|\eta_{ij}(s_1)\eta_{ij}(t_1)-\eta_{ij}(s_2)\eta_{ij}(t_2)|\\ &\le 2(|s_1-s_2|+|t_1-t_2|)\sup_{s\in[0,1]}|\dot{\eta}_{ij}(s)|\sup_{s\in[0,1]}|\eta_{ij}(s)|\end{aligned}$$

holds for any (s_1,t_1) and (s_2,t_2), the functional class $\{\eta_j(u)\eta_j(v):(u,v)\in[0,1]^2\}$ is a Vapnik and Cervonenkis (VC) class [13, 9]. Thus, it yields that (53) is true.

Finally, we can show that

$$\sup_{(s,t)} n^{-1}|\sum_{i=1}^{n}\Delta_{ij}(s)\Delta_{ij}(t)| = O_p((Mh_{2j})^{-1}+(\log n/n)^{1/2}+h_j^4+h_{1j}^{(2)4}). \tag{54}$$

With some calculations, we have

$$\begin{aligned}|\sum_{i=1}^{n}\Delta_{ij}(s)\Delta_{ij}(t)| \le{}& C_1\sup_{(s,t)}[|\sum_{i=1}^{n}\overline{\epsilon}_{ij}(s)\overline{\epsilon}_{ij}(t)|+|\sum_{i=1}^{n}\overline{\epsilon}_{ij}(s)\Delta\eta_{ij}(t)|+|\sum_{i=1}^{n}\Delta\eta_{ij}(t)\mathbf{x}_i^T\Delta B_j(s)|\\ &+|\sum_{i=1}^{n}\overline{\epsilon}_{ij}(s)\mathbf{x}_i^T\Delta B_j(t)|+|\sum_{i=1}^{n}\Delta\eta_{ij}(s)\Delta\eta_{ij}(t)|+|\sum_{i=1}^{n}\mathbf{x}_i^T\Delta B_j(s)\Delta B_j(t)\mathbf{x}_i|],\end{aligned} \tag{55}$$

for a positive constant C_1.

It follows from Lemma 7 that

$$\sup_{(s,t)} n^{-1}|\sum_{i=1}^{n}\overline{\epsilon}_{ij}(s)\overline{\epsilon}_{ij}(t)| = O_p((Mh_{2j})^{-1}+(\log n/n)^{1/2}),$$

$$\sup_{(s,t)} n^{-1}[|\sum_{i=1}^{n}\overline{\epsilon}_{ij}(s)\Delta\eta_{ij}(t)|+|\sum_{i=1}^{n}\Delta\eta_{ij}(t)\mathbf{x}_i^T\Delta B_j(s)|+|\sum_{i=1}^{n}\overline{\epsilon}_{ij}(s)\mathbf{x}_i^T\Delta B_j(t)|] = O_p((\log n/n)^{1/2}).$$

Since

$$\sup_{s\in[0,1]} |\Delta\eta_{ij}(s)| = C_2 \sup_{s\in[0,1]} |\ddot{\eta}_{ij}(s)| h_{2j}^2,$$

we have $\sup_{(s,t)} n^{-1} |\sum_{i=1}^n \Delta\eta_{ij}(s)\Delta\eta_{ij}(t)| = O(h_{1j}^{(2)4})$. Furthermore, since $\sup_{s\in[0,1]} ||\Delta\mathbf{B}(s)|| = O_p(n^{-1/2} + h_j^2)$, we have

$$n^{-1} |\sum_{i=1}^n \mathbf{x}_i^T \Delta B_j(s)\Delta B_j(t)\mathbf{x}_i| = O_p(n^{-1} + h_j^4).$$

Note that the arguments for (51)-(54) hold for $\hat{\Sigma}_{\eta,jj'}(\cdot,\cdot)$ for any $j \neq j'$. Thus, combining (51)-(54) leads to Theorem 3 (i).

To prove Theorem 3 (ii), we follow the same arguments in Lemma 6 of Li and Hsing [11]. For completion, we highlight several key steps below. We define

$$(\Delta\psi_{j,j})(s) = \int_0^1 [\hat{\Sigma}_{\eta,jj}(s,t) - \Sigma_{\eta,jj}(s,t)]\psi_{j,j}(t)dt. \tag{56}$$

Following Hall and Hosseini-Nasab [5] and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
& \{\int_0^1 [\hat{\psi}_{j,j}(s) - \psi_{j,j}(s)]^2 ds\}^{1/2} \\
\leq \;& C_2\{[\int_0^1 (\Delta\psi_{j,j})(s)^2 ds]^{1/2} + \int_0^1\int_0^1 [\hat{\Sigma}_{\eta,jj}(s,t) - \Sigma_{\eta,jj}(s,t)]^2 dsdt\} \\
\leq \;& C_2\{\int_0^1\int_0^1 [\hat{\Sigma}_{\eta,jj}(s,t) - \Sigma_{\eta,jj}(s,t)]^2 dsdt\}^{1/2}\{\int_0^1 [\psi_{j,j}(t)]^2 dt\}^{1/2} \\
& + \int_0^1\int_0^1 [\hat{\Sigma}_{\eta,jj}(s,t) - \Sigma_{\eta,jj}(s,t)]^2 dsdt \\
\leq \;& C_3 \sup_{(s,t)\in[0,1]^2} |\hat{\Sigma}_{\eta,jj}(s,t) - \Sigma_{\eta,jj}(s,t)|,
\end{aligned}$$

which yields Theorem 3 (ii.a).

Using (4.9) in Hall, Müller and Wang [6], we have

$$\begin{aligned}
& |\hat{\lambda}_{j,j} - \lambda_{j,j}| \leq |\int_0^1\int_0^1 [\hat{\Sigma}_{\eta,jj} - \Sigma_{\eta,jj}](s,t)\psi_{j,j}(s)\psi_{j,j}(t)dsdt + O(\int_0^1 (\Delta\psi_{j,j})(s)^2 ds) \\
\leq \;& C_4 \sup_{(s,t)\in[0,1]^2} |\hat{\Sigma}_{\eta,jj}(s,t) - \Sigma_{\eta,jj}(s,t)|,
\end{aligned}$$

which yields Theorem 3 (ii.b). This completes the proof.

We need to introduce some notation to establish the weak convergence of a sequence of stochastic processes indexed by $s \in [0,1]$ [13]. The uniform metric is used here to define the weak convergence. Let $\ell^\infty([0,1])$ be the space of all uniformly bounded, real functions on $[0,1]$, and endow $\ell^\infty([0,1])$ with the uniform metric. We consider $\mathcal{B}L_1(\ell^\infty([0,1]))$ to be the space of real-valued functions on $\ell^\infty([0,1])$ with Lipschitz norm bounded by 1; that is, for any $k(\cdot) \in \mathcal{B}L_1(\ell^\infty([0,1]))$, $\sup_{x(s)\in\ell^\infty([0,1])} |k(x)| \leq 1$ and $|k(x) - k(y)| \leq ||x-y||_{[0,1]}$. As $n \to \infty$, a stochastic process $G_j(\cdot)$ weakly converges to $X(\cdot)$ on $\ell^\infty([0,1])$ if and only if $\sup_{k\in\mathcal{B}L_1(\ell^\infty([0,1]))} |Ek(G_j) - Ek(X)| \to 0$.

Proof of Theorem 5. We define $\mathbf{r}_{ij}(s) = y_{ij}(s) - \mathbf{x}_i^T B_j(s)$ and

$$\tilde{G}_j(s)^{(g)} = \sqrt{n}[I_p \otimes (1,0)]\mathrm{vec}(\Sigma(s,h_{1j})^{-1} \sum_{i=1}^n \tau_i^{(g)} \sum_{m=1}^M \mathbf{x}_i \otimes H_{h_{1j}}(s_m - s)\mathbf{r}_{ij}(s_m)). \tag{57}$$

Following the arguments in Kosorok [8] and Zhu and Zhang [16], we will prove Theorem 5 in three steps. In Step 1, we will prove the unconditional weak convergence of $\tilde{G}_j(s)^{(g)}$. In Step 2, we will prove the weak convergence of $\tilde{G}_j(s)^{(g)}$ conditional on the data. In Step 3, we will prove the weak convergence of $G_j(s)^{(g)}$ conditional on the data by showing that $\tilde{G}_j(s)^{(g)}$ and $G_j(s)^{(g)}$ are asymptotically equivalent as $n \to \infty$.

In Step 1, we note that $\mathbf{r}_{ij}(s_m) = \eta_{ij}(s_m) + \epsilon_{ij}(s_m)$ and

$$\tilde{G}_j(s)^{(g)} = \sqrt{n}[I_p \otimes (1,0)]\mathrm{vec}(\Sigma(s,h_{1j})^{-1} \sum_{i=1}^n \tau_i^{(g)} \mathbf{x}_i \otimes \sum_{m=1}^M H_{h_{1j}}(s_m - s)[\eta_{ij}(s_m) + \epsilon_{ij}(s_m)]).$$

Therefore, by treating $\tau_i^{(g)}\mathbf{x}_i$ as the new 'covariate' vector, we can apply the same arguments in the proof of Theorem 1 to prove that $\tilde{G}_j^{(g)}$ converges to G_j in distribution; that is, $\tilde{G}_j^{(g)}$ is asymptotically measurable.

In Step 2, we define

$$\begin{aligned}
S_j(s,t) &= n^{-1}n_G^{-2} \sum_{i=1}^n \mathbf{x}_i^{\otimes 2} \otimes \sum_{m,m'=1}^M H_{h_{1j}}(s_m - s)H_{h_{1j}}(s_{m'} - t)\mathbf{r}_{ij}(s_m)\mathbf{r}_{ij}(s_{m'}),\\
S_{j,\eta\eta}(s,t) &= n^{-1}n_G^{-2} \sum_{i=1}^n \mathbf{x}_i^{\otimes 2} \otimes \sum_{m,m'=1}^M H_{h_{1j}}(s_m - s)H_{h_{1j}}(s_{m'} - t)\eta_{ij}(s_m)\eta_{ij}(s_{m'}),\\
S_{j,\eta\epsilon}(s,t) &= n^{-1}n_G^{-2} \sum_{i=1}^n \mathbf{x}_i^{\otimes 2} \otimes \sum_{m,m'=1}^M H_{h_{1j}}(s_m - s)H_{h_{1j}}(s_{m'} - t)\eta_{ij}(s_m)\epsilon_{ij}(s_{m'}),\\
S_{j,\epsilon\epsilon}(s,t) &= n^{-1}n_G^{-2} \sum_{i=1}^n \mathbf{x}_i^{\otimes 2} \otimes \sum_{m,m'=1}^M H_{h_{1j}}(s_m - s)H_{h_{1j}}(s_{m'} - t)\epsilon_{ij}(s_m)\epsilon_{ij}(s_{m'}).
\end{aligned}$$

Thus, conditioning on the data, $\tilde{G}_j(s)^{(g)}$ is a normal random vector with zero mean and covariance given by $(nM)^{-2}[I_p \otimes (1,0)]\Sigma(s,h_{1j})^{-1}S_j(s,t)\Sigma(s,h_{1j})^{-1}[I_p \otimes (1,0)^T]$. It is easy to see that

$$S_j(s,t) = S_{j,\eta\eta}(s,t) + S_{j,\eta\epsilon}(s,t) + S_{j,\eta\epsilon}(t,s) + S_{j,\epsilon\epsilon}(s,t). \tag{58}$$

Following the arguments of Lemmas 6 and 7, we can show that $S_{j,\eta\epsilon}(s,t) + S_{j,\eta\epsilon}(t,s) + S_{j,\epsilon\epsilon}(s,t) = o(1)$. Furthermore, it can be shown that $E[S_{j,\eta\eta}(s,t)] = \Omega_X \otimes \mathrm{diag}(1,0)\Sigma_{\eta,jj}(s,t) + O(h_{1j})$ and $\mathrm{Cov}[S_{j,\eta\eta}(s,t)] = O(n^{-1})$. Therefore, $\mathrm{Cov}_\tau[\tilde{G}_j(s)^{(g)}, \tilde{G}_j(t)^{(g)}]$ converges to $\Sigma_{\eta,jj}(s,t)\Omega_X^{-1}$ in probability, where the expectation is taken with respect to $\tau_i^{(g)}$ conditioning on the data. We can obtain the marginal convergence of $\tilde{G}_j(s)^{(g)}$ in the conditional central limit theorem by using the Cramer-Wald method.

For each $\delta > 0$, let $\tilde{\mathbf{S}}_\delta = \{l\delta : l = 0, \ldots, \delta^{-1}\}$ be an equally $\delta-$spaced grid and $[0,1]_\delta(s)$ assign to each $s \in [0,1]$ a closest element of $\tilde{\mathbf{S}}_\delta$. The finite convergence results yield

$$\sup_{k(\cdot)\in \mathcal{B}L_1(\ell^\infty([0,1]))} |E_\tau k(\tilde{G}_j^{(g)}([0,1]_\delta)) - Ek(G_j([0,1]_\delta))| \to 0$$

in probability, as $n \to \infty$. Due to the continuity of $G_j(s)$, we have $G_j([0,1]_\delta(s)) \to G_j(s)$ almost surely as $\delta \to 0$; that is $\lim_{\delta\to 0}\sup_{k(\cdot)\in\mathcal{B}L_1(\ell^\infty([0,1]))}|E_\tau k(\tilde{G}_j([0,1]_\delta) - E_\tau k(G_j([0,1]))| = 0$. Finally, we have

$$\sup_{k(\cdot)\in\mathcal{B}L_1(\ell^\infty([0,1]))}|E_\tau k(\tilde{G}_j^{(g)}([0,1]_\delta(\cdot))) - E_\tau k(\tilde{G}_j^{(g)}([0,1](\cdot)))| \le E_\tau(\sup_{|s-s'|_2\le\delta}|\tilde{G}_j^{(g)}(s) - \tilde{G}_j^{(g)}(s')|).$$

Thus, the expectation on the left side of the above equation is smaller than $E(\sup_{|s-s'|_2\le\delta}|\tilde{G}_j^{(g)}(s) - \tilde{G}_j^{(g)}(s')|)$, which was established by the unconditional weak convergence of $\tilde{G}_j^{(g)}(\cdot)$ in Step 1. This finishes the proof of Step 2.

In Step 3, following the arguments in Theorem 3 of Kosorok [8], we only need to prove that

$$\Delta_{n,B} = \sup_{s\in[0,1]} n^{-1}\sum_{i=1}^n \mathrm{tr}\{\mathbf{x}_i^{\otimes 2}\otimes\{M^{-1}\sum_{m=1}^M H_{h_{1j}}(s_m - s)\mathbf{x}_i^T[\hat{B}_j(s_m) - B_j(s_m)]\}^{\otimes 2}\} = o_p(1).$$

It follows from the proof of Theorem 3 that $\Delta_{n,B} = O_p(n^{-1} + \hat{h}_{1j}^4)$, which converges to zero in probability. This finishes the proof of Theorem 5.

We prove several key results for ensuring that Theorems 1 and 2 are valid for the case with fixed grid points.

Lemma 8 . Under Assumptions (C1), (C3), (C4b), (C5), and (C7b), we have the following results:

$$n^{-1/2}\sum_{i=1}^n \mathbf{x}_i\otimes\{M^{-1}\sum_{m=1}^M K_{h_{1j}}(s_m - s)\mathbf{z}_{h_{1j}}(s_m - s)\epsilon_{i,j}(s_m)\} = o_p(1), \tag{59}$$

$$\sup_{s\in[0,1]}\left|\int K_{h_{1j}}(u-s)\frac{(u-s)^r}{h_{1j}^r}d[\Pi_M(u) - \Pi(u)]\right| = O_p((Mh_{1j})^{-1}), \tag{60}$$

$$\sup_{s\in[0,1]}|\Pi_M(s) - \Pi(s)| = O_p(M^{-1}). \tag{61}$$

Proof of Lemma 8. To prove (59), we focus on the first component of $\mathbf{z}_{h_{1j}}(s_m - s)$ for the sake of space. We introduce some notation as follows:

$$\begin{aligned}
D_{\epsilon,j}(s,h) &= \sum_{i=1}^n \mathbf{x}_i \sum_{m=1}^M M^{-1}K_{h_{1j}}(s_m - s)\epsilon_{ij}(s_m),\\
F_L^{\gamma_n}(\boldsymbol{\epsilon}_{ij}, s, h) &= (Mh_{1j})^{-1}\sum_{m=1}^M \epsilon_{ij}(s_m)\mathbf{1}(|s_m - s|\le h)\mathbf{1}(|\epsilon_{ij}(s_m)|\le\gamma_n),\\
F_U^{\gamma_n}(\boldsymbol{\epsilon}_{ij}, s, h) &= (Mh_{1j})^{-1}\sum_{m=1}^M \epsilon_{ij}(s_m)\mathbf{1}(|s_m - s|\le h)\mathbf{1}(|\epsilon_{ij}(s_m)|>\gamma_n),\\
F_\epsilon^\lambda(\boldsymbol{\epsilon}_{ij}) &= M^{-1}\sum_{m=1}^M|\epsilon_{ij}(s_m)|^\lambda, \quad G_{\epsilon,j}^{\gamma_n}(s,h) = \sum_{i=1}^n \mathbf{x}_i F_L^{\gamma_n}(\boldsymbol{\epsilon}_{ij}, s, h),\\
\epsilon_{ij}(s_m,\gamma_n) &= \epsilon_{ij}(s_m)\mathbf{1}(|\epsilon_{ij}(s_m)|\le\gamma_n), \quad \mu_{j\epsilon}(s_m,\gamma_n) = E[\epsilon_{ij}(s_m)\mathbf{1}(|\epsilon_{ij}(s_m)|\le\gamma_n)],\\
E[G_{\epsilon,j}^{\gamma_n}(s,h)] &= n\mu_x(Mh_{1j})^{-1}\sum_{m=1}^M \mathbf{1}(|s_m - s|\le h)\mu_{j\epsilon}(s_m,\gamma_n),
\end{aligned}$$

where $\boldsymbol{\epsilon}_{ij} = (\epsilon_{ij}(s_1),\cdots,\epsilon_{ij}(s_m))^T$. It follows from Lemma 2 of Li and Hsing [11] that

$$\sup_{s\in[0,1]}|D_{\epsilon,j}(s,h)| \le C_1\sup_{s\in[0,1]}|G_{\epsilon,j}^\infty(s,h)|. \tag{62}$$

We will prove that $n^{-1/2}\sup_{s\in[0,1]}|G^{\infty}_{\epsilon,j}(s,h_{1j})|=o_p(1)$ by using two steps. In Step 1, for a certain sequence of $\gamma_n\to\infty$, we will show that

$$n^{-1/2}\sup_{s\in[0,1]}|G^{\infty}_{\epsilon,j}(s,h_{1j})-G^{\gamma_n}_{\epsilon,j}(s,h_{1j})+E[G^{\gamma_n}_{\epsilon,j}(s,h_{1j})]|=o_p(1). \tag{63}$$

In Step 2, we will show that

$$n^{-1/2}\sup_{s\in[0,1]}|G^{\gamma_n}_{\epsilon,j}(s,h_{1j})-E[G^{\gamma_n}_{\epsilon,j}(s,h_{1j})]|=o_p(1). \tag{64}$$

We first prove (63). It follows from the definition of $G^{\gamma_n}_{\epsilon,j}(s,h)$ and Assumption (C7b) that

$$\begin{aligned}
&n^{-1/2}\sup_{s\in[0,1]}|G^{\infty}_{\epsilon,j}(s,h_{1j})-G^{\gamma_n}_{\epsilon,j}(s,h_{1j})|=|n^{-1/2}\sum_{i=1}^n\mathbf{x}_iF_U^{\gamma_n}(\boldsymbol{\epsilon}_{ij},s,h_{1j})|\\
&\le n^{1/2}\gamma_n^{1-\lambda}n^{-1}\sum_{i=1}^n||\mathbf{x}_i||_2(Mh_{1j})^{-1}\sup_{s\in[0,1]}\sum_{m=1}^M|\epsilon_{ij}(s_m)|^{\lambda}\mathbf{1}(|s_m-s|\le h_{1j})\mathbf{1}(|\epsilon_{ij}(s_m)|>\gamma_n)\\
&\le n^{1/2}\gamma_n^{1-\lambda}h_{1j}^{-1}n^{-1}\sum_{i=1}^n||\mathbf{x}_i||_2F_\epsilon^{\lambda}(\boldsymbol{\epsilon}_{ij})=n^{1/2}\gamma_n^{1-\lambda}h_{1j}^{-1}O_p(1)=o_p(1).
\end{aligned}$$

Similarly, we can prove that $n^{-1/2}\sup_{s\in[0,1]}|E[G^{\gamma_n}_{\epsilon,j}(s,h_{1j})]|=o_p(1)$.

We prove (64) as follows. It is easy to show that $n^{-1/2}\{G^{\gamma_n}_{\epsilon,j}(s,h_{1j})-E[G^{\gamma_n}_{\epsilon,j}(s,h_{1j})]\}$ can be written as the sum of two terms given by

$$\begin{aligned}
I_1(s)&=\{n^{-1/2}\sum_{i=1}^n(\mathbf{x}_i-\mu_x)\}\{(Mh_{1j})^{-1}\sum_{m=1}^M\mu_{j\epsilon}(s_m,\gamma_n)\mathbf{1}(|s_m-s|\le h)\},\\
I_2(s)&=n^{-1/2}\sum_{i=1}^n\mathbf{x}_i(Mh_{1j})^{-1}\sum_{m=1}^M[\epsilon_{ij}(s_m)\mathbf{1}(|\epsilon_{ij}(s_m)|\le\gamma_n)-\mu_{j\epsilon}(s_m,\gamma_n)]\mathbf{1}(|s_m-s|\le h).
\end{aligned}$$

We prove that $\sup_{s\in[0,1]}I_1(s)=o_p(1)$ as follows. Since $E[\epsilon_{ij}(s_m)]=0$ and $n^{-1/2}\sum_{i=1}^n(\mathbf{x}_i-\mu_x)=O_p(1)$, it follows from Assumption (C1) that

$$\begin{aligned}
\max_{s_m}|\mu_{j\epsilon}(s_m,\gamma_n)| &= \max_{s_m}|E\{\epsilon_{ij}(s_m)\mathbf{1}(|\epsilon_{ij}(s_m)|>\gamma_n)\}|\\
&\le \max_{s_m}E\{|\epsilon_{ij}(s_m)|\mathbf{1}(|\epsilon_{ij}(s_m)|>\gamma_n)\}|\le\gamma_n^{1-q}\max_{s_m}E\{|\epsilon_{ij}(s_m)|^q\}=o(1),
\end{aligned}$$

which yields $\sup_{s\in[0,1]}I_1(s)=o_p(1)$.

We prove that $\sup_{s\in[0,1]}I_2(s)=o_p(1)$ as follows. We note that

$$||\mathbf{x}_i(Mh_{1j})^{-1}\sum_{m=1}^M[\epsilon_{ij}(s_m)\mathbf{1}(|\epsilon_{ij}(s_m)|\le\gamma_n)-\mu_{j\epsilon}(s_m,\gamma_n)]\mathbf{1}(|s_m-s|\le h)||_\infty\le2\gamma_n||\mathbf{x}_i||_\infty,$$

$$\begin{aligned}
&\mathrm{var}(\sum_{i=1}^n\mathbf{x}_i(Mh_{1j})^{-1}\sum_{m=1}^M\{\epsilon_{ij}(s_m,\gamma_n)-\mu_{j\epsilon}(s_m,\gamma_n)\}\mathbf{1}(|s_m-s|\le h))\\
&=\sum_{i=1}^n\mathrm{var}(\mathbf{x}_i)(Mh_{1j})^{-2}\sum_{m=1}^M E\{\epsilon_{ij}(s_m,\gamma_n)-\mu_{j\epsilon}(s_m,\gamma_n)\}^2\mathbf{1}(|s_m-s|\le h))\\
&=O(n(Mh_{1j})^{-1}).
\end{aligned}$$

Since $s_1 \leq \ldots \leq s_m$ are fixed grid points, the interval $(s-h, s+h)$ covers a consecutive grid point block $\{s_{i_1}, s_{i_1+1}, \ldots, s_{i_1+L(i_1)}\}$ such that

$$s_{i_1-1} < s-h, s_{i_1} \geq s-h, s_{i_1+L(i_1)} \leq s+h, \quad \text{and} \quad s_{i_1+L(i_1)+1} > s+h.$$

Therefore, there are about $O(M)$ number of such consecutive grid point blocks, and thus, $F_\epsilon^\lambda(\boldsymbol{\epsilon}_{ij}, s, h)$ has the same number of values as s varies in $[0,1]$. Without loss of generality, we assume that there are M consecutive grid point blocks. Therefore, it follows from Bernstein's inequality that

$$P(\sup_{s\in[0,1]} |I_2(s)| > x) \leq M \exp(-\frac{x^2}{C_1(Mh_{1j})^{-1} + C_2\gamma_n x/\sqrt{n}}). \tag{65}$$

Thus, a sufficient condition of $P(\sup_{s\in[0,1]} |I_2(s)| > x) = o(1)$ is that for sufficiently large C, we obtain a quadratic equation given by $x^2 = C\log(M)\{C_1(Mh_{1j})^{-1} + C_2\gamma_n x/\sqrt{n}\}$, whose positive solution is given by

$$x_* = 0.5\frac{CC_2\gamma_n \log(M)}{\sqrt{n}} + 0.5\sqrt{\frac{C^2C_2^2\gamma_n^2\log(M)^2}{n} + 4CC_1\frac{\log(M)}{Mh_{1j}}} \tag{66}$$

It follows from Assumption (C7b) that $x_* = o(1)$. By substituting x_* into (65), we have

$$P(\sup_{s\in[0,1]} |I_2(s)| > x_*) \leq M\exp(-C\log(M)) = \exp(-(C-1)\log(M)) = o(1), \tag{67}$$

which yields $\sup_{s\in[0,1]} I_2(s) = o_p(1)$. Thus, we finish the proof of (59).

We prove (60) as follows. Let $s_0 = 0$. It follows from Taylor's expansion and Assumption (C5) that

$$\begin{aligned}
& \int K_{h_{1j}}(u-s)\frac{(u-s)^r}{h_{1j}^r} d[\Pi_M(u) - \Pi(u)] \\
= & \sum_{m=1}^M K_{h_{1j}}(s_m - s)\frac{(s_m-s)^r}{h_{1j}^r}\int_{s_{m-1}}^{s_m}\pi(u)du - \sum_{m=1}^M \int_{s_{m-1}}^{s_m} K_{h_{1j}}(u-s)\frac{(u-s)^r}{h_{1j}^r}\pi(u)du \\
= & \sum_{m=1}^M \int_{s_{m-1}}^{s_m} [K_{h_{1j}}(s_m-s)\frac{(s_m-s)^r}{h_{1j}^r} - K_{h_{1j}}(u-s)\frac{(u-s)^r}{h_{1j}^r}]\pi(u)du = O((Mh_{1j})^{-1}).
\end{aligned}$$

By using the same arguments, we can prove (61) and thus we omit the details.

2. Additional Simulation Results. **Example 2 (continue).** This example is used to evaluate the coverage probabilities of SCB of the accuracy of the estimators of the eigenvalues and eigenfunctions of $\Sigma_\eta(\cdot,\cdot)$ and the estimators of (σ_1^2, σ_2^2). The data were generated from the model used in Example 1 under the same parameter values. We set $n = 500$ and $M = 25$, 50, and 75 and generated 200 datasets for each combination.

We estimated the eigenvalues λ_{11}, λ_{12}, λ_{21}, and λ_{22} and the variances σ_1^2 and σ_2^2 for each simulated data set for $M = 25$, 50 and 75. The accuracy of estimators improves with M. The performance of the estimators for $M = 50$ is almost as good as their performance for $M = 75$. Fig. 1 shows the boxplots of values of $\hat{\lambda}_{jl}$ and $\hat{\sigma}_j^2$ for $j = 1, 2$ and $l = 1, 2$. The estimated eigenvalues and variances should be compared with the true ones, which are $(1.2, 0.6, 1, 0.5, 0.2, 0.1)$. When M is large, the estimated eigenvalues and variances are very close to their true values. We summarized the estimated results on $\psi_{jl}(s)$ for $j = 1, 2$ and $l = 1, 2$ in Fig. 2, in which we plotted the mean and the pointwise 1^{st} and 99^{th} percentiles of the estimated eigenfunctions with the true eigenfunctions. The performance of the estimated eigenfunctions improves with M increasing as expected.

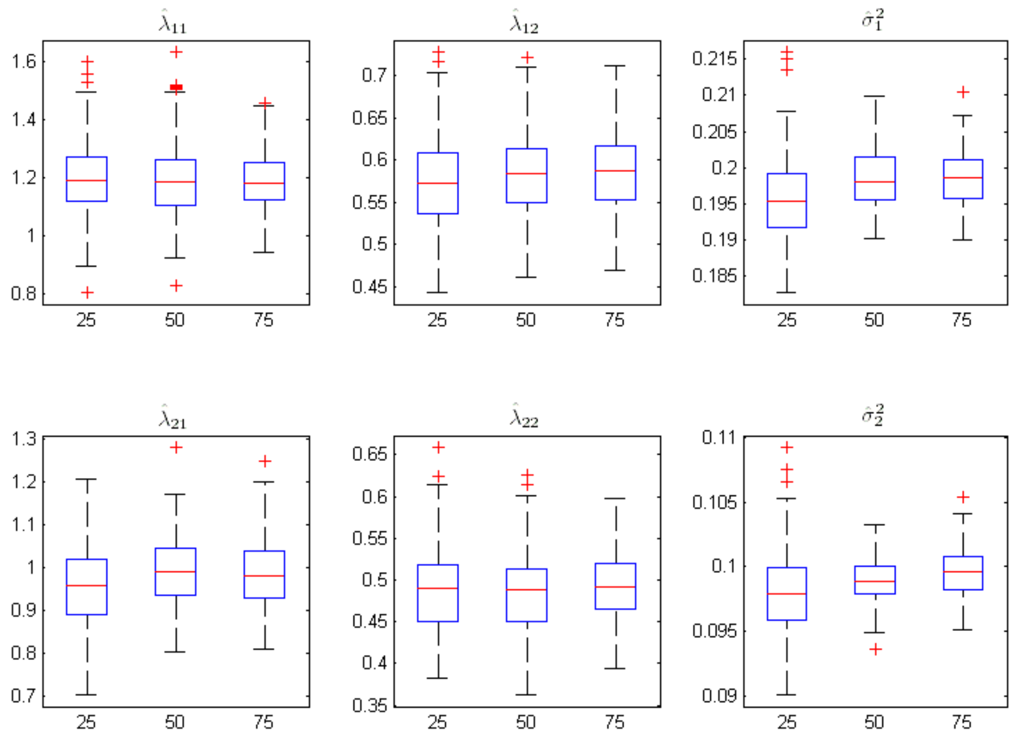

FIG 1. *Boxplot for the eigenvalues $\hat\lambda_{11}$, $\hat\lambda_{12}$, $\hat\lambda_{21}$, and $\hat\lambda_{22}$ and the variances $\hat\sigma_1^2$ and $\hat\sigma_2^2$, when $M = 25$, 50 and 75.*

3. Additional Real Data Example. Attention deficit hyperactivity disorder (ADHD) is one of the most common childhood disorders and can continue through adolescence and adulthood. Symptoms include difficulty staying focused and paying attention, difficulty controlling behavior, and hyperactivity (over-activity). ADHD has three common subtypes including predominantly hyperactive-impulsive, predominantly inattentive, and combined hyperactive-impulsive and inattentive. The resting-state fMRI (rs-fMRI) data set that we used here is part of ADHD-200 Global Competition data sets and was collected from the New York University (NYU) site. The data set consists of 170 subjects (98 normal controls and 72 combined hyperactive-impulsive subjects). Among them, there are 108 males whose mean age is 11.4 years with standard deviation 5.7 years and 62 females whose mean age is 11.9 years with standard deviation 6.0 years. Rs-fMRIs and T1-weighted images were acquired for each subject. For the rs-fMRI, a $T2^*$-weighted EPI sequence was used to acquire images. The imaging parameters were as follows: TR=2sec, TE=32 ms; 33 slices; and voxel size =4x4x4 mm^3. This sequence was repeated 150 times so as to providing time series images. The rs-fMRI data were preprocessed by standard steps including time shifting, motion correction, spatial smoothing (6-mm full width at half maximum Gaussian kernel), linear trend removal, and band-pass filtered with frequency range of 0.009-0.08 Hz. Subsequently, rs-fMRI data from each subject was co-registered to the automated anatomical labeling (AAL) atlas with 116 regions of interest (ROI).

For each individual subject, the average time series were calculated from each ROI and correlated with each other to define a 116×116 correlation matrix. To analyze the efficiency properties of brain functional network, each correlation matrix was thresholded to create an adjacency matrix G with

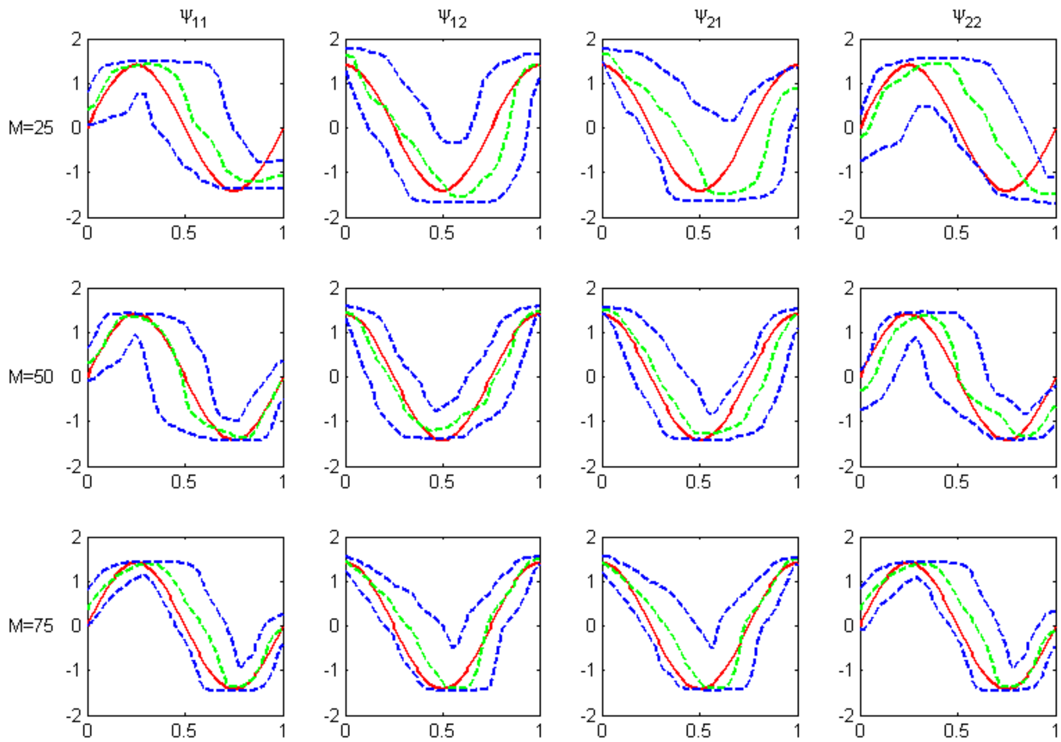

FIG 2. *Plot of eigenfunctions and their pointwise confidence intervals. The red solid line is the true eigenfunction, the middle green dashed line is the pointwise mean of estimated eigenfunctions and other two blue dashed lines are the pointwise 1% and 99% percentiles of the estimated eigenfunctions in* 200 *runs.*

elements of either 1 or 0 depending on whether the corresponding correlation value exceeds the threshold or not. Clearly the choice of threshold have major effects on the topology of the resulting network: conservative thresholds will produce sparsely connected graphs, which might eliminates true connections while more lenient thresholds will generate densely connected graph, which on the other hand might includes spurious connections. As a result, the adoption of any single threshold will inevitably raise the concern of possible bias associated with this unique value. Therefore, each correlation matrix was repeatedly thresholded over a range of significance levels from 0.01 to 1 to avoid such bias. Two connectivity and network complexity measures including local efficiency (LE) (Fig. 3 (c) and (e)) and characteristic path length (CPL) (Fig. 3 (d) and (f)) were calculated for each adjacency matrix G [10, 14, 7, 15, 1]. Finally, for each subject, we obtained two LE and CPL values at 100 evenly spaced grid points between [0, 1].

The aim of this analysis is to characterize the association between (LE, CPL) and covariates of interest including age and diagnosis status. We fitted model (1) to the LE and CPL values from all 170 subjects, in which $\mathbf{x}_i = (1, \mathrm{G}, \mathrm{Age}, \mathrm{D}, \mathrm{G} \times \mathrm{Age}, \mathrm{G} \times \mathrm{D}, \mathrm{Age} \times \mathrm{D})^T$, where G and D, respectively, represent gender and diagnostic status. We then applied the estimation and inference procedures to estimate $\mathbf{B}(s)$ and calculate S_n for each hypothesis test. We approximated the p-value of S_n using the resampling method with $G = 1,000$ replications. Finally, we constructed the 95% simultaneous confidence bands for the functional coefficients of $B_j(s)$ for $j = 1, 2$.

Fig. 4 presents the estimated coefficient functions corresponding to Age, D, and Age×D associ-

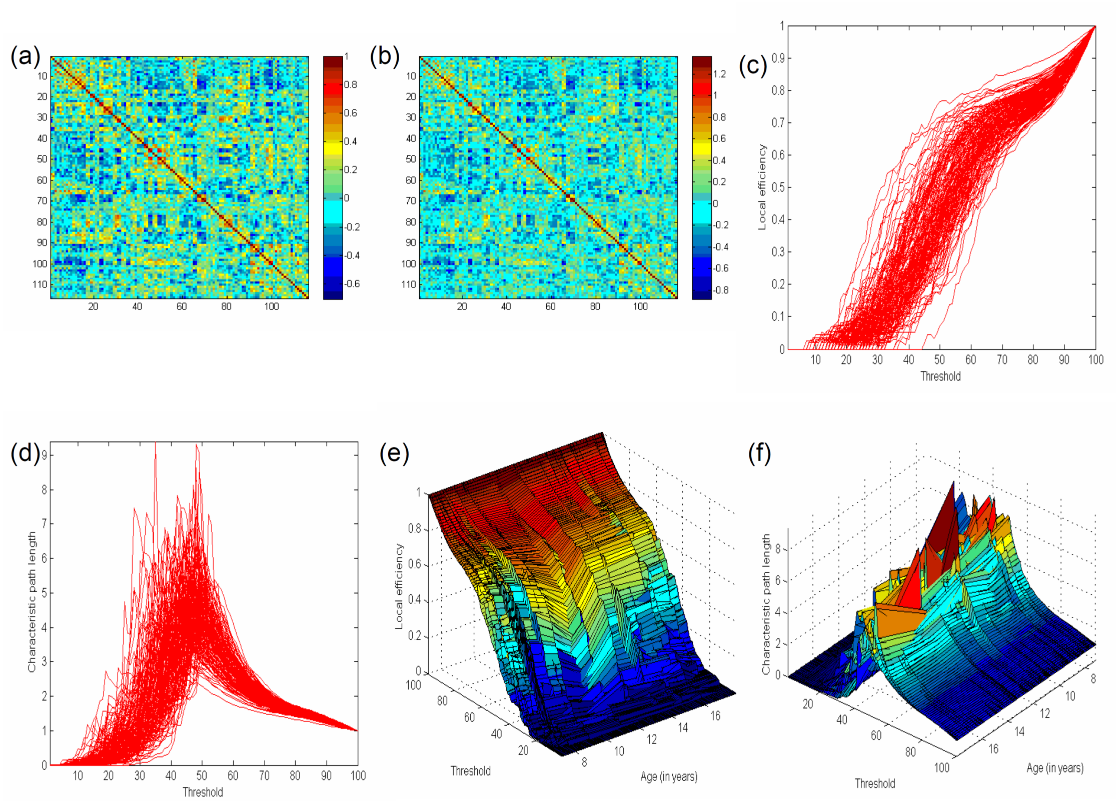

Fig 3. *ADHD data at NYU site: (a) the brain network using correlation matrix of time courses data, (b) the brain network using Fisher's z-transformed correlation matrix of time courses data, (c) local efficiency from 170 children, (d) characteristic path length from 170 children, (e) and (f) the 3D plots with age of local efficiency and characteristic path length from 170 children.*

ated with LE and CPL (blue solid lines in all panels of Fig. 4). The three effects for the two brain network properties are close to zero when the significance levels are either small or large, whereas they are significantly different from zero when the significance levels are moderate. We observe different change patterns in the coefficient functions of Age, D, and Age×D for LE and CPL. For moderate significance levels, the coefficient functions of Age, D, and Age×D for LE are almost consistently either positive or negative, while they swing between positive and negative for CPL. For example, for LE, the coefficient functions of diagnostic group (panel (b) of Fig. 4) are negative at most of the middle grid points, which may indicate that the local efficiency values of children without ADHD are greater than those of children with ADHD. The p-values of the global test for the interaction of age and diagnostic group are smaller than 0.001, indicating that the topological structure of brain network varies significantly across age and diagnostic groups. Furthermore, inspecting the SCBs of the Age×D interaction localizes correlation values around 0.3, where the LE of brain network differs significantly across age and diagnostic groups (Fig. 4 (c)), whereas we obtain important correlation values around 0.55 for the CPL of brain network (Fig. 4 (f)).

Fig. 5 presents the first 10 eigenvalues and 3 eigenfunctions of $\hat{\Sigma}_{\eta,jj}(s,t)$ for $j = 1, 2$. For the two network measures, the relative eigenvalues of $\hat{\Sigma}_{\eta,jj}$ defined as the ratios of the eigenvalues of $\hat{\Sigma}_{\eta,jj}(s,t)$ over their sum have similar distributional patterns (panel (a) of Fig. 5). We observe that the first three eigenvalues account for more than 90% of the total and the others quickly vanish to zero. The eigenfunctions of LE corresponding to the largest three eigenvalues (Fig. 5 (b)) are

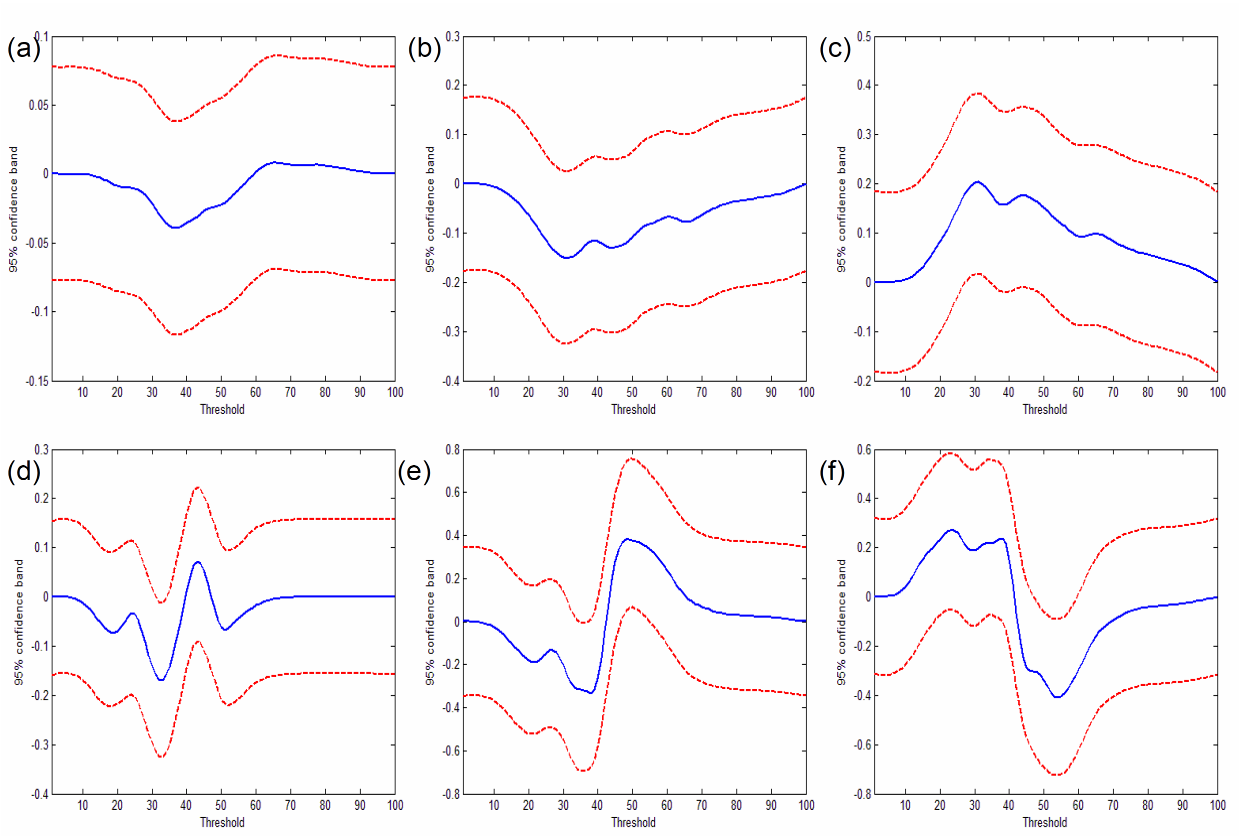

FIG 4. *Plot of estimated effects of age, diagnostic group, and age and diagnostic group interaction (from left to right) and their* 95% *confidence bands. The upper panels are for local efficiency and the lower panels are for characteristic path length. The blue solid curves are the estimated coefficient functions and the red dashed curves are the confidence bands.*

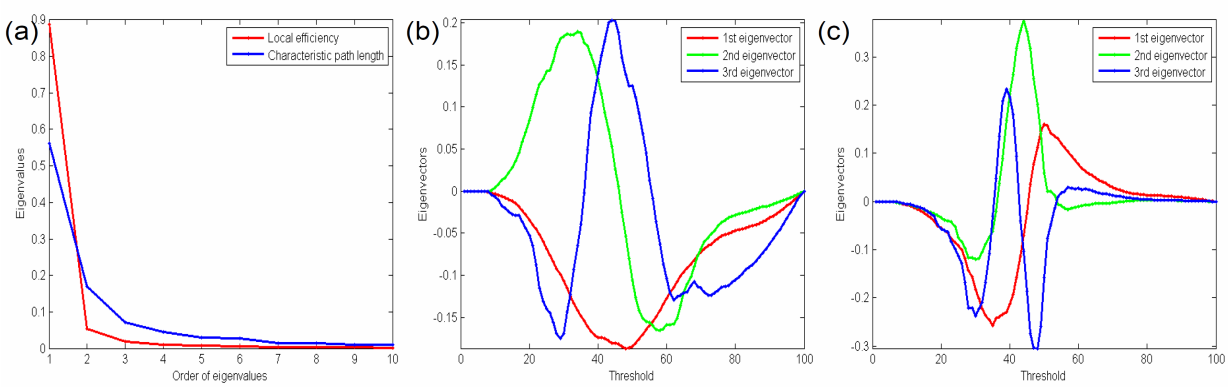

FIG 5. *Plot of the first* 10 *eigenvalues and the first* 3 *eigenfunctions.*

different from those of CPL (Fig. 5 (c)). For instance, for LE, the first eigenfunction is an weighted average at around the 50-th grid point; the second one is a weighted contrast between the 30-th grid point and the 60-th grid point; the third eigenfunction is a contrast between the 50-th grid point and the 30-th grid point together with the 60-th grid point. Here our interpretation of Fig. 5 is largely exploratory. How to statistically compare eigenvalues and eigenfunctions is an interesting topic for future research.

References.

[1] Alexander-Bloch, A., Lambiotte, R., Roberts, B., Giedd, J., Gogtay, N. and Bullmore, E. (2012). The discovery of population differences in network community structure: New methods and applications to brain functional networks in schizophrenia. *NeuroImage* **59** 3889-3900.
[2] DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*. Springer Verlag, New York.
[3] Einmahl, U. and Mason, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *Journal of Theoretical Probability* **13** 1-37.
[4] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.
[5] Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society B* **68** 109-126.
[6] Hall, P., Müller, H.-G. and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34** 1493–1517.
[7] Kaiser, M. (2011). A tutorial in connectome analysis: Topological and spatial features of brain networks. *NeuroImage* **57** 892-907.
[8] Kosorok, M. R. (2003). Bootstraps of sums of independent but not identically distributed stochastic processes. *J. Multivariate Anal.* **84** 299-318.
[9] Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer: New York.
[10] Latora, V. and Marchiori, M. (2001). Efficient behavior of small-world networks. *Phys. Rev. Lett.* **87** 198701.
[11] Li, Y. and Hsing, T. (2010). Uniform Convergence Rates for Nonparametric Regression and Principal Component Analysis in Functional/Longitudinal Data. *The Annals of Statistics* **38** 3321-3351.
[12] Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes,. *Ann. Probab* **22** 28-76.
[13] van der Vaar, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag Inc.
[14] Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature* **393** 440-442.
[15] Zalesky, A., Cocchi, L., Fornito, A., Murray, M. M. and Bullmore, E. (2012). Connectivity differences in brain networks Original Research Article. *NeuroImage* **60** 1055-1062.
[16] Zhu, H. T. and Zhang, H. P. (2006). Generalized score test of homogeneity for mixed effects models. *Annals of Statistics* **34** 1545–1569.