Empirical Likelihood Ratio Tests for Varying Coefficient Geo Models

Shuoyang Wang[a], Honglang Wang[b], Yichuan Zhao[c], Guanqun Cao[a], Yingru Li[d]

[a]*Auburn University,* [b]*Indiana University–Purdue University Indianapolis*

[c]*Georgia State University,* [d]*University of Central Florida*

Supplementary Material

This supplementary material provides regularity assumptions, technical lemmas, proofs for Proposition 1, Theorems 1, 2, and 3.

S1 Regularity Assumptions

Without loss of generality, let the area of Ω be 1. For the univariate splines, we consider equally-spaced knots in our theoretical derivation. For a univariate function $\psi(\cdot)$, denote $\psi'(\cdot)$, $\psi''(\cdot)$ and $\psi^{(v)}(\cdot)$ be its first, second and v-th order derivative, respectively. For any bivariate function g defined on Ω, let $\|g(\boldsymbol{s})\|_{\infty,\Omega} = \sup_{\mathbf{s}\in\Omega} |g(\boldsymbol{s})|$ be the supremum norm of g, and let $|g|_{v,\infty,\Omega} = \max_{i+j=v} \|\nabla^i_{s_1} \nabla^j_{s_2} g(\mathbf{s})\|_{\infty,\Omega}$ be the maximum norms of all the v-th order derivatives of g over Ω. Let v be a nonnegative integer, and $\delta \in (0,1]$ such that $\varrho =$

$\delta+v \geq 1$. Let $\mathcal{H}^{(\varrho)}([a,b])$ be the class of functions ψ on $[a,b]$ whose v-th derivative exists and satisfies a Lipschitz condition of order δ: $|\psi^{(v)}(x) - \psi^{(v)}(x')| \leq C_v|x-x'|^{\delta}$, for $x, x' \in [a,b]$. Let $\mathcal{D}^0([a,b]) = \{g : Eg(Z) = 0, Eg^2(Z) < \infty\}$ be the function space defined on $[a,b]$ and $\mathcal{W}^{d+1,\infty}(\Omega) = \{g : |g|_{k,\infty,\Omega} < \infty, 0 \leq k \leq d+1\}$ be the standard Sobolev space.

The following are the technical assumptions needed to facilitate the technical details,

(A1) For $k = 1, \ldots, p$, $\beta_{0k} \in \mathcal{H}^{(\varrho)} \cap \mathcal{D}^0$ and the true bivariate function $\alpha_0(\cdot) \in \mathcal{W}^{d+1,\infty}(\Omega)$.

(A2) The density function $f(\mathbf{x}, z, \boldsymbol{s})$ of $(X_1, \ldots, X_p, Z, \boldsymbol{S})$ satisfies

$$0 < c_f \leq \inf_{(\mathbf{x},z,\boldsymbol{s}) \in \mathbb{R}^{p+1}\times\Omega} f(\mathbf{x}, z, \boldsymbol{s}) \leq \sup_{(\mathbf{x},z,\boldsymbol{S}) \in \mathbb{R}^{p+1}\times\Omega} f(\mathbf{x}, z, \boldsymbol{s}) \leq C_f < \infty.$$

The marginal density function $f_z(\cdot)$ of Z is twice continuously differentiable and the marginal density function $f_{\boldsymbol{s}}(\cdot)$ of $\boldsymbol{S}$ is bounded away from zero and infinity on Ω.

(A3) Recall that $\mathbb{S}_d^r(\triangle)$ denotes the spline space of degree d and smoothness r over $\triangle$. For every $\alpha \in \mathbb{S}_{3r+2}^r$ and every $\tau \in \triangle$, there exists a positive constant F_1, independent of α and τ, such that

$$F_1\|\alpha\|_{\infty,\tau} \leq \left\{ \sum_{\boldsymbol{S}_i \in \tau, i \in \{1,\ldots,n\}} \alpha(\boldsymbol{S}_i)^2 \right\}^{1/2} \leq F_2\|\alpha\|_{\infty,\tau},$$

where $\|\alpha\|_{\infty,\tau}$ denotes the supremum norm of α over triangle τ, F_2 is the largest among the numbers of observations in triangles $\tau \in \triangle$ and $F_2/F_1 = O(1)$.

(A4) The errors satisfy

$$E\left\{\varepsilon_i|\mathbf{X}_i = \mathbf{x}_i, Z_i = z_i, \boldsymbol{S}_i = \boldsymbol{s}_i\right\} = 0$$

and

$$E\left\{\varepsilon_i^{2+\nu}|\mathbf{X}_i = \mathbf{x}_i, Z_i = z_i, \boldsymbol{S}_i = \boldsymbol{s}_i\right\} < \infty$$

for some $\nu \in (3, \infty)$.

(A5) For some positive constant π, $(\min_{\tau\in\triangle} T_\tau)^{-1} \leq |\triangle| \leq \pi$, where T_τ is the radius of the largest disk contained in τ.

(A6) The number of knots J_n for the univariate splines and the triangulation size $|\triangle|$ satisfy that $J_n \to \infty$, $|\triangle| \to 0$, and $J_n \ll |\triangle|^2 n \log^{-1}(n)$; and the smoothness penalty parameter $\lambda_n n^{-1}|\triangle|^{-3} \to 0$.

(A6′) $h = o(n^{-1/5})$. For some $\varrho \geq 1$ and $d \geq 2$, $|\triangle| \ll n^{-2/(5d+5)}$ and $|\triangle|^{1/(\varrho+1)} n^{2/(5\varrho+5)} \ll J_n \ll |\triangle|^2 n \log^{-1}(n)$ and $\lambda_n n^{-1}|\triangle|^{-3} n^{2/5} = o(1)$.

(A7) The kernel function $K(\cdot)$ is a symmetric probability density with bounded support in $[-1, 1]$.

(A8) $\mathbf{\Omega}(z) = E(\mathbf{X}_1\mathbf{X}_1^\top|Z = z)$ and $\mathbf{\Gamma}(z) = E(\mathbf{X}_1\mathbf{X}_1^\top\mathbf{X}_1^\top\mathbf{X}_1|Z = z)$ are twice continuously differentiable. $\mathbf{C}(z)$ is uniformly bounded in $[a, b]$.

The above assumptions are regularity conditions that can be satisfied in many practical situations. Assumption (A1) describes the requirement on the varying coefficient functions, which are frequently used in the literature of non and semi-parametric estimation. Assumptions (A1) and (A2) are similar to Assumptions (A1) and (A2) in Yu et al. (2020). Assumptions (A3) and (A5) are analogue to Assumptions (A2) and (A5) in Yu et al. (2020), which has been widely used in the triangulation based literature (Wang et al., 2020; Lai and Wang, 2013). Assumptions (A6) and (A6′) show the requirement of the number of interior knots and the size of triangulation to ensure the consistency property of spline estimator and to obtain the local linear estimator, respectively. Note that the Assumption (A6′) only provides the order of $h = o(n^{-1/5})$ to be satisfied. This upper bounds on the bandwidth h in Assumption (A6′), is adapted from Wang et al. (2018), which is a necessary condition for Proposition 1. The naive empirical log-likelihood ratio is asymptotically non-central if the optimal bandwidth is used, which has been discussed in Xue and Zhu (2007). To make the likelihood ratio asymptotically parameter free, we adopt the undersmoothing Assumption (A6′). Assumptions (A4), (A7) and (A8) which are analogue to conditions 1, 2 and 3 in Wang et al. (2018), are common regularity conditions in non-parametric smoothing literature.

S2 Preliminaries

In this section, we depict the following bivariate splines properties. We first introduce some notations. For any vector $\mathbf{a} = (a_1, ..., a_n)^\top \in \mathbb{R}^n$, denote the norm $\|\mathbf{a}\|_r = (|a_1|^r + \cdots + |a_n|^r)^{1/r}$, $1 \le r < +\infty$, $\|\mathbf{a}\|_\infty = \max(|a_1|, \ldots, |a_n|)$. For any matrix $\mathbf{A} = (a_{ij})_{i=1,j=1}^{m,n}$, denote its L_r norm as $\|\mathbf{A}\|_r = \max_{\mathbf{a}\in\mathbb{R}^n, \mathbf{a}\neq \mathbf{0}} \|\mathbf{A}\mathbf{a}\|_r \|\mathbf{a}\|_r^{-1}$, for $r < +\infty$ and $\|\mathbf{A}\|_r = \max_{1\le i\le m} \sum_{j=1}^n |a_{ij}|$, for $r = \infty$. Given sequences of positive numbers a_n and b_n, $a_n \lesssim b_n$ means a_n/b_n is bounded, and $a_n \asymp b_n$ means both $a_n \lesssim b_n$ and $a_n \gtrsim b_n$ hold. We define the norm on the space $\mathcal{G}$. For any functions $\phi_1, \phi_2 \in \mathcal{G}$, define their theoretical inner product as $\langle \phi_1, \phi_2 \rangle = E\phi_1(\mathbf{X}, Z, \boldsymbol{S})\phi_2(\mathbf{X}, Z, \boldsymbol{S})$. Define their empirical inner product as $\langle \phi_1, \phi_2 \rangle_n = \frac{1}{n}\sum_{i=1}^n \phi_1(\mathbf{X}_i, Z_i, \boldsymbol{S}_i)\phi_2(\mathbf{X}_i, Z_i, \boldsymbol{S}_i)$. Hence, $\|\phi\| = \sqrt{\langle \phi, \phi \rangle}$ and $\|\phi\|_n = \sqrt{\langle \phi, \phi \rangle_n}$.

Lemma 1. *(Theorem 10.2, Lai and Schumaker (2007)) Suppose that $|\triangle|$ is a π-quasi-uniform triangleation of a polygonal domain Ω, and $\phi(\cdot) \in \mathcal{W}^{d+1,\infty}(\Omega)$.*

(i) For bi-integer(α_1, α_2) with $0 \le a_1 + a_2 \le d$, there exists a spline $\phi^*(\cdot) \in \mathbb{S}_d^0(\triangle)$ such that $\|\nabla_{s_1}^{a_1}\nabla_{s_2}^{a_2}(\phi - \phi^*)\|_\infty \le C|\triangle|^{d+1-a_1-a_2}|\phi|_{d+1,\infty}$ where C is a constant depending on d and shape parameter π.

(ii) For bi-integer(α_1, α_2) with $0 \le a_1 + a_2 \le d$, there exists a spline $\phi^{**}(\cdot) \in$

$\mathbb{S}_d^0(\triangle)$ *($d \geq 3r+2$) such that* $\|\nabla_{s_1}^{a_1}\nabla_{s_2}^{a_2}(\phi - \phi^{**})\|_\infty \leq C|\triangle|^{d+1-a_1-a_2}|\phi|_{d+1,\infty}$ *where* C *is a constant depending on* d*,* r *and shape parameter* π*.*

Lemma 1 shows that $\mathbb{S}_d^0(\triangle)$ has full approximation power, and $\mathbb{S}_d^0(\triangle)$ also has full approximation power if $d \geq 3r+2$.

Lemma 2. *(Lemma B.4, Yu et al. (2020)) For any* $k = 1, \cdots, p$*,* $\phi_k \in \mathcal{H}^{(\varrho)} \cap \mathcal{D}_k^0$*, there exist a constant* c *and a function* $\phi_k^* \in \mathcal{U}_k^0$ *such that* $\|\phi_k - \phi_k^*\|_\infty \leq c\|\phi_k^{(\varrho+1)}\|_\infty J_n^{-\varrho-1}$.

Lemma 3. *Suppose that Assumptions (A2), (A5) and (A6) hold. Then*

$$\sup_{\phi_1,\phi_2 \in \mathcal{A}} \left| \frac{\langle \phi_1, \phi_2 \rangle_n - \langle \phi_1, \phi_2 \rangle}{\|\phi_1\|\|\phi_2\|} \right| = O_{a.s.}\left(J_n^{1/2} |\triangle|^{-1} n^{-1/2} \log^{1/2} n \right)$$

where $\mathcal{A} = \Big\{\phi : \phi(\mathbf{x}, z, \boldsymbol{S}) = \sum_{k=1}^p \sum_{j \in \mathcal{J}} \eta_{kj} U_{kj}(z) x_k + \sum_{m \in \mathcal{M}} \gamma_m B_m(\boldsymbol{s}),$ $x_k, z, \eta_{kj}, \gamma_m \in \mathbb{R}, \boldsymbol{s} \in \Omega\}$.

Proof. The proof is similar as the proof of Lemma B.7 in Yu et al. (2020). □

Lemma 4. *Under Assumptions (A2), (A5) and (A6), there exist constants* $0 < c_A < C_A < \infty$*, such that* $c_A \leq \lambda_{min}(n\mathbf{A}_{11}) \leq \lambda_{max}(n\mathbf{A}_{11}) \leq C_A$*, where* $\mathbf{A}_{11}$ *is given in (2.2).*

Proof. The proof is similar as the proof of Lemma B.8 in Yu et al. (2020). Details are omitted. □

S3 Proof of Theorem 1

Proof. We first prove the consistency of $\widehat{\alpha}$. Define $\mathbf{H}_w = \mathbf{I} - \mathbf{W}(\mathbf{W}^\top\mathbf{W})^{-1}\mathbf{W}^\top$. Note that

$$
\begin{aligned}
\widehat{\boldsymbol{\theta}} &= \mathbf{A}_{22}\mathbf{Q}_2^\top\mathbf{B}^\top\mathbf{H}_w\mathbf{Y} \\
&= \mathbf{A}_{22}\mathbf{Q}_2^\top\mathbf{B}^\top\mathbf{H}_w(\boldsymbol{\beta}_0^\top(Z)\mathbf{X} + \alpha_0(\boldsymbol{S})) + \mathbf{A}_{22}\mathbf{Q}_2^\top\mathbf{B}^\top\mathbf{H}_w\boldsymbol{\varepsilon} \\
&= \widetilde{\boldsymbol{\theta}}_\mu + \widetilde{\boldsymbol{\theta}}_\varepsilon.
\end{aligned}
$$

According to Lemmas 1 and 2, there exist $\alpha^*(\boldsymbol{S}) = \mathbf{B}(\boldsymbol{S})\mathbf{Q}_2\boldsymbol{\theta}_0$ and $\boldsymbol{\beta}^*(z) = U(z)\boldsymbol{\eta}_0$, which are the best approximation to α_0 and $\boldsymbol{\beta}_0$ with the approximation rate at $\|\alpha^* - \alpha_0\|_\infty \le C_\alpha|\triangle|^{d+1}|\alpha_0|_{d+1,\infty}$ and $\|\boldsymbol{\beta}_0(z) - U(z)\boldsymbol{\eta}_0\|_\infty \le C_\beta J_n^{-\varrho-1}$. Hence, it is easy to find that $\|\boldsymbol{\beta}_0(Z)^\top\mathbf{X} - \mathbf{W}\boldsymbol{\eta}_0\|_\infty = O_p(C_\beta J_n^{-\varrho-1})$. Denote by $\boldsymbol{\gamma}_0 = \mathbf{Q}_2\boldsymbol{\theta}_0$ the spline coefficients of α^*. We have the following decomposition: $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = \widetilde{\boldsymbol{\theta}}_\mu - \boldsymbol{\theta}_0 + \widetilde{\boldsymbol{\theta}}_\varepsilon$. Note that

$$
\begin{aligned}
\|\widetilde{\boldsymbol{\theta}}_\mu - \boldsymbol{\theta}_0\| \le\ & \left\|\mathbf{A}_{22}\mathbf{Q}_2^\top\mathbf{B}^\top\mathbf{H}_w\boldsymbol{\beta}_0^\top(Z)\mathbf{X}\right\| \\
& + \left\|\mathbf{A}_{22}\mathbf{Q}_2^\top\mathbf{B}^\top\mathbf{H}_w(\alpha_0 - \mathbf{B}\mathbf{Q}_2\boldsymbol{\theta}_0) - \lambda_n\mathbf{A}_{22}\mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2\boldsymbol{\theta}_0\right\|.
\end{aligned}
$$

For any vector $\mathbf{a}$, according to Lemma 4 and the proof of Theorem 2 in Wang et al. (2020), one has $n\mathbf{a}^\top\mathbf{A}_{22}\mathbf{a} \le C|\triangle|^{-2}$. Hence, we have

$$
\begin{aligned}
& \left\|\mathbf{A}_{22}\mathbf{Q}_2^\top\mathbf{B}^\top\mathbf{H}_w\boldsymbol{\beta}_0^\top(Z)\mathbf{X}\right\| \le C^{1/2}|\triangle|^{-1}n^{-1}\left\|\mathbf{B}^\top\mathbf{H}_w(\mathbf{W}\boldsymbol{\eta} + O_p(h^p)\mathbf{1})\right\| \\
\le\ & O_p\left(J_n^{-\varrho-1}\right)|\triangle|^{-1}n^{-1}\left[\sum_{m\in\mathcal{M}}\{\mathbf{B}_m^\top\mathbf{H}_w\mathbf{1}\}^2\right]^{1/2} = O_p\left(J_n^{-p}\right).
\end{aligned}
$$

Similarly,

$$\begin{aligned}
& \left\|\mathbf{A}_{22}\mathbf{Q}_2^\top\mathbf{B}^\top\mathbf{H}_w(\alpha_0-\mathbf{B}\mathbf{Q}_2\boldsymbol{\theta}_0)\right\| \\
\leq\ & C^{1/2}|\triangle|^{-1}n^{-1}\left[\sum_{m\in\mathcal{M}}\{\mathbf{B}_m^\top\mathbf{H}_w(\alpha_0-\mathbf{B}\mathbf{Q}_2\boldsymbol{\theta}_0)\}^2\right]^{1/2} \\
=\ & O_p\left(|\triangle|^d|\alpha_0|_{d+1,\infty}\right),
\end{aligned}$$

and $\lambda_n\|\mathbf{A}_{22}\mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2\boldsymbol{\theta}_0\| \leq \frac{\lambda_n}{n|\triangle|^4}\left(|\alpha_0|_{2,\infty}+|\triangle|^{d-1}|\alpha_0|_{d+1,\infty}\right)$.

Thus,

$$\|\widetilde{\boldsymbol{\theta}}_\mu-\boldsymbol{\theta}_0\| = O_p\left\{J_n^{-\varrho-1}+\frac{\lambda_n}{n|\triangle|^4}|\alpha_0|_{2,\infty}+\left(1+\frac{\lambda_n}{n|\triangle|^5}\right)|\triangle|^d|\alpha_0|_{d+1,\infty}\right\}.$$

For any $\boldsymbol{b}$ with $\|\boldsymbol{b}\|=1$, we have $\boldsymbol{b}^\top\widetilde{\boldsymbol{\theta}}_\varepsilon=\sum_{i=1}^n\alpha_i\varepsilon_i$ and

$$\alpha_i^2=\boldsymbol{b}^\top\mathbf{A}_{22}\mathbf{Q}_2\mathbf{B}^\top\mathbf{H}_w\mathbf{B}\mathbf{Q}_2\mathbf{A}_{22}\boldsymbol{b}.$$

Following the similar argument in Lemma S.7 in Wang et al. (2020), we have $\max_{1\leq i\leq n}\alpha_i^2=O_p\left(n^{-2}|\triangle|^{-2}\right)$. Thus,

$$\|\widetilde{\boldsymbol{\theta}}_\varepsilon\| \leq |\triangle|^{-1}|\boldsymbol{\alpha}^\top\widetilde{\boldsymbol{\theta}}_\varepsilon| = |\triangle|^{-1}\left|\sum_{i=1}^n\alpha_i\varepsilon_i\right| = O_p\left(n^{-1/2}|\triangle|^{-2}\right).$$

Hence,

$$\begin{aligned}
& \|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\| \\
=\ & O_p\left\{J_n^{-\varrho-1}+n^{-1/2}|\triangle|^{-2}+\frac{\lambda_n}{n|\triangle|^4}|\alpha_0|_{2,\infty}+\left(1+\frac{\lambda_n}{n|\triangle|^5}\right)|\triangle|^d|\alpha_0|_{d+1,\infty}\right\}.
\end{aligned}$$

Observing that $\widehat{\alpha}(\boldsymbol{S})=\mathbf{B}(\boldsymbol{S})\widehat{\boldsymbol{\gamma}}=\mathbf{B}(\boldsymbol{S})\mathbf{Q}_2\widehat{\boldsymbol{\theta}}$, we have

$$\|\widehat{\alpha}-\alpha_0\|_{L_2}$$

$$
\begin{aligned}
&\leq \|\widehat{\alpha}-\rho_{0,\alpha_0}\|_{L_2}+|\triangle|^{d+1}|\alpha_0|_{d+1,\infty}\\
&\leq C\left(|\triangle|\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|+|\triangle|^{d+1}|\alpha_0|_{d+1,\infty}\right)\\
&= O_p\left\{J_n^{-\varrho-1}|\triangle|+n^{-1/2}|\triangle|^{-1}+\frac{\lambda_n}{n|\triangle|^3}|\alpha_0|_{2,\infty}\right.\\
&\quad\left.+\left(1+\frac{\lambda_n}{n|\triangle|^5}\right)|\triangle|^{d+1}|\alpha_0|_{d+1,\infty}\right\}.
\end{aligned}
$$

Next, we prove the consistency for $\widehat{\boldsymbol{\beta}}$.

Define $\mathbf{H}_B = \mathbf{I}-\mathbf{B}\mathbf{Q}_2\left\{\mathbf{Q}_2^{\top}\left(\mathbf{B}^{\top}\mathbf{B}+\lambda_n\mathbf{P}\right)\mathbf{Q}_2\right\}^{-1}\mathbf{Q}_2^{\top}\mathbf{B}^{\top}$. Let $\boldsymbol{\alpha}_0 = (\alpha_0(\boldsymbol{S}_1),\ldots,\alpha_0(\boldsymbol{S}_n))^{\top}$ and note that

$$
\widehat{\boldsymbol{\eta}}=\mathbf{A}_{11}\mathbf{W}^{\top}\mathbf{H}_B\mathbf{Y}=\mathbf{A}_{11}\mathbf{W}^{\top}\mathbf{H}_B\left(\boldsymbol{\beta}_0^{\top}(Z)\mathbf{X}+\boldsymbol{\alpha}_0\right)+\mathbf{A}_{11}\mathbf{W}^{\top}\mathbf{H}_B\boldsymbol{\varepsilon}=\widetilde{\boldsymbol{\eta}}_{\mu}+\widetilde{\boldsymbol{\eta}}_{\varepsilon}.
$$

Note that,

$$
\begin{aligned}
\|\widetilde{\boldsymbol{\eta}}_{\mu}-\boldsymbol{\eta}_0\| &\leq \left\|\mathbf{A}_{11}\mathbf{W}^{\top}\mathbf{H}_B(\boldsymbol{\beta}_0^{\top}(Z)\mathbf{X}-\mathbf{W}\boldsymbol{\eta}_0)\right\|+\left\|\mathbf{A}_{11}\mathbf{W}^{\top}\mathbf{H}_B\boldsymbol{\alpha}_0\right\|\\
&\leq O_p\left(J_n^{-\varrho-1}\right)\|\mathbf{A}_{11}\mathbf{W}^{\top}\mathbf{H}_B\mathbf{1}\|+\left\|\mathbf{A}_{11}\mathbf{W}^{\top}\mathbf{H}_B\boldsymbol{\alpha}_0\right\|\\
&= O(1)\|\mathbf{A}_{11}\mathbf{W}^{\top}\mathbf{H}_B\boldsymbol{\alpha}_0\|.
\end{aligned}
$$

By the Lemma 4, there exist constants $0\leq c_A<C_A<\infty$, such that with probability approaching 1 as $n\to\infty$,

$$
c_A\mathbf{I}_{((J_n+\varrho+1))\times(J_n+\varrho+1)}\leq n\mathbf{A}_{11}\leq C_A\mathbf{I}_{(J_n+\varrho+1)\times(J_n+\varrho+1)}.
$$

Hence, we have

$$
\|\widetilde{\boldsymbol{\eta}}_{\mu}-\boldsymbol{\eta}_0\|
$$

$$\begin{aligned} &\leq O(1)\|n^{-1}\mathbf{W}^\top \left(\mathbf{I} - \mathbf{B}\mathbf{Q}_2\{\mathbf{Q}_2^\top(\mathbf{B}^\top\mathbf{B} + \lambda_n\mathbf{P})\mathbf{Q}_2\}^{-1}\mathbf{Q}_2^\top\mathbf{B}^\top\right)\boldsymbol{\alpha}_0\| \\ &= O(1)\|\mathbf{R}\|, \end{aligned}$$

where $\mathbf{R} = (R_1, \ldots, R_{p(J_n+\varrho+1)})^\top$, with

$$R_j = n^{-1}\mathbf{W}_j^\top \left[\boldsymbol{\alpha}_0 - \mathbf{B}\mathbf{Q}_2\{\mathbf{Q}_2^\top(\mathbf{B}^\top\mathbf{B} + \lambda_n\mathbf{P})\mathbf{Q}_2\}^{-1}\mathbf{Q}_2^\top\mathbf{B}^\top\boldsymbol{\alpha}_0\right]$$

for $\mathbf{W}_j^\top = (W_{1j}, \ldots, W_{nj})$. Next we derive the order of R_j, $j = 1, \ldots, p(J_n + \varrho+1)$. For any $\alpha_j \in \mathbb{S}$, we have $R_j = \langle w_j, \alpha_0 - \rho_{\lambda,\alpha_0}\rangle_n = \langle w_j - \alpha_j, \alpha_0 - \rho_{\lambda,\alpha_0}\rangle_n + \lambda_n n^{-1}\langle \rho_{\lambda,\alpha_0}, \alpha_j\rangle_{\mathcal{E}}$, where $\rho_{\lambda,\alpha_0} = \arg\min_{\rho\in\mathbb{S}} \sum_{i=1}^n \{\alpha_0(\boldsymbol{S}_i) - \rho(\boldsymbol{S}_i)\}^2 + \frac{\lambda}{2}\mathcal{E}(\rho)$ is the penalized least-squares splines of $\alpha(\cdot,\cdot)$.

By Assumptions (A1)-(A6) and Lemma S.6 in Wang et al. (2020), $|R_j| = o_p(n^{-1/2})$, for $j = 1, \ldots, p(J_n+\varrho+1)$. Therefore, $\|\widetilde{\boldsymbol{\eta}}_\mu - \boldsymbol{\eta}_0\| = O_p(n^{-1/2}J_n^{1/2})$.

Note that $\widetilde{\boldsymbol{\eta}}_\varepsilon = \mathbf{A}_{11}\mathbf{W}^\top\left(\mathbf{I} - \mathbf{B}\mathbf{Q}_2\mathbf{V}_{22}^{-1}\mathbf{Q}_2^\top\mathbf{B}^\top\right)\boldsymbol{\varepsilon}$. For any $\boldsymbol{b}$ with $\|\boldsymbol{b}\| = 1$, we have $\boldsymbol{b}^\top\widetilde{\boldsymbol{\eta}}_\varepsilon = \sum_{i=1}^n \alpha_i\varepsilon_i$ and

$$\alpha_i^2 = n^{-2}\boldsymbol{b}^\top(n\mathbf{A}_{11})\left(\mathbf{W}_i^\top - \mathbf{V}_{21}\mathbf{V}_{22}^{-1}\mathbf{Q}_2^\top\mathbf{B}_i\right)\left(\mathbf{W}_i - \mathbf{B}_i^\top\mathbf{Q}_2\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\right)(n\mathbf{A}_{11})\boldsymbol{b},$$

and conditioning on $\{(\mathbf{W}_i, \boldsymbol{S}_i), i = 1, \ldots, n\}$, $\alpha_i\varepsilon_i$'s are independent. By Lemma 4, we have that $\max_{1\leq i\leq n}\alpha_i^2 \leq Cn^{-2}\max_{1\leq i\leq n}\left\{\|\mathbf{W}_i\|^2 + \|\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{Q}_2^\top\mathbf{B}_i\|^2\right\}$, where for any $\mathbf{b} \in \mathbb{R}^p$,

$$\begin{aligned} &\mathbf{b}^\top\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{Q}_2^\top\mathbf{B}_i\mathbf{b} \\ &= n^{-1}\mathbf{b}^\top\mathbf{V}_{12}\left(\mathbf{Q}_2^\top\boldsymbol{\Gamma}_{n,\lambda}\mathbf{Q}_2\right)^{-1}\mathbf{Q}_2^\top\mathbf{B}_i\mathbf{b} \leq Cn^{-1}|\triangle|^{-2}\mathbf{b}^\top\mathbf{W}^\top\mathbf{B}\mathbf{B}_i\mathbf{b} \end{aligned}$$

and the j-th component of $n^{-1}\mathbf{W}^\top \mathbf{B}\mathbf{B}_i$ is

$$n^{-1}\sum_{i'=1}^{n} W_{i'j}\sum_{m\in\mathcal{M}} B_m(\boldsymbol{S}_{i'})B_m(\boldsymbol{S}_i).$$

Under Assumption (A2), we have

$$E\left\{n^{-1}\sum_{i'=1}^{n} W_{i'j}\sum_{m\in\mathcal{M}} B_m(\boldsymbol{S}_{i'})B_m(\boldsymbol{S}_i)\right\}^2 = O(1),$$

for large n. Thus with probability approaching 1,

$$\max_{1\le i\le n}\left|\frac{1}{n}\sum_{i'=1}^{n} W_{i'j}\sum_{m\in\mathcal{M}} B_m(\boldsymbol{S}_{i'})B_m(\boldsymbol{S}_i)\right| = O_p(1),$$
$$\max_{1\le i\le n}\|\mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{Q}_2^\top \mathbf{B}_i\|^2 = O_p(|\triangle|^{-2}).$$

Therefore, $\max_{1\le i\le n}\alpha_i^2 = O_p\{n^{-2}(|\triangle|^{-2}+J_n)\}$ and $\|\widetilde{\boldsymbol{\eta}}_\varepsilon\| = O_p(n^{-1}|\triangle|^{-1} + n^{-1}J_n^{1/2})$. Let $\widehat{\boldsymbol{\eta}} = (\widehat{\boldsymbol{\eta}}_1,\ldots,\widehat{\boldsymbol{\eta}}_p)$. $\|\boldsymbol{\beta}_0(z)-U(z)\boldsymbol{\eta}_0\|_\infty \le C_\beta J_n^{-\varrho-1}$ and observing that $\widehat{\beta}_k(Z) = \mathbf{U}_k^\top(Z)\widehat{\boldsymbol{\eta}}_k$, we have $\|\widehat{\beta}_k - \beta_{0k}\|_{L_2} \le C\left(\|\widehat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_{0k}\| + J_n^{-\varrho-1}\right) = O_p\left(n^{-1/2}J_n^{1/2} + n^{-1}|\triangle|^{-1} + J_n^{-\varrho-1}\right)$, and the consistency of $\widehat{\boldsymbol{\beta}}$ is proved. □

S4 Proof of Proposition 1

Proof. Recall that $\boldsymbol{\Omega}(z) = E\left(\mathbf{X}_i\mathbf{X}_i^\top|Z=z\right)$, $\boldsymbol{\Gamma}(z) = E\left(\mathbf{X}_i\mathbf{X}_i^\top\mathbf{X}_i^\top\mathbf{X}_i|Z=z\right)$. By the definition of $g_i\{\boldsymbol{\beta}_0(z)\}$, we have the following decomposition,

$$\begin{aligned} g_i\{\boldsymbol{\beta}_0(z)\} &= \left\{Y_i - \boldsymbol{\beta}_0^\top(z)\mathbf{X}_i - \widehat{\alpha}(\boldsymbol{S}_i)\right\}\mathbf{X}_iK_h(Z_i-z) \\ &= \left\{Y_i - \boldsymbol{\beta}_0^\top(Z_i)\mathbf{X}_i - \alpha_0(\boldsymbol{S}_i) + \boldsymbol{\beta}_0^\top(Z_i)\mathbf{X}_i - \boldsymbol{\beta}_0^\top(z)\mathbf{X}_i\right. \\ &\quad \left. +\alpha_0(\boldsymbol{S}_i) - \widehat{\alpha}(\boldsymbol{S}_i)\right\}\mathbf{X}_iK_h(Z_i-z) \end{aligned}$$

$$
\begin{aligned}
&= \left\{\epsilon_i + [\boldsymbol{\beta}_0(Z_i) - \boldsymbol{\beta}_0(z)]^\top \mathbf{X}_i + [\alpha_0(\boldsymbol{S}_i) - \widehat{\alpha}(\boldsymbol{S}_i)]\right\} \\
&\quad \times \mathbf{X}_i K_h(Z_i - z) \\
&= \epsilon_i \mathbf{X}_i K_h(Z_i - z) + \mathbf{X}_i \mathbf{X}_i^\top [\boldsymbol{\beta}_0(Z_i) - \boldsymbol{\beta}_0(z)] K_h(Z_i - z) \\
&\quad + [\alpha_0(\boldsymbol{S}_i) - \widehat{\alpha}(\boldsymbol{S}_i)] \mathbf{X}_i K_h(Z_i - z) \\
&:= \boldsymbol{\xi}_i + \mathbf{L}_{1i} + \mathbf{L}_{2i}.
\end{aligned}
$$

Denote $\boldsymbol{\beta}_0^{(1)}(\boldsymbol{z}) = \left(\beta'_{01}(z_1), \beta'_{02}(z_2), \ldots, \beta'_{0p}(z_p)\right)^\top$. By the smoothness of β_{0k}, $k = 1, 2, \ldots, p$, we have $\boldsymbol{\beta}_0^{(1)}(\boldsymbol{z}) = O(1)\mathbf{1}_{p\times 1}$ for all $\boldsymbol{z}$.

It is clear that $E\boldsymbol{\xi}_i = \mathbf{0}$, and we have

$$
\begin{aligned}
E(\mathbf{L}_{1i}) &= E\left\{E(\mathbf{X}_i \mathbf{X}_i^\top | Z_i)\left[\boldsymbol{\beta}_0^{(1)}(z^*)(Z_i - z)\right] K_h(Z_i - z)\right\} \\
&= \boldsymbol{\beta}_0^{(1)}(\boldsymbol{z}^*) E[\boldsymbol{\Omega}(Z_i)(Z_i - z) K_h(Z_i - z)]) \\
&= \boldsymbol{\beta}_0^{(1)}(\boldsymbol{z}^*) \int_a^b \boldsymbol{\Omega}(u)(u - z) K_h(u - z) f(u) du \\
&= \boldsymbol{\beta}_0^{(1)}(\boldsymbol{z}^*) \left[h \int_a^b v(\boldsymbol{\Omega}(z) + \boldsymbol{\Omega}'(z) h v + 1/2\boldsymbol{\Omega}'(z) h^2 v^2) K(v)(f(z)\right. \\
&\quad \left. + f'(z) h v + 1/2 f''(z) h^2 v^2) dv\right] \\
&= O(h^2)\mathbf{1}_{n\times 1}.
\end{aligned}
$$

According to the proof of Theorem 1, we also have

$$
\begin{aligned}
E(\mathbf{L}_{2i}) &= E\{\mathbf{X}_i K_h(Z_i - z)(\alpha_0(\boldsymbol{S}_i) - \widehat{\alpha}(\boldsymbol{S}_i))\} \\
&= E\left\{\mathbf{X}_i K_h(Z_i - z) E\left[(\alpha_0(\boldsymbol{S}_i) - \widehat{\alpha}(\boldsymbol{S}_i)) \mid \{\mathbf{X}_i, Z_i, \boldsymbol{S}_i\}_{i=1}^n\right]\right\} \\
&= E\left\{\mathbf{X}_i K_h(Z_i - z)\left(\alpha_0(\boldsymbol{S}_i) - E\left[\widehat{\alpha}(\boldsymbol{S}_i) \mid \{\mathbf{X}_i, Z_i, \boldsymbol{S}_i\}_{i=1}^n\right]\right)\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= E\left\{\mathbf{X}_i K_h(Z_i - z)\left(\alpha_0(\boldsymbol{S}_i) - \boldsymbol{B}(\boldsymbol{S}_i)\boldsymbol{Q}_2\widetilde{\boldsymbol{\theta}}_\mu\right)\right\} \\
&\leq E\left\{|\mathbf{X}_i K_h(Z_i - z)|\right\} \\
&\quad \times E\left(\left\|\alpha_0(\boldsymbol{S}_i) - \alpha^*(\boldsymbol{S}_i) + \boldsymbol{B}(\boldsymbol{S}_i)\boldsymbol{Q}_2\left(\boldsymbol{\theta}_0 - \widetilde{\boldsymbol{\theta}}_\mu\right)\right\|_\infty\right) \\
&\lesssim E\left\{|\mathbf{X}_i K_h(Z_i - z)|\right\} E\left(\|\alpha_0 - \alpha^*\|_\infty + O(|\triangle|)\|\boldsymbol{\theta}_0 - \widetilde{\boldsymbol{\theta}}_\mu\|_2\right),
\end{aligned}
$$

where by Theorem 1, we have

$$
E\left(\|\alpha_0(\boldsymbol{S}_i) - \alpha^*\|_\infty + |\Delta|\|\boldsymbol{\theta}_0 - \widetilde{\boldsymbol{\theta}}_\mu\|_2\right) = O\left(J_n^{-\varrho-1}|\triangle| + \frac{\lambda_n}{n|\triangle|^3} + |\triangle|^{d+1}\right),
$$

and $E\{|\mathbf{X}_i K_h(Z_i - z)|\} = E\left[E\{|\mathbf{X}_i|\big|Z_i\}K_h(Z_i - z)\right] \asymp E\left[K_h(Z_i - z)\right] \times \mathbf{1}_{p\times 1} = O(f(z))\mathbf{1}_{p\times 1}$. If $h = o(n^{-1/5})$, when $|\triangle| \ll n^{-2/(5d+5)}$ and $J_n \gg |\triangle|^{1/(\varrho+1)}n^{2/(5\varrho+5)}$, we have $E\mathbf{L}_{2i} = O(h^2)\mathbf{1}_{p\times 1}$ by Assumption (A6′). Therefore, we have $E\{g_i\{\boldsymbol{\beta}_0(z)\}\} = O(h^2)\mathbf{1}_{p\times 1}$. In the following, we calculate the variance of $g_i\{\boldsymbol{\beta}_0(z)\}$. Firstly, we have

$$
\begin{aligned}
E\left(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^\top\right) &= E\left\{\epsilon_i^2\mathbf{X}_i\mathbf{X}_i^\top K_h^2(Z_i - z)\right\} \\
&= \sigma^2 E\left[E\left(\mathbf{X}_i\mathbf{X}_i^\top|Z_i\right)K_h^2(Z_i - z)\right] \\
&= \sigma^2\boldsymbol{\Omega}(z)f(z)\mu_{20}h^{-1}\left(1 + o(1)\right).
\end{aligned}
$$

Secondly, we have

$$
\begin{aligned}
&E\left(\mathbf{L}_{1i}\mathbf{L}_{1i}^\top\right) \\
&= E\left\{\mathbf{X}_i\mathbf{X}_i^\top K_h^2(Z_i - z)\mathbf{X}_i^\top(\boldsymbol{\beta}_0(Z_i) - \boldsymbol{\beta}_0(z))(\boldsymbol{\beta}_0(Z_i) - \boldsymbol{\beta}_0(z))^\top\mathbf{X}_i\right\} \\
&= E\left\{\mathbf{X}_i\mathbf{X}_i^\top K_h^2(Z_i - z)\left[(Z_i - z)^2\mathbf{X}_i^\top\boldsymbol{\beta}_0^{(1)}(\boldsymbol{z}^*)\boldsymbol{\beta}_0^{(1)}(\boldsymbol{z}^*)^\top\mathbf{X}_i\right.\right.
\end{aligned}
$$

$$
\begin{aligned}
& + o(Z_i - z)^2 \mathbf{X}_i^\top \mathbf{X}_i \big] \big\} \\
= \;& E\left\{ E\left[\mathbf{X}_i \mathbf{X}_i^\top \mathbf{X}_i^\top \boldsymbol{\beta}_0^{(1)}(\boldsymbol{z}^*) \boldsymbol{\beta}_0^{(1)}(\boldsymbol{z}^*)^\top \mathbf{X}_i \middle| Z_i \right] K_h^2(Z_i - z)(Z_i - z)^2 \right\} \\
& \times \{1 + o(1)\} \\
\asymp \;& E\left\{ E\left[\mathbf{X}_i \mathbf{X}_i^\top \mathbf{X}_i^\top \mathbf{X}_i \middle| Z_i \right] K_h^2(Z_i - z)(Z_i - z)^2 \right\} (1 + o(1)) \\
= \;& E\left\{ \boldsymbol{\Gamma}(Z_i) K_h^2(Z_i - z)(Z_i - z)^2 \right\} (1 + o(1)) = \boldsymbol{\Gamma}(z) f(z) \mu_{22} h \left(1 + o(1)\right).
\end{aligned}
$$

Finally,

$$
\begin{aligned}
E\left(\mathbf{L}_{2i}\mathbf{L}_{2i}^\top\right) &= E\left\{ \mathbf{X}_i \mathbf{X}_i^\top (\alpha_0(\boldsymbol{S}_i) - \widehat{\alpha}(\boldsymbol{S}_i))^2 K_h^2(Z_i - z) \right\} \\
&= E\left\{ E\left[\mathbf{X}_i \mathbf{X}_i^\top (\alpha_0(\boldsymbol{S}_i) - \widehat{\alpha}(\boldsymbol{S}_i))^2 K_h^2(Z_i - z) | \left\{ \mathbf{X}_i, Z_i, \boldsymbol{S}_i \right\}_{i=1}^n \right] \right\} \\
&= E\left\{ E\left[(\alpha_0(\boldsymbol{S}_i) - \widehat{\alpha}(\boldsymbol{S}_i))^2 | \left\{ \mathbf{X}_i, Z_i, \boldsymbol{S}_i \right\}_{i=1}^n \right] \mathbf{X}_i \mathbf{X}_i^\top K_h^2(Z_i - z) \right\} \\
&= E\left\{ E\left[\mathbf{B}(\boldsymbol{S}_i) \mathbf{Q}_2 (\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}})(\boldsymbol{\theta}_0 - \widehat{\boldsymbol{\theta}})^\top \mathbf{Q}_2^\top \mathbf{B}^\top(\boldsymbol{S}_i) \right] \right. \\
&\quad \left. \times \mathbf{X}_i \mathbf{X}_i^\top K_h^2(Z_i - z) \right\} \\
&= E\left\{ \mathbf{X}_i \mathbf{X}_i^\top K_h^2(Z_i - z) \| \mathbf{B}(\boldsymbol{S}_i) \mathbf{Q}_2 (\boldsymbol{\theta}_0 - \widetilde{\boldsymbol{\theta}}_\mu) \|_2^2 \right\} \\
&\quad + \sigma^2 E\left\{ \mathbf{X}_i \mathbf{X}_i^\top K_h^2(Z_i - z) \| \mathbf{B}(\boldsymbol{S}_i) \mathbf{Q}_2 \boldsymbol{A}_{22} \mathbf{Q}_2^\top \mathbf{B}^\top \mathbf{H}_{\boldsymbol{\omega}} \|_2^2 \right\}.
\end{aligned}
$$

On the one hand, for (k, k')-th entry, k, $k' = 1, 2, \ldots, p$, we have

$$
\begin{aligned}
& E\left\{ X_{ik} X_{ik'} K_h^2(Z_i - z) \| \mathbf{B}(\boldsymbol{S}_i) \mathbf{Q}_2 (\boldsymbol{\theta}_0 - \widetilde{\boldsymbol{\theta}}_\mu) \|_2^2 \right\} \\
\leq \;& E\left\{ |\Delta|^2 X_{ik} X_{ik'} K_h^2(Z_i - z) \| \mathbf{Q}_2 (\boldsymbol{\theta}_0 - \widetilde{\boldsymbol{\theta}}_\mu) \|_2^2 \right\} \\
\leq \;& O\left(|\Delta|^2\right) \left(E\left\{ X_{ik}^2 X_{ik'}^2 K_h^4(Z_i - z) \right\} \right)^{1/2} \left(E \| \boldsymbol{\theta}_0 - \widetilde{\boldsymbol{\theta}}_\mu \|_2^4 \right)^{1/2} \\
= \;& O\left(|\Delta|^2 h^{-3/2}\right) \left(E\left\{ \| \boldsymbol{\theta}_0 - \widetilde{\boldsymbol{\theta}}_\mu \|_2^4 \right\} \right)^{1/2}. \qquad \text{(S4.1)}
\end{aligned}
$$

On the other hand, for (k,k')-th entry, k, $k'=1,2,\ldots,p$, we have

$$\begin{aligned} & E\left\{X_{ik}X_{ik'}K_h^2(Z_i-z)\|\mathbf{B}(\boldsymbol{S}_i)\mathbf{Q}_2\boldsymbol{A}_{22}\mathbf{Q}_2^\top\mathbf{B}^\top\mathbf{H}_{\boldsymbol{\omega}}\|_2^2\right\} \\ \leq\ & C|\Delta|^{-4}n^{-2}E\left\{X_{ik}X_{ik'}K_h^2(Z_i-z)\|\mathbf{B}(\boldsymbol{S}_i)\mathbf{B}^\top\mathbf{H}_{\boldsymbol{\omega}}\|_2^2\right\} \\ \leq\ & C|\Delta|^{-4}n^{-2}\left(E\left\{X_{ik}^2X_{ik'}^2K_h^4(Z_i-z)\right\}\right)^{1/2}\left(E\left\{\|\mathbf{B}(\boldsymbol{S}_i)\mathbf{B}^\top\mathbf{H}_{\boldsymbol{\omega}}\|_2^4\right\}\right)^{1/2} \\ =\ & O\left(n^{-1}h^{-3/2}\right). \end{aligned} \tag{S4.2}$$

Combining (S4.1) and (S4.2), we have $E\mathbf{L}_{2i}\mathbf{L}_{2i}^\top=O(|\Delta|^2h^{-3/2}+n^{-1}h^{-3/2})\mathbf{1}_{p\times p}$.

Hence, we have

$$\begin{aligned} Var\{g_i\{\boldsymbol{\beta}_0(z)\}\} &= E\left\{g_i\{\boldsymbol{\beta}_0(z)\}g_i\{\boldsymbol{\beta}_0(z)\}^\top\right\} \\ &= E\left(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^\top+\mathbf{L}_{1i}\mathbf{L}_{1i}^\top+\mathbf{L}_{2i}\mathbf{L}_{2i}^\top\right) \\ &= E\left(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^\top\right)(1+o(1)) \\ &= \sigma^2\boldsymbol{\Omega}(z)f(z)\mu_{20}h^{-1}(1+o(1)). \end{aligned}$$

□

S5 Proof of Theorem 2

Proof. First, for convenience we suppress the argument z in the functions such as $\boldsymbol{\beta}(z)$, $\boldsymbol{\Omega}(z)$ and so on, since we fix $z\in[a,b]$ in this proof.

For the minimization problem (3.3), we use the Lagrange multiplier method:

$$\min\frac{1}{n}\sum_{i=1}^n\log\left[1+\boldsymbol{\delta}^\top(z)g_i\{\boldsymbol{\beta}(z)\}\right]+\boldsymbol{\nu}^\top(z)H\{\boldsymbol{\beta}(z)\},$$

where $\boldsymbol{\nu}(z)$ is a $q \times 1$ vector of Lagrange multipliers. Define

$$M_{1n}(\boldsymbol{\beta}, \boldsymbol{\delta}) = \frac{1}{n} \sum_{i=1}^{n} \frac{g_i(\boldsymbol{\beta})}{1 + \boldsymbol{\delta}^\top(\boldsymbol{\beta}) g_i(\boldsymbol{\beta})}$$

and

$$M_{2n}(\boldsymbol{\beta}, \boldsymbol{\delta}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\frac{\partial g_i^\top(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \boldsymbol{\delta}}{1 + \boldsymbol{\delta}^\top(\boldsymbol{\beta}) g_i(\boldsymbol{\beta})} + \boldsymbol{\nu}^\top \mathbf{C}(\boldsymbol{\beta}).$$

We first obtain their derivatives with respect to the three variables $\boldsymbol{\beta}, \boldsymbol{\delta}$ and $\boldsymbol{\nu}$.

$$\frac{\partial M_{1n}(\boldsymbol{\beta}, \boldsymbol{\delta})}{\partial \boldsymbol{\beta}^\top} = \frac{1}{n} \sum_{i=1}^{n} \frac{\frac{\partial g_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top}(1 + \boldsymbol{\delta}^\top(\boldsymbol{\beta}) g_i(\boldsymbol{\beta})) - g_i(\boldsymbol{\beta}) \boldsymbol{\delta}^\top \frac{\partial g_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top}}{(1 + \boldsymbol{\delta}^\top(\boldsymbol{\beta}) g_i(\boldsymbol{\beta}))^2},$$

$$\frac{\partial M_{1n}(\boldsymbol{\beta}, \boldsymbol{\delta})}{\partial \boldsymbol{\delta}^\top} = -\frac{1}{n} \sum_{i=1}^{n} \frac{g_i(\boldsymbol{\beta}) g_i^\top(\boldsymbol{\beta})}{(1 + \boldsymbol{\delta}^\top(\boldsymbol{\beta}) g_i(\boldsymbol{\beta}))^2}, \quad \frac{\partial M_{1n}(\boldsymbol{\beta}, \boldsymbol{\delta})}{\partial \boldsymbol{\nu}^\top} = 0,$$

$$\begin{aligned} \frac{\partial M_{2n}(\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\nu})}{\partial \boldsymbol{\beta}^\top} &= \frac{1}{n} \sum_{i=1}^{n} \frac{\frac{\partial^2 g_i^\top(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top \partial \boldsymbol{\beta}} \boldsymbol{\delta}(1 + \boldsymbol{\delta}^\top(\boldsymbol{\beta}) g_i(\boldsymbol{\beta})) - \frac{\partial g_i^\top(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \boldsymbol{\delta} \boldsymbol{\delta}^\top \frac{\partial g_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top}}{(1 + \boldsymbol{\delta}^\top(\boldsymbol{\beta}) g_i(\boldsymbol{\beta}))^2} \\ &\quad + \frac{\partial \mathbf{C}^\top(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top} \boldsymbol{\nu}, \end{aligned}$$

$$\frac{\partial M_{2n}(\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\nu})}{\partial \boldsymbol{\delta}^\top} = \frac{1}{n} \sum_{i=1}^{n} \frac{\frac{\partial g_i^\top(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top} - \frac{\partial g_i^\top(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top} \boldsymbol{\delta} g_i^\top(\boldsymbol{\beta})}{(1 + \boldsymbol{\delta}^\top(\boldsymbol{\beta}) g_i(\boldsymbol{\beta}))^2}, \quad \frac{\partial M_{2n}(\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}^\top} = \mathbf{C}^\top(\boldsymbol{\beta}),$$

$$\frac{\partial H(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^\top} = \mathbf{C}(\boldsymbol{\beta}), \quad \frac{\partial H(\boldsymbol{\beta})}{\partial \boldsymbol{\delta}^\top} = 0, \quad \frac{\partial H(\boldsymbol{\beta})}{\partial \boldsymbol{\nu}^\top} = 0.$$

Hence, we have the following Taylor expansions of the system of equations at $(\boldsymbol{\beta}_0, 0, 0)$. Denote the solution to this equation system as $\left\{\widetilde{\boldsymbol{\beta}}(z), \widetilde{\boldsymbol{\delta}}(z), \widetilde{\boldsymbol{\nu}}(z)\right\}$. Let $\Delta_n = \|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\| + \|\widetilde{\boldsymbol{\delta}}\| + \|\widetilde{\boldsymbol{\nu}}\|$.

$0 = M_{1n}\left(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\delta}}, \widetilde{\boldsymbol{\nu}}\right)$

$$
\begin{aligned}
&= M_{1n}(\boldsymbol{\beta}_0, 0) + \frac{\partial M_{1n}(\boldsymbol{\beta}_0, 0)}{\partial \boldsymbol{\beta}^\top}\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) + \frac{\partial M_{1n}(\boldsymbol{\beta}_0, 0)}{\partial \boldsymbol{\delta}^\top}\left(\widetilde{\boldsymbol{\delta}} - 0\right) \\
&\quad + \frac{\partial M_{1n}(\boldsymbol{\beta}_0, 0)}{\partial \boldsymbol{\nu}^\top}\left(\widetilde{\boldsymbol{\nu}} - 0\right) + o_p(\Delta_n) \\
&= \frac{1}{n}\sum_{i=1}^n g_i(\boldsymbol{\beta}_0) + \frac{1}{n}\sum_{i=1}^n \frac{\partial g_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^\top}\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) - \frac{1}{n}\sum_{i=1}^n g_i(\boldsymbol{\beta}_0) g_i^\top(\boldsymbol{\beta}_0)\widetilde{\boldsymbol{\delta}} + o_p(\Delta_n),
\end{aligned}
$$

$$
\begin{aligned}
0 &= M_{2n}\left(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\delta}}, \widetilde{\boldsymbol{\nu}}\right) \\
&= M_{2n}(\boldsymbol{\beta}_0, 0, 0) + \frac{\partial M_{2n}(\boldsymbol{\beta}_0, 0, 0)}{\partial \boldsymbol{\beta}^\top}\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) + \frac{\partial M_{2n}(\boldsymbol{\beta}_0, 0, 0)}{\partial \boldsymbol{\delta}^\top}\left(\widetilde{\boldsymbol{\delta}} - 0\right) \\
&\quad + \frac{\partial M_{2n}(\boldsymbol{\beta}_0, 0, 0)}{\partial \boldsymbol{\nu}^\top}\left(\widetilde{\boldsymbol{\nu}} - 0\right) + o_p(\Delta_n) \\
&= \frac{1}{n}\sum_{i=1}^n \frac{\partial g_i^\top(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\widetilde{\boldsymbol{\delta}} + \mathbf{C}^\top(\boldsymbol{\beta}_0)\widetilde{\boldsymbol{\nu}} + o_p(\Delta_n),
\end{aligned}
$$

and $0 = H\left(\widetilde{\boldsymbol{\beta}}\right) = H\left(\boldsymbol{\beta}_0\right) + \mathbf{C}^\top(\boldsymbol{\beta}_0)\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) + o_p(\Delta_n) = \mathbf{C}^\top(\boldsymbol{\beta}_0)\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) + o_p(\Delta_n)$. Putting the above equations into a matrix form, we obtain

$$
\begin{pmatrix} -n^{-1}\sum_{i=1}^n g_i(\boldsymbol{\beta}_0) + o_p(\Delta_n) \\ o_p(\Delta_n) \\ -H(\boldsymbol{\beta}_0) + o_p(\Delta_n) \end{pmatrix} = \boldsymbol{\Sigma}_n \begin{pmatrix} C_n^2 n^{-1}\widetilde{\boldsymbol{\delta}} \\ \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widetilde{\boldsymbol{\nu}} \end{pmatrix}.
$$

where

$$
\boldsymbol{\Sigma}_n = \begin{pmatrix} -C_n^{-2}\sum_{i=1}^n g_i(\boldsymbol{\beta}_0) g_i^\top(\boldsymbol{\beta}_0) & n^{-1}\sum_{i=1}^n \frac{\partial g_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^\top} & 0 \\ n^{-1}\sum_{i=1}^n \frac{\partial g_i^\top(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} & 0 & \mathbf{C}^\top(\boldsymbol{\beta}_0) \\ 0 & \mathbf{C}(\boldsymbol{\beta}_0) & 0 \end{pmatrix}.
$$

Then we have $\boldsymbol{\Sigma}_n \xrightarrow{\mathcal{P}} \boldsymbol{\Sigma} = \begin{pmatrix} -\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & 0 \\ \boldsymbol{\Sigma}_{12} & 0 & \boldsymbol{\Sigma}_{23}^\top \\ 0 & \boldsymbol{\Sigma}_{23} & 0 \end{pmatrix}$, and $\boldsymbol{\Sigma}_{23} = \mathbf{C}(\boldsymbol{\beta}_0)$. By Proposition 1, it is easy to find that

$$\boldsymbol{\Sigma}_{11} = \sigma_0^2 \boldsymbol{\Omega}(z) f(z) \mu_{20}, \boldsymbol{\Sigma}_{12} = \boldsymbol{\Omega}(z) f(z). \tag{S5.3}$$

By the simple calculation, we have

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} -\boldsymbol{\Sigma}_{11}^{-1} + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Upsilon}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1} & \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Upsilon} & \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\mathbf{S}^\top \\ \boldsymbol{\Upsilon}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1} & \boldsymbol{\Upsilon} & \mathbf{S}^\top \\ \mathbf{S}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{S} & -\mathbf{R} \end{pmatrix},$$

where $\boldsymbol{\Upsilon} = \mathbf{V}\left(\mathbf{I} - \boldsymbol{\Sigma}_{23}^\top \mathbf{S}\right)$, $\mathbf{R} = \left(\boldsymbol{\Sigma}_{23}\mathbf{V}\boldsymbol{\Sigma}_{23}^\top\right)^{-1}$, $\mathbf{S} = \mathbf{R}\boldsymbol{\Sigma}_{23}\mathbf{V}$, and $\mathbf{V} = \left(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\right)^{-1}$. Thus, we have the following

$$\begin{pmatrix} C_n^2 n^{-1} \widetilde{\boldsymbol{\delta}} \\ \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widetilde{\boldsymbol{\nu}} \end{pmatrix} = \boldsymbol{\Sigma}^{-1} \begin{pmatrix} -n^{-1}\sum_{i=1}^n g_i(\boldsymbol{\beta}_0) \\ 0 \\ -H(\boldsymbol{\beta}_0) \end{pmatrix} + o_p(\Delta_n).$$

By this, under the local alternative hypothesis H_1, we could figure out that

$$\Delta_n = \left\| \begin{pmatrix} \widetilde{\boldsymbol{\delta}} \\ \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widetilde{\boldsymbol{\nu}} \end{pmatrix} \right\| \leq \left\| \begin{pmatrix} C_n^2 n^{-1} \widetilde{\boldsymbol{\delta}} \\ \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widetilde{\boldsymbol{\nu}} \end{pmatrix} \right\|$$

$$= \left\| \boldsymbol{\Sigma}^{-1} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \left\{ -\frac{1}{n} \sum_{i=1}^{n} g_i(\boldsymbol{\beta}_0) \right\} - \boldsymbol{\Sigma}^{-1} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} H(\boldsymbol{\beta}_0) + o_p(\Delta_n) \right\|$$

$$\leq O_p\left(n^{-1/2}h^{-1/2}\right) + o_p(\Delta_n),$$

which implies that $\Delta_n = O_p\left(n^{-1/2}h^{-1/2}\right)$.

Combining the above results, we have

$$\begin{pmatrix} C_n^2 n^{-1}\widetilde{\boldsymbol{\delta}} \\ \widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \\ \widetilde{\boldsymbol{\nu}} \end{pmatrix} = \begin{pmatrix} -\boldsymbol{\Sigma}_{11}^{-1} + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Upsilon}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1} \\ \boldsymbol{\Upsilon}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1} \\ \mathbf{S}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1} \end{pmatrix} \left\{ -\frac{1}{n} \sum_{i=1}^{n} g_i(\boldsymbol{\beta}_0) \right\}$$

$$- \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\mathbf{S}^{\top} \\ \mathbf{S}^{\top} \\ -\mathbf{R} \end{pmatrix} H(\boldsymbol{\beta}_0) + o_p\left(n^{-1/2}h^{-1/2}\right). \tag{S5.4}$$

Given the following results $-\mathbf{S}\boldsymbol{\Sigma}_{12}^{\top}\boldsymbol{\Sigma}_{11}^{-1} = -\mathbf{R}\boldsymbol{\Sigma}_{23}\mathbf{V}\mathbf{V}^{-1}\boldsymbol{\Sigma}_{12}^{-1} = -\mathbf{R}\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{12}^{-1}$

we have the asymptotic expression for $\widetilde{\boldsymbol{\nu}}$,

$$\begin{aligned} \widetilde{\boldsymbol{\nu}} &= -\mathbf{S}\boldsymbol{\Sigma}_{12}^{\top}\boldsymbol{\Sigma}_{11}^{-1}\left\{\tfrac{1}{n}\textstyle\sum_{i=1}^{n} g_i(\boldsymbol{\beta}_0)\right\} + \mathbf{R}H(\boldsymbol{\beta}_0) + o_p\left(n^{-1/2}h^{-1/2}\right) \\ &= -\mathbf{R}\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{12}^{-1}\left\{\tfrac{1}{n}\textstyle\sum_{i=1}^{n} g_i(\boldsymbol{\beta}_0)\right\} + \mathbf{R}H(\boldsymbol{\beta}_0) + o_p\left(n^{-1/2}h^{-1/2}\right). \end{aligned} \tag{S5.5}$$

By equation (S5.5), under the null hypothesis $H_0 : H\{\boldsymbol{\beta}_0(z)\} = 0$, we have

$$\widetilde{\boldsymbol{\nu}} = n^{-1}\mathbf{R}^{1/2}\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{12}^{-1}\sum_{i=1}^{n} g_i(\boldsymbol{\beta}_0) + o_p\left(n^{-1/2}h^{-1/2}\right).$$

Since

$$
\begin{aligned}
-\boldsymbol{\Upsilon}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1} &= -\boldsymbol{\Upsilon}\mathbf{V}^{-1}\boldsymbol{\Sigma}_{12}^{-1} = -\mathbf{V}(\mathbf{I}-\boldsymbol{\Sigma}_{23}^{\top}\mathbf{S})\mathbf{V}^{-1}\boldsymbol{\Sigma}_{12}^{-1} \\
&= -\boldsymbol{\Sigma}_{12}^{-1} + \mathbf{V}\boldsymbol{\Sigma}_{23}^{\top}\mathbf{S}\mathbf{V}^{-1}\boldsymbol{\Sigma}_{12}^{-1} = -\boldsymbol{\Sigma}_{12}^{-1} + \mathbf{V}\boldsymbol{\Sigma}_{23}^{\top}\mathbf{R}\boldsymbol{\Sigma}_{23}\mathbf{V}\mathbf{V}^{-1}\boldsymbol{\Sigma}_{12}^{-1} \\
&= -\boldsymbol{\Sigma}_{12}^{-1} + \mathbf{V}\boldsymbol{\Sigma}_{23}^{\top}\mathbf{R}\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{12}^{-1},
\end{aligned}
$$

for the asymptotic expression of $\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$, (S5.4) together with (S5.5) gives

$$
\begin{aligned}
\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 &= \left(-\boldsymbol{\Sigma}_{12}^{-1} + \mathbf{V}\boldsymbol{\Sigma}_{23}^{\top}\mathbf{R}\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{12}^{-1}\right)\left\{\frac{1}{n}\sum_{i=1}^{n} g_i(\boldsymbol{\beta}_0)\right\} + o_p\left(n^{-1/2}h^{-1/2}\right) \\
&= -\boldsymbol{\Sigma}_{12}^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n} g_i(\boldsymbol{\beta}_0)\right\} + \mathbf{V}\boldsymbol{\Sigma}_{23}^{\top}\mathbf{R}\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{12}^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n} g_i(\boldsymbol{\beta}_0)\right\} \\
&\quad + o_p\left(n^{-1/2}h^{-1/2}\right) \\
&= -\boldsymbol{\Sigma}_{12}^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n} g_i(\boldsymbol{\beta}_0)\right\} - \mathbf{V}\boldsymbol{\Sigma}_{23}^{\top}\widetilde{\boldsymbol{\nu}} + o_p\left(n^{-1/2}h^{-1/2}\right).
\end{aligned}
$$

Using the expression of $\widetilde{\boldsymbol{\delta}}$

$$
\begin{aligned}
\widetilde{\boldsymbol{\delta}} &= \left\{n^{-1}\sum_{i=1}^{n} g_i(\boldsymbol{\beta}_0)g_i^{\top}(\boldsymbol{\beta}_0)\right\}^{-1}\left\{n^{-1}\sum_{i=1}^{n} g_i(\boldsymbol{\beta}_0)\right\} \\
&\quad + \left\{n^{-1}\sum_{i=1}^{n} g_i(\boldsymbol{\beta}_0)g_i^{\top}(\boldsymbol{\beta}_0)\right\}^{-1}\left\{n^{-1}\sum_{i=1}^{n} g_i(\boldsymbol{\beta}_0)\frac{[\widetilde{\boldsymbol{\delta}}^{\top}g_i(\boldsymbol{\beta}_0)]^2}{1+\widetilde{\boldsymbol{\delta}}^{\top}g_i(\boldsymbol{\beta}_0)}\right\} \\
&= \left\{n^{-1}\sum_{i=1}^{n} g_i(\boldsymbol{\beta}_0)g_i^{\top}(\boldsymbol{\beta}_0)\right\}^{-1}\left\{n^{-1}\sum_{i=1}^{n} g_i(\boldsymbol{\beta}_0)\right\} + o_p\left(n^{-1/2}h^{-1/2}\right),
\end{aligned}
$$

and the above asymptotic expression for $\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$, the empirical log-likelihood ratio statistic can be written as

$$\begin{aligned}
2\ell(z) &= 2\sum_{i=1}^{n}\widetilde{\boldsymbol{\delta}}^{\top}g_i\left(\widetilde{\boldsymbol{\beta}}\right) - \sum_{i=1}^{n}\widetilde{\boldsymbol{\delta}}^{\top}g_i\left(\widetilde{\boldsymbol{\beta}}\right)g_i^{\top}\left(\widetilde{\boldsymbol{\beta}}\right)\widetilde{\boldsymbol{\delta}} + o_p(1)\\
&= 2n\left\{\frac{1}{n}\sum_{i=1}^{n}g_i^{\top}\left(\widetilde{\boldsymbol{\beta}}\right)\right\}\widetilde{\boldsymbol{\delta}} - n\widetilde{\boldsymbol{\delta}}^{\top}\left\{\frac{1}{n}\sum_{i=1}^{n}g_i\left(\widetilde{\boldsymbol{\beta}}\right)g_i^{\top}\left(\widetilde{\boldsymbol{\beta}}\right)\right\}\widetilde{\boldsymbol{\delta}} + o_p(1)\\
&= 2nh\left\{\frac{1}{n}\sum_{i=1}^{n}g_i^{\top}\left(\widetilde{\boldsymbol{\beta}}\right)\right\}\boldsymbol{\Sigma}_{11}^{-1}(\frac{1}{n}\sum_{i=1}^{n}g_i\left(\widetilde{\boldsymbol{\beta}}\right))\\
&\quad - nh\left\{\frac{1}{n}\sum_{i=1}^{n}g_i\left(\widetilde{\boldsymbol{\beta}}\right)\right\}^{\top}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{11}^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}g_i^{\top}\left(\widetilde{\boldsymbol{\beta}}\right)\right\} + o_p(1)\\
&= nh\left\{\frac{1}{n}\sum_{i=1}^{n}g_i^{\top}\left(\widetilde{\boldsymbol{\beta}}\right)\right\}\boldsymbol{\Sigma}_{11}^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}g_i^{\top}\left(\widetilde{\boldsymbol{\beta}}\right)\right\} + o_p(1)\\
&= nh\widetilde{\boldsymbol{\nu}}^{\top}\boldsymbol{\Sigma}_{23}\mathbf{V}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\mathbf{V}\boldsymbol{\Sigma}_{23}^{\top}\widetilde{\boldsymbol{\nu}} + o_p(1) = nh\widetilde{\boldsymbol{\nu}}^{\top}\mathbf{R}^{-1}\widetilde{\boldsymbol{\nu}} + o_p(1).
\end{aligned}$$

We see that $E\left(\mathbf{R}^{1/2}\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{12}^{-1}\sum_{i=1}^{n}g_i(\boldsymbol{\beta}_0)\right) = 0$ and as $n\to\infty$,

$$\begin{aligned}
C_n^{-1}\mathrm{Var}\left(\mathbf{R}^{1/2}\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{12}^{-1}\sum_{i=1}^{n}g_i(\boldsymbol{\beta}_0)\right) &\to \mathbf{R}^{1/2}\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{12}^{-1}\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{12}^{-1}\boldsymbol{\Sigma}_{23}^{\top}\mathbf{R}^{1/2}\\
&= \mathbf{R}^{1/2}\boldsymbol{\Sigma}_{23}(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{12})^{-1}\boldsymbol{\Sigma}_{23}^{\top}\mathbf{R}^{1/2}\\
&= \mathbf{R}^{1/2}\boldsymbol{\Sigma}_{23}\mathbf{V}\boldsymbol{\Sigma}_{23}^{\top}\mathbf{R}^{1/2}\\
&= \mathbf{R}^{1/2}\mathbf{R}^{-1}\mathbf{R}^{1/2}\\
&= \mathbf{I}_{q\times q}.
\end{aligned}$$

Thus, by the Central Limit Theorem, under the null hypothesis $H_0: H\{\boldsymbol{\beta}_0(z)\} = 0$, we have $n^{-1/2}h^{1/2}\mathbf{R}^{1/2}\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{12}^{-1}\sum_{i=1}^{n}g_i(\boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0},\mathbf{I}_q)$ which means

$$\sqrt{nh}\mathbf{R}^{-1/2}\widetilde{\boldsymbol{\nu}} \xrightarrow{d} N(\mathbf{0},\mathbf{I}_q).$$

Thus, $2\ell(z) \xrightarrow{d} \chi_q^2$. Under local alternative hypothesis $H_1 : H\{\boldsymbol{\beta}_0(z)\} =$

$(nh)^{1/2}\mathbf{d}(z)$, we have

$$\sqrt{nh}\mathbf{R}^{-1/2}\widetilde{\boldsymbol{\nu}} \xrightarrow{d} N(\mathbf{R}^{1/2}\mathbf{d}, \mathbf{I}_q).$$

Thus, $2\ell(z) = nh\widetilde{\boldsymbol{\nu}}^\top\mathbf{R}^{-1}\widetilde{\boldsymbol{\nu}} + o_p(1) \xrightarrow{d} \chi_q^2(\mathbf{d}^\top\mathbf{R}\mathbf{d})$. □

S6 Proof of Theorem 3

Proof. We first prove the asymptotic normality of D_n under the null hypothesis.

We have the decomposition for $D_n = D_{n1} + D_{n2}$, where

$$\begin{aligned}
D_{n1} &= C_n^{-2}\sum_{i=1}^n \int_a^b \boldsymbol{\xi}_i^\top(z)\mathbf{G}^\top(z)\mathbf{G}(z)\boldsymbol{\xi}_i(z)w(z)dz \\
&= C_n^{-2}\sum_{i=1}^n \epsilon_i^2 \int_a^b K_h(Z_i - z)^2\mathbf{X}_i^\top\mathbf{G}^\top(z)\mathbf{G}(z)\mathbf{X}_i w(z)dz, \\
D_{n2} &= C_n^{-2}\sum_{i=1}^n \sum_{k\neq i}^n \int_a^b \boldsymbol{\xi}_i^\top(z)\mathbf{G}^\top(z)\mathbf{G}(z)\boldsymbol{\xi}_k(z)w(z)dz \\
&= C_n^{-2}\sum_{i=1}^n \sum_{k\neq i}^n \epsilon_i\epsilon_k \int_a^b K_h(Z_i - z)K_h(Z_k - z)\mathbf{X}_i^\top\mathbf{G}^\top(z)\mathbf{G}(z)\mathbf{X}_k w(z)dz.
\end{aligned}$$

Note that,

$$\begin{aligned}
E\left\{\mathbf{X}_i^\top\mathbf{G}^\top(z)\mathbf{G}(z)\mathbf{X}_i \middle| Z_i = z\right\} &= tr(\mathbf{G}(z)\boldsymbol{\Omega}(z)\mathbf{G}^\top(z)) \\
&= tr\left[(\sigma^2\mu_{20}f(z))^{-1}\boldsymbol{I}_{q\times q}\right] = q(\sigma^2\mu_{20}f(z))^{-1},
\end{aligned}$$

and $ED_{n2} = 0$. We also have

$$ED_{n1} = h\sigma^2 \int_a^b E\left\{K_h(Z_i - z)^2\mathbf{X}_i^\top\mathbf{G}^\top(z)\mathbf{G}(z)\mathbf{X}_i\right\} w(z)dz$$

$$
\begin{aligned}
&= h\sigma^2 \int_a^b E\left\{E\left[\mathbf{X}_i^\top \mathbf{G}^\top(z)\mathbf{G}(z)\mathbf{X}_i \middle| Z_i\right] K_h(Z_i - z)^2\right\} w(z)dz \\
&= \sigma^2 \int_a^b \left[E\left\{\mathbf{X}_i^\top \mathbf{G}^\top(z)\mathbf{G}(z)\mathbf{X}_i \middle| Z_i = z\right\} f(z)\mu_{20} + O(h^2)\right] w(z)dz \\
&= q + O(h^2).
\end{aligned}
$$

Define $K^{(4)}(x) = \int K^2(y)K^2\left(y - x\right)dy$ and $I_{ik}(z) = \mathbf{X}_i^\top \mathbf{G}^\top(z)\mathbf{G}(z)\mathbf{X}_k$.

Thus, $\mathbf{X}_i^\top \mathbf{G}^\top(z_1)\mathbf{G}(z_1)\mathbf{X}_k\mathbf{X}_{i'}^\top \mathbf{G}^\top(z_2)\mathbf{G}(z_2)\mathbf{X}_{k'} = I_{ik}(z_1) \times I_{i'k'}(z_2)$. When $i \neq k$, we have

$$
\begin{aligned}
& E(I_{ii}(z_1)I_{kk}(z_2) \big| Z_i = z_1, Z_k = z_2) \\
&= E(I_{ii}(z_1) \big| Z_i = z_1)E(I_{kk}(z_2) \big| Z_k = z_2) \\
&= tr(E\left[\mathbf{G}^\top(z_1)\mathbf{X}_i^\top \mathbf{X}_i\mathbf{G}(z_1) \middle| Z_i = z_1\right]) \\
&\quad \times tr(E\left[\mathbf{G}^\top(z_2)\mathbf{X}_k^\top \mathbf{X}_k\mathbf{G}(z_2) \middle| Z_k = z_2\right]) \\
&= q^2\mu_{20}^{-2}\sigma^{-4}f^{-1}(z_1)f^{-1}(z_2).
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
& ED_{n1}^2 \\
&= h^2 n^{-2} \sum_{i=1}^n \sum_{i'=1}^n \int_a^b \int_a^b E\left\{\boldsymbol{\xi}_{i1}^\top \mathbf{G}^\top(z_1)\mathbf{G}(z_1)\boldsymbol{\xi}_{i1}\boldsymbol{\xi}_{i'2}^\top \mathbf{G}^\top(z_2)\mathbf{G}(z_2)\boldsymbol{\xi}_{i'2}\right\} \\
&\quad \times w(z_1)w(z_2)dz_1dz_2 \\
&= \sum_{i' \neq i}^n h^2 n^{-2} \int_a^b \int_a^b E\left\{\epsilon_i^2 \epsilon_{i'}^2 K_h(Z_i - z_1)^2 K_h(Z_{i'} - z_2)^2 I_{ii}(z_1)I_{i'i'}(z_2)\right\} \\
&\quad \times w(z_1)w(z_2)dz_1dz_2
\end{aligned}
$$

$$
\begin{aligned}
& +\sum_{i=1}^{n} h^2 n^{-2} \int_a^b \int_a^b E\left\{\epsilon_i^4 K_h(Z_i-z_1)^2 K_h(Z_i-z_2)^2 I_{ii}(z_1) I_{ii}(z_2)\right\} \\
& \times w(z_1)w(z_2)dz_1dz_2 \\
= \;& \sum_{i'\neq i}^{n} n^{-2}\sigma^4 \int_a^b \int_a^b \left\{E(I_{ii}(z_1)\times I_{kk}(z_2)\big|Z_i=z_1, Z_k=z_2)f(z_1)f(z_2)\mu_{20}^2\right. \\
& \left.+O(h^2)\right\} \times w(z_1)w(z_2)dz_1dz_2 \\
& +\sum_{i=1}^{n} n^{-2}E\epsilon_i^4 \int_a^b \int_a^b \left\{hE(I_{ii}(z_1)\times I_{ii}(z_2)\big|Z_i=z_1\right) f(z_1)K^{(4)}\left(\frac{z_2-z_1}{h}\right) \\
& \left.+o(h^2)\right\} \times w(z_1)w(z_2)dz_1dz_2 \\
\asymp \;& q^2 + O\left(h^2\right).
\end{aligned}
$$

Next, we calculate ED_{n2}^2. Note that

$$
\begin{aligned}
& ED_{n2}^2 \\
= \;& h^2n^{-2}\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{k=1}^{n}\sum_{k'=1}^{n}\int_a^b\int_a^b E\left\{\boldsymbol{\xi}_{i1}^\top \mathbf{G}^\top(z_1)\mathbf{G}(z_1)\boldsymbol{\xi}_{k1}\boldsymbol{\xi}_{i'2}^\top \mathbf{G}^\top(z_2)\mathbf{G}(z_2)\boldsymbol{\xi}_{k'2}\right\} \\
& \times w(z_1)w(z_2)dz_1dz_2.
\end{aligned}
$$

When $k \neq i$ and $k' \neq i'$, $E\left(\epsilon_i\epsilon_{i'}\epsilon_k\epsilon_{k'}\right) \neq 0$ only in two cases, the first one is $\{i=i', k=k'\}$, and the second one is $\{i=k', k=i'\}$. In particularly, we have $ED_{n2}^2 = h^2n^{-2}\sigma^4\sum_{i=1}^n\sum_{k\neq i}^n\int_a^b\int_a^b(\Pi_1+\Pi_2)w(z_1)w(z_2)dz_1dz_2$, where

$$
\begin{aligned}
\Pi_1 \;=\;& E\left\{I_{ik}(z_1)I_{ik}(z_2)K_h(Z_i-z_1)K_h(Z_k-z_1)K_h(Z_i-z_2)K_h(Z_k-z_2)\right\} \\
=\;& E\left\{E\left[I_{ik}(z_1)I_{ik}(z_2)\big|Z_i, Z_k\right]K_h(Z_i-z_1)K_h(Z_k-z_1)\right. \\
& \left.K_h(Z_i-z_2)K_h(Z_k-z_2)\right\} \\
=\;& \int_a^b\int_a^b E\left[I_{ik}(z_1)I_{ik}(z_2)\big|Z_i=x, Z_k=y\right]f(x)f(y)
\end{aligned}
$$

$$
\begin{aligned}
& \times K_h(x-z_1)K_h(x-z_2)K_h(y-z_1)K_h(y-z_2)dxdy \\
= & \int_a^b \left\{ h^{-1}E\left[I_{ik}(z_1)I_{ik}(z_2)\middle|Z_i=z_1,Z_k=y\right] f(z_1)K^{(2)}\left(\frac{z_2-z_1}{h}\right) \right. \\
& \left. +O(h)\right\} \times f(y)K_h(y-z_1)K_h(y-z_2)dy \\
= & \ h^{-2}E\left[I_{ik}(z_1)I_{ik}(z_2)\middle|Z_i=z_1,Z_k=z_2\right] f(z_1)f(z_2)\left\{K^{(2)}\left(\frac{z_2-z_1}{h}\right)\right\}^2 \\
& +O(1)
\end{aligned}
$$

and

$$
\begin{aligned}
\Pi_2 = & \ E\left\{I_{ik}(z_1)I_{ki}(z_2)K_h(Z_i-z_1)K_h(Z_k-z_1)K_h(Z_i-z_2)K_h(Z_k-z_2)\right\} \\
= & \ E\left\{E\left[I_{ik}(z_1)I_{ki}(z_2)\middle|Z_i,Z_k\right] K_h(Z_i-z_1)K_h(Z_k-z_1)K_h(Z_i-z_2)\right. \\
& \left. \times K_h(Z_k-z_2)\right\} \\
= & \int_a^b\int_a^b E\left[I_{ik}(z_1)I_{ki}(z_2)\middle|Z_i=x,Z_k=y\right] \\
& \times f(x)f(y)K_h(x-z_1)K_h(x-z_2)K_h(y-z_1)K_h(y-z_2)dxdy
\end{aligned}
$$

$$
\begin{aligned}
= & \int_a^b \left\{ h^{-1}E\left[I_{ik}(z_1)I_{ki}(z_2)\middle|Z_i=z_1,Z_k=y\right] f(z_1)K^{(2)}\left(\frac{z_2-z_1}{h}\right) + O(h)\right\} \\
& \times f(y)K_h(y-z_1)K_h(y-z_2)dy \\
= & \ h^{-2}E\left[I_{ik}(z_1)I_{ki}(z_2)\middle|Z_i=z_1,Z_k=z_2\right] f(z_1)f(z_2)\left\{K^{(2)}\left(\frac{z_2-z_1}{h}\right)\right\}^2 \\
& +O(1).
\end{aligned}
$$

Since

$$
E\left[I_{ik}(z_1)I_{ik}(z_2)\middle|Z_i=z_1,Z_k=z_2\right]
$$

$$
\begin{aligned}
&= E\left\{E\left[\mathbf{X}_i^\top \mathbf{G}^\top(z_1)\mathbf{G}(z_1)\mathbf{X}_k\mathbf{X}_i^\top \mathbf{G}^\top(z_2)\mathbf{G}(z_2)\mathbf{X}_k\right]|Z_i=z_1, Z_k=z_2\right\}\\
&= E\left\{\mathbf{X}_k^\top \mathbf{G}^\top(z_1)\mathbf{G}(z_1)E\left[\mathbf{X}_i\mathbf{X}_i^\top\right]\mathbf{G}^\top(z_2)\mathbf{G}(z_2)\mathbf{X}_k|Z_i=z_1, Z_k=z_2\right\}\\
&= tr\left\{E\left[\mathbf{X}_k^\top \mathbf{G}^\top(z_1)\mathbf{G}(z_1)E\left(\mathbf{X}_i\mathbf{X}_i^\top\right)\mathbf{G}^\top(z_2)\mathbf{G}(z_2)\mathbf{X}_k|Z_i=z_1, Z_k=z_2\right]\right\}\\
&= tr\left\{E\left[\mathbf{G}(z_1)E\left(\mathbf{X}_i\mathbf{X}_i^\top\right)\mathbf{G}^\top(z_2)\mathbf{G}(z_2)\mathbf{X}_k\mathbf{X}_k^\top\mathbf{G}^\top(z_1)|Z_i=z_1, Z_k=z_2\right]\right\}\\
&= \mathbf{G}(z_1)E\left(\mathbf{X}_i\mathbf{X}_i^\top|Z_i=z_1\right)\mathbf{G}^\top(z_2)\mathbf{G}(z_2)E\left(\mathbf{X}_k\mathbf{X}_k^\top|Z_k=z_2\right)\mathbf{G}^\top(z_1)\\
&= \mathbf{G}(z_1)\mathbf{\Omega}\left(z_1\right)\mathbf{G}^\top(z_2)\mathbf{G}(z_2)\mathbf{\Omega}\left(z_2\right)\mathbf{G}^\top(z_1)\asymp q,
\end{aligned}
$$

and similarly, we have $E\left[I_{ik}(z_1)I_{ki}(z_2)\middle|Z_i=z_1, Z_k=z_2\right]\asymp q$. Combining the results for Π_1 and Π_2, we have

$$
\begin{aligned}
ED_{n2}^2 &\asymp \sigma^4 h\int\int f(v)f(v+hu)w(v)w(v+hu)\left(K^{(2)}\left(u\right)\right)^2 dudv\\
&= h\int\left(K^{(2)}\left(u\right)\right)^2 du\int f^2(v)w^2(v)dv+O(h^2),
\end{aligned}
$$

Hence we have $Var(D_{n1})=o(Var(D_{n2}))$, and it follows that

$$
D_n-E(D_n)=D_{n2}\left\{1+o(1)\right\}.
$$

We can write D_{n2} as $D_{n2}=\frac{1}{n}\sum_{i\neq k}^n\int_a^b \mathcal{Z}_i^\top(z)\mathcal{Z}_k(z)w(z)dz$, where $\mathcal{Z}_i(z)=\sqrt{h}\mathbf{G}(z)\boldsymbol{\xi}_i(z)$. Let $\mathcal{U}_n=\frac{1}{n-1}D_{n2}=\frac{2}{n(n-1)}\sum_{1\leq i<k\leq n}\mathcal{K}(\mathcal{Z}_i,\mathcal{Z}_k)$, where $\mathcal{K}(\mathcal{Z}_i,\mathcal{Z}_k)$ $=\int_a^b \mathcal{Z}_i^\top(z)\mathcal{Z}_k(z)w(z)dz$. Define $A_{\mathcal{K}}$ as $A_{\mathcal{K}}g(x)=\int_\infty^\infty \mathcal{K}(x,y)g(y)dF(y)$, where F is the distribution of $\mathcal{Z}_i$. Then we have the associated eigenvalues and eigenfunctions, denoted as $\{\lambda_k,\psi_k\}_{k=1}^\infty$. The remaining proof for D_n under the null hypothesis test is analogous to the sparse case in Theorem 2 and Corollary 1 in Wang et al. (2018).

Secondly, we prove the asymptotic distribution for D_n under alternative hypothesis. Notice that $2\ell(z) = nh\widetilde{\boldsymbol{\nu}}^\top \mathbf{R}^{-1}\widetilde{\boldsymbol{\nu}} + o_p(1)$ as shown in Theorem 2 and by (S5.5) under local alternative,

$$\widetilde{\boldsymbol{\nu}} = -\mathbf{R}\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{12}^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n} g_i(\boldsymbol{\beta}_0)\right\} + \mathbf{R}H(\boldsymbol{\beta}_0) + o_p\left(n^{-1/2}h^{-1/2}\right).$$

The remaining proof is the same as the sparse case in Theorem 3 in Wang et al. (2018). □

Bibliography

Lai, M. and L. Wang (2013). Bivariate penalized splines for regression. *Statistica Sinica 23*, 1399–1417.

Lai, M.-J. and L. L. Schumaker (2007). *Spline functions on triangulations*, Volume 110 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge.

Wang, H., P.-S. Zhong, Y. Cui, and Y. Li (2018). Unified empirical likelihood ratio tests for functional concurrent linear models and the phase transition from sparse to dense functional data. *Journal of the Royal Statistical Society. Series B. 80*(2), 343–364.

Wang, L., G. Wang, M. Lai, and L. Gao (2020). Efficient estimation of partially linear models for data on complicated domains by bivariate penalized splines over triangulations. *Statistica Sinica 30*, 347–369.

Xue, L. and L. Zhu (2007). Empirical likelihood for a varying coefficient model with longitudinal data. *Journal of the American Statistical Association 102*(478), 642–654.

Yu, S., G. Wang, L. Wang, C. Liu, and L. Yang (2020). Estimation and inference for generalized geoadditive models. *J. Amer. Statist. Assoc. 115*(530), 761–774.