Received: 24 February 2023 | Revised: 11 November 2023 | Accepted: 25 February 2024

DOI: 10.1111/sjos.12717

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Empirical likelihood M-estimation for the varying-coefficient model with functional response

Xingcai Zhou[1] | **Dehan Kong**[2] | **Matthew Stephen Pietrosanu**[3] | **Linglong Kong**[3] | **Rohana J. Karunamuni**[3]

[1]School of Statistics and Data Science, Nanjing Audit University, Nanjing, China

[2]Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada

[3]Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada

Correspondence
Linglong Kong, Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB T6G 2G1, Canada.
Email: lkong@ualberta.ca

Funding information
Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: RGPIN-2018-04486; National Natural Science Fund of China, Grant/Award Number: 12171242

Abstract

This work is motivated by a gap in the functional data analysis literature, particularly in the context of neuroimaging, regarding the ability of functional models to robustly accommodate intra-observation dependence. In response, we propose an M-estimator based on generalized empirical likelihood for the varying-coefficient model with a functional response. We develop statistical inference procedures, simultaneous confidence regions, and a global general linear hypothesis test for the model's functional coefficient. Our theoretical results establish the weak convergence of the log-likelihood ratio process, a nonparametric version of Wilks' theorem for the log-likelihood ratio, and asymptotic properties of the proposed estimator. Through a simulation study, we show that the proposed confidence sets have close-to-nominal coverage probabilities. In a real-world application to a neuroimaging dataset, we show that mini-mental state examination score and apolipoprotein E genotype have significant associations with fractional anisotropy, while associations with gender and age are only present at high quantile levels.

KEYWORDS

functional data, global test, robust regression, simultaneous confidence band, weak convergence

1 | INTRODUCTION

Functional data measure a response as a curve or surface over a spatiotemporal region. This type of data is prevalent in neuroimaging through electroencephalography (EEG) and other modalities in the study of brain connectivity and neurodegeneration. Unsurprisingly, the varying-coefficient regression model, which estimates associations between a repeatedly measured response and covariates with a nonconstant effect (Zhu et al., 2012), has become a staple in this field. Since its introduction in Hastie and Tibshirani (1993), the model has been adapted for time series and longitudinal data (e.g., Wang et al., 2008), but has only recently been applied to functional data (e.g., Zhang & Chen, 2007; Zhu et al., 2012). Consequently, gaps exist in the literature pertaining to robustness and intra-observation dependence structures. This deficiency is a hindrance in neuroimaging data analysis, where sample sizes are small, imaging artifacts are common, and spatiotemporal correlations have significant impacts on statistical power. We are particularly interested in delineating the structure of the variability of the functional response and their association with a set of covariates of interest. In the real data analysis, we have applied our developed method to diffusion tensor imaging dataset from the Alzheimer's Disease Neuroimaging Initiative, and mined some valuable information, which is beneficial for the prevention and treatment of such diseases.

In this work, we focus on robust estimation and inference in the presence of intra-observation dependence for the varying-coefficient model

$$y_i(s) = x_i^\top \beta(s) + v_i(s) + \varepsilon_i(s), \quad i = 1, \ldots, n, \tag{1}$$

where s is a spatiotemporal location, y_i is a functional response, x_i is a p-vector of covariates, β is a vector-valued function of covariate effects, v_i is an observation-specific effect; and ε_i is a stochastic error. We assume that v_i and ε_i are independent stochastic processes (SP) and, respectively, copies of the low-frequency signal process $\mathrm{SP}(0, \gamma_v)$ and the noise process $\mathrm{SP}(0, \gamma_\varepsilon)$ with $\gamma_\varepsilon(s,t) = \sigma^2(s) I(s = t)$ for $s \in [0,1]$. Here, we do not need the distributions of the two stochastic processes to be symmetric. Our empirical likelihood M-estimation is suitable for outliers and symmetric/asymmetric error distributions. As typical in neuroimaging, we assume that every y_i is observed at the same set of s, namely, $(s_j)_{j=1}^m$. Similar settings to the model (1) have also been used for various other models. For example, multivariate varying coefficient model for functional responses (Zhu et al., 2012), single-index varying coefficient model for functional responses (Luo et al., 2016), functional varying-coefficient single index model for functional response data (Li et al., 2017), functional response quantile regression models (Zhou et al., 2023), and so on.

A few works have considered model (1) for functional data. Zhang and Chen (2007) estimated β with local polynomial kernel smoothing and used an L_2-based hypothesis-testing procedure for inference. Zhu et al. (2012) established the convergence of a local linear estimator and derived corresponding simultaneous confidence bands. Both of these approaches, however, are based on least-squares and are not robust to outliers or heavy-tailed error distributions. For longitudinal data, You et al. (2006) and Wang et al. (2010) used empirical likelihood to accommodate inter- and intra-observation dependence. This nonparametric, data-driven approach has numerous advantages over methods using generalized least squares or normal approximations. However, confidence sets based on empirical likelihood that does not account for robustness are sensitive to outliers (Owen, 2001, Chap. 2). Tang and Leng (2011) proposed inference procedures based on

quantile regression, which is known to be efficient in settings with heavy-tailed error and robust to outliers (Koenker & Bassett, 1978). Nonetheless, empirical likelihood remains nearly absent in functional data analysis: the unified tests in Wang et al. (2018) are the only relevant developments. Robust methods for estimation and inference do not yet exist in the literature.

We propose an M-estimation framework for model (1) that accommodates intra-observation dependence and naturally includes mean regression, quantile regression, and other robust approaches as special cases. The intra-observation dependence among the observations within the same subject, which is an inherent characteristic of function data, is introduced to improve the efficiency of statistical inference. For the estimation of models, many important works on functional data estimate the coefficient functions of model by kernel methods based on repeated measurement data, not considering the within-curve dependence (Li et al., 2017; Luo et al., 2016; Zhou et al., 2023; Zhu et al., 2012). Based on this simple estimation strategy, their estimators of varying coefficient functions can achieve $\sqrt{n}$-consistency and asymptotic normal distribution under the suitable $m = m(n) \to \infty$ as $n \to \infty$. For statistical inference such as the pointwise/simultaneous confidence regions and the local/global hypothesis tests, Zhu et al. (2012), Luo et al. (2016), Li et al. (2017) and Zhou et al. (2023) all applied the within-curve dependent structure of functional data, because ignoring the structure of the data may lead to the loss of efficiency in inference (Wang et al., 2010). In addition, for longitudinal/functional data, some distinguished works also gave the naive local linear estimation of the repeated measurement data without considering the within-curve dependence. For instance, Yao (2007) considered asymptotic distributions of nonparametric regression at $\sqrt{nmh^{\nu}}$-consistent rate, where $h = h(n, m)$ is bandwidth and the order $\nu > 0$; Li and Hsing (2010) gave uniform convergence rates for nonparametric regression in a unified framework, and $\sqrt{n}$-rates of convergence can be achieved if m is at least of the order $(n/\log n)^{1/4}$ for functional data.

Further, we establish the weak convergence of the log-likelihood ratio process, a nonparametric version of Wilks' theorem for the empirical log-likelihood ratio, and asymptotic properties of the proposed estimator. Notably, our $\hat{\beta}(s)$ also is $\sqrt{n}$-consistent and asymptotically Gaussian neither on m nor on bandwidths h, h_1, and h_2. In contrast, our proposed empirical likelihood M-estimation using local kernel smoothing avoids a stringent assumption of Zhang and Chen (2007) such that $n = O(m^{4/5})$ on the numbers of grid points and subjects. Readers are referred to Condition A.4 and Theorem 4 in Zhang and Chen (2007) for more details. Moreover, our m can be less than of the order $(n/\log n)^{1/4}$ with proper choices of bandwidths by Theorems 1 and 2. For example, we have the order $m = O((n/\log n)^{1/5})$, if take $h_2 = O((n/\log n)^{-1/6})$ and $q_3 = 7$ which satisfies Condition 9(c). This requirement for m is loose relative to the order $(n/\log n)^{1/4}$ of Li and Hsing (2010). Similar conclusions were also reached in Zhu et al. (2012), Luo et al. (2016) and Li et al. (2017) for the naive local linear kernel estimation approach. We leverage the latter to propose confidence sets and inference procedures for β. The paper is devoted to developing statistical inference procedures of simultaneous confidence regions and a global general linear hypothesis test for the model's functional coefficient via empirical likelihood M-estimation; meanwhile, the intra-observation dependence of functional data is accommodated in estimation and inference, rather than simply regarding observations as repeated measurement data.

This paper is organized as follows. In Section 2, We give the empirical likelihood M-estimation methods, including MELE, the ELL ratio construct, estimation of within-curve dependence and bandwidth selection. The theory of the estimator and the global test statistic are established in Section 3. In Section 4, we give a global general linear hypothesis test and simultaneous confidence bands/regions. Their proofs are given in Appendix. A Monte Carlo simulation study is

presented in Section 5. In Section 6, we illustrate the estimation and inference procedures on MELEs of varying coefficient models to diffusion tensor imaging data.

2 | ESTIMATION

2.1 | M-estimation

Under model (1), $y_i(s_j) = x_i^\top \beta(s_j) + \eta_i(s_j)$, where $\eta_i = v_i + \varepsilon_i$. Let $K(s)$ be a kernel function and h a bandwidth, and define $\tilde{K}_j^h(s) = h^{-1}K_j^h(s)$ with $K_j^h(s) = K((s_j - s)/h)$. Define $\hat{b}$ as an M-estimator of β, namely, where $\hat{b}(s)$ minimizes $\sum_{i=1}^n \sum_{j=1}^m \rho(y_i(s_j) - x_i^\top b(s))K_j^h(s)$ for some given convex function ρ. Equivalently, $\hat{b}(s)$ is a solution to the estimating equation

$$\sum_{i=1}^{n}\sum_{j=1}^{m} x_i \psi(y_i(s_j) - x_i^\top b(s))K_j^h(s) = 0,$$

for a suitable choice of ψ. Least-squares estimators, least-absolute-deviation estimators, and quantile estimators are included in this framework through the choice of ψ (Hample et al., 1986, Chap. 2.3a).

2.2 | Generalized empirical likelihood M-estimation

The covariance structure of η_i satisfies $\gamma_\eta(s,t) = \mathrm{cov}(\eta_i(s), \eta_i(t)) = \gamma_v(s,t) + \sigma^2(t)I(s=t)$. Thus, $\Gamma = (\gamma_\eta(s_k, s_l))_{k,l=1}^m$ is the covariance matrix for the observation vector $y_i = (y_i(s_j))_{j=1}^m$. We assume that Γ is positive definite. Let $X_i = (x_i, \ldots, x_i)^\top$ be an $m \times p$ matrix and define $D_K^h(s) = \mathrm{diag}(K_1^h(s), \ldots, K_m^h(s))$ and $\Psi_i(b) = (\psi(y_i(s_j) - x_i^\top b(s_j)))_{j=1}^m$. Define $(u_{i,j}(b,s))_{j=1}^m = \Gamma^{-1}D_K^h(s)\Psi_i(b)$ as an auxiliary random vector and $U_{i,j}(b,s) = (mh)^{-1}x_i u_{i,j}(b,s)$: we suppress the explicit dependence on h for convenience. To accommodate within-curve dependence, a defining feature of functional data, we consider estimators that satisfy the generalized estimating equation $\sum_{i=1}^n \sum_{j=1}^m U_{i,j}(b,s) = 0$. With the well-known relationship between empirical likelihood and estimating equations (Qin & Lawless, 1994) in mind, the corresponding empirical log-likelihood ratio function is

$$L(b,s) = -2\sup_{P\in\mathcal{P}}\left\{\sum_{i=1}^{n}\sum_{j=1}^{m}\log(nmp_{i,j}) \Big| \sum_{i=1}^{n}\sum_{j=1}^{m} p_{i,j}U_{i,j}(b,s) = 0\right\}, \tag{2}$$

where $\mathcal{P}$ is the set of $n \times m$ matrices $(p_{i,j})_{i,j}$ satisfying $p_{i,j} \geq 0$ and $\sum_{i=1}^n \sum_{j=1}^m p_{i,j} = 1$. Then the optimal weights $\{\hat{p}_{ij}\}$ are obtained by the Lagrange multiplier method from (2): $\hat{p}_{ij} = 1/\left[nm(1 + \lambda^T U_{ij}(b,s))\right]$ (See Subsec. 2.9 and 3.14 in Owen, 2001), and the function $L(b,s)$ can be written as

$$L(b,s) = -2\sum_{i=1}^{n}\sum_{j=1}^{m}\log(nm\hat{p}_{ij}) = 2\sum_{i=1}^{n}\sum_{j=1}^{m}\log\left(1 + \lambda^T U_{ij}(b,s)\right), \tag{3}$$

where λ satisfies $\frac{1}{nm}\sum_{i=1}^n \sum_{j=1}^m U_{ij}(b,s)/\left\{1 + \lambda^T U_{ij}(b,s)\right\} = 0$. We propose the maximum empirical likelihood estimator $\tilde{\beta}$ via $\tilde{\beta}(s) = \arg\min_b \ L(b,s)$.

In statistical inference of Section 4, the covariance matrix $\Gamma = (\gamma_\eta(s_k, s_l))_{k,l=1}^m$ in $L(b,s)$ needs to be estimated, which is given in Section 2.3. In Theorem 2 of Section 3.1, we show that the $\gamma_\eta(s,t)$ and its estimator $\hat{\gamma}_\eta(s,t)$ is uniformly consistent. Correspondingly, we have the $\hat{L}(b,s)$, and

$$\hat{\beta}(s) = \arg\min_b \ \hat{L}(b,s). \tag{4}$$

Remark The above approach takes an elementwise approach by using a distinct weight $p_{i,j}$ for each $y_i(s_j)$. A subjectwise approach with a single weight p_i for each y_i yields similar theoretical guarantees and empirical performance, so we do not discuss it further in this article.

Remark The covariance matrix Γ is a $m \times m$ matrix. If m is very large, the inverse of Γ has a high complexity of the order $O(m^3)$ on computation. Fortunately, we find that our model estimation and inference do not require very large m, which also is consistent with our theoretical results. By Theorems 1 and 2, a $\sqrt{n}$-consistency of our $\hat{\beta}$ can be achieved that the m can be less than of the order $(n/\log n)^{1/4}$ with proper choices of bandwidths, which implies that the m can be not too large. In practice, if m is indeed very large, we can reduce the number of grid points by some techniques, such as "smoothing first, then estimation" to get the number of data points available for empirical likelihood M-estimation. This has little impact on estimation and inference, since our theoretical results do not require too large m.

2.3 | Intra-observation dependence

We now wish to obtain a consistent estimator of Γ. Denote $\tilde{K}_j^{h_l}(s) = h_l^{-1}K((s_j - s)/h_l)$ for $j = 1,\ldots,m$ and $l = 1,2,3$. Define $z_{i,j}^{h_1}(s) = \left(x_i^\top, x_i^\top(s_j - s)/h_1\right)^\top$ for some bandwidth h_1. Take $\tilde{\beta}(s) = ((1,0) \otimes I_p)\tilde{a}(s)$ as a preliminary estimate of $\beta(s)$, where $\tilde{a}(s)$ minimizes the weighted least squares criterion $\sum_{i=1}^n \sum_{j=1}^m (y_i(s_j) - z_{i,j}^{h_1}(s)^\top a)^2 \tilde{K}_j^{h_1}(s)$ over a. Under some smoothness assumptions, we can estimate $v_i(s_j)$ through a Taylor expansion at s as $v_i(s_j) \approx d_i(s)^\top z_j^{h_2}(s)$, where $d_i(s) = \left(v_i(s), h_2 v_i(s)^\top\right)^\top$, $z_j^{h_2}(s) = (1, (s_j - s)/h_2)^\top$, and h_2 is a bandwidth. A weighted least-squares estimate of $d_i(s)$ is

$$\hat{d}_i(s) = \left(\sum_{j=1}^m \tilde{K}_j^{h_2}(s) z_j^{h_2}(s) z_j^{h_2}(s)^\top\right)^{-1} \sum_{j=1}^m \tilde{K}_j^{h_2}(s) z_j^{h_2}(s)(y_i(s_j) - x_i^\top \tilde{\beta}(s_j)).$$

We can then estimate $\gamma_v(s,t)$ via the empirical covariance of $\hat{v}_i(s) = (1,0)\hat{d}_i(s)$ as $\hat{\gamma}_v(s,t) = (n-p)^{-1}\sum_{i=1}^n \hat{v}_i(s)\hat{v}_i(t)$. A kernel estimator of the measurement error variance (Fan & Yao, 1998) is

$$\hat{\sigma}^2(s) = \left(n\sum_{j=1}^m \tilde{K}_j^{h_3}(s)\right)^{-1} \sum_{i=1}^n \sum_{j=1}^m \tilde{K}_j^{h_3}(s)\hat{\varepsilon}_i(s_j)^2,$$

where $\hat{\varepsilon}_i(s_j) = y_i(s_j) - x_i^\top \tilde{\beta}(s_j) - \hat{v}_i(s_j)$ and h_3 is a bandwidth. The above results suggest the covariance estimator $\hat{\gamma}_\eta(s,t) = \hat{\gamma}_v(s,t) + \hat{\sigma}^2(s)I(s=t)$. In Section 3, we show that $\hat{\gamma}_\eta(s,t)$ converges uniformly in probability to $\gamma_\eta(s,t)$ so that $\hat{\Gamma} = \left(\hat{\gamma}_\eta(s_k, s_l)\right)_{k,l=1}^m$ is a consistent estimator of Γ.

For the sake of computational efficiency, we implement a different estimator of Γ based on the spectral decomposition (Mercer, 1909) $\gamma_\eta(s,t) = \sum_{k=1}^\infty \lambda_k \phi_k(s)\phi_k(t)$, where $(\lambda_k)_{k=1}^\infty$ is a sequence of nonincreasing eigenvalues and $(\phi_k)_{k=1}^\infty$ is a sequence of corresponding orthogonal eigenfunctions

in $L_2[0,1]$. We estimate Γ by truncating the spectral decomposition above to k_0 terms that explain at least 85% of the variation in the data and then estimate $(\lambda_k)_{k=1}^{k_0}$ and $(\phi_k)_{k=1}^{k_0}$. For additional details, see Rice and Silverman (1991).

2.4 | Bandwidth

The choice of bandwidth is an important consideration for the proposed estimators, as in other methods for kernel smoothing. We adopt the tuning procedure in Zhu et al. (2012) to select h_1 and h_2. For h, we use the standard leave-one-curve-out cross-validation (CV) criterion

$$\mathrm{CV}(h) = (nm)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}\rho(y_i(s_j) - x_i^\top \hat{\beta}_{(-i)}^h(s_j)),$$

where $\hat{\beta}_{(-i)}^h$ is the maximum empirical likelihood estimator of β obtained using the bandwidth h after removing data from the ith observation. This procedure has previously been used for empirical likelihood estimators in varying coefficient models for longitudinal data (Xue & Zhu, 2007).

For h_3, we use a least-squares cross-validation (LSCV) procedure (Hall et al., 2004). In short, we choose h_3 to minimize

$$\mathrm{LSCV}(h_3) = \int_0^1 \hat{\sigma}(s)^2 \mathrm{d}s - 2\int_0^1 \hat{\sigma}^2(s)\tilde{\sigma}^2(s)\mathrm{d}s,$$

where

$$\tilde{\sigma}^2(s) = \left\{n(n-1)\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)\right\}^{-1}\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n}\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)(y_k(s_j) - x_k^\top\hat{\beta}_{(-i)}^h(s_j) - \hat{\nu}_k(s_j))^2,$$

is a leave-one-curve-out estimate of $\sigma^2(s)$. The criterion above can be derived from the integrated square error $\int_0^1(\hat{\sigma}^2(s) - \sigma^2(s))\mathrm{d}s$ after replacing σ^2 with $\tilde{\sigma}^2$.

3 | THEORY

3.1 | Estimator properties

We use the following conditions to establish the theoretical properties of the proposed estimator, which in turn support the inference procedures proposed in the following section. In particular, Theorem 2 implies that $\hat{\sigma}^2$ is a uniformly consistent estimator of σ^2 and, together with Theorem 3(i) of Zhu et al. (2012), that $\hat{\Gamma} = (\hat{\gamma}_\eta(s_k, s_l))_{k,l=1}^m$ is a consistent estimator of Γ.

Condition 1. Let CH denote the convex hull of a set. As $n, m \to \infty$, $pr(0 \in \mathrm{CH}\{U_{i,j}(\beta, s) | s \in [0,1]; i = 1,\ldots,n; j = 1,\ldots,m\}) \to 1$,

Condition 2. The x_is (over $i = 1,\ldots,n$) are independent and identically distributed copies of a random variable x. Additionally, $\|x\|_\infty = O_p(1)$ and $\Omega = E(xx^\top)$ is positive definite.

Condition 3. The functions β_l $(l = 1,\ldots,p)$ and η have continuous derivatives on $[0,1]$. The processes $(\psi(\eta(s)))_{s\in[0,1]}$ and $(x\psi(\eta(s)))_{s\in[0,1]}$ are Donsker classes. Moreover, $E[\sup_{s\in[0,1]}|v|^{r_1}] < \infty$ and $E\{\sup_{s\in[0,1]}[|v'(s)| + |v''(s)|]^{r_2}\} < \infty$ for some $r_1, r_2 \in (2,\infty)$.

Condition 4. The design points $S = \{s_j | j = 1,\ldots,m\}$ are sampled from a density function π that has a continuous derivative and a support of $[0,1]$.

Condition 5. The kernel function K is a Lipschitz-continuous, symmetric density function supported on $[-1,1]$.

Condition 6. The function ψ is a subgradient of ρ. For any $s \in [0,1]$, the distribution function of $\eta_i(s)$ vanishes at the discontinuity points of ψ.

Condition 7. Where $|s_j - s| = o_p(1)$ and $\sup_{s\in[0,1]}|u(s_j,s)| = o_p(1)$ for a function u, there exists a function A_1 such that $E(\psi(\eta(s_j) + u(s_j,s))|s_j) = A_1(s_j)u(s_j,s) + o_p(u(s_j,s))$ and $A_1(s+ht) = A_1(s) + O(ht)$ when $h \to 0$ and $|t| \le 1$. Additionally, $E(\psi(\eta(s_j) + u(s_j,s)) - \psi(\eta(s_j))|s_j)^2 = O_p(A_2(|u(s_j,s)|))$ holds uniformly over $s_j \in S$ for some function A_2 that is continuous at $u(s_j,s)$ and satisfies $A_2(0) = 0$. Lastly, $E(\psi(\eta(s))^2)$ is continuous and strictly positive for $s \in [0,1]$, and $E(\sup_{s\in[0,1]}\psi(\eta(s))^2) < \infty$.

Condition 8. Define $\omega_\psi(s,t) = E(\psi(\eta(s))\psi(\eta(t)))$ and $\Gamma^{-1} = (\varpi(s_k,s_l))_{k,l=1}^m$. There exist functions κ_1 and κ_2 satisfying $\sum_{j=1}^m \varpi(s_j,s)^k = \kappa_k(s)(1 + o_p(1))$ uniformly for $s \in [0,1]$ and $\sup_{s\in[0,1]}|\kappa_k(s)| < \infty$ for $k = 1,2$.

Condition 9. Assume the asymptotic setting with $n \to \infty$, $m \to \infty$, (a) $h \to 0$, $mh \to \infty$, and $h^{-1}|\log h|^{1-2/q_1} \le m^{1-2/q_1}$ for some $q_1 \in (2,4)$; (b) $h_1 \to 0$, $mh_1 \to \infty$ and $h_1^{-1}|\log h_1|^{1-2/q_2} \le m^{1-2/q_2}$ for some $q_2 \in (2,4)$; (c) $h_2 \to 0$, $mh_2 \to \infty$ and $h_2^{-4}(\log n/n)^{1-2/q_3} \to 0$ for $q_3 \in (2,\infty)$.

Condition 10. On $[0,1]$, σ^2 has a continuous derivative and $\mathrm{var}(\varepsilon(s)^2) < \infty$ for all $s \in [0,1]$.

Condition 11. On $[0,1]^2$, $\tilde{\omega}_\psi(s,t) = \omega_\psi(s,s)^{-1/2}\omega_\psi(t,t)^{-1/2}\omega_\psi(s,t)$ has a finite trace.

Remark Condition 1 is a conventional assumption for empirical likelihood (Owen, 2001; Wang et al., 2010). Condition 3 guarantees that $\beta(s_j)$ can be approximated by $\beta(s)$ when $|s_j - s| = o_p(1)$, and the conditions on v are similar to C2 and C9 in Zhu et al. (2012). Conditions 2,4–6, and 9 are used in Zhu et al. (2012). Conditions 6 and 7 are mild in robust regression (Huber & Ronchetti, 2009, Chap. 7). Also see Bai et al. (1992), Fan et al. (1994), and Huber and Ronchetti (2009). For example, the LSE case of $\psi(u) = 2u$, $A_1(s_j) = 1$ and $\omega_\psi(s) = 4\gamma_\eta(s,s)$; the LADE case of $\psi(u) = |u|$, then $A_1(s_j) = 2f_\eta(0;s_j)$ and $\omega_\psi(s) = \gamma_\eta(s,s)$, where $f_\eta(0;s_j) = F'_\eta(0;s_j)$ and $F_\eta(z;s_j) = P(\eta(s_j) < z)$; the Huber's (Huber, 1973) $\psi(\cdot)$, that is, $\psi(u) = u$ if $|u| \le c$, otherwise $c \cdot sign(u)$ if $|u| > c$ for a fixed constant $c > 0$, then $A_1(s_j) = \int_{-c}^{c} dF_\eta(z;s_j)$ and $\omega_\psi(s) = c^2 - \int_{-c}^{c}(c^2 - u^2)dF_\eta(u;s)$. Condition 11 aligns with a condition in Thm. 7 of Zhang and Chen (2007) and ensures that $\tilde{\omega}_\psi$ has a singular value decomposition.

Remark We know that $\gamma(s,t)$ is a continuous symmetric nonnegative definite kernel. By Mercer's theorem (Mercer, 1909), then there is an orthonormal basis $\{\phi_i\}$ of $L_2[0,1]$ and the corresponding sequence of nonnegative eigenvalues $\{\lambda_i\}$ such that

$$\int_0^1 \gamma(s,t)\phi_i(t)dt = \lambda_i\phi_i(s). \tag{5}$$

Let $\phi_i = (\phi_i(s_1),\ldots,\phi_i(s_m))^T$, and $\phi = (\phi_1,\ldots,\phi_m)$, $\Delta_j = s_j - s_{j-1}$, $j = 1,\ldots,m$. Note that $\max_j \Delta_j \to 0$ as $m \to \infty$. From (5), we have

$$\Gamma \approx \text{diag}(\Delta_1,\ldots,\Delta_m)\phi\text{diag}(\lambda_1,\ldots,\lambda_m)\phi^T.$$

Therefore,

$$\Gamma^{-1} \approx \phi\text{diag}(\lambda_1^{-1},\ldots,\lambda_m^{-1})\phi^T\text{diag}(\Delta_1^{-1},\ldots,\Delta_m^{-1}),$$

whose (k,l)th component is $\sum_{j=1}^m \lambda_j^{-1}\phi_j(s_k)\phi_j(s_l)\Delta_j$ and denotes $\varpi(s_k,s_l) = \sum_{j=1}^m \lambda_j\phi_j(s_k)\phi_j(s_l)\Delta_j$. So, Υ^{-1} has the form $(\varpi(s_k,s_l))_{m\times m}$.

Let $\mathcal{J} = \{j, |s_j - s| < c\}$. If there exists a constant c, $0 < c < 1$, such that $Card(\mathcal{J})$ is finite and $|\omega(s_j,s)| = O(m^{-1})$ for $j \in \overline{\mathcal{J}}$, where $\overline{\mathcal{J}}$ is a complement of $\mathcal{J}$. Then Condition (C8) holds.

From here on, let G_Σ and G_{ω_ψ} denote centered Gaussian processes indexed over $[0,1]$ with the covariance functions $\Sigma(s,t) = A_1(s)^{-1}A_1(t)^{-1}\omega_\psi(s,t)\Omega^{-1}$ and $\omega_\psi(s,t)I_p$, respectively.

Theorem 1. *Under Conditions 1–9, we have*
(i) $n^{1/2}(\hat\beta - \beta) \to G_\Sigma$ *and (ii)* $L(\beta,s) \to \kappa_1(s)^2\kappa_2(s)^{-1}\omega_\psi(s,s)^{-1}G_{\omega_\psi}(s)^\top G_{\omega_\psi}(s)$. *Furthermore,* $\kappa_1^{-2}(s)\kappa_2(s)L(\beta,s) \to \chi_p^2$.

Remark Theorems 1 gives the asymptotic results of MELE process $\hat\beta(s)$ and ELL ratio process $L(\beta(s),s)$ indexed over $s \in [0,1]$ according to three EL methods, which are the foundations for the above statistical inferences. Under the covariance structure Υ, the proposed three empirical likelihoods lead to the same asymptotically efficient point estimator. For any $s \in [0,1]$, based on the ELL ratio functions, we construct an asymptotic $\chi^2(p)$ statistic.

Theorem 2. *(i) Under Conditions 1–7, 9(b)(c) and 10 and the assumptions that* $\sup_{s\in[0,1]} E(\psi^4(\eta(s))) < \infty$ *and* $h_2 = O(m^{-1/5})$, *it follows that*

$$\sup_{s\in[0,1]} |\hat\sigma^2(s) - \sigma^2(s)| = O_p(|\log(m)|m^{-4/5} + h_1^4 + n^{-1} + (nmh_3)^{-1/2}(\log h_3)^{1/2} + h_3).$$

(ii) Under Conditions 1–7 and 9(b), it follows that

$$\sup_{(s,t)\in[0,1]^2} |\hat\gamma_v(s,t) - \gamma_v(s,t)| = O_p((mh_2)^{-1} + h_1^2 + h_2^2 + (\log n/n)^{1/2}).$$

Remark From Theorem 2, we see that $\hat\sigma^2(s)$ and $\hat\gamma_v(s,t)$ are the uniformly consistent estimators of $\sigma^2(s)$ and $\gamma_v(s,t)$, respectively. So $\hat\gamma_\eta(s,t)$ is a uniformly consistent estimator of $\gamma_\eta(s,t)$. Therefore, $\hat\Gamma$ formed by $\hat\gamma_\eta(s_k,s_l)$, $k,l = 1,\ldots,m$, can be used to statistical inferences in Section 3 since $\hat\Gamma = \Gamma + o_p(1)$.

3.2 | Global general linear hypothesis test

For a given $r \times p$ matrix C of full rank and a given vector of functions $c = (c_1,\ldots,c_r)$, consider the hypothesis test of $H_0 : C\beta(s) = c(s)$ for all s versus $H_1 : C\beta(s) \neq c(s)$ for some s. Toward a formal testing procedure, define $d_C(s) = n^{1/2}\hat A_1(s)\hat\omega_\psi(s,s)^{-1/2}\left(C\hat\Omega^{-1}C^\top\right)^{-1/2}(C\hat\beta(s) - c(s))$

and let $\hat{A}_1(s)$ be a consistent estimator of $A_1(s)$. Define $\hat{\omega}_\psi(s,s) = n^{-1}\sum_{i=1}^n \psi(\hat{\eta}_i(s)^2)$ and $\hat{\Omega} = n^{-1}\sum_{i=1}^n x_i x_i^\top$, both of which are consistent estimators. Consider the global test statistic $T_n = \int_0^1 d_C(s)^\top d_C(s)\mathrm{d}s$. By Theorem 1, $d_C(s)$ converges to the centred Gaussian process $G_C(s) = A_1(s)\omega_\psi(s)^{-1/2}(C\Omega^{-1}C^\top)^{-1/2}CG(s)$.

Theorem 3. *(a) Under Conditions 1–9, $T_n \to \int_0^1 G_C(s)^\top G_C(s)\mathrm{d}s + o_p(1)$. If Condition 11 additionally holds, then $T_n \to \sum_{l=1}^{l_0}\tau_l\chi^2_{r,l}$, where $(\chi^2_{r,l})_{l=1}^{l_0}$ are independent and identically distributed following a central χ^2 distribution with r degrees of freedom and $(\tau_l)_{l=1}^{l_0}$ is a sequence of strictly positive eigenvalues of $\tilde{\omega}_\psi$.*

Remark Theorem 3 presents the theoretical result for the global test statistic based on MELEs. The result of Theorem 4.5 shows that the distribution of T_n is asymptotically the same as that of a χ^2-type mixture. So, the χ^2-approximation and bootstrap methods can be used to approximate the null distribution of T_n; see Zhang and Chen (2007). In the paper, we approximate critical values of T_n using the bootstrapping. Note that we have to take a well-chosen integer $\hat{\imath}$ such that the eigenvalues τ_l, $l = 1,\ldots,\hat{\imath}$, explain a sufficiently large portion of the total variation $tr(\hat{\tilde{\omega}}_\psi(s,t)) = \sum_{l=1}^\infty \hat{\tau}_l$ and $\hat{\tau}_l$, $l = \hat{\imath}+1, \hat{\imath}+2,\ldots$ are essentially 0.

4 | PROCEDURES FOR STATISTICAL INFERENCE

4.1 | Global general linear hypothesis test

The form of T_n hinders an analytic study of its critical values, so we propose a three-step inferential approach based on the wild bootstrap (Zhu et al., 2012). First, obtain $\hat{\beta}$ and the $\hat{\eta}_i(s_j)$s under H_0. Second, for $b = 1,\ldots,B$, generate a random sample $\{\varsigma^{(b)}(s_j)|i = 1,\ldots,n; j = 1,\ldots,m\}$ from a standard normal distribution and compute $\hat{y}_i^{(b)}(s_j) = x_i^\top\hat{\beta}(s_j) + \varsigma_i^{(b)}(s_j)\hat{\eta}_i(s_j)$. Calculate $\hat{\beta}^{(b)}$ by fitting the model to the $\hat{y}_i^{(b)}$s. Further compute $T_n^{(b)}$, the corresponding value of the test statistic, using the resampled data. Third, compute the p-value for the test as

$$B^{-1}\sum_{b=1}^B I(T_n^{(b)} \geq T_n).$$

4.2 | Simultaneous confidence bands and regions based on coefficient estimates

A simultaneous confidence band for β takes the form $(L_l^\alpha(s), U_l^\alpha(s))$ over $s \in [0,1]$ and satisfies $pr(L_l^\alpha < \beta < U_l^\alpha) = 1-\alpha$ for $l = 1,\ldots,p$. By Theorem 1, $\sup_{s\in[0,1]}|n^{1/2}\{\hat{\beta}_l(s) - \beta_l(s)\}| \to \sup_{s\in[0,1]}|e_l^\top G_\Sigma(s)|$, so we choose a scalar C_l^α such that $pr(\sup_{s\in[0,1]}|e_l^\top G_\Sigma(s)| \leq C_l^\alpha) = 1-\alpha$ for a simultaneous confidence band of the form $(\hat{\beta}_l(s) - n^{-1/2}C_l^\alpha, \hat{\beta}_l(s) + n^{-1/2}C_l^\alpha)$. To estimate C_l^α, we take C_l^α as the empirical level-$(1-\alpha)$ quantile of $\{\sup_{s\in[0,1]}|e_l G_1^{(b)}(s)||b = 1,\ldots,B\}$ (Zhu et al., 2012), where

$$G_1^{(b)}(s) = \left\{(nmh)^{-1}\sum_{i=1}^n x_i x_i^\top J_m^\top\hat{\Gamma}^{-1}\hat{A}(s)\right\}^{-1} n^{-1/2}\sum_{i=1}^n\varsigma_i^{(b)}\sum_{j=1}^m U_{i,j}(\hat{\beta}, s),$$

$\{\varsigma_i^{(b)} | i = 1,\ldots,n; b = 1,\ldots,B\}$ is a random sample from a standard normal distribution, J_m is an m-vector of ones, and $\hat{A}(s) = (\hat{A}_1(s_1)K_1^h(s),\ldots,\hat{A}_1(s_m)K_m^h(s))^\top$.

We can also construct simultaneous confidence regions for β as a whole. By Theorem 1, $\sup_{s\in[0,1]} n(\hat{\beta}(s) - \beta(s))^\top(\hat{\beta}(s) - \beta(s)) \to \sup_{s\in[0,1]} G_\Sigma(s)^\top G_\Sigma(s)$, so we choose C^α such that $pr(\sup_{s\in[0,1]} G_\Sigma(s)^\top G_\Sigma(s) \le C^\alpha) = 1-\alpha$ for the simultaneous confidence region $\{b | \sup_{s\in[0,1]} n(\hat{\beta}(s) - b(s))^\top(\hat{\beta}(s) - b(s)) \le C^\alpha\}$. We can efficiently approximate C^α as the empirical level-$(1-\alpha)$ quantile of $\{\sup_{s\in[0,1]} |G_2^{(b)}(s)| | b = 1,\ldots,B\}$, where $G_2^{(b)}(s) = (\hat{\Phi}(s)g_2^{(b)}(s))^\top(\hat{\Phi}(s)g_2^{(b)}(s))$ with $g_2^{(b)}(s) = n^{-1/2}\sum_{i=1}^n \varsigma_i^{(b)} \sum_{j=1}^m U_{ij}(\hat{\beta}, s)$ and

$$\hat{\Phi}(s) = (nmh)^{-1} \sum_{i=1}^n x_i x_i^\top \sum_{j=1}^m \sum_{k=1}^m \varpi(s_j, s_k) J_m^\top \hat{\Gamma}^{-1} \hat{A}(s).$$

We refer to this approach as the Gaussian-approximation method.

4.3 | Confidence regions based on empirical log-likelihood

As shown in (A.9) of Appendix, $L(\beta, s) \to G_L(s)^\top \Sigma_L(s,s)^{-1} G_L(s)$, where G_L is a centered Gaussian process with the covariance function

$$\Sigma_L(s,t) = \kappa_1(s)\kappa_1(t)\mu_0(K;s,h)\mu_0(K;t,h)\pi(s)\pi(t)\omega_\psi(s,t)\Omega,$$

and $\mu_0(K;s,h) = \int_0^1 h^{-1}K((u-s)/h)\mathrm{d}u$. We choose a scalar C^α such that

$$pr\left(\sup_{s\in[0,1]} G_L(s)^\top \Sigma_L(s,s)^{-1} G_L(s) \le C^\alpha\right) = 1-\alpha,$$

for the confidence region $\{b | \sup_{s\in[0,1]} L(b,s) \le C^\alpha\}$. We can efficiently approximate C^α as the level-$(1-\alpha)$ empirical quantile of $\{\sup_{s\in[0,1]} |G_3^{(b)}| | b = 1,\ldots,B\}$, where $G_3^{(b)}(s) =$

TABLE 1 Empirical coverage probabilities of simultaneous confidence bands based on the least-squares estimator in the simulation study.

			$\alpha = 0.05$		$\alpha = 0.01$	
			$m = 40$	$m = 60$	$m = 40$	$m = 60$
Case I	$n = 100$	β_1	0.9633	0.9500	0.9867	0.9900
		β_2	0.9500	0.9300	0.9933	0.9733
	$n = 150$	β_1	0.9633	0.9533	0.9900	0.9867
		β_2	0.9368	0.9467	0.9967	0.9833
Case II	$n = 100$	β_1	0.9467	0.9600	0.9867	0.9900
		β_2	0.9700	0.9467	0.9967	0.9900
	$n = 150$	β_1	0.9467	0.9567	0.9900	0.9933
		β_2	0.9500	0.9433	0.9933	0.9867

$g_3^{(b)}(s)^\top \left(\sum_{i=1}^n \sum_{j=1}^m U_{i,j}(\hat{\beta}, s) U_{i,j}(\hat{\beta}, s)^\top \right)^{-1} g_3^{(b)}(s)$ and

$$g_3^{(b)}(s) = \sum_{i=1}^n \varsigma_i^{(b)} \sum_{j=1}^m U_{i,j}(\hat{\beta}, s).$$

Theorem 4. *Under Conditions 1–9 and the assumption that* $\sup_{s\in[0,1]} E(\psi^4(\eta(s))) < \infty$, *it follows that* $G_1^{(b)}(s) \to G_\Sigma(s)$, $G_2^{(b)}(s) \to G_\Sigma(s)^\top G_\Sigma(s)$, *and*

$$G_3^{(b)}(s) \to \kappa_1(s)^2 \kappa_2(s)^{-1} \omega_\psi(s)^{-1} G_{\omega_\psi}(s)^\top G_{\omega_\psi}(s, s).$$

Remark Theorem 4 validates the bootstrapped processes used in Section 3.

TABLE 2 Empirical coverage probabilities of simultaneous confidence bands based on the least-absolute-deviation estimator in the simulation study.

			$\alpha = 0.05$		$\alpha = 0.01$	
			$m = 40$	$m = 60$	$m = 40$	$m = 60$
Case I	$n = 100$	β_1	0.9667	0.9433	0.9867	0.9833
		β_2	0.9467	0.9467	0.9933	0.9800
	$n = 150$	β_1	0.9400	0.9700	0.9867	0.9833
		β_2	0.9500	0.9533	0.9867	0.9800
Case II	$n = 100$	β_1	0.9400	0.9400	0.9900	0.9967
		β_2	0.9667	0.9667	0.9867	0.9967
	$n = 150$	β_1	0.9467	0.9633	0.9866	0.9900
		β_2	0.9633	0.9400	0.9933	0.9900

TABLE 3 Empirical coverage probabilities of simultaneous confidence regions based on the least-squares estimator in the simulation study.

			$\alpha = 0.05$		$\alpha = 0.01$	
			$m = 40$	$m = 60$	$m = 40$	$m = 60$
Case I	$n = 100$	MELE	0.9500	0.9433	0.9933	0.9833
		ELLR	0.9533	0.9433	0.9933	0.9869
	$n = 150$	MELE	0.9467	0.9433	0.9800	0.9967
		ELLR	0.9500	0.9500	0.9767	0.9900
Case II	$n = 100$	MELE	0.9467	0.9467	0.9867	0.9900
		ELLR	0.9667	0.9600	0.9933	0.9966
	$n = 150$	MELE	0.9433	0.9600	0.9867	0.9867
		ELLR	0.9533	0.9633	0.9867	0.9933

TABLE 4 Empirical coverage probabilities of simultaneous confidence regions based on the least-absolute-deviation estimator in the simulation study.

			$\alpha = 0.05$		$\alpha = 0.01$	
			$m = 40$	$m = 60$	$m = 40$	$m = 60$
Case I	$n = 100$	MELE	0.9667	0.9400	0.9933	0.9933
		ELLR	0.9400	0.9500	0.9867	0.9800
	$n = 150$	MELE	0.9400	0.9467	0.9867	0.9800
		ELLR	0.9433	0.9633	0.9867	0.9867
Case II	$n = 100$	MELE	0.9400	0.9667	1.0000	0.9967
		ELLR	0.9400	0.9700	0.9967	0.9967
	$n = 150$	MELE	0.9533	0.9433	0.9867	0.9966
		ELLR	0.9667	0.9467	0.9900	0.9833

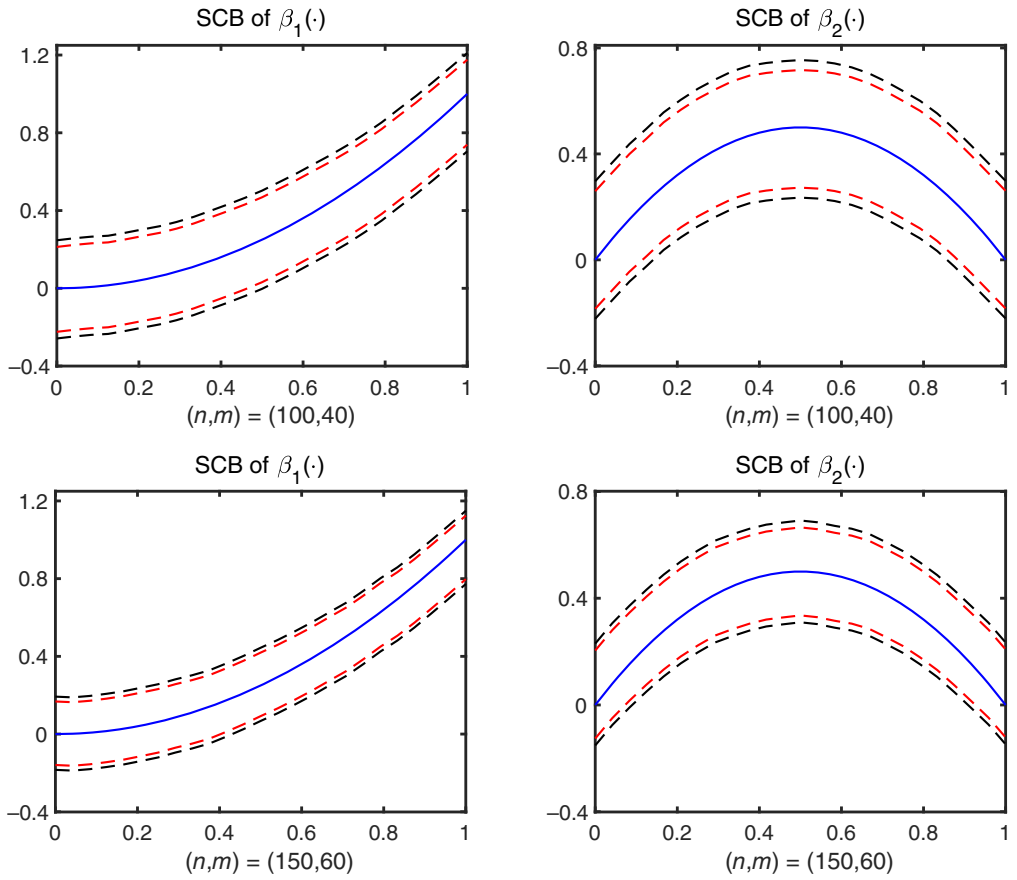

FIGURE 1 Typical simultaneous confidence bands for β_1 and β_2 with $(n, m) = (100, 40)$ (top) and $(n, m) = (150, 60)$ (bottom) for Case I of the simulation study. The bands are estimated using the estimate-based method under the least-squares loss. The solid blue curves denote the true coefficient functions, while the dashed red and black curves denote 95% and 99% simultaneous confidence bands, respectively.

5 | SIMULATION STUDY

The following simulation studies illustrate the performance of the proposed estimator and confidence sets. We generate data from the heteroscedastic model

$$y_i(s_j) = \beta_1(s_j) + x_i\beta_2(s_j) + \eta_i(s_j),$$

where the x_is are independent and identically distributed following the standard normal distribution. As Case I, we take $\beta_1(s) = s^2$ and $\beta_2(s) = 0.5\{1 - (2s-1)^2\}$. Additionally, we generate $(v_i(s_1),\ldots,v_i(s_m))^\top$ from a zero-mean, multivariate normal distribution such that $\mathrm{corr}(v_i(s_j), v_i(s_l)) = \exp(-|j-l|)$. As Case II, we take $\beta_1(s) = 4(s-0.5)^2$ and $\beta_2(s) = \sin(0.5\pi s)$ and generate $v_i(s) = \xi_{i,1}\varphi_1(s) + \xi_{i,2}\varphi_2(s)$, where $\xi_{i,1} \sim N(0, 1.2^2)$, $\xi_{i,2} \sim N(0, 1^2)$, $\varphi_1(s) = 2^{1/2}\sin(2\pi s)$, and $\varphi_2(s) = 2^{1/2}\cos(2\pi s)$. We obtain both least-squares and least-absolute-deviation estimates. For the former, $\eta_i(s_j) = v_i(s_j) + \varepsilon_i(s_j)$, where $\varepsilon_i(s_j) \sim N(0, 0.1^2)$. For the latter, $\eta_i(s_j) = \vartheta_i(s_j) - F^{-1}(0.5)$, where F is the marginal density function of $\vartheta_i(s_j) = v_i(s_j) + \varepsilon_i(s_j)$.

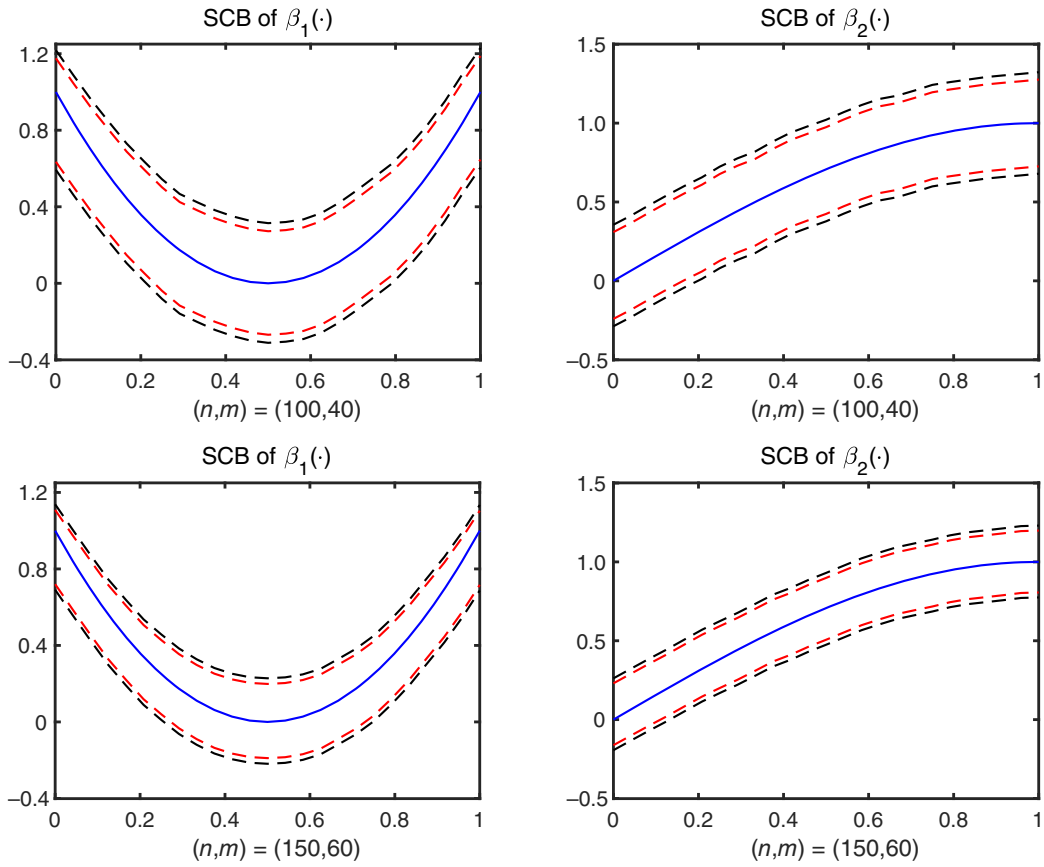

FIGURE 2 Typical simultaneous confidence bands for β_1 and β_2 with $(n, m) = (100, 40)$ (top) and $(n, m) = (150, 60)$ (bottom) for Case II of the simulation study. The bands are estimated using the estimate-based method under the least-squares loss. The solid blue curves denote the true coefficient functions, while the dashed red and black curves denote 95% and 99% simultaneous confidence bands, respectively.

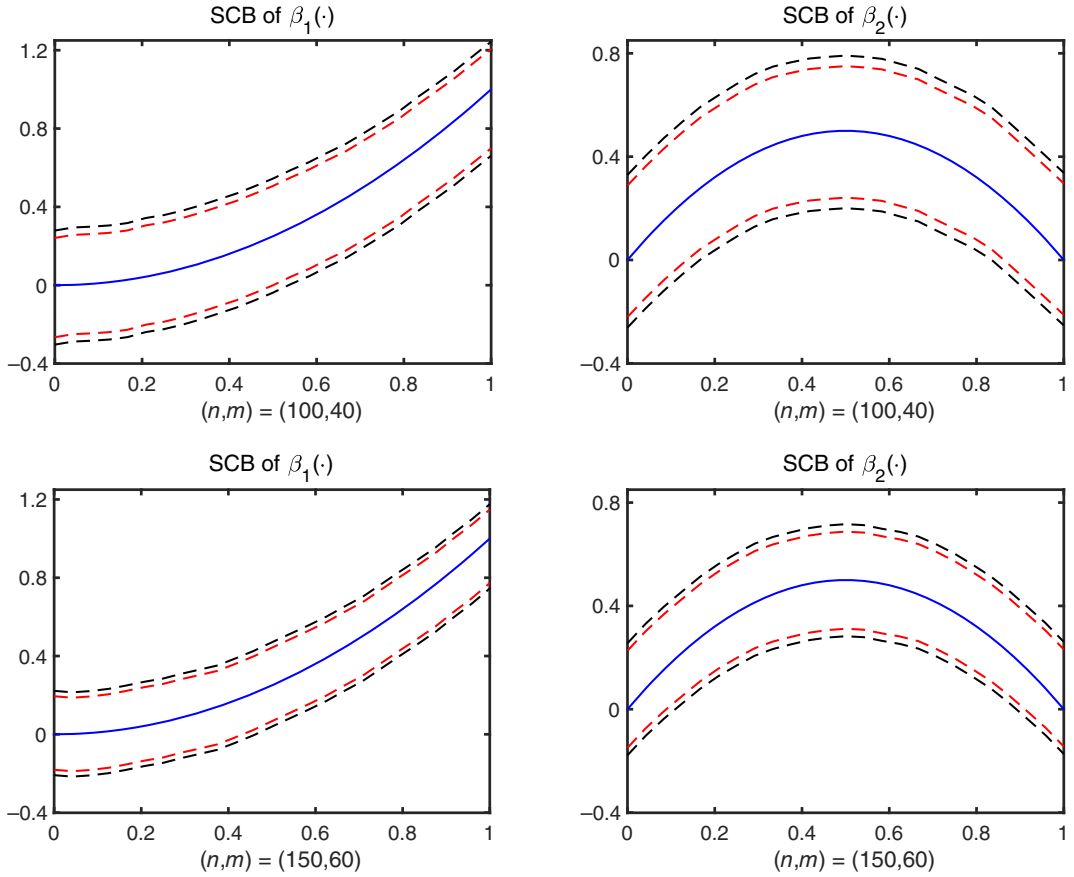

FIGURE 3 Simultaneous confidence bands for β_1 and β_2 with $(n, m) = (100, 40)$ (top) and $(n, m) = (150, 60)$ (bottom) for Case I of the simulation study. The bands are estimated using the estimate-based method under the least-absolute-deviation loss. The solid blue curves denote the true coefficient functions, while the dashed red and black curves denote 95% and 99% simultaneous confidence bands, respectively.

Tables 1 and 2 present empirical coverage probabilities of simultaneous confidence bands for the least square estimator and least-absolute-deviation estimator in Cases I and II based on 300 datasets with $n = 100{,}150$ and $m = 40, 60$, respectively. Tables 3 and 4 summarize the empirical coverage probabilities of simultaneous confidence regions for $(\beta_1(s), \beta_2(s))$. The coverage probabilities are close to the nominal level for both $\alpha = 0.01$ and $\alpha = 0.05$. Figures 1–4 depicts typical simultaneous confidence bands in the same setting. As expected, the bands narrow as n and m increase. As shown in Figures 5 and 6, our approach based on empirical log-likelihood yields smaller simultaneous confidence regions relative to the Gaussian-approximation method.

Thus, we see that our proposed methods based on empirical likelihood inference and robust regression work well on varying coefficient models for functional responses.

6 | REAL DATA ANALYSIS

We next consider a diffusion tensor imaging dataset from the Alzheimer's Disease Neuroimaging Initiative, which we preprocess following Zhu et al. (2012). See Acknowledgment. The analytic

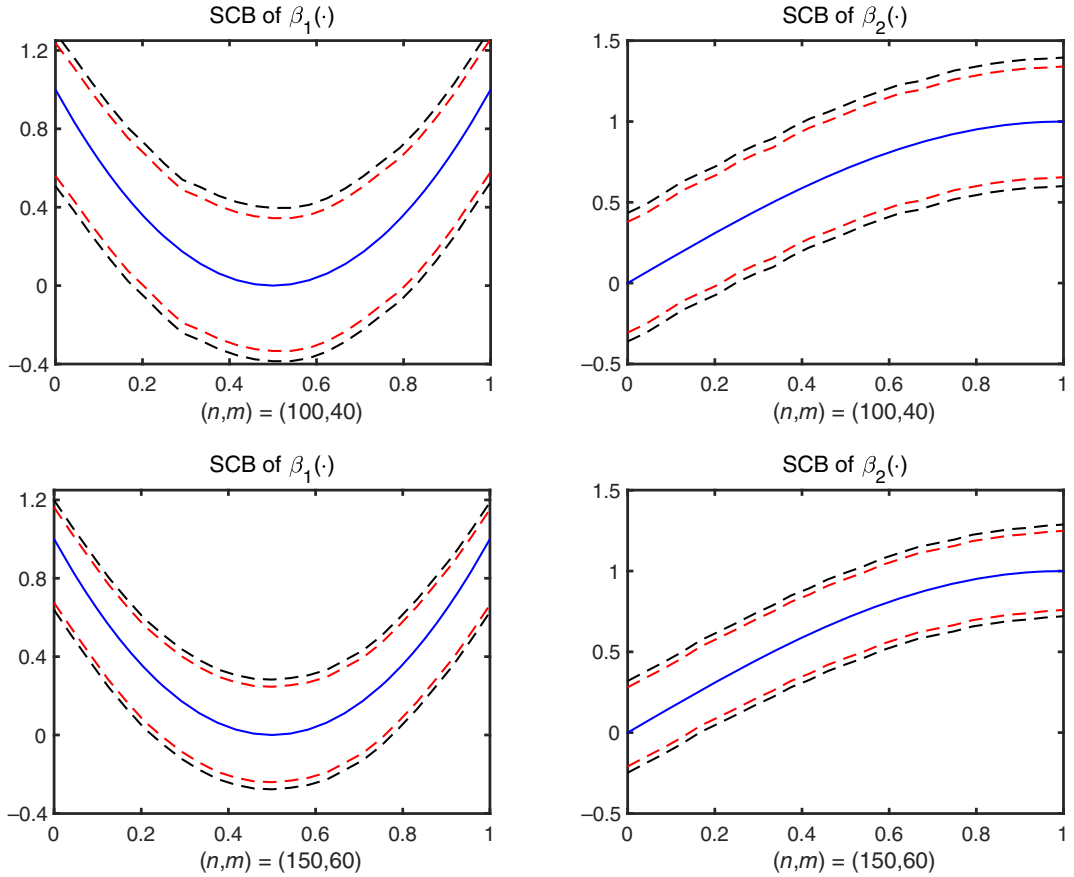

FIGURE 4 Simultaneous confidence bands for β_1 and β_2 with $(n, m) = (100, 40)$ (top) and $(n, m) = (150, 60)$ (bottom) for Case II of the simulation study under the least-absolute-deviation loss. The bands are estimated using the estimate-based method. The solid blue curves denote the true coefficient functions, while the dashed red and black curves respectively denote 95% and 99% simultaneous confidence bands.

sample consists of 203 subjects (119 males and 84 females) from the total 217 subjects after quality control and the exclusion of incomplete observations. Subject ages range from 48 to 91 years, with a mean of 73.3 years and a SD of 7.4 years.

We take fractional anisotropy, measured at 83 points as a function of arc length along the corpus callosum, as the functional response in the model. As scalar covariates, we include gender (female and male), age, years of education, diagnosis (normal control, mild cognitive impairment, Alzheimer's disease), mini-mental state examination score, and apolipoprotein E genotype (three levels). We take ρ as the quantile loss function due to the presence of outlying response values and consider the quantile levels $\tau = 0.25, 0.50, 0.75$.

At each quantile level, we estimate β and perform a global hypothesis test, with $B = 500$ replications, of the null hypothesis that $\beta_l = 0$. Figure 7 presents the coefficient function estimate for each covariate and the corresponding p-values. At the $\alpha = 0.10$ level, education level, diagnosis, mini-mental state examination score, and apolipoprotein E genotype have significant associations with all three conditional quantiles of fractional anisotropy. This is consistent with Zhu et al. (2012). On the other hand, the weak association between gender and mean fractional

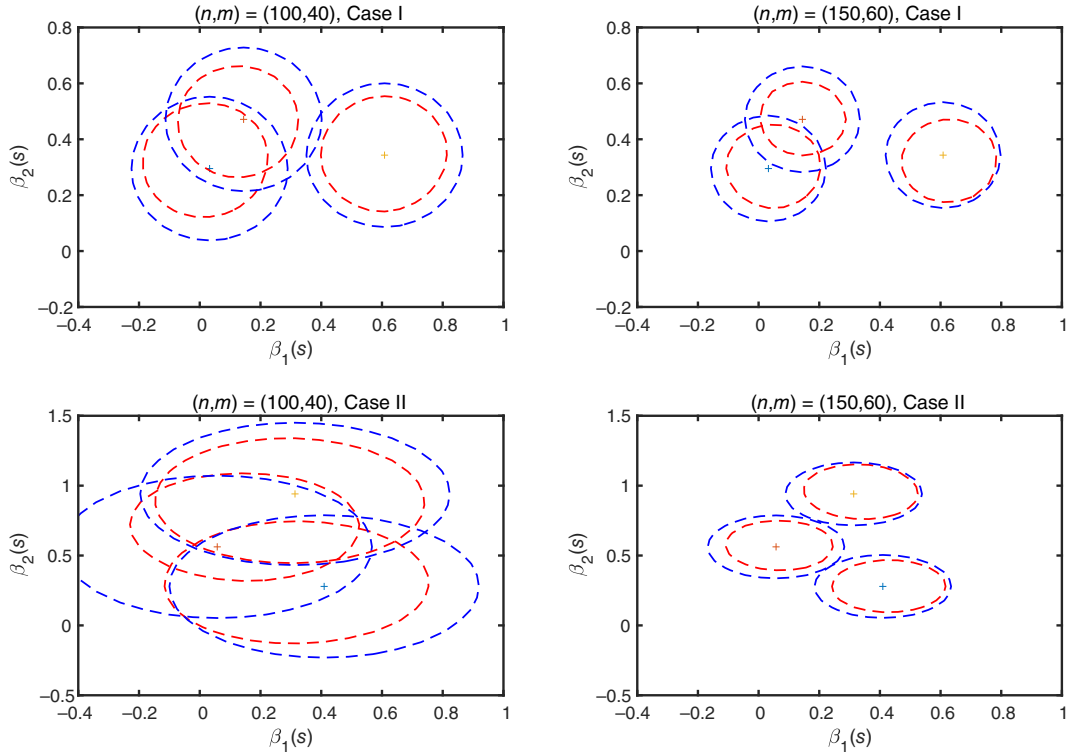

FIGURE 5 Slices of a typical 95% simultaneous confidence region for (β_1, β_2) at $s = 0.2, 0.4, 0.8$ m. The estimated regions are based on the empirical log-likelihood ratio under the least-squares loss (red) and the Gaussian approximation method (blue) for $(n, m) = (100, 40)$ (left), $(n, m) = (150, 60)$ (right), Case I (top), and Case II (bottom). The true coefficient function at each s is denoted by +.

anisotropy identified by the least-squares estimator in Zhu et al. (2012) is better contextualized by our results, which finds a significant association only at the high quantile level of $\tau = 0.75$. This result highlights the ability of quantile regression to provide more-specific insight into the distribution of the response of interest relative to mean regression as well as the general utility of our M-estimation framework.

7 | CONCLUSION

For our functional data model, it is very interesting to test whether a particular coefficient function is different from 0, or whether it is really varying or whether it is increasing in a compact subset. To address the latter two questions, it needs to construct simultaneous confidence bands, which is of particular interest in data analysis, instead of pointwise confidence bands. In many applications, the pointwise confidence bands are not satisfactory because they do not reflect the overall pattern of the coefficient functions. The construct of simultaneous confidence bands has been a very difficult problem when the model has dependent errors, such as within-subject/curve dependence in longitudinal data and functional data. In the paper, we propose an M-estimator

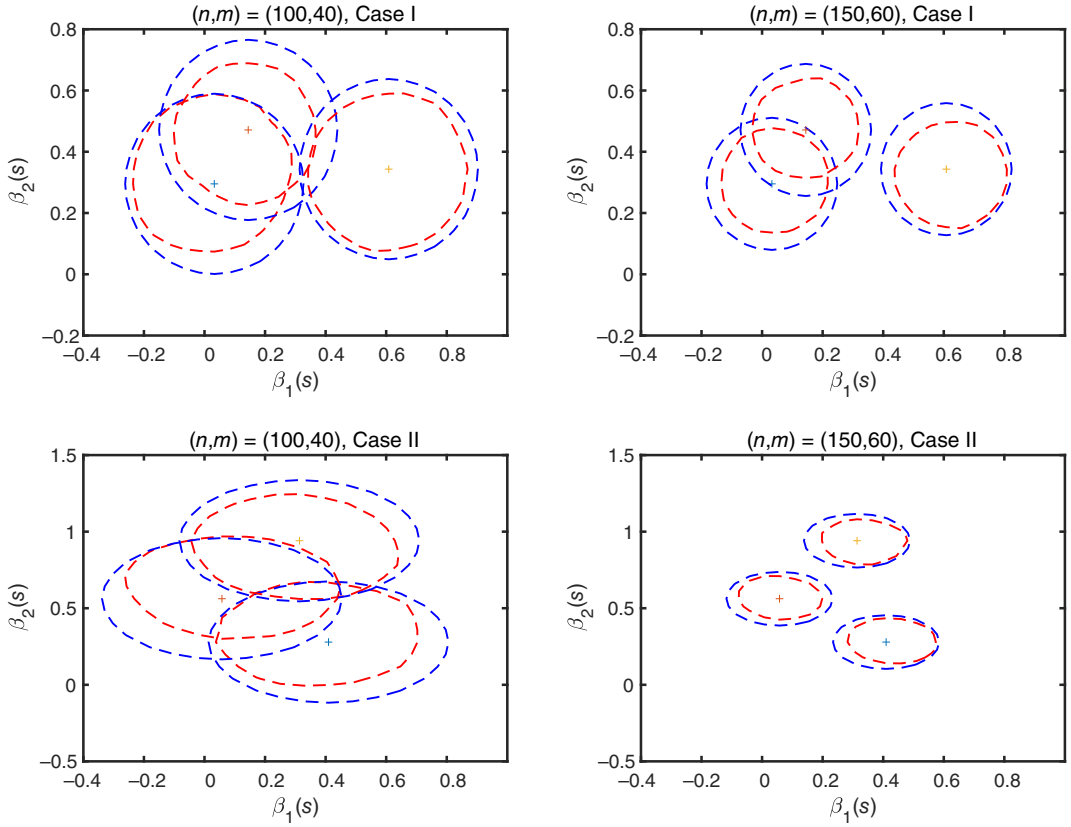

FIGURE 6 Slices of a typical 95% simultaneous confidence region for (β_1, β_2) at $s = 0.2, 0.4, 0.8$ m. The estimated regions are based on the empirical log-likelihood ratio under the least-absolute-deviation loss (red) and the Gaussian approximation method (blue) for $(n, m) = (100, 40)$ (left), $(n, m) = (150, 60)$ (right), Case I (top), and Case II (bottom). The true coefficient function at each s is denoted by +.

based on the generalized empirical likelihood for the varying-coefficient model with a functional response, and develop statistical inference procedures, simultaneous confidence regions, and a global general linear hypothesis test for the model's functional coefficient. Our theoretical results establish the weak convergence of the log-likelihood ratio process, a nonparametric version of Wilks' theorem for the log-likelihood ratio, and asymptotic properties of the proposed estimator.

Our approach based on empirical log-likelihood yields smaller simultaneous confidence regions relative to the Gaussian-approximation method. We analyze a diffusion tensor imaging dataset from the Alzheimer's Disease Neuroimaging Initiative based on our approach. Then, we have some important findings. For example, the weak association between gender and mean fractional anisotropy identified by the least-squares estimator in Zhu et al. (2012) is better contextualized by our results, which find a significant association only at the high quantile level. This result highlights the ability of quantile regression to provide more-specific insight into the distribution of the response of interest relative to mean regression as well as the general utility of our M-estimation framework.

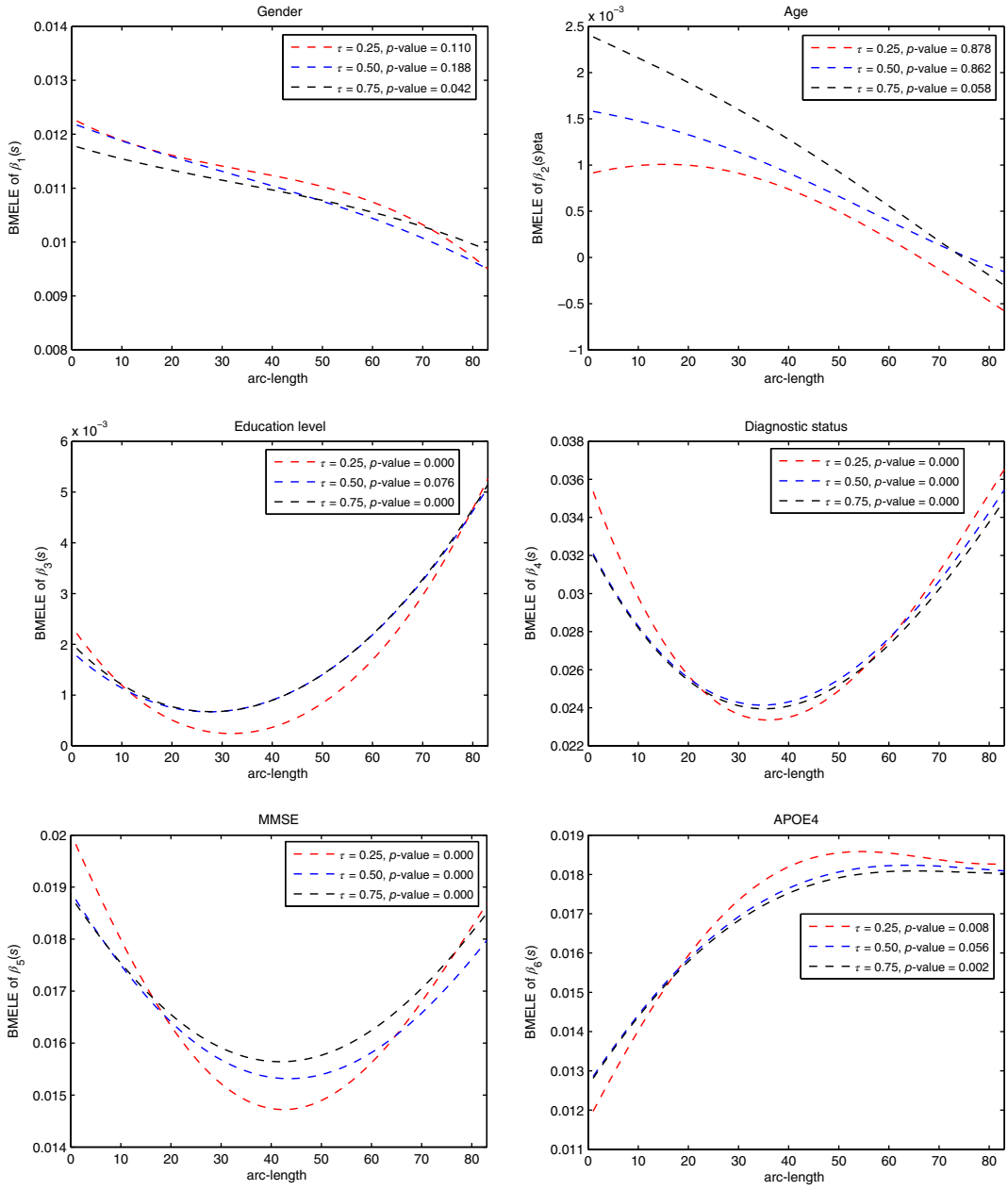

FIGURE 7 Estimated coefficients in the neuroimaging analysis. The red, blue, and black curves represent $\tau = 0.25$, $\tau = 0.5$, and $\tau = 0.75$, respectively. MMSE, mini-mental state examination score; APOE4, apolipoprotein E genotype.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Fund of China (12171242). Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see https://www.adni-info.org.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests.

ORCID

Matthew Stephen Pietrosanu https://orcid.org/0000-0003-2349-0535
Linglong Kong https://orcid.org/0000-0003-3011-9216

REFERENCES

Bai, Z., Rao, C., & Wu, Y. (1992). M-estimation of multivariate linerar regression parametrers under a convex discrepancy function. *Statistica Sinica*, *2*, 237–254.

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman & Hall.

Fan, J., Hu, T., & Truong, Y. (1994). Robust non-parametric function estimation. *Scandinavian Journal of Statistics*, *21*, 433–446.

Fan, J., & Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, *85*(3), 645–660.

Hall, P., Racine, J., & Li, Q. (2004). Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association*, *99*(468), 1015–1026.

Hample, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions*. Wiley.

Hastie, T., & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, *55*(4), 757–796.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and monte carlo. *The Annals of Statistics*, *1*, 799–821.

Huber, P. J., & Ronchetti, E. M. (2009). *Robust statistics* (2nd ed.). John Wiley & Sons.

Koenker, R., & Bassett, G. J. R. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50.

Kosorok, M. R. (2003). Bootstraps of sums of independent but not identically distributed stochastic processes. *Journal of Multivariate Analysis*, *84*(2), 299–318.

Li, J. L., Huang, C., Zhu, H. T., & Alzheimer's Disease Neuroimaging Initiative. (2017). A functional varying-coefficient single-index model for functional response data. *Journal of the American Statistical Association*, *112*(519), 1169–1181.

Li, Y., & Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, *38*, 3321–3351.

Luo, X. C., Zhu, L. X., & Zhu, H. T. (2016). Single-index varying coefficient model for functional responses. *Biometrics*, *72*, 1275–1284.

Mercer, J. (1909). Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *209*(1909), 415–446.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, *18*(1), 90–120.

Owen, A. B. (2001). *Empirical likelihood*. Chapman & Hall.

Qin, J., & Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, *22*(1), 300–325.

Rice, J. A., & Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society. Series B (Methodological)*, *53*(1), 233–243.

Tang, C. Y., & Leng, C. L. (2011). Empirical likelihood and quantile regression in longitudinal data analysis. *Biometrika*, *98*(4), 1001–1006.

van der Vaart, A. W., & Wellner, J. A. (1996). *Weak convergence and empirical processes: With applications to statistics*. Springer.

Wang, H., Zhong, P.-S., Cui, Y., & Li, Y. (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 (Statistical Methodology)*, *80*(2), 343–364.

Wang, L., Li, H., & Huang, J. Z. (2008). Variable selection in nonparametric varying coefficient models for analysis of repeated measurements. *Journal of the American Statistical Association*, *103*(484), 1556–1569.

Wang, S. J., Qian, L. F., & Carroll, R. J. (2010). Generalized empirical likelihood methods for analyzing longitudinal data. *Biometrika*, *97*(1), 79–93.

Xue, L. G., & Zhu, L. X. (2007). Empirical likelihood semiparametric regression analysis for longitudinal data. *Biometrika*, *94*(4), 921–937.

Yao, F. (2007). Asymptotic distributions of nonparametric regression estimators for longitudinal or functional data. *Journal of Multivariate Analysis*, *98*, 40–56.

You, J. H., Chen, G. M., & Zhou, Y. (2006). Block empirical likelihood for longitudinal partially linear regression models. *The Canadian Journal of Statistics*, *34*(1), 79–96.

Zhang, J. T., & Chen, J. (2007). Statistical inferences for functional data. *The Annals of Statistics*, *35*(3), 1052–1079.

Zhou, X. C., Kong, D. H., Kashlak, A. B., Kong, L. L., Karunamuni, R., & Zhu, H. T. (2023). Functional response quantile regression model. *Statistica Sinica*, *33*, 1–25.

Zhu, H. T., Li, R. Z., & Kong, L. (2012). Multivariate varying coefficient model for functional responses. *The Annals of Statistics*, *40*(5), 2634–2666.

How to cite this article: Zhou, X., Kong, D., Pietrosanu, M. S., Kong, L., & Karunamuni, R. J. (2024). Empirical likelihood M-estimation for the varying-coefficient model with functional response. *Scandinavian Journal of Statistics*, 1–31. https://doi.org/10.1111/sjos.12717

APPENDIX A

In this section, we give proof of this work's theoretical results. In the following proofs, we use the notation introduced in the main text. Additionally define $\theta(b,s) = b - \beta(s)$ and $\mu_r(K;s,h) = \int_0^1 (u-s)^r K((u-s)/h)h^{-(r+1)}\mathrm{d}u$ for $r \geq 0$.

A.1 PROOF OF THEOREM 1

Define

$$D(\beta, s) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} U_{i,j}(\beta, s),$$

$\Phi(s) = \kappa_1(s)A_1(s)\mu_0(K;s,h)\pi(s)\Omega$, and

$$\Sigma_L(s,t) = \kappa_1(s)\kappa_1(t)\mu_0(K;s,h)\mu_0(K;t,h)\pi(s)\pi(t)\omega_\psi(s,t)\Omega.$$

First, the proof of (i) consists of five steps.

Step 1. We show that

a. $n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(b,s)=O_p\left(n^{-1/2}\right)$,
b. $n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(b,s)U_{i,j}(b,s)^{\top}=\Sigma_L(s,s)\{1+o_p(1)\}$, and
c. $\lambda(b,s)=O_p\left(n^{-1/2}\right)$

All hold uniformly over $\Theta_n=\{(b,s)|s\in[0,1],\|\theta(b,s)\|\le d_n\}$, where $d_n\to 0$ as $n\to\infty$ and $\lambda(b,s)$ is a Lagrange multiplier in the problem of minimizing the empirical log-likelihood ratio function in (2) of the main text: that is, $\lambda(b,s)$ satisfies

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{U_{i,j}(b,s)}{1+\lambda(b,s)^{\top}U_{i,j}(b,s)}=0.$$

To establish (a), first note that

$$\begin{aligned}U_{i,j}(b,s)&=\frac{1}{mh}x_i u_{i,j}(b,s)\\&=\frac{1}{mh}x_i\sum_{k=1}^{m}\varpi(s_j,s_k)\psi(y_i(s_k)-x_i^{\top}b(s))K_k^h(s)\\&=\varpi(s_j,s)\mu_0(K;s,h)\pi(s)x_i\psi(\eta_i(s))(1+o_p(1)).\end{aligned}$$

Under Condition 3, it follows by the Donsker Theorem (van der Vaart & Wellner, 1996, Chap. 2.5) that $n^{-1/2}\sum_{i=1}^{n}x_i\psi(\eta_i(s))\to G(s)$, where G is a centered Gaussian process. Furthermore,

$$\sup_{s\in[0,1]}\left\|n^{-1/2}\sum_{i=1}^{n}x_i\psi(\eta_i(s))\right\|=O_p(1).$$

Therefore, by Lemma 2 of Zhu et al. (2012),

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(b,s)&=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\varpi(s_j,s)\mu_0(K;s,h)\pi(s)x_i\psi(\eta_i(s))(1+o_p(1))\\&=\kappa(s)\mu_0(K;s,h)\pi(s)\frac{1}{n}\sum_{i=1}^{n}x_i\psi(\eta_i(s))(1+o_p(1))=O_p\left(n^{-1/2}\right),\end{aligned}$$

uniformly on Θ_n.

To establish (b), we use a similar argument. By the weak law of large numbers,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(b,s)U_{i,j}(b,s)^{\top}\\&=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\varpi(s_j,s)^2\mu_0(K;s,h)^2\pi(s)^2x_ix_i^{\top}\psi(\eta_i(s))^2(1+o_p(1))\\&=\kappa_2(s)\mu_0(K;s,h)^2\pi(s)^2\frac{1}{n}\sum_{i=1}^{n}x_ix_i^{\top}\psi(\eta_i(s))^2(1+o_p(1))\\&=\Sigma_L(s,s)(1+o_p(1)),\end{aligned}$$

uniformly for $s\in[0,1]$.

To establish (c), let $\lambda(b,s)=\rho(b,s)\alpha$, where $\rho(b,s)>0$ and $\|\alpha\|=1$. By an argument similar to that in the proof of (2.12) of Owen (1990),

$$\frac{\rho(b,s)}{1+\rho(b,s)\max_{i,j}\sup_{(b,s)\in\Theta_n}\|U_{i,j}(b,s)\|}\alpha^\top\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(b,s)U_{i,j}(b,s)^\top\right)\alpha\le\left\|\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(b,s)\right\|.$$

From the proof of (a), we know that

$$\begin{aligned}\max_{i,j}\sup_{(b,s)\in\Theta_n}\|U_{i,j}(b,s)\|&=\max_{i,j}\sup_{s\in[0,1]}|\varpi(s_j,s)\psi(\eta_i(s))\mu_0(K;s,h)\pi(s)|\|x_i\|_\infty(1+o_p(1))\\&\le C\max_{i}\sup_{s\in[0,1]}|\psi(\eta_i(s))|(1+o_p(1)).\end{aligned}$$

Under Condition 3, $E(\sup_{s\in[0,1]}\psi^2(\eta_i(s)))<\infty$, and $\sup_{s\in[0,1]}|\psi(\eta_i(s))|$, for $i=1,\dots,n$, are independent and identically distributed. Thus, by Lemma 3 of Owen (1990), $\max_i\sup_{s\in[0,1]}|\psi(\eta_i(s))|=o_p(n^{1/2})$ and so

$$\max_{i,j}\sup_{(b,s)\in\Theta_n}\|U_{i,j}(b,s)\|=o_p(n^{1/2}). \tag{A.1}$$

By (A.1) and (a)–(b), $\sup_{(b,s)\in\Theta_n}\rho(b,s)=O_p\left(n^{-1/2}\right)$. Thus,

$$\sup_{(b,s)\in\Theta_n}\|\lambda(b,s)\|=\sup_{(b,s)\in\Theta_n}\rho(b,s)=O_p\left(n^{-1/2}\right). \tag{A.2}$$

Step 2. We show that

$$\lambda(b,s)=\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(b,s)U_{i,j}(b,s)^\top\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(b,s)\right)+o_p\left(n^{-1/2}\right),$$

uniformly over Θ_n.

By applying (A.1) and (A.2) and noting that

$$\begin{aligned}\max_{i,j}\sup_{(b,s)\in\Theta_n}\|\lambda(b,s)^\top U_{i,j}(b,s)\|&=\sup_{(b,s)\in\Theta_n}\|\lambda(b,s)\|\max_{i,j}\sup_{(b,s)\in\Theta_n}\|U_{i,j}(b,s)\|\\&=O_p\left(n^{-1/2}\right)o_p(n^{1/2})=o_p(1),\end{aligned} \tag{A.3}$$

it follows that

$$\begin{aligned}0&=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{U_{i,j}(b,s)}{1+\lambda(b,s)^\top U_{i,j}(b,s)}\\&=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(b,s)-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(b,s)U_{i,j}(b,s)^\top\lambda(b,s)\\&\quad+\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{U_{i,j}(b,s)(\lambda(b,s)^\top U_{i,j}(b,s))^2}{1+\lambda(b,s)^\top U_{i,j}(b,s)}.\end{aligned} \tag{A.4}$$

By (A.2) and (A.3),

$$\begin{aligned}
&\sup_{(b,s)\in\Theta_n}\left\|\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{U_{i,j}(b,s)(\lambda(b,s)^\top U_{i,j}(b,s))^2}{1+\lambda(b,s)^\top U_{i,j}(b,s)}\right\| \\
&\quad= O_p(1)\max_{i,j}\sup_{(b,s)\in\Theta_n}\|U_{i,j}(b,s)\|\sup_{(b,s)\in\Theta_n}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\|U_{i,j}(b,s)\|^2\sup_{(b,s)\in\Theta_n}\|\lambda(b,s)\|^2 \\
&\quad= o_p(n^{1/2})O_p(1)O_p\left(n^{-1}\right) \\
&\quad= o_p\left(n^{-1/2}\right). \qquad (A.5)
\end{aligned}$$

The claim in this step holds by (A.4) and (A.5).

Step 3. We show that

$$n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(\beta,s)\to G_L(s), \qquad (A.6)$$

where G_L is a centered Gaussian process with the covariance function $\Sigma_L(s,t)$.

We know that

$$n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(\beta,s)=n^{-1/2}\sum_{i=1}^{n}x_i\psi(\eta_i(s))\mu_0(K;s,h)\kappa(s)\pi(s)(1+o_p(1)).$$

Note that $\sup_{s\in[0,1]}|\mu_0(K;s,h)\pi(s)|=O(1)$ and that $\{x_i\psi(\eta_i(s))|s\in[0,1]\}$ is a Donsker class. The claim in (A.6) follows by Chap. 2.5 of van der Vaart and Wellner (1996). It is easy to verify that

$$\operatorname{cov}\left(n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(\beta,s),n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(\beta,t)\right)=\Sigma_L(s,t)(1+o(1)),$$

which completes the proof for this step.

Step 4. We show that

$$\tilde{\beta}(s)-\beta(s)=\frac{1}{n}\Phi(s)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(\beta,s)+o_p(1),$$

uniformly for $s\in[0,1]$.

First, we have that

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(b,s)=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(\beta,s)-\Phi(s)\theta(b,s)(1+o_p(1)).$$

By (A.3),

$$L(b,s)=2\left\{\sum_{i=1}^{n}\sum_{j=1}^{m}\lambda(b,s)^\top U_{i,j}(b,s)-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}(\lambda(b,s)^\top U_{i,j}(b,s))^2\right\}+2\sum_{i=1}^{n}\sum_{j=1}^{m}\xi_{i,j}, \qquad (A.7)$$

where

$$pr\left(|\xi_{i,j}| \le \sup_{(b,s)\in\Theta_n} |\lambda(b,s)^\top U_{i,j}(b,s)|^3 \text{ for } i=1,\ldots,n; j=1,\ldots,m\right) \to 1,$$

as $n \to \infty$.

Note that

$$\left|\sum_{i=1}^{n}\sum_{j=1}^{m}\xi_i\right| \le \sup_{(b,s)\in\Theta_n} \|\lambda(b,s)\|^3 \sup_{(b,s)\in\Theta_n} \sum_{i=1}^{n}\sum_{j=1}^{m}\|U_{i,j}(b,s)\|^3 = O_p(n^{-3/2})o_p(n^{3/2}) = o_p(1),$$

so by the result of Step 2, we can rewrite (A.7) as

$$\begin{aligned} L(b,s) &= 2\left(\sum_{i=1}^{n}\sum_{j=1}^{m}\lambda(b,s)^\top U_{i,j}(b,s) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}(\lambda(b,s)^\top U_{i,j}(b,s))^2\right) + o_p(1) \\ &= n\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(b,s)\right)^\top \left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(b,s)U_{i,j}(b,s)^\top\right)^{-1} \\ &\quad \times \left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(b,s)\right) + o_p(1) \\ &= (D(\beta,s) - \Phi(s)\theta(s))^\top \Sigma_L(s,s)^{-1}(D(\beta,s) - \Phi(s)\theta(s)) + o_p(1). \end{aligned} \tag{A.8}$$

By a similar argument,

$$L(\beta,s) = nD(\beta,s)^\top \Sigma_L(s,s)^{-1} D(\beta,s) + o_p(1). \tag{A.9}$$

By (A.8) and (A.9),

$$L(b,s) - L(\beta,s) = -2nD(\beta,s)^\top \Sigma_L(s,s)^{-1}\Phi(s)\theta(s) + n\theta(s)^\top \Phi(s)\Sigma_L(s,s)^{-1}\Phi(s)\theta(s) + o_p(1), \tag{A.10}$$

uniformly over Θ_n. We know that $L(\tilde{\beta},s) - L(\beta,s) \le 0$, so by (A.10),

$$n\|\tilde{\theta}(s)\|^2\|\Phi(s)\Sigma_L(s,s)^{-1}\Phi(s)\| + o_p(1) \le 2n\|D(\beta,s)\|\|\Sigma_L(s,s)^{-1}\Phi(s)\|\|\tilde{\theta}(s)\|,$$

where $\tilde{\theta}(s) = \tilde{\beta}(s) - \beta(s)$.

By the result of Step 3, $\sup_{(b,s)\in\Theta_n}\|D(b,s)\| = O_p\left(n^{-1/2}\right)$. Note that

$$\sup_{s\in[0,1]}\|\Phi(s)\Sigma_L(s,s)^{-1}\Phi(s)\| \le C\|\Omega\| = O_p(1),$$

and $\sup_{s\in[0,1]}\|\Sigma_L(s,s)^{-1}\Phi(s)\| = O_p(1)$. Therefore, $\sup_{s\in[0,1]}\|\tilde{\beta}(s) - \beta(s)\| = O_p\left(n^{-1/2}\right)$.

Let $\tilde{\tilde{\beta}}(s) = \beta(s) + (\Phi(s)\Sigma_L(s,s)^{-1}\Phi(s))^{-1}\Phi(s)\Sigma_L(s,s)^{-1}D(\beta,s)$. By (A.8) and (A.9),

$$L(\tilde{\beta},s) - L(\beta,s) = -2nD(\beta,s)^\top \Sigma_L(s,s)^{-1}\Phi(s)\tilde{\theta}(s) + n\tilde{\theta}(s)^\top \Phi(s)\Sigma_L(s,s)^{-1}\Phi(s)\tilde{\theta}(s) + o_p(1). \tag{A.11}$$

Similarly,

$$L(\tilde{\tilde{\beta}},s) - L(\beta,s) = -2nD(\beta,s)^\top \Sigma_L(s,s)^{-1}\Phi(s)\tilde{\tilde{\theta}}(s) + n\tilde{\tilde{\theta}}(s)^\top \Phi(s)\Sigma_L(s,s)^{-1}\Phi(s)\tilde{\tilde{\theta}}(s) + o_p(1). \tag{A.12}$$

By (A.11) and (A.12),

$$0 \geq L(\tilde{\beta}, s) - L(\tilde{\tilde{\beta}}, s) = n(\tilde{\beta}(s) - \tilde{\tilde{\beta}}(s))^{\top}\Phi(s)\Sigma_L(s,s)^{-1}\Phi(s)(\hat{\beta}(s) - \tilde{\beta}(s)) + o_p(1).$$

Therefore,

$$\tilde{\beta}(s) - \beta(s) = \tilde{\beta}(s) - \tilde{\tilde{\beta}}(s) + o_p(1) = \Phi(s)^{-1}D(\beta, s) + o_p(1). \tag{A.13}$$

Step 5. We now compute the covariance matrix of $n^{1/2}(\tilde{\beta}(s) - \beta(s))$. By (A.13),

$$n\mathrm{cov}(\tilde{\beta}(s) - \beta(s), \tilde{\beta}(t) - \beta(t)) = n\Phi(s)^{-1}\mathrm{cov}(D(\beta, s), D(\beta, t))\Phi(t)^{-1} = \Phi(s)^{-1}\Sigma_L(s,t)\Phi(t)^{-1}.$$

Thus, we obtain $n^{1/2}(\tilde{\beta} - \beta) \to G_\Sigma$. By Remark for Theorem 2, we have $\hat{\Gamma} = \Gamma + o_p(1)$. Together with the definition of $\hat{\beta}$ and $\hat{L}(b, s)$, we have $n^{1/2}(\hat{\beta} - \beta) \to G_\Sigma$. This completes the proof of Theorem 1(i).

Second, for any $s \in [0, 1]$,

$$\begin{aligned} L(\beta, s) = & n\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(\beta, s)\right)^{\top}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(\beta, s)U_{i,j}(\beta, s)^{\top}\right)^{-1} \\ & \times\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(\beta, s)\right) + o_p(1) \\ \to & G_L(s)^{\top}\Sigma_L(s,s)^{-1}G_L(s) = \kappa_1(s)^2\kappa_2(s)^{-1}\omega_\psi(s,s)^{-1}G_{\omega_\psi}(s)^{\top}G_{\omega_\psi}(s). \end{aligned}$$

Thus, for any $s \in [0, 1]$, $L(\beta, s) \to \kappa_1(s)^2\kappa_2(s)^{-1}\chi_p^2$. This completes the proof of Theorem 1.

APPENDIX B. PROOF OF THEOREM 2

(i) By the definition of $\hat{\sigma}^2$in the main text,

$$\begin{aligned} \hat{\sigma}^2(s) = & \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)(y_i(s_j) - x_i^{\top}\hat{\beta}(s_j) - \hat{v}_i(s_j))^2}{n\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)} \\ = & \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)\{x_i^{\top}(\beta(s) - \hat{\beta}(s_j)) + (v_i(s_j) - \hat{v}_i(s_j)) + \varepsilon_i(s_j)\}^2}{n\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)} \\ = & \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)\varepsilon_i(s_j)^2}{n\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)} + \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)\{x_i^{\top}(\beta(s) - \hat{\beta}(s_j))\}^2}{n\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)} \\ & + \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)(v_i(s_j) - \hat{v}_i(s_j))^2}{n\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)} + \frac{2\sum_{i=1}^{n}\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)x_i^{\top}(\beta(s) - \hat{\beta}(s_j))\varepsilon_i(s_j)}{n\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)} \\ & + \frac{2\sum_{i=1}^{n}\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)x_i^{\top}(\beta(s) - \hat{\beta}(s_j))(v_i(s_j) - \hat{v}_i(s_j))}{n\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)} \end{aligned}$$

$$+\frac{2\sum_{i=1}^{n}\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)(v_i(s_j)-\hat{v}_i(s_j))\epsilon_i(s_j)}{n\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)}$$
$$=J_1(s)+\cdots+J_6(s),$$

where the last equality serves to define $J_1,\ldots,J_6$ in the obvious manner. Note that $\sup_{s\in[0,1]}\|\hat{\beta}(s)-\beta(s)\|=O_p\left(n^{-1/2}\right)$ and $\sup_{s\in[0,1]}|\hat{v}_i(s)-v_i(s)|=O_p(|\log m|^{1/2}m^{-2/5}+h_1^2+n^{-1/2})$ if $h_2=O(m^{-1/5})$ by Thms 1 and 2 of Zhu et al. (2012). Furthermore,

$$J_1(s)=\sigma^2(s)+O_p(h_3)\mu_1(K;s,h_3)+O_p(h_3^2)\mu_2(K;s,h_3),$$

$$\sup_{s\in[0,1]}|J_2(s)|=O_p\left(n^{-1}\right),$$

$$\sup_{s\in[0,1]}|J_3(s)|=O_p(|\log m|m^{-4/5}+h_1^4+n^{-1}),$$

$$\sup_{s\in[0,1]}|J_4(s)|=O_p\left(n^{-1/2}\right)\frac{\left|\sum_{i=1}^{n}\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)\epsilon_i(s_j)\right|}{n\sum_{j=1}^{m}\tilde{K}_j^{h_3}(s)}=O_p\left(\frac{\log h_3}{nmh_3}\right)^{1/2},$$

$$\sup_{s\in[0,1]}|J_5(s)|=O_p(n^{-1/2}(|\log m|^{1/2}m^{-2/5}+h_1^2+n^{-1/2})),$$

$$\sup_{s\in[0,1]}|J_6(s)|=O_p\left(\left(\frac{\log h_3}{mh_3}\right)^{1/2}(|\log m|^{1/2}m^{-2/5}+h_1^2+n^{-1/2})\right).$$

Therefore,

$$\sup_{s\in[0,1]}|\hat{\sigma}^2(s)-\sigma^2(s)|=O_p\left(|\log m|m^{-4/5}+h_1^4+n^{-1}+\frac{\log h_3}{nmh_3}+h_3\right).$$

Thus, we complete the proof of Theorem 2(i).

(ii) Let $\overline{K}_j^{h_2}(s)=[1\ 0]\left(\sum_{j=1}^{m}\tilde{K}_j^{h_2}(s)z_j^{h_2}(s)z_j^{h_2}(s)^\top\right)^{-1}\tilde{K}_j^{h_2}(s)z_j^{h_2}(s)$, which is the empirical equivalent kernel (Fan & Gijbels, 1996). We define $\overline{\epsilon}_i(s)=\sum_{j=1}^{m}\overline{K}_j^{h_2}(s)\epsilon_i(s_j)$, $\Delta v_i(s)=\sum_{j=1}^{m}\overline{K}_j^{h_2}(s)[v_i(s_j)-v_i(s)]$, $\Delta\beta(s)=\sum_{j=1}^{m}\overline{K}_j^{h_2}(s)[\beta(s_j)-\tilde{\beta}(s_j)]$, $\Delta_i(s)=\overline{\epsilon}_i(s)+\Delta v_i(s)+x_i^\top\Delta\beta(s)$. From the definition of $\hat{v}_i(s)$, we have $\hat{v}_i(s)=v_i(s)+\Delta_i(s)$. So

$$\frac{1}{n-p}\sum_{i=1}^{n}\hat{v}_i(s)\hat{v}_i(t)=\frac{1}{n-p}\sum_{i=1}^{n}\Delta_i(s)\Delta_i(t)+\frac{1}{n-p}\sum_{i=1}^{n}v_i(s)\Delta_i(t)$$
$$+\frac{1}{n-p}\sum_{i=1}^{n}\Delta_i(s)v_i(t)+\frac{1}{n-p}\sum_{i=1}^{n}v_i(s)v_i(t). \tag{B.1}$$

The proof consists of two steps. The first step is to prove that the first three terms on the right side of (B.1) converge to zero uniformly for all $(s,t)\in[0,1]^2$ in probability; the second step is to prove the $\frac{1}{n-p}\sum_{i=1}^{n}v_i(s)v_i(t)$ converges to $\gamma_v(s,t)$ uniformly over $(s,t)\in[0,1]^2$ in probability.

For Step 1, we first show that the first term

$$\sup_{(s,t)}\frac{1}{n-p}\left|\sum_{i=1}^{n}\Delta_i(s)\Delta_i(t)\right| = O_p\big((mh_2)^{-1}+(\log n/n)^{1/2}+h_1^4+h_2^4\big). \tag{B.2}$$

For a positive constant C, we have

$$\begin{aligned}\left|\sum_{i=1}^{n}\Delta_i(s)\Delta_i(t)\right| = C\sup_{(s,t)}\Bigg[&\left|\sum_{i=1}^{n}\bar{\varepsilon}_i(s)\bar{\varepsilon}_i(t)\right|+\left|\sum_{i=1}^{n}\bar{\varepsilon}_i(s)\Delta v_i(t)\right|+\left|\sum_{i=1}^{n}\Delta v_i(t)x_i^{\top}\Delta\beta(s)\right|\\&+\left|\sum_{i=1}^{n}\bar{\varepsilon}_i(s)x_i^{\top}\Delta\beta(t)\right|+\left|\sum_{i=1}^{n}\Delta v_i(s)\Delta v_i(t)\right|+\left|\sum_{i=1}^{n}x_i^{\top}\Delta\beta(s)\Delta\beta(t)^{\top}x_i\right|\Bigg].\end{aligned} \tag{B.3}$$

By Thm. 1 of Zhu et al. (2012), $\sup_{s\in[0,1]}\|\Delta\beta(s)\|_2 = O_p(n^{-1/2}+h_1^2)$. Notice that $\sup_{s\in[0,1]}|\Delta v_i(s)| = \sup_{s\in[0,1]}|v_i''(s)|O_p(h_2^2) = O_p(h_2^2)$. Further, from Lemmas 6 and 7 in Zhu et al. (2012) or Lemma 4 in Luo et al. (2016), it follows that

$$\sup_{(s,t)}\frac{1}{n-p}\left|\sum_{i=1}^{n}\bar{\varepsilon}_i(s)\bar{\varepsilon}_i(t)\right| = O_p\left((mh_2)^{-1}+(\log n/n)^{1/2}\right),$$

$$\sup_{(s,t)}\frac{1}{n-p}\left|\sum_{i=1}^{n}\bar{\varepsilon}_i(s)\Delta v_i(t)\right| = O_p\left((\log n/n)^{1/2}\right),$$

$$\sup_{(s,t)}\frac{1}{n-p}\left|\sum_{i=1}^{n}\Delta v_i(t)x_i^{\top}\Delta\beta(s)\right| = \sup_{s\in[0,1]}\|\Delta\beta(s)\|_2\sup_{t\in[0,1]}|\Delta v_i(t)|\frac{1}{n-p}\sum_{i=1}^{n}\|x_i\|_2 = O_p\left((n^{-1/2}+h_1^2)h_2^2\right),$$

$$\begin{aligned}\sup_{(s,t)}\frac{1}{n-p}\left|\sum_{i=1}^{n}\bar{\varepsilon}_i(s)x_i^{\top}\Delta\beta(s)\right| &= \sup_{s\in[0,1]}\|\Delta\beta(s)\|_2\sup_{t\in[0,1]}|\bar{\varepsilon}_i(t)|\frac{1}{n-p}\sum_{i=1}^{n}\|x_i\|_2\\&= O_p\left(n^{-1/2}+h_1^2\right)O_p\left(\sqrt{\frac{\log h_2}{mh_2}}\right),\end{aligned}$$

$$\sup_{(s,t)}\frac{1}{n-p}\left|\sum_{i=1}^{n}\Delta v_i(s)\Delta v_i(t)\right| = O(1)\sup_{s\in[0,1]}(\Delta v_i(s))^2 = O_p\left(h_2^4\right),$$

$$\sup_{(s,t)}\frac{1}{n-p}\left|\sum_{i=1}^{n}x_i^{\top}\Delta\beta(s)\Delta\beta(t)^{\top}x_i\right| = O_p(n^{-1}+h_1^4).$$

Therefore, together with (B.3), we obtain (B.2). Then, we show that the second term

$$\sup_{(s,t)}\frac{1}{n-p}\left|\sum_{i=1}^{n}\Delta_i(s)v_i(t)\right| = O_p\left(n^{-1/2}+h_1^2+h_2^2+(\log n/n)^{1/2}\right). \tag{B.4}$$

Since

$$\left|\sum_{i=1}^{n}\Delta_i(s)v_i(t)\right| \le \left|\sum_{i=1}^{n}\bar{\varepsilon}_i(s)v_i(t)\right|+\left|\sum_{i=1}^{n}\Delta v_i(s)v_i(t)\right|+\left|\sum_{i=1}^{n}x_i^{\top}\Delta\beta(s)v_i(t)\right|, \tag{B.5}$$

Furthermore, we have

$$\sup_{(s,t)} \frac{1}{n-p} \left| \sum_{i=1}^{n} \bar{\varepsilon}_i(s) v_i(t) \right| = O\left((\log n/n)^{1/2} \right),$$

from Lemma 6 in Zhu et al. (2012) or Lemma 4 in Luo et al. (2016),

$$\sup_{(s,t)} \frac{1}{n-p} \left| \sum_{i=1}^{n} \Delta v_i(s) v_i(t) \right| \le \frac{1}{n-p} \sum_{i=1}^{n} \sup_{(s,t)} |\Delta v_i(s) v_i(t)| = O_p(h_2^2),$$

$$\sup_{(s,t)} \frac{1}{n-p} \left| \sum_{i=1}^{n} x_i^\top \Delta\beta(s) v_i(t) \right| \le \frac{1}{n-p} \sum_{i=1}^{n} \|x_i\|_2 \sup_{s\in[0,1]} \|\Delta\beta(s)\|_2 \sup_{t\in[0,1]} |v_i(t)| = O_p(n^{-1/2} + h_1^2).$$

So, together with (B.5), we obtain (B.4). Similarly, we have the 3rd term $\sup_{(s,t)} (n-p)^{-1} \sum_{i=1}^{n} \Delta_i(t) v_i(s) = O_p(n^{-1/2} + h_1^2 + h_2^2 + (\log n/n)^{1/2})$. Thus, we complete the proofs of Step 1.

For Step 2, we show that

$$\sup_{(s,t)} \left| \frac{1}{n-p} \sum_{i=1}^{n} [v_i(s) v_i(t) - \gamma_v(s,t)] \right| = O_p\left(n^{-1/2} \right). \tag{B.6}$$

This can be argued by noting that

$$|v_i(s_1) v_i(t_1) - v_i(s_2) v_i(t_2)| \le 2(|s_1 - s_2| + |t_1 - t_2|) \sup_{s\in[0,1]} |v_i'(s)| \sup_{s\in[0,1]} |v_i(s)|$$

holds for any (s_1, t_1) and (s_2, t_2). Therefore the functional class $\{v(s)v(t) : (s,t) \in [0,1]^2\}$ is a VC class (van der Vaart & Wellner, 1996). So, it yields that (B.6) is true. Thus, we complete the proof of Theorem 2(ii).

APPENDIX C: PROOF OF THEOREM 3

Recall that $d_C(s) = \hat{A}_1(s)\hat{\omega}_\psi(s,s)^{-1/2} \left(C\hat{\Omega}^{-1} C^\top \right)^{-1/2} d(s)$ with $d(s) = n^{1/2}(C\hat{\beta}(s) - c(s))$. Let $\Delta(s) = d_C(s) - G_C(s)$. Thus,

$$\begin{aligned}
\Delta(s) &= \hat{A}_1(s)\hat{\omega}_\psi(s,s)^{-1/2} \left(C\hat{\Omega}^{-1} C^\top \right)^{-1/2} d(s) - A_1(s)\omega_\psi(s,s)^{-1/2} (C\Omega^{-1}C^\top)^{-1/2} CG(s) \\
&= \{\hat{A}_1(s)\hat{\omega}_\psi(s,s)^{-1/2} \left(C\hat{\Omega}^{-1} C^\top \right)^{-1/2} - A_1(s)\omega_\psi(s,s)^{-1/2} (C\Omega^{-1}C^\top)^{-1/2}\} d(s) \\
&\quad + A_1(s)\omega_\psi(s,s)^{-1/2} (C\Omega^{-1}C^\top)^{-1/2} (d(s) - CG(s)) \\
&= o_p(1)O_p(1) + O_p(1)o_p(1) = o_p(1).
\end{aligned}$$

Therefore,

$$\begin{aligned}
T_n &= \int_0^1 G_C(s)^\top G_C(s) \mathrm{d}s + 2\int_0^1 G_C(s)^\top \Delta(s) \mathrm{d}s + \int_0^1 \Delta(s)^\top \Delta(s) \mathrm{d}s \\
&= \int_0^1 G_C(s)^\top G_C(s) \mathrm{d}s + o_p(1),
\end{aligned}$$

and so the first part of the claim holds.

To complete the proof, it is sufficient to show that $\int_0^1 G_C(s)^\top G_C(s)\mathrm{d}s \to \sum_{l=1}^{l_0}\tau_l\chi^2_{r,l}$. Recall that $G_C(s)$ is a centered Gaussian process with the covariance function

$$\begin{aligned}\Sigma_C(s,t) &= A_1(s)\omega_\psi(s,s)^{-1/2}(C\Omega^{-1}C^\top)^{-1/2}C\Sigma_L(s,t)C^\top A_1(t)\omega_\psi(t,t)^{-1/2}(C\Omega^{-1}C^\top)^{-1/2}\\ &= \omega_\psi(s,s)^{-1/2}\omega_\psi(t,t)^{-1/2}\omega_\psi(s,t)I_r\\ &= \tilde{\omega}_\psi(s,t)I_r.\end{aligned}$$

Let $G_{C,l}$ denote the components of G_C so that $G_C(s) = (G_{C,1}(s),\ldots,G_{C,r}(s))$. By the above result, it is clear that the $G_{C,l}(s)$, over $l = 1,\ldots,r$, are independent, and that $G_{C,l}(s)$ follows an asymptotic Gaussian process distribution with a mean function of 0 and a covariance function of $\tilde{\omega}_\psi$.

By Condition 11, $\int_0^1\int_0^1\tilde{\omega}_\psi(s,t)\mathrm{d}s\mathrm{d}t < \infty$, so by the Cauchy–Schwarz inequality, $\tilde{\omega}_\psi$ has an eigendecomposition of the form $\tilde{\omega}_\psi(s,t) = \sum_{l=1}^{l_0}\tau_l\varphi_l(s)\varphi_l(t)$, where $(\tau_l)_{l=1}^{l_0}$ is a nonincreasing sequence of eigenvalues and $(\varphi_l)_{l=1}^{l_0}$ is a corresponding sequence of orthogonal eigenfunctions.

By the above eigendecomposition, $G_{C,l}(s) = \sum_{k=1}^{l_0}\zeta_{l,k}\varphi_k(s)$, where

$$\zeta_{l,k} = \int_0^1 G_{C,l}(s)\varphi_k(s)\mathrm{d}s \sim \mathrm{AN}(0,\tau_k),$$

and AN denotes an asymptotic normal distribution.

Consequently,

$$\begin{aligned}\int_0^1 G_C(s)^\top G_C(s)\mathrm{d}s = \int_0^1 G_C(s)^\top G_C(s)\mathrm{d}s &= \sum_{l=1}^{r}\int_0^1 G_{C,l}(s)^2\mathrm{d}s\\ &= \sum_{l=1}^{r}\sum_{k=1}^{l_0}\zeta_{l,k}^2 = \sum_{k=1}^{l_0}\sum_{l=1}^{r}\zeta_{l,k}^2 \to \sum_{k=1}^{l_0}\tau_l\chi^2_{r,l},\end{aligned}$$

which completes the proof of the theorem.

APPENDIX D: PROOF OF THEOREM 4

The following proof considers only the claim pertaining to $G_3^{(b)}$: proofs for the claims regarding $G_1^{(b)}$ and $G_2^{(b)}$ are similar.

Let $D(\hat{\beta},s) = n^{-1/2}\sum_{i=1}^n\sum_{j=1}^m U_{i,j}(\hat{\beta},s)$, $D^{(b)}(\hat{\beta},s) = n^{-1/2}\sum_{i=1}^n\varsigma_i^{(b)}\sum_{j=1}^m U_{i,j}(\hat{\beta},s)$, and

$$V(\hat{\beta},s) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m}U_{i,j}(\hat{\beta},s)U_{i,j}(\hat{\beta},s)^\top.$$

Recall that $G_3^{(b)}(s) = D^{(b)}(\hat{\beta},s)^\top V(\hat{\beta},s)^{-1}D^{(b)}(\hat{\beta},s)$ and that

$$\kappa_1(s)^2\kappa_2(s)^{-1}\omega_\psi(s,s)^{-1}G_{\omega_\psi}(s,s)^\top G_{\omega_\psi}(s,s) = G_L(s)^\top\Sigma_L(s,s)^{-1}G_L(s).$$

We will prove that $G_3^{(b)}(s) \to \kappa_1(s)^2\kappa_2(s)^{-1}\omega_\psi(s,s)^{-1}G_{\omega_\psi}(s,s)^\top G_{\omega_\psi}(s,s)$. It is sufficient to prove that $D^{(b)}(\hat{\beta},s) \to G_L(s)$ and that $V(\hat{\beta},s) = \Sigma_L(s,s) + o_p(1)$. The latter statement follows immediately

from the proof of Theorem 1. We now prove the former claim by following the arguments in Kosorok (2003) and Zhu et al. (2012).

First, we establish the weak convergence of $D^{(b)}(\hat{\beta}, s)$. Observe that

$$\begin{aligned} D^{(b)}(\hat{\beta}, s) &= n^{-1/2}\sum_{i=1}^{n}\varsigma_i^{(b)}\sum_{j=1}^{m}(mh)^{-1}x_i\sum_{k=1}^{m}\varpi(s_j,s_k)\psi(y_i(s_k)-x_i^{\top}\hat{\beta}(s))K_k^h(s) \\ &= n^{-1/2}\sum_{i=1}^{n}\varsigma_i^{(b)}x_i\sum_{j=1}^{m}(mh)^{-1}\sum_{k=1}^{m}\varpi(s_j,s_k)\psi(\eta_i(s_k))K_k^h(s) \\ &\quad + n^{-1/2}\sum_{i=1}^{n}\varsigma_i^{(b)}\sum_{j=1}^{m}(mh)^{-1}x_i\sum_{k=1}^{m}\varpi(s_j,s_k)A_1(s_k)x_i^{\top}(\beta(s_k)-\hat{\beta}(s))K_k^h(s) \\ &= J_1(s)+J_2(s), \end{aligned}$$

where the last equality defines $J_1(s)$ and $J_2(s)$ in the obvious manner. Regarding $J_1(s)$, if we treat $\varsigma_i^{(b)}$ as a "covariate vector," then similar to the proof of Theorem 1, we see that $J_1(s)$ converges to $G_L(s)$. Regarding $J_2(s)$, we have that, uniformly, $J_2(s) = O_p(h)n^{-1/2}\sum_{i=1}^{n}\varsigma_i^{(b)}x_ix_i^{\top}$ and $E(J_2(s)) = 0$. Thus, $J_2(s) = o_p(1)$ uniformly for $s \in [0,1]$. We can thus conclude that $D^{(b)}(\hat{\beta}, s)$ converges to $G_L(s)$ in distribution.

We now consider the weak convergence of $D^{(b)}(\hat{\beta}, s)$ conditional on the data $\mathcal{D} = \{\mathcal{X}, \mathcal{S}\}$, where $\mathcal{X} = \{x_1,\ldots,x_n\}$. By standard calculations, $D^{(b)}(\hat{\beta}, \cdot)$ is a Gaussian process with a mean function of 0 and a covariance function that evaluates to $\mathrm{cov}(J_1(s), J_1(t)|\mathcal{D}) + \mathrm{cov}(J_1(s), J_2(t)|\mathcal{D}) + \mathrm{cov}(J_2(s), J_1(t)|\mathcal{D}) + \mathrm{cov}(J_2(s), J_2(t)|\mathcal{D})$ at (s,t). Now,

$$\begin{aligned} &\mathrm{cov}(J_1(s), J_1(t)|\mathcal{D}) \\ &= \frac{1}{n}\sum_{i=1}^{n}x_ix_i^{\top}\left\{\sum_{j=1}^{m}(mh)^{-1}\sum_{k=1}^{m}\varpi(s_j,s_k)\psi(\eta_i(s_k))K_k^h(s)\right\} \\ &\quad\times\left\{\sum_{j=1}^{m}(mh)^{-1}\sum_{k=1}^{m}\varpi(s_j,s_k)\psi(\eta_i(s_k))K_k^h(t)\right\} \\ &= \frac{1}{n}\sum_{i=1}^{n}x_ix_i^{\top}\kappa_1(s)\pi(s)\mu_0(K;s,h)\psi(\eta_i(s))\kappa_1(t)\pi(t)\mu_0(K;t,h)\psi(\eta_i(t))(1+o_p(1)). \end{aligned}$$

Furthermore, $E(\mathrm{cov}(J_1(s), J_1(t)|\mathcal{D})) = \Sigma_L(s,t)$ and $\mathrm{cov}(\mathrm{cov}(J_1(s), J_1(t)|\mathcal{D})) = O\left(n^{-1}\right)$. Thus, $\mathrm{cov}_{\varsigma}(D^{(b)}(s), D^{(b)}(t)|\mathcal{D})$, calculated with respect to the ς_is, converges to $\Sigma_L(s,t)$ in probability. We can then establish the marginal convergence of $D^{(b)}(\hat{\beta}, s)$ in the conditional central limit theorem by applying the Cramér–Wald method.

For $\delta > 0$, let $\mathcal{S}_\delta = \{l\delta | l = 0,\ldots,\lfloor\delta^{-1}\rfloor\}$ be a δ-grid and let $[0,1]_\delta(s)$, for any $s \in [0,1]$, denote an element of $\mathcal{S}_\delta$ that is closest to s. For a metric space $(\mathcal{M}, d)$, let $\mathrm{BL}_1(\mathcal{M})$ denote the space of real-valued functions on $\mathcal{M}$ with a Lipschitz norm bounded by one, i.e., for any $g \in \mathrm{BL}_1(\mathcal{M})$, $\sup_{x\in\mathcal{M}}|g(x)| \le 1$ and $|g(x) - g(y)| \le d(x,y)$ for all $x, y \in \mathcal{M}$. We have that

$$\sup_{g\in\mathrm{BL}_1(\ell^\infty[0,1])}|E_\varsigma(g(D^{(b)}([0,1]_\delta))) - E(g(G_L([0,1]_\delta)))| \to 0,$$

in probability as $n \to \infty$. Since G_L is continuous, $G_L([0,1]_\delta(s)) \to G_L(s)$ almost surely as $\delta \to 0$, that is,

$$\lim_{\delta \to 0^+} \sup_{g \in \mathrm{BL}_1(\ell^\infty[0,1])} |E_\varsigma(g(D^{(b)}([0,1]_\delta))) - E(g(G_L([0,1])))| = 0.$$

Additionally,

$$\sup_{g \in \mathrm{BL}_1(\ell^\infty[0,1])} |E_\varsigma(g(D^{(b)}([0,1]_\delta(\cdot)))) - E_\varsigma(g(D^{(b)}([0,1](\cdot))))|$$
$$\leq E_\varsigma\left(\sup_{\substack{s,t \in [0,1] \\ |s-t|<\delta}} |D^{(b)}(s) - D^{(b)}(t)| \right).$$

The expectation on the right-hand side of the above inequality was computed previously in this proof. This completes the proof of Theorem 4.