The Annals of Statistics
2013, Vol. 41, No. 5, 2608–2638
DOI: 10.1214/13-AOS1164

LOCAL AND GLOBAL ASYMPTOTIC INFERENCE IN SMOOTHING SPLINE MODELS

BY ZUOFENG SHANG AND GUANG CHENG[1]

University of Notre Dame and Purdue University

This article studies local and global inference for smoothing spline estimation in a unified asymptotic framework. We first introduce a new technical tool called functional Bahadur representation, which significantly generalizes the traditional Bahadur representation in parametric models, that is, Bahadur [*Ann. Inst. Statist. Math.* **37** (1966) 577–580]. Equipped with this tool, we develop four interconnected procedures for inference: (i) pointwise confidence interval; (ii) local likelihood ratio testing; (iii) simultaneous confidence band; (iv) global likelihood ratio testing. In particular, our confidence intervals are proved to be asymptotically valid at any point in the support, and they are shorter on average than the Bayesian confidence intervals proposed by Wahba [*J. R. Stat. Soc. Ser. B Stat. Methodol.* **45** (1983) 133–150] and Nychka [*J. Amer. Statist. Assoc.* **83** (1988) 1134–1143]. We also discuss a version of the Wilks phenomenon arising from local/global likelihood ratio testing. It is also worth noting that our simultaneous confidence bands are the first ones applicable to general quasi-likelihood models. Furthermore, issues relating to optimality and efficiency are carefully addressed. As a by-product, we discover a surprising relationship between periodic and nonperiodic smoothing splines in terms of inference.

1. Introduction. As a flexible modeling tool, smoothing splines provide a general framework for statistical analysis in a variety of fields; see [13, 41, 42]. The asymptotic studies on smoothing splines in the literature focus primarily on the estimation performance, and in particular the global convergence. However, in practice it is often of great interest to conduct *asymptotic inference* on the unknown functions. The procedures for inference developed in this article, together with their rigorously derived asymptotic properties, fill this long-standing gap in the smoothing spline literature.

Received August 2013.
[1]Sponsored by NSF Grant DMS-09-06497 and CAREER Award DMS-11-51692.
AMS 2000 subject classifications. Primary 62G20, 62F25; secondary 62F15, 62F12.
Key words and phrases. Asymptotic normality, functional Bahadur representation, local/global likelihood ratio test, simultaneous confidence band, smoothing spline.

As an illustration, consider two popular nonparametric regression models: (i) normal regression: $Y \mid Z = z \sim N(g_0(z), \sigma^2)$ for some unknown $\sigma^2 > 0$; (ii) logistic regression: $P(Y = 1 \mid Z = z) = \exp(g_0(z))/(1 + \exp(g_0(z)))$. The function g_0 is assumed to be smooth in both models. Our goal in this paper is to develop asymptotic theory for constructing pointwise confidence intervals and simultaneous confidence bands for g_0, testing on the value of $g_0(z_0)$ at any given point z_0, and testing whether g_0 satisfies certain global properties such as linearity. Pointwise confidence intervals and tests on a local value are known as local inference. Simultaneous confidence bands and tests on a global property are known as global inference. To the best of our knowledge, there has been little systematic and rigorous theoretical study of asymptotic inference. This is partly because of the technical restrictions of the widely used equivalent kernel method. The *functional Bahadur representation* (FBR) developed in this paper makes several important contributions to this area. Our main contribution is a set of procedures for local and global inference for a univariate smooth function in a general class of nonparametric regression models that cover both the aforementioned cases. Moreover, we investigate issues relating to optimality and efficiency that have not been treated in the existing smoothing spline literature.

The equivalent kernel has long been used as a standard tool for handling the asymptotic properties of smoothing spline estimators, but this method is restricted to least square regression; see [26, 35]. Furthermore, the equivalent kernel only "approximates" the reproducing kernel function, and the approximation formula becomes extremely complicated when the penalty order increases or the design points are nonuniform. To analyze the smoothing spline estimate in a more effective way, we employ empirical process theory to develop a new technical tool, the functional Bahadur representation, which directly handles the "exact" reproducing kernel, and makes it possible to study asymptotic inference in a broader range of nonparametric models. An immediate consequence of the FBR is the asymptotic normality of the smoothing spline estimate. This naturally leads to the construction of pointwise asymptotic confidence intervals (CIs). The classical Bayesian CIs in the literature [28, 40] are valid on average over the observed covariates. However, our CIs are proved to be theoretically valid at any point, and they even have shorter lengths than the Bayesian CIs. We next introduce a likelihood ratio method for testing the local value of a regression function. It is shown that the null limiting distribution is a scaled Chi-square with one degree of freedom, and that the scaling constant converges to one as the smoothness level of the regression function increases. Therefore, we have discovered an interesting Wilks phenomenon (meaning that the asymptotic null distribution is free of nuisance parameters) arising from the proposed nonparametric local testing.

Procedures for global inference are also useful. Simultaneous confidence bands (SCBs) accurately depict the global behavior of the regression function, and they have been extensively studied in the literature. However, most of the efforts were devoted to simple regression models with additive Gaussian errors based on kernel or local polynomial estimates; see [5, 11, 17, 38, 44]. By incorporating the reproducing kernel Hilbert space (RKHS) theory into [2], we obtain an SCB applicable to general nonparametric regression models, and we demonstrate its theoretical validity based on strong approximation techniques. To the best of our knowledge, this is the first SCB ever developed for a general nonparametric regression model in smoothing spline settings. We further demonstrate that our SCB is optimal in the sense that the minimum width of the SCB achieves the lower bound established by [12]. Model assessment is another important aspect of global inference. Fan et al. [9] used local polynomial estimates for testing nonparametric regression models, namely the generalized likelihood ratio test (GLRT). Based on smoothing spline estimates, we propose an alternative method called the penalized likelihood ratio test (PLRT), and we identify its null limiting distribution as nearly Chi-square with diverging degrees of freedom. Therefore, the Wilks phenomenon holds for the global test as well. More importantly, we show that the PLRT achieves the minimax rate of testing in the sense of [19]. In comparison, other smoothing-spline-based tests such as the locally most powerful (LMP) test, the generalized cross validation (GCV) test and the generalized maximum likelihood ratio (GML) test (see [4, 6, 20, 23, 30, 41]) either lead to complicated null distributions with nuisance parameters or are not known to be optimal.

As a by-product, we derive the asymptotic equivalence of the proposed procedures based on periodic and nonperiodic smoothing splines under mild conditions; see Remark 5.2. In general, our findings reveal an intrinsic connection between the two rather different basis structures, which in turn facilitates the implementation of inference.

Our paper is mainly devoted to theoretical studies. However, a few practical issues are noteworthy. GCV is currently used for the empirical tuning of the smoothing parameter, and it is known to result in biased estimates if the true function is spatially inhomogeneous with peaks and troughs. Moreover, the penalty order is prespecified rather than data-driven. Future research is needed to develop an efficient method for choosing a suitable smoothing parameter for bias reduction and an empirical method for quantifying the penalty order through data. We also note that some of our asymptotic procedures are not fully automatic since certain quantities need to be estimated; see Example 6.3. A large sample size may be necessary for the benefits of our asymptotic methods to become apparent. Finally, we want to mention that extensions to more complicated models such as multivariate smoothing spline models and semiparametric models are conceptually feasible by applying similar FBR techniques and likelihood-based approaches.

The rest of this paper is organized as follows. Section 2 introduces the basic notation, the model assumptions, and some preliminary RKHS results. Section 3 presents the FBR and the local asymptotic results. In Sections 4 and 5, several procedures for local and global inference together with their theoretical properties are formally discussed. In Section 6, we give three concrete examples to illustrate our theory. Numerical studies are also provided for both periodic and nonperiodic splines. The proofs are included in an online supplementary document [33].

2. Preliminaries.

2.1. *Notation and assumptions.* Suppose that the data $T_i = (Y_i, Z_i)$, $i = 1, \ldots, n$, are i.i.d. copies of $T = (Y, Z)$, where $Y \in \mathcal{Y} \subseteq \mathbb{R}$ is the response variable, $Z \in \mathbb{I}$ is the covariate variable and $\mathbb{I} = [0, 1]$. Consider a general class of nonparametric regression models under the primary assumption

$$\mu_0(Z) \equiv E(Y \mid Z) = F(g_0(Z)), \tag{2.1}$$

where $g_0(\cdot)$ is some unknown smooth function and $F(\cdot)$ is a known link function. This framework covers two subclasses of statistical interest. The first subclass assumes that the data are modeled by $y_i \mid z_i \sim p(y_i; \mu_0(z_i))$ for a conditional distribution $p(y; \mu_0(z))$ unknown up to μ_0. Instead of assuming known distributions, the second subclass specifies the relation between the conditional mean and variance as $\mathrm{Var}(Y \mid Z) = \mathcal{V}(\mu_0(Z))$, where $\mathcal{V}$ is a known positive-valued function. The nonparametric estimation of g in the second situation uses the quasi-likelihood $Q(y; \mu) \equiv \int_y^\mu (y - s)/\mathcal{V}(s)\, ds$ as an objective function (see [43]), where $\mu = F(g)$. Despite distinct modeling principles, the two subclasses have a large overlap since $Q(y; \mu)$ coincides with $\log p(y; \mu)$ under many common combinations of $(F, \mathcal{V})$; see Table 2.1 of [25].

From now on, we focus on a smooth criterion function $\ell(y; a) : \mathcal{Y} \times \mathbb{R} \mapsto \mathbb{R}$ that covers the above two cases, that is, $\ell(y; a) = Q(y; F(a))$ or $\log p(y; F(a))$. Throughout this paper, we define the functional parameter space $\mathcal{H}$ as the mth-order Sobolev space:

$$H^m(\mathbb{I}) \equiv \{g : \mathbb{I} \mapsto \mathbb{R} \mid g^{(j)} \text{ is absolutely continuous}$$
$$\text{for } j = 0, 1, \ldots, m-1 \text{ and } g^{(m)} \in L_2(\mathbb{I})\},$$

where m is assumed to be known and larger than $1/2$. With some abuse of notation, $\mathcal{H}$ may also refer to the homogeneous subspace $H_0^m(\mathbb{I})$ of $H^m(\mathbb{I})$. The space $H_0^m(\mathbb{I})$ is also known as the class of periodic functions such that a function $g \in H_0^m(\mathbb{I})$ has the additional restrictions $g^{(j)}(0) = g^{(j)}(1)$ for $j = 0, 1, \ldots, m-1$. Let $J(g, \widetilde{g}) = \int_{\mathbb{I}} g^{(m)}(z) \widetilde{g}^{(m)}(z)\, dz$. Consider the penalized nonparametric estimate $\widehat{g}_{n,\lambda}$:

$$\begin{aligned}\widehat{g}_{n,\lambda} &= \arg\max_{g\in\mathcal{H}} \ell_{n,\lambda}(g) \\ &= \arg\max_{g\in\mathcal{H}} \left\{ \frac{1}{n}\sum_{i=1}^{n} \ell(Y_i; g(Z_i)) - (\lambda/2)J(g,g) \right\},\end{aligned} \tag{2.2}$$

where $J(g,g)$ is the roughness penalty and λ is the smoothing parameter, which converges to zero as $n \to \infty$. We use $\lambda/2$ (rather than λ) to simplify future expressions. The existence and uniqueness of $\widehat{g}_{n,\lambda}$ are guaranteed by Theorem 2.9 of [13] when the null space $\mathcal{N}_m \equiv \{g \in \mathcal{H} : J(g,g) = 0\}$ is finite dimensional and $\ell(y;a)$ is concave and continuous w.r.t. a.

We next assume a set of model conditions. Let $\mathcal{I}_0$ be the range of g_0, which is obviously compact. Denote the first-, second- and third-order derivatives of $\ell(y;a)$ w.r.t. a by $\dot{\ell}_a(y;a)$, $\ddot{\ell}_a(y;a)$ and $\ell_a'''(y;a)$, respectively. We assume the following smoothness and tail conditions on ℓ:

ASSUMPTION A.1. (a) $\ell(y;a)$ is three times continuously differentiable and concave w.r.t. a. There exists a bounded open interval $\mathcal{I} \supset \mathcal{I}_0$ and positive constants C_0 and C_1 s.t.

$$E\left\{\exp\left(\sup_{a\in\mathcal{I}} |\ddot{\ell}_a(Y;a)|/C_0\right) \mid Z\right\} \le C_0 \qquad \text{a.s.} \tag{2.3}$$

and

$$E\left\{\exp\left(\sup_{a\in\mathcal{I}} |\ell_a'''(Y;a)|/C_0\right) \mid Z\right\} \le C_1 \qquad \text{a.s.} \tag{2.4}$$

(b) There exists a positive constant C_2 such that $C_2^{-1} \le I(Z) \equiv -E(\ddot{\ell}_a(Y; g_0(Z)) \mid Z) \le C_2$ a.s.

(c) $\epsilon \equiv \dot{\ell}_a(Y; g_0(Z))$ satisfies $E(\epsilon \mid Z) = 0$ and $E(\epsilon^2 \mid Z) = I(Z)$ a.s.

Assumption A.1(a) implies the slow diverging rate $O_P(\log n)$ of

$$\max_{1\le i\le n} \sup_{a\in\mathcal{I}} |\ddot{\ell}_a(Y_i;a)| \quad \text{and} \quad \max_{1\le i\le n} \sup_{a\in\mathcal{I}} |\ell_a'''(Y_i;a)|.$$

When $\ell(y;a) = \log p(y;a)$, Assumption A.1(b) imposes boundedness and positive definiteness of the Fisher information, and Assumption A.1(c) trivially holds if p satisfies certain regularity conditions. When $\ell(y;a) = Q(y; F(a))$, we have

$$\ddot{\ell}_a(Y;a) = F_1(a) + \varepsilon F_2(a) \quad \text{and} \quad \ell_a'''(Y;a) = \dot{F}_1(a) + \varepsilon \dot{F}_2(a), \tag{2.5}$$

where $\varepsilon = Y - \mu_0(Z)$, $F_1(a) = -|\dot{F}(a)|^2/\mathcal{V}(F(a)) + (F(g_0(Z)) - F(a))F_2(a)$ and $F_2(a) = (\ddot{F}(a)\mathcal{V}(F(a)) - \dot{\mathcal{V}}(F(a))|\dot{F}(a)|^2)/\mathcal{V}^2(F(a))$. Hence, Assumption A.1(a) holds if $F_j(a)$, $\dot{F}_j(a)$, $j = 1, 2$, are all bounded over $a \in \mathcal{I}$ and

$$E\{\exp(|\varepsilon|/C_0) \mid Z\} \le C_1 \qquad \text{a.s.} \tag{2.6}$$

By (2.5), we have $I(Z) = |\dot{F}(g_0(Z))|^2/\mathcal{V}(F(g_0(Z)))$. Thus, Assumption A.1(b) holds if

$$1/C_2 \le \frac{|\dot{F}(a)|^2}{\mathcal{V}(F(a))} \le C_2 \qquad \text{for all } a \in \mathcal{I}_0 \text{ a.s.} \tag{2.7}$$

Assumption A.1(c) follows from the definition of $\mathcal{V}(\cdot)$. The sub-exponential tail condition (2.6) and the boundedness condition (2.7) are very mild quasi-likelihood model assumptions (e.g., also assumed in [24]). The assumption that F_j and $\dot{F}_j$ are both bounded over $\mathcal{I}$ could be restrictive and can be removed in many cases, such as the binary logistic regression model, by applying empirical process arguments similar to those in Section 7 of [24].

2.2. *Reproducing kernel Hilbert space.* We now introduce a number of RKHS results as extensions of [7] and [29]. It is well known that, when $m > 1/2$, $\mathcal{H} = H^m(\mathbb{I})$ [or $H_0^m(\mathbb{I})$] is an RKHS endowed with the inner product $\langle g, \widetilde{g}\rangle = E\{I(Z)g(Z)\widetilde{g}(Z)\} + \lambda J(g, \widetilde{g})$ and the norm

$$\|g\|^2 = \langle g, g\rangle. \tag{2.8}$$

The reproducing kernel $K(z_1, z_2)$ defined on $\mathbb{I} \times \mathbb{I}$ is known to have the following property:

$$K_z(\cdot) \equiv K(z, \cdot) \in \mathcal{H} \quad \text{and} \quad \langle K_z, g\rangle = g(z) \qquad \text{for any } z \in \mathbb{I} \text{ and } g \in \mathcal{H}.$$

Obviously, K is symmetric with $K(z_1, z_2) = K(z_2, z_1)$. We further introduce a positive definite self-adjoint operator $W_\lambda : \mathcal{H} \mapsto \mathcal{H}$ such that

$$\langle W_\lambda g, \widetilde{g}\rangle = \lambda J(g, \widetilde{g}) \tag{2.9}$$

for any $g, \widetilde{g} \in \mathcal{H}$. Let $V(g, \widetilde{g}) = E\{I(Z)g(Z)\widetilde{g}(Z)\}$. Then $\langle g, \widetilde{g}\rangle = V(g, \widetilde{g}) + \langle W_\lambda g, \widetilde{g}\rangle$ and $V(g, \widetilde{g}) = \langle (\mathrm{id} - W_\lambda)g, \widetilde{g}\rangle$, where id denotes the identity operator.

Next, we assume that there exists a sequence of basis functions in the space $\mathcal{H}$ that simultaneously diagonalizes the bilinear forms V and J. Such eigenvalue/eigenfunction assumptions are typical in the smoothing spline literature, and they are critical to control the local behavior of the penalized estimates. Hereafter, we denote positive sequences a_μ and b_μ as $a_\mu \asymp b_\mu$ if they satisfy $\lim_{\mu\to\infty}(a_\mu/b_\mu) = c > 0$. If $c = 1$, we write $a_\mu \sim b_\mu$. Let $\sum_\nu$ denote the sum over $\nu \in \mathbb{N} = \{0, 1, 2, \ldots\}$ for convenience. Denote the sup-norm of $g \in \mathcal{H}$ as $\|g\|_{\sup} = \sup_{z\in\mathbb{I}} |g(z)|$.

ASSUMPTION A.2. There exists a sequence of eigenfunctions $h_\nu \in \mathcal{H}$ satisfying $\sup_{\nu\in\mathbb{N}} \|h_\nu\|_{\sup} < \infty$, and a nondecreasing sequence of eigenvalues $\gamma_\nu \asymp \nu^{2m}$ such that

$$V(h_\mu, h_\nu) = \delta_{\mu\nu}, \qquad J(h_\mu, h_\nu) = \gamma_\mu \delta_{\mu\nu}, \qquad \mu, \nu \in \mathbb{N}, \tag{2.10}$$

where $\delta_{\mu\nu}$ is the Kronecker's delta. In particular, any $g \in \mathcal{H}$ admits a Fourier expansion $g = \sum_\nu V(g, h_\nu)h_\nu$ with convergence in the $\|\cdot\|$-norm.

Assumption A.2 enables us to derive explicit expressions for $\|g\|$, $K_z(\cdot)$ and $W_\lambda h_\nu(\cdot)$ for any $g \in \mathcal{H}$ and $z \in \mathbb{I}$; see Proposition 2.1 below.

PROPOSITION 2.1. *For any $g \in \mathcal{H}$ and $z \in \mathbb{I}$, we have $\|g\|^2 = \sum_\nu |V(g, h_\nu)|^2(1+\lambda\gamma_\nu)$, $K_z(\cdot) = \sum_\nu \frac{h_\nu(z)}{1+\lambda\gamma_\nu} h_\nu(\cdot)$ and $W_\lambda h_\nu(\cdot) = \frac{\lambda\gamma_\nu}{1+\lambda\gamma_\nu} h_\nu(\cdot)$ under Assumption A.2.*

For future theoretical derivations, we need to figure out the underlying eigensystem that implies Assumption A.2. For example, when $\ell(y;a) = -(y-a)^2/2$ and $\mathcal{H} = H_0^m(\mathbb{I})$, Assumption A.2 is known to be satisfied if (γ_ν, h_ν) is chosen as the trigonometric polynomial basis specified in (6.2) of Example 6.1. For general $\ell(y;a)$ with $\mathcal{H} = H^m(\mathbb{I})$, Proposition 2.2 below says that Assumption A.2 is still valid if (γ_ν, h_ν) is chosen as the (normalized) solution of the following equations:

$$(-1)^m h_\nu^{(2m)}(\cdot) = \gamma_\nu I(\cdot)\pi(\cdot)h_\nu(\cdot), \qquad h_\nu^{(j)}(0) = h_\nu^{(j)}(1) = 0, \quad j = m, m+1, \ldots, 2m-1, \tag{2.11}$$

where $\pi(\cdot)$ is the marginal density of the covariate Z. Proposition 2.2 can be viewed as a nontrivial extension of [39], which assumes $I = \pi = 1$. The proof relies substantially on the ODE techniques developed in [3, 36]. Let $C^m(\mathbb{I})$ be the class of the mth-order continuously differentiable functions over $\mathbb{I}$.

PROPOSITION 2.2. *If $\pi(z), I(z) \in C^{2m-1}(\mathbb{I})$ are both bounded away from zero and infinity over $\mathbb{I}$, then the eigenvalues γ_ν and the corresponding eigenfunctions h_ν, found from (2.11) and normalized to $V(h_\nu, h_\nu) = 1$, satisfy Assumption A.2.*

Finally, for later use we summarize the notation for Fréchet derivatives. Let Δg, $\Delta g_j \in \mathcal{H}$ for $j = 1, 2, 3$. The Fréchet derivative of $\ell_{n,\lambda}$ can be identified as

$$\begin{aligned} D\ell_{n,\lambda}(g)\Delta g &= \frac{1}{n}\sum_{i=1}^n \dot{\ell}_a(Y_i; g(Z_i))\langle K_{Z_i}, \Delta g\rangle - \langle W_\lambda g, \Delta g\rangle \\ &\equiv \langle S_n(g), \Delta g\rangle - \langle W_\lambda g, \Delta g\rangle \\ &\equiv \langle S_{n,\lambda}(g), \Delta g\rangle. \end{aligned}$$

Note that $S_{n,\lambda}(\widehat{g}_{n,\lambda}) = 0$ and $S_{n,\lambda}(g_0)$ can be expressed as

$$S_{n,\lambda}(g_0) = \frac{1}{n}\sum_{i=1}^n \epsilon_i K_{Z_i} - W_\lambda g_0. \tag{2.12}$$

The Fréchet derivative of $S_{n,\lambda}$ ($DS_{n,\lambda}$) is denoted $DS_{n,\lambda}(g)\Delta g_1 \Delta g_2 (D^2 S_{n,\lambda}(g) \times \Delta g_1 \Delta g_2 \Delta g_3)$. These derivatives can be explicitly written as $D^2\ell_{n,\lambda}(g)\Delta g_1 \Delta g_2 =$

$n^{-1}\sum_{i=1}^{n}\ddot{\ell}_a(Y_i;g(Z_i))\langle K_{Z_i},\Delta g_1\rangle\langle K_{Z_i},\Delta g_2\rangle - \langle W_\lambda \Delta g_1,\Delta g_2\rangle$ [or $D^3\ell_{n,\lambda}(g)\times \Delta g_1\Delta g_2\Delta g_3 = n^{-1}\sum_{i=1}^{n}\ell_a'''(Y_i;g(Z_i))\langle K_{Z_i},\Delta g_1\rangle\langle K_{Z_i},\Delta g_2\rangle\langle K_{Z_i},\Delta g_3\rangle$].

Define $S(g)=E\{S_n(g)\}$, $S_\lambda(g)=S(g)-W_\lambda g$ and $DS_\lambda(g)=DS(g)-W_\lambda$, where $DS(g)\Delta g_1\Delta g_2 = E\{\ddot{\ell}_a(Y;g(Z))\langle K_Z,\Delta g_1\rangle\langle K_Z,\Delta g_2\rangle\}$. Since $\langle DS_\lambda(g_0)f, g\rangle = -\langle f,g\rangle$ for any $f,g\in\mathcal{H}$, we have the following result:

PROPOSITION 2.3. *$DS_\lambda(g_0)=-\mathrm{id}$, where id is the identity operator on $\mathcal{H}$.*

3. Functional Bahadur representation.

In this section, we first develop the key technical tool of this paper: *functional Bahadur representation*, and we then present the local asymptotics of the smoothing spline estimate as a straightforward application. In fact, FBR provides a rigorous theoretical foundation for the procedures for inference developed in Sections 4 and 5.

3.1. *Functional Bahadur representation.*

We first present a relationship between the norms $\|\cdot\|_{\sup}$ and $\|\cdot\|$ in Lemma 3.1 below, and we then derive a *concentration inequality* in Lemma 3.2 as the preliminary step for obtaining the FBR. For convenience, we denote $\lambda^{1/(2m)}$ as h.

LEMMA 3.1. *There exists a constant $c_m>0$ s.t. $|g(z)|\le c_m h^{-1/2}\|g\|$ for any $z\in\mathbb{I}$ and $g\in\mathcal{H}$. In particular, c_m is not dependent on the choice of z and g. Hence, $\|g\|_{\sup}\le c_m h^{-1/2}\|g\|$.*

Define $\mathcal{G}=\{g\in\mathcal{H}: \|g\|_{\sup}\le 1, J(g,g)\le c_m^{-2}h\lambda^{-1}\}$, where c_m is specified in Lemma 3.1. Recall that $\mathcal{T}$ denotes the domain of the full data variable $T=(Y,Z)$. We now prove a concentration inequality on the empirical process $Z_n(g)$ defined, for any $g\in\mathcal{G}$ and $z\in\mathbb{I}$ as

$$Z_n(g)(z)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\psi_n(T_i;g)K_{Z_i}(z)-E(\psi_n(T;g)K_Z(z))], \tag{3.1}$$

where $\psi_n(T;g)$ is a real-valued function (possibly depending on n) defined on $\mathcal{T}\times\mathcal{G}$.

LEMMA 3.2. *Suppose that ψ_n satisfies the following Lipschitz continuity condition:*

$$|\psi_n(T;f)-\psi_n(T;g)|\le c_m^{-1}h^{1/2}\|f-g\|_{\sup} \qquad \text{for any } f,g\in\mathcal{G}, \tag{3.2}$$

where c_m is specified in Lemma 3.1. Then we have

$$\lim_{n\to\infty} P\left(\sup_{g\in\mathcal{G}}\frac{\|Z_n(g)\|}{h^{-(2m-1)/(4m)}\|g\|_{\sup}^{1-1/(2m)}+n^{-1/2}}\le(5\log\log n)^{1/2}\right)=1.$$

To obtain the FBR, we need to further assume a proper convergence rate for $\widehat{g}_{n,\lambda}$:

ASSUMPTION A.3. $\|\widehat{g}_{n,\lambda} - g_0\| = O_P((nh)^{-1/2} + h^m)$.

Simple (but not necessarily the weakest) sufficient conditions for Assumption A.3 are provided in Proposition 3.3 below. Before stating this result, we introduce another norm on the space $\mathcal{H}$, that is, more commonly used in functional analysis. For any $g \in \mathcal{H}$, define

$$\|g\|_{\mathcal{H}}^2 = E\{I(Z)g(Z)^2\} + J(g,g). \tag{3.3}$$

When $\lambda \le 1$, $\|\cdot\|_{\mathcal{H}}$ is a type of Sobolev norm dominating $\|\cdot\|$ defined in (2.8). Denote

$$\lambda^* \asymp n^{-2m/(2m+1)} \qquad \text{or equivalently, } h^* \asymp n^{-1/(2m+1)}. \tag{3.4}$$

Note that λ^* is known as the optimal order when we estimate $g_0 \in \mathcal{H}$.

PROPOSITION 3.3. *Suppose that Assumption A.1 holds, and further $\|\widehat{g}_{n,\lambda} - g_0\|_{\mathcal{H}} = o_P(1)$. If h satisfies $(n^{1/2}h)^{-1}(\log\log n)^{m/(2m-1)}(\log n)^{2m/(2m-1)} = o(1)$, then Assumption A.3 is satisfied. In particular, $\widehat{g}_{n,\lambda}$ achieves the optimal rate of convergence, that is, $O_P(n^{-m/(2m+1)})$, when $\lambda = \lambda^*$.*

Classical results on rates of convergence are obtained through either linearization techniques, for example, [7], or quadratic approximation devices, for example, [13, 14]. However, the proof of Proposition 3.3 relies on empirical process techniques. Hence, it is not surprising that Proposition 3.3 requires a different set of conditions than those used in [7, 13, 14], although the derived convergence rates are the same and in all approaches the optimal rate is achieved when $\lambda = \lambda^*$. For example, Cox and O'Sullivan [7] assumed a weaker smoothness condition on the likelihood function but a more restrictive condition on h, that is, $(n^{1/2}h\lambda^{\alpha})^{-1} = o(1)$ for some $\alpha > 0$.

Now we are ready to present the key technical tool: *functional Bahadur representation*, which is also of independent interest. Shang [32] developed a different form of Bahadur representation, which is of limited use in practice. This is due to the intractable form of the inverse operator $DS_\lambda(g_0)^{-1}$, constructed based on a different type of Sobolev norm. However, by incorporating λ into the norm (2.8), we can show $DS_\lambda(g_0)^{-1} = -\text{id}$ based on Proposition 2.3, and thus obtain a more refined version of the representation of [32] that naturally applies to our general setting for inference purposes.

THEOREM 3.4 (Functional Bahadur representation). *Suppose that Assumptions A.1–A.3 hold, $h = o(1)$ and $nh^2 \to \infty$. Recall that $S_{n,\lambda}(g_0)$ is defined in (2.12). Then we have*

$$\|\widehat{g}_{n,\lambda} - g_0 - S_{n,\lambda}(g_0)\| = O_P(a_n \log n), \tag{3.5}$$

where $a_n = n^{-1/2}((nh)^{-1/2} + h^m)h^{-(6m-1)/(4m)}(\log\log n)^{1/2} + C_\ell h^{-1/2}((nh)^{-1} + h^{2m})/\log n$ *and* $C_\ell = \sup_{z\in\mathbb{I}} E\{\sup_{a\in\mathcal{I}} |\ell_a'''(Y;a)| \mid Z = z\}$. *When* $h = h^*$, *the RHS of (3.5) is* $o_P(n^{-m/(2m+1)})$.

3.2. *Local asymptotic behavior.* In this section, we obtain the pointwise asymptotics of $\widehat{g}_{n,\lambda}$ as a direct application of the FBR. The equivalent kernel method may be used for this purpose, but it is restricted to L_2 regression, for example, [35]. However, the FBR-based proof applies to more general regression. Notably, our results reveal that several well-known global convergence properties continue to hold locally.

THEOREM 3.5 (General regression). *Assume Assumptions A.1–A.3, and suppose* $h = o(1)$, $nh^2 \to \infty$ *and* $a_n \log n = o(n^{-1/2})$, *where* a_n *is defined in Theorem 3.4, as* $n \to \infty$. *Furthermore, for any* $z_0 \in \mathbb{I}$,

$$hV(K_{z_0}, K_{z_0}) \to \sigma_{z_0}^2 \qquad \text{as } n \to \infty. \tag{3.6}$$

Let $g_0^* = (\mathrm{id} - W_\lambda)g_0$ *be the biased "true parameter." Then we have*

$$\sqrt{nh}(\widehat{g}_{n,\lambda}(z_0) - g_0^*(z_0)) \xrightarrow{d} N(0, \sigma_{z_0}^2), \tag{3.7}$$

where

$$\sigma_{z_0}^2 = \lim_{h\to 0} \sum_\nu \frac{h|h_\nu(z_0)|^2}{(1+\lambda\gamma_\nu)^2}. \tag{3.8}$$

From Theorem 3.5, we immediately obtain the following result.

COROLLARY 3.6. *Suppose that the conditions in Theorem 3.5 hold and*

$$\lim_{n\to\infty} (nh)^{1/2}(W_\lambda g_0)(z_0) = -b_{z_0}. \tag{3.9}$$

Then we have

$$\sqrt{nh}(\widehat{g}_{n,\lambda}(z_0) - g_0(z_0)) \xrightarrow{d} N(b_{z_0}, \sigma_{z_0}^2), \tag{3.10}$$

where $\sigma_{z_0}^2$ *is defined as in (3.8).*

To illustrate Corollary 3.6 in detail, we consider L_2 regression in which $W_\lambda g_0(z_0)$ (also b_{z_0}) has an explicit expression under the additional boundary conditions:

$$g_0^{(j)}(0) = g_0^{(j)}(1) = 0 \qquad \text{for } j = m, \dots, 2m-1. \tag{3.11}$$

In fact, (3.11) is also the price we pay for obtaining the boundary results, that is, $z_0 = 0, 1$. However, (3.11) could be too strong in practice. Therefore, we provide an alternative set of conditions in (3.14) below, which can be implied by the so-called "exponential envelope condition" introduced in [29]. In Corollary 3.7 below, we consider two different cases: $b_{z_0} \neq 0$ and $b_{z_0} = 0$.

COROLLARY 3.7 (L_2 regression). *Let* $m > (3+\sqrt{5})/4 \approx 1.309$ *and* $\ell(y;a) = -(y-a)^2/2$. *Suppose that Assumption A.3 and (3.6) hold, and the normalized eigenfunctions* h_ν *satisfy (2.11). Assume that* $g_0 \in H^{2m}(\mathbb{I})$ *satisfies* $\sum_\nu |V(g_0^{(2m)}, h_\nu)h_\nu(z_0)| < \infty$.

(i) *Suppose* g_0 *satisfies the boundary conditions (3.11). If* $h/n^{-1/(4m+1)} \to c > 0$, *then we have, for any* $z_0 \in [0,1]$,

$$\sqrt{nh}(\widehat{g}_{n,\lambda}(z_0) - g_0(z_0)) \xrightarrow{d} N((-1)^{m-1}c^{2m}g_0^{(2m)}(z_0)/\pi(z_0), \sigma_{z_0}^2). \tag{3.12}$$

If $h \asymp n^{-d}$ *for some* $\frac{1}{4m+1} < d \le \frac{2m}{8m-1}$, *then we have, for any* $z_0 \in [0,1]$,

$$\sqrt{nh}(\widehat{g}_{n,\lambda}(z_0) - g_0(z_0)) \xrightarrow{d} N(0, \sigma_{z_0}^2). \tag{3.13}$$

(ii) *If we replace the boundary conditions (3.11) by the following reproducing kernel conditions: for any* $z_0 \in (0,1)$, *as* $h \to 0$

$$\frac{\partial^j}{\partial z^j}K_{z_0}(z)\bigg|_{z=0} = o(1), \qquad \frac{\partial^j}{\partial z^j}K_{z_0}(z)\bigg|_{z=1} = o(1) \qquad \text{for } j = 0,\ldots,m-1, \tag{3.14}$$

then (3.12) and (3.13) hold for any $z_0 \in (0,1)$.

We note that in (3.12) the asymptotic bias is proportional to $g_0^{(2m)}(z_0)$, and the asymptotic variance can be expressed as a weighted sum of squares of the (*infinitely many*) terms $h_\nu(z_0)$; see (3.8). These observations are consistent with those in the polynomial spline setting insofar as the bias is proportional to $g_0^{(2m)}(z_0)$, and the variance is a weighted sum of squares of (*finitely many*) terms involving the normalized B-spline basis functions evaluated at z_0; see [45]. Furthermore, (3.13) describes how to remove the asymptotic bias through undersmoothing, although the corresponding smoothing parameter yields suboptimal estimates in terms of the convergence rate.

The existing smoothing spline literature is mostly concerned with the global convergence properties of the estimates. For example, Nychka [29] and Rice and Rosenblatt [31] derived global convergence rates in terms of the (integrated) mean squared error. Instead, Theorem 3.5 and Corollaries 3.6 and 3.7 mainly focus on local asymptotics, and they conclude that the well-known global results on the rates of convergence also hold in the *local* sense.

4. Local asymptotic inference. We consider inferring $g(\cdot)$ *locally* by constructing the pointwise asymptotic CI in Section 4.1 and testing the local hypothesis in Section 4.2.

4.1. *Pointwise confidence interval.* We consider the confidence interval for some real-valued smooth function of $g_0(z_0)$ at any fixed $z_0 \in \mathbb{I}$, denoted $\rho_0 = \rho(g_0(z_0))$, for example, $\rho_0 = \exp(g_0(z_0))/(1+\exp(g_0(z_0)))$ in logistic regression. Corollary 3.6 together with the Delta method immediately implies Proposition 4.1 on the pointwise CI where the asymptotic estimation bias is assumed to be removed by undersmoothing.

PROPOSITION 4.1 (Pointwise confidence interval). *Suppose that the assumptions in Corollary 3.6 hold and that the estimation bias asymptotically vanishes, that is,* $\lim_{n\to\infty}(nh)^{1/2}(W_\lambda g_0)(z_0) = 0$. *Let* $\dot{\rho}(\cdot)$ *be the first derivative of* $\rho(\cdot)$. *If* $\dot{\rho}(g_0(z_0)) \neq 0$, *we have*

$$P\left(\rho_0 \in \left[\rho(\widehat{g}_{n,\lambda}(z_0)) \pm \Phi(\alpha/2)\frac{\dot{\rho}(g_0(z_0))\sigma_{z_0}}{\sqrt{nh}}\right]\right) \longrightarrow 1-\alpha,$$

where $\Phi(\alpha)$ *is the lower* α*th quantile of* $N(0,1)$.

From now on, we focus on the pointwise CI for $g_0(z_0)$ and compare it with the classical *Bayesian confidence intervals* proposed by Wahba [40] and Nychka [28]. For simplicity, we consider $\ell(y;a) = -(y-a)^2/(2\sigma^2)$, $Z \sim \mathrm{Unif}[0,1]$ and $\mathcal{H} = H_0^m(\mathbb{I})$ under which Proposition 4.1 implies the following asymptotic 95% CI for $g_0(z_0)$:

$$\widehat{g}_{n,\lambda}(z_0) \pm 1.96\sigma\sqrt{I_2/(n\pi h^\dagger)}, \tag{4.1}$$

where $h^\dagger = h\sigma^{1/m}$ and $I_l = \int_0^1 (1+x^{2m})^{-l}\,dx$ for $l = 1,2$; see case (I) of Example 6.1 for the derivations. When σ is unknown, we may replace it by any consistent estimate. As far as we are aware, (4.1) is the first rigorously proven pointwise CI for smoothing spline. It is well known that the Bayesian type CI only approximately achieves the 95% nominal level on average rather than pointwise. Specifically, its average coverage probability over the observed covariates is *not* exactly 95% even asymptotically. Furthermore, the Bayesian type CI ignores the important issue of coverage uniformity across the design space, and thus it may not be reliable if only evaluated at peaks or troughs, as pointed out in [28]. However, the asymptotic CI (4.1) is proved to be valid at any point, and is even shorter than the Bayesian CIs proposed in [28, 40].

As an illustration, we perform a detailed comparison of the three CIs for the special case $m = 2$. We first derive the asymptotically equivalent versions of the Bayesian CIs. Wahba [40] proposed the following heuristic CI under a Bayesian framework:

$$\widehat{g}_{n,\lambda}(z_0) \pm 1.96\sigma\sqrt{a(h^\dagger)}, \tag{4.2}$$

where $a(h^\dagger) = n^{-1}(1 + (1+(\pi n h^\dagger))^{-4} + 2\sum_{\nu=1}^{n/2-1}(1+(2\pi\nu h^\dagger))^{-4})$. Under the assumptions $h^\dagger = o(1)$ and $(nh^\dagger)^{-1} = o(1)$, Lemma 6.1 in Example 6.1

implies $2\sum_{\nu=1}^{n/2-1}(1+(2\pi\nu h^\dagger))^{-4}\sim I_1/(\pi h^\dagger)=4I_2/(3\pi h^\dagger)$, since $I_2/I_1=3/4$ when $m=2$. Hence, we obtain an asymptotically equivalent version of Wahba's Bayesian CI as

$$\widehat{g}_{n,\lambda}(z_0)\pm 1.96\sigma\sqrt{(4/3)\cdot I_2/(n\pi h^\dagger)}. \tag{4.3}$$

Nychka [28] further shortened (4.2) by proposing

$$\widehat{g}_{n,\lambda}(z_0)\pm 1.96\sqrt{\mathrm{Var}(b(z_0))+\mathrm{Var}(v(z_0))}, \tag{4.4}$$

where $b(z_0)=E\{\widehat{g}_{n,\lambda}(z_0)\}-g_0(z_0)$ and $v(z_0)=\widehat{g}_{n,\lambda}(z_0)-E\{\widehat{g}_{n,\lambda}(z_0)\}$, and showed that

$$\sigma^2 a(h^\dagger)/(\mathrm{Var}(b(z_0))+\mathrm{Var}(v(z_0)))\to 32/27 \quad \text{as } n\to\infty \text{ and } \mathrm{Var}(v(z_0))=8\,\mathrm{Var}(b(z_0)), \tag{4.5}$$

see his equation (2.3) and the Appendix. Hence, we have

$$\begin{aligned}&\mathrm{Var}(v(z_0))\sim\sigma^2\cdot(I_2/(n\pi h^\dagger)) \quad \text{and}\\ &\mathrm{Var}(b(z_0))\sim(\sigma^2/8)\cdot(I_2/(n\pi h^\dagger)).\end{aligned} \tag{4.6}$$

Therefore, Nychka's Bayesian CI (4.4) is asymptotically equivalent to

$$\widehat{g}_{n,\lambda}(z_0)\pm 1.96\sigma\sqrt{(9/8)\cdot I_2/(n\pi h^\dagger)}. \tag{4.7}$$

In view of (4.3) and (4.7), we find that the Bayesian CIs of Wahba and Nychka are asymptotically 15.4% and 6.1%, respectively, wider than (4.1). Meanwhile, by (4.6) we find that (4.1) turns out to be a corrected version of Nychka's CI (4.4) by removing the random bias term $b(z_0)$. The inclusion of $b(z_0)$ in (4.4) might be problematic in that (i) it makes the pointwise limit distribution nonnormal and thus leads to biased coverage probability; and (ii) it introduces additional variance, which unnecessarily increases the length of the interval. By removing $b(z_0)$, we can achieve both pointwise consistency and a shorter length. Similar considerations apply when $m>2$. Furthermore, the simulation results in Example 6.1 demonstrate the superior performance of our CI in both periodic and nonperiodic splines.

4.2. *Local likelihood ratio test.* In this section, we propose a likelihood ratio method for testing the value of $g_0(z_0)$ at any $z_0\in\mathbb{I}$. First, we show that the null limiting distribution is a scaled noncentral Chi-square with one degree of freedom. Second, by removing the estimation bias, we obtain a more useful central Chi-square limit distribution. We also note that as the smoothness order m approaches infinity, the scaling constant eventually converges to one. Therefore, we have unveiled an interesting Wilks phenomenon arising from the proposed nonparametric local testing. A relevant study was conducted by Banerjee [1], who considered a likelihood ratio test for *monotone* functions, but his estimation method and null limiting distribution are fundamentally different from ours.

For some prespecified point (z_0, w_0), we consider the following hypothesis:

$$H_0 : g(z_0) = w_0 \quad \text{versus} \quad H_1 : g(z_0) \neq w_0. \tag{4.8}$$

The "constrained" penalized log-likelihood is defined as $L_{n,\lambda}(g) = n^{-1} \sum_{i=1}^n \ell(Y_i; w_0 + g(Z_i)) - (\lambda/2) J(g,g)$, where $g \in \mathcal{H}_0 = \{g \in \mathcal{H} : g(z_0) = 0\}$. We consider the likelihood ratio test (LRT) statistic defined as

$$\mathrm{LRT}_{n,\lambda} = \ell_{n,\lambda}(w_0 + \widehat{g}^0_{n,\lambda}) - \ell_{n,\lambda}(\widehat{g}_{n,\lambda}), \tag{4.9}$$

where $\widehat{g}^0_{n,\lambda}$ is the MLE of g under the local restriction, that is,

$$\widehat{g}^0_{n,\lambda} = \arg\max_{g \in \mathcal{H}_0} L_{n,\lambda}(g).$$

Endowed with the norm $\|\cdot\|$, $\mathcal{H}_0$ is a closed subset in $\mathcal{H}$, and thus a Hilbert space. Proposition 4.2 below says that $\mathcal{H}_0$ also inherits the reproducing kernel and penalty operator from $\mathcal{H}$. The proof is trivial and thus omitted.

PROPOSITION 4.2. (a) *Recall that $K(z_1, z_2)$ is the reproducing kernel for $\mathcal{H}$ under $\langle\cdot,\cdot\rangle$. The bivariate function $K^*(z_1, z_2) = K(z_1, z_2) - (K(z_1, z_0)K(z_0, z_2))/K(z_0, z_0)$ is a reproducing kernel for $(\mathcal{H}_0, \langle\cdot,\cdot\rangle)$. That is, for any $z' \in \mathbb{I}$ and $g \in \mathcal{H}_0$, we have $K^*_{z'} \equiv K^*(z', \cdot) \in \mathcal{H}_0$ and $\langle K^*_{z'}, g\rangle = g(z')$.*
(b) *The operator W^*_λ defined by $W^*_\lambda g = W_\lambda g - [(W_\lambda g)(z_0)/K(z_0, z_0)] \cdot K_{z_0}$ is bounded linear from $\mathcal{H}_0$ to $\mathcal{H}_0$ and satisfies $\langle W^*_\lambda g, \widetilde{g}\rangle = \lambda J(g, \widetilde{g})$, for any $g, \widetilde{g} \in \mathcal{H}_0$.*

On the basis of Proposition 4.2, we derive the *restricted* FBR for $\widehat{g}^0_{n,\lambda}$, which will be used to obtain the null limiting distribution. By straightforward calculation we can find the Fréchet derivatives of $L_{n,\lambda}$ (under $\mathcal{H}_0$). Let $\Delta g, \Delta g_j \in \mathcal{H}_0$ for $j = 1, 2, 3$. The first-order Fréchet derivative of $L_{n,\lambda}$ is

$$\begin{aligned} DL_{n,\lambda}(g)\Delta g &= n^{-1} \sum_{i=1}^n \dot{\ell}_a(Y_i; w_0 + g(Z_i)) \langle K^*_{Z_i}, \Delta g\rangle - \langle W^*_\lambda g, \Delta g\rangle \\ &\equiv \langle S^0_n(g), \Delta g\rangle - \langle W^*_\lambda g, \Delta g\rangle \\ &\equiv \langle S^0_{n,\lambda}(g), \Delta g\rangle. \end{aligned}$$

Clearly, we have $S^0_{n,\lambda}(\widehat{g}^0_{n,\lambda}) = 0$. Define $S^0(g)\Delta g = E\{\langle S^0_n(g), \Delta g\rangle\}$ and $S^0_\lambda(g)\Delta g = S^0(g)\Delta g - \langle W^*_\lambda g, \Delta g\rangle$. The second-order derivatives are $DS^0_{n,\lambda}(g) \times \Delta g_1 \Delta g_2 = D^2 L_{n,\lambda}(g)\Delta g_1 \Delta g_2$ and $DS^0_\lambda(g)\Delta g_1 \Delta g_2 = DS^0(g)\Delta g_1 \Delta g_2 - \langle W^*_\lambda \Delta g_1, g_2\rangle$, where

$$DS^0(g)\Delta g_1 \Delta g_2 = E\{\ddot{\ell}_a(Y; w_0 + g(Z))\langle K^*_Z, \Delta g_1\rangle \langle K^*_Z, \Delta g_2\rangle\}.$$

The third-order Fréchet derivative of $L_{n,\lambda}$ is

$$D^3 L_{n,\lambda}(g)\Delta g_1 \Delta g_2 \Delta g_3 = n^{-1}\sum_{i=1}^{n} \ell'''_a(Y_i; w_0 + g(Z_i))\langle K^*_{Z_i}, \Delta g_1\rangle \langle K^*_{Z_i}, \Delta g_2\rangle \langle K^*_{Z_i}, \Delta g_3\rangle.$$

Similarly to Theorem 3.4, we need an additional assumption on the convergence rate of $\widehat{g}^0_{n,\lambda}$:

ASSUMPTION A.4. Under H_0, $\|\widehat{g}^0_{n,\lambda} - g^0_0\| = O_P((nh)^{-1/2} + h^m)$, where $g^0_0(\cdot) = (g_0(\cdot) - w_0) \in \mathcal{H}_0$.

Assumption A.4 is easy to verify by assuming (2.3), (2.4) and $\|\widehat{g}^0_{n,\lambda} - g^0_0\|_{\mathcal{H}} = o_P(1)$. The proof is similar to that of Proposition 3.3 by replacing $\mathcal{H}$ with the subspace $\mathcal{H}_0$.

THEOREM 4.3 (Restricted FBR). *Suppose that Assumptions A.1, A.2, A.4 and H_0 are satisfied. If $h = o(1)$ and $nh^2 \to \infty$, then $\|\widehat{g}^0_{n,\lambda} - g^0_0 - S^0_{n,\lambda}(g^0_0)\| = O_P(a_n \log n)$.*

Our main result on the local LRT is presented below. Define $r_n = (nh)^{-1/2} + h^m$.

THEOREM 4.4 (Local likelihood ratio test). *Suppose that Assumptions A.1–A.4 are satisfied. Also assume $h = o(1)$, $nh^2 \to \infty$, $a_n = o(\min\{r_n, n^{-1}r_n^{-1}(\log n)^{-1}, n^{-1/2}(\log n)^{-1}\})$ and $r_n^2 h^{-1/2} = o(a_n)$. Furthermore, for any $z_0 \in [0,1]$, $n^{1/2}(W_\lambda g_0)(z_0)/\sqrt{K(z_0,z_0)} \to -c_{z_0}$,*

$$\lim_{h\to 0} hV(K_{z_0}, K_{z_0}) \to \sigma^2_{z_0} > 0 \quad \text{and}$$
$$\lim_{\lambda\to 0} E\{I(Z)|K_{z_0}(Z)|^2\}/K(z_0,z_0) \equiv c_0 \in (0,1]. \tag{4.10}$$

Under H_0, we show: (i) $\|\widehat{g}_{n,\lambda} - \widehat{g}^0_{n,\lambda} - w_0\| = O_P(n^{-1/2})$; (ii) $-2n \cdot \mathrm{LRT}_{n,\lambda} = n\|\widehat{g}_{n,\lambda} - \widehat{g}^0_{n,\lambda} - w_0\|^2 + o_P(1)$; *and*

$$\text{(iii)} \quad -2n \cdot \mathrm{LRT}_{n,\lambda} \xrightarrow{d} c_0 \chi^2_1(c^2_{z_0}/c_0) \tag{4.11}$$

with noncentrality parameter $c^2_{z_0}/c_0$.

Note that the parametric convergence rate stated in (i) of Theorem 4.4 is reasonable since the restriction is local. By Proposition 2.1, it can be explicitly shown that

$$c_0 = \lim_{\lambda\to 0} \frac{Q_2(\lambda, z_0)}{Q_1(\lambda, z_0)}, \tag{4.12}$$
$$\text{where } Q_l(\lambda, z) \equiv \sum_{\nu\in\mathbb{N}} \frac{|h_\nu(z)|^2}{(1+\lambda\gamma_\nu)^l} \qquad \text{for } l = 1, 2.$$

The reproducing kernel K, if it exists, is uniquely determined by the corresponding RKHS; see [8]. Therefore, c_0 defined in (4.10) depends only on the parameter space. Hence, different choices of (γ_ν, h_ν) in (4.12) will give exactly the same value of c_0, although certain choices can facilitate the calculations. For example, when $\mathcal{H} = H_0^m(\mathbb{I})$, we can explicitly calculate the value of c_0 as 0.75 (0.83) when $m = 2$ (3) by choosing the trigonometric polynomial basis (6.2). Interestingly, when $\mathcal{H} = H^2(\mathbb{I})$, we can obtain the same value of c_0 even without specifying its (rather different) eigensystem; see Remark 5.2 for more details. In contrast, the value of c_{z_0} in (4.11) depends on the asymptotic bias specified in (3.9), whose estimation is notoriously difficult. Fortunately, under various undersmoothing conditions, we can show $c_{z_0} = 0$ and thus establish a central Chi-square limit distribution. For example, we can assume higher order smoothness on the true function, as in Corollary 4.5 below.

Corollary 4.5. *Suppose that Assumptions A.1–A.4 are satisfied and H_0 holds. Let $m > 1 + \sqrt{3}/2 \approx 1.866$. Also assume that the Fourier coefficients $\{V(g_0, h_\nu)\}_{\nu\in\mathbb{N}}$ of g_0 satisfy $\sum_\nu |V(g_0, h_\nu)|^2\gamma_\nu^d$ for some $d > 1 + 1/(2m)$, which holds if $g_0 \in H^{md}(\mathbb{I})$. Furthermore, if (4.10) is satisfied for any $z_0 \in [0, 1]$, then (4.11) holds with the limiting distribution $c_0\chi_1^2$ under $\lambda = \lambda^*$.*

Corollary 4.5 demonstrates a nonparametric type of the Wilks phenomenon, which approaches the parametric type as $m \to \infty$ since $\lim_{m\to\infty} c_0 = 1$. This result provides a theoretical insight for nonparametric local hypothesis testing; see its *global* counterpart in Section 5.2.

5. Global asymptotic inference.

Depicting the global behavior of a smooth function is crucial in practice. In Sections 5.1 and 5.2, we develop the *global* counterparts of Section 4 by constructing simultaneous confidence bands and testing a set of global hypotheses.

5.1. *Simultaneous confidence band.*

In this section, we construct the SCBs for g using the approach of [2]. We demonstrate the theoretical validity of the proposed SCB based on the FBR and strong approximation techniques. The approach of [2] was originally developed in the context of density estimation, and it was then extended to M-estimation by [17] and local polynomial estimation by [5, 11, 44]. The volume-of-tube method [38] is another approach, but it requires the error distribution to be symmetric; see [22, 45]. Sun et al. [37] relaxed the restrictive error assumption in generalized linear models, but they had to translate the nonparametric estimation into parametric estimation. Our SCBs work for a general class of nonparametric models including normal regression and logistic regression. Additionally, the minimum width of the proposed SCB is shown to achieve the lower bound established by [12]; see Remark 5.3. An interesting by-product is that, under the equivalent kernel conditions given in this section, the local asymptotic inference procedures developed from cubic splines and periodic splines are essentially the same despite the intrinsic difference in their eigensystems; see Remark 5.2 for technical details.

The key conditions assumed in this section are the equivalent kernel conditions (5.1)–(5.3). Specifically, we assume that there exists a real-valued function $\omega(\cdot)$ defined on $\mathbb{R}$ satisfying, for any fixed $0 < \varphi < 1$, $h^\varphi \le z \le 1 - h^\varphi$ and $t \in \mathbb{I}$,

$$\left|\frac{d^j}{dt^j}(h^{-1}\omega((z-t)/h) - K(z,t))\right| \le C_K h^{-(j+1)} \exp(-C_2 h^{-1+\varphi}) \qquad \text{for } j = 0, 1, \tag{5.1}$$

where C_2, C_K are positive constants. Condition (5.1) implies that ω is an equivalent kernel of the reproducing kernel function K with a certain degree of approximation accuracy. We also require two regularity conditions on ω:

$$|\omega(u)| \le C_\omega \exp(-|u|/C_3), \qquad |\omega'(u)| \le C_\omega \exp(-|u|/C_3) \qquad \text{for any } u \in \mathbb{R}, \tag{5.2}$$

and there exists a constant $0 < \rho \le 2$ s.t.

$$\int_{-\infty}^{\infty} \omega(t)\omega(t+z)\,dt = \sigma_\omega^2 - C_\rho |z|^\rho + o(|z|^\rho) \qquad \text{as } |z| \to \infty, \tag{5.3}$$

where C_3, C_ω, C_ρ are positive constants and $\sigma_\omega^2 = \int_{\mathbb{R}} \omega(t)^2\,dt$. An example of ω satisfying (5.1)–(5.3) will be given in Proposition 5.2. The following exponential envelope condition is also needed:

$$\sup_{z,t\in\mathbb{I}} \left|\frac{\partial}{\partial z} K(z,t)\right| = O(h^{-2}). \tag{5.4}$$

THEOREM 5.1 (Simultaneous confidence band). *Suppose Assumptions A.1–A.3 are satisfied, and Z is uniform on $\mathbb{I}$. Let* $m > (3+\sqrt{5})/4 \approx 1.3091$

and $h = n^{-\delta}$ *for any* $\delta \in (0, 2m/(8m-1))$. *Furthermore,* $E\{\exp(|\epsilon|/C_0) \mid Z\} \le C_1$, *a.s., and (5.1)–(5.4) hold. The conditional density of* ϵ *given* $Z = z$, *denoted* $\pi(\epsilon \mid z)$, *satisfies the following: for some positive constants* ρ_1 *and* ρ_2,

$$\left|\frac{d}{dz}\log\pi(\epsilon \mid z)\right| \le \rho_1(1 + |\epsilon|^{\rho_2}) \qquad \text{for any } \epsilon \in \mathbb{R} \text{ and } z \in \mathbb{I}. \tag{5.5}$$

Then, for any $0 < \varphi < 1$ *and* $u \in \mathbb{R}$,

$$\begin{aligned} P\Big((2\delta\log n)^{1/2}\Big\{&\sup_{h^\varphi \le z \le 1-h^\varphi}(nh)^{1/2}\sigma_\omega^{-1}I(z)^{-1/2} \\ &\times |\widehat{g}_{n,\lambda}(z) - g_0(z) + (W_\lambda g_0)(z)| - d_n\Big\} \le u\Big) \\ &\longrightarrow \exp(-2\exp(-u)), \end{aligned} \tag{5.6}$$

where d_n *relies only on* h, ρ, φ *and* C_ρ.

The FBR developed in Section 3.1 and the strong approximation techniques developed by [2] are crucial to the proof of Theorem 5.1. The uniform distribution on Z is assumed only for simplicity, and this condition can be relaxed by requiring that the density is bounded away from zero and infinity. Condition (5.5) holds in various situations such as the conditional normal model $\epsilon \mid Z = z \sim N(0, \sigma^2(z))$, where $\sigma^2(z)$ satisfies $\inf_z \sigma^2(z) > 0$, and $\sigma(z)$ and $\sigma'(z)$ both have finite upper bounds. The existence of the bias term $W_\lambda g_0(z)$ in (5.6) may result in poor finite-sample performance. We address this issue by assuming undersmoothing, which is advocated by [15, 16, 27]; they showed that undersmoothing is more efficient than explicit bias correction when the goal is to minimize the coverage error. Specifically, the bias term will asymptotically vanish if we assume that

$$\lim_{n\to\infty}\Big\{\sup_{h^\varphi \le z \le 1-h^\varphi}\sqrt{nh\log n}|W_\lambda g_0(z)|\Big\} = 0. \tag{5.7}$$

Condition (5.7) is slightly stronger than the undersmoothing condition $\sqrt{nh}(W_\lambda g_0)(z_0) = o(1)$ assumed in Proposition 4.1. By the generalized Fourier expansion of $W_\lambda g_0$ and the uniform boundedness of h_ν (see Assumption A.2), we can show that (5.7) holds if we properly increase the amount of smoothness on g_0 or choose a suboptimal λ, as in Corollaries 3.7 and 4.5.

Proposition 5.2 below demonstrates the validity of Conditions (5.1)–(5.3) in L_2 regression. The proof relies on an explicit construction of the equivalent kernel function obtained by [26]. We consider only $m = 2$ for simplicity.

PROPOSITION 5.2 (L_2 regression). *Let* $\ell(y; a) = -(y-a)^2/(2\sigma^2)$, $Z \sim \mathrm{Unif}[0,1]$ *and* $\mathcal{H} = H^2(\mathbb{I})$, *that is,* $m = 2$. *Then, (5.1)–(5.3) hold with* $\omega(t) = \sigma^{2-1/m}\omega_0(\sigma^{-1/m}t)$ *for* $t \in \mathbb{R}$, *where* $\omega_0(t) = \frac{1}{2\sqrt{2}}\exp(-|t|/\sqrt{2})(\cos(t/\sqrt{2}) + \sin(|t|/\sqrt{2}))$. *In particular, (5.3) holds for arbitrary* $\rho \in (0, 2]$ *and* $C_\rho = 0$.

REMARK 5.1. In the setting of Proposition 5.2, we are able to explicitly find the constants σ_ω^2 and d_n in Theorem 5.1. Specifically, by direct calculation it can be found that $\sigma_\omega^2 = 0.265165\sigma^{7/2}$ since $\sigma_{\omega_0}^2 = \int_{-\infty}^{\infty} |\omega_0(t)|^2\,dt = 0.265165$ when $m = 2$. Choose $C_\rho = 0$ for arbitrary $\rho \in (0, 2]$. By the formula of $B(t)$ given in Theorem A.1 of [2], we know that

$$(5.8)\quad d_n = (2\log(h^{-1} - 2h^{\varphi-1}))^{1/2} + \frac{(1/\rho - 1/2)\log\log(h^{-1} - 2h^{\varphi-1})}{(2\log(h^{-1} - 2h^{\varphi-1}))^{1/2}}.$$

When $\rho = 2$, the above d_n is simplified as $(2\log(h^{-1} - 2h^{\varphi-1}))^{1/2}$. In general, we observe that $d_n \sim (-2\log h)^{1/2} \asymp \sqrt{\log n}$ for sufficiently large n since $h = n^{-\delta}$. Given that the estimation bias is removed, for example, under (5.7), we obtain the following $100 \times (1-\alpha)\%$ SCB:

$$(5.9)\quad \begin{aligned} \{[\widehat{g}_{n,\lambda}(z) \pm 0.5149418(nh)^{-1/2}\widehat{\sigma}^{3/4}(c_\alpha^*/\sqrt{-2\log h} + d_n)]: \\ h^\varphi \le z \le 1 - h^\varphi\}, \end{aligned}$$

where $d_n = (-2\log h)^{1/2}$, $c_\alpha^* = -\log(-\log(1-\alpha)/2)$ and $\widehat{\sigma}$ is a consistent estimate of σ. Therefore, to obtain uniform coverage, we have to increase the bandwidth up to an order of $\sqrt{\log n}$ over the length of the pointwise CI given in (4.1). Note that we have excluded the boundary points in (5.9).

REMARK 5.2. An interesting by-product we discover in the setting of Proposition 5.2 is that the pointwise asymptotic CIs for $g_0(z_0)$ based on cubic splines and periodic splines share the same length at any $z_0 \in (0, 1)$. This result is surprising since the two splines have intrinsically different structures. Under (5.1), it can be shown that

$$\begin{aligned} \sigma_{z_0}^2 &\sim \sigma^{-2} h \int_0^1 |K(z_0, z)|^2\,dz \\ &\sim \sigma^{-2} h^{-1} \int_0^1 \left|\omega\left(\frac{z - z_0}{h}\right)\right|^2 dz \\ &= \sigma^{-2} \int_{-z_0/h}^{(1-z_0)/h} |\omega(s)|^2\,ds \sim \sigma^{-2} \int_{\mathbb{R}} |\omega(s)|^2\,ds = \sigma^{3/2}\sigma_{\omega_0}^2, \end{aligned}$$

given the choice of ω in Proposition 5.2. Thus, Corollary 3.6 implies the following 95% CI:

$$(5.10)\quad \widehat{g}_{n,\lambda}(z_0) \pm 1.96(nh)^{-1/2}\sigma^{3/4}\sigma_{\omega_0} = \widehat{g}_{n,\lambda}(z_0) \pm 1.96(nh^\dagger)^{-1/2}\sigma\sigma_{\omega_0}.$$

Since $\sigma_{\omega_0}^2 = I_2/\pi$, the lengths of the CIs (4.1) (periodic spline) and (5.10) (cubic spline) coincide with each other. The above calculation of $\sigma_{z_0}^2$ relies on L_2 regression. For general models such as logistic regression, one can instead use a weighted version of (2.2) with the weights $B(Z_i)^{-1}$ to obtain the exact formula. Another application of Proposition 5.2 is to find the value of c_0 in Theorem 4.3 for the construction of the local LRT test when $\mathcal{H} = H^2(\mathbb{I})$. According to the definition of c_0, that is, (4.10), we have $c_0 \sim \sigma_{z_0}^2/(hK(z_0, z_0))$. Under (5.1), we have $K(z_0, z_0) \sim h^{-1}\omega(0) = h^{-1}\sigma^{3/2}\omega_0(0) = 0.3535534h^{-1}\sigma^{3/2}$. Since $\sigma_{z_0}^2 \sim \sigma^{3/2}\sigma_{\omega_0}^2$ and $\sigma_{\omega_0}^2 = I_2/\pi$, we have $c_0 = 0.75$. This value coincides with that found in periodic splines, that is, $\mathcal{H} = H_0^2(\mathbb{I})$. These intriguing phenomena have never been observed in the literature and may be useful for simplifying the construction of CIs and local LRT.

REMARK 5.3. Genovese and Wasserman [12] showed that when g_0 belongs to an mth-order Sobolev ball, the lower bound for the average width of an SCB is proportional to $b_n n^{-m/(2m+1)}$, where b_n depends only on $\log n$. We next show that the (minimum) bandwidth of the proposed SCB can achieve this lower bound with $b_n = (\log n)^{(m+1)/(2m+1)}$. Based on Theorem 5.1, the width of the SCB is of order $d_n(nh)^{-1/2}$, where $d_n \asymp \sqrt{\log n}$; see Remark 5.1. Meanwhile, Condition (5.7) is crucial for our band to maintain the desired coverage probability. Suppose that the Fourier coefficients of g_0 satisfy $\sum_\nu |V(g_0, h_\nu)|\gamma_\nu^{1/2} < \infty$. It can be verified that (5.7) holds when $nh^{2m+1}\log n = O(1)$, which sets an upper bound for h, that is, $O(n \log n)^{-1/(2m+1)}$. When h is chosen as the above upper bound and $d_n \asymp \sqrt{\log n}$, our SCB achieves the minimum order of bandwidth $n^{-m/(2m+1)}(\log n)^{(m+1)/(2m+1)}$, which turns out to be optimal according to [12].

In practice, the construction of our SCB requires a delicate choice of (h, φ). Otherwise, over-coverage or undercoverage of the true function may occur near the boundary points. There is no practical or theoretical guideline on how to find the optimal (h, φ), although, as noted by [2], one can choose a proper h to make the band as thin as possible. Hence, in the next section, we propose a more straightforward likelihood-ratio-based approach for testing the global behavior, which requires only tuning h.

5.2. *Global likelihood ratio test.* There is a vast literature dealing with nonparametric hypothesis testing, among which the GLRT proposed by Fan et al. [9] stands out. Because of the technical complexity, they focused on the local polynomial fitting; see [10] for a sieve version. Based on smoothing spline estimation, we propose the PLRT, which is applicable to both simple and composite hypotheses. The null limiting distribution is identified to be nearly Chi-square with diverging degrees of freedom. The degrees of freedom depend only on the functional parameter space, while the null limiting distribution of the GLRT depends on the choice of kernel functions; see Table 2 in [9]. Furthermore, the PLRT is shown to achieve the minimax rate of testing in the sense of [19]. As demonstrated in our simulations, the PLRT performs better than the GLRT in terms of power, especially in small-sample situations. Other smoothing-spline-based testing such as LMP, GCV and GML (see [4, 6, 20, 23, 30, 41]) use ad-hoc discrepancy measures leading to complicated null distributions involving nuisance parameters; see a thorough review in [23].

Consider the following "global" hypothesis:

$$H_0^{\text{global}}: g = g_0 \quad \text{versus} \quad H_1^{\text{global}}: g \in \mathcal{H} - \{g_0\}, \tag{5.11}$$

where $g_0 \in \mathcal{H}$ can be either known or unknown. The PLRT statistic is defined to be

$$\text{PLRT}_{n,\lambda} = \ell_{n,\lambda}(g_0) - \ell_{n,\lambda}(\widehat{g}_{n,\lambda}). \tag{5.12}$$

Theorem 5.3 below derives the null limiting distribution of $\text{PLRT}_{n,\lambda}$. We remark that the null limiting distribution remains the same even when the hypothesized value g_0 is unknown (whether its dimension is finite or infinite). This nice property can be used to test the composite hypothesis; see Remark 5.4.

THEOREM 5.3 (Penalized likelihood ratio test). *Let Assumptions A.1–A.3 be satisfied. Also assume* $nh^{2m+1} = O(1)$, $nh^2 \to \infty$, $a_n = o(\min\{r_n, n^{-1}r_n^{-1}h^{-1/2}(\log n)^{-1}, n^{-1/2}(\log n)^{-1}\})$ *and* $r_n^2 h^{-1/2} = o(a_n)$. *Furthermore, under* H_0^{global}, $E\{\epsilon^4 \mid Z\} \le C$, *a.s., for some constant* $C > 0$, *where* $\epsilon = \dot{\ell}_a(Y; g_0(Z))$ *represents the "model error." Under* H_0^{global}, *we have*

$$(2u_n)^{-1/2}(-2nr_K \cdot \text{PLRT}_{n,\lambda} - nr_K\|W_\lambda g_0\|^2 - u_n) \xrightarrow{d} N(0,1), \tag{5.13}$$

where $u_n = h^{-1}\sigma_K^4/\rho_K^2$, $r_K = \sigma_K^2/\rho_K^2$,

$$\begin{aligned} \sigma_K^2 &= hE\{\epsilon^2 K(Z,Z)\} = \sum_\nu \frac{h}{(1+\lambda\gamma_\nu)}, \\ \rho_K^2 &= hE\{\epsilon_1^2\epsilon_2^2 K(Z_1,Z_2)^2\} = \sum_\nu \frac{h}{(1+\lambda\gamma_\nu)^2} \end{aligned} \tag{5.14}$$

and (ϵ_i, Z_i), $i = 1, 2$ *are i.i.d. copies of* (ϵ, Z).

A direct examination reveals that $h \asymp n^{-d}$ with $\frac{1}{2m+1} \le d < \frac{2m}{8m-1}$ satisfies the rate conditions required by Theorem 5.3 when $m > (3+\sqrt{5})/4 \approx 1.309$. By the proof of Theorem 5.3, it can be shown that $n\|W_\lambda g_0\|^2 = o(h^{-1}) = o(u_n)$. Therefore, $-2nr_K \cdot \text{PLRT}_{n,\lambda}$ is asymptotically $N(u_n, 2u_n)$. As n approaches infinity, $N(u_n, 2u_n)$ is nearly $\chi^2_{u_n}$. Hence, $-2nr_K \cdot \text{PLRT}_{n,\lambda}$ is approximately distributed as $\chi^2_{u_n}$, denoted

$$-2nr_K \cdot \text{PLRT}_{n,\lambda} \overset{a}{\sim} \chi^2_{u_n}. \tag{5.15}$$

That is, the Wilks phenomenon holds for the PLRT. The specifications of (5.15), that is, σ_K^2 and ρ_K^2, are determined only by the parameter space and model setup. We also note that undersmoothing is not required for our global test.

We next discuss the calculation of (r_K, u_n). In the setting of Proposition 5.2, it can be shown by the equivalent kernel conditions that $\sigma_K^2 = h\sigma^{-2}\int_0^1 K(z,z)\,dz \sim h\sigma^{-2}(h^{-1}\omega(0)) = \sigma^{-1/2}\omega_0(0) = 0.3535534\sigma^{-1/2}$ and $\rho_K^2 \sim \sigma^{-1/2}\sigma_{\omega_0}^2 = 0.265165\sigma^{-1/2}$. So $r_K = 1.3333$ and $u_n = 0.4714h^{-1}\sigma^{-1/2}$. If we replace $H^2(\mathbb{I})$ by $H_0^2(\mathbb{I})$, direct calculation in case (I) of Example 6.1 reveals that (r_K, u_n) have exactly the same values. When $\mathcal{H} = H_0^m(\mathbb{I})$, we have $2r_K \to 2$ as m tends to infinity. This limit is consistent with the scaling constant two in the parametric likelihood ratio theory. In L_2 regression, the possibly unknown parameter σ in u_n can be profiled out without changing the null limiting distribution. In practice, by the wild bootstrap we can directly simulate the null limiting distribution by fixing the nuisance parameters at some reasonable values or estimates without finding the values of (r_K, u_n). This is a major advantage of the Wilks type of results.

REMARK 5.4. We discuss composite hypothesis testing via the PLRT. Specifically, we test whether g belongs to some finite-dimensional class of functions, which is much larger than the null space $\mathcal{N}_m$ considered in the literature. For instance, for any integer $q \geq 0$, consider the null hypothesis

$$H_0^{\text{global}}: g \in \mathcal{L}_q(\mathbb{I}), \tag{5.16}$$

where $\mathcal{L}_q(\mathbb{I}) \equiv \{g(z) = \sum_{l=0}^q a_l z^l : a = (a_0, a_1, \ldots, a_q)^T \in \mathbb{R}^{q+1}\}$ is the class of the qth-order polynomials. Let $\widehat{a}_* = \arg\max_{a\in\mathbb{R}^{q+1}}\{(1/n)\sum_{i=1}^n \ell(Y_i; \sum_{l=0}^q a_l Z_i^l) - (\lambda/2)a^T Da\}$, where

$$D = \int_0^1 (0,0,2,6z,\ldots,q(q-1)z^{q-2})^T(0,0,2,6z,\ldots,q(q-1)z^{q-2})\,dz$$

is a $(q+1)\times(q+1)$ matrix. Hence, under H_0^{global}, the penalized MLE is $\widehat{g}_*(z) = \sum_{l=0}^q \widehat{a}_{*l} z^l$. Let g_{0q} be an unknown "true parameter" in $\mathcal{L}_q(\mathbb{I})$ corresponding to a vector of polynomial coefficients $a^0 = (a_0^0, a_1^0, \ldots, a_q^0)^T$. To test (5.16), we decompose the PLRT statistic as $\text{PLRT}_{n,\lambda}^{\text{com}} = L_{n1} - L_{n2}$, where $L_{n1} = \ell_{n,\lambda}(g_{0q}) - \ell_{n,\lambda}(\widehat{g}_{n,\lambda})$ and $L_{n2} = \ell_{n,\lambda}(g_{0q}) - \ell_{n,\lambda}(\widehat{g}_*)$. When we formulate

$$H_0': a = a^0 \quad \text{versus} \quad H_1': a \neq a^0,$$

L_{n2} appears to be the PLRT statistic in the parametric setup. It can be shown that $L_{n2} = O_P(n^{-1})$ whether $q < m$ (by applying the parametric theory in [34]) or $q \geq m$ (by slightly modifying the proof of Theorem 4.4). On the other hand, L_{n1} is exactly the PLRT for testing

$$H_0' : g = g_{0q} \quad \text{versus} \quad H_1^{\text{global}} : g \neq g_{0q}.$$

By Theorem 5.3, L_{n1} follows the limit distribution specified in (5.15). In summary, under (5.16), $\text{PLRT}_{n,\lambda}^{\text{com}}$ has the same limit distribution since $L_{n2} = O_P(n^{-1})$ is negligible.

To conclude this section, we show that the PLRT achieves the optimal minimax rate of testing specified in Ingster [19] based on a *uniform* version of the FBR. For convenience, we consider only $\ell(Y; a) = -(Y - a)^2/2$. Extensions to a more general setup can be found in the supplementary document [33] under stronger assumptions, for example, a more restrictive alternative set.

Write the local alternative as $H_{1n} : g = g_{n0}$, where $g_{n0} = g_0 + g_n$, $g_0 \in \mathcal{H}$ and g_n belongs to the alternative value set $\mathcal{G}_a \equiv \{g \in \mathcal{H} \mid \text{Var}(g(Z)^2) \leq \zeta E^2\{g(Z)^2\}, J(g, g) \leq \zeta\}$ for some constant $\zeta > 0$.

THEOREM 5.4. *Let $m > (3 + \sqrt{5})/4 \approx 1.309$ and $h \asymp n^{-d}$ for $\frac{1}{2m+1} \leq d < \frac{2m}{8m-1}$. Suppose that Assumption A.2 is satisfied, and uniformly over $g_n \in \mathcal{G}_a$, $\|\widehat{g}_{n,\lambda} - g_{n0}\| = O_P(r_n)$ holds under $H_{1n} : g = g_{n0}$. Then for any $\delta \in (0, 1)$, there exist positive constants C and N such that*

$$\inf_{n \geq N} \inf_{\substack{g_n \in \mathcal{G}_a \\ \|g_n\| \geq C\eta_n}} P(\text{reject } H_0^{\text{global}} \mid H_{1n} \text{ is true}) \geq 1 - \delta, \tag{5.17}$$

where $\eta_n \geq \sqrt{h^{2m} + (nh^{1/2})^{-1}}$. The minimal lower bound of η_n, that is, $n^{-2m/(4m+1)}$, is achieved when $h = h^{**} \equiv n^{-2/(4m+1)}$.

The condition "uniformly over $g_n \in \mathcal{G}_a$, $\|\widehat{g}_{n,\lambda} - g_{n0}\| = O_P(r_n)$ holds under $H_{1n} : g = g_{n0}$" means that for any $\widetilde{\delta} > 0$, there exist constants $\widetilde{C}$ and $\widetilde{N}$, both unrelated to $g_n \in \mathcal{G}_a$, such that $\inf_{n \geq \widetilde{N}} \inf_{g_n \in \mathcal{G}_a} P_{g_{n0}}(\|\widehat{g}_{n,\lambda} - g_{n0}\| \leq \widetilde{C} r_n) \geq 1 - \widetilde{\delta}$.

Theorem 5.4 proves that, when $h = h^{**}$, the PLRT can detect any local alternatives with separation rates no faster than $n^{-2m/(4m+1)}$, which turns out to be the minimax rate of testing in the sense of Ingster [19]; see Remark 5.5 below.

REMARK 5.5. The minimax rate of testing established in Ingster [19] is under the usual $\|\cdot\|_{L_2}$-norm (w.r.t. the Lebesgue measure). However, the separation rate derived under the $\|\cdot\|$-norm is still optimal because of the trivial domination of $\|\cdot\|$ over $\|\cdot\|_{L_2}$ (under the conditions of Theorem 5.4). Next, we heuristically explain why the minimax rates of testing associated with $\|\cdot\|$, denoted b_n', and with $\|\cdot\|_{L_2}$, denoted b_n, are the same. By definition, whenever $\|g_n\| \geq b_n'$ or $\|g_n\|_{L_2} \geq b_n$, H_0^{global} can be rejected with a large probability, or equivalently, the local alternatives can be detected. b_n' and b_n are the minimum rates that satisfy this property. Ingster [19] has shown that $b_n \asymp n^{-2m/(4m+1)}$. Since $\|g_n\|_{L_2} \geq b_n'$ implies $\|g_n\| \geq b_n'$, H_0^{global} is rejected. This means b_n' is an upper bound for detecting the local alternatives in terms of $\|\cdot\|_{L_2}$ and so $b_n \leq b_n'$. On the other hand, suppose $h = h^{**} \asymp n^{-2/(4m+1)}$ and $\|g_n\| \geq Cn^{-2m/(4m+1)} \asymp b_n$ for some large $C > \zeta^{1/2}$. Since $\lambda J(g_n, g_n) \leq \zeta\lambda \asymp \zeta n^{-4m/(4m+1)}$, it follows that $\|g_n\|_{L_2} \geq (C^2 - \zeta)^{1/2} n^{-2m/(4m+1)} \asymp b_n$. This means b_n is a upper bound for detecting the local alternatives in terms of $\|\cdot\|$ and so $b_n' \leq b_n$. Therefore, b_n' and b_n are of the same order.

6. Examples. In this section, we provide three concrete examples together with simulations.

EXAMPLE 6.1 (L_2 regression). We consider the regression model with an additive error

$$Y = g_0(Z) + \epsilon, \tag{6.1}$$

where $\epsilon \sim N(0, \sigma^2)$ with an unknown variance σ^2. Hence, $I(Z) = \sigma^{-2}$ and $V(g, \widetilde{g}) = \sigma^{-2} E\{g(Z)\widetilde{g}(Z)\}$. For simplicity, Z was generated uniformly over $\mathbb{I}$. The function *ssr*( ) in the R package *assist* was used to select the smoothing parameter λ based on CV or GCV; see [21]. We first consider $\mathcal{H} = H_0^m(\mathbb{I})$ in case (I) and then $\mathcal{H} = H^m(\mathbb{I})$ in case (II).

Case (I). $\mathcal{H} = H_0^m(\mathbb{I})$: In this case, we choose the basis functions as

$$h_\mu(z) = \begin{cases} \sigma, & \mu = 0, \\ \sqrt{2}\sigma\cos(2\pi k z), & \mu = 2k, k = 1, 2, \ldots, \\ \sqrt{2}\sigma\sin(2\pi k z), & \mu = 2k-1, k = 1, 2, \ldots, \end{cases} \tag{6.2}$$

with the corresponding eigenvalues $\gamma_{2k-1} = \gamma_{2k} = \sigma^2(2\pi k)^{2m}$ for $k \geq 1$ and $\gamma_0 = 0$. Assumption A.2 trivially holds for this choice of (h_μ, γ_μ). The lemma below is useful for identifying the critical quantities for inference.

LEMMA 6.1. *Let $I_l = \int_0^\infty (1 + x^{2m})^{-l}\,dx$ for $l = 1, 2$ and $h^\dagger = h\sigma^{1/m}$. Then*

$$\sum_{k=1}^\infty \frac{1}{(1 + (2\pi h^\dagger k)^{2m})^l} \sim \frac{I_l}{2\pi h^\dagger}. \tag{6.3}$$

By Proposition 4.1, the asymptotic 95% pointwise CI for $g(z_0)$ is $\widehat{g}_{n,\lambda}(z_0) \pm 1.96\sigma_{z_0}/\sqrt{nh}$ when ignoring the bias. By the definition of $\sigma^2_{z_0}$ and Lemma 6.1, we have

$$\sigma^2_{z_0} \sim hV(K_{z_0}, K_{z_0}) = \sigma^2 h\left(1 + 2\sum_{k=1}^{\infty}(1 + (2\pi h^\dagger k)^{2m})^{-2}\right) \sim (I_2\sigma^{2-1/m})/\pi.$$

Hence, the CI becomes

$$\widehat{g}_{n,\lambda}(z_0) \pm 1.96\widehat{\sigma}^{1-1/(2m)}\sqrt{I_2/(\pi nh)}, \tag{6.4}$$

where $\widehat{\sigma}^2 = \sum_i (Y_i - \widehat{g}_{n,\lambda}(Z_i))^2/(n - \mathrm{trace}(A(\lambda)))$ is a consistent estimate of σ^2 and $A(\lambda)$ denotes the smoothing matrix; see [41]. By (4.12) and (6.2), for $l = 1, 2$,

$$\begin{aligned} Q_l(\lambda, z_0) &= \sigma^2 + \sum_{k\ge 1}\left\{\frac{|h_{2k}(z_0)|^2}{(1 + \lambda\sigma^2(2\pi k)^{2m})^l} + \frac{|h_{2k-1}(z_0)|^2}{(1 + \lambda\sigma^2(2\pi k)^{2m})^l}\right\} \\ &= \sigma^2 + 2\sigma^2\sum_{k\ge 1}\frac{1}{(1 + \lambda\sigma^2(2\pi k)^{2m})^l} \\ &= \sigma^2 + 2\sigma^2\sum_{k\ge 1}\frac{1}{(1 + (2\pi h^\dagger k)^{2m})^l}. \end{aligned}$$

By Lemma 6.1, we have $c_0 = I_2/I_1$. In particular, $c_0 = 0.75$ (0.83) when $m = 2$ (3).

To examine the pointwise asymptotic CI, we considered the true function $g_0(z) = 0.6\beta_{30,17}(z) + 0.4\beta_{3,11}(z)$, where $\beta_{a,b}$ is the density function for $\mathrm{Beta}(a, b)$, and estimated it using periodic splines with $m = 2$; σ was chosen as 0.05. In Figure 1, we compare the coverage probability (CP) of our asymptotic CI (6.4), denoted ACI, Wahba's Bayesian CI (4.3), denoted WCI and Nychka's Bayesian CI (4.7), denoted NCI, at thirty equally spaced grid points of $\mathbb{I}$. The CP was computed as the proportion of the CIs that cover g_0 at each point based on 1000 replications. We observe that, in general, all CIs exhibit similar patterns, for example, undercoverage near peaks or troughs. However, when the sample size is sufficiently large, for example, $n = 2000$, the CP of ACI is uniformly closer to 95% than that of WCI and NCI in smooth regions such as $[0.1, 0.4]$ and $[0.8, 0.9]$. We also report the average lengths of the three CIs in the titles of the plots. The ACI is the shortest, as indicated in Figure 1.

In Figure 3 of the supplementary document [33], we construct the SCB for g based on formula (5.9) by taking $d_n = (-2\log h)^{1/2}$. We compare it with the *pointwise* confidence bands constructed by linking the endpoints of the ACI, WCI and NCI at each observed covariate, denoted ACB, BCB1 and BCB2, respectively. The data were generated under the same setup as

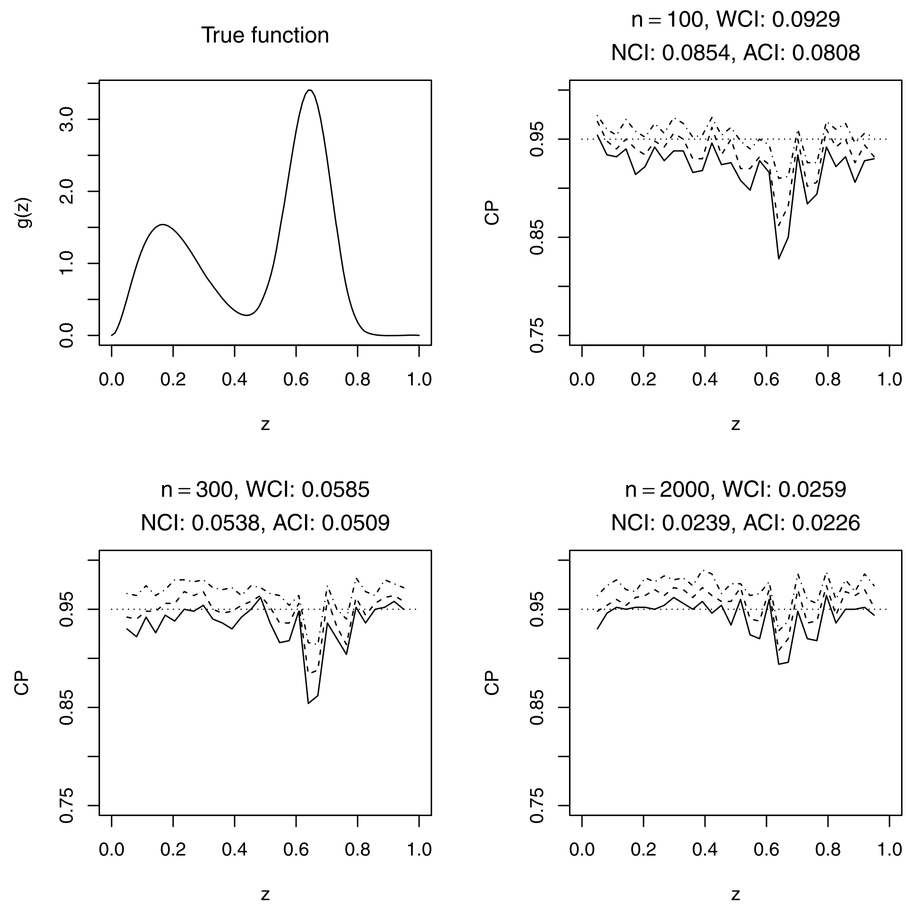

FIG. 1. *The first panel displays the true function $g_0(z) = 0.6\beta_{30,17}(z) + 0.4\beta_{3,11}(z)$ used in case* (I) *of Example 6.1. The other panels contain the coverage probabilities (CPs) of ACI (solid), NCI (dashed) and WCI (dotted dashed), and the average lengths of the three CIs (numbers in the plot titles). The CIs were built upon thirty equally spaced covariates.*

above. We observe that the coverage properties of all bands are reasonably good, and they become better as n grows. Meanwhile, the band areas, that is, the areas covered by the bands, shrink to zero as n grows. We also note that the ACB has the smallest band area, while the SCB has the largest. This is because of the d_n factor in the construction of SCB; see Remark 5.1 for more details.

To conclude case (I), we tested $H_0 : g$ *is linear* at the 95% significance level by the PLRT and GLRT. By Lemma 6.1 and (6.2), direct calculation leads to $r_K = 1.3333$ and $u_n = 0.4714(h\sigma^{1/2})^{-1}$ when $m = 2$. The data were generated under the same setup except that different test functions $g(z) = -0.5 + z + c(\sin(\pi z) - 0.5)$, $c = 0, 0.5, 1.5, 2$, were used for the purpose

TABLE 1
Power comparison of the PLRT and the GLRT in case (I) *of Example 6.1 where the test function is* $g_0(z) = -0.5 + z + c(\sin(\pi z) - 0.5)$ *with various* c *values. The significance level is 95%*

n	$c=0$ PLRT	$c=0$ GLRT	$c=0.5$ PLRT	$c=0.5$ GLRT	$c=1.5$ PLRT	$c=1.5$ GLRT	$c=2$ PLRT	$c=2$ GLRT
				100 × Power%				
20	18.60	20.10	28.40	30.10	89.60	86.30	97.30	96.10
30	13.60	14.40	33.00	30.60	98.10	96.80	99.60	99.60
70	8.30	9.40	54.40	48.40	100	100	100	100
200	5.20	5.50	95.10	92.70	100	100	100	100

of the power comparison. For the GLRT method, the R function *glkerns*( ) provided in the *lokern* package (see [18]) was used for the local polynomial fitting based on the Epanechnikov kernel. For the PLRT method, GCV was used to select the smoothing parameter. Table 1 compares the power (the proportion of rejections based on 1000 replications) for $n = 20, 30, 70, 200$. When $c \geq 1.5$ ($c = 0$) and $n = 70$ or larger, both testing methods achieve 100% power (5% correct level). We also observe that (i) the power increases as c increases, that is, the test function becomes more nonlinear; and (ii) the PLRT shows moderate advantages over the GLRT, especially in small samples such as $n = 20$. An intuitive reason for (ii) is that the smoothing spline estimate in the PLRT uses the full data information, while the local polynomial estimate employed in the GLRT uses only local data information. Of course, as n grows, this difference rapidly vanishes because of the increasing data information.

Case (II). $\mathcal{H} = H^m(\mathbb{I})$: We used cubic splines and repeated most of the procedures in case (I). A different true function $g_0(z) = \sin(2.8\pi z)$ was chosen to examine the CIs. Figure 4 in the supplementary document [33] summarizes the SCB and the pointwise bands constructed by ACB, BCB1 and BCB2. In particular, BCB1 was computed by (4.2) and BCB2 was constructed by scaling the length of BCB1 by the factor $\sqrt{27/32} \approx 0.919$. We also tested the linearity of g_0 at the 95% significance level, using the test functions $g_0(z) = -0.5 + z + c(\sin(2.8\pi z) - 0.5)$, for $c = 0, 0.5, 1.5, 2$. Table 2 compares the power of the PLRT and GLRT. From Figure 4 and Table 2, we conclude that all findings in case (I) are also true in case (II).

EXAMPLE 6.2 (Nonparametric gamma model). Consider a two-parameter exponential model

$$Y \mid Z \sim \mathrm{Gamma}(\alpha, \exp(g_0(Z))),$$

TABLE 2
Power comparison of the PLRT and the GLRT in case (II) *of Example 6.1 where the test function is* $g_0(z) = -0.5 + z + c(\sin(2.8\pi z) - 0.5)$ *with various* c *values. The significance level is 95%*

n	$c=0$ PLRT	$c=0$ GLRT	$c=0.5$ PLRT	$c=0.5$ GLRT	$c=1.5$ PLRT	$c=1.5$ GLRT	$c=2$ PLRT	$c=2$ GLRT
				$100\times$ Power%				
20	16.00	17.40	71.10	67.60	100	100	100	100
30	12.70	14.00	83.20	81.20	100	100	100	100
70	6.50	7.40	99.80	99.70	100	100	100	100
200	5.10	5.30	100	100	100	100	100	100

where $\alpha > 0$, $g_0 \in H_0^m(\mathbb{I})$ and Z is uniform over $[0,1]$. This framework leads to $\ell(y; g(z)) = \alpha g(z) + (\alpha - 1)\log y - y\exp(g(z))$. Thus, $I(z) = \alpha$, leading us to choose the trigonometric polynomial basis defined as in (6.2) with σ replaced with $\alpha^{-1/2}$, and the eigenvalues $\gamma_0 = 0$ and $\gamma_{2k} = \gamma_{2k-1} = \alpha^{-1}(2\pi k)^{2m}$ for $k \geq 1$. Local and global inference can be conducted similarly to Example 6.1.

EXAMPLE 6.3 (Nonparametric logistic regression). In this example, we consider the binary response $Y \in \{0, 1\}$ modeled by the logistic relationship

$$P(Y = 1 \mid Z = z) = \frac{\exp(g_0(z))}{1 + \exp(g_0(z))}, \tag{6.5}$$

where $g_0 \in H^m(\mathbb{I})$. Given the length of this paper, we conducted simulations only for the ACI and PLRT. A straightforward calculation gives $I(z) = \frac{\exp(g_0(z))}{(1+\exp(g_0(z)))^2}$, which can be estimated by $\widehat{I}(z) = \frac{\exp(\widehat{g}_{n,\lambda}(z))}{(1+\exp(\widehat{g}_{n,\lambda}(z)))^2}$. Given the estimate $\widehat{I}(z)$ and the marginal density estimate $\widehat{\pi}(z)$, we find the approximate eigenvalues and eigenfunctions via (2.11).

The results are based on 1000 replicated data sets drawn from (6.5), with $n = 70, 100, 300, 500$. To test whether g is linear, we considered two test functions, $g_0(z) = -0.5 + z + c(\sin(\pi z) - 0.5)$ and $g_0(z) = -0.5 + z + c(\sin(2.8\pi z) - 0.5)$, for $c = 0, 1, 1.5, 2$. We use $m = 2$. Numerical calculations reveal that the eigenvalues are $\gamma_\nu \approx (\alpha\nu)^{2m}$, where $\alpha = 4.40, 4.41, 4.47, 4.52$ and $\alpha = 4.40, 4.44, 4.71, 4.91$ corresponding to the two test functions and the four values of c. This simplifies the calculations of σ_K^2 and ρ_K^2 defined in Theorem 5.3. For instance, when $\gamma_\nu \approx (4.40\nu)^{2m}$, using a result analogous to Lemma 6.1 we have $\sigma_K^2 \approx 0.25$ and $\rho_K^2 \approx 0.19$. Then the quantities r_K and u_n are found for the PLRT method. To evaluate ACI, we considered the true function $g_0(z) = (0.15)10^6 z^{11}(1-z)^6 + (0.5)10^4 z^3(1-z)^{10} - 1$. The CP and the average lengths of the ACI are calculated at thirty evenly spaced points in $\mathbb{I}$ under three sample sizes, $n = 200, 500, 2000$.

Table 3
Power of PLRT in Example 6.3 where the test function is $g_0(z) = -0.5 + z + c(\sin(\pi z) - 0.5)$ *with various c values. The significance level is 95%*

n	$c=0$	$c=1$	$c=1.5$	$c=2$
		100 × Power%		
70	4.10	16.90	30.20	50.80
100	4.50	17.30	38.90	63.40
300	5.00	52.50	92.00	99.30
500	5.00	79.70	99.30	100

Table 4
Power of PLRT in Example 6.3 where the test function is $g_0(z) = -0.5 + z + c(\sin(2.8\pi z) - 0.5)$ *with various c values. The significance level is 95%*

n	$c=0$	$c=1$	$c=1.5$	$c=2$
		100 × Power%		
70	4.10	56.20	90.10	99.00
100	5.00	71.90	96.90	100
300	5.00	99.80	100	100
500	5.00	100	100	100

The results on the power of the PLRT are summarized in Tables 3 and 4, which demonstrate the validity of the proposed testing method. Specifically, when $c = 0$, the power reduces to the desired size 0.05; when $c \geq 1.5$ and $n \geq 300$, the power approaches one. The results for the CPs and average lengths of ACIs are summarized in Figure 2. The CP uniformly approaches the desired 95% confidence level as n grows, showing the validity of the intervals.

Acknowledgement. We are grateful for helpful discussions with Professor Chong Gu and careful proofreading of Professor Pang Du. The authors also thank the Co-Editor Peter Hall, the Associate Editor, and two referees for insightful comments that led to important improvements in the paper.

SUPPLEMENTARY MATERIAL

Supplement to "Local and global asymptotic inference in smoothing spline models" (DOI: 10.1214/13-AOS1164SUPP; .pdf). The supplementary materials contain all the proofs of the theoretical results in the present paper.

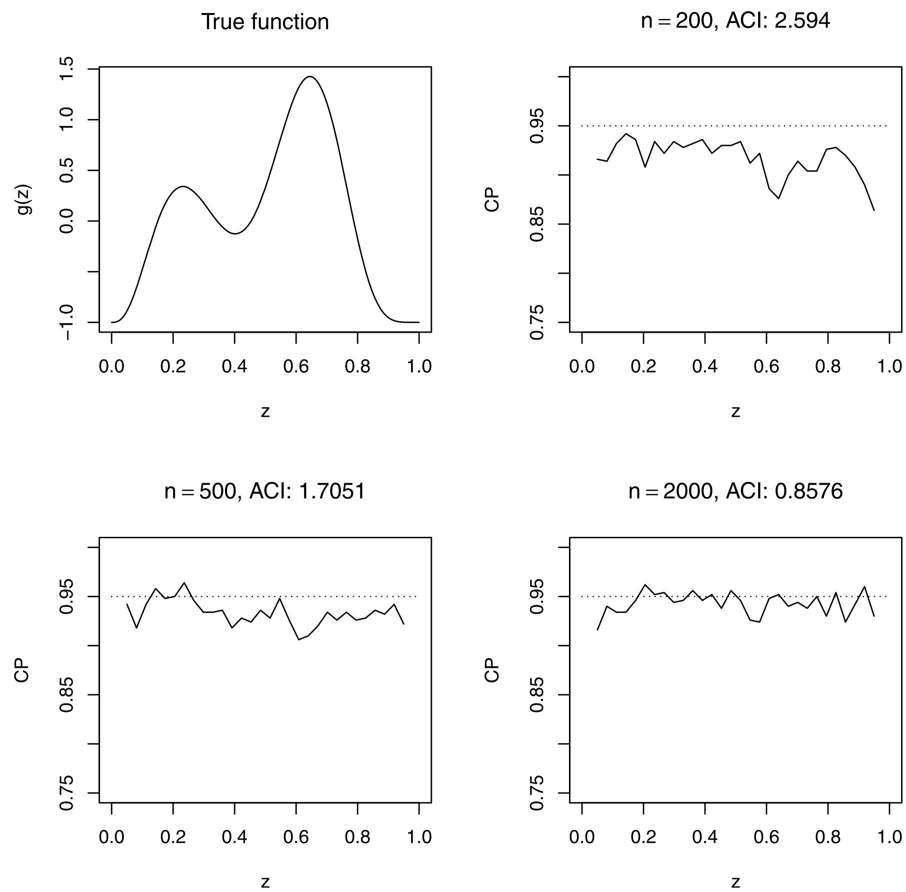

Fig. 2. *The first panel displays the true function* $g_0(z) = (0.15)10^6 z^{11}(1-z)^6 + (0.5)10^4 z^3(1-z)^{10} - 1$ *used in Example 6.3. The other panels contain the CP and average length (number in the plot title) of each ACI. The ACIs were built upon thirty equally spaced covariates.*

REFERENCES

[1] Banerjee, M. (2007). Likelihood based inference for monotone response models. *Ann. Statist.* **35** 931–956. MR2341693

[2] Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095. MR0348906

[3] Birkhoff, G. D. (1908). Boundary value and expansion problems of ordinary linear differential equations. *Trans. Amer. Math. Soc.* **9** 373–395. MR1500818

[4] Chen, J. C. (1994). Testing goodness of fit of polynomial models via spline smoothing techniques. *Statist. Probab. Lett.* **19** 65–76. MR1253314

[5] Claeskens, G. and Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31** 1852–1884. MR2036392

[6] Cox, D., Koh, E., Wahba, G. and Yandell, B. S. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Statist.* **16** 113–119. MR0924859

[7] Cox, D. D. and O'Sullivan, F. (1990). Asymptotic analysis of penalized likelihood and related estimators. *Ann. Statist.* **18** 1676–1695. MR1074429

[8] Davis, P. J. (1963). *Interpolation and Approximation*. Blaisdell, New York. MR0157156

[9] Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29** 153–193. MR1833962

[10] Fan, J. and Zhang, J. (2004). Sieve empirical likelihood ratio tests for nonparametric functions. *Ann. Statist.* **32** 1858–1907. MR2102496

[11] Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Stat.* **27** 715–731. MR1804172

[12] Genovese, C. and Wasserman, L. (2008). Adaptive confidence bands. *Ann. Statist.* **36** 875–905. MR2396818

[13] Gu, C. (2002). *Smoothing Spline ANOVA Models*. Springer, New York. MR1876599

[14] Gu, C. and Qiu, C. (1993). Smoothing spline density estimation: Theory. *Ann. Statist.* **21** 217–234. MR1212174

[15] Hall, P. (1991). Edgeworth expansions for nonparametric density estimators, with applications. *Statistics* **22** 215–232. MR1097375

[16] Hall, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist.* **20** 675–694. MR1165587

[17] Härdle, W. (1989). Asymptotic maximal deviation of M-smoothers. *J. Multivariate Anal.* **29** 163–179. MR1004333

[18] Herrmann, E. (1997). Local bandwidth choice in kernel regression estimation. *J. Comput. Graph. Statist.* **6** 35–54. MR1451989

[19] Ingster, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives I–III. *Math. Methods Statist.* **2** 85–114; **3** 171–189; **4**, 249–268.

[20] Jayasuriya, B. R. (1996). Testing for polynomial regression using nonparametric regression techniques. *J. Amer. Statist. Assoc.* **91** 1626–1631. MR1439103

[21] Ke, C. and Wang, Y. (2002). ASSIST: A suite of S-plus functions implementing spline smoothing techniques. Preprint.

[22] Krivobokova, T., Kneib, T. and Claeskens, G. (2010). Simultaneous confidence bands for penalized spline estimators. *J. Amer. Statist. Assoc.* **105** 852–863. MR2724866

[23] Liu, A. and Wang, Y. (2004). Hypothesis testing in smoothing spline models. *J. Stat. Comput. Simul.* **74** 581–597. MR2074614

[24] Mammen, E. and van de Geer, S. (1997). Penalized quasi-likelihood estimation in partial linear models. *Ann. Statist.* **25** 1014–1035. MR1447739

[25] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. Chapman & Hall, London. MR0727836

[26] Messer, K. and Goldstein, L. (1993). A new class of kernels for nonparametric curve estimation. *Ann. Statist.* **21** 179–195. MR1212172

[27] Neumann, M. H. and Polzehl, J. (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *J. Nonparametr. Stat.* **9** 307–333. MR1646905

[28] Nychka, D. (1988). Bayesian confidence intervals for smoothing splines. *J. Amer. Statist. Assoc.* **83** 1134–1143. MR0997592

[29] Nychka, D. (1995). Splines as local smoothers. *Ann. Statist.* **23** 1175–1197. MR1353501

[30] Ramil Novo, L. A. and González Manteiga, W. (2000). F tests and regression analysis of variance based on smoothing spline estimators. *Statist. Sinica* **10** 819–837. MR1787781

[31] Rice, J. and Rosenblatt, M. (1983). Smoothing splines: Regression, derivatives and deconvolution. *Ann. Statist.* **11** 141–156. MR0684872

[32] Shang, Z. (2010). Convergence rate and Bahadur type representation of general smoothing spline M-estimates. *Electron. J. Stat.* **4** 1411–1442. MR2741207

[33] Shang, Z. and Cheng, G. (2013). Supplement to "Local and global asymptotic inference in smoothing spline models." DOI:10.1214/13-AOS1164SUPP.

[34] Shao, J. (2003). *Mathematical Statistics*, 2nd ed. Springer, New York. MR2002723

[35] Silverman, B. W. (1984). Spline smoothing: The equivalent variable kernel method. *Ann. Statist.* **12** 898–916. MR0751281

[36] Stone, M. H. (1926). A comparison of the series of Fourier and Birkhoff. *Trans. Amer. Math. Soc.* **28** 695–761. MR1501372

[37] Sun, J., Loader, C. and McCormick, W. P. (2000). Confidence bands in generalized linear models. *Ann. Statist.* **28** 429–460. MR1790004

[38] Sun, J. and Loader, C. R. (1994). Simultaneous confidence bands for linear regression and smoothing. *Ann. Statist.* **22** 1328–1345. MR1311978

[39] Utreras, F. I. (1988). Boundary effects on convergence rates for Tikhonov regularization. *J. Approx. Theory* **54** 235–249. MR0960047

[40] Wahba, G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **45** 133–150. MR0701084

[41] Wahba, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia, PA. MR1045442

[42] Wang, Y. (2011). *Smoothing Splines: Methods and Applications*. *Monographs on Statistics and Applied Probability* **121**. Chapman & Hall/CRC Press, Boca Raton, FL. MR2814838

[43] Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika* **61** 439–447. MR0375592

[44] Zhang, W. and Peng, H. (2010). Simultaneous confidence band and hypothesis test in generalised varying-coefficient models. *J. Multivariate Anal.* **101** 1656–1680. MR2610738

[45] Zhou, S., Shen, X. and Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Ann. Statist.* **26** 1760–1782. MR1673277

University of Notre Dame
Notre Dame, Indiana 46556
USA
E-mail: zuofengshang@gmail.com

Department of Statistics
Purdue University
250 N. University St.
West Lafayette, Indiana 47906
USA
E-mail: chengg@stat.purdue.edu

Submitted to the Annals of Statistics

Supplementary document to

LOCAL AND GLOBAL ASYMPTOTIC INFERENCE IN SMOOTHING SPLINE MODELS

By Zuofeng Shang[‡] and Guang Cheng[§]

University of Notre Dame and Purdue University

August 27, 2013

In this document, we first give a table that lists the notation of the paper, and then give the proofs. In particular, we describe and prove the minimax results for the PLRT.

We organize this document as follows:

- In Section S.1, we give a table defining our notation, and indicating the page numbers of the first occurences.
- Sections S.2, S.3, and S.4 include the proofs of Proposition 2.1, Proposition 2.2, and Lemma 3.1.
- In Section S.5, we prove the concentration inequality in Lemma 3.2.
- In Section S.6, we prove Proposition 3.3, i.e., the convergence rate of $\widehat{g}_{n,\lambda}$.
- In Section S.7, we prove Theorem 3.4, i.e., the FBR. In Section S.8, we prove the pointwise asymptotic normality of the smoothing spline estimate in Theorem 3.5.
- In Section S.9, we prove Corollary 3.7 on the pointwise asymptotic normality of $\widehat{g}_{n,\lambda}$ in L_2 regression.
- In Section S.10, we sketch the proof of another technical tool, i.e., the restricted FBR in Theorem 4.3, which is used to establish the asymptotic null distribution of the local LRT.
- In Section S.11, we show the Wilks phenomenon of the local LRT.
- In Section S.12, we prove Corollary 4.5.
- In Section S.13, we demonstrate the validity of the proposed SCB.
- In Section S.14, we prove Proposition 5.2, i.e., the equivalent kernel conditions.
- In Section S.15, we derive the null limiting distribution of the PLRT.

[‡]Postdoctoral Fellow.

[§]Corresponding Author. Associate Professor. Research Sponsored by NSF, DMS-0906497, and CAREER Award DMS-1151692

- In Section S.16, we prove Theorem 5.4, i.e., when the data are normal, the PLRT attains the optimal minimax rates of testing. We further extend this result to a more general modeling framework in Section S.17.
- In Section S.18, we show Lemma 6.1.
- We give two figures for Example 6.1 showing the performance of the proposed ACI and SCB.

The reference labels of the equations, theorems, propositions, and lemmas in this document are consistent with those in the main text of the paper.

Notation	Meaning	First appearing in page No.
Y	response	4
Z	covariate	4
T	$T = (Y, Z)$ denotes the full data variable	4
$\mathcal{Y}$	range of Y	4
$\mathbb{I}$	$\mathbb{I} = [0, 1]$ denotes the range of Z	4
F	the link function	4
g_0	the true function	4
$Q(y; \mu)$	the quasi-likelihood function	4
$\ell(y; a)$	the general criteria function	4
m	the roughness degree of g_0	4
$J(\cdot, \cdot)$	the penalty function	4
$\mathcal{H}$	the parameter space for g	4
$H^m(\mathbb{I})$	the Sobolev space on $[0, 1]$	4
$H_0^m(\mathbb{I})$	the homogeneous Sobolev space on $[0, 1]$	4
$\ell_{n,\lambda}$	the penalized empirical criteria function	4
$\widehat{g}_{n,\lambda}$	the unconstrained estimate of g	4
λ	the smoothing parameter	4
$\mathcal{N}_m$	the null space for J	4
$\mathcal{I}_0$	the range of g_0	4
$\mathcal{I}$	a bounded open interval including $\mathcal{I}_0$	4
$\dot{\ell}_a, \ddot{\ell}_a, \ell_a'''$	the first-, second- and third-order partial derivatives w.r.t. a	4
$I(Z)$	$-E\{\ddot{\ell}_a(Y; g_0(Z))\|Z\}$	5
$\langle\cdot, \cdot\rangle, \\|\cdot\\|$	the inner product and norm	5
K	the reproducing kernel function	6
K_z	the function determined by $K_z(\cdot) = K(z, \cdot)$	6
V	$V(g, \widetilde{g}) = E\{I(Z)g(Z)\widetilde{g}(Z)\}$	6
W_λ	the penalty operator	6
id	the identity operator	6
$a_\nu \asymp b_\nu$	asymptotically equivalent	6
$a_\nu \sim b_\nu$	asymptotically equal	6
$\\|\cdot\\|_{\sup}$	the supremum norm	6
γ_ν	the eigenvalue	6
h_ν	the eigenfunction function corresponding to γ_ν	6
D	the Fréchet differential operator	7
$S_{n,\lambda}$	$= D\ell_{n,\lambda}$	7
h	$= \lambda^{1/(2m)}$	7
$\\|\cdot\\|_{\mathcal{H}}$	a norm stronger than $\\|\cdot\\|$	8
λ^*	a sequence satisfying $\lambda^* \asymp n^{-2m/(2m+1)}$	8
h^*	$= (\lambda^*)^{1/(2m)}$	8
g_0^*	$= (id - W_\lambda)g_0$	9
$\sigma_{z_0}^2$	the asymptotic variance of $\widehat{g}_{n,\lambda}(z_0)$	9
b_{z_0}	the asymptotic estimation bias of $\widehat{g}_{n,\lambda}(z_0)$	9
$h^\dagger$	$= h\sigma^{1/m}$	11
I_1, I_2	two explicit integrals	11
$a(h^\dagger)$	a quantity introduced by Wahba (1983)	12
$\widehat{g}_{n,\lambda}^0$	the optimizer constrained with $g(z_0) = 0$	13
$\mathcal{H}_0$	a subspace of $\mathcal{H}$ with restriction $g(z_0) = 0$	13
K^*	a restricted version of K in $\mathcal{H}_0$	13
W_λ^*	a restricted version of W_λ in $\mathcal{H}_0$	13
r_n	rate of convergence	14
c_0	a scaling constant in local LRT	14
$\omega(\cdot)$	an equivalent kernel function	15
ω_0	an equivalent kernel function in the specific L_2 regression	17
$r_K, \rho_K, \sigma_K^2, u_n$	constants in the PLRT	19
$\widehat{\sigma}^2$	estimated variance	22

TABLE 5

A table that lists all useful notation, their meanings and where they first appear.

S.1. *Notation Table.*

S.2. *Proof of Proposition 2.1.* Based on the definition (2.8), we can write $\|g\|^2 = V(g,g) + \lambda J(g,g)$, and then plug in the Fourier expansion of g to obtain the explicit expression of $\|g\|^2$. A direct calculation reveals that

$$\langle g, h_\nu \rangle = \langle \sum_\mu V(g,h_\mu)h_\mu, h_\nu \rangle = V(g,h_\nu)(1+\lambda\gamma_\nu), \tag{S.1}$$

for any $g \in \mathcal{H}$ and $\nu \in \mathbb{N}$. It follows by (S.1) that $V(K_z, h_\nu) = \langle K_z, h_\nu \rangle/(1+\lambda\gamma_\nu) = h_\nu(z)/(1+\lambda\gamma_\nu)$. Hence, we can obtain the expression of $K_z(\cdot)$ by using $K_z(\cdot) = \sum_\nu V(K_z,h_\nu)h_\nu(\cdot)$. Furthermore, (S.1) implies that $V(W_\lambda h_\nu, h_\mu) = \langle W_\lambda h_\nu, h_\mu \rangle/(1+\lambda\gamma_\mu) = \lambda\gamma_\mu\delta_{\mu\nu}/(1+\lambda\gamma_\mu)$, for any $\nu, \mu \in \mathbb{N}$. Finally, we can conclude the proof of Proposition 2.1 by using $W_\lambda h_\nu(\cdot) = \sum_\mu V(W_\lambda h_\nu, h_\mu)h_\mu$.

S.3. *Proof of Proposition 2.2.* The usual L_2-inner product is defined to be $\langle g, \xi \rangle_{L_2} = \int_0^1 g(z)\xi(z)dz$. Let D be the differential operator, i.e., $D\phi = \frac{d}{dz}\phi$, and $\omega = 1/(I\pi)$. Thus, $\omega \in C^m(\mathbb{I})$ is positive and finitely upper bounded. It follows from [5] that the growing rate for γ_ν is of order ν^{2m}. Since the operator $L_0 = (-1)^m\omega D^{2m}$ is self-adjoint under the inner product V, that is, $V(L_0 g, \xi) = V(g, L_0\xi)$ for any $\xi, g \in C^{2m}(\mathbb{I})$ satisfying the boundary conditions in (2.11), orthogonality and completeness of h_ν under V thus follow from Theorem 2.1 (pp. 189) and Theorem 4.2 (pp. 199) of [8] with the usual L_2-inner product $\langle , \rangle_{L_2}$ replaced with V. Therefore, when h_ν are normalized to $V(h_\nu,h_\nu) = 1$, they form an orthonormal and complete set in $L_2(\mathbb{I}; V)$.

Next we show that $h_\nu^{(m)}$, $\nu \geq m$, are complete in $L_2(\mathbb{I})$ under $\langle , \rangle_{L_2}$. The idea follows from arguments on page 147 of [31]. The eigenspace corresponding to the zero eigenvalue contains functions ϕ that satisfy $(-1)^m\phi^{(2m)} = 0$ with boundary conditions $\phi^{(j)}(0) = \phi^{(j)}(1) = 0$ for $j = m, \ldots, 2m-1$, thus it follows from [49] that this eigenspace is $\mathcal{P}_{m-1}$, i.e., the set of all polynomials of degree at most $m-1$. Let h_ν, $\nu = 0, \ldots, m-1$, be the orthonormal basis (under V) of $\mathcal{P}_{m-1}$ corresponding to $\gamma_0 = \ldots = \gamma_{m-1} = 0$. Note $\gamma_\nu > 0$ for $\nu \geq m$. For $g \in L_2(\mathbb{I})$ such that for any $\nu \geq m$, $\int_0^1 g h_\nu^{(m)} = 0$, let ξ be a solution of $\xi^{(m)} = g$, then using integration by parts we have $0 = \int_0^1 \xi h_\nu^{(2m)} = (-1)^m\gamma_\nu V(\xi, h_\nu)$. Therefore $V(\xi, h_\nu) = 0$ for any $\nu \geq m$. By completeness of h_ν, ξ must be a linear combination of $h_0, \ldots, h_{m-1}$, a polynomial with degree at most $m-1$. So $g = \xi^{(m)} = 0$ implying the completeness of $h_\nu^{(m)}/\gamma_\nu^{1/2}$ for $\nu \geq m$ in $L_2(\mathbb{I})$ under $\langle , \rangle_{L_2}$. Now, for any $\tilde{g} \in \mathcal{H}$, by completeness of h_ν in $L_2(\mathbb{I})$ under V-norm, $\tilde{g} = \sum_{\nu\in\mathbb{N}} V(\tilde{g}, h_\nu)h_\nu$ with convergence in V-norm. Since $V(\tilde{g}, h_\nu) = \int_0^1 \tilde{g}^{(m)}h_\nu^{(m)}/\gamma_\nu$, by completeness of $h_\nu^{(m)}/\gamma_\nu^{1/2}$ for $\nu \geq m$ in $L_2(\mathbb{I})$ in usual $\|\cdot\|_{L_2}$-norm, $\tilde{g}^{(m)} = \sum_{\nu\geq m}\langle \tilde{g}^{(m)}, h_\nu^{(m)}\rangle_{L_2} h_\nu^{(m)}/\gamma_\nu = \sum_{\nu\geq m} V(\tilde{g}, h_\nu)h_\nu^{(m)}$ with convergence in usual L_2-norm. This implies $\tilde{g} = \sum_\nu V(\tilde{g}, h_\nu)h_\nu$ converges in $\|\cdot\|$.

Next we show the uniform boundedness of h_ν. We only consider those h_ν corresponding to nonzero γ_ν. If $\gamma_\nu \neq 0$ and h_ν satisfy $(-1)^m h_\nu^{(2m)} = \gamma_\nu I\pi h_\nu$ and $V(h_\nu, h_\nu) = 1$, then using the boundary conditions in (2.11) and integration by parts, one can check that $J(h_\nu, h_\nu) = \gamma_\nu$. Dividing $I\pi$ and taking m-order derivatives on both sides, one obtains $Lh_\nu^{(m)} = \gamma_\nu h_\nu^{(m)}$ with $h_\nu^{(m+j)}(0) = h_\nu^{(m+j)}(1) = 0$, $j = 0, \ldots, m-1$, where $L = (-1)^m \sum_{j=0}^m \binom{m}{j} \omega^{(j)} D^{2m-j}$. Therefore, $h_\nu^{(m)}$ is an eigenfunction of L with eigenvalue γ_ν. Denote the eigenfunctions and eigenvalues of L to be ψ_ν and λ_ν subject to $\psi_\nu^{(j)}(0) = \psi_\nu^{(j)}(1) = 0$, $j = 0, \ldots, m-1$. We need to transform L to normal form. Let $t(z) = \int_0^z [I(s)\pi(s)]^{1/(2m)} ds/C$ and $C = \int_0^1 [I(z)\pi(z)]^{1/(2m)} dz$. Define $\phi_\nu(t(z)) = \psi_\nu(z)$. Then by a direct examination, ϕ_ν satisfies the following differential equation

(S.2)
$$\phi_\nu^{(2m)}(t) + q_{2m-1}(t)\phi_\nu^{(2m-1)}(t) + \ldots + q_0(t)\phi_\nu(t) = \rho_\nu \phi_\nu(t), \phi_\nu^{(j)}(0) = \phi_\nu^{(j)}(1) = 0, j = 0, \ldots, m-1$$

where q_j, $j = 0, \ldots, 2m-1$, are coefficient functions depending only on $I\pi$ and m, and $\rho_\nu = \lambda_\nu C^{2m}$. In general the form of q_j are complicated though they can be determined by Faá di Bruno's formula ([25]). As an illustration, when $m = 2$, $q_0(t) = 0$, $q_3(t) = -(C/4)\omega^{(1)}(z(t))\omega(z(t))^{-3/4}$, $q_2(t) = -(C^2/4)(\omega^{(1)}(z(t)))^2\omega(z(t))^{-3/2}$, and

$$q_1(t) = C^3(-5\omega(z(t))^{-9/4}(\omega^{(1)}(z(t)))^3/64 + 3\omega(z(t))^{-5/4}\omega^{(1)}(z(t))\omega^{(2)}(z(t))/16 - \omega(z(t))^{-1/4}\omega^{(3)}(z(t))),$$

where $z(t)$ is the inverse function of $t(z)$ and $b_2(z) = [I(z)\pi(z)]^{1/4}$. Define

$$u_\nu(t) = \phi_\nu(t) \exp\left(\frac{1}{2m}\int_0^t q_{2m-1}(s) ds\right), \tag{S.3}$$

then (S.2) is equivalent to

$$\tilde{L}u_\nu \equiv u_\nu^{(2m)}(t) + p_{2m-2}(t)u_\nu^{(2m-2)} + \ldots + p_0(t)u_\nu(t) = \rho_\nu u_\nu(t), \tag{S.4}$$

with the boundary conditions $u_\nu^{(j)}(0) = u_\nu^{(j)}(1) = 0$, $j = 0, \ldots, m-1$. Note (S.4) is the classic form of differential systems discussed in [5]. According to [5], ρ_ν are simple due to the regular boundary conditions, and the residual of the Green function $G(z_1, z_2; \rho)$ for $\tilde{L} - \rho I$ at pole ρ_ν is given by $\frac{u_\nu(t_1)u_\nu(t_2)}{\|u_\nu\|_{L_2}^2}$, where $\|\cdot\|_{L_2}$ denotes the usual L_2-norm. On the other hand, the residue can also be represented by $\frac{1}{2\pi\sqrt{-1}}\int_{\Gamma_{\rho_\nu}} 2m\zeta^{2m-1}G(t_1, t_2, \zeta^{2m})d\zeta$ (pp. 722, [45]), where $\zeta = \rho^{1/(2m)}$, Γ_{ρ_ν} denotes the contour centered around pole ρ_ν with suitably small radius. By equation (56) and the discussions below in [5], $2m\zeta^{2m-1}G(t_1, t_2; \zeta^{2m})$ is uniformly bounded for $t_1, t_2 \in \mathbb{I}$, thus the residue is uniformly bounded for all t_1, t_2. In particular, letting $t_1 = t_2 = t$, we get $|u_\nu(t)| \le c\|u_\nu\|_{L_2}$ for any $t \in \mathbb{I}$ with a universal constant $c > 0$. Since q_{2m-1} achieves finite upper and lower bounds on $\mathbb{I}$, by (S.3), there is a universal constant $c_1 > 0$ such that for any ν, $\|\phi_\nu\|_{\sup} \le c_1\|\phi_\nu\|_{L_2}$. Now, using $\phi_\nu(t(z)) = \psi_\nu(z)$, we get

$$\|\psi_\nu\|_{\sup}^2 = \|\phi_\nu\|_{\sup}^2 \le c_1^2\|\phi_\nu\|_{L_2}^2 = c_1^2\int_0^1 |\phi_\nu(t)|^2 dt = c_1^2\int_0^1 |\phi_\nu(t(z))|^2 |I(z)\pi(z)|^{1/(2m)} dz \le c_1^2 c_{I\pi}^2 \|\psi_\nu\|_{L_2}^2,$$

where $c_{I\pi}$ is a constant depending only on $I\pi$ and m. So $\|\psi_\nu\|_{\sup} \le c_1 c_{I\pi}\|\psi_\nu\|_{L_2}$. Letting $\psi_\nu = h_\nu^{(m)}$ and using the fact that $\|h_\nu^{(m)}\|_{L_2}^2 = \gamma_\nu$, we have $\|h_\nu^{(m)}\|_{\sup} \le c_1 c_{I\pi}\gamma_\nu^{1/2}$, for any $\nu \in \mathbb{N}$.

By the Sobolev embedding theorem ([1]), $\|h_\nu\|_{\sup}^2 \le c^2(V(h_\nu, h_\nu) + J(h_\nu, h_\nu)) = c^2(1+\gamma_\nu)$. By Theorem 5 of [49], for any $j = 1, \ldots, m$, there is a constant $C_j > 0$ such that $\|h_\nu^{(j)}\|_{\sup} \le C_j(1+\gamma_\nu)^{1/2}$, $\forall \nu \in \mathbb{N}$. Therefore, taking the mth-order derivative on both sides of $(-1)^m h_\nu^{(2m)} = \gamma_\nu I\pi h_\nu$, one can show that $\|h_\nu^{(3m)}\|_{\sup} \le \gamma_\nu \sum_{j=0}^m \binom{m}{j}\|(I\pi)^{(m-j)}\|_{\sup} \cdot \|h_\nu^{(j)}\|_{\sup} \le c_2(1+\gamma_\nu)^{3/2}$ for some constant $c_2 > 0$ and any ν. Again, by Theorem 5 of [49] for $h_\nu^{(m)}$ and $\epsilon = \gamma_\nu^{-1/(2m)}$, we have $\|h_\nu^{(2m)}\|_{\sup} \le C'_m(1+\gamma_\nu)$, which implies $\|h_\nu\|_{\sup} \le C'_m(\inf_z |I(z)|)^{-1}(1+\gamma_\nu)/\gamma_\nu \le C''_m$, with a universal constant C''_m unrelated to ν. This proves the desired uniform boundedness of h_ν.

S.4. *Proof of Lemma 3.1.* For any $z \in \mathbb{I}$, $|\langle K_z, g\rangle| \le \|K_z\| \cdot \|g\|$. So, we only need to find the upper bound for $\|K_z\|$. By Proposition 2.1 and the boundedness of h_μ, we have

$$\|K_z\|^2 = K(z,z) = \sum_{\mu\in\mathbb{Z}} \frac{|h_\mu(z)|^2}{1+\lambda\gamma_\mu} \le C \sum_{\mu\in\mathbb{Z}} \frac{1}{1+\lambda\gamma_\mu} \le c_m^2 \lambda^{-1/(2m)} = c_m^2 h^{-1}, \tag{S.5}$$

where $c_m > 0$ is a constant that does not rely on z and h. So $\|K_z\| \le c_m h^{-1/2}$.

S.5. *Proof of Lemma 3.2.* For any $g, f \in \mathcal{G}$, by Lemma 3.1,

$$\begin{aligned}\|(\psi_n(T;f) - \psi_n(T;g))K_Z\| &\le c_m^{-1} h^{1/2}\|f-g\|_{\sup} \cdot \|K_Z\| \\ &\le c_m^{-1} h^{1/2}\|f-g\|_{\sup} \cdot c_m h^{-1/2} = \|f-g\|_{\sup}.\end{aligned}$$

By Theorem 3.5 of [38], for any $t > 0$, $P\left(\|Z_n(f) - Z_n(g)\| \ge t\right) \le 2\exp\left(-\frac{t^2}{8\|f-g\|_{\sup}^2}\right)$. Then by Lemma 8.1 in [27], we have $\|\|Z_n(g) - Z_n(f)\|\|_{\psi_2} \le 8\|g-f\|_{\sup}$, where $\|\cdot\|_{\psi_2}$ denotes the Orlicz norm associated with $\psi_2(s) \equiv \exp(s^2) - 1$. It follows by Theorem 8.4 of [27] that for arbitrary $\delta > 0$,

$$\begin{aligned}\left\| \sup_{\substack{g,f\in\mathcal{G}\\ \|g-f\|_{\sup}\le\delta}} \|Z_n(g) - Z_n(f)\| \right\|_{\psi_2} &\le C'\left(\int_0^\delta \sqrt{\log(1+N(\delta,\mathcal{G},\|\cdot\|_{\sup}))} + \delta\sqrt{\log(1+N(\delta,\mathcal{G},\|\cdot\|_{\sup})^2)}\right) \\ &\asymp h^{-(2m-1)/(4m)}\delta^{1-1/(2m)}.\end{aligned}$$

So, again, by Lemma 8.1 in [27],

$$P\left(\sup_{\substack{g\in\mathcal{G}\\ \|g\|_{\sup}\le\delta}} \|Z_n(g)\| \ge t\right) \le 2\exp(-h^{(2m-1)/(2m)}\delta^{-2+1/m}t^2). \tag{S.6}$$

Let $b_n = n^{1/2}h^{-(2m-1)/(4m)}$, $\varepsilon = b_n^{-1}$, $\gamma = 1 - 1/(2m)$, $T_n = (5\log\log n)^{1/2}$, and $Q_\varepsilon = [-\log\varepsilon - 1]$, where $[a]$ denotes the integer part of a. Then by (S.6),

$$
\begin{aligned}
P\left(\sup_{g\in\mathcal{G}}\frac{\sqrt{n}\|Z_n(g)\|}{a_n\|g\|_{\sup}^{\gamma}+1}\geq T_n\right) &\leq P\left(\sup_{\substack{g\in\mathcal{G}\\ \|g\|_{\sup}\leq\varepsilon^{1/\gamma}}}\frac{\sqrt{n}\|Z_n(g)\|}{a_n\|g\|_{\sup}^{\gamma}+1}\geq T_n\right)\\
&\quad+\sum_{l=0}^{Q_\varepsilon}P\left(\sup_{\substack{g\in\mathcal{G}\\ (2^l\varepsilon)^{1/\gamma}\leq\|g\|_{\sup}\leq(2^{l+1}\varepsilon)^{1/\gamma}}}\frac{\sqrt{n}\|Z_n(g)\|}{a_n\|g\|_{\sup}^{\gamma}+1}\geq T_n\right)\\
&\leq P\left(\sup_{\substack{g\in\mathcal{G}\\ \|g\|_{\sup}\leq\varepsilon^{1/\gamma}}}\sqrt{n}\|Z_n(g)\|\geq T_n\right)\\
&\quad+\sum_{l=0}^{Q_\varepsilon}P\left(\sup_{\substack{g\in\mathcal{G}\\ \|g\|_{\sup}\leq(2^{l+1}\varepsilon)^{1/\gamma}}}\sqrt{n}\|Z_n(g)\|\geq(1+2^l)T_n\right)\\
&\leq 2\exp\left(-h^{(2m-1)/(2m)}(\varepsilon^{1/\gamma})^{-2+1/m}T_n^2/n\right)\\
&\quad+\sum_{l=0}^{Q_\varepsilon}2\exp\left(-h^{(2m-1)/(2m)}[(2^{l+1}\varepsilon)^{1/\gamma}]^{-2+1/m}T_n^2(2^l+1)^2/n\right)\\
&= 2\exp\left(-T_n^2\right)+\sum_{l=0}^{Q_\varepsilon}2\exp\left(-2^{-2(l+1)}T_n^2(2^l+1)^2\right)\\
&\leq 2(Q_\varepsilon+2)\exp\left(-T^2/4\right)\leq \mathrm{const}\cdot\log n(\log n)^{-5/4}\to 0,
\end{aligned}
$$

as $n\to\infty$. This proves the result.

S.6. *Proof of Proposition 3.3.* To prove Proposition 3.3, we first need the following Lemma. Denote $N(\delta,\mathcal{G},\|\cdot\|_{\sup})$ as the δ-covering number of the function class $\mathcal{G}$ in terms of the uniform norm.

Lemma S.1. *Suppose that $c_m^{-2}h\lambda^{-1}>1$. Then, for any $\delta>0$, we have*

$$\log N(\delta,\mathcal{G},\|\cdot\|_{\sup})\leq C(h\lambda^{-1})^{1/(2m)}\delta^{-1/m},$$

where $C>0$ is an universal constant.

Proof of Lemma S.1. Note that by $c_m^{-2}h\lambda^{-1}>1$,

$$\mathcal{G}=(c_m^{-2}h\lambda^{-1})^{1/2}\cdot\{g\in\mathcal{H}|\|g\|_{\sup}\leq(c_m^{-2}h\lambda^{-1})^{-1/2},J(g,g)\leq 1\}\subset(c_m^{-2}h\lambda^{-1})^{1/2}\mathcal{T},$$

where $\mathcal{T} = \{g \in \mathcal{H} | \|g\|_{\sup} \leq 1, J(g,g) \leq 1\}$. So, by [27], we have

$$\begin{aligned} \log N(\delta, \mathcal{G}, \|\cdot\|_{\sup}) &\leq \log N(\delta, (c_m^{-2} h\lambda^{-1})^{1/2}\mathcal{T}, \|\cdot\|_{\sup}) \\ &= \log N((c_m^{-2} h\lambda^{-1})^{-1/2}\delta, \mathcal{T}, \|\cdot\|_{\sup}) \\ &\leq c((c_m^{-2} h\lambda^{-1})^{-1/2}\delta)^{-1/m} = c c_m^{-1/m}(h\lambda^{-1})^{1/(2m)}\delta^{-1/m}. \end{aligned}$$

□

Consider the function class $\mathcal{F} = \{g \in \mathcal{H} | \|g\|_{\sup} \leq 1, J(g,g) \leq 1\}$. By Lemma S.1, for any $\delta > 0$, we know $\log N(\delta, \mathcal{F}, \|\cdot\|_{\sup}) \leq c\delta^{-1/m}$, where c is some universal constant. Then a modification of Lemma 3.2 leads to

Lemma S.2. *Suppose that ψ_n satisfies the following Lipschitz continuity condition, i.e.,*

$$|\psi_n(T;f) - \psi_n(T;g)| \leq c_m^{-1} h^{1/2} \|f - g\|_{\sup}, \text{ for all } f, g \in \mathcal{F}, \tag{S.7}$$

where c_m is specified in Lemma 3.1. Then we have

$$\lim_{n\to\infty} P\left(\sup_{g\in\mathcal{F}} \frac{\|Z_n(g)\|}{\|g\|_{\sup}^{1-1/(2m)} + n^{-1/2}} \leq (5\log\log n)^{1/2}\right) = 1,$$

where the empirical process $Z_n(g)$ is defined in (3.1).

Denote $g = \widehat{g}_{n,\lambda} - g_0$. By consistency of $\widehat{g}_{n,\lambda}$ in $\|\cdot\|_{\mathcal{H}}$-norm and Sobolev embedding Theorem (see [1]), we know that $\widehat{g}_{n,\lambda}(z)$ falls in $\mathcal{I}$ for any $z \in \mathbb{I}$ and large enough n. By Taylor's expansion,

$$\ell_{n,\lambda}(g_0 + g) - \ell_{n,\lambda}(g_0) = S_{n,\lambda}(g_0)g + \frac{1}{2}DS_{n,\lambda}(g_0)gg + \frac{1}{6}D^2 S_{n,\lambda}(g^\star)ggg \geq 0,$$

where $g^\star = g_0 + t^\star g$ for some $t^\star \in [0,1]$. Denote the three terms on the right side of the above equation by I_1, I_2, I_3. Next we will study the rates for these terms. Denote $A_i = \{\sup_{a\in\mathcal{I}} |\ddot{\ell}_a(Y_i;a)| + \sup_{a\in\mathcal{I}} |\ell_a'''(Y_i;a)| \leq C\log n\}$. By (2.4), we may choose C to be large so that $\cap_i A_i$ has large probability and $P(A_i^c) = O(n^{-2})$. Then, on $\cap_i A_i$, we have

$$\begin{aligned} |6I_3| &\leq \frac{1}{n}\sum_{i=1}^n \sup_{a\in\mathcal{I}} |\ell_a'''(Y_i;a)| \cdot |g(Z_i)|^3 \\ &\leq \frac{1}{n}\|g\|_{\sup} \sum_{i=1}^n \sup_{a\in\mathcal{I}} |\ell_a'''(Y_i;a)| \cdot g(Z_i)^2 \\ &= \frac{1}{n}\|g\|_{\sup} \langle \sum_{i=1}^n \psi(T_i;g) K_{Z_i}, g\rangle \\ &= \frac{1}{n}\|g\|_{\sup} \langle \sum_{i=1}^n [\psi(T_i;g)K_{Z_i} - E\{\psi(T;g)K_Z\}], g\rangle + \|g\|_{\sup} E\{\psi(T;g)g(Z)\}, \end{aligned}$$

where $\psi(T_i; g) = \sup_{a\in\mathcal{I}} |\ell'''_a(Y_i; a)| g(Z_i) I_{A_i}$. Let $\psi_n(T_i; g) = (C\log n)^{-1} c_m^{-1} h^{1/2}\psi(T_i; g)$, which satisfies (S.7). Thus, by Lemma S.2, we have, with large probability give large n,

$$\|\sum_{i=1}^n [\psi_n(T_i; g)K_{Z_i} - E\{\psi_n(T; g)K_Z\}]\| \le (n^{1/2}\|g\|_{\sup}^{1-1/(2m)} + 1)(5\log\log n)^{1/2}.$$

So, by Cauchy's inequality, we have

$$\left|\langle\sum_{i=1}^n [\psi(T_i; g)K_{Z_i} - E\{\psi(T; g)K_Z\}], g\rangle\right| \le \|g\| \cdot (n^{1/2}\|g\|_{\sup}^{1-1/(2m)} + 1)(5\log\log n)^{1/2}.$$

On the other hand, by Assumption A.1 (a), we have

$$E\{\psi(T; g)g(Z)\} \le E\{\sup_{a\in\mathcal{I}} |\ell'''_a(Y; a)| g(Z)^2\} \le 2C_0^2 C_1 \|g\|^2.$$

By $(n^{1/2}h)^{-1}(\log\log n)^{m/(2m-1)}(\log n)^{2m/(2m-1)} = o(1)$, which implies $(n^{1/2}h)^{-1}(\log\log n)^{1/2}\log n = o(1)$, we have

$$\begin{aligned} |6I_3| &\le \frac{1}{n}\|g\|_{\sup} \cdot \|g\|(C\log n)c_m h^{-1/2}(n^{1/2}\|g\|_{\sup}^{1-1/(2m)} + 1)(5\log\log n)^{1/2} + 2C_0^2C_1\|g\|_{\sup} \cdot \|g\|^2 \\ &= c_m^2 C'(n^{1/2}h)^{-1}(\log\log n)^{1/2}(\log n)\|g\|^2 + 2C_0^2 C_1\|g\|_{\sup} \cdot \|g\|^2 \\ &= o_P(1) \cdot \|g\|^2. \qquad \text{(S.8)} \end{aligned}$$

To approximate I_2, by Cauchy's inequality, we have

$$\begin{aligned} \left|E\{\ddot{\ell}_a(Y; g_0(Z))I_{A^c} g(Z)^2\}\right| &\le E\{|\ddot{\ell}_a(Y; g_0(Z))|^2 I_{A^c} g(Z)^4\}^{1/2} \cdot P(A^c)^{1/2} \\ &\le O(1) \cdot (\log n)\|g\|_{\sup}\|g\| n^{-1} = \|g\|^2 O((nh)^{-1/2}) = o(1)\|g\|^2. \end{aligned}$$

By changing ψ and ψ_n in the proof of (S.8) to $\psi(T_i; g) = \ddot{\ell}_a(Y_i; g_0(Z_i))gI_{A_i}$ and $\psi_n(T_i; g) = (C\log n)^{-1}c_m^{-1}h^{1/2}\psi(T_i; g)$, and using an argument similar to the proof of (S.8), we have

$$\begin{aligned} |[DS_{n,\lambda}(g_0) - E\{DS_{n,\lambda}(g_0)\}]gg| &\le Cc_m h^{-1+1/(4m)} n^{-1/2}(\log\log n)^{1/2}(\log n)\|g\|^{2-1/(2m)} \\ &\quad + Cc_m(nh^{1/2})^{-1}(\log\log n)^{1/2}(\log n)\|g\| + o_P(1)\|g\|^2. \end{aligned}$$

Thus,

$$\begin{aligned} 2I_2 &= -\|g\|^2 + Cc_m h^{-1+1/(4m)} n^{-1/2}(\log\log n)^{1/2}(\log n)\|g\|^{2-1/(2m)} \\ &\quad + Cc_m(nh^{1/2})^{-1}(\log\log n)^{1/2}(\log n)\|g\| + o_P(1)\|g\|^2. \qquad \text{(S.9)} \end{aligned}$$

Note that $E\{\|\sum_{i=1}^n \epsilon_i K_{Z_i}\|^2\} = O(nh^{-1})$. By (S.44) in the main paper, we have

$$\|S_{n,\lambda}(g_0)\| = O_P((nh)^{-1/2} + \lambda^{1/2}). \qquad \text{(S.10)}$$

Combining (S.8), (S.9), and (S.10), and by $(nh^{1/2})^{-1}(\log\log n)^{1/2}(\log n) = o((nh)^{-1/2})$, we have for some large C'

$$(1+o_P(1))\|g\|^2 \le C'((nh)^{-1/2}+\lambda^{1/2})\|g\| + Cc_m h^{-1+1/(4m)} n^{-1/2}(\log\log n)^{1/2}(\log n)\|g\|^{2-1/(2m)}.$$

Solving this inequality, and using $(n^{1/2}h)^{-1}(\log\log n)^{m/(2m-1)}(\log n)^{2m/(2m-1)} = o(1)$, we get $\|g\| = O_P((nh)^{-1/2}+\lambda^{1/2})$.

S.7. *Proof of Theorem 3.4.* By Assumption A.1 (a), it is not difficult to check the following

$$\max_{1\le i\le n}\sup_{a\in\mathcal{I}}|\ddot{\ell}_a(Y_i;a)| = O_P(\log n). \tag{S.11}$$

By (S.11), we can let $C > C_0$ be sufficiently large so that the event $B_{n1} = \{\max_{1\le i\le n}\sup_{a\in\mathcal{I}}|\ddot{\ell}_a(Y_i;a)| \le C\log n\}$ has large probability. Denote $g = \widehat{g}_{n,\lambda} - g_0$. By Assumption A.3, the event $B_{n2} = \{\|g\| \le r_n \equiv M((nh)^{-1/2}+h^m)\}$ has large probability with some pre-selected large M. So $B_n = B_{n1}\cap B_{n2}$ has a large probability. Define $\widetilde{g} = d_n^{-1}g$, where $d_n = c_m r_n h^{-1/2}$. Since $h = o(1)$ and $nh^2 \to \infty$, $d_n = o(1)$. Then by Lemma 3.1, on B_n, $\|\widetilde{g}\|_{\sup} \le 1$. Note that $J(\widetilde{g},\widetilde{g}) = d_n^{-2}\lambda^{-1}(\lambda J(g,g)) \le d_n^{-2}\lambda^{-1}\|g\|^2 \le d_n^{-2}\lambda^{-1}r_n^2 \le c_m^{-2}h\lambda^{-1}$. Thus, when the event B_n holds, $\widetilde{g}$ is an element in $\mathcal{G}$.

Define $\psi(T;g) = \dot{\ell}_a(Y;g(Z)+g_0(Z)) - \dot{\ell}_a(Y;g_0(Z))$. By the definition of S_n and a direct calculation, one can verify that $S_n(g+g_0) - S(g+g_0) - (S_n(g_0) - S(g_0)) = \frac{1}{n}\sum_{i=1}^n[\psi(T_i;g)K_{Z_i} - E\{\psi(T;g)K_Z\}]$. Let $\widetilde{\psi}_n(T;\widetilde{g}) = C^{-1}c_m^{-1}(\log n)^{-1}h^{1/2}d_n^{-1}\psi(T;d_n\widetilde{g})$ and $\psi_n(T_i;\widetilde{g}) = \widetilde{\psi}_n(T_i;\widetilde{g})I_{A_i}$, where $A_i = \{\sup_{a\in\mathcal{I}}|\ddot{\ell}_a(Y_i;a)| \le C\log n\}$ for $i = 1,\ldots,n$. Observe that B_n implies $\cap_i A_i$.

Next we show that ψ_n satisfies (3.2). For any g_1, $g_2 \in \mathcal{G}$, and $z\in\mathbb{I}$, both $g_0(z)+d_n g_1(z)$ and $g_0(z)+d_n g_2(z)$ fall in $\mathcal{I}$ when n is sufficiently large since $g_0(z)\in\mathcal{I}_0$ and $d_n = o(1)$. Recall that $\mathcal{I}_0$ and $\mathcal{I}$ are specified in Assumption A.1. Therefore,

$$\begin{aligned}
|\psi_n(T_i;d_n g_1) - \psi_n(T_i;d_n g_2)| &= C^{-1}c_m^{-1}(\log n)^{-1}h^{1/2}d_n^{-1}|\psi(T_i;g_1) - \psi(T_i;g_2)|\cdot I_{A_i}\\
&= C^{-1}c_m^{-1}(\log n)^{-1}h^{1/2}d_n^{-1}\Big|\int_{g_0(Z_i)}^{g_0(Z_i)+d_n g_1(Z_i)}\ddot{\ell}_a(Y_i;a)\cdot I_{A_i}da\\
&\quad -\int_{g_0(Z_i)}^{g_0(Z_i)+d_n g_1(Z_i)}\ddot{\ell}_a(Y_i;a)\cdot I_{A_i}da\Big|\\
&\le C^{-1}c_m^{-1}(\log n)^{-1}h^{1/2}d_n^{-1}\cdot d_n\|g_1-g_2\|_{\sup}\cdot\sup_{a\in\mathcal{I}}|\ddot{\ell}_a(Y_i;a)|\cdot I_{A_i}\\
&\le C^{-1}c_m^{-1}(\log n)^{-1}h^{1/2}d_n^{-1}\cdot d_n\cdot C\log n\cdot\|g_1-g_2\|_{\sup}\\
&= c_m^{-1}h^{1/2}\|g_1-g_2\|_{\sup}.
\end{aligned}$$

Thus, ψ_n satisfies (3.2). By Lemma 3.2, with large probability

$$\|\sum_{i=1}^n[\psi_n(T_i;\widetilde{g})K_{Z_i} - E\{\psi_n(T;\widetilde{g})K_Z\}]\| \le (n^{1/2}h^{-(2m-1)/(4m)}+1)(5\log\log n)^{1/2}. \tag{S.12}$$

On the other hand, by Chebyshev's inequality

$$P(A_i^c) = \exp(-(C/C_0)\log n)E\{\exp(\sup_{a\in\mathcal{I}}|\ddot{\ell}_a(Y_i;a)|/C_0)\} \le C_1 n^{-C/C_0}.$$

Since $h = o(1)$ and $nh^2 \to \infty$, we may choose C to be large enough so that $2^{1/2}C^{-1}C_0C_1(\log n)^{-1}n^{-C/(2C_0)} < a_n' h^{1/2} d_n^{-1}$, where $a_n' = n^{-1/2}((nh)^{-1/2} + h^m)h^{-(6m-1)/(4m)}(\log\log n)^{1/2}$. By (2.3), which implies $E\{\sup_{a\in\mathcal{I}}|\ddot{\ell}_a(Y_i;a)|\big|Z_i\} \le 2C_1C_0^2$, we have $E\{|\psi(T;d_n\widetilde{g})|^2\} \le 2C_1C_0^2d_n^2$ on B_n, where the expectation is taken with respect to $T=(Y,Z)$. So when n is large, by Chebyshev's inequality, we have on B_n,

$$\begin{aligned}
\|E\{\psi_n(T_i;\widetilde{g})K_{Z_i}\} - E\{\widetilde{\psi}_n(T_i;\widetilde{g})K_{Z_i}\}\| &= \|E\{\widetilde{\psi}_n(T_i;\widetilde{g})K_{Z_i}\cdot I_{A_i^c}\}\| \\
&\le C^{-1}(\log n)^{-1}d_n^{-1}\left(E\{|\psi(T_i;d_n\widetilde{g})|^2\}\right)^{1/2}P(A_i^c)^{1/2} \\
&\le 2^{1/2}C^{-1}C_0C_1(\log n)^{-1}n^{-C/(2C_0)} \\
&\le a_n' h^{1/2} d_n^{-1},
\end{aligned}$$

where the expectation is taken with respect to T_i. Therefore, by (S.12) and on B_n, we have

$$\begin{aligned}
&\|S_n(g+g_0) - S(g+g_0) - (S_n(g_0) - S(g_0))\| \\
=\ & \frac{Cc_m(\log n)h^{-1/2}d_n}{n}\|\sum_{i=1}^n[\widetilde{\psi}_n(T_i;\widetilde{g})K_{Z_i} - E\{\widetilde{\psi}_n(T;\widetilde{g})K_Z\}]\| \\
\le\ & \frac{Cc_m(\log n)h^{-1/2}d_n}{n}(\|\sum_{i=1}^n[\psi_n(T_i;\widetilde{g})K_{Z_i} - E\{\psi_n(T;\widetilde{g})K_Z\}]\| \\
&+n\|E\{\psi_n(T_i;\widetilde{g})K_{Z_i}\} - E\{\widetilde{\psi}_n(T_i;\widetilde{g})K_{Z_i}\}\|) \\
\le\ & \frac{Cc_m(\log n)h^{-1/2}d_n}{n}\cdot[(n^{1/2}h^{-(2m-1)/(4m)}+1)(5\log\log n)^{1/2} + na_n'h^{1/2}d_n^{-1}] \\
\le\ & C'c_m a_n' \log n, \qquad \text{(S.13)}
\end{aligned}$$

for some constant $C' > 0$ that only depends on C, c_m, M.

By Taylor's expansion, the fact $S_{n,\lambda}(g+g_0) = 0$, and Proposition 2.3, we have

$$\begin{aligned}
&\|S_n(g+g_0) - S(g+g_0) - (S_n(g_0) - S(g_0))\| \\
=\ & \|S_{n,\lambda}(g+g_0) - S_\lambda(g+g_0) - S_{n,\lambda}(g_0) + S_\lambda(g_0)\| \\
=\ & \|S_\lambda(g+g_0) + S_{n,\lambda}(g_0) - S_\lambda(g_0)\| \\
=\ & \|DS_\lambda(g_0)g + \int_0^1\int_0^1 sD^2S_\lambda(g_0+ss'g)gg\,ds\,ds' + S_{n,\lambda}(g_0)\| \\
=\ & \|-g + \int_0^1\int_0^1 sD^2S_\lambda(g_0+ss'g)gg\,ds\,ds' + S_{n,\lambda}(g_0)\| \\
\ge\ & \|-g + S_{n,\lambda}(g_0)\| - \|\int_0^1\int_0^1 sD^2S_\lambda(g_0+ss'g)gg\,ds\,ds'\|.
\end{aligned}$$

Therefore,

$$
\begin{aligned}
& \|g - S_{n,\lambda}(g_0)\| \\
\leq\ & \|S_n(g+g_0) - S(g+g_0) - (S_n(g_0) - S(g_0))\| + \|\int_0^1 \int_0^1 sD^2 S_\lambda(g_0 + ss'g)ggdsds'\| \\
\leq\ & \|S_n(g+g_0) - S(g+g_0) - (S_n(g_0) - S(g_0))\| + \int_0^1 \int_0^1 s\|D^2 S_\lambda(g_0 + ss'g)gg\|dsds'. \qquad \text{(S.14)}
\end{aligned}
$$

Next we find an upper bound for $\|D^2 S_\lambda(g_0 + ss'g)gg\|$. The Fréchet derivative of DS_λ is found to be $D^2 S_\lambda = D^2 S$ Therefore, $D^2 S_\lambda(g_0+ss'g)gg = D^2 S(g_0+ss'g)gg = E\{\ell_a'''(Y;(g_0+ss'g)(Z))g(Z)^2 K_Z\}$, where the expectation is taken with respect to T. Hence, by (2.4) and on B_n, we have

$$
\begin{aligned}
\|D^2 S_\lambda(g_0 + ss'g)gg\| = \|E\{\ell_a'''(Y;(g_0+ss'g)(Z))g(Z)^2 K_Z\}\| &\leq E\{E\{\sup_{a\in\mathcal{I}} |\ell_a'''(Y;a)|\big|Z\}g(Z)^2\|K_Z\|\} \\
&\leq C_\ell c_m h^{-1/2}\|g\|^2, \qquad \text{(S.15)}
\end{aligned}
$$

where $C_\ell = \sup_{z\in\mathbb{I}} E\{\sup_{a\in\mathcal{I}} |\ell_a'''(Y;a)|\big|Z = z\}$. Thus, from (S.13), (S.14) and (S.15), with large probability, $\|g - S_{n,\lambda}(g_0)\| \leq C' c_m a_n' \log n + C_\ell c_m h^{-1/2}((nh)^{-1/2} + h^m)^2$. This completes the proof of Theorem 3.4.

S.8. *Proof of Theorem 3.5.* Define $Rem_n = \widehat{g}_{n,\lambda} - g_0^* - \frac{1}{n}\sum_{i=1}^n \epsilon_i K_{Z_i}$. By Theorem 3.4, Rem_n satisfies $\|Rem_n\| = O_P(a_n \log n)$. By assumption $a_n \log n = o(n^{-1/2})$, we have $\|Rem_n\| = o_P(n^{-1/2})$. Since $E\{\|\sum_{i=1}^n \epsilon_i K_{Z_i}\|^2\} = nE\{\epsilon^2\|K_Z\|^2\} = O(nh^{-1})$, we have $\|n^{-1}\sum_{i=1}^n \epsilon_i K_{Z_i}\| = O_P((nh)^{-1/2})$. Thus, Rem_n is negligible compared with $\sum_{i=1}^n \epsilon_i K_{Z_i}$.

Next we show the limiting distribution of $(nh)^{1/2}(\widehat{g}_{n,\lambda}(z_0) - g_0^*(z_0))$. Note that this is equal to $(nh)^{1/2}\langle K_{z_0}, \widehat{g}_{n,\lambda} - g_0^*\rangle$. Using the fact

$$
\begin{aligned}
|(nh)^{1/2}\langle K_{z_0}, \widehat{g}_{n,\lambda} - g_0^* - \frac{1}{n}\sum_{i=1}^n \epsilon_i K_{Z_i}\rangle| &\leq (nh)^{1/2}\|K_{z_0}\| \cdot \|Rem_n\| \\
&= O_P((nh)^{1/2}h^{-1/2}a_n \log n) = o_P(1),
\end{aligned}
$$

we just need to find the limiting distribution of $(nh)^{1/2}\langle K_{z_0}, \frac{1}{n}\sum_{i=1}^n \epsilon_i K_{Z_i}\rangle = (nh^{-1})^{-1/2}\sum_{i=1}^n \epsilon_i K_{Z_i}(z_0)$.

By Assumption A.1 (c), i.e., $E\{\epsilon^2|Z\} = I(Z)$, we have

$$
Var(\sum_{i=1}^n \epsilon_i K_{Z_i}(z_0)) = nE\{\epsilon^2|K_Z(z_0)|^2\} = nE\{E\{\epsilon^2|Z\}|K_Z(z_0)|^2\} = nE\{I(Z)|K_Z(z_0)|^2\} = nV(K_{z_0}, K_{z_0}).
$$

By the definition, as $h \to 0$, $hV(K_{z_0}, K_{z_0}) \to \sigma_{z_0}^2$. Thus, $(nh^{-1})^{-1/2}\sum_{i=1}^n \epsilon_i K_{Z_i}(z_0) \xrightarrow{d} N(0, \sigma_{z_0}^2)$ by CLT. The expression of $\sigma_{z_0}^2$, i.e., (3.8), follows from Proposition 2.1. This completes the proof.

S.9. *Proof of Corollary 3.7.* By Proposition 2.2, Assumption A.2 holds. We first show part (i). By $\ell_a'''(y;a) = 0$ for any y and a, i.e.,, in (3.5) $C_\ell = 0$, we obtain $a_n = n^{-1/2}((nh)^{-1/2} + h^m)h^{-(6m-1)/(4m)}(\log\log n)^{1/2}$. Since $h \asymp n^{-1/(4m+1)}$, we have $h = o(1)$ and $nh^2 \to \infty$. By $m > (3+\sqrt{5})/4$, it can be verified that $a_n \log n = o(n^{-1/2})$.

On the other hand, by expression of K in terms of h_ν (see Proposition 2.1), as $h \to 0$, we have

$$\begin{aligned}\int_0^1 g_0^{(2m)}(z)K(z_0,z)dz - g_0^{(2m)}(z_0)/\pi(z_0) &= \sum_\nu \frac{1}{1+\lambda\gamma_\nu}V(g_0^{(2m)}/\pi, h_\nu)h_\nu(z_0) - \sum_\nu V(g_0^{(2m)}/\pi, h_\nu)h_\nu(z_0)\\ &= -\sum_\nu \frac{\lambda\gamma_\nu}{1+\lambda\gamma_\nu}V(g_0^{(2m)}/\pi, h_\nu)h_\nu(z_0) \to 0, \qquad \text{(S.16)}\end{aligned}$$

where the limit in (S.16) follows from $\sum_\nu |V(g_0^{(2m)}, h_\nu)h_\nu(z_0)| < \infty$ and the dominated convergence theorem. Then, by (3.11) and integration by parts, it can be shown that

$$\begin{aligned}&(W_\lambda g_0)(z_0) = \langle W_\lambda g_0, K_{z_0}\rangle = \lambda J(g_0, K_{z_0})\\ &= (-1)^m h^{2m}\int_0^1 g_0^{(2m)}(z)K(z_0,z)dz = (-1)^m h^{2m}(g_0^{(2m)}(z_0)/\pi(z_0) + o(1)). \qquad \text{(S.17)}\end{aligned}$$

So, as $n \to \infty$, $(nh)^{1/2}(W_\lambda g_0)(z_0) \to (-1)^m g_0^{(2m)}(z_0)/\pi(z_0)$. Therefore all assumptions in Theorem 3.6 hold. Then (3.12) directly follows from (3.10).

The proof of (3.13) is similar to that of (3.12). One only notes, by (S.17) and $h \asymp n^{-d}$ for $\frac{1}{4m+1} < d \le \frac{2m}{8m-1}$, $(nh)^{1/2}(W_\lambda g_0)(z_0) = O((nh)^{1/2}h^{2m}) = o(1)$. Then (3.13) follows from (3.10).

The proof of part (ii) is similar in spirit to that of part (i). The only difference is that (S.17) should be replaced by the following, by integration by parts,

$$(W_\lambda g_0)(z_0) = h^{2m}\sum_{j=1}^m (-1)^{j-1}\left[\left(\frac{\partial^{m-j}}{\partial z^{m-j}}K_{z_0}^{m-j}(z)\right)\cdot g_0^{(m+j-1)}(z)\Big|_0^1\right] + (-1)^m h^{2m}\int_0^1 g_0^{(2m)}(z)K(z_0,z)dz \qquad \text{(S.18)}$$

since g_0 does not satisfy the boundary conditions. The first sum is $o(h^{2m})$ by (3.14). The second sum is $(-1)^m h^{2m}(g_0^{(2m)}(z_0)/\pi(z_0)+o(1))$ by (S.16). Thus, $(W_\lambda g_0)(z_0) = (-1)^m h^{2m} g_0^{(2m)}(z_0)/\pi(z_0) + o(h^{2m})$. Note this is not true for $z_0 = 0$ or 1. Then the proof can be finished by similar arguments in the proof of part (i).

S.10. *Proof of Theorem 4.3.* The proof is similar to those in Theorem 3.4, so we only sketch the basic idea. Let $g = \widehat{g}_{n,\lambda}^0 - g_0^0$. Assumption A.4 guarantees that with large probability, $\|g\| \le r_n \equiv M((nh)^{-1/2} + h^m)$ for a suitably large M. By modifying the proof of Lemma 3.2, we have the following lemma.

Lemma S.3. *Suppose that ψ_n satisfies the following Lipschitz continuity condition:*

$$|\psi_n(T;g_1)-\psi_n(T;g_2)| \leq c_m^{-1}h^{1/2}\|g_1-g_2\|_{\sup}, \text{ for all } g_1,g_2\in\mathcal{H}_0, \tag{S.19}$$

where $T=(Y,Z)$ denotes the full data variable. Then we have

$$\lim_{n\to\infty} P\left(\sup_{g\in\mathcal{G}_0}\frac{\|Z_n^0(g)\|}{n^{1/2}h^{-(2m-1)/(4m)}\|g\|_{\sup}^{1-1/(2m)}+1} \leq (5\log\log n)^{1/2}\right)=1,$$

where $\mathcal{G}_0=\{g\in\mathcal{H}_0 | \|g\|_{\sup}\leq 1, J(g,g)\leq c_m^{-2}h\lambda^{-1}\}$ *and* $Z_n^0(g)=\sum_{i=1}^n[\psi_n(T_i;g)K_{Z_i}^*-E\{\psi_n(T;g)K_Z^*\}]$.

By reexamining the proof of Theorem 3.4, we have $g\in\mathcal{G}_0$ and ψ_n satisfies Lipschitz continuity (S.19) with large probability, where $\psi_n(T;g)=C^{-1}c_m^{-1}(\log n)^{-1}h^{1/2}d_n^{-1}\{\dot{\ell}_a(Y;g_0(Z)+d_ng(Z))-\dot{\ell}_a(Y;g_0(Z))\}$ and $d_n=c_mr_nh^{-1/2}$. This leads to, with large probability,

$$\|\sum_{i=1}^n[\psi_n(T_i;g)K_{Z_i}^*-E\{\psi_n(T;g)K_Z^*\}]\| \leq (n^{1/2}h^{-(2m-1)/(4m)}+1)(5\log\log n)^{1/2}. \tag{S.20}$$

The remainder of the proof follows by (S.11), and by an argument similar to (S.13) – (S.15).

S.11. *Proof of Theorem 4.4.* For notational convenience, denote $\widehat{g}=\widehat{g}_{n,\lambda}$, $\widehat{g}^0=\widehat{g}^0_{n,\lambda}$, $g=w_0+\widehat{g}^0-\widehat{g}$. By Assumptions A.3 and A.4, with large probability, $\|g\|=O_P(r_n)$, where $r_n=M((nh)^{-1/2}+h^m)$ for some large M. By Assumption A.1 (a), for some large constant $C>0$, the event $B_{n1}\cap B_{n2}$ has large probability, where $B_{n1}=\{\max_{1\leq i\leq n}\sup_{a\in\mathcal{I}}|\ddot{\ell}_a(Y_i;a)|\leq C\log n\}$ and $B_{n2}=\{\max_{1\leq i\leq n}\sup_{a\in\mathcal{I}}|\ell_a'''(Y_i;a)|\leq C\log n\}$. Let a_n be defined as in (3.5).

By Taylor expansion,

$$\begin{aligned} LRT_{n,\lambda} &= \ell_{n,\lambda}(w_0+\widehat{g}^0)-\ell_{n,\lambda}(\widehat{g}) \\ &= S_{n,\lambda}(\widehat{g})g+\int_0^1\int_0^1 sDS_{n,\lambda}(\widehat{g}+ss'g)ggdsds' \\ &= \int_0^1\int_0^1 sDS_{n,\lambda}(\widehat{g}+ss'g)ggdsds' \\ &= \int_0^1\int_0^1 s\{DS_{n,\lambda}(\widehat{g}+ss'g)gg-DS_{n,\lambda}(g_0)gg\}dsds' \\ &\quad +\frac{1}{2}(DS_{n,\lambda}(g_0)gg-E\{DS_{n,\lambda}(g_0)gg\})+\frac{1}{2}E\{DS_{n,\lambda}(g_0)gg\}. \end{aligned} \tag{S.21}$$

Denote the above three terms by I_1, I_2 and I_3. Next we will study the asymptotic behavior of these terms. Denote $\widetilde{g}=\widehat{g}+ss'g-g_0$, for any $0\leq s,s'\leq 1$. So $\|\widetilde{g}\|=O_P(r_n)$.

We first study I_1. By calculations of the Fréchet derivatives, we have

$$DS_{n,\lambda}(\widehat{g}+ss'g)gg=DS_{n,\lambda}(\widetilde{g}+g_0)gg=\frac{1}{n}\sum_{i=1}^n\ddot{\ell}_a(Y_i;g_0(Z_i)+\widetilde{g}(Z_i))g(Z_i)^2-\langle W_\lambda g,g\rangle/2,$$

and $DS_{n,\lambda}(g_0)gg = \frac{1}{n}\sum_{i=1}^n \ddot{\ell}_a(Y_i; g_0(Z_i))g(Z_i)^2 - \langle W_\lambda g, g\rangle/2$. On $B_{n1}\cap B_{n2}$,

$$\begin{aligned} & |DS_{n,\lambda}(\widehat{g}+ss'g)gg - DS_{n,\lambda}(g_0)gg| \\ \leq\ & \frac{1}{n}C(\log n)\|\widetilde{g}\|_{\sup}\sum_{i=1}^n g(Z_i)^2 \\ =\ & C(\log n)\|\widetilde{g}\|_{\sup}\langle\frac{1}{n}\sum_{i=1}^n g(Z_i)K_{Z_i}, g\rangle \\ =\ & C(\log n)\|\widetilde{g}\|_{\sup}\langle\frac{1}{n}\sum_{i=1}^n g(Z_i)K_{Z_i} - E\{g(Z)K_Z\}, g\rangle + C(\log n)\|\widetilde{g}\|_{\sup}E\{g(Z)^2\}, \end{aligned} \tag{S.22}$$

where the expectations are taken with respect to Z. Now we examine $\frac{1}{n}\|\sum_{i=1}^n g(Z_i)K_{Z_i} - E\{g(Z)K_Z\}\|$. Let $d_n = c_m h^{-1/2} r_n$ and $\bar{g} = d_n^{-1}g$. Consider $\psi(T; g) = g(Z)$ and $\psi_n(T;\bar{g}) = c_m^{-1}h^{1/2}d_n^{-1}\psi(T; d_n\bar{g})$ (which satisfies (3.2)). Then by Lemma 3.2,

$$\left\|\frac{1}{n}\sum_{i=1}^n[g(Z_i)K_{Z_i} - E\{g(Z)K_Z\}]\right\| = \frac{c_m h^{-1/2}d_n}{n}\left\|\sum_{i=1}^n[\psi_n(T_i;\bar{g})K_{Z_i} - E\{\psi_n(T;\bar{g})K_Z\}]\right\| = O_P(a_n'), \tag{S.23}$$

where $a_n' = n^{-1/2}((nh)^{-1/2} + h^m)h^{-(6m-1)/(4m)}(\log\log n)^{1/2}$. Obviously, $E\{g(Z)^2\} = O(\|g\|^2) = O_P(r_n^2)$. So, by $a_n' = o(r_n)$, we have

$$\begin{aligned} |DS_{n,\lambda}(\widehat{g}+ss'g)gg - DS_{n,\lambda}(g_0)gg| &= \|\widetilde{g}\|_{\sup}(O_P(a_n' r_n\log n) + O_P(r_n^2\log n)) \\ &= h^{-1/2}r_n O_P(r_n^2\log n) \\ &= O_P(r_n^3 h^{-1/2}\log n). \end{aligned} \tag{S.24}$$

Thus, $|I_1| = O_P(r_n^3 h^{-1/2}\log n)$.

Next we study I_2. By an argument similar to (S.13), it can be shown that

$$\frac{1}{n}\|\sum_{i=1}^n \ddot{\ell}_a(Y_i; g_0(Z_i))g(Z_i)K_{Z_i} - E\{\ddot{\ell}_a(Y_i; g_0(Z))g(Z)K_Z\}\| = O_P(a_n'\log n). \tag{S.25}$$

Thus, $|I_2| = O_P(a_n' r_n\log n)$.

Note $I_3 = -\|g\|^2/2$. Therefore, combining the above approximations of I_1 and I_2, we have $-2n\cdot LRT_{n,\lambda} = n\|w_0 + \widehat{g}^0 - \widehat{g}\|^2 + O_P(nr_n a_n'\log n + nr_n^3 h^{-1/2}\log n) = n\|w_0 + \widehat{g}^0 - \widehat{g}\|^2 + O_P(nr_n a_n\log n + nr_n^3 h^{-1/2}\log n)$. By $r_n^2 h^{-1/2} = o(a_n)$ and $nr_n a_n = o((\log n)^{-1})$, it is easy to see that $O_P(nr_n a_n\log n + nr_n^3 h^{-1/2}\log n) = o_P(1)$. Thus, part (ii) holds. So, to find the limiting distribution of the LRT test, we only focus on $n\|w_0 + \widehat{g}^0 - \widehat{g}\|^2$. By Theorems 3.4 and 4.3,

$$n^{1/2}\|w_0 + \widehat{g}^0 - \widehat{g} - S_{n,\lambda}^0(g_0^0) + S_{n,\lambda}(g_0)\| = O_P(n^{1/2}a_n\log n) = o_P(1), \tag{S.26}$$

so we just have to focus on $n^{1/2}\{S^0_{n,\lambda}(g^0_0) - S_{n,\lambda}(g_0)\}$. Recall that

$$\begin{aligned} S^0_{n,\lambda}(g^0_0) &= \frac{1}{n}\sum_{i=1}^n \epsilon_i K^*_{Z_i} - W^*_\lambda g^0_0 \\ &= \frac{1}{n}\sum_{i=1}^n \epsilon_i (K_{Z_i} - K_{Z_i}(z_0)K_{z_0}/K(z_0,z_0)) - W_\lambda g_0 + (W_\lambda g_0)(z_0)K_{z_0}/K(z_0,z_0), \end{aligned}$$

where $\epsilon_i = \dot{\ell}_a(Y_i; g_0(Z_i))$, and $S_{n,\lambda}(g_0) = \frac{1}{n}\sum_{i=1}^n \epsilon_i K_{Z_i} - W_\lambda g_0$. Thus,

$$S^0_{n,\lambda}(g^0_0) - S_{n,\lambda}(g_0) = \left(-\frac{1}{n}\sum_{i=1}^n \epsilon_i K_{Z_i}(z_0) + (W_\lambda g_0)(z_0)\right) K_{z_0}/K(z_0,z_0). \tag{S.27}$$

So $n\|S^0_{n,\lambda}(g^0_0) - S_{n,\lambda}(g_0)\|^2 = |\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i K_{Z_i}(z_0)/\sqrt{K(z_0,z_0)} - \sqrt{n}(W_\lambda g_0)(z_0)/\sqrt{K(z_0,z_0)}|^2$. By central limit theorem, (4.10) and $\sqrt{n}(W_\lambda g_0)(z_0)/\sqrt{K(z_0,z_0)} \to -c_{z_0}$, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i K_{Z_i}(z_0)/\sqrt{K(z_0,z_0)} - \sqrt{n}(W_\lambda g_0)(z_0)/\sqrt{K(z_0,z_0)} \xrightarrow{d} N(c_{z_0}, c_0). \tag{S.28}$$

It follows by (S.26)–(S.28) that $-2n \cdot LRT_{n,\lambda} \xrightarrow{d} c_0\chi^2_1(c^2_{z_0}/c_0)$, the scaled noncentral χ^2 distribution with one degree of freedom and noncentrality parameter $c^2_{z_0}/c_0$, which shows (iii). It follows immediately that $\|w_0 + \widehat{g}^0 - \widehat{g}\| = O_P(n^{-1/2})$, i.e., part (i) holds. This completes the proof.

S.12. *Proof of Corollary 4.5.* By Fourier expansion of g_0 and $W_\lambda h_\nu = \frac{\lambda\gamma_\nu}{1+\lambda\gamma_\nu}$, we have $(W_\lambda g_0)(z_0) = \sum_\nu V(g_0, h_\nu)\frac{\lambda\gamma_\nu}{1+\lambda\gamma_\nu}h_\nu(z_0)$. By the assumption that $\sum_\nu |V(g_0,h_\nu)|^2\gamma_\nu^d < \infty$, one obtains the bound $|(W_\lambda g_0)(z_0)| = O((\lambda^d h^{-1})^{1/2}) = O(h^{md-1/2})$ by using Cauchy's inequality. Thus, by $h \asymp n^{-1/(2m+1)}$ and $d > 1 + 1/(2m)$, $(nh)^{1/2}(W_\lambda g_0)(z_0) = o(1)$. Direct calculations verify $h = o(1)$, $nh^2 \to \infty$, $a_n = o((nh)^{-1/2} + h^m)$, $a_n = o(n^{-1/2}(\log n)^{-1})$, $a_n = o(n^{-1}((nh)^{-1/2} + h^m)^{-1}(\log n)^{-1})$, and $a_n \gg ((nh)^{-1/2} + h^m)^2 h^{-1/2}$. Thus, the desired result follows from Theorem 4.3.

S.13. *Proof of Theorem 5.1.* By Theorem 3.4 and Lemma 3.1,

$$\|\widehat{g} - g^*_0 - \frac{1}{n}\sum_{i=1}^n \epsilon_i K_{Z_i}\|_{\sup} = O_P(a_n h^{-1/2}\log n). \tag{S.29}$$

So the key is to study the leading process $H_n(z) = n^{-1/2}\sum_{i=1}^n \epsilon_i K_{Z_i}(z)$.

Since $E\{\exp(|\epsilon|/C_1)|Z\} \le C_2$, a.s., we may fix a sufficiently large constant $C > (1+3\delta)C_1$ such that the event $E_n = \{\max_{1\le i\le n}|\epsilon_i| \le b_n = C\log n\}$ has large probability. Define $H^b_n(z) = n^{-1/2}\sum_{i=1}^n \epsilon_i I(|\epsilon_i| \le b_n)K_{Z_i}(z)$. Write $H_n(z) = H_n(z) - H^b_n(z) - E\{H_n(z) - H^b_n(z)\} + H^b_n(z) - E\{H^b_n(z)\}$. Obviously, $H_n(z) - H^b_n(z) = 0$ on E_n. By Chebyshev's inequality and Lemma 3.1, we have

$$\begin{aligned}|E\{H_n(z)-H_n^b(z)\}| &= n^{1/2}|E\{\epsilon I(|\epsilon|\geq b_n)K_Z(z))\}|\\ &\leq O(1)h^{-1/2}n^{1/2}E\{|\epsilon|\cdot I(|\epsilon|\geq b_n)\}\\ &\leq O(1)h^{-1/2}n^{1/2}E\{|\epsilon|^2\}^{1/2}P(|\epsilon|>b_n)^{1/2}\\ &= O(h^{-1/2}n^{1/2}\exp(-b_n/(2C_1))).\end{aligned}$$

Thus,

$$\sup_{z\in\mathbb{I}}|H_n(z)-H_n^b(z)-E\{H_n(z)-H_n^b(z)\}|=O_P(h^{-1/2}n^{1/2}\exp(-b_n/(2C_1))). \tag{S.30}$$

Denote $R_n(z)=H_n^b(z)-E\{H_n^b(z)\}$. Then by (S.30), we have

$$\sup_{z\in\mathbb{I}}|H_n(z)-R_n(z)|=O_P(h^{-1/2}n^{1/2}\exp(-b_n/(2C_1))). \tag{S.31}$$

Let $Z_n(\epsilon,z)=n^{1/2}(P_n(\epsilon,z)-P(\epsilon,z))$, where $P_n(\epsilon,z)$ and $P(\epsilon,z)$ are empirical and population distribution of (ϵ,Z). Then by Theorem 1 of [48], $\sup_{\epsilon\in\mathbb{R},z\in\mathbb{I}}|Z_n(\epsilon,z)-W(\tau(\epsilon,z))|=O_P(n^{-1/2}(\log n)^2)$, where W is Brownian bridge indexed by $[0,1]\times[0,1]$, $\tau(\epsilon,z)=(P_Z(z),P_{\epsilon|Z}(\epsilon|z))$, P_Z is the marginal distribution of Z, and $P_{\epsilon|Z}$ is the conditional distribution of ϵ given Z. Write

$$R_n^b(z)=\int_0^1\int_{-b_n}^{b_n}\epsilon K(z,t)dZ_n(\epsilon,t)=\int_0^1 K(z,t)dV_n(t),\quad\text{and}$$

$$R_n^0(z)=\int_0^1\int_{-b_n}^{b_n}\epsilon K(z,t)dW(\tau(\epsilon,t))=\int_0^1 K(z,t)dV_n^0(t),$$

where $V_n(t)=\int_{-b_n}^{b_n}\epsilon dZ_n(\epsilon,t)$ and $V_n^0(t)=\int_{-b_n}^{b_n}\epsilon dW(\tau(\epsilon,t))$. By integration by parts,

$$V_n(t)=Z_n(\epsilon,t)\epsilon\Big|_{-b_n}^{b_n}-\int_{-b_n}^{b_n}Z_n(\epsilon,t)d\epsilon,\quad\text{and}$$

$$V_n^0(t)=W(\tau(\epsilon,t))\epsilon\Big|_{-b_n}^{b_n}-\int_{-b_n}^{b_n}W(\tau(\epsilon,t))d\epsilon.$$

So $\sup_{t\in\mathbb{I}}|V_n(t)-V_n^0(t)|=O_P(b_nn^{-1/2}(\log n)^2)$.

By integration by parts again, we have

$$R_n(z)=V_n(t)K(z,t)\Big|_{t=0}^{1}-\int_0^1 V_n(t)\frac{d}{dt}K(z,t)dt,\quad\text{and}$$

$$R_n^0(z)=V_n^0(t)K(z,t)\Big|_{t=0}^{1}-\int_0^1 V_n^0(t)\frac{d}{dt}K(z,t)dt.$$

Therefore, by assumption $\sup_{z,t} |\frac{d}{dt}K(z,t)| = O(h^{-2})$, we have

$$\sup_{z\in\mathbb{I}} |R_n(z) - R_n^0(z)| = O_P(h^{-2}b_n n^{-1/2}(\log n)^2). \tag{S.32}$$

Write $W(t_1,t_2) = B(t_1,t_2) - t_1t_2B(1,1)$, where B is standard Brownian motion indexed on $[0,1]\times[0,1]$. Define $\bar{R}_n^0(z) = \int_0^1 K(z,t)d\bar{U}_n^0(t)$, where $\bar{U}_n^0(t) = \int_{-b_n}^{b_n} \epsilon dB(\tau(\epsilon,t))$. Direct calculations lead to $R_n^0(z) - \bar{R}_n^0(z) = B(1,1)\int_0^1 K(z,t)\int_{-b_n}^{b_n} \epsilon\, dP(\epsilon,t)$. Therefore, by Lemma 3.1 and the finite exponential moment of $|\epsilon|$, we have

$$\begin{aligned}
\sup_{z\in\mathbb{I}} |R_n^0(z) - \bar{R}_n^0(z)| &= |B(1,1)|\cdot\sup_{z\in\mathbb{I}}|\int_0^1 K(z,t)\int_{-b_n}^{b_n} \epsilon\; dP_{\epsilon|Z}(\epsilon|t)dP_Z(t)| \\
&= |B(1,1)|\cdot\sup_{z\in\mathbb{I}}|\int_0^1 K(z,t)E\{\epsilon I(|\epsilon|\le b_n)|Z=t\}dP_Z(t)| \\
&= |B(1,1)|\cdot\sup_{z\in\mathbb{I}}|\int_0^1 K(z,t)E\{\epsilon I(|\epsilon|> b_n)|Z=t\}dP_Z(t)| \\
&\le c_m^2 h^{-1}|B(1,1)|E\{|\epsilon|I(|\epsilon|>b_n)\} \\
&= O_P(h^{-1}\exp(-b_n/(2C_1))). \qquad \text{(S.33)}
\end{aligned}$$

Define $\tilde{R}_n^0(z) = \int_0^1 h^{-1}\omega((z-t)/h)d\bar{U}_n^0(t)$. Using integration by parts, we get $\bar{U}_n^0(t) = B(\tau(\epsilon,t))\epsilon\Big|_{\epsilon=-b_n}^{b_n} - \int_{-b_n}^{b_n} B(\tau(\epsilon,t))d\epsilon$, so we have $\sup_{t\in\mathbb{I}}|\bar{U}_n^0(t)| = O_P(b_n)$. Again, by integration by parts, we have $\bar{R}_n^0(z)-\tilde{R}_n^0(z) = \bar{U}_n^0(t)\left(h^{-1}\omega((z-t)/h) - K(z,t)\right)\Big|_{t=0}^{1} - \int_0^1 \bar{U}_n^0(t)\frac{d}{dt}\left(h^{-1}\omega((z-t)/h) - K(z,t)\right)dt$, which, by assumption (5.1), leads to

$$\sup_{h^\varphi\le z\le 1-h^\varphi} |\bar{R}_n^0(z) - \tilde{R}_n^0(z)| = O_P(h^{-2}b_n\exp(-C_2h^{-1+\varphi})). \tag{S.34}$$

By the proof of Lemma 3.7 in [20], the process $\tilde{R}_n^0(z)$ is Gaussian with mean zero and has the same distribution as the process $Y_z^{(n)} = h^{-1}\int_0^1 (I_n(t))^{1/2}\omega((z-t)/h)dW(t)$, where W is standard one-dimensional Brownian motion indexed on $\mathbb{R}$ and $I_n(z) = E\{\epsilon^2 I(|\epsilon|\le b_n)|Z=z\}$. Define $Y_{0,z}^{(n)} = h^{-1}\int_0^1 (I(t))^{1/2}\omega((z-t)/h)dW(t)$. Obviously, $\sup_{z\in\mathbb{I}}|I(z)-I_n(z)| = O(\exp(-b_n/(2C_1)))$. It follows from the assumption (5.5) and $E\{\exp(|\epsilon|/C_1)|Z\}\le C$, a.s., that

$$\begin{aligned}
\sup_{z\in\mathbb{I}}\left|\frac{d}{dz}(I(z)-I_n(z))\right| &= \sup_{z\in\mathbb{I}}\left|\int_{|\epsilon|>b_n}\epsilon^2\frac{d}{dz}\pi(\epsilon|z)d\epsilon\right| \\
&\le \sup_{z\in\mathbb{I}}\int_{|\epsilon|>b_n}\epsilon^2\rho_1(1+|\epsilon|^{\rho_2})\pi(\epsilon|z)d\epsilon \\
&= \sup_{z\in\mathbb{I}}\rho_1 E\{\epsilon^2(1+|\epsilon|^{\rho_2})I(|\epsilon|>b_n)|Z=z\} = O(\exp(-b_n/(2C_1))).
\end{aligned}$$

By (5.5) and trivial calculations, it can be shown that $\sup_{t\in\mathbb{I}}|\frac{d}{dt}I(t)|<\infty$. Since, when n is large, both I and I_n are bounded below from zero, we have

$$\begin{aligned}\left|\frac{d}{dt}\left(I(t)^{1/2}-I_n(t)^{1/2}\right)\right| &= (1/2)\left|\frac{I(t)'I_n(t)^{1/2}-I_n(t)'I(t)^{1/2}}{I(t)^{1/2}I_n(t)^{1/2}}\right| \\ &\leq (1/2)\frac{|I(t)'|\cdot|I(t)^{1/2}-I_n(t)^{1/2}|+I(t)^{1/2}|I(t)'-I_n(t)'|}{I(t)^{1/2}I_n(t)^{1/2}} \\ &= O(\exp(-b_n/(2C_1))),\end{aligned}$$

where for convenience we denote $I'(t)$ to be the derivative of $I(t)$. By integration by parts,

$$\begin{aligned}Y_{0,z}^{(n)}-Y_z^{(n)} &= h^{-1}W(t)(I(t)^{1/2}-I_n(t)^{1/2})\omega((z-t)/h)\Big|_{t=0}^{1} \\ &\quad -h^{-1}\int_0^1 W(t)\frac{d}{dt}\left((I(t)^{1/2}-I_n(t)^{1/2})\omega((z-t)/h)\right)dt \\ &= h^{-1}W(t)(I(t)^{1/2}-I_n(t)^{1/2})\omega((z-t)/h)\Big|_{t=0}^{1} \\ &\quad -h^{-1}\int_0^1 W(t)\frac{d}{dt}\left(I(t)^{1/2}-I_n(t)^{1/2}\right)\cdot\omega((z-t)/h) \\ &\quad +h^{-2}\int_0^1 W(t)\left(I(t)^{1/2}-I_n(t)^{1/2}\right)\cdot\omega'((z-t)/h)dt,\end{aligned}$$

for which we have

$$\sup_{z\in\mathbb{Z}}|Y_{0,z}^{(n)}-Y_z^{(n)}| = O_P(h^{-2}\exp(-b_n/(2C_1))). \tag{S.35}$$

Next we define $\bar{Y}_{0,z}^{(n)} = h^{-1}I(z)^{1/2}\int_0^1\omega((z-t)/h)dW(t)$. Then we have

$$\begin{aligned}Y_{0,z}^{(n)}-\bar{Y}_{0,z}^{(n)} &= h^{-1}\int_0^1(I(t)^{1/2}-I(z)^{1/2})\omega((z-t)/h)dW(t) \\ &= h^{-1}\int_{z/h}^{(z-1)/h}(I(z-sh)^{1/2}-I(z)^{1/2})\omega(s)dW(z-sh) \\ &= h^{-1}W(z-sh)(I(z-sh)^{1/2}-I(z)^{1/2})\omega(s)\Big|_{s=z/h}^{(z-1)/h} \\ &\quad -h^{-1}\int_{z/h}^{(z-1)/h}W(z-sh)\frac{d}{ds}\left((I(z-sh)^{1/2}-I(z)^{1/2})\omega(s)\right)ds.\end{aligned}$$

Based on the facts that $|I(z-sh)^{1/2}-I(z)^{1/2}|\leq C_I|s|h$ for some positive constant C_I and any $z,s\in\mathbb{I}$, that $|\omega(s)|\leq C_\omega\exp(-|s|/C_3)$ implying $|\omega(z/h)|\leq C_\omega\exp(-h^{\varphi-1}/C_3)=O(h)$ and $|\omega((z-1)/h)|\leq C_\omega\exp(-h^{\varphi-1}/C_3)=O(h)$ for $h^\varphi\leq z\leq 1-h^\varphi$, and that ω' is bounded, it can be verified that

$$\sup_{h^\varphi\leq z\leq 1-h^\varphi}|Y_{0,z}^{(n)}-\bar{Y}_{0,z}^{(n)}| = O_P(1). \tag{S.36}$$

The last random process we will consider is $L_z^{(n)} = h^{-1}I(z)^{1/2}\int_{\mathbb{R}}\omega((z-t)/h)dW(t)$. We will establish the rate of convergence for $\sup_{h^\varphi\le z\le 1-h^\varphi}|L_z^{(n)} - \bar{Y}_{0,z}^{(n)}|$. For this purpose, we need the following result.

Lemma S.4. *For any $\kappa > 1/2$, $\lim_{d\to\infty} P\left(\sup_{s\in\mathbb{R}}\frac{|W(s)|}{(1+|s|)^\kappa} > d\right) = 0$.*

Proof of Lemma S.4. Let $D_\kappa = \sup_{s>0}\frac{|W(s)|}{(1+s)^\kappa}$. We will only show $\lim_{d\to\infty}P(D_\kappa > d) = 0$. The proof for $\sup_{s\le 0}\frac{|W(s)|}{(1+s)^\kappa}$ is similar. Let $\mathbb{Z}_+ = \{0,1,\ldots\}$ be the set of nonnegative integers. Note $\sup_{s>0}\frac{|W(s)|}{(1+s)^\kappa} = \sup_{m\in\mathbb{Z}_+}\sup_{m<s\le m+1}\frac{|W(s)|}{(1+s)^\kappa}$. Choose a constant $\beta > 0$ such that $(\beta+1)(\kappa-1/2) > 1$. Then

$$\begin{aligned}
P(D_\kappa > d) &= P\left(\sup_{m\in\mathbb{Z}_+}\sup_{m<s\le m+1}\frac{|W(s)|}{(1+s)^\kappa} > d\right)\\
&\le \sum_{m=0}^{\infty} P\left(\sup_{m<s\le m+1}\frac{|W(s)|}{(1+s)^\kappa} > d\right)\\
&\le \sum_{m=0}^{\infty} P\left(\sup_{m<s\le m+1}|W(s)| > (1+m)^\kappa d\right)\\
&\le \sum_{m=0}^{\infty} P\left(\sup_{0<s\le m+1}|W(s)| > (1+m)^\kappa d\right)\\
&\le \frac{4}{(2\pi)^{1/2}}\sum_{m=0}^{\infty}\frac{\exp(-(d(1+m)^{\kappa-1/2})^2/2)}{d(1+m)^{\kappa-1/2}} \qquad \text{(S.37)}\\
&\le \frac{4}{(2\pi)^{1/2}}\sum_{m=0}^{\infty}\frac{1}{(d(1+m)^{\kappa-1/2})^{\beta+1}} = O(d^{-(\beta+1)}),
\end{aligned}$$

where (S.37) follows by [26]. Therefore, the desired result holds. □

Now define $E_{n,1} = \left\{\sup_{s\in\mathbb{R}}\frac{|W(s)|}{(1+|s|)^\kappa} \le d\right\}$ for some fixed $d > 0$ so that $E_{n,1}$ has large probability. By integration by parts and a straightforward calculation, we have

$$\begin{aligned}
L_z^{(n)} - \bar{Y}_{0,z}^{(n)} &= h^{-1}I(z)^{1/2}\left(\int_{-\infty}^0 \omega((z-t)/h)dW(t) + \int_1^\infty \omega((z-t)/h)dW(t)\right)\\
&= h^{-1}I(z)^{1/2}\left(W(t)\omega((z-t)/h)\Big|_{t=-\infty}^0 - \int_{-\infty}^0 W(t)\omega'((z-t)/h)h^{-1}dt\right)\\
&\quad + h^{-1}I(z)^{1/2}\left(W(t)\omega((z-t)/h)\Big|_{t=1}^\infty - \int_1^\infty W(t)\omega'((z-t)/h)h^{-1}dt\right).
\end{aligned}$$

On $E_{1,n}$, for any z, we have $|W(z)| \le d(1+|z|)^\kappa$. By assumption (5.1), $|\omega((z-t)/h))| \le C_\omega\exp(-|z-t|/(hC_3))$, $|\omega'((z-t)/h))| \le C_\omega\exp(-|z-t|/(hC_3))$. Thus we have, for any fixed z, as $|t|\to\infty$

$$|W(t)\omega((z-t)/h)| \le dC_\omega(1+|t|)^\kappa\exp(-|z-t|/(hC_3)) \longrightarrow 0.$$

Meanwhile, on $E_{1,n}$, we have $|W(1)\omega((z-1)/h)| \le 2d\exp(-h^{\varphi-1}/C_3)$, and

$$\begin{aligned}\left|\int_1^\infty W(t)\omega'((z-t)/h)dt\right| &\le \int_1^\infty d(1+t)^\kappa \cdot C_\omega \exp(-|z-t|/(C_3h))dt\\ &= \int_1^\infty d(1+t)^\kappa \cdot C_\omega \exp(-(t-z)/(C_3h))dt\\ &= \int_{1-z}^\infty d(1+t+z)^\kappa \cdot C_\omega \exp(-t/(C_3h))dt\\ &\le \int_{h^\varphi}^\infty d(2+t)^\kappa \cdot C_\omega \exp(-t/(C_3h))dt\\ &= h^\varphi \int_1^\infty d(2+h^\varphi t)^\kappa \cdot C_\omega \exp(-t/(C_3h^{1-\varphi}))dt\\ &\le h^\varphi \int_1^\infty d(2+t)^\kappa \cdot C_\omega (t/(C_3h^{1-\varphi}))^{-a}dt\\ &= C_3^a d C_\omega h^{\varphi+a(1-\varphi)} \int_1^\infty (2+t)^\kappa t^{-a}dt = O(h^{\varphi+a(1-\varphi)}) = O(h^3),\end{aligned}$$

where a is constant with $a > \kappa+2$ and $\varphi + a(1-\varphi) > 3$. Using a similar technique, one can show that on $E_{1,n}$, $\left|\int_{-\infty}^0 W(t)\omega'((z-t)/h)dt\right| \le O(h\exp(-h^{\varphi-1}/C_3))$. Consequently,

$$\sup_{h^\varphi \le z \le 1-h^\varphi} |L_z^{(n)} - \bar{Y}_{0,z}^{(n)}| = O_P(h). \tag{S.38}$$

Since $h^{1/2}L_z^{(n)}I(z)^{-1/2}/\sigma_\omega = h^{-1/2}\int \omega((t-z)/h)dW(t)/\sigma_\omega$ is stationary Gaussian with mean zero, the process $h^{1/2}L_{hz}^{(n)}I(hz)^{-1/2}/\sigma_\omega$ is Gaussian with mean zero and covariance function $\int_{-\infty}^\infty \omega(t)\omega(t+\cdot)dt/\sigma_\omega^2$. Then by [4], we have as $n\to\infty$,

$$\begin{aligned}&P\left((2\delta\log n)^{1/2}\left\{\sup_{h^\varphi\le z\le 1-h^\varphi}|h^{1/2}L_z^{(n)}I(z)^{-1/2}\sigma_\omega^{-1}| - d_n\right\}\le u\right)\\ =\ &P\left((2\delta\log n)^{1/2}\left\{\sup_{0\le z\le 1-2h^\varphi}|h^{1/2}L_z^{(n)}I(z)^{-1/2}\sigma_\omega^{-1}| - d_n\right\}\le u\right)\\ =\ &P\left((2\delta\log n)^{1/2}\left\{\sup_{0\le z\le h^{-1}(1-2h^\varphi)}|h^{1/2}L_{hz}^{(n)}I(hz)^{-1/2}\sigma_\omega^{-1}| - d_n\right\}\le u\right)\\ \to\ &\exp(-\exp(-2u)),\end{aligned} \tag{S.39}$$

where $\sigma_\omega = (\int_{\mathbb{R}}\omega(u)^2du)^{1/2}$. By assumption $C > (3\delta+1)C_1$, $m > (3+\sqrt{5})/4$ and $0 < \delta < 2m/(8m-1)$, the remainders in (S.29), (S.31)–(S.38) are all $o_P((h\log n)^{-1/2})$. Thus the desired conclusion holds.

S.14. *Proof of Proposition 5.2.* We first consider (5.2) and (5.3). (5.2) trivially holds. By boundedness and absolute integrability of ω, for any $\rho\in(0,2]$, $\lim_{|z|\to\infty}\int_{-\infty}^\infty \frac{\omega(t)(\omega(t+z)-\omega(t))}{|z|^\rho}dt = 0$, implying C_ρ in (5.3) is actually zero.

For general m, let $\tilde{h}_\nu$ and $\tilde{\gamma}_\nu$ be the normalized (with respect to the usual L_2-norm) eigenfunctions and eigenvalues of the boundary value problem $(-1)^m\tilde{h}_\nu^{(2m)} = \tilde{\gamma}_\nu\tilde{h}_\nu$, $\widetilde{h}_\nu^{(j)}(0) = \widetilde{h}_\nu^{(j)}(1) = 0$, $j = m, m+1, \ldots, 2m-1$. Thus, it is easy to see that $h_\nu = \sigma\widetilde{h}_\nu$ and $\gamma_\nu = \sigma^2\widetilde{\gamma}_\nu$ satisfy (2.11) with $\pi(z)I(z) \equiv \sigma^{-2}$, implying that h_ν and γ_ν form an effective eigensystem in $\mathcal{H}$. Let $\lambda^\dagger = \sigma^2\lambda$ and $h^\dagger = \sigma^{1/m}h$. Define $\widetilde{K}(s,t) = \sum_\nu \frac{\widetilde{h}_\nu(s)\widetilde{h}_\nu(t)}{1+\lambda^\dagger\widetilde{\gamma}_\nu}$. Then $\widetilde{K}$ is the reproducing kernel function associated with the inner product $\langle f, g\rangle_1 = \int_0^1 f(t)g(t)dt + \lambda^\dagger \int_0^1 f^{(m)}(t)g^{(m)}(t)dt$. Thus, $\widetilde{K}$ is the Green's function associated with the differential equation (2.1) in [37], with the penalty parameter therein replaced by $\lambda^\dagger$.

Next we restrict $m = 2$. By Theorem 4.1 in [34], for $j = 0, 1$, we have

$$\text{(S.40)} \qquad \sup_{s,t\in\mathbb{I}} \left| \frac{d^j}{dt^j}\left(\widetilde{K}(s,t) - \bar{K}(s,t)\right)\right| \leq C_K'(h^\dagger)^{-(j+1)}\exp(-\sin(\pi/(2m))/h^\dagger),$$

where by equation (6) in [34], $\bar{K}$ satisfies for any $s, t \in \mathbb{I}$ and $j = 0, 1$,

$$\text{(S.41)} \quad \left|\frac{d^j}{dt^j}\left(\bar{K}(s,t) - \frac{1}{h^\dagger}\omega_0\left(\frac{s-t}{h^\dagger}\right)\right)\right| \leq C_K''(h^\dagger)^{-(j+1)}(\exp(-|1-s|/(\sqrt{2}h^\dagger)) + \exp(-|s|/(\sqrt{2}h^\dagger))),$$

with C_K', C_K'' both being positive constants. By (S.40) and (S.41), it is easy to see that for any $s, t \in \mathbb{I}$ and $j = 0, 1$,

$$\text{(S.42)} \quad \begin{aligned} &\left|\frac{d^j}{dt^j}\left(\widetilde{K}(s,t) - \frac{1}{h^\dagger}\omega_0\left(\frac{s-t}{h^\dagger}\right)\right)\right| \\ \leq\ & C_K(h^\dagger)^{-(j+1)}(\exp(-\sin(\pi/(2m))/h^\dagger) + \exp(-|1-s|/(\sqrt{2}h^\dagger)) + \exp(-|s|/(\sqrt{2}h^\dagger))), \end{aligned}$$

where C', C_K are positive constant. By Proposition 2.1, $K(s,t) = \sum_\nu \frac{h_\nu(s)h_\nu(t)}{1+\lambda\gamma_\nu} = \sigma^2\widetilde{K}(s,t)$. Therefore, $K(s,t) - h^{-1}\omega((s-t)/h) = \sigma^2(\widetilde{K}(s,t) - (h^\dagger)^{-1}\omega_0((s-t)/h^\dagger))$. It can thus be shown that, by (S.42), Condition (5.1) holds.

S.15. *Proof of Theorem 5.3.* For simplicity, denote $\widehat{g} = \widehat{g}_{n,\lambda}$ and $g = \widehat{g} - g_0$. Using arguments similar to (S.21), (S.22) and (S.25), and by assumption $a_n = o(r_n)$, $nr_na_n\log n = o(h^{-1/2})$, $nr_n^3h^{-1/2}\log n = o(nr_na_n\log n) = o(h^{-1/2})$, it can be shown that

$$\text{(S.43)} \quad -2n\cdot PLRT_{n,\lambda} = n\|\widehat{g}-g_0\|^2 + O_P(nr_na_n\log n + nr_n^3h^{-1/2}\log n) = n\|\widehat{g}-g_0\|^2 + o_P(h^{-1/2}).$$

Under the hypothesis H_0^{global} that g_0 is the "true" parameter, by Theorem 3.4, we have $\|\widehat{g} - g_0 - S_{n,\lambda}(g_0)\| = O_P(a_n\log n)$, where a_n is defined as in (3.5). It thus follows from $n^{1/2}a_n\log n = o(1)$ that $n^{1/2}\|\widehat{g} - g_0\| = n^{1/2}\|S_{n,\lambda}(g_0)\| + o_P(1)$.

Next we study the leading term $\|S_{n,\lambda}(g_0)\|$. We first approximate $\|W_\lambda g_0\|$. By Proposition 2.1 and the dominated convergence theorem, it can be established that

$$\text{(S.44)} \qquad \|W_\lambda g_0\|^2 = o(\lambda).$$

To see (S.44), define $f_\lambda(\nu) = |V(g_0, h_\nu)|^2\gamma_\nu\frac{\lambda\gamma_\nu}{1+\lambda\gamma_\nu}$, for $\nu = 0, 1, \ldots,$, $\lambda > 0$. Then f_λ is a sequence of functions satisfying $|f_\lambda(\nu)| \leq |V(g_0, h_\nu)|^2\gamma_\nu \equiv f(\nu)$. From $g_0 \in \mathcal{H}$, we have $\sum_{\nu\in\mathbb{N}} |V(g_0, h_\nu)|^2\gamma_\nu = \int_{\mathbb{N}} f(\nu)dm(\nu) < \infty$, where $m(\cdot)$ denotes the discrete measure over $\mathbb{N}$. So $f(\nu)$ is an integrable function over $\mathbb{N}$ which dominates $f_\lambda(\nu)$. Since $\lim_{\lambda\to 0} f_\lambda(\nu) = 0$, we have $\sum_\nu |V(g_0, h_\nu)|^2\frac{\lambda\gamma_\nu^2}{1+\lambda\gamma_\nu} = \int_{\mathbb{N}} f_\lambda(\nu)dm(\nu) \to 0$ based on the Lebesgue dominated convergence theorem. That is, $\|W_\lambda g_0\|^2 = \sum_\nu |V(g_0, h_\nu)|^2\frac{\lambda^2\gamma_\nu^2}{1+\lambda\gamma_\nu} = o(\lambda)$.

By (2.12), $n\|S_{n,\lambda}(g_0)\|^2 = n^{-1}\|\sum_{i=1}^n \epsilon_i K_{Z_i}\|^2 - 2\sum_{i=1}^n \epsilon_i(W_\lambda g_0)(Z_i) + n\|W_\lambda g_0\|^2$. It follows by the Fourier expansion of g_0 and Proposition 2.1 that

$$\begin{aligned} & E\left\{|\sum_{i=1}^n \epsilon_i(W_\lambda g_0)(Z_i)|^2\right\} \\ = \quad & nE\{\epsilon^2|(W_\lambda g_0)(Z)|^2\} = nV(W_\lambda g_0, W_\lambda g_0) = n\sum_\nu |V(g_0, h_\nu)|^2\left(\frac{\lambda\gamma_\nu}{1+\lambda\gamma_\nu}\right)^2 = o(n\lambda), \end{aligned}$$

where the last equality follows by $\sum_\nu |V(g_0, h_\nu)|^2\gamma_\nu < \infty$ and the dominated convergence theorem; see (S.44) for similar arguments examining $\|W_\lambda g_0\|$. So $\sum_{i=1}^n \epsilon_i(W_\lambda g_0)(Z_i) = o_P((n\lambda)^{1/2}) = o_P(h^{-1/2})$. Thus, $n\|W_\lambda g_0\|^2 = o(n\lambda)$. Consequently, $n\|S_{n,\lambda}(g_0)\|^2 = n^{-1}\|\sum_{i=1}^n \epsilon_i K_{Z_i}\|^2 + n\|W_\lambda g_0\|^2 + o_P(h^{-1/2}) = n^{-1}\|\sum_{i=1}^n \epsilon_i K_{Z_i}\|^2 + o(n\lambda) + o_P(h^{-1/2})$. In what follows, we study the limiting property of $n^{-1}\|\sum_{i=1}^n \epsilon_i K_{Z_i}\|^2$.

Write $n^{-1}\|\sum_{i=1}^n \epsilon_i K_{Z_i}\|^2 = n^{-1}\sum_{i=1}^n \epsilon_i^2 K(Z_i, Z_i) + n^{-1}W(n)$, where $W(n) = \sum_{i\neq j}\epsilon_i\epsilon_j K(Z_i, Z_j)$. If we denote $W_{ij} = 2\epsilon_i\epsilon_j K(Z_i, Z_j)$, then we can rewrite $W(n)$ as $\sum_{1\leq i<j\leq n} W_{ij}$ so that $W(n)$ is *clean* (see [12]). Next we will derive the limiting distribution for $W(n)$. Let $\sigma(n)^2 = Var(W(n))$, and G_I, G_{II}, G_{IV} be defined as

$$G_I = \sum_{i<j} E\{W_{ij}^4\},$$

$$G_{II} = \sum_{i<j<k} (E\{W_{ij}^2 W_{ik}^2\} + E\{W_{ji}^2 W_{jk}^2\} + E\{W_{ki}^2 W_{kj}^2\}), \quad \text{and}$$

$$G_{IV} = \sum_{i<j<k<l} (E\{W_{ij}W_{ik}W_{lj}W_{lk}\} + E\{W_{ij}W_{il}W_{kj}W_{kl}\} + E\{W_{ik}W_{il}W_{jk}W_{jl}\}).$$

It follows from Proposition 3.2 of [12] that, to show $\sigma(n)^{-1}W(n) \xrightarrow{d} N(0,1)$, it is sufficient to show that G_I, G_{II}, G_{IV} are of lower order than $\sigma(n)^4$. By assumption $E\{\epsilon^4|Z\} \leq C$, a.s., we have $E\{\epsilon^4|Z\} \leq C \leq CC_2 I(Z)$, a.s. It then follows from (S.5) that $E\{W_{ij}^4\} = 16E\{\epsilon_i^4\epsilon_j^4 K(Z_i, Z_j)^4\} = O(h^{-4})$, which implies $G_I = O(n^2h^{-4})$. Obviously, $E\{W_{ij}^2 W_{ik}^2\} \leq E\{W_{ij}^4\} = O(h^{-4})$, implying $G_{II} = O(n^3h^{-4})$.

To approximate G_{IV}, for pairwise different i,j,k,l, we have

$$\begin{aligned} E\{W_{ij}W_{ik}W_{lj}W_{lk}\} &= 16E\{\epsilon_i^2\epsilon_j^2\epsilon_l^2\epsilon_k^2K(Z_i,Z_j)K(Z_i,Z_k)K(Z_l,Z_j)K(Z_l,Z_k)\} \\ &= \sum_\nu \frac{1}{(1+\lambda\gamma_\nu)^4} = O(h^{-1}) \end{aligned}$$

based on the direct examination. Therefore, $G_{IV} = O(n^4h^{-1})$.

Next we obtain the exact order of $\sigma(n)^4$, which is n^4h^{-2}. This follows from the observation $E\{W_{ij}^2\} = 4E\{\epsilon_i^2\epsilon_j^2K(Z_i,Z_j)^2\} = 4h^{-1}\rho_K^2$. Thus, $\sigma(n)^4 = \left(\binom{n}{2}E\{W_{ij}^2\}\right)^2$ has the same order as $4n^4h^{-2}\rho_K^4$. It follows by $h=o(1)$ and $(nh^2)^{-1}=o(1)$ that G_I, G_{II} and G_{IV} are of lower order than $\sigma(n)^4$, which implies by Proposition 3.2 of [12] that

$$\frac{1}{\sqrt{2h^{-1}}n\rho_K}W(n) \xrightarrow{d} N(0,1). \tag{S.45}$$

To conclude, we approximate the term $\sum_{i=1}^n \epsilon_i^2K(Z_i,Z_i)$. By $E\{\epsilon^4|Z\}\le C$, a.s., we have $E\{\epsilon^4K(Z,Z)^2\} = O(h^{-2})$. Therefore, a direct calculation leads to $E\{|\sum_{i=1}^n[\epsilon_i^2K(Z_i,Z_i)-h^{-1}\sigma_K^2]|^2\}\le nE\{\epsilon^4K(Z,Z)^2\} = O(nh^{-2})$, where $\sigma_K^2 = hE\{\epsilon^2K(Z,Z)\}$. This implies $\sum_{i=1}^n[\epsilon_i^2K(Z_i,Z_i)-h^{-1}\sigma_K^2] = O_P(n^{1/2}h^{-1})$. Therefore,

$$n^{-1}\sum_{i=1}^n\epsilon_i^2K(Z_i,Z_i) = h^{-1}\sigma_K^2 + O_P(n^{-1/2}h^{-1}) = h^{-1}\sigma_K^2 + O_P(1). \tag{S.46}$$

From (S.45) and (S.46), $(h/n)\|\sum_{i=1}^n\epsilon_iK_{Z_i}\|^2 = \sigma_K^2 + o_P(1)$, which implies $n\|S_{n,\lambda}(g_0)\|^2 = O_P(h^{-1} + n\lambda + h^{-1/2}) = O_P(h^{-1})$, and hence $n^{1/2}\|S_{n,\lambda}(g_0)\| = O_P(h^{-1/2})$. Thus,

$$\begin{aligned} -2n\cdot PLRT_{n,\lambda} &= n\|\widehat{g}-g_0\|^2 + o_P(h^{-1/2}) \\ &= \left(n^{1/2}\|S_{n,\lambda}(g_0)\| + o_P(1)\right)^2 + o_P(h^{-1/2}) \\ &= n\|S_{n,\lambda}(g_0)\|^2 + 2n^{1/2}\|S_{n,\lambda}(g_0)\|\cdot o_P(1) + o_P(h^{-1/2}) \\ &= n^{-1}\|\sum_{i=1}^n\epsilon_iK_{Z_i}\|^2 + n\|W_\lambda g_0\|^2 + o_P(h^{-1/2}). \end{aligned} \tag{S.47}$$

It follows by (S.45)–(S.47) and Slutsky's theorem that $(2h^{-1}\sigma_K^4/\rho_K^2)^{-1/2}(-2nr_K\cdot PLRT_{n,\lambda} - nr_K\|W_\lambda g_0\|^2 - h^{-1}\sigma_K^4/\rho_K^2) \xrightarrow{d} N(0,1)$.

S.16. *Proof of Theorem 5.4*. First of all, by direct calculations, one can verify by $\frac{1}{2m+1}\le d < \frac{2m}{8m-1}$ and $m > \frac{3+\sqrt{5}}{4}$ that $h \asymp n^{-d}$ satisfies the conditions in Theorem 5.3.

Next we prove our theorem. We write

$$-2n\cdot PLRT_{n,\lambda} = -2n(\ell_{n,\lambda}(g_0) - \ell_{n,\lambda}(g_{n0})) - 2n(\ell_{n,\lambda}(g_{n0}) - \ell_{n,\lambda}(\widehat{g}_{n,\lambda})). \tag{S.48}$$

The proof proceeds in two parts. Firstly. We note that $-2n \cdot PLRT' \equiv -2n(\ell_{n,\lambda}(g_{n0}) - \ell_{n,\lambda}(\widehat{g}_{n,\lambda}))$ is actually the PLRT test for testing H_{1n} against H_1^{global}. Under H_{1n}, $-2n \cdot PLRT'$ has the same asymptotic distribution as in Theorem 5.3, but uniformly for all $g_n \in \mathcal{G}_a$. That is to say, $(2u_n)^{-1/2}(-2nr_K \cdot PLRT' - n\|W_\lambda g_{n0}\|^2 - u_n) = O_P(1)$ uniformly for $g_n \in \mathcal{G}_a$, where $u_n = h^{-1}\sigma_K^4/\rho_K^2$ with σ_K^2 and ρ_K^2 given in (5.14). Secondly, we show that $-2n(\ell_{n,\lambda}(g_0) - \ell_{n,\lambda}(g_{n0})) = n\|g_n\|^2 + O_P(n^{1/2}\|g_n\| + n^{1/2}\|g_n\|^2 + n\lambda)$. Then $(2u_n)^{-1/2}(-2nr_K \cdot PLRT - u_n) \geq n(2u_n)^{-1/2}\|g_n\|^2(1 + O_P(n^{-1/2}\|g_n\|^{-1} + n^{-1/2} + \lambda\|g_n\|^{-2})) + (2u_n)^{-1/2}n\|W_\lambda g_{n0}\|^2 + O_P(1) \geq n(2u_n)^{-1/2}\|g_n\|^2(1 + O_P(n^{-1/2}\|g_n\|^{-1} + n^{-1/2} + \lambda\|g_n\|^{-2})) + O_P(1)$, where $O_P(\cdot)$ holds uniformly for $g_n \in \mathcal{G}_a$. Let $n^{-1/2}\|g_n\|^{-1} \leq 1/C$, $\lambda\|g_n\|^{-2} \leq 1/C$ and $\|g_n\|^2 \geq C(nh^{1/2})$ for sufficiently large C, which implies that $|\frac{-2nr_K \cdot PLRT - u_n}{(2u_n)^{1/2}}| \geq c_\alpha$ with large probability, where c_α is the cutoff value (based on $N(0,1)$) for rejecting H_0^{global} at level α. This means that we have to assume $\|g_n\|^2 \geq C(\lambda + (nh^{1/2})^{-1})$ to achieve large power.

Next we complete the above two parts. Firstly, it can be established that the following "uniform" FBR holds, i.e., for any $\delta \in (0,1)$, there exist positive constants $\widetilde{C}$ and N such that

$$\inf_{n \geq N} \inf_{g_n \in \mathcal{G}_a} P_{g_{n0}}\left(\|\widehat{g}_{n,\lambda} - g_{n0} - S_{n,\lambda}(g_{n0})\| \leq \widetilde{C}a_n\right) \geq 1 - \delta, \tag{S.49}$$

where a_n is defined in (3.5), The proof of (S.49) follows by a careful reexamination of Theorem 3.4. Specifically, one can choose C and M (to be unrelated to $g_n \in \mathcal{G}_a$) to be large so that the event $B_{n1} \cap B_{n2}$, which is defined in the proof of Theorem 3.4, has probability greater than $1 - \frac{\delta}{4}$. Then by going through exactly the same proof, it can be shown that when $n \geq N$ for some suitably selected N, (S.13) holds with probability greater than $1 - \delta/2$ for any $g_n \in \mathcal{G}_a$ (by properly tuning the probability), with the constant C' therein only depending on C, M, c_m. By going through the proof of (S.14) and (S.15), it can be shown that for $n \geq N$ and $g_n \in \mathcal{G}_a$, with probability larger than $1 - \delta$, $\|\widehat{g}_{n,\lambda} - g_{n0} - S_{n,\lambda}(g_{n0})\| \leq \widetilde{C}a_n$, where the constant $\widetilde{C}$ and N are unrelated to $g_n \in \mathcal{G}_a$. Using (S.49) and by exactly the same proof of Theorem 5.3, it can be shown that $-2n \cdot PLRT'$ follows the same asymptotic normal distribution under $H_{1n} : g = g_{n0}$ as in Theorem 5.3, uniformly for $g_n \in \mathcal{G}_a$.

For notational simplicity, denote $R_i = \ell(Y_i; g_0(Z_i)) - \ell(Y_i; g_{n0}(Z_i))$ for $i = 1, \ldots, n$. Secondly, we have

$$E\{|\sum_{i=1}^n [R_i - E(R_i)]|^2\} \leq nE\{R_i^2\} = nE\{|\epsilon_i g_n(Z_i) + g_n(Z_i)^2|^2\} = O(n\|g_n\|^2 + n\|g_n\|^4).$$

Therefore, uniformly over $g_n \in \mathcal{G}_a$, $n(\ell_{n,\lambda}(g_0) - \ell_{n,\lambda}(g_{n0}) - E\{\ell_{n,\lambda}(g_0) - \ell_{n,\lambda}(g_{n0})\}) = O_P(n^{1/2}\|g_n\| + n^{1/2}\|g_n\|^2)$. On the other hand, $E\{DS_{n,\lambda}(g_{n0})g_n g_n\} = -E\{|g_n(Z)|^2\} - \lambda J(g_n, g_n) = -\|g_n\|^2$. Therefore,

$$E\{\ell_{n,\lambda}(g_0)-\ell_{n,\lambda}(g_{n0})\}=E\{S_{n,\lambda}(g_{n0})(-g_n)+(1/2)DS_{n,\lambda}(g_{n0})g_ng_n\}=\lambda J(g_{n0},g_n)-\|g_n\|^2/2.$$

Since $|J(g_{n0},g_n)|\leq|J(g_0,g_n)|+J(g_n,g_n)\leq J(g_0,g_0)^{1/2}\zeta^{1/2}+\zeta$, we get that $2n(\ell_{n,\lambda}(g_0)-\ell_{n,\lambda}(g_{n0}))=-n\|g_n\|^2+O_P(n\lambda+n^{1/2}\|g_n\|+n^{1/2}\|g_n\|^2)$ uniformly for $g_n\in\mathcal{G}_a$. This completes the proof.

S.17. *Minimax rate of the PLRT test in general modeling framework.* In this section, we remark that PLRT achieves the optimal minimax rate of hypothesis testing specified in [23] in a more general modeling framework. The proofs are similar to those of Theorem 5.4 but require a deeper technical tool, i.e., the mapping principle, which builds equivalence between the eigenvalues obtained under null and contiguous alternatives. Write the local alternative as $H_{1n}: g=g_{n0}$, where $g_{n0}=g_0+g_n$, $g_0\in\mathcal{H}$ and g_n belongs to some alternative value set $\mathcal{G}_a$ below.

THEOREM S.2. *Let $m>(3+\sqrt{5})/4\approx 1.309$ and $h\asymp n^{-d}$ for $\frac{1}{2m+1}\leq d<\frac{2m}{8m-1}$. Let Assumption A.1 (a) hold for constants C_0, C_1, a compact interval $\mathcal{I}_0$ and an open interval $\mathcal{I}$ with $\mathcal{I}_0\subset\mathcal{I}$. Assume that there is a constant $C_2>0$ such that $1/C_2\leq-\ddot{\ell}_a(Y;a)\leq C_2$ holds for any $a\in\mathcal{I}$. The value of $2g_0$ belongs to $\mathcal{I}_0$. Consider the alternative value set*

$$\mathcal{G}_a=\{g\in H^m(\mathbb{I})|2g(z)\in\mathcal{I}_0 \text{ for any } z\in\mathbb{I},\ \|g\|_{\sup}\leq\zeta,\ J(g,g)\leq M\},$$

where $\zeta=1/(2C_0C_1C_2)$ and M is a positive constant. Suppose, under $H_{1n}: g=g_{n0}$ for $g_n\in\mathcal{G}_a$, Assumptions A.1 (c) and A.2 hold (with g_0 therein replaced by g_{n0}), $E\{\epsilon_{n0}^4|Z\}\leq C$, a.s., for some constant $C>0$ with $\epsilon_{n0}=\dot{\ell}_a(Y;g_{n0}(Z))$, and $\|\widehat{g}_{n,\lambda}-g_{n0}\|=O_P(r_n)$ holds under $H_{1n}: g=g_{n0}$ uniformly over $g_n\in\mathcal{G}_a$. Then for any $\delta\in(0,1)$, there exist positive constants C' and N such that

$$\inf_{n\geq N}\inf_{\substack{g_n\in\mathcal{G}_a\\ \|g_n\|\geq C'\eta_n}} P\left(\textit{reject } H_0^{global}|H_{1n} \textit{ is true}\right)\geq 1-\delta, \tag{S.50}$$

where $\eta_n\geq\sqrt{h^{2m}+(nh^{1/2})^{-1}}$. The minimal lower bound of η_n, i.e., $n^{-2m/(4m+1)}$, is achieved when $h=h^{**}\equiv n^{-2/(4m+1)}$.

PROOF OF THEOREM S.2. First of all, by direct calculations, one can verify by $\frac{1}{2m+1}\leq d<\frac{2m}{8m-1}$ and $m>\frac{3+\sqrt{5}}{4}$ that $h\asymp n^{-d}$ satisfies the conditions in Theorem 5.3. Throughout, we only consider $g_{n0}=g_0+g_n$ for $g_n\in\mathcal{G}_a$.

Next we prove our theorem. We write

$$-2n\cdot PLRT_{n,\lambda}=-2n(\ell_{n,\lambda}(g_0)-\ell_{n,\lambda}(g_{n0}))-2n(\ell_{n,\lambda}(g_{n0})-\ell_{n,\lambda}(\widehat{g}_{n,\lambda})). \tag{S.51}$$

The proof proceeds in two parts. We first note that $-2n \cdot PLRT' \equiv -2n(\ell_{n,\lambda}(g_{n0}) - \ell_{n,\lambda}(\widehat{g}_{n,\lambda}))$ is actually the PLRT test for testing H_{1n} against H_1^{global}. Under H_{1n}, $-2n \cdot PLRT'$ has the same asymptotic distribution as described in Theorem 5.3, but uniformly for all $g_n \in \mathcal{G}_a$. That is to say, $(2u_{n0})^{-1/2}(-2n \cdot PLRT'_{n,\lambda} - n\|W_\lambda g_{n0}\|^2 - h^{-1}\sigma^2_{Kn0}) = O_P(1)$ uniformly for $g_{n0} = g_0 + g_n$ with $g_n \in \mathcal{G}_a$, where $u_{n0} = h^{-1}\sigma^4_{Kn0}/\rho^2_{Kn0}$ under $g = g_{n0}$ and σ^2_{Kn0}, ρ^2_{Kn0} are given in (5.14) with eigenvalues therein derived under $g = g_{n0}$. Denote $u_n = h^{-1}\sigma^4_K/\rho^2_K$ under $g = g_0$ with σ^2_K, ρ^2_K given in (5.14). Let $V_{g_{n0}}$ and V_{g_0} be the V functionals defined as in Section 2.2 under $g = g_{n0}$ and $g = g_0$ respectively. Then, for any $f \in \mathcal{H}$, by Assumption A.1 (a) and (b), we have

$$\begin{aligned}|V_{g_{n0}}(f,f) - V_{g_0}(f,f)| &= |E\{[\ddot{\ell}_a(Y; g_{n0}(Z)) - \ddot{\ell}_a(Y; g_0(Z))]|f(Z)|^2\}| \\ &\leq E\{\sup_{a\in\mathcal{I}} |\ell'''_a(Y;a)| \cdot |g_n(Z)| \cdot |f(Z)|^2\} \\ &\leq C_0C_1C_2\|g_n\|_{\sup} V_{g_{n0}}(f,f) = \zeta_0\|g_n\|_{\sup} V_{g_{n0}}(f,f),\end{aligned}$$

where $\zeta_0 = C_0C_1C_2 = 1/(2\zeta)$ is a universal constant. Therefore, $(1 - \zeta_0\|g_n\|_{\sup})V_{g_{n0}}(f,f) \leq V_{g_0}(f,f) \leq (1 + \zeta_0\|g_n\|_{\sup})V_{g_{n0}}(f,f)$. By the mapping principle (see Theorem 6.1 in [54]), the eigenvalues induced by the functional pairs $(V_{g_{n0}}, J)$ and (V_{g_0}, J) are thus equivalent in the sense that $(1-\zeta_0\|g_n\|_{\sup})\gamma^{n0}_\nu \leq \gamma_\nu \leq (1+\zeta_0\|g_n\|_{\sup})\gamma^{n0}_\nu$ for any $\nu \in \mathbb{N}$, where γ^{n0}_ν denotes the eigenvalue corresponding to $V_{g_{n0}}$ and γ_ν is the eigenvalue corresponding to V_{g_0}. Therefore, uniformly for g_{n0},

$$\sigma^2_{Kn0} - \sigma^2_K = \sum_\nu \frac{h\lambda(\gamma_\nu - \gamma^{n0}_\nu)}{(1+\lambda\gamma^{n0}_\nu)(1+\lambda\gamma_\nu)} = O(\|g_n\|_{\sup}) = O(h^{-1/2}\|g_n\|).$$

Secondly, we show that $-2n(\ell_{n,\lambda}(g_0) - \ell_{n,\lambda}(g_{n0})) \geq nC'\|g_n\|^2 + O_P(n^{1/2}\|g_n\| + n\lambda)$, where C' is some positive constant unrelated to f. Then

$$\begin{aligned}&(2u_n)^{-1/2}(-2nr_K \cdot PLRT - u_n) \\ =\ & r_K(2u_n)^{-1/2}(-2n \cdot PLRT'_{n,\lambda} - n\|W_\lambda g_{n0}\|^2 - h^{-1}\sigma^2_{Kn0}) + r_K(2u_n)^{-1/2}n\|W_\lambda g_{n0}\|^2 \\ & -r_K(2u_n)^{-1/2} \cdot 2n(\ell_{n,\lambda}(g_0) - \ell_{n,\lambda}(g_{n0})) + r_K(2u_n)^{-1/2}h^{-1}(\sigma^2_{Kn0} - \sigma^2_K) \\ \geq\ & O_P(1) + nC'r_K(2u_n)^{-1/2}\|g_n\|^2(1 + O_P(n^{-1/2}\|g_n\|^{-1} + \lambda\|g_n\|^{-2})) + O(h^{-1}\|g_n\|),\end{aligned}$$

where $O_P(\cdot)$ holds uniformly for $g_n \in \mathcal{G}_a$. Let $n^{-1/2}\|g_n\|^{-1} \leq 1/C$, $\lambda\|g_n\|^{-2} \leq 1/C$, $Ch^{-1}\|g_n\| \leq nh^{1/2}\|g_n\|^2$, and $\|g_n\|^2 \geq C(nh^{1/2})^{-1}$ for sufficiently large C, which implies that $|\frac{-2nr_K \cdot PLRT - u_n}{(2u_n)^{1/2}}| \geq c_\alpha$ with large probability, where c_α is the cutoff value (based on $N(0,1)$) for rejecting H_0^{global} at nominal level α . This means we have to assume $\|g_n\|^2 \geq C(\lambda + (nh^{1/2})^{-1})$ to achieve large power.

Next we complete the above two parts. Firstly, it can be established that the following "uniform" FBR holds, i.e., for any $\delta \in (0,1)$, there exist positive constants $\widetilde{C}$ and N such that

$$\inf_{n\geq N} \inf_{g_n\in\mathcal{G}_a} P_{g_{n0}}\left(\|\widehat{g}_{n,\lambda} - g_{n0} - S_{n,\lambda}(g_{n0})\| \leq \widetilde{C}a_n\right) \geq 1 - \delta, \tag{S.52}$$

where a_n is defined in (3.5), The proof of (S.52) follows by a careful reexamination of Theorem 3.4. Specifically, one can choose C and M (to be unrelated to $g_n \in \mathcal{G}_a$) to be large so that the event $B_{n1} \cap B_{n2}$, which is defined in the proof of Theorem 3.4, has probability greater than $1 - \frac{\delta}{4}$. Then by going through exactly the same proof, it can be shown that when $n \geq N$ for some suitably selected N, for any $g_n \in \mathcal{G}_a$, (S.13) holds with probability greater than $1 - \delta/2$ (by properly tuning the probability), with the constant C' therein only depending on C, M, c_m. By going through the proofs of (S.14) and (S.15), it can be shown that for $n \geq N$ and $g_n \in \mathcal{G}_a$, with probability larger than $1 - \delta$, $\|\widehat{g}_{n,\lambda} - g_{n0} - S_{n,\lambda}(g_{n0})\| \leq \widetilde{C} a_n$, where the constant $\widetilde{C}$ and N are unrelated to $g_n \in \mathcal{G}_a$. Using (S.52) and by exactly the same proof of Theorem 5.3, it can be shown that $-2n \cdot PLRT'$ follows the same asymptotic normal distribution under $H_{1n} : g = g_{n0}$ as in Theorem 5.3, uniformly for $g_n \in \mathcal{G}_a$.

For simplicity, denote $R_i = \ell(Y_i; g_0(Z_i)) - \ell(Y_i; g_{n0}(Z_i))$ for $i = 1, \ldots, n$. Then

$$E\{|\sum_{i=1}^n [R_i - E(R_i)]|^2\} \leq nE\{R_i^2\} = nE\{| - \epsilon_i g_n(Z_i) + \ddot{\ell}_a(Y_i; g_{n0}^*(Z_i)) g_n(Z_i)^2|^2\}, \tag{S.53}$$

where $g_{n0}^*(z) = g_0(z) + t^* g_n(z)$ for $t^* \in (0,1)$, implying $g_{n0}^*(z) \in \mathcal{I}_0$ for any z. By Assumption A.1, we get that (S.53) is uniformly $O(n\|g_n\|^2)$ over $g_n \in \mathcal{G}_a$. Therefore, uniformly over $g_n \in \mathcal{G}_a$, $n(\ell_{n,\lambda}(g_0) - \ell_{n,\lambda}(g_{n0}) - E\{\ell_{n,\lambda}(g_0) - \ell_{n,\lambda}(g_{n0})\}) = O_P(n^{1/2}\|g_n\|)$.

On the other hand, by $\sup_{a \in \mathcal{I}} \ddot{\ell}_a(Y; a) < 0$, we can find $C' > 0$ (unrelated to $g_n \in \mathcal{G}_a$) such that $E\{DS_{n,\lambda}(g_{n0}^*) g_n g_n\} = E\{\ddot{\ell}_a(Y; g_{n0}^*(Z)) |g_n(Z)|^2\} - \lambda J(g_n, g_n) \leq -C'\|g_n\|^2/2$. Therefore,

$$\begin{aligned} E\{\ell_{n,\lambda}(g_0) - \ell_{n,\lambda}(g_{n0})\} &= E\{S_{n,\lambda}(g_{n0})(-g_n) + (1/2) DS_{n,\lambda}(g_{n0}^*) g_n g_n\} \\ &\leq \lambda J(g_{n0}, g_n) - C'\|g_n\|^2/2 = O(\lambda) - C'\|g_n\|^2/2, \end{aligned}$$

where the last equality holds by $J(g_n, g_n) \leq M$ and $|J(g_{n0}, g_n)| \leq |J(g_0, g_n)| + J(g_n, g_n) \leq J(g_0, g_0)^{1/2} M^{1/2} + M$. Consequently, $2n(\ell_{n,\lambda}(g_0) - \ell_{n,\lambda}(g_{n0})) \leq -nC'\|g_n\|^2 + O_P(n\lambda + n^{1/2}\|g_n\|)$. This completes the proof.

S.18. *Proof of Lemma 6.1.* We need the following two inequalities to establishing (6.3):

$$\int_0^\infty \frac{1}{(1 + x^{2m})^l} dx = \sum_{k=0}^\infty \int_{2\pi h^\dagger k}^{2\pi h^\dagger (k+1)} \frac{1}{(1 + x^{2m})^l} dx \leq \sum_{k=0}^\infty \frac{2\pi h^\dagger}{(1 + (2\pi h^\dagger k)^{2m})^l},$$

and by a similar argument, $\int_0^\infty \frac{1}{(1+x^{2m})^l} dx \geq \sum_{k=1}^\infty \frac{2\pi h^\dagger}{(1+(2\pi h^\dagger k)^{2m})^l}$. This completes the proof.

University of Notre Dame
Notre Dame, IN 46545
Email: zuofengshang@gmail.com

Department of Statistics
Purdue University
250 N. University Street
West Lafayette, IN 47906
Email: chengg@stat.purdue.edu

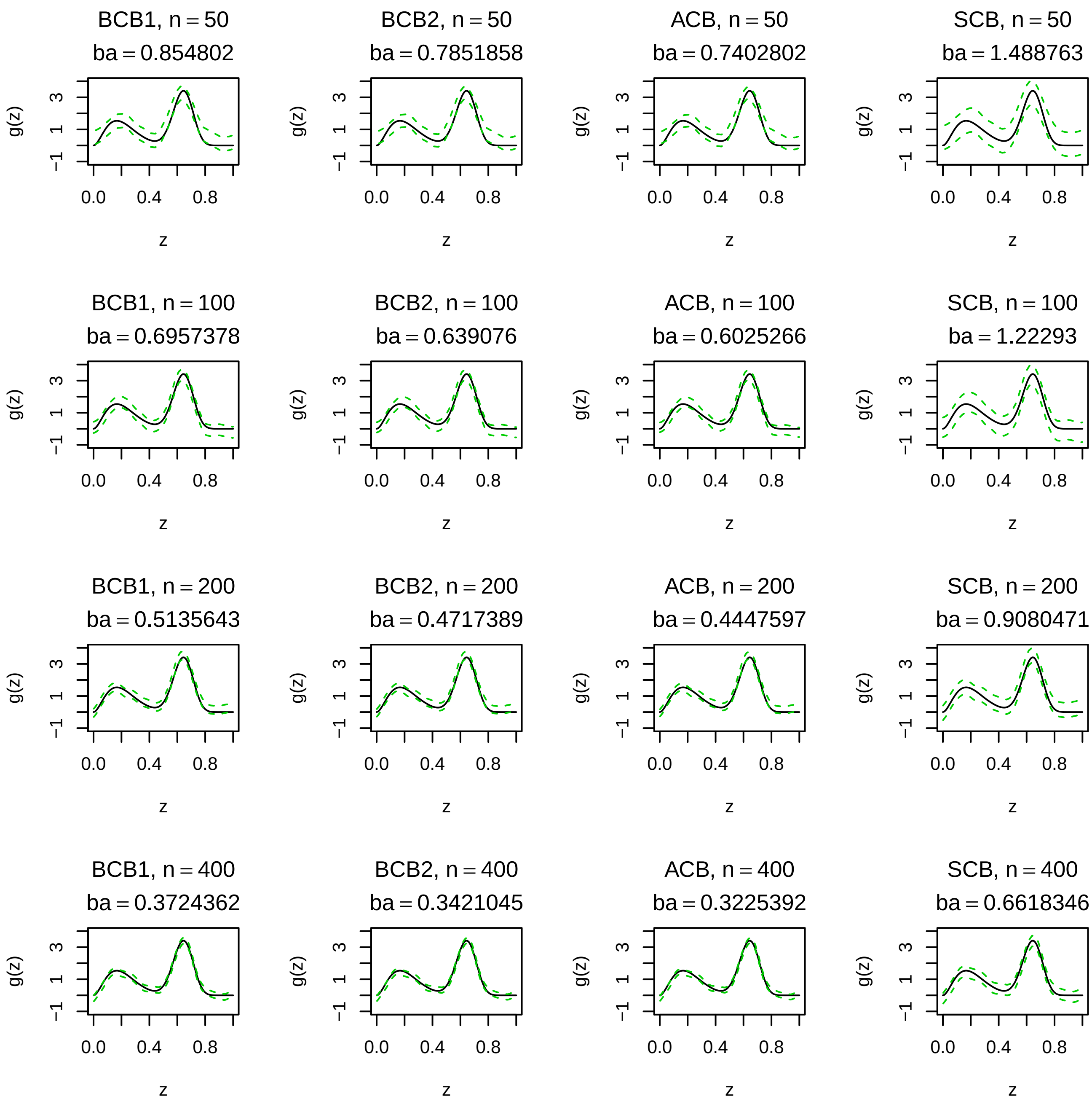

FIG 3. *95% pointwise and simultaneous confidence bands for periodic $g_0(z) = 0.6\beta_{30,17}(z) + 0.4\beta_{3,11}(z)$ in Case (I) of Example 6.1. The upper and lower bands are indicated by green curves, while the central black curve represents the true function. The numerical band area is denoted as "ba".*

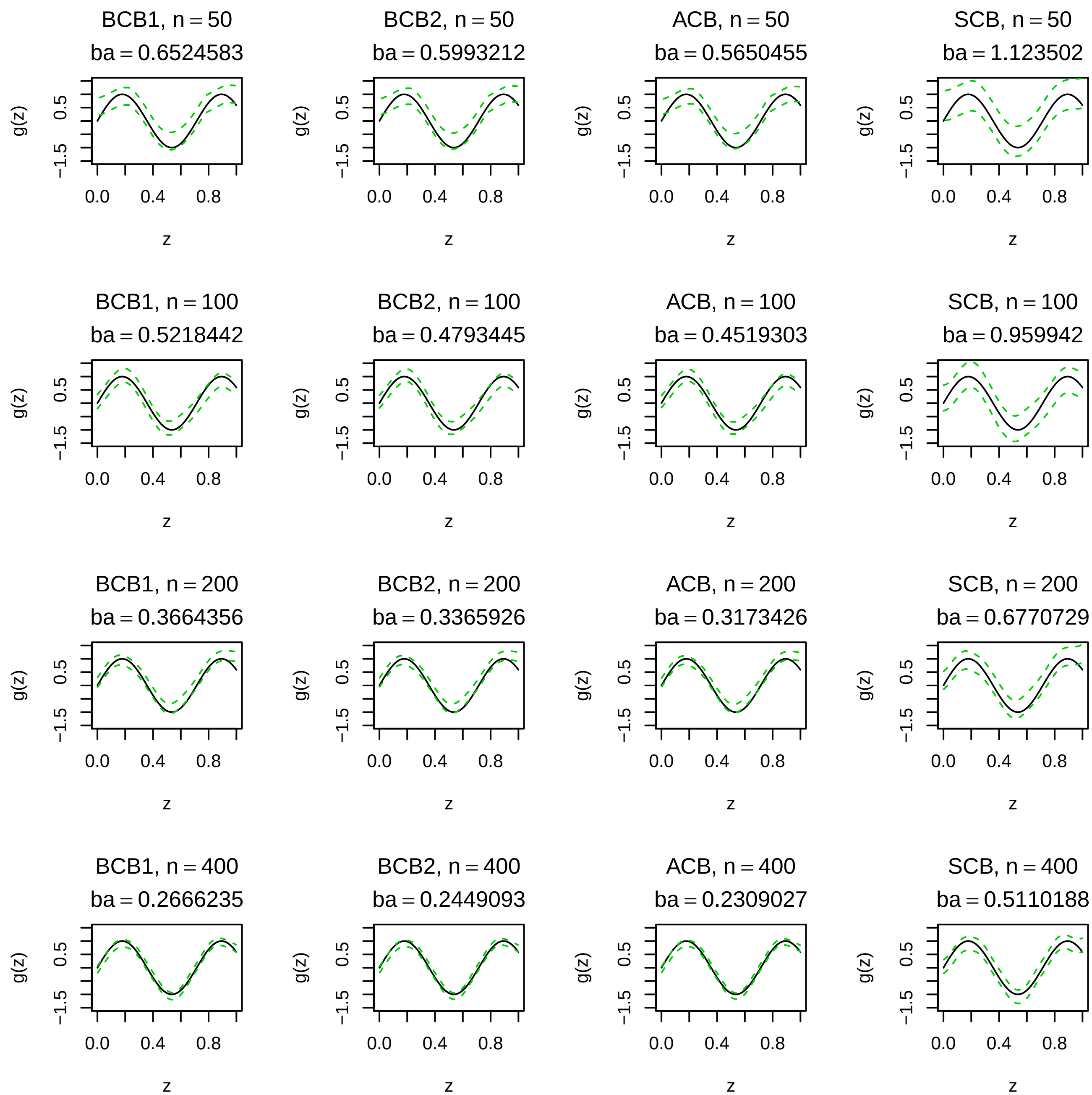

FIG 4. *95% pointwise and simultaneous confidence bands for nonperiodic* $g_0(z) = \sin(2.8\pi z)$ *in Case (II) of Example 6.1. The upper and lower bands are indicated by green curves, while the central black curve represents the true function. The numerical band area is denoted as "ba".*