Local and Global Asymptotic Inferences for the Smoothing Spline Estimate

By

Zuofeng Shang
University of Notre Dame

Guang Cheng
Purdue University

Technical Report #13-01

Department of Statistics
Purdue University

January 2013

LOCAL AND GLOBAL ASYMPTOTIC INFERENCES FOR THE SMOOTHING SPLINE ESTIMATE

BY ZUOFENG SHANG* AND GUANG CHENG †

University of Notre Dame and Purdue University

DECEMBER 27, 2012

This article presents the first comprehensive studies on the local and global inferences for the smoothing spline estimate in a unified asymptotic framework. The novel functional Bahadur representation is developed as the theoretical foundation of this article, and is also of independent interest. Based on that, we establish four interconnected inference procedures: (i) Point-wise Confidence Interval; (ii) Local Likelihood Ratio Testing; (iii) Simultaneous Confidence Band (SCB); (iv) Global Likelihood Ratio Testing. In particular, our C.I. is proven to be asymptotically valid at any point over the support, and is extraordinarily shorter than the classical Bayesian C.I. (Wahba, 1983). We also unveil new Wilk's phenomena arising from the local/global likelihood ratio testing, and further show that the global testing is more powerful/efficient than the local one in terms of the smaller minimum separation rate. It is also worthy noting that our SCB is the first one applicable to the general quasi-likelihood models. Furthermore, the inference optimality/efficiency issues are carefully addressed. As a by-product of this article, we discover some surprising asymptotic equivalence phenomenon between the periodic and non-periodic smoothing splines in terms of inferences.

1. Introduction. Smoothing spline models provide a very general framework for data analysis, modeling and learning in a variety of fields; see [57, 58, 21]. As far as we are aware, the existing literature are mostly concerned about the global convergence properties or methodological studies of smoothing spline estimate. Unfortunately, a systematic and rigorous theoretical study on their *asymptotic inferences* is almost nonexistent. This is partly due to the technical restrictions of the widely used equivalent kernel method. The novel Functional Bahadur Representation (FBR) we develop brings several major breakthroughs into the inference studies. The main purpose of this paper is to propose a series of local and global inference procedures for a univariate smooth curve based on FBR as the theoretical foundation. Moreover, we carefully investigate the inference optimality/efficiency that has not been well treated in the smoothing spline literature.

In this paper, we consider a general class of nonparametric regression models that covers the least square regression and logistic regression. The equivalent kernel method has long been used as a standard tool in dealing with the asymptotics of the smoothing splines, but it is only restricted to the simple least square regression; see [48, 38]. Moreover, this classical method only "approximates" the reproducing kernel function and the approximation formula becomes extremely complicated as

*Postdoctoral Fellow.

†Corresponding Author. Assistant Professor. Research Sponsored by NSF, DMS-0906497 and CAREER Award DMS-1151692

AMS 2000 subject classifications: Primary 62G20, 62F25; secondary 62F15, 62F12

Keywords and phrases: asymptotic normality, functional Bahadur representation, local/global likelihood ratio test, simultaneous confidence band, smoothing spline

the smoothness of regression function increases and the design points are not uniform. To analyze the smoothing spline estimates in a more feasible way, we develop a novel technical tool via the empirical processes techniques, i.e., *Functional Bahadur Representation*, which directly deals with the "exact" reproducing kernel, and thus makes the systematic inference studies possible in the general nonparametric regression models. Several new theoretical insights are also obtained through its applications. An immediate consequence of FBR is the local behaviors of the smoothing splines, i.e., asymptotic normality, which naturally leads to our construction of approximate point-wise confidence intervals. The classical Bayesian C.I. in the literature ([56, 40]) is only valid in an average sense over the observed covariates, and may not be reliable if only evaluated at peaks or troughs as pointed out by Nychka (1988). However, our *frequentist* C.I. is proven to be theoretically valid at any point, and even possesses the surprisingly shorter length. We next introduce the local likelihood ratio method in testing the value of the regression function at any point of interest. It is shown that the null limit distribution is a scaled Chi-square distribution with degree of freedom one, and its scaling constant converges to one as the regression function becomes more and more smooth. Therefore, we have unveiled an interesting Wilk's phenomenon arising from this nonparametric local testing, which injects new theoretical insight into the literature. The very tricky testing sensitivity issue has also been studied by characterizing its power behaviors under a sequence of local alternatives. One relevant work is for the *monotone* function but with rather different null limit distribution; see [3].

In practice, the global inferences are arguably more useful. The simultaneous confidence band depicts the global behaviors of the regression function with sufficient accuracy, and its construction has been extensively studied in the literature. However, most of the efforts were devoted to the simple regression models with either symmetric errors, i.e., the volume of tube method ([51]), or additive Gaussian errors based on the kernel or local polynomial estimates, e.g., [22, 10, 17, 60]. By incorporating the approach of [5] into the Reproducing Kernel Hilbert Space (RKHS) framework, we are able to construct the first SCB applicable to the general nonparametric regression models, and prove its theoretical validity based on the strong approximation techniques. We further demonstrate that the minimum bandwidth order of our SCB has achieved the lower bound established in Genovese and Wasserman (2008). Model assessment forms another crucial component of global inferences; see [23]. Fan et al (2001) explored the use of local polynomial estimate in testing nonparametric regression models by the Generalized Likelihood Ratio Testing (GLRT). Based on the smoothing spline estimate, we propose an alternative method called as the Penalized Likelihood Ratio Testing (PLRT), and prove its null limit distribution as the nearly Chi-square with diverging degree of freedom. Therefore, the Wilk's phenomenon previously established for the local testing continues to hold for the global one but in a more nonparametric manner. Moreover, we demonstrate that the PLRT achieves the optimal minimax rate for the nonparametric hypothesis testing in the sense of Ingster (1993), and also discover that this global testing is more powerful/efficient than the local one in terms of the smaller minimum separation rate. Note that most other smoothing spline based tests, e.g., LMP and GML tests ([13, 57, 27, 8, 43]), either lead to complicated null distributions with nuisance parameters, or have not addressed the optimality issues. One major advantage of our PLRT over GLRT is that the specifications of the former null limit distribution are only determined by the parameter space, while the latter heavily depends on the choice of kernel function. In other words, our PLRT tests the nonparametric models in a more fundamental way.

In the end, we would like to reiterate the highlights of this paper:

(i). Our asymptotic C.I. has the point-wise consistency and shorter length than the Bayesian C.I.;
(ii). Our SCB is the first one applicable to the general class of nonparametric regression models;
(iii). Our local and global likelihood ratio testing both yield the Wilk's phenomenon. More importantly, we prove that the global testing is more sensitive/powerful than the local one in terms

of the smaller minimum separation rate.

As an important by-product of this paper, we derive the asymptotic equivalence of inferences between the periodic and non-periodic smoothing splines under mild conditions; see Remark 5.2. In general, our discoveries reveal an intrinsic connection between two rather different basis structures, which in turn may be used to facilitate the practical implementations. The under-smoothing is usually needed in the nonparametric inferences, e.g., [39, 24, 25], and its amount has been precisely quantified for the inference procedures considered in this paper. We also give three general under-smoothing rules in Remarks 3.1 – 3.3. However, we also note that the under-smoothing is actually not needed in the global testing, i.e., PLRT. As it will be seen, the innovative FBR is an ideal theoretical tool for studying the above inference problems.

Our paper is mainly devoted to theoretical studies, and leaves the more practical issues, e.g., tuning of the smoothing parameter, and the more challenging adaptive inferences as future topics. The general class of nonparametric models in consideration is fundamentally important so that the inferences on the more complicated models, e.g., multivariate extension, become conceptually simple by applying similar likelihood based approach and FBR techniques. In particular, the semiparametric extension has been investigated in [9]. The rest is organized as follows. Section 2 introduces the basic notations, model assumptions, and some preliminary RKHS results. Section 3 presents the key technical tool of this paper, i.e., FBR, and the local asymptotics of the smoothing spline as its trivial application. In Sections 4 and 5, two local and two global inference procedures together with their theoretical properties are formally discussed, respectively. In Section 6, we give three concrete examples showing the validity of our theories. Numerical studies are also provided for both periodic and non-periodic splines. All the technical arguments are included in Appendix or Online Supplementary ([46]).

2. Preliminaries.

2.1. *Notations 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 models under the primary assumption that

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

where $g_0(\cdot)$ is some unknown smooth function and $F(\cdot)$ is some known link function. It covers two sub-classes of statistical interest. The first sub-class assumes that the data are modelled by $y_i|z_i \sim p(y_i; \mu_0(z_i))$ for some conditional distribution p unknown upto the parameter μ_0. Instead of assuming the distributional knowledge, the second sub-class only specifies the moment relation in the sense that there exists some known positive function $\mathcal{V}(\cdot)$ such that $Var(Y|Z) = \mathcal{V}(\mu_0(Z))$. The nonparametric estimation of g in the latter is engaged by using the quasi-likelihood $Q(y; \mu) \equiv \int_y^\mu (y - s)/\mathcal{V}(s) ds$, where $\mu = F(g)$, as an objective function ([59]). Despite distinct modelling principles, these two sub-classes have a large overlap since $Q(y; \mu)$ coincides with several commonly used distributions $\log p(y; \mu)$ under various combinations of $(F, \mathcal{V})$ as summarized in Table 1 below.

p	Normal	Logistic	Gamma(α, β)	Poisson	Inverse Gaussian
$F(a)$	a	$\frac{\exp(a)}{1+\exp(a)}$	$\exp(a)$	$\exp(a)$	$\exp(a)$
$\mathcal{V}(s)$	1	$s(1-s)$	$\alpha^{-1}s^2$	s	s^3

TABLE 1

Five commonly used distributions together with their mean and variance functions.

From now on, we focus on the smooth criterion function $\ell(y;a):\mathcal{Y}\times\mathbb{R}\mapsto\mathbb{R}$, and allow it to cover the above two statistical classes, i.e., $\ell(y;a)=Q(y;F(a))$ or $\log p(y;F(a))$. Denote the parameter space $\mathcal{H}$ as the m-th order Sobolev space:

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

where m is assumed to be known and larger than $1/2$. In some cases, $\mathcal{H}$ is also defined as a subclass of $H^m(\mathbb{I})$, i.e., homogeneous Sobolev space $H_0^m(\mathbb{I})$, which has an additional restriction $g^{(j)}(0)=g^{(j)}(1)$ for $j=0,1,\ldots,m-1$ (also known as the class of periodic functions). Let $J(g,\widetilde{g})=\int_{\mathbb{I}}g^{(m)}(z)\widetilde{g}^{(m)}(z)dz$. Consider the penalized nonparametric estimate $\widehat{g}_{n,\lambda}$:

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

where $J(g,g)$ is the roughness penalty of order m and λ is the smoothing parameter converging to zero as n. We use $\lambda/2$ (rather than λ) here only for the simplicity of future expressions. The existence and uniqueness of $\widehat{g}_{n,\lambda}$ are guaranteed by Theorem 2.9 of [21] when the null space $\mathcal{N}_m\equiv\{g\in H^m(\mathbb{I}):J(g,g)=0\}$ is finite dimensional and $\ell(y;a)$ is concave and continuous w.r.t. a.

We next assume some basic model conditions. Let $\mathcal{I}_0$ be the range for $g_0(z)$, 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 first 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.*

$$E\left\{\exp(\sup_{a\in\mathcal{I}}|\ddot{\ell}_a(Y;a)|/C_0)\Big|Z\right\}\leq C_1,\ \ a.s.,\tag{2.3}$$

and

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

(b). There exists a positive constant C_2 such that $C_2^{-1}\leq I(Z)\equiv -E(\ddot{\ell}_a(Y;g_0(Z))|Z)\leq C_2$ a.s..
(c). $\epsilon\equiv\dot{\ell}_a(Y;g_0(Z))$ satisfies $E(\epsilon|Z)=0$ and $E(\epsilon^2|Z)=I(Z)$, a.s.

Assumption A.1 (a) implies the slow diverging rate, i.e., $O_P(\log n)$, of $\max_{1\leq i\leq n}\sup_{a\in\mathcal{I}}|\ddot{\ell}_a(Y_i;a)\vee\ell_a'''(Y_i;a)|$. In the case that $\ell(y;a)=\log p(y;a)$, Assumption A.1 (b) imposes the boundedness and positive definiteness of the Fisher information, and Assumption A.1 (c) trivially holds if p satisfies some regularity conditions. However, when $\ell(y;a)=Q(y;a)$, 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

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

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

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

Assumption A.1 (c) follows from the definition of $\mathcal{V}(\cdot)$. Sub-exponential tail Condition (2.6) and boundedness Condition (2.7) are very mild quasi-likelihood model assumptions (also assumed in [37]). The assumption that F_j and $\dot{F}_j$ are both bounded over $\mathcal{I}$ could be restrictive without assuming the estimation consistency. However, we can remove it in most models, e.g., binary logistic regression, by applying similar empirical processes arguments as in Section 7 of [37].

2.2. *Reproducing Kernel Hilbert Space (RKHS).* Some RKHS results are introduced into our general model framework as slight extensions of [12] and [41], e.g., an important Sobolev norm (2.8). It is well known that, when $m > 1/2$, $\mathcal{H} = H^m(\mathbb{I})$ is a RKHS in which we endow the inner product and norm as, respectively, $\langle g, \widetilde{g}\rangle = E\{I(Z)g(Z)\widetilde{g}(Z)\} + \lambda J(g, \widetilde{g})$ and

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

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

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

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

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

for any $g, \widetilde{g} \in H^m(\mathbb{I})$. Denote $V(g, \widetilde{g}) = E\{I(Z)g(Z)\widetilde{g}(Z)\}$. Hence, $\langle g, \widetilde{g}\rangle = V(g, \widetilde{g}) + \langle W_\lambda g, \widetilde{g}\rangle$, which implies $V(g, \widetilde{g}) = \langle (id - W_\lambda)g, \widetilde{g}\rangle$, where id denotes the identity operator.

In the below, we assume that there exists a sequence of basis functions in the space $H^m(\mathbb{I})$, which can simultaneously diagonalize the bilinear forms V and J. Such an eigenvalue/eigenfunction assumption is typical in the smoothing spline literature, and is critical to control the local behaviors of our penalized estimates. Hereinafter, positive sequences a_μ and b_μ satisfying $\lim_{\mu\to\infty}(a_\mu/b_\mu) = c > 0$ is denoted as $a_\mu \asymp b_\mu$. If $c = 1$, we denote $a_\mu \sim b_\mu$. Let $\sum_\nu$ denote the sum over $\nu \in \mathbb{N} = \{0, 1, 2, \ldots\}$ for convenience. Denote the sup-norm of $g \in H^m(\mathbb{I})$ as $\|g\|_{\sup} = \sup_{z\in\mathbb{I}} |g(z)|$.

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

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

where $\delta_{\mu\nu}$ is the Kronecker's delta. Furthermore, for any $g \in H^m(\mathbb{I})$, it admits the Fourier expansion $g = \sum_\nu V(g, h_\nu)h_\nu$ with the convergence held under $\|\cdot\|$-norm.

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

PROPOSITION 2.1. *For any $g \in H^m(\mathbb{I})$ 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, it is crucially important to give sufficient conditions on Assumption A.2 in terms of the underlying eigensystem. When $\ell(y;a) = -(y-a)^2/2$ and $\mathcal{H} = H_0^m(\mathbb{I})$, Assumption A.2 is known to satisfy if (γ_ν, h_ν) is chosen as the trigonometric basis (6.2) specified in Example 6.1. However, in the more general $\ell(y;a)$ with $\mathcal{H} = H^m(\mathbb{I})$, we will show that Assumption A.2 is still valid if (γ_ν, h_ν)'s are chosen as the (normalized) solutions of the problem

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

where $\pi(\cdot)$ is the marginal density function of covariate Z. Our proof heavily relies on the ODE techniques developed in [6, 50].

Let $C^m(\mathbb{I})$ be the class of m-th order continuously differentiable functions over $\mathbb{I}$.

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

Our Proposition 2.2 can be viewed as a nontrivial extension of Utreras (1988) in which $I = \pi = 1$.

In the end, we summarize the notations on Frechét derivatives to be used later. The Frechét derivatives of $\ell_{n,\lambda}$ can be shown to be, for any $\Delta g, \Delta g_j \in H^m(\mathbb{I})$ and $j = 1, 2, 3$,

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

Note that $S_{n,\lambda}(\widehat{g}_{n,\lambda}) = 0$. In particular, $S_{n,\lambda}(g_0)$ is of interest, which can be expressed by

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

The Frechét derivative of $S_{n,\lambda}$ ($DS_{n,\lambda}$) is defined as $DS_{n,\lambda}(g)\Delta g_1 \Delta g_2$ ($D^2 S_{n,\lambda}(g)\Delta g_1\Delta g_2\Delta g_3$), and can be written as $D^2\ell_{n,\lambda}(g)\Delta g_1\Delta g_2 = n^{-1}\sum_{i=1}^n \ddot{\ell}_a(Y_i; g(Z_i))\langle K_{Z_i}, \Delta g_1\rangle\langle K_{Z_i}, \Delta g_2\rangle - \langle W_\lambda \Delta g_1, \Delta g_2\rangle$ ($D^3\ell_{n,\lambda}(g)\Delta g_1\Delta g_2\Delta g_3 = n^{-1}\sum_{i=1}^n \ell_a'''(Y_i; g(Z_i))\langle K_{Z_i}, \Delta g_1\rangle\langle K_{Z_i}, \Delta g_2\rangle\langle K_{Z_i}, \Delta g_3\rangle$).

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

PROPOSITION 2.3. *$DS_\lambda(g_0) = -id$, where recall that id is the identity operator on $H^m(\mathbb{I})$.*

3. Functional Bahadur Representation. In this section, we first develop the key technical tool of this paper: *Functional Bahadur Representation*, and then present the local asymptotics of the smoothing spline estimate as its straightforward application. In fact, FBR provides the rigorous theoretical foundation for the series of inference tools to be established in Sections 4 and 5.

3.1. *Functional Bahadur Representation.* We first state the relationship between the $\|\cdot\|_{\sup}$-norm and $\|\cdot\|$-norm in Lemma 3.1 below, and then derive a *concentration inequality* in Lemma 3.2 as the preliminary step in obtaining FBR. Denote h as $\lambda^{1/(2m)}$.

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 $g \in H^m(\mathbb{I})$. 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

$$\mathcal{G} = \{g(z) \in H^m(\mathbb{I}) : \|g\|_{\sup} \leq 1, J(g,g) \leq c_m^{-2} h\lambda^{-1}\},$$

where the constant c_m is specified in Lemma 3.1. Recall here that $T_i = (Y_i, Z_i)$s denote the full data variables with domain $\mathcal{T}$. Our Lemma 3.2 below proves a concentration inequality on the empirical processes $Z_n(g)$ defined as, for any $g \in \mathcal{G}$ and $z \in \mathbb{I}$,

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

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:*

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

where c_m is specified in Lemma 3.1. Then we have

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

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

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

A set of simple (but unnecessarily weakest) sufficient conditions for Assumption A.3 is provided in Proposition 3.3 below. Before stating this result, we introduce another norm in the space $\mathcal{H}$ which is more commonly used in the 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 $\lambda \leq 1$, $\|\cdot\|_{\mathcal{H}}$ is one type of Sobolev norm dominating $\|\cdot\|$ defined in (2.8). Denote

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

Note that λ^* is known as the optimal order of smoothing parameter when estimating $g_0 \in H^m(\mathbb{I})$.

PROPOSITION 3.3. *Suppose that Assumption A.1 holds, and further that $\|\widehat{g}_{n,\lambda} - g_0\|_{\mathcal{H}} = o_P(1)$. 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 valid. In particular, $\widehat{g}_{n,\lambda}$ achieves the optimal rate of convergence, i.e., $O_P(n^{-m/(2m+1)})$, when $\lambda = \lambda^*$.*

Now we are ready to present the key technical tool: *Functional Bahadur Representation*, which is also of independent interest. By incorporating λ into the norm (2.8), we obtain a more powerful version of Shang (2010) 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 $nh^2 \to \infty$ are satisfied. 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 $C_\ell = \sup_{z \in \mathbb{I}} E\{\sup_{a \in \mathcal{I}} |\ell_a'''(Y; a)| | Z = z\}$. Also, the RHS of (3.5) is $o_P(n^{-m/(2m+1)})$ when $h = h^*$.

3.2. *Local Asymptotic Behaviors.* In this section, we obtain the point-wise asymptotics of $\widehat{g}_{n,\lambda}$ as a direct application of FBR. The equivalent kernel idea may be used for deriving similar results but only restricted to the L_2 regression, e.g., [48]. In contrast, our FBR-based proof applies to the more general regression and tackles the problems from a totally new perspective. Notably, our results reveal that some well known global convergence properties continue to hold in the local sense; see Remarks 3.1, and three types of under-smoothing conditions are summarized in Remarks 3.1 – 3.3.

THEOREM 3.5. *(General Regression) Let the Assumptions A.1 through A.3 be satisfied. Suppose that $h = o(1)$, $nh^2 \to \infty$, and $a_n \log n = o(n^{-1/2})$, where a_n is defined in Theorem 3.4, as $n \to \infty$. Furthermore, assume that, for any $z_0 \in \mathbb{I}$,*

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

Denote $g_0^* = (id - W_\lambda)g_0$ as 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 have the following result.

COROLLARY 3.6. *Suppose that 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).

We want to emphasize that our Theorem 3.5 covers a general class of nonparametric models under penalized estimation. To illustrate Corollary 3.6 in more details, we consider the L_2-regression in which $W_\lambda g_0(z_0)$ (also b_{z_0}) has an explicit expression under the additional boundary condition:

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

Specifically, we consider two separate cases, i.e., $b_{z_0} \neq 0$ and $b_{z_0} = 0$. Our results also apply to the boundary points after paying the price of boundary conditions (3.11). To gain more flexibility, we provide an alternative set of conditions to (3.11), i.e., (3.14), which can be implied by the so-called "exponential envelop condition" in [41].

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 also the normalized eigenfunctions h_νs satisfy (2.11). Assume that $g_0 \in H^{2m}(\mathbb{I})$ and satisfies $\sum_\nu |V(g_0^{(2m)}, h_\nu)h_\nu(z_0)| < \infty$.*

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

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

If $h \asymp n^{-d}$ for some $\frac{1}{4m+1} < d \leq \frac{2m}{8m-1}$, 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 condition (3.11) by the following reproducing kernel conditions that, for any $z_0 \in (0,1)$, as $h \to 0$

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

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

In (3.12), we note that the asymptotic bias b_{z_0} is proportional to $g_0^{(2m)}(z_0)$, and the asymptotic variance $\sigma_{z_0}^2$ can be expressed as a weighted sum of squares of the *(infinitely many)* basis functions $h_\nu(z_0)$s, i.e., (3.8). It is worthy pointing out that these observations are consistent with those in the polynomial spline setting that the former is proportional to $g_0^{(2m)}(z_0)$, and the latter is a weighted sum of squares of the *(finitely many)* normalized B-spline basis functions evaluated at z_0; see [61].

REMARK 3.1. *The existing smoothing spline literature are mostly concerned about global convergence properties of the estimate. For example, Nychka (1995) (Rice and Rosenblatt (1983)) derived the global convergence rate in terms of the (integrated) mean squared error. However, we mainly focus on the local asymptotic behaviors here, and find that those well known global results actually hold in the local (point-wise) sense as well. Stone (1982) showed that, when $g_0 \in H^{m'}(\mathbb{I})$, the optimal convergence rate of $\widehat{g}_{n,\lambda}$ (in the global sense) is $O_P(n^{-m'/(2m'+1)})$. However, to achieve the above optimal rate, the order of λ has to be chosen according to the degree of the regularization. Specifically, λ needs to be chosen as $h^{2m} \asymp n^{-2m/(2m'+1)}$ under the m-th order Sobolev penalization. Under the setting of Corollary 3.7 where $g_0 \in H^{2m}(\mathbb{I})$ and the m-th order penalty is used, our local result (3.12) shows that $\widehat{g}_{n,\lambda}(z_0)$ has achieved the point-wise rate $O_P(n^{-2m/(4m+1)})$, which turns out to be the optimal global rate, when $\lambda \asymp n^{-2m/(4m+1)}$. To further remove the asymptotic estimation bias, we have to sacrifice the convergence rate of $\widehat{g}_{n,\lambda}(z_0)$ in (3.13) by choosing some faster convergent λ. This further coincides with the under-smoothing procedure known in the literature.*

REMARK 3.2. *In practice, it might be more convenient to fix $g_0 \in H^m(\mathbb{I})$ and properly tune the smoothing parameter for removing the estimation bias. For example, in the general regression, we can achieve this purpose by choosing some faster convergent λ than the optimal $\lambda^* \asymp n^{-2m/(2m+1)}$, i.e., $h^* \asymp n^{-1/(2m+1)}$. Specifically, we can choose $h \asymp n^{-d}$ with $\frac{1}{2m} < d < \frac{2m}{8m-1}$ when $m > 1 + \sqrt{3}/2 \approx 1.866$. It can be checked that the above h satisfies the conditions in Theorem 3.5. By reproducing kernel property and (2.9), $|W_\lambda g_0(z_0)| = |\langle W_\lambda g_0, K_{z_0}\rangle| = \lambda |J(g_0, K_{z_0})| = O(\lambda J(K_{z_0}, K_{z_0})^{1/2})$. By Proposition 2.1 and Lemma 2.2 of [12], $J(K_{z_0}, K_{z_0}) = \sum_\nu \frac{|h_\nu(z_0)|^2\gamma_\nu}{(1+\lambda\gamma_\nu)^2} \asymp h^{-(2m+1)}$, which implies $W_\lambda g_0(z_0) = O(\lambda h^{-m-1/2}) = O(h^{m-1/2})$ for any $z_0 \in \mathbb{I}$. Thus, $\sqrt{nh}W_\lambda g_0(z_0) = O(n^{1/2}h^m) = o(1)$, i.e., $b_{z_0} = 0$ in (3.9), implied by the above range of h.*

REMARK 3.3. *In fact, we can also remove the estimation bias while fixing $\mathcal{H} = H^m(\mathbb{I})$ and employing $\lambda = \lambda^*$ by assuming $\sum_\nu |V(g_0, h_\nu)|\gamma_\nu^{1/2} < \infty$, which might be "the weakest possible conditions". Below are the explanations. By Proposition 2.1, we have*

$$|W_\lambda g_0(z_0)| \leq \sup_\nu \|h_\nu\|_{\sup} \cdot \sum_\nu |V(g_0, h_\nu)| \frac{\lambda\gamma_\nu}{1+\lambda\gamma_\nu} = \lambda^{1/2} \sup_\nu \|h_\nu\|_{\sup} \cdot \sum_\nu |V(g_0, h_\nu)|\gamma_\nu^{1/2} \frac{(\lambda\gamma_\nu)^{1/2}}{1+\lambda\gamma_\nu}.$$

Under the above assumption and the inequality that $|V(g_0, h_\nu)|\gamma_\nu^{1/2}\frac{(\lambda\gamma_\nu)^{1/2}}{1+\lambda\gamma_\nu} \leq |V(g_0, h_\nu)|\gamma_\nu^{1/2}$, dominated convergence theorem implies $\sum_\nu |V(g_0, h_\nu)|\gamma_\nu^{1/2}\frac{(\lambda\gamma_\nu)^{1/2}}{1+\lambda\gamma_\nu} = o(1)$, as $\lambda \to 0$, and thus $\sqrt{nh}W_\lambda g_0(z_0) = o(\sqrt{nh}\lambda^{1/2}) = o(1)$ when $\lambda = \lambda^*$. Since $V(g_0^{(m)}, h_\nu) = (-1)^m \nu^m V(g_0, h_\nu)$, the above assumption holds when $V(g_0^{(m)}, h_\nu)$ are absolutely summable, e.g., $g_0^{(m)} \in Lip_\alpha(\mathbb{I})$ with $\alpha \in (1/2, 1]$, the Lipschitz functional class with index α; see the so-called Wiener algebra in [30].

4. Local Asymptotic Inferences.

We consider inferring $g(\cdot)$ *locally* by constructing the point-wise C.I. in Section 4.1 and testing the local hypothesis via likelihood ratio in Section 4.2. In particular, the related inference optimality will also be discussed; see Remark 4.1 and Theorem 4.6.

4.1. *Point-wise 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 as $\rho_0 = \rho(g_0(z_0))$, e.g., $\rho_0 = F(g_0(z_0)) = E(Y|Z = z_0)$. An instance is $\rho_0 = \exp(g_0(z_0))/(1 + \exp(g_0(z_0)))$ for the logistic regression model. Corollary 3.6 together with the Delta method immediately implies Proposition 4.1 on the point-wise C.I. in which the asymptotic estimation bias is assumed to be removed, e.g., by under-smoothing.

PROPOSITION 4.1. *(Point-wise Confidence Interval) Suppose that Assumptions in Corollary 3.6 hold and the estimation bias asymptotically vanishes, i.e., $\lim_{n\to\infty}(nh)^{1/2}(W_\lambda g_0)(z_0) = 0$. If $\rho'(g_0(z_0)) \neq 0$, 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 $N(0, 1)$ and $\dot{\rho}(\cdot)$ is the first derivative of $\rho(\cdot)$.*

From now on, we focus on the point-wise C.I. for $g_0(z_0)$ and discuss its optimality in the end. For simplicity, we consider the setting that $\ell(y; a) = -(y - a)^2/(2\sigma^2)$, $Z \sim Unif[0, 1]$ and $\mathcal{H} = H_0^m(\mathbb{I})$ under which Proposition 4.1 implies the following asymptotic 95% C.I. 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) in Example 6.1 for the derivation of (4.1). When σ is unknown, we may replace it by any consistent estimate. Under mild conditions, we further prove in Remark 5.2 that the same form of C.I. (4.1) also holds for the cubic spline, i.e., $\mathcal{H} = H^2(\mathbb{I})$, although the center $\widehat{g}_{n,\lambda}(z_0)$ is different. As far as we are aware, (4.1) is the first rigorously proven point-wise C.I. for the smoothing spline. However, the major contribution of this section is the surprising comparison between (4.1) and the classical *Bayesian Confidence Interval* proposed (studied) by Wahba (1983) (Nychka (1988)) even they are constructed based on different principles, i.e., frequentist v.s. Bayesian. Firstly, we would like to emphasize that the Bayesian C.I. is only shown to approximately achieve the 95% nominal level in an average sense. In other words, its average coverage probability over the observed covariates is *not* exactly 95% even asymptotically. Secondly, the Bayesian C.I. ignores the important issue of uniformity of coverage across the design space, and thus may not be reliable if only evaluated at peaks or troughs as pointed out in [40]. However, our asymptotic C.I. (4.1) is proven to be valid at any point. A more striking fact is that (4.1) even possesses the shorter length than those of Wahba's and Nychka's Bayesian C.I.s at the same time. See the discussions below.

For the purpose of comparison, we first derive the asymptotic equivalent versions of the Bayesian C.I.s. Wahba (1983) heuristically proposed the following Bayesian C.I.:

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

where $a(h^\dagger) = n^{-1}\left(1 + (1+(\pi n h^\dagger))^{-4} + 2\sum_{\nu=1}^{n/2-1}(1+(2\pi\nu h^\dagger))^{-4}\right)$. Under the assumption that $h^\dagger = o(1)$ and $(nh^\dagger)^{-1} = o(1)$, Lemma 6.1 in Example 6.1 implies that $2\sum_{\nu=1}^{n/2-1}(1+(2\pi\nu h^\dagger))^{-4} \sim I_1/(\pi h^\dagger) = 4I_2/(3\pi h^\dagger)$ since $I_2/I_1 = 3/4$ when $m = 2$. The asymptotic equivalent version of Wahba's Bayesian C.I. is thus

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

Nychka (1988) further shortened the Wahba's version (4.2) by proposing

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

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 also showed that

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

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

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

Therefore, Nychka's Bayesian C.I. (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 have discovered that Wahba's Bayesian C.I. and Nychka's Byesian C.I. are asymptotically 15.4% and 6.1% wider than our C.I. (4.1), respectively. Similar conclusion holds for $m > 2$. Furthermore, the simulations performed in Example 6.1 empirically verify the superior performance of our C.I. in both periodic and non-periodic splines. Interestingly, we also realize that our *frequentist* C.I. (4.1) turns out to be the corrected version of Nychka's *Bayesian* C.I. (4.4) by removing its random bias term $b(z_0)$; see (4.6). The inclusion of $b(z_0)$ in Nychka's C.I. is problematic in the sense that: (i) it makes the point-wise limit distribution non-normal leading to the biased coverage probability; and (ii) it introduces additional variance unnecessarily enlarging the interval length. Therefore, by removing $b(z_0)$ from (4.7), we are able to achieve both the point-wise consistency and shorter interval in (4.1) without adding any computational burden.

REMARK 4.1. *It follows from Cai and Low (2004) that the lower bound on the length of the point-wise C.I. relies on the modulous of continuity over the parameter space. When the parameter space is $H^m(\mathbb{I})$, Donoho and Liu (1991) showed that the modulous of continuity is of order $n^{-m/(2m+1)}$. The length of our C.I. achieves this lower bound by adding a mild restriction on $g_0 \in H^m(\mathbb{I})$, i.e., $g_0^{(m)}$ has absolutely summable Fourier coefficients, and choosing $\lambda = \lambda^*$; see Remark 3.3.*

4.2. *Local Likelihood Ratio Test.* In this section, we propose the likelihood ratio method in testing the value of $g_0(z_0)$ at any point of interest $z_0 \in \mathbb{I}$. We first show that the null limit distribution is a scaled non-central Chi-square distribution with degree of freedom one, whose specification is jointly determined by the reproducing kernel and estimation bias, and then establish the central Chi-square limit distribution after removing the estimation bias. We also note that, as the smoothness of the regression function increases, the scaling constant will eventually converge to one. Therefore, we have unveiled an interesting Wilk's phenomenon (meaning that the asymptotic null distribution is independent of any nuisance parameters) arising from this nonparametric local testing, which injects new theoretical insight into the literature. Hence, the inversion of likelihood ratio test can be conveniently used to constructing the point-wise C.I. for $g_0(z_0)$, and also $F(g_0(z_0))$ due to the monotonicity of $F(\cdot)$; see Table 1. We further address the tricky testing efficiency/senstivity issue by studying its power behaviors under a sequence of local alternatives. The above issue is technically challenging since the testing sensitivity relies on the whole estimated curve even the test itself is local; see Theorem 4.6. An interesting testing sensitivity comparison will be made between the local LRT and its global counterpart in Section 5.2 in terms of their minimum separation rates. One related reference is Banerjee (2007) who considered similar test for the *monotone* functions, but his estimation method and null limit distribution are different from ours.

For some pre-fixed point (z_0, w_0), we consider the following hypothesis:

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

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 $g \in \mathcal{H}_0 = \{g \in H^m(\mathbb{I}) : g(z_0) = 0\}$. We consider the LRT statistic defined as

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

where $\widehat{g}^0_{n,\lambda}$ is the MLE of g under the local restriction, i.e., $\widehat{g}^0_{n,\lambda} = \arg\max_{g\in\mathcal{H}_0} L_{n,\lambda}(g)$.

Endowed with the norm associated with the inner product $\langle\cdot,\cdot\rangle$, $\mathcal{H}_0$ is a closed subset in $\mathcal{H} = H^m(\mathbb{I})$, and thus a Hilbert space. Proposition 4.2 below says that it also inherits the reproducing kernel and the penalty operator from $\mathcal{H}$. Its proof is trivial, and thus omitted.

PROPOSITION 4.2. *(a). Recall that $K(z_1, z_2)$ is the reproducing kernel for $H^m(\mathbb{I})$ 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 in $(\mathcal{H}_0, \langle\cdot,\cdot\rangle)$. 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 $g, \widetilde{g} \in \mathcal{H}_0$.*

Given Proposition 4.2, we are ready to derive the *restricted* FBR for $\widehat{g}^0_{n,\lambda}$ that is used to obtaining the null limit distribution. We first define the Frechét derivatives of $L_{n,\lambda}$ (under $\mathcal{H}_0$) by modifying those of $\ell_{n,\lambda}$ as follows: replace g, K_{Z_i} and W_λ by $w_0 + g$, $K^*_{Z_i}$ and W^*_λ, respectively. For example,

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

Similarly, we have $S^0_{n,\lambda}(\widehat{g}^0_{n,\lambda}) = 0$. Also define $S^0(g)\Delta g = E\{\langle S^0_n(g), \Delta g\rangle\}$ and $S^0_\lambda(g)\Delta g = S^0(g)\Delta g - \langle W^*_\lambda g, \Delta g\rangle$. As for the second derivatives, we have $DS^0_{n,\lambda}(g)\Delta g_1\Delta g_2 = D^2L_{n,\lambda}(g)\Delta g_1\Delta g_2$ 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

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

Similar as Theorem 3.4, we need an additional rate assumption for the restricted FBR result:

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 argument is similar as the proof of Proposition 3.3 by replacing $\mathcal{H}$ with its 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 Testing) Suppose that Assumptions A.1 through A.4 are satisfied. Also assume that $h = o(1)$, $nh^2 \to \infty$, $a_n = o(\min\{r_n, n^{-1}r_n^{-1}(\log n)^{-1}, n^{-1/2}(\log n)^{-1}\})$, and $r_n^2 h^{-1/2} = o(a_n)$. Furthermore, for any $z_0 \in [0,1]$, if $n^{1/2}(W_\lambda g_0)(z_0)/\sqrt{K(z_0,z_0)} \to -c_{z_0}$,*

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

then we obtain: (i). $\|\widehat{g}_{n,\lambda}-\widehat{g}^0_{n,\lambda}-w_0\| = O_P(n^{-1/2})$; (ii). $-2n\cdot LRT_{n,\lambda} = n\|\widehat{g}_{n,\lambda}-\widehat{g}^0_{n,\lambda}-w_0\|^2+o_P(1)$;

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

with non-centrality parameter $c^2_{z_0}/c_0$; under H_0.

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

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

The reproducing kernel K is uniquely determined for any Hilbert space if it exists; see [14]. So, c_0 defined in (4.10) is only determined by the parameter space. Hence, different choices of (γ_ν, h_ν) in (4.12) will give exactly the same value of c_0 although some particular choice will facilitate the computation of c_0. For example, when $\mathcal{H} = H^m_0(\mathbb{I})$, we can explicitly calculate the value of c_0 as 0.75 (0.83) when $m = 2$ (3) by choosing the trigonometric basis (6.2). Interestingly, in the more general $H^2(\mathbb{I})$, we can obtain the same value of c_0 even without specifying its (rather different) eigensystem under mild conditions; see Remark 5.2. However, the value of c_{z_0} in (4.11) partly depends on the asymptotic estimation bias (see (3.9)), whose estimation is notoriously difficult. Fortunately, under various under-smoothing conditions, we can show $c_{z_0} = 0$, and thus establish the central Chi-square limit distribution. For example, we can choose faster convergent smoothing parameter when fixing $g_0 \in H^m(\mathbb{I})$ as in Remark 3.2. Alternatively, we can also insist using λ^* but assume the parameter space with more smoothness; see Remark 3.1. In Corollary 4.5, we explore the latter approach in more details.

COROLLARY 4.5. *Suppose that Assumptions A.1 through 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 $\sum_\nu |V(g_0,h_\nu)|^2\gamma^d_\nu$ for some $d > 1+1/(2m)$, which is implied by $g_0 \in H^{md}(\mathbb{I})$. Furthermore, if (4.10) is satisfied for any $z_0 \in [0,1]$, then (4.11) holds with limiting distribution $c_0\chi^2_1$ given $\lambda = \lambda^*$.*

Corollary 4.5 discovers a nonparametric type of Wilk's phenomenon arising from the local hypothesis testing, which further converts into the classical one in the parametric setup as $m \to \infty$ since $\lim_{m\to\infty} c_0 = 1$. Our result delivers new theoretical insight into the nonparametric local hypothesis testing; see it *global* counterpart in Section 5.2.

In the end, we discuss the efficiency/sensitivity of our local LRT by characterizing the limiting power under a sequence of local alternatives converging to the null. Let η_n be any positive sequence converging to zero. Consider the local alternative: $H_{1n} : g = g_{n0}$, where $g_{n0} \equiv g_* + \eta_n f_n$, and $g_*, f_n \in H^m(\mathbb{I})$ satisfying $g_*(z_0) = w_0$, and as $n \to \infty$,

$$|f_n(z_0)| \to \infty, J(f_n, f_n) \le C_a(n\lambda\eta_n^2)^{-1} \text{ and } n\eta_n^2 V(f_n, f_n) \to \tau_{z_0}^2, \tag{4.13}$$

under $g = g_*$ for some constants $C_a > 0$ and $\tau_{z_0}^2$. Under the above design of H_{1n}, we have $g = g_{n0} \in H^m(\mathbb{I})$ and $g(z_0) = g_{n0}(z_0) \neq w_0$ (asymptotically), i.e., H_0 does not hold. This well constructed sequence of local alternatives can be used to examine how much deviation from $g_{n0}(z_0)$ to $g_*(z_0)$ (or equivalently, w_0) within $H^m(\mathbb{I})$ can trigger the rejection of H_0 using $LRT_{n,\lambda}$. Theorem 4.6 explicitly says that H_{1n} can be detected when $g_{n0}(z_0)$ and w_0 are separated by a distance converging to zero at some rate no faster than $n^{-m/(2m+1)}$, which is further proven to be a sharp bound. This minimum separation rate $n^{-m/(2m+1)}$ is achieved under the smoothing parameter of the same order as the optimal one in the estimation, i.e., $\lambda = \lambda^*$. We also note that the minimum separate rate coincides with the minimum length of the point-wise C.I. established in [36]; see Remark 4.1. As for the global likelihood ratio testing, Theorem 5.4 derives a faster minimal separation rate, i.e., $n^{-2m/(4m+1)}$, indicating that the global testing is actually more powerful/sensitive. The above surprising difference in the minimum separation rates turns out to be reasonable after a second thought. This is because, in the local testing, the data information is not as fully used as in the global one, which leads to a slower minimum separation rate as a compensation/tradeoff.

THEOREM 4.6. *Let $m > 1 + \sqrt{3}/2 \approx 1.866$, $h \asymp n^{-d}$ for $\frac{1}{2m+1} \le d < \frac{2m}{10m-1}$ and $\eta_n \ge (nh)^{-1/2} + h^m$. Assume that $\ell(Y; g)$ is the log-density. Suppose, under both $g = g_*$ and $g = g_{n0}$, Assumptions A.1 through A.4 are satisfied, e.g., Assumption A.1 holds with g_0 therein replaced by g_* and g_{n0}, respectively, and $\sum_\nu |V(g_*, h_\nu)|\gamma_\nu^{1/2} \le C_*$ for some positive constant C_* unrelated to n, and (4.10) holds. Then for any $\delta \in (0, 1)$, there exists a sufficiently large constant N such that*

$$\inf_{n\ge N} P\left(\textit{reject } H_0 | H_{1n} \textit{ is true}\right) \ge 1 - \delta. \tag{4.14}$$

The lower bound of η_n, i.e., $n^{-m/(2m+1)}$, is achieved when $h = h^*$. If $\eta_n = o(n^{-m/(2m+1)})$, then we can find a sequence of functions f_n satisfying (4.13) such that (4.14) does not hold. Thus, $n^{-m/(2m+1)}$ is the minimum separation rate for the local LRT to detect H_{1n}.

The log-density condition in Theorem 4.6 is only assumed for simplicity, and can be easily relaxed by assuming that $P_{g_{n0}}^n$ is contiguous with respect to $P_{g_*}^n$, where P_g^n is denoted as the distribution function under the model parameter g. The above contiguity assumption can be verified using Le Cam's first lemma, i.e., Theorem 3.10.2 of [55]. We want to point out that the techniques in the proof of Theorem 4.6 are very generic and can be applied to derive the minimum separation rate in the local testing based on other test statistic, e.g., $T_{n,\lambda} = \sqrt{nh}(\widehat{g}_{n,\lambda}(z_0) - w_0)/\sigma_{z_0}$, which is essentially the same.

5. Global Asymptotic Inferences.

Depicting the global behavior of a smooth function is crucially important in practice. In Sections 5.1 and 5.2, we develop the *global* counterparts of Section 4 by constructing the simultaneous confidence band and testing global hypothesis via likelihood ratio. Again, the FBR is the key ingredient in the theoretical studies.

5.1. *Simultaneous Confidence Band.* In this section, we establish the simultaneous confidence band (SCB) for $g(z)$ following the approach of Bickel and Rosenblatt (1973). The proposed SCB centers around $\widehat{g}_{n,\lambda}(z)$ with the $\sqrt{\log n}$-wider bandwidth than the asymptotic point-wise C.I., and is proven to be asymptotically valid over any compact subset in $(0,1)$ based on the FBR and strong approximation techniques. The approach of Bickel and Rosenblatt (1973) was originally developed in the density estimation context, and then has been extended to M-estimation ([22]) and local polynomial estimation ([10]). For example, SCB is constructed for (generalized) varying-coefficient models based on the latter method; see [17, 60]. The volume of tube method ([51]) is another approach, but requires the error distribution to be symmetric; see its application to [61, 32]. All the models considered above require the error to be additive and Gaussian. Sun, Loader and McCormick (2000) relaxed the restrictive error assumption of [51] in generalized linear models, but had to translate the nonparametric estimation into the parametric one. As far as we are aware, we construct the first SCB for the general class of nonparametric models including the logistic regression. In particular, the minimum bandwidth of our SCB is shown to achieve the lower bound established in Genovese and Wasserman (2008). In addition, the equivalent kernel conditions assumed in this section imply an interesting by-product that the asymptotic lengths of our point-wise C.I.s (also scaling constants in the null limit distribution (4.11)) based on the cubic spline and periodic spline are actually the same despite their different eigensystems; see Remark 5.2.

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

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

where C_2, C_K are some positive constants. Condition (5.1) implies that ω is an equivalent kernel of the reproducing kernel function K with certain degree of approximation accuracy. Meanwhile, we also require some regularity conditions on ω. In particular, we assume that

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

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

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

where C_3, C_ω, C_ρ are some positive constants and $\sigma_\omega^2=\int_{\mathbb{R}}\omega(t)^2dt$. The following exponential envelop 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 through 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, assume that there exist positive constants C_0 and C_1 such that $E\{\exp(|\epsilon|/C_1)|Z\}\le C_0$, a.s., and that (5.1) – (5.4) hold. The conditional density of ϵ given $Z=z$, namely $\pi(\epsilon|z)$, is assumed to be satisfied for some positive constants ρ_1 and ρ_2,*

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

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

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

where d_n is some constant relying merely on h, ρ, φ and C_ρ.

The FBR developed in Section 3.1 and the strong approximation techniques ([5]) are crucial to the proof of Theorem 5.1. The uniform distribution condition on Z is only assumed for simplicity, and can be relaxed to the density that is bounded away from zero and infinity. Condition (5.5) is easy to check in various situations. For example, it holds for the conditional normal model, i.e., $\epsilon | Z = z \sim N(0, \sigma^2(z))$, if $\sigma^2(z)$ satisfies $\inf_z \sigma(z)^2 > 0$, and $\sigma(z)$ and $\sigma'(z)$ both have finite upper bounds. The existence of the bias term $W_\lambda g_0(z)$ in the SCB (5.6) may lead to poor small sample performances. We avoid the bias estimation by a slight under-smoothing which is also advocated by [39], following earlier results of [24, 25] where it is shown that under-smoothing is more efficient than explicit bias correction when the goal is to minimize the coverage error. Specifically, this bias effect will asymptotically disappear if we assume:

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

Condition (5.7) is slightly stronger than the under-smoothing Condition that $\sqrt{nh}(W_\lambda g_0)(z_0) = o(1)$ assumed for the C.I. in Proposition 4.1. Due to the uniform boundedness of h_ν's in Assumption A.2 and the generalized Fourier expansion of $W_\lambda g_0$, it is easy to show that (5.7) is satisfied if we (i) increase the smoothness of g_0; (ii) choose some suboptimal smoothing parameter; or (iii) assume slightly stronger conditions on $g_0^{(m)}$; see Remarks 3.1 – 3.3.

Proposition 5.2 reveals the validity of Conditions (5.1) – (5.3) in the setting of L_2 regression. The proof relies on the explicit construction of an equivalent kernel for various m in [38]. Here we only consider $m = 2$ for simplicity.

PROPOSITION 5.2. *(L_2 regression) Consider the setting that $\ell(y;a) = -(y-a)^2/(2\sigma^2)$, $Z \sim Unif[0,1]$ and $\mathcal{H} = H^2(\mathbb{I})$, i.e., $m = 2$. 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 $\omega_0(t) = \frac{1}{2\sqrt{2}}\exp(-|t|/\sqrt{2})\left(\cos(t/\sqrt{2})+\sin(|t|/\sqrt{2})\right)$. 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, it is trivial to calculate that $\sigma_\omega^2 = 0.265165\sigma^{7/2}$ since $\sigma_{\omega_0}^2 = \int_{-\infty}^{\infty}|\omega_0(t)|^2dt = 0.265165$ and $m = 2$. Since $C_\rho = 0$ for arbitrary $\rho \in (0,2]$, by the formula $B(t)$ in Theorem A1 of [5], we know that*

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

When $\rho = 2$, the above d_n is simplified as $(2\log(h^{-1}-2h^{\varphi-1}))^{1/2}$. In general, we know 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, e.g., under (5.7), we have the following $100 \times (1-\alpha)\%$ SCB:

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

where d_n is given in (5.8), $c_\alpha^* = -\log(-\log(1-\alpha)/2)$ and $\widehat{\sigma}$ is a consistent estimate. Note that we exclude the boundary points in (5.9). To obtain the uniform coverage, we have to sacrifice a bit by increasing the bandwidth upto $\sqrt{\log n}$-order over the length of the point-wise C.I., e.g., (4.1).

REMARK 5.2. *One interesting by-product we discover in the setting of Proposition 5.2 is that the point-wise C.I.s for $g_0(z_0)$ based on the cubic spline and periodic spline share the asymptotic equivalent length at any fixed $z_0 \in (0,1)$. This result is a bit surprising since these two splines have very distinct eigensystem. 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% C.I.

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

Since $\sigma_{\omega_0}^2 = I_2/\pi$, the lengths of C.I.s (4.1) (based on periodic spline) and (5.10) (based on cubic spline) surprisingly coincide with each other. Another useful application of Proposition 5.2 is to find the value of c_0 needed in the local LRT test when $\mathcal{H} = H^2(\mathbb{I})$; see Theorem 4.3. According to the definition of c_0 in (4.10), we have $c_0 \sim \sigma_{z_0}^2/(hK(z_0, z_0))$. Under (5.1), we can show $K(z_0, z_0) \sim h^{-1}\omega(0) = h^{-1}\sigma^{3/2}\omega_0(0) = 0.3535534 h^{-1}\sigma^{3/2}$. 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 the one found in periodic splines, i.e., $\mathcal{H} = H_0^2(\mathbb{I})$. These somewhat amazing phenomena have never been observed in the literature and may be used to facilitate the construction of C.I. and local LRT in practice.

REMARK 5.3. *Genovese and Wasserman (2008) showed that when g_0 belongs to a m-order Sobolev ball, the lower bound for the average length of the SCB is proportional to $b_n n^{-m/(2m+1)}$ with b_n merely depending on $\log n$. We next show that the (minimum) bandwidth of our SCB can achieve this lower bound with $b_n = (\log n)^{(m+1)/(2m+1)}$. Based on Theorem 5.1, the bandwidth of our SCB has the shrinking rate $d_n(nh)^{-1/2}$, where d_n is of the order $\sqrt{\log n}$; 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 the condition in Remark 3.3. It can be verified that (5.7) holds when $nh^{2m+1}\log n = O(1)$ which sets an upper bound for h. When h is chosen as the above upper bound, i.e., $O(n\log n)^{-1/(2m+1)}$, and $d_n \asymp \sqrt{\log n}$, our SCB has achieved its minimum bandwidth, i.e., $n^{-m/(2m+1)}(\log n)^{(m+1)/(2m+1)}$, which turns out to be rate optimal according to [20].*

In practice, the construction of our SCB requires a delicate choice of (h, φ). Otherwise, over/under-coverage of the true function may occur near the boundary points. Unfortunately, as pointed by [5], there is no practical/theoretical guideline on how to find the optimal (h, φ), although one can choose proper h to make the band as thin as possible. Hence, in next section, we propose a more practically feasible approach to explore the global behaviors, which only requires the tuning of h. Moreover, we are able to specify an optimal h under which our likelihood-ratio-based approach achieves the optimal minimax rate of hypothesis testing specified by Ingster (1993).

5.2. *Global Likelihood Ratio Test.* Nonparametric hypothesis testing is of equal importance in studying the global behaviors; see an overview and references in [23]. There is a vast literature dealing with this problem among which the Generalized Likelihood Ratio Testing (GLRT) ([18]) arises as a fundamental approach. Due to the technical tractability, Fan et al (2001) only focused on the local polynomial fitting in the GLRT; also see [19] for the sieve extension. Based on the smoothing spline estimate, we propose an alternative method called as the Penalized Likelihood Ratio Testing (PLRT), which not only applies to the simple hypothesis but also to a very general class of composite hypothesis; see Remark 5.4. The null limit distribution is proven to be nearly χ^2 with diverging degree of freedom. Therefore, the Wilk's phenomenon observed in local LRT continues to hold in nonparametric *penalized* likelihood but with a more nonparametric form. Besides the much more concise assumptions, one major advantage of our PLRT over GLRT is that the specifications of the former null limit distribution are only determined by the parameter space, while the latter heavily depends on the choice of kernel function; see Table 2 in [18]. In other words, the PLRT is closer to the nature of nonparametric models. Furthermore, we show that the PLRT achieves the optimal minimax rate for hypothesis testing in the sense of Ingster (1993). In practice, the power performances of PLRT are superior and better than those of GLRT for small sample sizes in both periodic and non-periodic splines; see Example 6.1. In summary, our PLRT is not only intuitive to use but also powerful to apply. However, most other smoothing spline based tests, e.g., LMP and GML tests ([13, 57, 27, 8, 43]), use ad-hoc discrepancy measure leading to complicated null distributions with nuisance parameters, and have not addressed the optimality issues at all. Hence, their applicability is restricted; see more review in [34].

Consider the following "global" hypothesis:

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

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

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

Even when g_0 is unknown, the limit distribution of PLRT under H_0^{global} can still be derived, though the value of test statistic is not calculable. More importantly, this nice property can be used to test composite hypothesis; see Remark 5.4.

Theorem 5.3 below derives the null limiting distribution of $PLRT_{n,\lambda}$ based on the FBR result.

THEOREM 5.3. *Let the Assumptions A.1 through A.3 be satisfied. Also assume that* $nh^{2m+1} = O(1)$, $nh^2 \to \infty$, $a_n = o(\min\{r_n, n^{-1}r_n^{-1}h^{-1/2}(\log n)^{-1}, n^{-1/2}(\log n)^{-1}\})$ *and* $r_n^2 h^{-1/2} = o(a_n)$. *Furthermore, under* H_0^{global}, $E\{\epsilon^4|Z\} \le C$, *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*

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

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

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

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

Direct examination reveals that $h \asymp n^{-d}$ with $\frac{1}{2m+1} \le d < \frac{2m}{8m-1}$ satisfies the rate conditions required in the above Theorem when $m > (3+\sqrt{5})/4 \approx 1.309$. In Theorem 5.4, we further show that some particular choice of h in the above range will guarantee the minimax optimality of PLRT.

Theorem 5.3 implies that $-2nr_K \cdot PLRT_{n,\lambda}$ is asymptotically $N(u_n, 2u_n)$ since $n\|W_\lambda g_0\|^2 = o(h^{-1}) = o(u_n)$ implied by the proof of Theorem 5.3 and the definition of u_n. As n approaches ∞, i.e., $u_n \to \infty$, we know that $N(u_n, 2u_n)$ is nearly the same as $\chi^2_{u_n}$ in distribution. Hence, $-2nr_K \cdot PLRT_{n,\lambda}$ is approximately distributed as $\chi^2_{u_n}$, denoted as

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

Therefore, we claim that the fundamental Wilk's phenomenon also holds under nonparametric penalized estimation but with a more nonparametric form, i.e., the diverging degree of freedom. Obviously, the specifications of (5.15), i.e., σ_K^2 and ρ_K^2, are only determined by the parameter space and model setup. This is in stark contrast with the null limit distribution of GLRT whose specifications vary with the used kernel functions; see Table 2 of [18]. Unfortunately, there is no theoretical guideline in choosing the most suitable kernel function. Hence, our PLRT tests the nonparametric models in a more fundamental way. In addition, we find that the under-smoothing is not needed in carrying out the valid global testing, i.e., (5.15), unlike the other inference procedures.

We next discuss the calculation of (r_K, u_n) and its implications in some important setup. In the setting of Proposition 5.2, we can show $\sigma_K^2 = h\sigma^{-2}\int_0^1 K(z,z)dz \sim h\sigma^{-2}(h^{-1}\omega(0)) = \sigma^{-1/2}\omega_0(0) = 0.3535534\sigma^{-1/2}$ by applying this Proposition. Similarly, we have $\rho_K^2 \sim \sigma^{-1/2}\sigma^2_{\omega_0} = 0.265165\sigma^{-1/2}$. So $r_K = 1.3333$ and $u_n = 0.4714h^{-1}\sigma^{-1/2}$. Surprisingly, if we replace $H^2(\mathbb{I})$ by $H_0^2(\mathbb{I})$ in the above setup, our direct calculations in Case (I) of Example 6.1 reveal that (r_K, u_n) share exactly the same values. We also note that $r_K \to 1$ when $\mathcal{H} = H_0^m(\mathbb{I})$ as the degree of smoothness m tends to ∞. This is consistent with the scaling constant 2 in the classical likelihood ratio theory. Note that the possibly unknown parameter σ in u_n can be essentially profiled out without affecting the null limit distribution. We keep it here only for the consistency with our general modeling framework. Alternatively, we can directly simulate the null limit distribution by fixing the nuisance parameters, e.g., the null value g_0, at reasonable values or estimates (e.g., by wild bootstrap) even without calculating the values of (r_K, u_n). This is one major advantage of the Wilk's type of results.

REMARK 5.4. *In this Remark, we will discuss the composite hypothesis testing via PLRT and the related Wilk's phenomenon. Specifically, we are able to test whether g belongs to some finite dimensional class of functions with bounded Sobolev norm, which is much larger than the null space $\mathcal{N}_m$ considered in the literature. As an example, we consider testing, for any integer $q \geq 0$,*

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

where $\mathcal{L}_q(\mathbb{I}) \equiv \{g(z) = \sum_{l=0}^q a_l z^l : a = (a_0, a_1, \ldots, a_q)^T \in \mathbb{R}^{q+1}\}$ represents the class of q-th polynomials over $\mathbb{I}$. 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 $\widehat{g}_*(z) = \sum_{l=0}^q \widehat{a}_{*l} z^l$. Let g_{0q} denote some unknown "true" parameter in $\mathcal{L}_q(\mathbb{I})$ with some polynomial coefficient $a^0 = (a_0^0, a_1^0, \ldots, a_q^0)^T$. For testing the composite hypothesis (5.16), we first decompose the PLRT statistic $PLRT_{n,\lambda}^{com}$ as $L_{n1} - L_{n2}$, where $L_{n1} = \ell_{n,\lambda}(g_{0q}) - \ell_{n,\lambda}(\widehat{g}_{n,\lambda})$ and $L_{n2} = \ell_{n,\lambda}(g_{0q}) - \ell_{n,\lambda}(\widehat{g}_*)$. By formulating

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

we notice that L_{n2} appears to be the PLRT in the parametric setup. We can prove the order of L_{n2} as $O_P(n^{-1})$ no matter $q < m$ (by applying the parametric theory in [47]) 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

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

Since Theorem 5.3 also applies to the unknown null value g_{0q}, L_{n1} follows the limit distribution (5.15). So does $PLRT^{com}_{n,\lambda}$ under the composite hypothesis (5.16) considering $L_{n2} = O_P(n^{-1})$.

To the end of this section, we remark that PLRT achieves the optimal minimax rate of hypothesis testing specified in Ingster (1993). By developing the *uniform* version of FBR, we rigorously prove the above claim in Theorem 5.4. For convenience, we only consider $\ell(Y;a) = -(Y-a)^2/2$. The extension to the more general setup can be found in [46] under stronger assumptions, e.g., more restrictive $\mathcal{G}_a$ defined below. Write the local alternative as $H_{1n}: g = g_{n0}$, where $g_{n0} = g_0 + g_n$, $g_0 \in H^m(\mathbb{I})$ and g_n belongs to some alternative value set $\mathcal{G}_a \equiv \{g \in H^m(\mathbb{I}) | Var(g(Z)^2) \leq \zeta E^2\{g(Z)^2\}, J(g,g) \leq \zeta\}$ 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*

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

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 $h = h^{**} \equiv n^{-2/(4m+1)}$.

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

Theorem 5.4 proves that, when $h = h^{**}$, PLRT can detect any local alternatives with a separation rate no faster than $n^{-2m/(4m+1)}$, which turns out to be the optimal minimax rate in the sense of [26]; see more discussions in Remark 5.5. The above rates are consistent with those derived in the local polynomial estimation ([18]) although our nonparametric models are more general and conditions in Theorem 5.4 are more concise. In contrast with the local LRT studied in Theorem 4.6, we note an interesting fact that two different smoothing parameters are employed for obtaining the minimum separation rates, i.e., $\lambda = \lambda^* = n^{-2m/(2m+1)}$ for the local testing and $\lambda = \lambda^{**} \equiv n^{-4m/(4m+1)}$ for the global testing. Such a distinction might be caused by the different nature of these two testing, i.e., local v.s. global, which is reflected by their different minimum separation rates; see the discussions right below Theorem 4.6. In Example 6.1, a simulation study was conducted to compare the powers of PLRT and GLRT for both periodic and non-periodic splines; see Tables 3 & 4. As n grows, we find that the powers in both tests rapidly approach to one, and, more interestingly, that PLRT appears to be more powerful in the small sample size such as $n = 20$.

REMARK 5.5. *We note that the optimal minimax rate of testing established in Ingster (1993) is under the usual $\|\cdot\|_{L_2}$-norm (w.r.t. Lebesgue measure). However, our minimum separation rate derived under $\|\cdot\|$-norm is still optimal due to the trivial domination of $\|\cdot\|$ over $\|\cdot\|_{L_2}$ (under conditions of Theorem 5.4). We next heuristically explain why the minimax rates of testing associated with $\|\cdot\|$, denoted as b_n', and with $\|\cdot\|_{L_2}$, denoted as b_n, are essentially the same under conditions of Theorem 5.4, which may not be easy to see. By definition, whenever $\|g_n\| \geq b_n'$ or $\|g_n\|_{L_2} \geq b_n$, H_0^{global} can be rejected with large probability, or equivalently, the local alternatives can be detected. Note that b_n' and b_n are the minimum rates that satisfy this property. Ingster (1993) 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 local alternatives in terms of* $\|\cdot\|_{L_2}$*, and so* $b_n \le b_n'$*. On the other hand, suppose* $h = h^{**} \asymp n^{-2/(4m+1)}$ *and* $\|g_n\| \ge Cn^{-2m/(4m+1)} \asymp b_n$ *for some large* $C > \zeta^{1/2}$*. Since* $\lambda J(g_n, g_n) \le \zeta\lambda \asymp \zeta n^{-4m/(4m+1)}$*, it follows that* $\|g_n\|_{L_2} \ge (C^2-\zeta)^{1/2} n^{-2m/(4m+1)} \asymp b_n$*. This means* b_n *is a upper bound for detecting the local alternatives in terms of* $\|\cdot\|$*, and so* $b_n' \le b_n$*. Therefore,* b_n' *and* b_n *are of the same order.*

6. Examples. This section illustrates the applicability of our theories with three examples, and demonstrates the empirical performances of our inference procedures via some simulations.

EXAMPLE 6.1. (L_2 *Regression*) Consider the nonparametric regression model

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

where $\epsilon \sim N(0, \sigma^2)$ with unknown σ^2. Hence, we have $I(Z) = \sigma^{-2}$ and $V(g, \widetilde{g}) = \sigma^{-2}E\{g(Z)\widetilde{g}(Z)\}$. For simplicity, we assume that the true value of σ is one and Z is uniformly distributed over $\mathbb{I}$. In the simulations, the unknown σ can be either consistently estimated or profiled out as in [18]. The function "ssr()" in R package *assist* was used to select the smoothing parameter λ, i.e., h, based on CV or GCV; see [58]. Note that, in the simulations, we implicitly perform the under-smoothing using the GCV-selected smoothing parameter since the employed test function is sufficiently smooth; see Remark 3.1. 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 can choose the basis functions h_μ's as

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

with the eigenvalues $\gamma_{2k-1} = \gamma_{2k} = \sigma^2(2\pi k)^{2m}$ for $k \ge 1$ and $\gamma_0 = 0$. Assumption A.2 is trivially satisfied for the above choice of (h_μ, γ_μ)'s. We first prove a useful Lemma below.

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

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

Proposition 4.1 implies the asymptotic 95% point-wise C.I. for $g(z_0)$ as $\widehat{g}_{n,\lambda}(z_0) \pm 1.96\sigma_{z_0}/\sqrt{nh}$ by choosing proper h; see (3.13). To obtain an explicit form of $\sigma_{z_0}^2$, which is the limit of $hV(K_{z_0}, K_{z_0})$ as $h \to 0$, we note that $hV(K_{z_0}, K_{z_0}) = \sigma^2 h \left(1 + 2\sum_{k=1}^\infty (1 + (2\pi h^\dagger k)^{2m})^{-2}\right) \sim (I_2\sigma^{2-1/m})/\pi$ based on Lemma 6.1. Hence, in practice, we use

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

where $\widehat{\sigma}^2 = n^{-1}\sum_i (Y_i - \widehat{g}_{n,\lambda}(Z_i))^2$. Alternatively, according to Theorem 4.4, we can also establish the asymptotic C.I. by inverting the local likelihood ratio. The above trigonometric basis (6.2) gives

$$\begin{aligned} Q_l(\lambda, z_0) &= \sigma^2 + \sum_{k\ge1}\left\{\frac{|h_{2k}(z_0)|^2}{(1+\lambda\sigma^2(2\pi k)^{2m})^l} + \frac{|h_{2k-1}(z_0)|^2}{(1+\lambda\sigma^2(2\pi k)^{2m})^l}\right\} \\ &= \sigma^2 + 2\sigma^2\sum_{k\ge1}\frac{1}{(1+\lambda\sigma^2(2\pi k)^{2m})^l} = \sigma^2 + 2\sigma^2\sum_{k\ge1}\frac{1}{(1+(2\pi h^\dagger k)^{2m})^l}. \end{aligned}$$

Combining (4.12) with Lemma 6.1, we have $c_0 = I_2/I_1$. Hence, $c_0 = 0.75$ (0.83) when $m = 2$ (3).

In Table 2 below, we compare the coverage probability (CP) between our asymptotic C.I. (6.4), denoted as ACI, and Nychka's Bayesian C.I. (4.7), denoted as NCI, at three quartiles (Q_1, Q_2, Q_3) of the observed covariates Z. We assume the true periodic function $g_0(z) = 12\sin(\pi z)$ and estimate it using periodic spline under $m = 2$. The CP was computed as the proportion of the C.I.s that cover g_0 at that point over $10,000$ replications. From Table 2, it is observed that the CPs of ACIs and NCIs are both reasonably close to the 95% nominal level. However, as n grows, the CPs of the ACIs are getting closer to 95% while those of NCIs always stay a bit above 95% with the increasing gap, in particular when $n = 800$. This somewhat unsatisfactory performance of NCI is consistent with the observations in [40]. Except for better CP, our ACI also has shorter length; see Table 2. Our simulation results empirically verify our claim in Section 4.1 that the Bayesian C.I. has biased coverage probability and larger interval length.

	Q_1		Q_2		Q_3	
n	NCI	ACI	NCI	ACI	NCI	ACI
100	95.12	93.74	95.43	94.17	95.33	93.99
200	95.94	94.64	95.75	94.47	95.79	94.51
300	95.81	94.60	95.97	94.74	95.92	94.62
400	95.93	94.60	96.03	94.90	95.92	94.60
800	96.20	94.75	96.14	94.94	96.34	95.15

TABLE 2

Comparison of 100× CP% of CIs in Case (I). The lengths of the NCI are $1.14, 0.88, 0.75, 0.68, 0.52$*, and those of ACIs are* $1.08, 0.83, 0.71, 0.64, 0.49$*, for* $n = 100, 200, 300, 400, 800$*. Nominal level is 95%.*

In Figure 1, we constructed the SCB for g over $(0, 1)$ based on (5.9) by taking $d_n = (-2\log h)^{1/2}$, and compared it with three so-called *point-wise* confidence bands constructed by linking the end-points of the ACI (6.4), Wahba's Bayesian C.I. (4.3) and NCI (4.7) at each observed covariate, denoted as ACB, BCB1 and BCB2, respectively. Data were generated under the same setup as above. From Figure 1, it is observed that the coverage properties of all the confidence bands are reasonably good, and getting better as n grows. Meanwhile, all band areas clearly shrink to zero as n. We also note that the ACBs possess the smallest band area, while the SCBs have the largest one, which is not surprising by its definition. The more technical reason is due to the d_n factor in the construction of SCB, which is of $\sqrt{\log n}$-order; see Remark 5.1.

In the end of Case (I), we considered testing $H_0 : g$ *is linear* at the 95% significance level by both our PLRT and the GLRT ([18]). By Lemma 6.1 and (6.2), some direct calculations reveal that $r_K = 1.3333$ and $u_n = 0.4714(h\sigma^{1/2})^{-1}$ in (5.15) when $m = 2$. In the simulations, we replaced σ by $\widehat{\sigma}$ defined above. Data were generated under the same setup except that a more linear true function $g(z) = 3.2\sin(\pi z)$ (than the previous $g(z) = 12\sin(\pi z)$) was used for the purpose of power comparison. For the GLRT method, the Epanechnikov kernel function is used under the R function "glkerns()". For PLRT method, GCV was used to select the smoothing parameter considering the *slight* difference between h^* and h^{**}. Table 3 compares the powers (proportions of rejections in $10,000$ replications) for four sample sizes. When $n = 40$ or larger, both test methods achieve almost 100% power. We also note that PLRT shows moderate advantage in smaller sample even though the chosen smoothing parameter (by GCV) is not optimal in terms of testing. An intuitive reason is that the smoothing spline estimate in PLRT uses the full data information; while the local polynomial estimate used in GLRT only uses local data information, which might not be sufficient when sample size is small. Of course, as n grows, such difference rapidly vanishes due to the increasing data information.

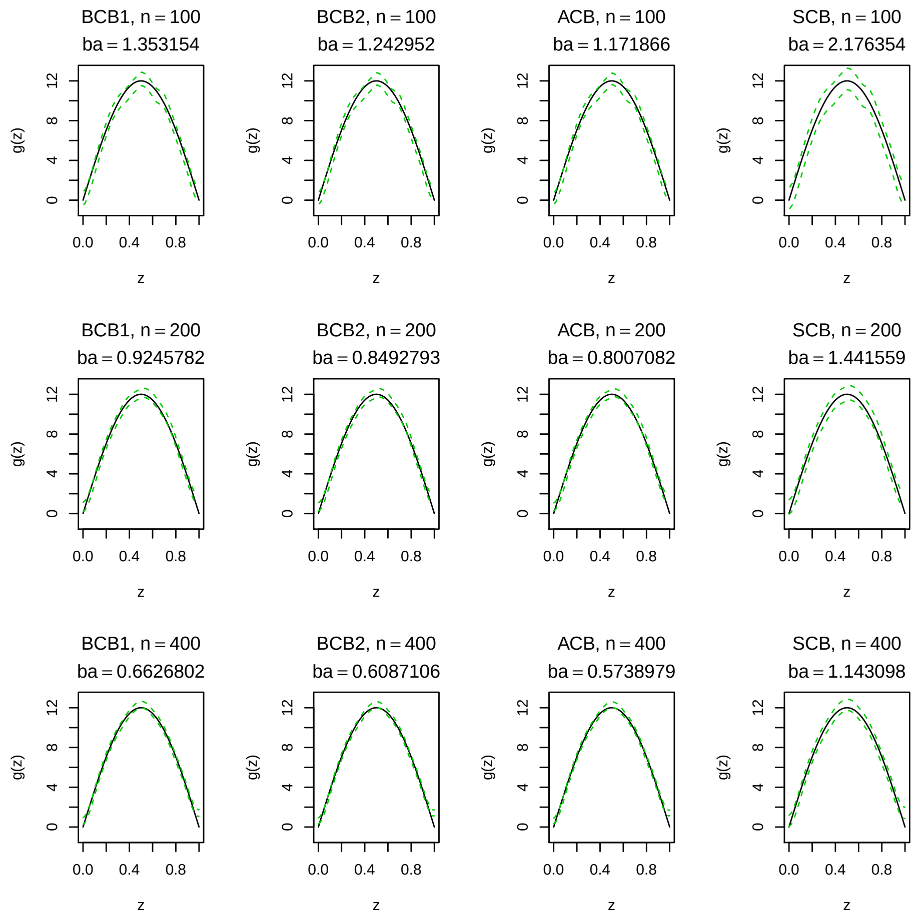

FIG 1. *95% point-wise and simultaneous confidence bands for periodic g in Case (I). The upper and lower bands are indicated by green curves, while the central black curve represents the true function. The numerical band area is denoted as "ba".*

Case (II). $\mathcal{H} = H^m(\mathbb{I})$**:** For this more general $\mathcal{H}$, we repeated most of the inference procedures in Case (I) by assuming the non-periodic true function $g(z) = 6\sin(2.8\pi z)$ and using the cubic spline for estimation. Hence, we only point out the differences. Figure 2 summarizes the simultaneous confidence band and point-wise confidence bands in which BCB1 was computed by (4.2) and BCB2 was constructed by scaling the length of the BCB1 by a factor $\sqrt{27/32} \approx 0.919$. We tested the linearity of g at significance level 95%, and assumed $g(z) = 1.5\sin(2.8\pi z)$. Table 4 summarizes the powers of the PLRT and GLRT. From Figure 2 and Table 4, we conclude that all the observations and findings for the periodic spline in Case (I) remain the same for the non-periodic spline.

EXAMPLE 6.2. (*Nonparametric Gamma Model*) Consider the two-parameter exponential model

$$Y|Z \sim \text{Gamma}(\alpha, \exp(g_0(Z))),$$

where $\alpha > 0$, $g_0 \in H_0^m(\mathbb{I})$ and Z is uniform over $[0,1]$. This framework corresponds to $\ell(y; g(z)) = \alpha g(z) + (\alpha - 1)\log y - y\exp(g(z))$. Thus, it can be shown that $I(z) = \alpha$, leading us to choose the basis functions to be h_νs defined as in (6.2) with $\sigma = \alpha^{-1/2}$, and the eigenvalues to be $\gamma_{2k} =$

	100× Power%			
	$n = 20$	$n = 30$	$n = 40$	$n = 100$
PLRT	92.7	98.6	99.7	100
GLRT	90.1	98.2	99.7	100

TABLE 3

Power comparison of the PLRT and GLRT for four sample sizes in Case (I). Significance level is 95%.

	100 × Power%			
	$n = 20$	$n = 30$	$n = 40$	$n = 100$
PLRT	92.5	99.1	99.7	100
GLRT	90.3	98.9	99.5	100

TABLE 4

Power comparison of the PLRT and GLRT for four sample sizes in Case (II). Significance level is 95%.

$\gamma_{2k-1} = \alpha^{-1}(2\pi k)^{2m}$ for $k \geq 1$, and $\gamma_0 = 0$. One can conduct the local and global inferences in the similar manner as Case (I) of Example 6.1.

EXAMPLE 6.3. (*Nonparametric Logistic Regression*) In this example, we consider the binary response $Y \in \{0, 1\}$ modeled by the following logistic model

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

where $g_0 \in H^m(\mathbb{I})$. A straightforward calculation gives $I(z) = \frac{\exp(g_0(z))}{(1+\exp(g_0(z)))^2}$. In this example, c_0 has no explicit form since the pair (h_μ, γ_μ) has no explicit form. Therefore, we have to find an accurate estimate of c_0. To achieve this, we will use (2.11) to approximate h_νs and γ_νs. Thus, accurate estimates $\widehat{I}(z)$ and $\widehat{\pi}(z)$ are needed. Observe that $I(z) = P(Y = 1|Z = z)P(Y = 0|Z = z)$. To approximate $I(z)$, we thus have to plug in an estimate of $P(Y = 1|Z = z)$. Note $P(Y = 1|Z = z) = [P(Z = z|Y = 1)P(Y = 1)]/P(Z = z)$. Denote $\pi_1(z) = P(Z = z|Y = 1)$, $r = P(Y = 1)$ and $\pi(z) = P(Z = z)$. Let $\widehat{\pi}_1$ and $\widehat{\pi}$ be consistent estimate of π_1 and π, such as the kernel density estimators. Let $\widehat{r}$ be the proportion of $Y = 1$, which is a consistent estimate of r. Then we can approximate $I(z)$ by $\widehat{I}(z) = \frac{\widehat{\pi}_1(z)\widehat{r}}{\widehat{\pi}(z)}\left(1 - \frac{\widehat{\pi}_1(z)\widehat{r}}{\widehat{\pi}(z)}\right)$. One may find the approximated eigensystem $(\widehat{h}_\mu, \widehat{\lambda}_\mu)$s by solving the approximate version of (2.11) in which $I(\cdot)$ and $\pi(\cdot)$ are replaced by $\widehat{I}(\cdot)$ and $\widehat{\pi}(\cdot)$, respectively. Obviously, the approximated eigensystem are needed in the local and global inferences. For example, to perform PLRT test based on Theorem 5.3, we can use $(\widehat{h}_\mu, \widehat{\lambda}_\mu)$s to specify the null limiting distribution and the theoretical 95% cutoff value in (5.15). Meanwhile, we are also aware that solving the approximated eigensystem could be computationally tricky. Fortunately, in the PLRT, it can be avoided by directly simulating the null limit distributions, e.g., by the wild bootstrap in [37], as long as the Wilk's type of results holds.

Acknowledge: We appreciate helpful discussions with Professor Chong Gu.

APPENDIX

A.1. *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

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

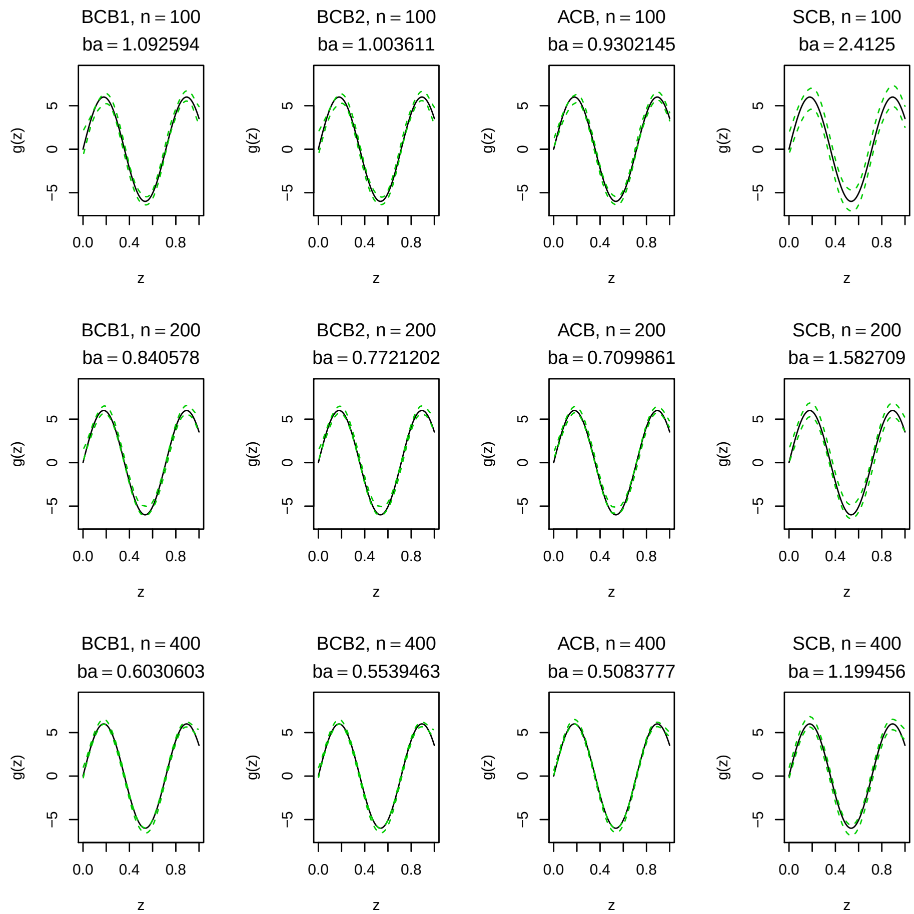

FIG 2. *95% point-wise and simultaneous confidence bands for non-periodic g in Case (II). The upper and lower bands are indicated by green curves, while the central black curve represents the true function. The numerical band area is denoted as "ba".*

for any $g \in H^m(\mathbb{I})$ and $\nu \in \mathbb{N}$. It follows by (A.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 considering $K_z(\cdot) = \sum_\nu V(K_z, h_\nu)h_\nu(\cdot)$. Furthermore, (A.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 $\nu, \mu \in \mathbb{N}$. In the end, we can conclude the proof of Proposition 2.1 by considering $W_\lambda h_\nu(\cdot) = \sum_\mu V(W_\lambda h_\nu, h_\mu)h_\mu$.

A.2. *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 [6] that the growing rates 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), the orthogonality and completeness of h_νs under V thus follow from Theorem 2.1 (pp. 189) and Theorem 4.2 (pp. 199) of [11] with the usual L_2-inner product $\langle,\rangle_{L_2}$ replaced with V. Therefore, when h_νs 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 by arguments in page 147 of [35]. The eigenspace corresponding to zero eigenvalue contains functions ϕs 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 [54] that this eigenspace is $\mathcal{P}_{m-1}$, 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 $\nu \geq m$. If $g \in L_2(\mathbb{I})$ 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_νs, 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 $h_\nu^{(m)}/\gamma_\nu^{1/2}$, $\nu \geq m$, in $L_2(\mathbb{I})$ under $\langle, \rangle_{L_2}$. Now, for any $\tilde{g} \in H^m(\mathbb{I})$, by completeness of h_νs 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 $h_\nu^{(m)}/\gamma_\nu^{1/2}$, $\nu \geq m$ 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, implying $\tilde{g} = \sum_\nu V(\tilde{g}, h_\nu) h_\nu$ converges in $\|\cdot\|$.

Next we show the uniform boundedness of h_ν. We only consider those h_νs corresponding to nonzero γ_νs. 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 boundary conditions in (2.11) and integration by parts one can check that $J(h_\nu, h_\nu) = \gamma_\nu$. On both sides, dividing $I\pi$ and taking m-order derivatives 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 $t(z) = \int_0^z [I(s)\pi(s)]^{1/(2m)} ds/C$, $C = \int_0^1 [I(z)\pi(z)]^{1/(2m)} dz$. Define $\phi_\nu(t(z)) = \psi_\nu(z)$. Then by a direct examination, ϕ_ν satisfies the following differential equation

(A.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_js, $j = 0, \ldots, 2m-1$, are coefficient functions depending only on $I\pi$ and m, and $\rho_\nu = \lambda_\nu C^{2m}$. In general the forms of q_js are complicated though they can be determined by Faá di Bruno's formula ([28]). As an illustration, when $m = 2$, $q_0(t) = 0$, $q_3(t) = -(K/4)\omega^{(1)}(z(t))\omega(z(t))^{-3/4}$, $q_2(t) = -(K^2/4)(\omega^{(1)}(z(t)))^2 \omega(z(t))^{-3/2}$, and

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

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

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

then (A.2) is equivalent to

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

with the boundary conditions $u_\nu^{(j)}(0) = u_\nu^{(j)}(1) = 0$, $j = 0, \ldots, m-1$. Note (A.4) is the classic form of differential systems discussed in [6]. According to [6], ρ_νs are simple due to the regular boundary conditions, and the residue 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, [50]), 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 [6], $2m\zeta^{2m-1}G(t_1, t_2; \zeta^{2m})$ is uniformly bounded for $t_1, t_2 \in \mathbb{I}$, thus, the residue is uniformly abounded 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 (A.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 use $\phi_\nu(t(z)) = \psi_\nu(z)$ 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 Sobolev embedding theorem ([1]), $\|h_\nu\|_{\sup}^2 \le c^2(V(h_\nu, h_\nu) + J(h_\nu, h_\nu)) = c^2(1+\gamma_\nu)$. Using Theorem 5 of [54], for any $j = 1, \ldots, m$, there is 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 m-order derivative on both sides of $(-1)^m h_\nu^{(2m)} = \gamma_\nu I\pi h_\nu$, one has for some constant $c_2 > 0$, for any ν, $\|h_\nu^{(3m)}\|_{\sup} \le \gamma_\nu \sum_{j=0}^m \binom{m}{j}\|(I\pi)^{(m-j)}\|_{\sup} \cdot \|h_\nu^{(j)}\|_{\sup} \le c_2(1+\gamma_\nu)^{3/2}$.

Again, by Theorem 5 of [54] 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_νs.

A.3. *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_μs,

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

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

A.4. *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 [42], for any $t > 0$, $P(\|Z_n(f) - Z_n(g)\| \ge t) \le 2\exp\left(-\frac{t^2}{8\|f-g\|_{\sup}^2}\right)$. Then by Lemma 8.1 in [31], 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 [31] 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 [31],

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

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 (A.6),

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

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

A.5. *Proof of Theorem 3.4.* By Assumption A.1 (a), it is not difficult to check the following

$$\max_{1\leq i\leq n}\sup_{a\in\mathcal{I}} |\ddot{\ell}_a(Y_i;a)| = O_P(\log n). \tag{A.7}$$

By (A.7) we can let $C > C_0$ be sufficiently large so that the event $B_{n1} = \{\max_{1\leq i\leq n}\sup_{a\in\mathcal{I}}|\ddot{\ell}_a(Y_i;a)| \leq C\log n\}$ has large probability.

Denote $g = \widehat{g}_{n,\lambda} - g_0$. By Assumption A.3, the event $B_{n2} = \{\|g\| \leq r_n \equiv M((nh)^{-1/2} + h^m)\}$ has large probability with some preselected large M, so $B_n = B_{n1}\cap B_{n2}$ has 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, $\|\widetilde{g}\|_{\sup} \leq 1$. Note that $J(\widetilde{g},\widetilde{g}) = d_n^{-2}\lambda^{-1}(\lambda J(g,g)) \leq d_n^{-2}\lambda^{-1}\|g\|^2 \leq d_n^{-2}\lambda^{-1}r_n^2 \leq c_m^{-2}h\lambda^{-1}$. Thus, when 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 $A_i = \{\sup_{a\in\mathcal{I}}|\ddot{\ell}_a(Y_i;a)| \leq C\log n\}$ 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 $z\in\mathbb{I}$, since $g_0(z)\in\mathcal{I}_0$ and $d_n = o(1)$, both $g_0(z)+d_ng_1(z)$ and $g_0(z)+d_ng_2(z)$ fall in $\mathcal{I}$ when n is sufficiently large (recall that

$\mathcal{I}_0$ and $\mathcal{I}$ are specified in Assumption A.1). Therefore,

$$\begin{aligned}|\psi_n(T_i;d_ng_1)-\psi_n(T_i;d_ng_2)| &= C^{-1}c_m^{-1}(\log n)^{-1}h^{1/2}d_n^{-1}|\psi(T_i;g_1)-\psi(T_i;g_2)|\cdot I_{A_i}\\ &= C^{-1}c_m^{-1}(\log n)^{-1}h^{1/2}d_n^{-1}\Big|\int_{g_0(Z_i)}^{g_0(Z_i)+d_ng_1(Z_i)}\ddot{\ell}_a(Y_i;a)\cdot I_{A_i}da\\ &\quad -\int_{g_0(Z_i)}^{g_0(Z_i)+d_ng_1(Z_i)}\ddot{\ell}_a(Y_i;a)\cdot I_{A_i}da\Big|\\ &\le C^{-1}c_m^{-1}(\log n)^{-1}h^{1/2}d_n^{-1}\cdot d_n\|g_1-g_2\|_{\sup}\cdot\sup_{a\in\mathcal{I}}|\ddot{\ell}_a(Y_i;a)|\cdot I_{A_i}\\ &\le C^{-1}c_m^{-1}(\log n)^{-1}h^{1/2}d_n^{-1}\cdot d_n\cdot C\log n\cdot\|g_1-g_2\|_{\sup}\\ &= c_m^{-1}h^{1/2}\|g_1-g_2\|_{\sup}.\end{aligned}$$

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

$$\|\sum_{i=1}^n[\psi_n(T_i;\widetilde{g})K_{Z_i}-E\{\psi_n(T;\widetilde{g})K_Z\}]\|\le(n^{1/2}h^{-(2m-1)/(4m)}+1)(5\log\log n)^{1/2}. \tag{A.8}$$

On the other hand, by Chebyshev's inequality

$$P(A_i^c)=\exp(-(C/C_0)\log n)E\{\exp(\sup_{a\in\mathcal{I}}|\ddot{\ell}_a(Y_i;a)|/C_0)\}\le C_1n^{-C/C_0}.$$

Since $h=o(1)$ and $nh^2\to\infty$, we may choose C to be large 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 $E\{\sup_{a\in\mathcal{I}}|\ddot{\ell}_a(Y_i;a)||Z_i\}\le 2C_1C_0^2$, we have, on B_n, $E\{|\psi(T;d_n\widetilde{g})|^2\}\le 2C_1C_0^2d_n^2$, where expectation is taken with respect to $T=(Y,Z)$. So when n is large, on B_n, by Chebyshev's inequality

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

where the expectation is taken with respect to T_i. Therefore, by (A.8) and on B_n,

$$\begin{aligned}&\|S_n(g+g_0)-S(g+g_0)-(S_n(g_0)-S(g_0))\|\\ =\;&\frac{Cc_m(\log n)h^{-1/2}d_n}{n}\|\sum_{i=1}^n[\widetilde{\psi}_n(T_i;\widetilde{g})K_{Z_i}-E\{\widetilde{\psi}_n(T;\widetilde{g})K_Z\}]\|\\ \le\;&\frac{Cc_m(\log n)h^{-1/2}d_n}{n}(\|\sum_{i=1}^n[\psi_n(T_i;\widetilde{g})K_{Z_i}-E\{\psi_n(T;\widetilde{g})K_Z\}]\|\\ &+n\|E\{\psi_n(T_i;\widetilde{g})K_{Z_i}\}-E\{\widetilde{\psi}_n(T_i;\widetilde{g})K_{Z_i}\}\|)\\ \le\;&\frac{Cc_m(\log n)h^{-1/2}d_n}{n}\cdot[(n^{1/2}h^{-(2m-1)/(4m)}+1)(5\log\log n)^{1/2}+na_n'h^{1/2}d_n^{-1}]\\ \le\;&C'c_ma_n'\log n, \qquad\qquad (A.9)\end{aligned}$$

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

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

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

Therefore,

$$\begin{aligned}
& \|g-S_{n,\lambda}(g_0)\| \\
\leq & \|S_n(g+g_0)-S(g+g_0)-(S_n(g_0)-S(g_0))\|+\|\int_0^1\int_0^1 sD^2S_\lambda(g_0+ss'g)ggdsds'\| \\
\leq & \|S_n(g+g_0)-S(g+g_0)-(S_n(g_0)-S(g_0))\|+\int_0^1\int_0^1 s\|D^2S_\lambda(g_0+ss'g)gg\|dsds'. \qquad \text{(A.10)}
\end{aligned}$$

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

$$\begin{aligned}
\|D^2S_\lambda(g_0+ss'g)gg\|=\|E\{\ell_a'''(Y;(g_0+ss'g)(Z))g(Z)^2K_Z\}\| \leq & E\{E\{\sup_{a\in\mathcal{I}}|\ell_a'''(Y;a)||Z\}g(Z)^2\|K_Z\|\} \\
\leq & C_\ell c_m h^{-1/2}\|g\|^2, \qquad \text{(A.11)}
\end{aligned}$$

where $C_\ell=\sup_{z\in\mathbb{I}}E\{\sup_{a\in\mathcal{I}}|\ell_a'''(Y;a)||Z=z\}$. Thus, from (A.9), (A.10) and (A.11), 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.

A.6. *Proof of Theorem 3.5.* Define $Rem_n=\widehat{g}_{n,\lambda}-g_0^*-\frac{1}{n}\sum_{i=1}^n\epsilon_iK_{Z_i}$. By Theorem 3.4, Rem_n satisfies $\|Rem_n\|=O_P(a_n\log n)$. By assumption $a_n\log n=o(n^{-1/2})$, $\|Rem_n\|=o_P(n^{-1/2})$. Since $E\{\|\sum_{i=1}^n\epsilon_iK_{Z_i}\|^2\}=nE\{\epsilon^2\|K_Z\|^2\}=O(nh^{-1})$, $\|\frac{1}{n}\sum_{i=1}^n\epsilon_iK_{Z_i}\|=O_P((nh)^{-1/2})$. Thus, Rem_n is ignorable compared with $\sum_{i=1}^n\epsilon_iK_{Z_i}$.

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

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

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

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

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

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

A.7. *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 $B_{n1} = \{\max_{1\le i\le n} \sup_{a\in\mathcal{I}} |\ddot{\ell}_a(Y_i; a)| \le C\log n\}$ and $B_{n2} = \{\max_{1\le i\le n} \sup_{a\in\mathcal{I}} |\ell'''_a(Y_i; a)| \le C\log n\}$. Let a_n be defined as in (3.5).

By Taylor expansion,

$$
\begin{aligned}
LRT_{n,\lambda} &= \ell_{n,\lambda}(w_0 + \widehat{g}^0) - \ell_{n,\lambda}(\widehat{g}) \\
&= S_{n,\lambda}(\widehat{g})g + \int_0^1\int_0^1 sDS_{n,\lambda}(\widehat{g} + ss'g)ggdsds' \\
&= \int_0^1\int_0^1 sDS_{n,\lambda}(\widehat{g} + ss'g)ggdsds' \\
&= \int_0^1\int_0^1 s\{DS_{n,\lambda}(\widehat{g} + ss'g)gg - DS_{n,\lambda}(g_0)gg\}dsds' \\
&\quad + \frac{1}{2}(DS_{n,\lambda}(g_0)gg - E\{DS_{n,\lambda}(g_0)gg\}) + \frac{1}{2}E\{DS_{n,\lambda}(g_0)gg\}, \qquad \text{(A.12)}
\end{aligned}
$$

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

We first study I_1. By calculations of the Frechét 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 $B_{n1} \cap B_{n2}$,

$$
\begin{aligned}
&|DS_{n,\lambda}(\widehat{g} + ss'g)gg - DS_{n,\lambda}(g_0)gg| \\
\le\ & \frac{1}{n}C(\log n)\|\widetilde{g}\|_{\sup}\sum_{i=1}^n g(Z_i)^2 \\
=\ & C(\log n)\|\widetilde{g}\|_{\sup}\langle \frac{1}{n}\sum_{i=1}^n g(Z_i)K_{Z_i}, g\rangle \\
=\ & C(\log n)\|\widetilde{g}\|_{\sup}\langle \frac{1}{n}\sum_{i=1}^n g(Z_i)K_{Z_i} - E\{g(Z)K_Z\}, g\rangle + C(\log n)\|\widetilde{g}\|_{\sup}E\{g(Z)^2\}, \qquad \text{(A.13)}
\end{aligned}
$$

where the expectations are taken with respect to Z. Now we look at $\frac{1}{n}\|\sum_{i=1}^n g(Z_i)K_{Z_i} - E\{g(Z)K_Z\}\|$. 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,

$$
\left\|\frac{1}{n}\sum_{i=1}^n [g(Z_i)K_{Z_i} - E\{g(Z)K_Z\}]\right\| = \frac{c_m h^{-1/2} d_n}{n}\left\|\sum_{i=1}^n [\psi_n(T_i; \bar{g})K_{Z_i} - E\{\psi_n(T; \bar{g})K_Z\}]\right\| = O_P(a'_n), \tag{A.14}
$$

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 $a'_n = o(r_n)$,

$$
\begin{aligned}
|DS_{n,\lambda}(\widehat{g} + ss'g)gg - DS_{n,\lambda}(g_0)gg| &= \|\widetilde{g}\|_{\sup}(O_P(a'_n r_n \log n) + O_P(r_n^2 \log n)) \\
&= h^{-1/2} r_n O_P(r_n^2 \log n) \\
&= O_P(r_n^3 h^{-1/2} \log n). \qquad \text{(A.15)}
\end{aligned}
$$

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

Next we study I_2. By an argument similar to (A.9), it can be shown that

$$\frac{1}{n}\|\sum_{i=1}^n \ddot{\ell}_a(Y_i; g_0(Z_i))g(Z_i)K_{Z_i} - E\{\ddot{\ell}_a(Y_i; g_0(Z))g(Z)K_Z\}\| = O_P(a_n' \log n). \tag{A.16}$$

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,

$$n^{1/2}\|w_0 + \widehat{g}^0 - \widehat{g} - S^0_{n,\lambda}(g^0_0) + S_{n,\lambda}(g_0)\| = O_P(n^{1/2} a_n \log n) = o_P(1), \tag{A.17}$$

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,

$$S^0_{n,\lambda}(g^0_0) - S_{n,\lambda}(g_0) = \left(-\frac{1}{n}\sum_{i=1}^n \epsilon_i K_{Z_i}(z_0) + (W_\lambda g_0)(z_0)\right) K_{z_0}/K(z_0, z_0). \tag{A.18}$$

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

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i K_{Z_i}(z_0)/\sqrt{K(z_0, z_0)} - \sqrt{n}(W_\lambda g_0)(z_0)/\sqrt{K(z_0, z_0)} \xrightarrow{d} N(c_{z_0}, c_0). \tag{A.19}$$

It follows by (A.17)–(A.19) that $-2n \cdot LRT_{n,\lambda} \xrightarrow{d} c_0\chi^2_1(c^2_{z_0}/c_0)$, the scaled non-central χ^2 distribution with degree of freedom one 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.

A.8. *Proof of Theorem 5.1.* By Theorem 3.4 and Lemma 3.1,

$$\|\widehat{g} - g^*_0 - \frac{1}{n}\sum_{i=1}^n \epsilon_i K_{Z_i}\|_{\sup} = O_P(a_n h^{-1/2} \log n). \tag{A.20}$$

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 $H_n(z) = H_n(z) - H^b_n(z) - E\{H_n(z) - H^b_n(z)\} + H^b_n(z) -$

$E\{H_n^b(z)\}$. Obviously, on E_n, $H_n(z) - H_n^b(z) = 0$. 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,

$$\sup_{z\in\mathbb{I}} |H_n(z) - H_n^b(z) - E\{H_n(z) - H_n^b(z)\}| = O_P(h^{-1/2}n^{1/2}\exp(-b_n/(2C_1))). \tag{A.21}$$

Denote $R_n(z) = H_n^b(z) - E\{H_n^b(z)\}$, then by (A.21),

$$\sup_{z\in\mathbb{I}} |H_n(z) - R_n(z)| = O_P(h^{-1/2}n^{1/2}\exp(-b_n/(2C_1))). \tag{A.22}$$

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 [53], $\sup_{\epsilon\in\mathbb{R}, z\in\mathbb{I}} |Z_n(\epsilon, z) - W(\tau(\epsilon, z))| = O_P(n^{-1/2}(\log n)^2)$, 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 $V_n(t) = \int_{-b_n}^{b_n} \epsilon d_\epsilon Z_n(\epsilon, t)$ and $V_n^0(t) = \int_{-b_n}^{b_n} \epsilon d_\epsilon W(\tau(\epsilon, t))$. 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 $\sup_{t\in\mathbb{I}} |V_n(t) - V_n^0(t)| = O_P(b_n n^{-1/2}(\log n)^2)$.

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

$$\sup_{z\in\mathbb{I}} |R_n(z) - R_n^0(z)| = O_P(h^{-2}b_n n^{-1/2}(\log n)^2). \tag{A.23}$$

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 $\bar{U}_n^0(t) = \int_{-b_n}^{b_n}\epsilon d_\epsilon B(\tau(\epsilon,t))$. 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 $|\epsilon|$,

$$\begin{aligned}
\sup_{z\in\mathbb{I}}|R_n^0(z)-\bar{R}_n^0(z)| &= |B(1,1)|\cdot\sup_{z\in\mathbb{I}}\left|\int_0^1 K(z,t)\int_{-b_n}^{b_n}\epsilon\, dP_{\epsilon|Z}(\epsilon|t)dP_Z(t)\right| \\
&= |B(1,1)|\cdot\sup_{z\in\mathbb{I}}\left|\int_0^1 K(z,t)E\{\epsilon I(|\epsilon|\le b_n)|Z=t\}dP_Z(t)\right| \\
&= |B(1,1)|\cdot\sup_{z\in\mathbb{I}}\left|\int_0^1 K(z,t)E\{\epsilon I(|\epsilon|> b_n)|Z=t\}dP_Z(t)\right| \\
&\le c_m^2h^{-1}|B(1,1)|E\{|\epsilon|I(|\epsilon|>b_n)\} \\
&= O_P(h^{-1}\exp(-b_n/(2C_1))). \qquad \text{(A.24)}
\end{aligned}$$

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, $\bar{R}_n^0(z) - \tilde{R}_n^0(z) = \bar{U}_n^0(t)\left(h^{-1}\omega((z-t)/h) - K(z,t)\right)\Big|_{t=0}^{1} - \int_0^1\bar{U}_n^0(t)\frac{d}{dt}\left(h^{-1}\omega((z-t)/h)-K(z,t)\right)dt$, which, by assumption (5.1), leads to

$$\sup_{h^\varphi\le z\le 1-h^\varphi}|\bar{R}_n^0(z)-\tilde{R}_n^0(z)| = O_P(h^{-2}b_n\exp(-C_2h^{-1+\varphi})). \qquad \text{(A.25)}$$

By proof of Lemma 3.7 in [22], 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 $I_n(z) = E\{\epsilon^2I(|\epsilon|\le b_n)|Z=z\}$. 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

$$\begin{aligned}
\sup_{z\in\mathbb{I}}\left|\frac{d}{dz}(I(z)-I_n(z))\right| &= \sup_{z\in\mathbb{I}}\left|\int_{|\epsilon|>b_n}\epsilon^2\frac{d}{dz}\pi(\epsilon|z)d\epsilon\right| \\
&\le \sup_{z\in\mathbb{I}}\int_{|\epsilon|>b_n}\epsilon^2\rho_1(1+|\epsilon|^{\rho_2})\pi(\epsilon|z)d\epsilon \\
&= \sup_{z\in\mathbb{I}}\rho_1E\{\epsilon^2(1+|\epsilon|^{\rho_2})I(|\epsilon|>b_n)|Z=z\} = O(\exp(-b_n/(2C_1))).
\end{aligned}$$

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,

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

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

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

for which we have

$$\sup_{z\in\mathbb{Z}} |Y_{0,z}^{(n)} - Y_z^{(n)}| = O_P(h^{-2}\exp(-b_n/(2C_1))). \tag{A.26}$$

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

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

Using the fact 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 $|\omega(s)| \leq C_\omega \exp(-|s|/C_3)$ which implies $|\omega(z/h)| \leq C_\omega \exp(-h^{\varphi-1}/C_3) = O(h)$ 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

$$\sup_{h^\varphi \leq z \leq 1-h^\varphi} |Y_{0,z}^{(n)} - \bar{Y}_{0,z}^{(n)}| = O_P(1). \tag{A.27}$$

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 $\sup_{h^\varphi \leq z \leq 1-h^\varphi} |L_z^{(n)} - \bar{Y}_{0,z}^{(n)}|$. For this purpose, we need the following result.

Lemma A.1. *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 A.1. 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 $\sup_{s\leq 0} \frac{|W(s)|}{(1+s)^\kappa}$ 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

$$
\begin{aligned}
P(D_\kappa > d) &= P\left(\sup_{m\in\mathbb{Z}_+} \sup_{m<s\le m+1} \frac{|W(s)|}{(1+s)^\kappa} > d\right) \\
&\le \sum_{m=0}^{\infty} P\left(\sup_{m<s\le m+1} \frac{|W(s)|}{(1+s)^\kappa} > d\right) \\
&\le \sum_{m=0}^{\infty} P\left(\sup_{m<s\le m+1} |W(s)| > (1+m)^\kappa d\right) \\
&\le \sum_{m=0}^{\infty} P\left(\sup_{0<s\le m+1} |W(s)| > (1+m)^\kappa d\right) \\
&\le \frac{4}{(2\pi)^{1/2}} \sum_{m=0}^{\infty} \frac{\exp(-(d(1+m)^{\kappa-1/2})^2/2)}{d(1+m)^{\kappa-1/2}} \qquad \text{(A.28)} \\
&\le \frac{4}{(2\pi)^{1/2}} \sum_{m=0}^{\infty} \frac{1}{(d(1+m)^{\kappa-1/2})^{\beta+1}} = O(d^{-(\beta+1)}),
\end{aligned}
$$

where (A.28) follows by [29]. 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,

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

On $E_{1,n}$, for any z, $|W(z)| \le d(1+|z|)^\kappa$. 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 $|t| \to \infty$

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

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

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

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

$$\sup_{h^\varphi \leq z \leq 1-h^\varphi} |L_z^{(n)} - \bar{Y}_{0,z}^{(n)}| = O_P(h). \tag{A.29}$$

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 [5], we have as $n \to \infty$,

$$\begin{aligned}
& P\left((2\delta\log n)^{1/2}\left\{\sup_{h^\varphi \leq z \leq 1-h^\varphi} |h^{1/2}L_z^{(n)}I(z)^{-1/2}\sigma_\omega^{-1}| - d_n\right\} \leq u\right) \\
= \; & P\left((2\delta\log n)^{1/2}\left\{\sup_{0 \leq z \leq 1-2h^\varphi} |h^{1/2}L_z^{(n)}I(z)^{-1/2}\sigma_\omega^{-1}| - d_n\right\} \leq u\right) \\
= \; & P\left((2\delta\log n)^{1/2}\left\{\sup_{0 \leq z \leq h^{-1}(1-2h^\varphi)} |h^{1/2}L_{hz}^{(n)}I(hz)^{-1/2}\sigma_\omega^{-1}| - d_n\right\} \leq u\right) \\
\to \; & \exp(-\exp(-2u)),
\end{aligned} \tag{A.30}$$

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 (A.20), (A.22)–(A.29) are all $o_P((h\log n)^{-1/2})$. Thus the desired conclusion holds.

A.9. *Proof of Theorem 5.3.* For simplicity, denote $\widehat{g} = \widehat{g}_{n,\lambda}$ and $g = \widehat{g} - g_0$. Using arguments similar to (A.12), (A.13) and (A.16), 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

$$-2n \cdot PLRT_{n,\lambda} = n\|\widehat{g}-g_0\|^2 + O_P(nr_na_n\log n + nr_n^3h^{-1/2}\log n) = n\|\widehat{g}-g_0\|^2 + o_P(h^{-1/2}). \tag{A.31}$$

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 dominated convergence theorem, it can be established that

$$\|W_\lambda g_0\|^2 = o(\lambda). \tag{A.32}$$

To see (A.32), 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 $g_0 \in H^m(\mathbb{I})$, $\sum_{\nu\in\mathbb{N}} |V(g_0, h_\nu)|^2\gamma_\nu = \int_{\mathbb{N}} f(\nu)dm(\nu) < \infty$, where recall that $\mathbb{N} = \{0, 1, 2, \ldots\}$ and $m(\cdot)$ denotes the discrete measure over $\mathbb{N}$. So f is an integrable function over $\mathbb{N}$ which dominates $f_\lambda(\nu)$. Since $\lim_{\lambda\to 0} f_\lambda(\nu) = 0$, from Lebesgue dominated convergence theorem $\sum_\nu |V(g_0, h_\nu)|^2 \frac{\lambda\gamma_\nu^2}{1+\lambda\gamma_\nu} = \int_{\mathbb{N}} f_\lambda(\nu)dm(\nu) \to 0$. 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

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

where the last equality follows by $\sum_\nu |V(g_0, h_\nu)|^2\gamma_\nu < \infty$ and dominated convergence theorem; see (A.32) 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 [15]). 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 by Proposition 3.2 of [15] 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 (A.5) that $E\{W_{ij}^4\} = 16E\{\epsilon_i^4\epsilon_j^4 K(Z_i, Z_j)^4\} = O(h^{-4})$, implying $G_I = O(n^2h^{-4})$. 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, from direct examinations, we have

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

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 $E\{W_{ij}^2\} = 4E\{\epsilon_i^2\epsilon_j^2 K(Z_i, Z_j)^2\} = 4h^{-1}\rho_K^2$. 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 [15] that

$$\frac{1}{\sqrt{2h^{-1}}n\rho_K}W(n)\xrightarrow{d}N(0,1). \tag{A.33}$$

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 recall that $\sigma_K^2=hE\{\epsilon^2K(Z,Z)\}$. 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,

$$n^{-1}\sum_{i=1}^n\epsilon_i^2K(Z_i,Z_i)=h^{-1}\sigma_K^2+O_P(n^{-1/2}h^{-1})=h^{-1}\sigma_K^2+O_P(1). \tag{A.34}$$

From (A.33) and (A.34), $(h/n)\|\sum_{i=1}^n\epsilon_iK_{Z_i}\|^2=\sigma_K^2+o_P(1)$, implying $n\|S_{n,\lambda}(g_0)\|^2=O_P(h^{-1}+n\lambda+h^{-1/2})=O_P(h^{-1})$, and hence $n^{1/2}\|S_{n,\lambda}(g_0)\|=O_P(h^{-1/2})$. Thus,

$$\begin{aligned}
-2n\cdot PLRT_{n,\lambda} &= n\|\widehat{g}-g_0\|^2+o_P(h^{-1/2})\\
&= \Big(n^{1/2}\|S_{n,\lambda}(g_0)\|+o_P(1)\Big)^2+o_P(h^{-1/2})\\
&= n\|S_{n,\lambda}(g_0)\|^2+2n^{1/2}\|S_{n,\lambda}(g_0)\|\cdot o_P(1)+o_P(h^{-1/2})\\
&= n^{-1}\|\sum_{i=1}^n\epsilon_iK_{Z_i}\|^2+n\|W_\lambda g_0\|^2+o_P(h^{-1/2}). \qquad \text{(A.35)}
\end{aligned}$$

It follows by (A.33)–(A.35) and Slutsky's theorem that $(2h^{-1}\sigma_K^4/\rho_K^2)^{-1/2}(-2nr_K\cdot PLRT_{n,\lambda}-nr_K\|W_\lambda g_0\|^2-h^{-1}\sigma_K^4/\rho_K^2)\xrightarrow{d}N(0,1)$.

A.10. *Proof of Lemma 6.1.* We need the following two inequalities in establishing (6.3):

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

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

REFERENCES

[1] Adams, R. A. (1975). *Sobolev Spaces.* Academic Press, New York-London. Pure and Applied Mathematics, Vol. 65.

[2] Banerjee, M. (2005). Likelihood ratio tests under local and fixed alternatives in monotone function problems. *Scandinavian Journal of Statistics*, **32**, 507–525.

[3] Banerjee, M. (2007). Likelihood based inference for monotone response models. *Annals of Statistics*, **35**, 931–956.

[4] Bickel, P. and Kwon, J. (2001). Inference for semiparametric models: some questions and an answer. *Statistica Sinica*, **11**, 863–960.

[5] Bickel, P. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Annals of Statistics*, *1*, 1071–1095.

[6] Birkhoff, D. (1908). Boundary value and expansion problems of ordinary linear differential equations. *Transactions of the American Mathematical Society*, **9**, 373–395.

[7] Cai, T.T. and Low, M.G. (2004). An adaptive theory for nonparametric confidence interval. *Annals of Statistics*, **32**, 1805–1840.

[8] Chen, J.C. (1994). Testing goodness of fit of polynomial models via spline smoothing techniques. *Statistics & Probability Letters*, **19**, 65–76.

[9] Cheng, G. and Shang, Z. (2012). Joint asymptotics and inferences for semi-nonparametric models. *Purdue Technical Report.*

[10] Claeskens, G. and Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Annals of Statistics*, **6**, 1852–1884.

[11] Coddington, E. A. and Levinson, N. (1987). *Theory of Ordinary Differential Equations.* Tata McGraw-Hill Publishing CO. LTD, New Delhi.

[12] Cox, D. and O'Sullivan, F. (1990). Asymptotic analysis of penalized likelihood and related estimators. *Annals of Statistics*, **18**, 1676–1695.

[13] Cox, D., Koh, E., Wahba, G. and Yandell, B. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Annals of Statistics*, **16**, 113–119.

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

[15] de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probability Theory & Related Fields*, **75**, 261–277.

[16] Donoho, D.L. and Liu, R.C. (1991). Geometrizing rates of convergence, III. *Annals of Statistics*, **19**, 668 - 701.

[17] Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics*, **27**, 715–731.

[18] Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Annals of Statistics*, **1**, 153–193.

[19] Fan, J. and Zhang, J. (2004). Sieve empirical likelihood ratio tests for nonparametric functions. *Annals of Statistics*, **32**, 1858–1907.

[20] Genovese, C. and Wasserman, L. (2008). Adaptive confidence bands. *Annals of Statistics*, **36**, 875–905.

[21] Gu, C. (2002). *Smoothing Spline ANOVA Models.* Springer-Verlag.

[22] Härdle, W. (1989). Asymptotic maximal deviation of M-smoothers. *Journal of Multivariate Analysis*, **29**, 163–179.

[23] Hart, J.D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests.* Springer, New York.

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

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

[26] Ingster, Yu I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives I–III. *Mathematical Methods of Statistics*, **2**, 85–114; **3**, 171–189; **4**, 249–268.

[27] Jayasuriya, B.R. (1996). Testing for Polynomial Regression using Nonparametric Regression Techniques. *Journal of the American Statistical Association*, **91**, 1626–1630.

[28] Johnson, W. P. (2002). The curious history of Faá di Bruno's formula.*American Mathematical Monthly*, **109**, 217–234.

[29] Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd eds. Springer.

[30] Katznelson, Y. (2004). *An Introduction to Harmonic Analysis*, 3rd eds. New York: Cambridge Mathematical Library.

[31] Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer: New York.

[32] Krivobokova, T., Kneib, T. and Claeskens, G. (2010). Simultaneous confidence bands for penalized spline estimators. *Journal of the American Statistical Association*, **105**, 852–863.

[33] Lepski, O. V. and Tsybakov, A. B. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probability Theory & Related Fields*, **117**, 17–48.

[34] Liu, A. and Wang, Y. (2002). Hypothesis testing in smoothing spline models. *Journal of Statistical Computation and Simulation*, **74**, 581–597.

[35] Lorentz, G. G. (1966). *Approximation of Functions.* Holt, Rinehart and Winston, Inc.

[36] Low, M. G. (1997). On nonparametric confidence intervals. *Annals of Statistics*, **25**, 2547–2554.

[37] Mammen, E. and van de Geer, S. (1997). Penalized quasi-likelihood estimation in partial linear models. *Annals of Statistics*, **25**, 1014–1035.

[38] Messer, K. and Goldstein, L. (1993). A new class of kernels for nonparametric curve estimation. *Annals of Statistics*, **21**, 179–195.

[39] Neumann, H. and Polzehl, J. (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *Journal of Nonparametric Statistics*, **9**, 307–333.

[40] Nychka, D. (1988). Bayesian confidence intervals for smoothing splines. *Journal of the American Statistical Association*, **83**, 1134–1143.

[41] Nychka, D. (1995). Splines as local smoothers. *Annals of Statistics*, **23**, 1175–1197.

[42] Pinelis, I. (1994). Optimum bounds for the distributions of martingales in Banach spaces. *Annals of Probability*, **22**, 1679–1706.

[43] Ramil-Novo, L.A. and GonzKalez-Manteiga, W. (2000). F-tests and regression ANOVA based on smoothing spline estimators. *Statistica Sinica*, **10**, 819 - 837.

[44] Rice, J. and Rosenblatt, M. (1983). Smoothing splines: regression, derivatives and deconvolution. *Annals of Statistics*, **11**, 141–156.

[45] Shang, Z. (2010). Convergence rate and Bahadur type representation of general smoothing spline M-Estimates. *Electronic Journal of Statistics*, **4**, 1411–1442.

[46] Shang, Z. and Cheng, G. (2012). Online supplementary to "local and global asymptotic inferences for the smoothing spline estimate".

[47] Shao, J. (2003). *Mathematical Statistics*, 2nd Ed. Springer Texts in Statistics. Springer, New York.

[48] Silverman, B.W. (1984). Spline smoothing: the equivalent variable kernel method. *Annals of Statistics*, **12**, 898 - 916.

[49] Stone, C. (1982). Optimal global rates of convergence for nonparametric regression. *Annals of Statistics*, **10**, 1040–1053.

[50] Stone, M. H. (1926). A comparison of the series of Fourier and Birkhoff. *Transactions of the American Mathematical Society*, **28**, 695–761.

[51] Sun, J. and Loader, C. (1994). Simultaneous confidence bands for linear regression and smoothing. *Annals of Statistics*, **3**, 1328–1345.

[52] Sun, J., Loader, C. and McCormick, W. P. (2000). Confidence bands in generalized linear models. *Annals of Statistics*, **28**, 429–460.

[53] Tusnady, G. (1977). A remark on the approximation of the sample distribution function in the multidimensional case. *Periodica Mathematica Hungarica*, **8**, 53–55.

[54] Utreras, F. I. (1988). Boundary effects on convergence rates for Tikhonov regularization. *Journal of Approximation Theory*, **54**, 235–249.

[55] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer.

[56] Wahba, G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline. *Journal of the Royal Statistical Society, Series B*, **45**, 133-150.

[57] Wahba, G. (1990). *Spline Models for Observational Data*. SIAM, Philidelphia.

[58] Wang, Y. (2011). *Smoothing Splines: Methods and Applications Monographs on Statistics & Applied Probability.* Chapman & Hall/CRC.

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

[60] Zhang, W. and Peng, H. (2010). Simultaneous confidence band and hypothesis test in generalized varying-coefficient models. *Journal of Multivariate Analysis*, **101**, 1656–1680.

[61] Zhou, S., Shen, X. and D. A. Wolfe. (1998). Local asymptotics for regression splines and confidence regions. *Annals of Statistics*, **5**, 1760–1782.

Supplementary materials to

LOCAL AND GLOBAL ASYMPTOTIC INFERENCES FOR THE PENALIZED NONPARAMETRIC ESTIMATE

BY ZUOFENG SHANG AND GUANG CHENG

University of Notre Dame and Purdue University
DECEMBER 27, 2012

In this document, we give the proofs of several results that were not included in Appendix. We also give a minimax rate result of PLRT testing in the more general modeling framework. The reference labels of the equations, Theorems, Propositions and Lemmas in this document are consistent with those in the main text of the paper.

We organize this document as follows. In Section S.1, we prove Proposition 3.3, i.e., the rates of convergence of $\widehat{g}_{n,\lambda}$. In Section S.2, we prove Corollary 3.7 on the pointwise asymptotic normality of $\widehat{g}_{n,\lambda}$ in a special setting. In Section S.3, we sketch the proof of another technical tool, the restricted FBR, which is used to establish the asymptotic null distribution of the local LRT test. In Section S.4, we prove Corollary 4.5. In Section S.5, we prove Theorem 4.6, that is, our local LRT attains minimum separation rates in a general framework. In Section S.6, we prove Proposition 5.2, i.e., the equivalent kernel conditions in cubic spline. In Section S.7, we prove Theorem 5.4, i.e., when data are normal, the PLRT attains minimax rates of testing. We further extend this result to a more general modeling framework in Section S.8.

S.1. *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 $\delta > 0$, $\log N(\delta, \mathcal{G}, \|\cdot\|_{\sup}) \leq C(h\lambda^{-1})^{1/(2m)}\delta^{-1/m}$, where $C > 0$ is an universal constant.*

PROOF OF LEMMA S.1. Note that by $c_m^{-2}h\lambda^{-1} > 1$,

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

where $\mathcal{T} = \{g \in H^m(\mathbb{I}) | \|g\|_{\sup} \leq 1, J(g,g) \leq 1\}$. So by [31],

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

□

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

Lemma S.2. *Suppose that ψ_n satisfies Lipschitz continuity, namely,*

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

where c_m is specified in Lemma 3.1. Then we have

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

where the empirical process $Z_n(f)$ 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 sums 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 $\cap_i A_i$,

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

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.1). Thus, by Lemma S.2, for large n and with large probability,

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

So by Cauchy's inequality,

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

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

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

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

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

To approximate I_2, by Cauchy's inequality we have

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

By changing ψ and ψ_n in the proof of (S.2) 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.2), we have

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

Thus,

$$
\begin{aligned}
2I_2 &= -\|g\|^2 + Cc_mh^{-1+1/(4m)}n^{-1/2}(\log\log n)^{1/2}(\log n)\|g\|^{2-1/(2m)}\\
&\quad + Cc_m(nh^{1/2})^{-1}(\log\log n)^{1/2}(\log n)\|g\| + o_P(1)\|g\|^2. \qquad \text{(S.3)}
\end{aligned}
$$

Note that $E\{\|\sum_{i=1}^n \epsilon_i K_{Z_i}\|^2\} = O(nh^{-1})$, by (A.32) in the main paper, we have

$$
\|S_{n,\lambda}(g_0)\| = O_P((nh)^{-1/2}+\lambda^{1/2}). \qquad \text{(S.4)}
$$

Combining (S.2), (S.3), and (S.4), and by $(nh^{1/2})^{-1}(\log\log n)^{1/2}(\log n) = o((nh)^{-1/2})$, we have for some large C'

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

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.2. *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, that is, 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_νs (see Proposition 2.1), as $h \to 0$,

$$
\begin{aligned}
\int_0^1 g_0^{(2m)}(z)K(z_0,z)dz - g_0^{(2m)}(z_0)/\pi(z_0) &= \sum_\nu \frac{1}{1+\lambda\gamma_\nu}V(g_0^{(2m)}/\pi,h_\nu)h_\nu(z_0) - \sum_\nu V(g_0^{(2m)}/\pi,h_\nu)h_\nu(z_0)\\
&= -\sum_\nu \frac{\lambda\gamma_\nu}{1+\lambda\gamma_\nu}V(g_0^{(2m)}/\pi,h_\nu)h_\nu(z_0) \to 0, \qquad \text{(S.5)}
\end{aligned}
$$

where the limit in (S.5) follows from $\sum_\nu |V(g_0^{(2m)}, h_\nu)h_\nu(z_0)| < \infty$ and dominated convergence theorem. Then, by (3.11) and integration by parts, it can be shown that

$$
\begin{aligned}
(W_\lambda g_0)(z_0) &= \langle W_\lambda g_0, K_{z_0}\rangle = \lambda J(g_0, K_{z_0}) \\
&= (-1)^m h^{2m}\int_0^1 g_0^{(2m)}(z)K(z_0,z)dz = (-1)^m h^{2m}(g_0^{(2m)}(z_0)/\pi(z_0) + o(1)).
\end{aligned}
\tag{S.6}
$$

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 the 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.6) 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 since g_0 does not satisfy the boundary conditions, and by integration by parts, (S.6) should be replaced by the following

$$
(W_\lambda g_0)(z_0) = h^{2m}\sum_{j=1}^m(-1)^{j-1}\left[\left(\frac{\partial^{m-j}}{\partial z^{m-j}}K_{z_0}^{m-j}(z)\right)\cdot g_0^{(m+j-1)}(z)\Big|_0^1\right] + (-1)^m h^{2m}\int_0^1 g_0^{(2m)}(z)K(z_0,z)dz.
\tag{S.7}
$$

The first sum, by (3.14), is $o(h^{2m})$. The second sum, by (S.5), is $(-1)^m h^{2m}(g_0^{(2m)}(z_0)/\pi(z_0) + o(1))$. 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.3. *Proof of Theorem 4.3.* The proof is similar to those in Theorem 3.4, so we only sketch the 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 proper large M. By a modification of the proof of Lemma 3.2, we have the following lemma.

Lemma S.3. *Suppose that ψ_n satisfies Lipschitz continuity, namely, there exists a constant $C_\psi > 0$ such that*

$$
|\psi_n(T; g_1) - \psi_n(T; g_2)| \le c_m^{-1}h^{1/2}\|g_1 - g_2\|_{\sup}, \text{ for all } g_1, g_2 \in \mathcal{H}_0,
\tag{S.8}
$$

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

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

where $\mathcal{G}_0 = \{g \in \mathcal{H}_0 | \|g\|_{\sup} \le 1, J(g,g) \le c_m^{-2}h\lambda^{-1}\}$ 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 a reexamination of the proof of Theorem 3.4, we have, with large probability, $g \in \mathcal{G}_0$ and ψ_n satisfies Lipschitz continuity (S.8), where $\psi_n(T; g) = C^{-1}c_m^{-1}(\log n)^{-1}h^{1/2}d_n^{-1}\{\dot{\ell}_a(Y; g_0(Z) + d_n g(Z)) - \dot{\ell}_a(Y; g_0(Z))\}$, and $d_n = c_m r_n h^{-1/2}$. This leads to, with large probability,

$$
\|\sum_{i=1}^n[\psi_n(T_i; g)K_{Z_i}^* - E\{\psi_n(T; g)K_Z^*\}]\| \le (n^{1/2}h^{-(2m-1)/(4m)} + 1)(5\log\log n)^{1/2}.
\tag{S.9}
$$

The remainder of the proof follows by (A.7), and by an argument similar to (A.9) – (A.11).

S.4. *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.5. *Proof of Theorem 4.6.* It is easy to check that $h \asymp n^{-d}$ with $1/(2m+1) \le d < 2m/(10m-1)$ and $m > 1+\sqrt{3}/2$ satisfies all the rate conditions on a_n and h stated in Theorem 4.3. Before formally giving the proof, we establish the contiguity between $P^n_{g_{n0}}$ and $P^n_{g_*}$. It can be shown that the log likelihood ratio

(S.10)
$$\log(P^n_{g_{n0}}/P^n_{g_*}) = n^{-1/2}\sum_{i=1}^n \dot\ell_a(Y_i; g_*(Z_i))n^{1/2}\eta_n f_n(Z_i) + (1/2)n^{-1}\sum_{i=1}^n \ddot\ell_a(Y_i; g_*(Z_i))n\eta_n^2|f_n(Z_i)|^2 + o_{P^n_{g_*}}(1).$$

Thus, under $g = g_*$ and using (4.13), $P^n_{g_{n0}}/P^n_{g_*} \xrightarrow{d} \exp(\xi)$, where $\xi \sim N(-\tau_{z_0}^2/2, \tau_{z_0}^2)$. Since $E\{\exp(\xi)\} = 1$, by Theorem 3.10.2 of [55], $P^n_{g_{n0}}$ is contiguous with respect to $P^n_{g_*}$.

Next we prove the theorem. For notational convenience, denote $\widehat{g} = \widehat{g}_{n,\lambda}$ and $\widehat{g}^0 = \widehat{g}^0_{n,\lambda}$. Under $g = g_*$, since $g_*(z_0) = w_0$, $H_0 : g(z_0) = w_0$ automatically holds. It then follows from Assumptions A.3 and A.4, and the proof of Theorem 4.3 that under $g = g_*$, $-2n\cdot LRT_{n,\lambda} = n\|w_0 + \widehat{g}^0 - \widehat{g}\|^2 + o_{P^n_{g_*}}(1)$. Applying Theorems 3.4 and 4.3, we have $-2n\cdot LRT_{n,\lambda} = n\|S_{n,\lambda}(g_*) - S^0_{n,\lambda}(g^0_*)\|^2 + o_{P^n_{g_*}}(1)$, under $g = g_*$, where recall $g^0_* = g_* - w_0$. Also, under $g = g_*$, $\frac{1}{n}\sum_{i=1}^n\left(\dot\ell_a(Y_i; g_*(Z_i))K_{z_0}(Z_i) - E\{\dot\ell_a(Y; g_*(Z))K_{z_0}(Z)\}\right) = O_{P^n_{g_*}}((nh)^{-1/2})$. By contiguity between $P^n_{g_{n0}}$ and $P^n_{g_*}$, under $g = g_{n0}$, $-2n\cdot LRT_{n,\lambda} = n\|S_{n,\lambda}(g_*) - S^0_{n,\lambda}(g^0_*)\|^2 + o_{P^n_{g_{n0}}}(1)$, and $\frac{1}{n}\sum_{i=1}^n\left(\dot\ell_a(Y_i; g_*(Z_i))K_{z_0}(Z_i) - E\{\dot\ell_a(Y; g_*(Z))K_{z_0}(Z)\}\right) = O_{P^n_{g_{n0}}}((nh)^{-1/2})$.

On the other hand, using $hK(z_0, z_0) \asymp \sigma^2_{z_0}$ and a direct examination leads to

$$\begin{aligned} n\|S_{n,\lambda}(g_*) - S^0_{n,\lambda}(g^0_*)\|^2 &= \frac{n}{K(z_0,z_0)}\Big|\frac{1}{n}\sum_{i=1}^n \dot\ell_a(Y_i; g_*(Z_i))K_{z_0}(Z_i) - (W_\lambda g_*)(z_0)\Big|^2 \\ &\asymp \sigma_{z_0}^{-2}\Big|\frac{(nh)^{1/2}}{n}\sum_{i=1}^n \dot\ell_a(Y_i; g_*(Z_i))K_{z_0}(Z_i) - (nh)^{1/2}(W_\lambda g_*)(z_0)\Big|^2. \qquad \text{(S.11)} \end{aligned}$$

Note under $g = g_{n0}$, by assumptions of the theorem and Taylor's expansion,

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^n \dot\ell_a(Y_i; g_*(Z_i))K_{z_0}(Z_i) \\ =\ & \frac{1}{n}\sum_{i=1}^n\left(\dot\ell_a(Y_i; g_*(Z_i))K_{z_0}(Z_i) - E\{\dot\ell_a(Y; g_*(Z))K_{z_0}(Z)\}\right) + E\left\{\left(\dot\ell_a(Y; g_*(Z)) - \dot\ell_a(Y; g_{n0}(Z))\right)K_{z_0}(Z)\right\} \\ =\ & O_{P^n_{g_{n0}}}((nh)^{-1/2}) + E\{\int_0^1 \ddot\ell_a(Y; g_{n0}(Z) - s\eta_n f_n(Z))\eta_n f_n(Z)K_{z_0}(Z)ds\} \\ =\ & O_{P^n_{g_{n0}}}((nh)^{-1/2}) + E\{\int_0^1 [\ddot\ell_a(Y; g_{n0}(Z) - s\eta_n f_n(Z)) - \ddot\ell_a(Y; g_{n0}(Z))]\eta_n f_n(Z)K_{z_0}(Z)ds\} - \eta_n V(f, K_{z_0}) \\ =\ & O_{P^n_{g_{n0}}}((nh)^{-1/2}) + \eta_n^2 h^{-1}\|f_n\|^2_{L_2}O(1) - \eta_n f_n(z_0) + \eta_n(W_\lambda f_n)(z_0). \end{aligned}$$
(S.12)

Since $n\eta_n^2 V(f_n, f_n) = O(1)$ under $g = g_*$, which also holds under $g = g_{n0}$ by contiguity between $P_{g_{n0}}^n$ and $P_{g_*}^n$, we have $\eta_n^2 h^{-1}\|f_n\|_{L_2}^2 = O((nh)^{-1}) = O((nh)^{-1/2})$. Since $J(f_n, f_n) \le C_a(n\lambda\eta_n^2)^{-1}$, by Fourier expansion and Cauchy's inequality, it can be shown that

$$\eta_n|(W_\lambda f_n)(z_0)| \le \eta_n\sqrt{J(f_n, f_n)}(\lambda h^{-1})^{1/2}O(1) = O((nh)^{-1/2}).$$

Therefore, $\frac{1}{n}\sum_{i=1}^n \dot{\ell}_a(Y_i; g_*(Z_i))K_{z_0}(Z_i) = O_{P_{g_{n0}}^n}((nh)^{-1/2}) - n^{-1/2}f_n(z_0)$. Since $\sum_\nu |V(g_*, h_\nu)|\gamma_\nu^{1/2} < \infty$ implying $(W_\lambda g_*)(z_0) = O(h^m)$ (see Remark 4.1), $(nh)^{1/2}(W_\lambda g_*)(z_0) = O((nh)^{1/2}h^m)$. By assumption $\eta_n \ge (nh)^{-1/2} + h^m$ and $|f_n(z_0)| \to \infty$ as $n \to \infty$, and by (S.11) and (S.12), the leading term in the approximation of $-2n \cdot LRT_{n,\lambda}$ is $\sigma_{z_0}^{-2}(nh\eta_n^2)|f_n(z_0)|^2$ which goes to infinity as $n \to \infty$. Therefore, there exists some sufficiently large N such that for any $n \ge N$, under $g = g_{n0}$, $-2n \cdot LRT_{n,\lambda} > c_\alpha$ with probability (in terms of $P_{g_{n0}}^n$) greater than $1 - \delta$, where c_α is the α-cutoff associated with the limiting distribution described in Theorem 4.4. Balancing the lower bound of η_n one obtains when $h = h^*$ the minimum rate $\eta_n = n^{-m/(2m+1)}$ is achieved.

To show $n^{-m/(2m+1)}$ is the sharp lower bound for η_n, assume otherwise $\eta_n \ll n^{-m/(2m+1)}$. Let ω be a function defined over $\mathbb{R}$ satisfying $\omega(0) = 1$, ω and $\omega^{(m)}$ are square integrable. Since $\eta_n \ll n^{-m/(2m+1)}$ and $(nh)^{1/2} = O(n^{m/(2m+1)})$, we have $\eta_n(nh)^{1/2} = o(1)$. Choose c_n such that, as $n \to \infty$, $c_n \to \infty$, $c_n\eta_n n^{1/2} \to \infty$ and $c_n\eta_n(nh)^{1/2} = o(1)$. For instance, one can choose $c_n = \max\{(\eta_n n^{1/2})^{-1}h^{-1/4}, [\eta_n(nh)^{1/2}]^{-1/2}\}$. Define $f_n(z) = c_n\omega(c_n^2 n\eta_n^2(z - z_0))$ for $z \in \mathbb{I}$. By a direct calculation, $J(f_n, f_n) = c_n^2(c_n^2 n\eta_n^2)^{2m}\int_0^1 |\omega^{(m)}(c_n n\eta_n^2(z - z_0))|^2 dz = O(c_n^{4m}(n\eta_n^2)^{2m-1})$. Since $c_n\eta_n(nh)^{1/2} = o(1)$, we have $J(f_n, f_n) = o(\eta_n^{-4m}(nh)^{-2m}(n\eta_n^2)^{2m-1}) = o((n\lambda\eta_n^2)^{-1})$. Clearly, $f_n(z_0) = c_n \to \infty$. Since $c_n\eta_n n^{1/2} \to \infty$, we have under $g = g_*$,

$$\begin{aligned} n\eta_n^2 V(f_n, f_n) &= c_n^2 n\eta_n^2 \int_0^1 I(z)\pi(z)|\omega(c_n^2 n\eta_n^2(z - z_0))|^2 \\ &= \int_{-c_n^2 n\eta_n^2 z_0}^{c_n^2 n\eta_n^2(1-z_0)} (I\pi)(z_0 + (c_n^2 n\eta_n^2)t)|\omega(t)|^2 dt \\ &\to I(z_0)\pi(z_0)\int_{\mathbb{R}} |\omega(t)|^2 dt, \end{aligned}$$

where recall that π is the density of Z and $I(z) = -E\{\ddot{\ell}_a(Y; g_*(Z))|Z = z\}$. Therefore, f_n satisfies (4.13). Following (S.10) and the arguments below, it can be shown that $P_{g_*}^n$ and $P_{g_{n0}^n}$ are contiguous. Then by the proofs of (S.11) and (S.12), under $g = g_{n0} = g_* + \eta_n f_n$, $-2n \cdot LRT_{n,\lambda} = (hK(z_0, z_0))^{-1}|O_{P_{g_{n0}}^n}(1) + (nh)^{1/2}\eta_n f_n(z_0) + O((nh^{2m+1})^{1/2})|^2 + o_{P_{g_{n0}}^n}(1)$. Since $h \asymp n^{-d}$ with $d \in [1/(2m+1), 2m/(10m-1)]$, $(nh^{2m+1})^{1/2} = O(1)$. Note when $n \to \infty$, $(nh)^{1/2}\eta_n f_n(z_0) = c_n\eta_n(nh)^{1/2} = o(1)$ eventually vanishes. So $-2n \cdot LRT_{n,\lambda} = O_{P_{g_{n0}}^n}(1)$. This means, $-2n \cdot LRT_{n,\lambda} > c_\alpha$ with probability (in terms of $P_{g_{n0}}^n$) bounded by $1 - \delta_0$ for some $\delta_0 \in (0, 1)$ unrelated to n. This proves the sharpness of the lower bound $n^{1/(2(2m+1))}$ for η_n.

S.6. *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 $\tilde{h}_\nu$s and $\tilde{\gamma}_\nu$s be the normalized (with respect to the usual L_2-norm) eigenfunctions and eigenvalues of the boundary value problem $(-1)^m\tilde{h}_\nu^{(2m)} = \tilde{\gamma}_\nu\tilde{h}_\nu$, $\tilde{h}_\nu^{(j)}(0) = \tilde{h}_\nu^{(j)}(1) = 0$, $j = m, m+1, \ldots, 2m-1$. Thus, it is easy to see that $h_\nu = \sigma\tilde{h}_\nu$ and $\gamma_\nu = \sigma^2\tilde{\gamma}_\nu$ satisfy (2.11) with $\pi(z)I(z) \equiv \sigma^{-2}$, implying that h_νs and γ_νs form an effective eigensystem in $H^m(\mathbb{I})$. 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 [41], with the penalty parameter therein replaced by $\lambda^\dagger$.

Next we restrict $m = 2$. By Theorem 4.1 in [38], for $j = 0,1$,

$$\sup_{s,t\in\mathbb{I}}\left|\frac{d^j}{dt^j}\left(\widetilde{K}(s,t)-\bar{K}(s,t)\right)\right| \le C_K'(h^\dagger)^{-(j+1)}\exp(-\sin(\pi/(2m))/h^\dagger), \tag{S.13}$$

where by equation (6) in [38], $\bar{K}$ satisfies for any $s,t\in\mathbb{I}$ and $j=0,1$,

$$\left|\frac{d^j}{dt^j}\left(\bar{K}(s,t)-\frac{1}{h^\dagger}\omega_0\left(\frac{s-t}{h^\dagger}\right)\right)\right| \le C_K''(h^\dagger)^{-(j+1)}(\exp(-|1-s|/(\sqrt{2}h^\dagger))+\exp(-|s|/(\sqrt{2}h^\dagger))), \tag{S.14}$$

with C_K', C_K'' both being positive constants. By (S.13) and (S.14), it is easy to see that for any $s,t\in\mathbb{I}$ and $j=0,1$,

$$\begin{aligned}&\left|\frac{d^j}{dt^j}\left(\widetilde{K}(s,t)-\frac{1}{h^\dagger}\omega_0\left(\frac{s-t}{h^\dagger}\right)\right)\right| \\ \le\ & C_K(h^\dagger)^{-(j+1)}(\exp(-\sin(\pi/(2m))/h^\dagger)+\exp(-|1-s|/(\sqrt{2}h^\dagger))+\exp(-|s|/(\sqrt{2}h^\dagger))),\end{aligned} \tag{S.15}$$

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.15), Condition (5.1) holds.

S.7. *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

$$-2n\cdot PLRT_{n,\lambda} = -2n(\ell_{n,\lambda}(g_0)-\ell_{n,\lambda}(g_{n0})) - 2n(\ell_{n,\lambda}(g_{n0})-\ell_{n,\lambda}(\widehat{g}_{n,\lambda})). \tag{S.16}$$

The proof proceeds by 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 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). Second, 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 $(2u_n)^{-1/2}(-2nr_K\cdot PLRT - u_n) \ge n(2u_n)^{-1/2}\|g_n\|^2(1+O_P(n^{-1/2}\|g_n\|^{-1} + n^{-1/2}+\lambda\|g_n\|^{-2})) + (2u_n)^{-1/2}n\|W_\lambda g_0\|^2 + O_P(1) \ge n(2u_n)^{-1/2}\|g_n\|^2(1+O_P(n^{-1/2}\|g_n\|^{-1}+n^{-1/2}+\lambda\|g_n\|^{-2})) + O_P(1)$, where $O_P(\cdot)$ holds uniformly for $g_{n0}\in\mathcal{G}_a$. Let $n^{-1/2}\|g_n\|^{-1}\le 1/C$, $\lambda\|g_n\|^{-2}\le 1/C$ and $\|g_n\|^2 \ge C(nh^{1/2})$ for sufficiently large C, which implies that $|\frac{-2nr_K\cdot PLRT-u_n}{(2u_n)^{1/2}}| \ge c_\alpha$ 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 we have to assume $\|g_n\|^2 \ge C(\lambda + (nh^{1/2})^{-1})$ to achieve large power.

Next we complete the above two parts. First, 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

$$\inf_{n\ge N}\inf_{g_{n0}\in\mathcal{G}_a} P_{g_{n0}}\left(\|\widehat{g}_{n,\lambda}-g_{n0}-S_{n,\lambda}(g_{n0})\| \le \widetilde{C}a_n\right) \ge 1-\delta, \tag{S.17}$$

where recall that a_n is defined as in (3.5), The proof of (S.17) 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 $B_{n1} \cap B_{n2}$, 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 $g_n \in \mathcal{G}_a$, (A.9) holds with probability greater than $1-\delta/2$ (by properly tunning the probability), with the constant C' therein only depending on C, M, c_m. By going through the proofs of (A.10) and (A.11), 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.17) 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 $g_n \in \mathcal{G}_a$.

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

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

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,

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

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.8. *Minimax separation rates of PLRT test in general modeling framework.* To the end of this supplement document, we remark that in a more general modeling framework PLRT achieves the optimal minimax rate of hypothesis testing specified in Ingster (1993). The proofs are similar to those of Theorem 5.4 but requires a deeper tachnical tool, i.e., the mapping principle which builds equivalence between the eigenvalues obtained under null and contiguous alternatives. We still write the local alternative as $H_{1n} : g = g_{n0}$, where $g_{n0} = g_0 + g_n$, $g_0 \in H^m(\mathbb{I})$ and g_n belongs to some alternative value set $\mathcal{G}_a$.

THEOREM 6.2. *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 $\mathcal{I}_0 \subset \mathcal{I}$. 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 values of $2g_0$ belong 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 uniformly over $g_{n0} \in \mathcal{G}_a$, $\|\widehat{g}_{n,\lambda} - g_{n0}\| = O_P(r_n)$ holds under $H_{1n} : g = g_{n0}$. Then for any $\delta \in (0,1)$, there exist positive constants C' and N such that

$$\inf_{n \geq N} \inf_{\substack{g_n \in \mathcal{G}_a \\ \|g_n\| \geq C'\eta_n}} P\left(\text{reject } H_0^{global} | H_{1n} \text{ is true}\right) \geq 1 - \delta, \tag{S.18}$$

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 6.2. 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. Throughout, we only consider $g_{n0} = g_0 + g_n$ for $g_n \in \mathcal{G}_a$.

Next we prove our theorem. We write

$$-2n \cdot PLRT_{n,\lambda} = -2n(\ell_{n,\lambda}(g_0) - \ell_{n,\lambda}(g_{n0})) - 2n(\ell_{n,\lambda}(g_{n0}) - \ell_{n,\lambda}(\widehat{g}_{n,\lambda})). \tag{S.19}$$

The proof proceeds by 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 $f \in H^m(\mathbb{I})$, by Assumption A.1 (a) and (b)

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

where $\zeta_0 = C_0C_1C_2 = 1/(2\zeta)$ is a universal constant. Therefore, $(1 - \zeta_0\|g_n\|_{\sup})V_{g_{n0}}(f,f) \le V_{g_0}(f,f) \le (1 + \zeta_0\|g_n\|_{\sup})V_{g_{n0}}(f,f)$. By mapping principle (see Theorem 6.1 in [59]), the eigenvalues induced by the functional pairs $(V_{g_{n0}}, J)$ and (V_{g_0}, J) are thus equivalent in the sense that $(1 - \zeta_0\|g_n\|_{\sup})\gamma_\nu^{n0} \le \gamma_\nu \le (1 + \zeta_0\|g_n\|_{\sup})\gamma_\nu^{n0}$ for any $\nu \in \mathbb{N}$, where γ_ν^{n0} denotes the eigenvalue corresponding to $V_{g_{n0}}$ and γ_ν is the eigenvalue corresponding to V_{g_0}. Therefore, uniformly for g_{n0},

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

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

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

where $O_P(\cdot)$ holds uniformly for $g_n \in \mathcal{G}_a$. Let $n^{-1/2}\|g_n\|^{-1} \le 1/C$, $\lambda\|g_n\|^{-2} \le 1/C$, $Ch^{-1}\|g_n\| \le nh^{1/2}\|g_n\|^2$, and $\|g_n\|^2 \ge C(nh^{1/2})^{-1}$ for sufficiently large C, which implies that $|\frac{-2nr_K \cdot PLRT - u_n}{(2u_n)^{1/2}}| \ge c_\alpha$ 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 $\|g_n\|^2 \ge C(\lambda + (nh^{1/2})^{-1})$ to achieve large power.

Next we complete the above two parts. First, 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

$$\inf_{n\ge N} \inf_{g_n \in \mathcal{G}_a} P_{g_{n0}}\left(\|\widehat{g}_{n,\lambda} - g_{n0} - S_{n,\lambda}(g_{n0})\| \le \widetilde{C}a_n\right) \ge 1 - \delta, \tag{S.20}$$

where recall that a_n is defined as in (3.5), The proof of (S.20) 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 $B_{n1} \cap B_{n2}$, 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 $g_n \in \mathcal{G}_a$, (A.9) holds with probability greater than $1 - \delta/2$ (by properly tunning the probability), with the constant C' therein only depending on C, M, c_m. By going through the proofs of (A.10) and (A.11), 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.20) 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

$$E\{|\sum_{i=1}^{n}[R_i - E(R_i)]|^2\} \leq nE\{R_i^2\} = nE\{|-\epsilon_i g_n(Z_i) + \ddot{\ell}_a(Y_i; g_{n0}^*(Z_i))g_n(Z_i)^2|^2\}, \tag{S.21}$$

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.21) 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.

UNIVERSITY OF NOTRE DAME
NOTRE DAME, IN 46545
EMAIL: ZSHANG@ND.EDU

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
250 N. UNIVERSITY STREET
WEST LAFAYETTE, IN 47906
EMAIL: CHENGG@STAT.PURDUE.EDU