Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Journal of Statistical Planning and Inference 126 (2004) 73–95

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Assessing the equivalence of nonparametric regression tests based on spline and local polynomial smoothers

Chunming Zhang*

Department of Statistics, University of Wisconsin, Madison, WI 53706-1685, USA

Received 28 August 2002; accepted 26 July 2003

Abstract

It is widely known that, in a certain sense, a smoothing spline estimate of the regression function is asymptotically equivalent to a kernel regression estimate. However, little information has been available about the equivalence between nonparametric regression tests, based on the smoothing spline and local polynomial regression methods. To assess their relative behaviors and to facilitate illustration, we consider in this paper the "generalized likelihood ratio" (GLR) test statistic, constructed from each type of smoother. For the resulting test statistics, we first establish their equivalent forms of the asymptotic distributions, under the null hypothesis. After that, we derive their equivalent optimal rates of smoothing parameters for nonparametric testing. Furthermore, we evaluate their relative asymptotic efficiency and characterize its relation to the magnitude of the smoothing parameters. Finally, we illustrate the large-sample behaviors of the cubic smoothing spline and local linear regression methods, in the GLR tests, with results from small-scale simulations.

MSC: primary 62G10; secondary 62G05; 62G20

Keywords: Generalized likelihood ratio test; Goodness-of-fit; Local polynomial regression; Smoothing parameter; Smoothing spline

1. Introduction

Model assessment forms a crucial component of statistical analysis. Driven by the rapid advances in computing technology, nonparametric regression techniques have

* Tel.: +1-608-2620084; fax: +1-608-2620032.
E-mail address: cmzhang@stat.wisc.edu (C. Zhang).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2003.07.013

emerged as powerful tools for assessing model agreement. In this article we consider an important special case of the general problem–testing the mean function in a nonparametric regression model.

Among the existing methods for mean estimation, smoothing splines have received considerable research efforts in the literature on nonparametric specification tests. The work includes Yanagimoto and Yanagimoto (1987), Cox et al. (1988), Cox and Koh (1989), Eubank and Spiegelman (1990), Eubank and LaRiccia (1993), Jayasuriya (1996), Ramil-Novo and González-Manteiga (2000), and many others. Meanwhile, the local polynomial regression, as an alternative method of curve fitting, has also been under impressive development in those tests of Azzalini et al. (1989), Hart and Wehrly (1992), Azzalini and Bowman (1993), Härdle and Mammen (1993), Young and Bowman (1995), and Bowman and Young (1996), among others. These work together provides users with flexible choices of nonparametric diagnostic tools. However, a research issue regarding the effect of the curve fitting methods on the testing procedures has not yet been addressed in the literature.

In particular, a number of relevant questions have frequently arisen including the following. (i) Whether the associated nonparametric regression tests, based on the smoothing spline and local polynomial estimation techniques, have equivalent forms of the asymptotic null distributions? (ii) Since the performance of each type of smoothing method relies on a specific choice of smoothing parameter, how to determine the optimal selection of smoothing parameters for the resulting tests? (iii) Whether the associated nonparametric tests perform in similar manners against alternatives? or which fitting method will lead to a more effective test to detect departures from the null hypothesis? (iv) How to determine the optimal kernel function for local polynomial-based nonparametric regression tests? Given the wide application domain of the smoothing spline and local polynomial fitting methods in many scientific disciplines nowadays, these kinds of issues are not only important for correctly understanding their relative performances in nonparametric inference, but also helpful in offering valuable guidance on practical implementation of these techniques.

Addressing the questions raised above is a nontrivial task, however. From some previous results of equivalence, established in the setting of nonparametric function estimation, it may be deduced only that, in a certain sense, a smoothing spline estimate corresponds approximately to a kernel regression estimate (Cox, 1983; Silverman, 1984a), and that conversely, the kernel regression and local linear regression estimators can be interpreted as equivalent roughness-penalty estimators (Huang, 2001); none of these study results has immediate implications for the foregoing issues. In this paper, to facilitate illustration, we will consider the generalized likelihood ratio (abbreviated as GLR) testing procedure, introduced in Fan et al. (2001), although the basic ideas of the theoretical work, developed in later sections, apply to other kinds of nonparametric regression tests. For expositional convenience, we call $\mathscr{G}_{n,L}$ the GLR test based on the local polynomial regression estimator, and $\mathscr{G}_{n,S}$ the GLR test based on the smoothing spline estimator. Curiously, the simulation study in Section 5, which is designed to compare the power of the $\mathscr{G}_{n,S}$ test based on the cubic smoothing spline method and that of the $\mathscr{G}_{n,L}$ test based on the local linear regression method, tends to suggest the superiority of the cubic smoothing spline over the local linear method, despite of the fact that choices of smoothing parameters are taken to be equivalent for estimating the mean function. Actually, conclusions based on this suggestion is incomplete and would be misleading. Nevertheless, little published information exists to explain this empirical comparison, and hence a more careful study is needed.

In this paper, three major parts of equivalence results will be investigated. In response to the first question above, we will establish explicitly the equivalence between the asymptotic null distributions of $\mathscr{G}_{n,S}$ and $\mathscr{G}_{n,L}$. To address the second issue, we will derive the optimal rate of the smoothing parameter for the $\mathscr{G}_{n,S}$ test. The "optimal" here is defined in the sense, such that the contiguous alternatives with the fastest rate of convergence to the null can be detected consistently. As we shall demonstrate in Section 3.2, this optimal rate is asymptotically equivalent to the optimal rate of bandwidth, obtained in Fan et al. (2001) from the $\mathscr{G}_{n,L}$ test. To our knowledge, theoretical investigations of the optimal rate of smoothing parameter, in the context of spline-based tests, have been largely overlooked in the literature. This paper fills that gap in the literature. Moreover, we put forward an empirical method for obtaining the optimum smoothing parameter. Regarding the third question, our derivations indicate that the $\mathscr{G}_{n,S}$ and $\mathscr{G}_{n,L}$ tests, have the asymptotic normal distributions under the null and local alternatives. In this case, the Pitman-type asymptotic relative efficiency of the two tests can be evaluated. This result, based on comparing the magnitude of the smoothing parameters, will be particularly helpful for us to know, for e.g., when the $\mathscr{G}_{n,S}$ test is asymptotically more powerful than the $\mathscr{G}_{n,L}$ test and when it is not.

The establishment of equivalence results has several useful consequences. For instance, translating the power expression of the $\mathscr{G}_{n,S}$ test into that of the $\mathscr{G}_{n,L}$ test indicates that the optimal choice of kernel function solves a constrained variational problem. Formulating this variational criterion, within the context of the $\mathscr{G}_{n,L}$ test itself, will be unnecessarily complicated without applications of the equivalence result.

Nonparametric regression model serves as a building block for various complicated statistical models. It is thus anticipated that the basic conclusions of this article can be extended to more general settings, such as the varying coefficient model (Hastie and Tibshirani, 1993).

This paper is organized as follows. Section 2 briefly reviews relevant aspects of the $\mathscr{G}_{n,L}$ statistic. Section 3 presents the main results on the asymptotic distribution, power, and the optimal rate of smoothing parameter of the $\mathscr{G}_{n,S}$ test. The relative power of the two tests is examined in Section 4, followed by the optimal choice of kernel function. Section 5 reports simulation studies. Section 6 summarizes the paper and outlines some possible extensions. Proofs are collected in the Appendix.

2. Background

2.1. The GLR test

To begin with, we first briefly outline the GLR test proposed in Fan et al. (2001). Suppose that we are given independent observations, $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, from a nonparametric regression model

$$Y = m(X) + \varepsilon \tag{2.1}$$

in which the error ε, conditional on the design variable X, has a normal distribution with mean zero and unknown variance σ^2. The mean regression function, $m(x) = E(Y|X = x)$, is assumed to belong to a smooth class of functions $\mathscr{M}$; a parametric form is not assumed. For a fixed integer $p \geqslant 1$, let $\Pi_{p-1} = \{\theta_0 + \theta_1 x + \cdots + \theta_{p-1}x^{p-1}\colon \boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_{p-1})^{\mathrm{T}} \in \mathbb{R}^p\}$ denote the set of polynomial regression functions of degree $p-1$, where the superscript T stands for the transpose of a vector or matrix. Suppose we are interested in testing

$$H_0 : m(x) \in \Pi_{p-1} \quad \text{versus} \quad H_1 : m(x) \in \mathscr{M} \setminus \Pi_{p-1}. \tag{2.2}$$

To derive the GLR statistic, consider from (2.1) the conditional log-likelihood function, expressed as

$$\ell_n = -n \log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \{Y_i - m(X_i)\}^2.$$

Let $m_{\hat{\boldsymbol{\theta}}}(\cdot)$ stand for the maximum likelihood estimator (MLE) under H_0, in which $\hat{\boldsymbol{\theta}}$ denotes MLE of the unknown vector of parameters $\boldsymbol{\theta}$. Usually, the MLE of $m(\cdot)$ may not exist under H_1. In such instances, one could carry out the nonparametric fit, denoted by $\hat{m}$. Denote by RSS_0 the residual sum of squares under H_0, and by RSS_1 under H_1; that is, $\mathrm{RSS}_0 = \sum_{i=1}^{n} \{Y_i - m_{\hat{\boldsymbol{\theta}}}(X_i)\}^2$, and $\mathrm{RSS}_1 = \sum_{i=1}^{n} \{Y_i - \hat{m}(X_i)\}^2$. Then the logarithm of the conditional nonparametric likelihood ratio statistic for (2.2), given by

$$\mathscr{G}_n = \ell_n(H_1) - \ell_n(H_0) = (n/2)\log\{\mathrm{RSS}_0/\mathrm{RSS}_1\} \tag{2.3}$$

is called a GLR statistic. It is asymptotically equivalent to the Azzalini and Bowman (1993) statistic. We make another remark here; that is, even if we drop the normality assumption in (2.1), $\mathscr{G}_n$ itself, as a valid statistic, can still be utilized to assess the goodness-of-fit of a polynomial regression.

Remark 1. Testing the polynomial regression in the null hypothesis (2.2) is only for the sake of technical simplicity.

2.2. The GLR test based on local polynomial regression estimation

We now start by describing briefly the case where $\hat{m}$ employed in the GLR statistic is the pointwise local polynomial regression estimate. Assume that we wish to estimate the regression curve at a fitting point x_0. According to the Taylor's expansion, any smooth function $m(x)$ with the $q+1$th derivative, around a neighborhood of x_0, can be locally approximated by a qth degree polynomial

$$m(x) \approx m(x_0) + (x - x_0)m^{(1)}(x_0) + \cdots + (x - x_0)^q m^{(q)}(x_0)/q!.$$

Denote $\beta_j = m^{(j)}(x_0)/j!,\ j = 0, \ldots, q$, where the dependence of β_j's on x_0 is suppressed. Then the local polynomial regression estimates $\hat{\beta}_j$, of degree q, are defined to be the minimizers of the weighted residual sum of squares,

$$\sum_{i=1}^{n}\{Y_i-\beta_0-(X_i-x_0)\beta_1-\cdots-(X_i-x_0)^q\beta_q\}^2K_h(X_i-x_0) \tag{2.4}$$

with respect to values of β_j. Here the weight function $K_h(\cdot)=K(\cdot/h)/h$ is re-scaled from a kernel function $K(\cdot)$, and the smoothing parameter, $h>0$, is called the bandwidth which governs the local amount of data smoothing. The resulting $\hat{\beta}_0$ gives the qth degree local polynomial regression estimate $\hat{m}_h(x_0)$. In particular, the conventional kernel regression and local linear regression estimates correspond to locally fitting a constant function (with degree $q=0$) and a straight line (with degree $q=1$), respectively.

Fan et al. (2001) showed the Wilks-type asymptotic null distribution of the $\mathscr{G}_{n,L}(h)$ statistic, in which the local polynomial estimator $\hat{m}_h$ is used. Their results are applicable to both random- and fixed-design variables, as well as non-Gaussian random errors. Namely, assume that $E(\varepsilon|X)=0$ and $E(\varepsilon^2|X)=\sigma^2$ in (2.1), and suppose that Condition (A) in the Appendix holds. Define by $|\Omega|$ the length of the support of the design variable X, and let $\xrightarrow{\mathscr{L}}$ indicate converges in distribution. Then under the null hypothesis in (2.2), as $q\geqslant p-1$, $n\to\infty$, and $h\to 0$ such that $nh^{3/2}\to\infty$, it follows that

$$\frac{r_{\mathscr{K}}\mathscr{G}_{n,L}(h)-r_{\mathscr{K}}c_{\mathscr{K}}|\Omega|h^{-1}}{\{2r_{\mathscr{K}}c_{\mathscr{K}}|\Omega|h^{-1}\}^{1/2}}\xrightarrow{\mathscr{L}}\mathrm{N}(0,1), \tag{2.5}$$

where through the convolution operator $*$, the quantities $r_{\mathscr{K}}$ and $c_{\mathscr{K}}$ are represented by

$$r_{\mathscr{K}}=\frac{\mathscr{K}(0)-2^{-1}\mathscr{K}*\mathscr{K}(0)}{\int(\mathscr{K}-2^{-1}\mathscr{K}*\mathscr{K})^2(t)\,\mathrm{d}t} \quad \text{and} \quad c_{\mathscr{K}}=\mathscr{K}(0)-2^{-1}\mathscr{K}*\mathscr{K}(0). \tag{2.6}$$

The function $\mathscr{K}(t)=\mathscr{K}(t;q)$ above is defined as the equivalent kernel function (Fan and Gijbels, 1996, p. 64), induced from the qth degree local polynomial smoother with the basic kernel $K(t)$ in (2.4).

Moreover, it was shown in Fan et al. (2001) that applying the $(p-1)$th degree local polynomial smoother with a bandwidth of rate,

$$h_{\text{opt}}=n^{-2/(4p+1)}, \tag{2.7}$$

the resulting GLR statistic is capable of detecting alternatives converging to the null at a nonparametric rate, $\rho_n=n^{-2p/(4p+1)}$, which is the optimal rate of convergence of nonparametric testing (Ingster, 1982; Lepski and Spokoiny, 1999). For practical implementation, an empirical choice of the optimal bandwidth for $\mathscr{G}_{n,L}$ was suggested and tested in Zhang (2003): first re-scale the range of the observed X_i to the interval $[0,1]$, denoted by X_i^*, then for $\{(X_i^*,Y_i)_{i=1}^n\}$ take

$$h_{\text{opt;e}}=\eta\times\text{std}\{X_1^*,\ldots,X_n^*\}\times n^{-2/(4p+1)} \tag{2.8}$$

with a constant $\eta=1.5$. This empirical rule will be revisited later in Section 3.2.

Under the conditions specified before (2.5), conclusion (2.5) holds for a class of kernel functions that satisfy Condition (A3). This raises a question about how to assess the influence of a kernel function on the performance of the GLR test; refer to Section 4 for further discussion.

Recall that the GLR test statistics introduced in Section 2.1 could be carried out via any other type of nonparametric regression method. In next section we shall explore one such possibility, based on the smoothing splines.

3. The GLR test based on spline-smoothing

To ease technical manipulation, the fixed-design points, $X_i = x_i$, are considered in this paper for the smoothing spline estimator, denoted by $\hat{m}_\lambda$, which minimizes the penalized sum of squared errors

$$n^{-1}\sum_{i=1}^{n}\{Y_i - m(x_i)\}^2 + \lambda\int_0^1 \{m^{(p)}(x)\}^2\,dx, \qquad \lambda > 0 \tag{3.1}$$

over all functions $m \in W_2^p[0,1]$, where $W_2^p[0,1]$, termed the pth order Sobolev space, is defined as

$$W_2^p[0,1] = \left\{ m : m^{(j)} \text{ is absolutely continuous for } j = 0,1,\dots,p-1; \int_0^1 \{m^{(p)}(x)\}^2\,dx < \infty \right\}$$

for some fixed integer $p \geqslant 1$. The commonly-used cubic smoothing spline corresponds to $p=2$. The support of design points x_i, taken to be $\Omega=[0,1]$, is merely for simplicity. The smoothing parameter or the penalty factor λ, upon which the smoothing spline estimator depends, regulates the "rate of exchange" between fidelity to the data and smoothness of the fitted curve. See Eubank (1988) and Wahba (1990) for the detailed descriptions of smoothing splines.

Both the local polynomial and smoothing spline smoothers are linear estimators of the regression curve. According to (2.4), this feature is evident for the local polynomial regression estimators. For spline smoothing, we now verify this "linear smoother" property. For convenience, we assume that the points x_i are distinct and have been ordered, so that $x_1 < \cdots < x_n$. It is well known (see, e.g., Reinsch, 1967) that $\hat{m}_\lambda$ belongs to $\mathscr{S}_n^p$, the n-dimensional space of natural splines:

$$\mathscr{S}_n^p = \{m : m \in C^{2p-2}[0,1],\ m \text{ is a polynomial of degree } 2p-1 \text{ on } [x_i, x_{i+1}],\ i = 1,\dots,n-1, \text{ and of degree } p-1 \text{ on } [0,x_1] \text{ and } [x_n,1]\},$$

in which $C^{2p-2}[0,1]$ denotes the space of all functions on $[0,1]$ that have $2p-2$ continuous derivatives. An explicit expression for $\hat{m}_\lambda(x)$ can be obtained via the basis functions $\{\phi_{jn},\ j=1,\dots,n\}$ of $\mathscr{S}_n^p$ introduced by Demmler and Reinsch (1975). These functions satisfy the conditions

$$\frac{1}{n}\sum_{i=1}^{n}\phi_{jn}(x_i)\phi_{kn}(x_i)=\delta_{jk},$$

$$\int_0^1 \phi_{jn}^{(p)}(x)\phi_{kn}^{(p)}(x)\,\mathrm{d}x=\gamma_{kn}\delta_{jk}$$

for $j,k=1,\ldots,n$, with $0=\gamma_{1n}=\cdots=\gamma_{pn}<\gamma_{(p+1)n}\leqslant\cdots\leqslant\gamma_{nn}$, and δ_{jk} the Kronecker's delta. Denote by $\boldsymbol{\phi}_{jn}=(\phi_{jn}(x_1),\ldots,\phi_{jn}(x_n))^{\mathrm{T}},\ j=1,\ldots,n$, the basis vectors evaluated at the design observations, and by $\mathbf{y}=(Y_1,\ldots,Y_n)^{\mathrm{T}}$ the vector of responses. Then the solution of (3.1) can be expressed as

$$\hat{m}_\lambda(x)=\sum_{j=1}^{p}\hat{\theta}_{j-1}x^{j-1}+\sum_{j=p+1}^{n}\frac{n^{-1}\boldsymbol{\phi}_{jn}^{\mathrm{T}}\mathbf{y}}{1+\lambda\gamma_{jn}}\,\phi_{jn}(x),$$

where $\{\hat{\theta}_j\}_{j=0}^{p-1}$ stand for the ordinary least-squares estimates of parameters under H_0 in (2.2). The smoother matrix H_λ, associated with $\hat{m}_\lambda$, such that $(\hat{m}_\lambda(x_1),\ldots,\hat{m}_\lambda(x_n))^{\mathrm{T}}=H_\lambda\mathbf{y}$, can be written in the form

$$H_\lambda=\boldsymbol{X}\,\mathrm{diag}\{(1+\lambda\gamma_{jn})^{-1}\}_{j=1}^{n}\boldsymbol{X}^{\mathrm{T}} \tag{3.2}$$

in which the square matrix $\boldsymbol{X}=n^{-1/2}(\boldsymbol{\phi}_{1n},\ldots,\boldsymbol{\phi}_{nn})$ is orthonormal. Now define by $T=(x_i^{j-1})_{i=1,\ldots,n}^{j=1,\ldots,p}$ the design matrix, and by P the projection matrix generated by T. Set $\boldsymbol{X}_1=n^{-1/2}(\boldsymbol{\phi}_{1n},\ldots,\boldsymbol{\phi}_{pn})$ and $\boldsymbol{X}_2=n^{-1/2}(\boldsymbol{\phi}_{(p+1)n},\ldots,\boldsymbol{\phi}_{nn})$. Then it follows that

$$P=\boldsymbol{X}_1\boldsymbol{X}_1^{\mathrm{T}}\quad\text{and}\quad H_\lambda=P+\boldsymbol{X}_2\,\mathrm{diag}\{(1+\lambda\gamma_{jn})^{-1}\}_{j=p+1}^{n}\boldsymbol{X}_2^{\mathrm{T}}. \tag{3.3}$$

There is an extensive literature on the use of smoothing splines to nonparametric tests of the regression function. We now briefly review those results that have particular relevance here. Often, they are based upon comparing the residual sum of squares of a nonparametric regression fit, $H_\lambda\mathbf{y}$, versus that of a parametric fit, $P\mathbf{y}$, via the quadratic form

$$\mathrm{RSS}(H_\lambda,P)=\mathbf{y}^{\mathrm{T}}(H_\lambda-P)^{\mathrm{T}}(H_\lambda-P)\mathbf{y} \tag{3.4}$$

or equivalently, versus that of a smoothed parametric fit, $H_\lambda P\mathbf{y}$, via $\mathrm{RSS}(H_\lambda,H_\lambda P)=\mathbf{y}^{\mathrm{T}}(H_\lambda-H_\lambda P)^{\mathrm{T}}(H_\lambda-H_\lambda P)\mathbf{y}$. A χ^2-test, based on (3.4), was developed in Jayasuriya (1996), which involves calculation of the trace of H_λ^k, $\mathrm{tr}(H_\lambda^k)$, for $k=2,4$, and an estimate of the noise variance σ^2. This approach is a generalization of the Eubank and Spiegelman (1990) test for linearity under Gaussian models, to non-Gaussian models and to testing for the adequacy of a $(p-1)$th degree polynomial regression. An alternative proposal of the χ^2-test, based on the roughness-penalty of the fitted spline function,

$$\int_0^1\{\hat{m}_\lambda^{(p)}(x)\}^2\,\mathrm{d}x=(n\lambda)^{-1}\mathbf{y}^{\mathrm{T}}(H_\lambda-H_\lambda^2)\mathbf{y} \tag{3.5}$$

was developed in Chen (1994). Recently, two F-tests, one based on (3.4) and the other based on (3.5), are constructed in Ramil-Novo and González-Manteiga (2000). As have been shown in these articles, the χ^2- and F-tests, based on (3.4), can detect local alternatives converging at a rate, $\delta_{n;1} = \{n\lambda^{1/(4p)}\}^{-1/2}$, or slower to the null model (Jayasuriya, 1996, p. 1628; Ramil-Novo and González-Manteiga, 2000, p. 824), whereas the χ^2-test, based on (3.5), is only able to distinguish local alternatives that are of order, $\delta_{n;2} = \{n\lambda^{(4p+1)/(4p)}\}^{-1/2}$, or slower from the null hypothesis (Chen, 1994, p. 68). While these "local asymptotic" results are useful for explaining the slower nonparametric rate of convergence compared with the parametric $n^{-1/2}$-rate, they cannot be applied directly to obtain the optimum amount of smoothing parameter λ. This issue will be addressed in our study of Section 3.2.

3.1. Asymptotic null distribution

Hereafter, we denote by $\mathscr{G}_{n,S}(\lambda)$ the GLR statistic, based on the smoothing spline estimator $\hat{m}_\lambda$. We first establish in Theorem 1 the asymptotic normal distribution of $\mathscr{G}_{n,S}(\lambda)$, under the null hypothesis; an application of this result will be utilized in Section 3.2, where the asymptotic power of $\mathscr{G}_{n,S}(\lambda)$ is studied.

Before displaying Theorem 1, its technical assumptions on λ, in non-Gaussian models (called (Case 1)-models) and Gaussian models (called (Case 2)-models), deserve some explanations. Basically, these two sets of assumptions are similar to those of Ramil-Novo and González-Manteiga (2000, pp. 823–824), in which the asymptotic distributions of their F-tests in both (Case 1)- and (Case II)-models are derived under the null and local alternatives converging to the null at the rate $\delta_{n;1}$ specified above. Thus, in (Case 2)-models, we will simply use the same set of assumptions as they have imposed for Gaussian models. In (Case 1)-models, however, one condition among the set of assumptions they put for non-Gaussian models, $n\lambda^{(4p+1)/(4p)} \to \infty$ is not required for deriving the limiting distributions of their F-tests under the null hypothesis alone; we notice that a weaker condition, $n\lambda \to \infty$, will suffice. Therefore in our Theorem 1 below, which only deals with the asymptotic null distribution of the $\mathscr{G}_{n,S}(\lambda)$ statistic, we will put the weaker condition $n\lambda \to \infty$ to (Case 1)-models.

Moreover, the eigenvalues of the matrix H_λ, as given in (3.2), are of great importance in studying the asymptotic properties of smoothing splines and the behaviors of the associated tests. Applying Speckman (1981)'s result and assuming suitable conditions, we could show in Lemma 2 that

$$\lambda^{1/(2p)}\,\mathrm{tr}(H_\lambda^r) = c(f)(2\pi)^{-1}\int_{-\infty}^{+\infty}(1+t^{2p})^{-r}\,\mathrm{d}t + \mathrm{o}(1) \tag{3.6}$$

for each integer $r \geqslant 2$, where $c(f) = \int_0^1 f(t)^{1/(2p)}\,\mathrm{d}t$ and f represents the design density that fulfills Condition (B3) in the Appendix. A similar type of approximation, when $r = 1$, can be found in Eubank (1988, p. 327). In general, the approximation error term, as denoted by o(1) in (3.6), without further specification of its magnitude, will suffice for most of the asymptotic investigations in which $\mathrm{tr}(H_\lambda^r)$ is incorporated. To facilitate technical manipulation of our Theorem 1, we shall put in (3.6) an additional assumption, $o(1)=o\{\lambda^{1/(4p)}\}$, when $r=1,2$. To see the validity of this assumption, let us consider two typical instances. In one situation, arising from periodic smoothing splines with uniform designs, the error term is shown (Wahba, 1975, p. 388) to be $O\{\lambda^{1/(2p)}\}+O\{(n\lambda^{1/(2p)})^{-(2p-1)}\}$ when $r=1$, and analogously, to be $O\{\lambda^{1/(2p)}\}+O\{(n\lambda^{1/(2p)})^{-2p}\}$ when $r=2$. In another situation arising from cubic smoothing splines ($p=2$) and uniform designs, Kou (2000) obtained explicit bounds on the eigenvalues of H_λ; these bounds could (by a typical graphical argument) lead to the same approximation error term, $o\{\lambda^{1/(4p)}\}$, for $r=1$ and 2. Therefore our condition, imposed on the term o(1), does not appear to be particularly restrictive.

Theorem 1. *Let* $\boldsymbol{K}(x)=(2\pi)^{-1}\int_{-\infty}^{+\infty}(1+t^{2p})^{-1}\exp(-\mathrm{i}tx)\,\mathrm{d}t$, *and set* $c(f)=\int_0^1 f(t)^{1/(2p)}\,\mathrm{d}t$. *Suppose* $\lambda^{1/(2p)}\operatorname{tr}(H_\lambda^r)=c(f)(2\pi)^{-1}\int_{-\infty}^{+\infty}(1+t^{2p})^{-r}\,\mathrm{d}t+o\{\lambda^{1/(4p)}\}$, *for* $r=1,2$. *Assume that either*

Case 1: *Condition* (B) *in the Appendix holds, and* $p\geqslant 2$, $n\to\infty$, $\lambda\to 0$ *such that* $n\lambda\to\infty$ *and* $n\lambda^{(3+4\zeta)/(2p)}\to\infty$;

or

Case 2: *Condition* (C) *in the Appendix holds, and* $p\geqslant 2$, $n\to\infty$, $\lambda\to 0$ *such that* $n\lambda\to\infty$.

Then under H_0, *it holds that*

$$\frac{r_{\boldsymbol{K}}\mathscr{G}_{n,S}(\lambda)-r_{\boldsymbol{K}}c_{\boldsymbol{K}}\lambda^{-1/(2p)}c(f)}{\{2r_{\boldsymbol{K}}c_{\boldsymbol{K}}\lambda^{-1/(2p)}c(f)\}^{1/2}}\xrightarrow{\mathscr{L}}\mathrm{N}(0,1) \tag{3.7}$$

with

$$r_{\boldsymbol{K}}=\frac{2p+1}{2p-1}\,\frac{48p^2}{24p^2+14p+1}\quad \textit{and}\quad c_{\boldsymbol{K}}=\frac{2p+1}{8p^2\sin\{\pi/(2p)\}}. \tag{3.8}$$

Equivalently, the expressions for the two constants above can be rewritten as

$$r_{\boldsymbol{K}}=\frac{\boldsymbol{K}(0)-2^{-1}\boldsymbol{K}*\boldsymbol{K}(0)}{\int(\boldsymbol{K}-2^{-1}\boldsymbol{K}*\boldsymbol{K})^2(t)\,\mathrm{d}t}\quad \textit{and}\quad c_{\boldsymbol{K}}=\boldsymbol{K}(0)-2^{-1}\boldsymbol{K}*\boldsymbol{K}(0). \tag{3.9}$$

The conclusion drawn from Theorem 1 reveals that a parallel version of result (2.5), based on the local polynomial regression fit, continues to hold for the smoothing spline method. In contrast to the equivalent kernel function $\mathscr{K}$ in (2.6), the function $\boldsymbol{K}$ specified in (3.9) is defined by Silverman (1984a) as the Fourier transform of $(1+t^{2p})^{-1}$. Thus if we interpret this particular $\boldsymbol{K}$ as a kernel function, make the identification $|\Omega|=1$ (see Condition (B3) in the Appendix), and place

$$h=\lambda^{1/(2p)}/c(f), \tag{3.10}$$

we observe that Theorem 1 could have been predicted from (2.5) and (2.6), which are developed from the local polynomial curve-fitting method. In this sense, the asymptotic null distributions of the two GLR test statistics, i.e., $\mathscr{G}_{n,S}(\lambda)$ based on smoothing splines and $\mathscr{G}_{n,L}(h)$ based on local polynomial smoothing, can indeed be regarded as equivalent. Results of this type complement those obtained, in regression function estimation, by Silverman (1984a, pp. 901–902), who showed that under appropriate conditions, the cubic smoothing spline estimator ($p=2$), at an interior point x, behaves roughly as a kernel regression fit to the data $\{(x_i, Y_i/f(x_i))\}_{i=1}^n$ with kernel $\mathscr{K}$ and variable bandwidth $h(x)=\{\lambda/f(x)\}^{1/(2p)}$. As we can clearly see, this expression bears a close resemblance to (3.10), and is exactly identical to (3.10) when uniform designs on $[0,1]$ are used. Interestingly, we shall find in Section 4.1 that, even if we use the equivalent choices of λ and h as given in (3.10), the $\mathscr{G}_{n,S}(\lambda)$ and $\mathscr{G}_{n,L}(h)$ tests do not necessarily have equivalent powers.

Apart from the equivalence of sampling null distributions between $\mathscr{G}_{n,S}(\lambda)$ and its counterpart $\mathscr{G}_{n,L}(h)$, Theorem 1 indicates that $\mathscr{G}_{n,S}(\lambda)$ enjoys a substantial computational advantage over other spline-based tests. This viewpoint is clearly demonstrated in (3.7)–(3.9), where the expressions for $r_{\boldsymbol{K}}$ and $c_{\boldsymbol{K}}$ can be computed explicitly, and $c(f)$ equals 1 in uniform designs. In contrast, the preceding spline-based test statistics require heavy computation of $\mathrm{tr}(H_\lambda^r)$, which can potentially become a problem for large and huge sample sizes. The only trivial computational price, paid for using $\mathscr{G}_{n,S}(\lambda)$, is to approximate the functional $c(f)$ in (3.7), when f is unknown or non-uniform, through the use of a nonparametric kernel density estimate (Silverman, 1986). Silverman (1984b, p. 586) described and analyzed an efficient estimation algorithm for the case $p=2$; this procedure can be extended straightforwardly to the general cases of $p \geqslant 2$.

3.2. Power study and optimal selection of smoothing parameter

It is well known in the literature on smoothing splines that the optimal choice of λ, for estimating the regression curve, minimizes the mean squared error of the spline estimator, leading to the optimal rate $\mathrm{O}(n^{-2p/(2p+1)})$; a data-driven method, based on minimizing the generalized cross-validation (GCV) criterion (Craven and Wahba, 1979), produces a consistent estimate of such optimal λ. However, studies on the optimal rate of λ in the context of spline-based tests have received much less attention in the literature (see the aforementioned work on spline-based test statistics). For this issue, we would like to stress that it is intimately related to examining the power of a spline-based test, rather than assessing the fit of a spline estimator. As a consequence, it is not surprising that the optimal rates of smoothing parameters will differ between testing and estimation.

The large-sample power of tests is often studied by considering their behavior under local alternatives to the null hypothesis. In what follows, the departures of alternative sets from the hypothetical polynomial model, denoted by g, are assumed to be smooth and lie in a general class of functions, $L_2^f[0,1] \setminus \{1,\dots,x^{p-1}\}$, which consists of all functions g on $[0,1]$ such that $\|g\|_f^2 = \int_0^1 g^2(x) f(x)\,\mathrm{d}x < \infty$ and $\int_0^1 x^j g(x) f(x)\,\mathrm{d}x = 0$ for all $j=0,\dots,p-1$. With these assumptions, we shall show in Theorem 2 that the GLR test based on smoothing splines achieves the optimal rate of convergence for nonparametric hypothesis testing, provided that an appropriate choice of smoothing parameter is taken.

Theorem 2. *Suppose* $\lambda^{1/(2p)}\,\mathrm{tr}(H_\lambda^r) = c(f)(2\pi)^{-1}\int_{-\infty}^{+\infty}(1+t^{2p})^{-r}\,\mathrm{d}t + \mathrm{o}\{\lambda^{1/(4p)}\}$, *for* $r = 1, 2$. *Assume that either*

Case 1: *Condition* (B) *in the Appendix holds, and* $p \geqslant 2$, $n \to \infty$, $\lambda \to 0$ *such that* $n\lambda \to \infty$ *and* $n\lambda^{(3+4\zeta)/(2p)} \to \infty$ *with* $0 < \zeta < (4p-5)/8$;

or

Case 2: *Condition* (C) *in the Appendix holds, and* $p \geqslant 2$, $n \to \infty$, $\lambda \to 0$ *such that* $n\lambda \to \infty$.

Let $m(x) = \sum_{j=0}^{p-1} \theta_j x^j + g(x)$, *with* $g \in L_2^f[0,1] \setminus \{1, x, \ldots, x^{p-1}\}$ *and* $g \in W_2^p[0,1]$. *Then* $\mathcal{G}_{n,S}(\lambda)$ *can detect alternatives of order* $\{n\lambda^{1/(4p)}\}^{-1/2}$ *or slower from the null model. Set* $\rho_n = n^{-2p/(4p+1)}$, *and take* λ *of rate* $n^{-4p/(4p+1)}$. *Then for any sequence* $c_n \to \infty$, *the power function of the* $\mathcal{G}_{n,S}(\lambda)$ *test is asymptotically one, i.e.*,

$$\lim_{n\to\infty} \inf_{g \in L_2^f[0,1]\cap W_2^p[0,1]\setminus\{1,x,\ldots,x^{p-1}\}:\ \|g\|_f \geqslant c_n\rho_n} P_g\left[\frac{r_{\boldsymbol{K}}\mathcal{G}_{n,S}(\lambda) - r_{\boldsymbol{K}}c_{\boldsymbol{K}}\lambda^{-1/(2p)}c(f)}{\{2r_{\boldsymbol{K}}c_{\boldsymbol{K}}\lambda^{-1/(2p)}c(f)\}^{1/2}} \geqslant z_\alpha\right] = 1,$$

where z_α *denotes the* $100(1-\alpha)$th *percentile of the standard normal distribution, and* P_g *denotes the probability calculated under the alternative with a departure* g *from the null.*

Theorem 2 has a number of consequences. Firstly, it implies that the $\mathcal{G}_{n,S}(\lambda)$ test, when λ is of the rate $n^{-4p/(4p+1)}$, is uniformly consistent against a smooth alternative that deviates from the null with the distance of order $\rho_n = n^{-2p/(4p+1)}$. Secondly, Theorem 2 indicates that the optimal rate of smoothing parameter of the $\mathcal{G}_{n,S}(\lambda)$ test is

$$\lambda_{\mathrm{opt}} = n^{-4p/(4p+1)}. \tag{3.11}$$

In this way we formally verify that, in the context of nonparametric testing, the optimal rate of smoothing parameter λ_{opt} in (3.11) is equivalent, via the mapping (3.10), to h_{opt} in (2.7). Furthermore, we suggest from (3.10) an empirical choice of the optimal λ,

$$\lambda_{\mathrm{opt;e}} = \{h_{\mathrm{opt;e}}c(f)\}^{2p} \tag{3.12}$$

in which $h_{\mathrm{opt;e}}$ is given in (2.8). Thirdly, the theorem reveals that λ_{opt} approaches zero at a rate faster than its counterpart, $\lambda = \mathrm{O}\{n^{-2p/(2p+1)}\}$, obtained from minimizing the conventional generalized cross-validation criterion. Fourthly, from a practical point of view, the magnitude of the optimal rate $n^{-4p/(4p+1)}$ for testing and the magnitude of the optimal rate $n^{-2p/(2p+1)}$ for estimation, do not seem to differ substantially. Therefore, for a $\mathcal{G}_{n,S}(\lambda)$ test, the GCV criterion may be employed for choosing the degree of smoothing, although power loss of the resulting test may arise. Finally, the arguments used for verifying Theorem 2 can be extended to other types of spline-based test statistics.

4. Power comparison

Given the two types of estimation methods that we have considered in the likelihood-ratio based tests, it is of interest to compare them from an efficiency standpoint. Their relative powers against local alternatives are presented in Theorem 3.

4.1. Power comparison of GLR tests based on smoothing spline and local polynomial smoother

Theorem 3. *Assume that either Case* 1 *assumption or Case* 2 *assumption of Theorem* 2 *holds.*

(1) *Let* $h = C_1 n^{-2/(4p+1)}$ *and let* $\lambda = C_2^{2p} n^{-4p/(4p+1)}\{c(f)\}^{2p}$ *with positive constants* C_1 *and* C_2. *Then the Pitman asymptotic relative efficiency of* $\mathscr{G}_{n,S}(\lambda)$ *to* $\mathscr{G}_{n,L}(h)$ *is given by*

$$\{(C_2/C_1)(r_{\boldsymbol{K}}/c_{\boldsymbol{K}})/(r_{\mathscr{K}}/c_{\mathscr{K}})\}^{(4p+1)/(8p)}. \tag{4.1}$$

(2) *Let* $h = C_1 n^{-1/(2p+1)}$ *and let* $\lambda = C_2^{2p} n^{-2p/(2p+1)}\{c(f)\}^{2p}$ *with positive constants* C_1 *and* C_2. *Then the Pitman asymptotic relative efficiency of* $\mathscr{G}_{n,S}(\lambda)$ *to* $\mathscr{G}_{n,L}(h)$ *is given by*

$$\{(C_2/C_1)(r_{\boldsymbol{K}}/c_{\boldsymbol{K}})/(r_{\mathscr{K}}/c_{\mathscr{K}})\}^{(4p+2)/(8p+2)}. \tag{4.2}$$

Theorem 3 compares $\mathscr{G}_{n,S}(\lambda)$ and $\mathscr{G}_{n,L}(h)$, under what can be regarded as "optimal" rates of λ and h, for testing and for estimation. Clearly, the relative efficiency, in either case, depends on the ratio of $r_{\boldsymbol{K}}/c_{\boldsymbol{K}}$ to $r_{\mathscr{K}}/c_{\mathscr{K}}$. To calibrate the efficiency numerically, we present in Table 1 the values of $r_{\boldsymbol{K}}/c_{\boldsymbol{K}}$ and $r_{\mathscr{K}}/c_{\mathscr{K}}$. According to Eq. (3.8), the ratio $r_{\boldsymbol{K}}/c_{\boldsymbol{K}}$ will quickly approach its limit $4\pi \approx 12.57$ as p increases, and thus we take $p = 2$ up to 6 in Table 1. With respect to $\mathscr{G}_{n,L}(h)$, the commonly used multi-weight

Table 1
Constants $r_{\boldsymbol{K}}/c_{\boldsymbol{K}}$ from the smoothing splines that minimize (3.1) and constants $r_{\mathscr{K}}/c_{\mathscr{K}}$ from the $(p-1)$th degree local polynomial smoother

p	Smoothing spline $r_{\boldsymbol{K}}/c_{\boldsymbol{K}}$	Epanechnikov kernel $r_{\mathscr{K}}/c_{\mathscr{K}}$	Biweight kernel $r_{\mathscr{K}}/c_{\mathscr{K}}$	Triweight kernel $r_{\mathscr{K}}/c_{\mathscr{K}}$
2	11.5852	4.7007	3.9733	3.4700
3	12.0093	2.5288	2.2714	2.0545
4	12.1864	2.5288	2.2714	2.0545
5	12.2809	1.7510	1.6217	1.4984
6	12.3388	1.7510	1.6217	1.4984

kernel functions, of the form

$$\{\mathrm{Beta}(\tfrac{1}{2}, \ell + 1)\}^{-1}(1 - t^2)^{\ell} I(|t| \leqslant 1), \quad \ell = 1, 2, \ldots$$

are considered, where $\mathrm{Beta}(\cdot,\cdot)$ denotes a Beta function and $I(\cdot)$ is an indicator function. The Epanechnikov, Biweight and Triweight kernel functions correspond to indexes $\ell=1$, $\ell=2$, and $\ell=3$, respectively. Hence, if we take the equivalent form of parameters λ and h, as given in (3.10), namely, $C_1 = C_2$, it appears that, from (4.1) (or (4.2)) and Table 1, the spline approach seems to be more efficient than the local polynomial regression method, because $(r_{\boldsymbol{K}}/c_{\boldsymbol{K}})/(r_{\mathscr{K}}/c_{\mathscr{K}}) > 1$. For example, with these choices of λ and h, the $\mathscr{G}_{n,S}(\lambda)$ test of linearity based on the cubic smoothing splines will be asymptotically 1.7 times as efficient as the $\mathscr{G}_{n,L}(h)$ test of linearity based on the local linear estimation (combined with the Epanechnikov kernel).

It should be pointed out, however, that the relative efficiency in Theorem 3 also depends on the relative magnitude of the constants C_1 and C_2 in conjunction with the optimal rates of smoothing parameters. To summarize, the analytical criterion below we put on the relative magnitude of the smoothing parameters h and λ directly determine the relative power of the $\mathscr{G}_{n,L}(h)$ and $\mathscr{G}_{n,S}(\lambda)$ tests.

Case I: if $C_2/C_1 < (r_{\mathscr{K}}/c_{\mathscr{K}})/(r_{\boldsymbol{K}}/c_{\boldsymbol{K}})$ or $\lambda < \{h\,c(f)(r_{\mathscr{K}}/c_{\mathscr{K}})/(r_{\boldsymbol{K}}/c_{\boldsymbol{K}})\}^{2p}$, the $\mathscr{G}_{n,S}(\lambda)$ test is less powerful than the $\mathscr{G}_{n,L}(h)$ test.

Case II: if $C_2/C_1 = (r_{\mathscr{K}}/c_{\mathscr{K}})/(r_{\boldsymbol{K}}/c_{\boldsymbol{K}})$ or $\lambda = \{h\,c(f)(r_{\mathscr{K}}/c_{\mathscr{K}})/(r_{\boldsymbol{K}}/c_{\boldsymbol{K}})\}^{2p}$, the $\mathscr{G}_{n,S}(\lambda)$ test is as powerful as the $\mathscr{G}_{n,L}(h)$ test.

Case III: if $C_2/C_1 > (r_{\mathscr{K}}/c_{\mathscr{K}})/(r_{\boldsymbol{K}}/c_{\boldsymbol{K}})$ or $\lambda > \{h\,c(f)(r_{\mathscr{K}}/c_{\mathscr{K}})/(r_{\boldsymbol{K}}/c_{\boldsymbol{K}})\}^{2p}$, the $\mathscr{G}_{n,S}(\lambda)$ test is more powerful than the $\mathscr{G}_{n,L}(h)$ test.

Hence neither the $\mathscr{G}_{n,S}(\lambda)$ test nor its competitor $\mathscr{G}_{n,L}(h)$ can consistently outperform the other in terms of powers. Furthermore, since the equivalent choice of h and λ, for function estimation under the uniform design, is the same as that given in (3.10), Case III above implies that, for these choices of λ and h, the $\mathscr{G}_{n,S}(\lambda)$ test will be more powerful than the $\mathscr{G}_{n,L}(h)$ test.

4.2. Optimal choice of kernel function

Unlike nonparametric curve fitting, the literature on nonparametric testing has not addressed the issue of optimal kernel function. Interestingly, the equivalence results between the $\mathscr{G}_{n,S}(\lambda)$ and $\mathscr{G}_{n,L}(h)$ tests deliver quantitative information about choices of kernel function. This is due to the fact that the arguments for Theorem 2 regarding the $\mathscr{G}_{n,S}(\lambda)$ test will go through, after appropriate modifications, to the $\mathscr{G}_{n,L}(h)$ test. Particularly, examination of the second term on the right hand side of Eq. (6.10) (see Appendix) suggests that for the $\mathscr{G}_{n,L}(h)$ test, the asymptotic power is determined in rate by $nh^{1/2}$, and in proportionality constant by a functional, $r_{\mathscr{K}}/c_{\mathscr{K}}$, depending on the kernel function. It follows naturally that the larger the ratio $r_{\mathscr{K}}/c_{\mathscr{K}} = 1/\int(\mathscr{K} - 2^{-1}\mathscr{K} * \mathscr{K})^2(t)\,\mathrm{d}t$ is, the larger the resulting asymptotic power is. According to this optimality criterion, Table 1 indicates that the Epanechnikov kernel performs the best, followed by the Biweight and Triweight kernels. Seeking the closed-form expression for the optimal kernel function along this line will be an interesting future work.

5. Simulation

In this section we summarize numerical information about the extent to which large-sample properties of the previous section are reflected in finite sample situations. To make a fair power comparison between the $\mathscr{G}_{n,L}(h)$ and $\mathscr{G}_{n,S}(\lambda)$ tests, fixed-design points are considered. We simulate responses from the alternative models of two different forms of regression functions. They are

$$\text{Example 1: } Y_i = 1 + 2(4x_i - 2) + \theta(4x_i - 2)^2 + \varepsilon_i, \quad i = 1, \dots, n, \ \ \theta \in [0, 0.9], \tag{5.1}$$

$$\text{Example 2: } Y_i = 1 + 2(4x_i - 2) + \theta(4x_i - 2)\exp(2x_i) + \varepsilon_i, \quad i = 1, \dots, n, \ \ \theta \in [0, 0.6], \tag{5.2}$$

with $x_i = (i - 0.5)/n$, where the ε_i are uncorrelated standard normal random errors. For simplicity, we choose 20 values of θ equally spaced in the intervals above; $\theta = 0$

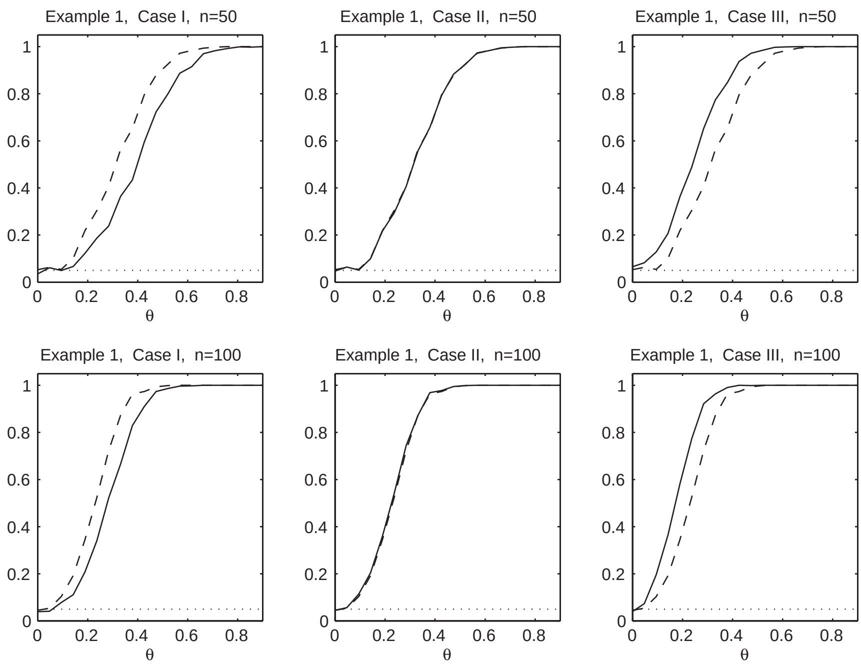

Fig. 1. Power comparison of $\mathscr{G}_{n,S}(\lambda)$ and $\mathscr{G}_{n,L}(h)$ against alternative models (5.1), where $h = C_1 n^{-2/(4p+1)}$ with $C_1 = 1.5 \times 0.29$, and $\lambda = C_2^{2p} n^{-4p/(4p+1)}$. Solid curve: $\mathscr{G}_{n,S}(\lambda)$ based on cubic smoothing spline; dashed curve: $\mathscr{G}_{n,L}(h)$ based on local linear estimation with the Epanechnikov kernel. The bottom dotted lines denote the norminal 5% significance level. Case I: $C_2/C_1 = 2^{-1}(r_{\mathscr{K}}/c_{\mathscr{K}})/(r_{\boldsymbol{K}}/c_{\boldsymbol{K}})$; Case II: $C_2/C_1 = (r_{\mathscr{K}}/c_{\mathscr{K}})/(r_{\boldsymbol{K}}/c_{\boldsymbol{K}})$; Case III: $C_2/C_1 = 1$.

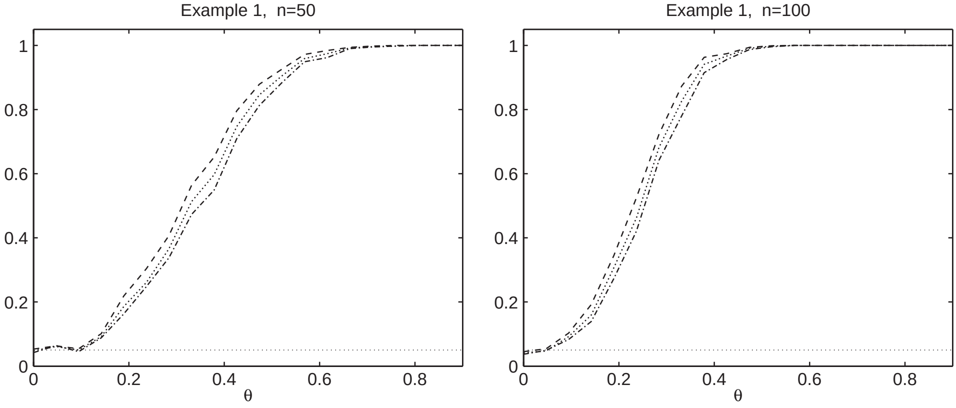

Fig. 2. Power comparison of $\mathscr{G}_{n,L}(h)$, using different kernel-based smoothers, against alternative models (5.1). Dashed curve: Epanechnikov kernel; dotted curve: Biweight kernel; dash dotted curve: Triweight kernel. The bottom dotted lines denote the norminal 5% significance level.

corresponds to the null model, that is, our goal is to test for the validity of a linear model (with $p=2$ in (2.2)) under a nonparametric regression model (2.1). The simulation study consists of generating 1000 independent samples from each of the alternative models. The empirical powers are estimated by the proportion of times the observed test statistics exceed their upper 5% critical values. The critical values are obtained from the 95th sample percentiles across 1000 independent samples, generated from the null model. Cubic smoothing splines, local linear smoothing with the Epanechnikov kernel are treated, respectively.

The empirical power curves of the $\mathscr{G}_{n,S}(\lambda)$ and $\mathscr{G}_{n,L}(h)$ tests against the alternatives in (5.1) are presented in Fig. 1, in which sample sizes 50 and 100 are considered. There the bandwidth $h = C_1 n^{-2/(4p+1)}$, with the proportionality constant $C_1 = 1.5 \times 0.29$, is chosen based on the empirical formula given in (2.8); the choice $\lambda = C_2^{2p} n^{-4p/(4p+1)}$ follows the first part of Theorem 3. (The power curves will be similar if the rates of h and λ are replaced by $n^{-1/(2p+1)}$ and $n^{-2p/(2p+1)}$ as in the second part of Theorem 3.) The three cases in Fig. 1 correspond to the ratios C_2/C_1 equal to $2^{-1}(r_{\mathscr{K}}/c_{\mathscr{K}})/(r_{\boldsymbol{K}}/c_{\boldsymbol{K}})$, $(r_{\mathscr{K}}/c_{\mathscr{K}})/(r_{\boldsymbol{K}}/c_{\boldsymbol{K}})$, and 1, respectively. According to the summary in the last paragraph of Section 4.1, $\mathscr{G}_{n,S}(\lambda)$ is less powerful in Case I than $\mathscr{G}_{n,L}(h)$, is as powerful in Case II as $\mathscr{G}_{n,L}(h)$, and is more powerful in Case III than $\mathscr{G}_{n,L}(h)$. Indeed, these theoretical conclusions are well supported by the simulated power plots in Fig. 1.

Furthermore, we present in Fig. 2 the power comparison of the $\mathscr{G}_{n,L}(h)$ test, in which the Epanechnikov, Biweight, and Triweight kernels are used respectively. Clearly, Fig. 2 lends convincing support to an asymptotic result obtained in Section 4.2; that is, the Epanechinikov kernel outperforms the other multiweight kernels.

Similarly, when the alternative model is (5.2), the power comparison between $\mathscr{G}_{n,L}(h)$ and $\mathscr{G}_{n,S}(\lambda)$ is presented in Fig. 3, while Fig. 4 displays the power comparison between different choices of kernel functions.

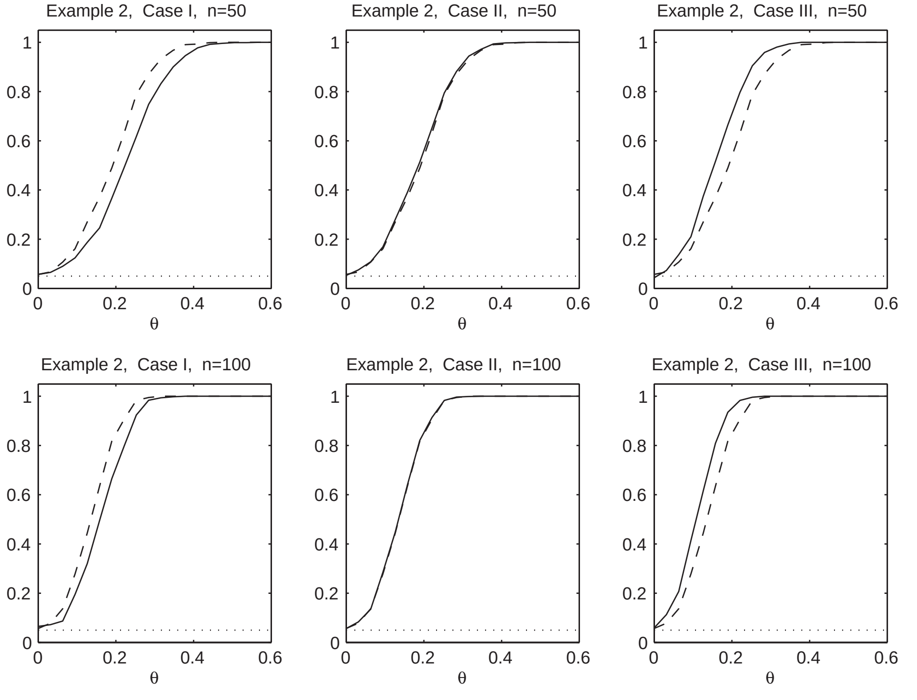

Fig. 3. Power comparison of $\mathscr{G}_{n,S}(\lambda)$ and $\mathscr{G}_{n,L}(h)$ against alternative models (5.2), where $h = C_1 n^{-2/(4p+1)}$ with $C_1 = 1.5 \times 0.29$, and $\lambda = C_2^{2p} n^{-4p/(4p+1)}$. Solid curve: $\mathscr{G}_{n,S}(\lambda)$ based on cubic smoothing spline; dashed curve: $\mathscr{G}_{n,L}(h)$ based on local linear estimation with the Epanechnikov kernel. The bottom dotted lines denote the norminal 5% significance level. Case I: $C_2/C_1 = 2^{-1}(r_{\mathscr{K}}/c_{\mathscr{K}})/(r_{\boldsymbol{K}}/c_{\boldsymbol{K}})$; Case II: $C_2/C_1 = (r_{\mathscr{K}}/c_{\mathscr{K}})/(r_{\boldsymbol{K}}/c_{\boldsymbol{K}})$; Case III: $C_2/C_1 = 1$.

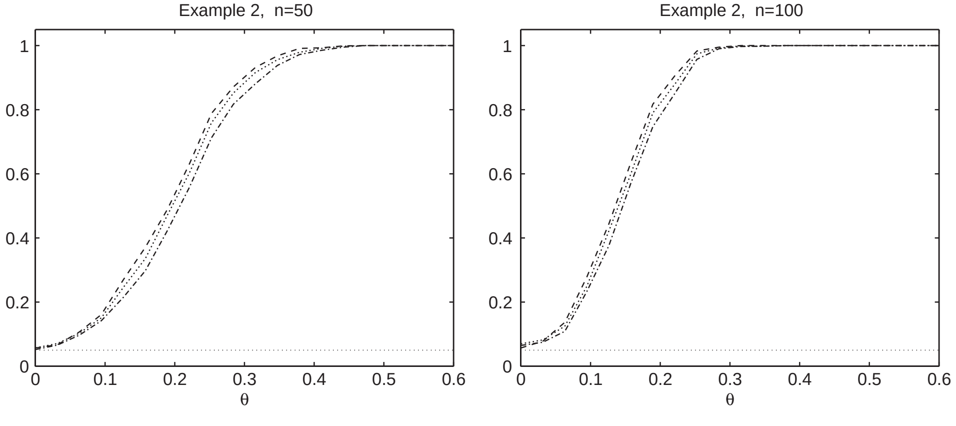

Fig. 4. Power comparison of $\mathscr{G}_{n,L}(h)$, using different kernel-based smoothers, against alternative models (5.2). Dashed curve: Epanechnikov kernel; dotted curve: Biweight kernel; dash dotted curve: Triweight kernel. The bottom dotted lines denote the norminal 5% significance level.

6. Summary and extensions

In nonparametric testing, it is of both theoretical and practical interest to compare tests based on different types of curve estimation methods. In this article, we show that the relative performances of the GLR tests based on the spline and local polynomial smoothers depend largely on the smoothing parameters. We provide a simple method, based on comparing the magnitude of the smoothing parameters, to gauge when the spline-based test is more powerful than the local polynomial regression-based test and when it is not. We also suggest in (2.8) and (3.12) the empirical methods for choosing the optimal smoothing parameters. Regarding aspects on model assumptions, local polynomial-based tests are comparatively design adaptive, and require milder assumptions on smoothing parameters and random errors. To solve real problems at hand, it would be advisable to choose the most convenient smoothing method.

In conclusion, we point out that the results in Sections 3 and 4 can be extended in several directions. For instance, in generalized linear models, the likelihood function can be obtained via exponential family distribution of the response variable given the design variable. The GLR statistic can be formulated analogously. Using local linear regression technique, inference for the generalized linear model based on the $\mathscr{G}_{n,L}(h)$ test has been given in Cai et al. (2000). In this case, our Theorems 1 and 2 could be extended to the version of the $\mathscr{G}_{n,S}(\lambda)$ test based on spline method, without too much technical difficulty. For multivariate modeling, in which the response variable is related to the covariates by a varying coefficient model, one may also be interested in assessing whether the varying coefficient functions are really varying with certain covariates. To this end, the varying-coefficient functions can be fitted by cubic smoothing splines as proposed in Hastie and Tibshirani (1993); the test statistic $\mathscr{G}_{n,S}(\lambda)$ can be constructed analogously. It may be possible to extend the results of $\mathscr{G}_{n,L}(h)$ given by Fan et al. (2001) to those of $\mathscr{G}_{n,S}(\lambda)$. An alternative approach to accomplish this is by fitting a cubic smoothing spline to the residuals obtained from parametric fits. Then the testing problem reduces to evaluate whether the residuals are significantly different from zero, namely, testing for no effect of predictor variables. Future research will focus more on this aspect of the extension.

Acknowledgements

The research was partially supported by the Wisconsin Alumni Research Foundation. The author is grateful to the associate editor and two anonymous referees for constructive comments and suggestions that have significantly improved this article, and to Grace Wahba for suggesting this research problem.

Appendix

We first list Conditions (A)–(C), which are used in this paper. Assumptions in Conditions (A) and (B) are not the weakest possible and may be relaxed.

Condition (A)

(A1) The marginal density f of design variable is Lipschitz continuous and bounded away from 0; the design variable has a bounded support Ω.
(A2) $m(x)$ has the continuous $(q+1)$th derivative.
(A3) The kernel function $K(t)$ is a symmetric probability density function with bounded support, and is Lipschitz continuous.
(A4) $0 < E(\varepsilon^4) < \infty$.

Condition (B)

(B1) Let $G_\lambda(s,t)$ be the Green's function for the differential operator $(-1)^p \lambda D^{2p} + f$ with domain $\{g \in C^{2p}[0,1] : g^{(v)}(0) = g^{(v)}(1), \text{ for } p \leqslant v \leqslant 2p-1\}$, where $C^{2p}[0,1] = \{g : g \text{ has } 2p \text{ continuous derivatives}\}$. Let $\rho = \lambda^{1/(2p)}$. There exist finite, positive constants α, ζ and k such that for all $s,t \in [0,1]$, $|G_\lambda(s,t)| \leqslant (k/\rho^{1+\zeta}) \exp(-\alpha|s-t|/\rho)$, and $|\partial G_\lambda(s,t)/\partial s| \leqslant (k/\rho^{2+\zeta}) \exp(-\alpha|s-t|/\rho)$. If $s \neq t$, then $|\partial^2 G_\lambda(s,t)/\partial s\, \partial t| \leqslant (k/\rho^{3+\zeta}) \exp(-\alpha|s-t|/\rho)$. Either $\partial^2 G_\lambda(s,t)/\partial s \partial t$ exists for $s=t$, in which case the last inequality holds, or for all continuous function g on $[0,1]$, $|\partial[\int_0^1 \{\partial G_\lambda(s,t)/\partial t\} g(t)\,\mathrm{d}t]/\partial s| \leqslant (k/\rho^{3+\zeta})\{\int_0^1 (1/2)\exp(-\alpha|s-t|/\rho)|g(t)|\,\mathrm{d}t + |g(s)|\}$.
(B2) $0 < E(\varepsilon^4) < \infty$.
(B3) The design points x_i, $1 \leqslant i \leqslant n$, are generated from a continuous and strictly positive density f, on a finite interval $[0,1]$ without loss of generality, through the relation $\int_0^{x_i} f(x)\,\mathrm{d}x = (i-0.5)/n$.

Remark 2. Condition (B1) follows from Chen (1994, p. 67) and Jayasuriya (1996, p. 1627), and is used to provide an upper bound on $|H_\lambda(i,j)|$, uniformly in $i,j = 1,\dots,n$. Condition (B3) on the design points follows the typical assumptions that have been frequently used in literature on spline-based tests (Eubank and Spiegelman, 1990, p. 388; Eubank and LaRiccia, 1993, p. 2; Chen, 1994, p. 68; Ramil-Novo and González-Manteiga, 2000, p. 819); Jayasuriya (1996, p. 1628) points out the possibility of relaxing these design assumptions.

Condition (C). The error ε has a normal distribution.
Before proving the theorems, we require two lemmas.

Lemma 1. *Let* $\boldsymbol{K}(x) = (2\pi)^{-1}\int_{-\infty}^{+\infty}(1+t^{2p})^{-1}\exp(-\mathrm{i}tx)\,\mathrm{d}t$, *with* $p = 1,2,\dots$. *Denote* $\overbrace{\boldsymbol{K} * \cdots * \boldsymbol{K}}^{r}(x)$ *as the r-times convolution product of* $\boldsymbol{K}(x)$. *Then*

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}(1+t^{2p})^{-r}\,\mathrm{d}t = \overbrace{\boldsymbol{K} * \cdots * \boldsymbol{K}}^{r}(0), \qquad r = 1,2,\dots .$$

In particular, $(2\pi)^{-1}\int_{-\infty}^{+\infty}(1+t^{2p})^{-2s}\,\mathrm{d}t = \int(\overbrace{\boldsymbol{K} * \cdots * \boldsymbol{K}}^{s})^2(x)\,\mathrm{d}x$ *for* $s = 1, 2,\dots$.

The proof of Lemma 1 can be found in Ramil-Novo and González-Manteiga (2000). We now use this result to prove the following asymptotic representations of $\operatorname{tr}(H_\lambda^r)$, in terms of Silverman's kernel function.

Lemma 2. *Let* $\boldsymbol{K}(x)=(2\pi)^{-1}\int_{-\infty}^{+\infty}(1+t^{2p})^{-1}\exp(-\mathrm{i}tx)\,\mathrm{d}t$. *Set* $c(f)=\int_0^1 f(t)^{1/(2p)}\,\mathrm{d}t$. *Then for* $p\geqslant 2$, *as* $n\to\infty$, $\lambda\to 0$, *and* $n\lambda\to\infty$, *it holds that for* P *and* H_λ *given in* (3.3)

$$\operatorname{tr}\{(I-P)^2-(I-H_\lambda)^2\}^2$$
$$=4\lambda^{-1/(2p)}c(f)\int(\boldsymbol{K}-2^{-1}\boldsymbol{K}*\boldsymbol{K})^2(x)\,\mathrm{d}x\{1+\mathrm{o}(1)\}. \tag{6.1}$$

Proof. We first observe that from (3.2), $\operatorname{tr}(H_\lambda^r)=\sum_{j=1}^n(1+\lambda\gamma_{jn})^{-r}$, $r=1,2,\ldots$. Speckman (1981) showed that, if $p\geqslant 2$, $\gamma_{jn}=j^{2p}c_0\{1+\mathrm{o}(1)\}$ for $1\leqslant j\leqslant n$, where $c_0=\pi^{2p}\{\int_0^1 f(t)^{1/(2p)}\,\mathrm{d}t\}^{-2p}$, and the o(1) term is uniform for $j=\mathrm{o}(n^{2/5})$. Combining this result, it follows that

$$\sum_{1\leqslant j\leqslant n^{3/(4p)}}(1+\lambda\gamma_{jn})^{-r}=\sum_{1\leqslant j\leqslant n^{3/(4p)}}(1+\lambda c_0 j^{2p})^{-r}\{1+\mathrm{o}(1)\}$$
$$=(\lambda c_0)^{-1/(2p)}\int_0^\infty\frac{\mathrm{d}t}{(1+t^{2p})^r}\{1+\mathrm{o}(1)\}.$$

On the other hand, $\{\gamma_{jn}\}_{j=1}^n$ are nondecreasing and therefore $\gamma_{jn}\geqslant \mathrm{O}(n^{3/2})$ for $j\geqslant n^{3/(4p)}$, so that

$$\sum_{n^{3/(4p)}<j\leqslant n}(1+\lambda\gamma_{jn})^{-r}\leqslant \mathrm{O}\{n(n^{3/2}\lambda)^{-r}\}. \tag{6.2}$$

The upper bound in (6.2) is thus $\mathrm{o}\{\lambda^{-1/(2p)}\}$ if $r\geqslant 2$. Hence $\operatorname{tr}(H_\lambda^r)=\lambda^{-1/(2p)}c(f)$ $(2\pi)^{-1}\int_{-\infty}^{+\infty}(1+t^{2p})^{-r}\,\mathrm{d}t\,\{1+\mathrm{o}(1)\}$ for $r\geqslant 2$, and its dominating term can be expressed from Lemma 1 in terms of the kernel function $\boldsymbol{K}$. The proof is completed by observing $(I-P)^2-(I-H_\lambda)^2=(2H_\lambda-H_\lambda^2)-P$ and $\{(I-P)^2-(I-H_\lambda)^2\}^2=(2H_\lambda-H_\lambda^2)^2-P$. □

Proof of Theorem 1. Some additional notations will be introduced first. Put $S_n=2^{-1}(\mathrm{RSS}_0-\mathrm{RSS}_1)/\sigma^2$ and $T_n=2^{-1}n(\mathrm{RSS}_0-\mathrm{RSS}_1)/\mathrm{RSS}_1$. Then under H_0, following Theorem 1 of Ramil-Novo and González-Manteiga (2000), we see that

$$\frac{(\mathrm{RSS}_0-\mathrm{RSS}_1)/\sigma^2-\operatorname{tr}\{(I-P)^2-(I-H_\lambda)^2\}}{[2\operatorname{tr}\{(I-P)^2-(I-H_\lambda)^2\}^2]^{1/2}}\xrightarrow{\mathscr{L}}\mathrm{N}(0,1). \tag{6.3}$$

Applying Lemma 2 along with assumptions made on the orders of the approximation errors of $\mathrm{tr}(H_\lambda)$ and $\mathrm{tr}(H_\lambda^2)$, we arrive at

$$\frac{S_n - \lambda^{-1/(2p)}c(f)\{\boldsymbol{K}(0) - 2^{-1}\boldsymbol{K} * \boldsymbol{K}(0)\}}{\{2\lambda^{-1/(2p)}c(f)\int(\boldsymbol{K} - 2^{-1}\boldsymbol{K} * \boldsymbol{K})^2(t)\,\mathrm{d}t\}^{1/2}} \xrightarrow{\mathscr{L}} \mathrm{N}(0,1). \tag{6.4}$$

Furthermore it is easy to verify that under the null hypothesis, $\mathrm{RSS}_1/n = \sigma^2 + \mathrm{O_P}(n^{-1/2}) + \mathrm{O_P}\{(n\lambda^{1/(2p)})^{-1}\}$, which together with (6.4) leads to

$$\frac{T_n - \lambda^{-1/(2p)}c(f)\{\boldsymbol{K}(0) - 2^{-1}\boldsymbol{K} * \boldsymbol{K}(0)\}}{\{2\lambda^{-1/(2p)}c(f)\int(\boldsymbol{K} - 2^{-1}\boldsymbol{K} * \boldsymbol{K})^2(t)\,\mathrm{d}t\}^{1/2}} \xrightarrow{\mathscr{L}} \mathrm{N}(0,1). \tag{6.5}$$

Using the inequality $x/(1+x) \leqslant \log(1+x) \leqslant x$ for $x > -1$, we obtain

$$\mathscr{G}_{n,S}(\lambda) = (n/2)\{(\mathrm{RSS}_0 - \mathrm{RSS}_1)/\mathrm{RSS}_1 + \mathrm{O_P}(n^{-2}\lambda^{-1/p})\} = T_n + \mathrm{O_P}(n^{-1}\lambda^{-1/p}),$$

which means that (6.5) also holds when T_n is replaced by $\mathscr{G}_{n,S}(\lambda)$. The proof is then completed by applying Lemma 1 and the identities

$$\int_0^\infty \frac{\mathrm{d}y}{(1+y^{2p})} = \frac{1}{2p\sin\{\pi/(2p)\}}\pi, \qquad \int_0^\infty \frac{\mathrm{d}y}{(1+y^{2p})^2} = \frac{(2p-1)}{4p^2\sin\{\pi/(2p)\}}\pi,$$

$$\int_0^\infty \frac{\mathrm{d}y}{(1+y^{2p})^3} = \frac{(2p-1)(4p-1)}{16p^3\sin\{\pi/(2p)\}}\pi,$$

$$\int_0^\infty \frac{\mathrm{d}y}{(1+y^{2p})^4} = \frac{(2p-1)(4p-1)(6p-1)}{96p^4\sin\{\pi/(2p)\}}\pi. \qquad \square$$

Proof of Theorem 2. Under the alternative hypothesis, we can obtain $\mathrm{RSS}_1/n = \sigma^2 + \mathrm{O_P}(n^{-1/2}) + \mathrm{O_P}\{n^{-1}\lambda^{-1/(2p)}\} + \mathrm{O}(\lambda)$. Arguing as in the proof of Theorem 1, it suffices to consider first the term $S_n = 2^{-1}(\mathrm{RSS}_0 - \mathrm{RSS}_1)/\sigma^2$. Put $\mathbf{g}_n = (g(x_1),\ldots,g(x_n))^{\mathrm{T}}$ and $\boldsymbol{\varepsilon}_n = (\varepsilon_1,\ldots,\varepsilon_n)^{\mathrm{T}}$. Then we obtain

$$\begin{aligned} S_n = S_n^0 &+ [\mathbf{g}_n^{\mathrm{T}}\{(I-P)^2 - (I-H_\lambda)^2\}\mathbf{g}_n \\ &+ 2\mathbf{g}_n^{\mathrm{T}}\{(I-P)^2 - (I-H_\lambda)^2\}\boldsymbol{\varepsilon}_n]/(2\sigma^2), \end{aligned} \tag{6.6}$$

where the distribution of S_n^0 is identical to that of S_n under the null hypothesis.

We now deal with the latter two terms in (6.6). Let $g_{n\lambda}$ denote the solution to the variational problem,

$$\min_{s \in W_2^p[0,1]} \left[n^{-1}\sum_{i=1}^n \{g(x_i) - s(x_i)\}^2 + \lambda \int_0^1 \{s^{(p)}(x)\}^2\,\mathrm{d}x \right]$$

and define $\mathbf{g}_{n\lambda} = (g_{n\lambda}(x_1),\ldots,g_{n\lambda}(x_n))^{\mathrm{T}}$. Then it follows that

$$n^{-1}\mathbf{g}_n^{\mathrm{T}}(H_\lambda - P)^2\mathbf{g}_n = n^{-1}\mathbf{g}_{n\lambda}^{\mathrm{T}}\mathbf{g}_{n\lambda} - n^{-1}\mathbf{g}_n^{\mathrm{T}}P\mathbf{g}_n. \tag{6.7}$$

The smoothness condition of g and the regularity assumption of the design ensure that $n^{-1}\mathbf{g}_n^{\mathrm{T}}\mathbf{g}_n = \|g\|_f^2 + \mathrm{o}(1)$ and $n^{-1}\mathbf{g}_n^{\mathrm{T}}P\mathbf{g}_n = \mathrm{o}(1)$. Application of Lemma 4.1 in Craven and Wahba (1979) gives that $n^{-1}(\mathbf{g}_{n\lambda} - \mathbf{g}_n)^{\mathrm{T}}(\mathbf{g}_{n\lambda} - \mathbf{g}_n) \leqslant \lambda \int_0^1 \{g^{(p)}(x)\}^2\,\mathrm{d}x$. Combining this inequality with the triangle inequality leads to

$$\begin{aligned} n^{-1}\mathbf{g}_{n\lambda}^{\mathrm{T}}\mathbf{g}_{n\lambda} &\geqslant 2^{-1}n^{-1}\mathbf{g}_n^{\mathrm{T}}\mathbf{g}_n - n^{-1}(\mathbf{g}_{n\lambda} - \mathbf{g}_n)^{\mathrm{T}}(\mathbf{g}_{n\lambda} - \mathbf{g}_n) \\ &\geqslant 2^{-1}\|g\|_f^2 - \lambda \int_0^1 \{g^{(p)}(x)\}^2\,\mathrm{d}x + \mathrm{o}(1) \end{aligned} \tag{6.8}$$

in which the negligible term, denoted by o(1), is independent of λ. We also deduce from Chen (1994), the representation expressed as

$$n^{-1}\mathbf{g}_n^{\mathrm{T}}(H_\lambda - H_\lambda^2)\mathbf{g}_n = \lambda\|g_{n\lambda}^{(p)}\|_f^2\{1 + \mathrm{o}(1)\}. \tag{6.9}$$

Furthermore, it can easily be shown that the cross term $\mathbf{g}_n^{\mathrm{T}}\{(I-P)^2 - (I-H_\lambda)^2\}\boldsymbol{\varepsilon}_n = \|g\|_f \mathrm{O_P}(n^{1/2})$. This along with (6.6), (6.7) and (6.9) yields

$$\begin{aligned} \frac{r_{\boldsymbol{K}}S_n - a_n}{(2a_n)^{1/2}} &= \frac{r_{\boldsymbol{K}}S_n^0 - a_n}{(2a_n)^{1/2}} + r_{\boldsymbol{K}}n/\{2\sigma^2(2a_n)^{1/2}\}[n^{-1}\mathbf{g}_{n\lambda}^{\mathrm{T}}\mathbf{g}_{n\lambda} + \mathrm{o}(1) \\ &\quad + 2\lambda\|g_{n\lambda}^{(p)}\|_f^2\{1 + \mathrm{o}(1)\} + 2\|g\|_f \mathrm{O_P}(n^{-1/2})], \end{aligned} \tag{6.10}$$

where $a_n = r_{\boldsymbol{K}}c_{\boldsymbol{K}}\lambda^{-1/(2p)}c(f)$, and $\frac{r_{\boldsymbol{K}}S_n^0 - a_n}{(2a_n)^{1/2}} \xrightarrow{\mathscr{L}} \mathrm{N}(0,1)$ following directly from the proof of Theorem 1. This completes the proof of the first part.

Hence (6.8) and (6.10) imply that the test statistics $(r_{\boldsymbol{K}}S_n - a_n)/(2a_n)^{1/2}$ or $(r_{\boldsymbol{K}}\mathscr{G}_{n,S} - a_n)/(2a_n)^{1/2}$ can detect a signal $g \in L_2^f[0,1] \backslash \{1, x, \ldots, x^{p-1}\}$ with $g \in W_2^p[0,1]$ from the alternative set, if

$$\|g\|_f^2 \geqslant C\{\lambda + n^{-1}\lambda^{-1/(4p)}\} \tag{6.11}$$

holds for some sufficiently large C. In this case, minimization over λ in the lower bound above leads to the optimal rate of smoothing parameter $\lambda = \mathrm{O}\{n^{-4p/(4p+1)}\}$, and thus (6.11) is satisfied for $\|g\|_f \geqslant C_1 n^{-2p/(4p+1)}$ for some sufficiently large C_1. This finishes the proof. □

Proof of Theorem 3. The arguments that we use here are similar to those of Serfling (1980, Chapter 10) or Eubank and LaRiccia (1993).

To compute the relative efficiency, suppose that the spline-based test uses n_2 observations, whereas the test using local polynomial smoother is to be based on n_1 observations. Denote by $\lambda_2 = C_2^{2p} n_2^{-4p/(4p+1)}\{c(f)\}^{2p}$ and $h_1 = C_1 n_1^{-2/(4p+1)}$ the corresponding smoothing parameters. Now consider the local alternative for $\mathscr{G}_{n_2,S}(\lambda_2)$ to be of the form $\sum_{j=0}^{p-1} \theta_j x^j + \{n_2\lambda_2^{1/(4p)}/\{c(f)\}^{1/2}\}^{-1/2} g_2(x)$, and let the local alternative for $\mathscr{G}_{n_1,L}(h_1)$ be of the form $\sum_{j=0}^{p-1} \theta_j x^j + \{n_1 h_1^{1/2}\}^{-1/2} g_1(x)$, with $g_i \in L_2^f[0,1] \cap W_2^p[0,1] \setminus \{1, \ldots, x^{p-1}\}$, $i = 1, 2$. In order to compare the powers of the two tests, the alternatives must coincide asymptotically, which leads to the requirement

$$\lim \left\{ \frac{n_2 \lambda_2^{1/(4p)} / \{c(f)\}^{1/2}}{n_1 h_1^{1/2}} \right\}^{-1/2} g_2(x) = g_1(x) \tag{6.12}$$

for all $x \in [0,1]$.

Under the local alternative above, the asymptotic power of $\mathscr{G}_{n_2,S}(\lambda_2)$, by Eq. (6.10), is $1-\Phi\{z_\alpha - (r_{\boldsymbol{K}}/c_{\boldsymbol{K}})^{1/2}\|g_2\|_f^2/(2\sigma^2\sqrt{2})\}$, with Φ the standard normal distribution function. Similarly, the $\mathscr{G}_{n_1,L}(h_1)$ test using n_1 observations has the limiting power, given by $1-\Phi\{z_\alpha - (r_{\mathscr{K}}/c_{\mathscr{K}})^{1/2}\|g_1\|_f^2/(2\sigma^2\sqrt{2})\}$. If we equate these two limiting powers and make use of (6.12), we obtain

$$\lim(n_1/n_2) = \left\{ (C_2/C_1)^{1/2} \frac{(r_{\boldsymbol{K}}/c_{\boldsymbol{K}})^{1/2}}{(r_{\mathscr{K}}/c_{\mathscr{K}})^{1/2}} \right\}^{(4p+1)/(4p)}.$$

This establishes the conclusion of (4.1). Similar arguments can be applied to the proof of (4.2). □

References

Azzalini, A., Bowman, A.N., 1993. On the use of nonparametric regression for checking linear relationships. J. Roy. Statist. Soc. Ser. B 55, 549–557.

Azzalini, A., Bowman, A.N., Härdle, W., 1989. On the use of nonparametric regression for model checking. Biometrika 76, 1–11.

Bowman, A., Young, S., 1996. Graphical comparison of nonparametric curves. Appl. Statist. 45, 83–98.

Cai, Z., Fan, J., Li, R., 2000. Efficient estimation and inferences for varying–coefficient models. J. Amer. Statist. Assoc. 95, 888–902.

Chen, J.C., 1994. Testing goodness of fit of polynomial models via spline smoothing techniques. Statist. Probab. Lett. 19, 65–76.

Cox, D., 1983. Asymptotics of M-type smoothing splines. Ann. Statist. 11, 530–551.

Cox, D., Koh, E., 1989. A smoothing spline based test of model adequacy in polynomial regression. Ann. Inst. Statist. Math. 41, 383–400.

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

Craven, P., Wahba, G., 1979. Smoothing noisy data with spline functions. Numer. Math. 31, 377–403.

Demmler, A., Reinsch, C., 1975. Oscillation matrices with spline functions. Numer. Math. 24, 375–382.

Eubank, R.L., 1988. Spline Smoothing and Nonparametric Regression. Marcel Dekker, New York.

Eubank, R.L., LaRiccia, V.N., 1993. Testing for no effect in nonparametric regression. J. Statist. Plann. Inference 36, 1–14.

Eubank, R.L., Spiegelman, C.H., 1990. Testing the goodness of fit of a linear model via nonparametric regression techniques. J. Amer. Statist. Assoc. 85, 387–392.

Fan, J., Gijbels, I., 1996. Local Polynomial Modeling and its Applications. Chapman & Hall, London.

Fan, J., Zhang, C.M., Zhang, J., 2001. Generalized likelihood ratio statistics and Wilks phenomenon. Ann. Statist. 29, 153–193.

Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. Ann. Statist. 21, 1926–1947.

Hart, J.D., Wehrly, T.S., 1992. Kernel regression when the boundary region is large, with application to testing the adequacy of polynomial model. J. Amer. Statist. Assoc. 87, 1018–1024.

Hastie, T.J., Tibshirani, R.J., 1993. Varying-coefficient models (with discussion). J. Roy. Statist. Soc. Ser. B 55, 757–796.

Huang, L.S., 2001. A roughness–penalty view of kernel smoothing. Statist. Probab. Lett. 52, 85–89.

Ingster, Yu.I., 1982. Minimax nonparametric detection of signals in white Gaussian noise. Problems Inform. Transmission 18, 130–140.

Jayasuriya, B.R., 1996. Testing for polynomial regression using nonparametric regression techniques. J. Amer. Statist. Assoc. 91, 1626–1630.

Kou, S.C., 2000. Bounding the eigenvalues of smoothing splines and its application to Bayesian efficiency. Department of Statistics, Stanford University, unpublished.

Lepski, O.V., Spokoiny, V.G., 1999. Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. Bernoulli 5, 333–358.

Ramil-Novo, L.A., González-Manteiga, W., 2000. F-tests and regression ANOVA based on smoothing spline estimators. Statist. Sinica 10, 819–837.

Reinsch, C., 1967. Smoothing by spline functions. Numer. Math. 10, 177–183.

Serfling, R., 1980. Approximation Theorems of Mathematical Statistics. Wiley, New York.

Silverman, B.W., 1984a. Spline smoothing: the equivalent variable kernel method. Ann. Statist. 19, 898–916.

Silverman, B.W., 1984b. A fast and efficient cross-validation method for smoothing parameter choice in spline regression. J. Amer. Statist. Assoc. 79, 584–589.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall, London.

Speckman, P., 1981. The asymptotic integrated mean squared error for smoothing noisy data by spline functions. Department of Statistics, University of Missouri, unpublished.

Wahba, G., 1975. Smoothing noisy data with spline functions. Numer. Math. 24, 383–393.

Wahba, G., 1990. Spline Models for Observational Data. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Yanagimoto, T., Yanagimoto, M., 1987. The use of marginal likelihood for a diagnostic test for the goodness of fit of the simple linear regression model. Technometrics 29, 95–101.

Young, S.G., Bowman, A.W., 1995. Non-parametric analysis of covariance. Biometrics 51, 920–931.

Zhang, C.M., 2003. Adaptive tests of regression functions via multi-scale generalized likelihood ratios. Canad. J. Statist. 31, 151–157.