

HYPOTHESIS TESTING IN SMOOTHING SPLINE MODELS

ANNA LIU¹ and YUEDONG WANG²

*Department of Statistics & Applied Probability, University of California,
Santa Barbara, CA 93106, USA*

(Revised 8 June 2003; In final form 11 September 2003)

Nonparametric regression models are often used to check or suggest a parametric model. Several methods have been proposed to test the hypothesis of a parametric regression function against an alternative smoothing spline model. Some tests such as the locally most powerful (LMP) test by Cox *et al.* (Cox, D., Koh, E., Wahba, G. and Yandell, B. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Stat.*, **16**, 113–119.), the generalized maximum likelihood (GML) ratio test and the generalized cross validation (GCV) test by Wahba (Wahba, G. (1990). Spline models for observational data. *CBMS-NSF Regional Conference Series in Applied Mathematics*, SIAM.) were developed from the corresponding Bayesian models. Their frequentist properties have not been studied. We conduct simulations to evaluate and compare finite sample performances. Simulation results show that the performances of these tests depend on the shape of the true function. The LMP and GML tests are more powerful for low frequency functions while the GCV test is more powerful for high frequency functions. For all test statistics, distributions under the null hypothesis are complicated. Computationally intensive Monte Carlo methods can be used to calculate null distributions. We also propose approximations to these null distributions and evaluate their performances by simulations.

Keywords: Bayesian models for smoothing splines; Connections between linear mixed effects models and smoothing splines; GCV test; GML test; F-test; LMP test; Symmetrized Kullback-Leibler test

1 INTRODUCTION

As a popular nonparametric regression method, spline smoothing has attracted a great deal of attention. Most research in the literature concentrates on estimation, while inference, especially hypothesis testing, has received less attention. Several test procedures were developed only for simple hypotheses of simple spline models. Their properties and performances are not well understood. The aim of this study is to evaluate and compare some existing smoothing spline based tests using simulations.

Consider the univariate nonparametric regression model

$$y_i = f(t_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad 0 \leq t_i \leq 1, \quad (1)$$

where y_i is the response, ε_i is the random error and $\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$. f is assumed to be in an infinite dimensional model space to be specified later.

¹E-mail: annaliu@pstat.ucsb.edu

²E-mail: yuedong@pstat.ucsb.edu

One of the most useful applications of the nonparametric regression models is to check or suggest a parametric model. Parametric models, especially linear models, are preferred in practice because of their simplicity and interpretability. Diagnostic investigations of the departures from these parametric assumptions are necessary to avoid misleading results. If some specific alternative form is suspected, a simple lack of fit test can be performed. However, this kind of test would not perform well for other departures from the parametric model, especially those orthogonal to the suspected alternative. For example, to detect departure from a linear model, one may consider a quadratic polynomial as the alternative. Then the higher order departure may be missed. Tests performing well for general departures in a large model space are desirable.

Most existing methods for testing general departures from a parametric model are based on nonparametric regression models such as kernel estimation (Azzalini and Bowman, 1993), local polynomial regression (Cleveland and Devlin, 1988) and smoothing spline. In this paper we focus on the tests based on smoothing spline models. The connection between smoothing spline models and Bayesian models (or mixed effects models) simplifies certain hypothesis tests. Also, the general form of smoothing spline models allows us to consider many different situations in a unified fashion.

Cox *et al.* (1988) showed that for the hypothesis of f being a polynomial of degree m ($m \geq 0$) versus f being smooth, there is no uniformly most powerful (UMP) test. Thus they proposed to use a locally most powerful (LMP) test. Wahba (1990) proposed two tests based on the generalized maximum likelihood (GML) and the generalized cross validation (GCV) scores. For nonGaussian data, Xiang and Wahba (1995) developed the symmetrized Kullback-Leibler (SKL) test based on the SKL distance between the function estimated under the null hypothesis and the function estimated under the alternative. We are going to examine the performance of these tests for Gaussian data.

Raz (1990) developed a permutation test for the hypothesis of independence between the response and the covariates without assuming any particular error distribution. Two generalized F tests were studied by Raz (1990) in the context of general nonparametric regression. However, their performances were not discussed.

In Section 2, a brief introduction to smoothing splines is given. In Section 3, we review some existing tests and develop approximations to null distributions. We evaluate and compare these tests and approximations in Section 4. Section 5 concludes with some remarks and potential research topics.

2 SMOOTHING SPLINE MODELS

In this section we briefly review smoothing spline models, their corresponding Bayesian models and connections with linear mixed effects models (LMM). For simplicity, we limit our discussions to polynomial splines on $[0, 1]$. All tests in this study can be written in terms of general spline models on arbitrary domains (Wahba, 1990). Thus these tests can be used to test more complicated hypothesis under general spline models.

In model (1), assume that $f \in W_m$, where

$$W_m = \{g | g, \dots, g^{(m-1)} \text{ are absolutely continuous, } g^{(m)} \in \mathcal{L}_2[0, 1]\}.$$

The smoothing spline estimate of f , \hat{f}_λ , is the minimizer of

$$\frac{1}{n} \sum_{i=1}^n (y_i - f(t_i))^2 + \lambda \int_0^1 (f^{(m)}(u))^2 du, \quad (2)$$

where λ is a smoothing parameter which controls the trade-off between the goodness-of-fit and the smoothness of the estimate.

Let $\mathbf{y} = (y_1, \dots, y_n)'$, $\phi_v(t) = t^{v-1}/(v-1)!$, $v = 1, \dots, m$, and $R^1(s, t) = \int_0^{\min(s,t)} (s-u)^{m-1}(t-u)^{m-1} du / ((m-1)!)^2$. Consider $T_{n \times m} = \{\phi_v(t_i)\}_{i=1}^n \}_{v=1}^m$ and $\Sigma_{n \times n} = \{R^1(t_i, t_j)\}_{i=1}^n \}_{j=1}^n$. Kimeldorf and Wahba (1971) showed that the solution to Eq. (2) has the form

$$\hat{f}_\lambda(t) = \sum_{v=1}^m d_v \phi_v(t) + \sum_{i=1}^n c_i R^1(t, t_i),$$

where $\mathbf{c} = (c_1, \dots, c_n)'$ and $\mathbf{d} = (d_1, \dots, d_m)'$ are solutions to

$$\begin{pmatrix} T & \Sigma + n\lambda I \\ \mathbf{0} & T' \end{pmatrix} \begin{pmatrix} \mathbf{d} \\ \mathbf{c} \end{pmatrix} = \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix}. \tag{3}$$

The system (3) is definite when T is of full column rank, which we assume to be true in this paper. Thus $\hat{\mathbf{f}}_\lambda = (\hat{f}_\lambda(t_1), \dots, \hat{f}_\lambda(t_n))' = T\mathbf{d} + \Sigma\mathbf{c}$ is always unique. Let

$$T = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} R \\ 0 \end{pmatrix}$$

be the QR decomposition of T . One may check that $\hat{\mathbf{f}}_\lambda$ is a linear function of \mathbf{y} : $\hat{\mathbf{f}}_\lambda = A(\lambda)\mathbf{y}$, where $A(\lambda)$ is the ‘hat’ matrix. It can be verified that

$$A(\lambda) = I - n\lambda Q_2(Q_2'(\Sigma + n\lambda I)Q_2)^{-1}Q_2'. \tag{4}$$

Note that $A(\lambda)$ is symmetric, but usually not idempotent.

The smoothing spline estimate can be obtained from the Bayesian point of view. Assume the following prior to f

$$F(t) = \sum_{v=1}^m \theta_v \phi_v(t) + b^{1/2}X(t),$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)' \sim N(0, aI)$, a and b are positive constants, and $X(t)$ is a zero mean Gaussian stochastic process independent of $\boldsymbol{\theta}$ with covariance $EX(s)X(t) = R^1(s, t)$.

Consider

$$y_i = F(t_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad t_i \in [0, 1], \tag{5}$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)' \sim N(0, \sigma^2 I)$ and is independent of F . Wahba (1990) showed that with $\lambda = \sigma^2/nb$,

$$\lim_{a \rightarrow \infty} E(F(t)|\mathbf{y}) = \hat{f}_\lambda(t).$$

With $a \rightarrow \infty$, diffuse priors are assumed for the coefficients of the polynomials of degree less than m .

Smoothing spline models can also be connected to certain LMM. Consider the following LMM

$$\mathbf{y} = T\mathbf{d} + \mathbf{u} + \boldsymbol{\varepsilon}, \tag{6}$$

where \mathbf{d} are the fixed effects, \mathbf{u} are the random effects and $\mathbf{u} \sim N(0, b\Sigma)$, $\boldsymbol{\varepsilon}$ are random errors and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2I)$, and \mathbf{u} and $\boldsymbol{\varepsilon}$ are independent. Wang (1998a) showed that the smoothing spline estimate evaluated at the design points, $\hat{\mathbf{f}}_\lambda$, is the same as the best linear unbiased prediction (BLUP) estimate in Eq. (6).

3 EXISTING TEST METHODS

Let $\mathcal{H}_0 = \text{span}\{\phi_v, v = 1, \dots, m\}$. Often we are interested in testing the hypothesis that f is a polynomial of degree $m - 1$ or less

$$H_0: f \in \mathcal{H}_0, \quad H_1: f \in W_m \quad \text{and} \quad f \notin \mathcal{H}_0. \tag{7}$$

It is easy to see that $\lambda = \infty$ in Eq. (2), or equivalently $b = 0$ in the corresponding Bayesian and mixed effects models, leads to $f \in \mathcal{H}_0$. Thus the hypothesis (7) can be re-expressed as

$$H_0: \lambda = \infty, \quad H_1: \lambda < \infty, \tag{8}$$

or

$$H_0: b = 0, \quad H_1: b > 0. \tag{9}$$

Notice that $\mathbf{y} \sim N(0, aTT' + b\Sigma + \sigma^2I)$ under the Bayesian model (5) and $\mathbf{y} \sim N(T\mathbf{d}, b\Sigma + \sigma^2I)$ under the mixed effects model (6). Let $\mathbf{w} = Q_2'\mathbf{y}$, then $\mathbf{w} \sim N(0, bQ_2'\Sigma Q_2 + \sigma^2I)$ under both the mixed effects and the Bayesian models. It is clear that the transformation $Q_2'\mathbf{y}$ eliminates contribution from the model under the null hypothesis. Thus \mathbf{w} reflects signals, if any, from $W_m \ominus \mathcal{H}_0$.

Let the spectral decomposition of $Q_2'\Sigma Q_2$ be UDU' , where $D = \text{diag}(\lambda_{vn}, v = 1, \dots, n - m)$, and λ_{vn} 's are the eigenvalues of $Q_2'\Sigma Q_2$ ordered so that $\lambda_{1n} \geq \lambda_{2n} \geq \dots \geq \lambda_{n-mn}$. Let $\mathbf{z} = U'\mathbf{w}$, then

$$\mathbf{z} \sim N(0, bD + \sigma^2I). \tag{10}$$

Note that z_v , the v th component of \mathbf{z} , is the projection of \mathbf{w} in the direction of the v th column (eigenvector) of $U(Q_2'\Sigma Q_2)$.

3.1 LMP Tests

Cox *et al.* (1988) showed that the UMP test does not exist for hypothesis (9) under model (5). When σ^2 is known, they proposed an LMP test which rejects the null hypothesis for large values of

$$t_{\text{LMP}} = \sum_{v=1}^{n-m} \lambda_{vn} z_v^2. \tag{11}$$

When σ^2 is unknown they proposed an approximate LMP which rejects the null hypothesis for large values of

$$t_{\text{appLMP}} = \frac{\sum_{v=1}^{n-m} \lambda_{vn} z_v^2}{\sum_{v=1}^{n-m} z_v^2}. \quad (12)$$

Let $l(b, \sigma^2 | \mathbf{z})$ denote the log-likelihood of b and σ^2 given \mathbf{z} . Then

$$l(b, \sigma^2 | \mathbf{z}) = -\frac{n-m}{2} \log(2\pi) - \frac{1}{2} \sum_{v=1}^{n-m} \log(b\lambda_{vn} + \sigma^2) - \frac{1}{2} \sum_{v=1}^{n-m} \frac{z_v^2}{b\lambda_{vn} + \sigma^2}.$$

Note that \mathbf{z} is based on an orthogonal contrast of the original observations which eliminates the fixed effects. Thus $l(b, \sigma^2 | \mathbf{z})$ is the so-called restricted log-likelihood in the mixed effects literature.

It is not difficult to check that the LMP test is equivalent to the score test (Cox and Hinkley, 1974) defined by

$$t_{\text{score}} = \frac{U_b(0, \sigma^2)}{\sqrt{I_{bb}(0, \sigma^2)}}, \quad (13)$$

where $U_b(b, \sigma^2)$ is the efficient score defined as $\partial l(b, \sigma^2 | \mathbf{z}) / \partial b$ and $I_{bb}(b, \sigma^2)$ is the Fisher information of b . When it is unknown, σ^2 can be replaced by the maximum likelihood estimate (MLE) under the null hypothesis (9), $\hat{\sigma}^2 = \sum_{v=1}^{n-m} z_v^2 / (n-m)$, which leads to the approximate LMP test.

The test statistic t_{appLMP} does not follow a simple distribution under H_0 . It is straightforward to simulate the null distribution (Wahba, 1990). The p -value, $P(t_{\text{appLMP}} > t_{\text{appLMP}}^{\text{obs}}) = P(\sum_{v=1}^{n-m} (\lambda_{vn} - t_{\text{appLMP}}^{\text{obs}}) z_v^2 > 0)$, can also be calculated numerically using the algorithm in Davies (1980). We find that this nonsimulation-based approximation method is very fast and agrees with the results from the Monte Carlo method.

3.2 GML Test

Since $b = \sigma^2 / n\lambda$, the log-likelihood from \mathbf{z} can be re-expressed as

$$l(\lambda, b | \mathbf{z}) = -\frac{1}{2}(n-m) \log b - \frac{1}{2} \sum_{v=1}^{n-m} \log(\lambda_{vn} + n\lambda) - \frac{1}{2b} \sum_{v=1}^{n-m} \frac{z_v^2}{\lambda_{vn} + n\lambda} + C,$$

where C is a constant.

For fixed λ , maximizing the log-likelihood with respect to b , we have

$$\hat{b}_\lambda = \frac{1}{n-m} \sum_{v=1}^{n-m} \frac{z_v^2}{\lambda_{vn} + n\lambda}.$$

Then the profiled likelihood of λ is

$$L(\lambda | \mathbf{z}) = \exp(l(\lambda, \hat{b}_\lambda | \mathbf{z})) = C_1 \left(\frac{\sum_{v=1}^{n-m} z_v^2 / (\lambda_{vn} + n\lambda)}{\prod_{v=1}^{n-m} (\lambda_{vn} + n\lambda)^{-1/(n-m)}} \right)^{-(n-m)/2}, \quad (14)$$

where C_1 is a constant.

The GML estimate of λ , $\hat{\lambda}_{\text{GML}}$, is the maximizer of Eq. (14). Wahba (1990) defined the GML test statistic for the hypothesis (8) as

$$t_{\text{GML}} = \left[\frac{L(\hat{\lambda}_{\text{GML}}|\mathbf{z})}{L(\infty|\mathbf{z})} \right]^{-2/(n-m)} = \frac{\sum_{v=1}^{n-m} z_v^2 / (\lambda_{vn} + n\hat{\lambda}_{\text{GML}})}{\prod_{v=1}^{n-m} (\lambda_{vn} + n\hat{\lambda}_{\text{GML}})^{-1/(n-m)} \sum_{v=1}^{n-m} z_v^2}. \tag{15}$$

The null hypothesis is rejected when t_{GML} is too small.

It is difficult to derive the null distribution for t_{GML} . The standard theory for likelihood ratio tests does not apply because the parameter is on the boundary under the null hypothesis. The nonstandard asymptotic theory, developed by Self and Liang (1987), which states that $-(n - m) \log t_{\text{GML}}$ has an asymptotic null distribution of a 50:50 mixture of χ_1^2 and χ_0^2 does not apply either because of the lack of replicated observations. Crainiceanu *et al.* (2003) reported the same finding for P-spline models. In a subsequent paper, Crainiceanu and Ruppert (2003) provided the asymptotic distributions of likelihood ratio tests for linear mixed models. Monte Carlo methods are used to obtain quantiles of these asymptotic distributions.

The direct Monte Carlo method simulates l samples of $-(n - m) \log t_{\text{GML}}$ under the null hypothesis. Denote $-(n - m) \log t_{\text{GML}}$ based on data as x_0 and suppose that $x_0 > 0$. Then the true p -value is

$$p = P(-(n - m) \log t_{\text{GML}} > x_0 | H_0).$$

We generate l samples of \mathbf{z} from $N(0, I)$ (without loss of generality, we set $\sigma^2 = 1$), calculate $\hat{\lambda}_{\text{GML}}$ for each sample, and construct t_{GML} for each sample. Let x_1, \dots, x_l denote the l samples of $-(n - m) \log t_{\text{GML}}$. Then p is estimated by

$$\hat{p} = \frac{1}{l} \sum_{i=1}^l I(x_i > x_0),$$

where $I(\cdot)$ is the indicator function. Then $E\hat{p} = p$ and $\text{Var}(\hat{p}) = p(1 - p)/l$. This approach usually requires a very large l . For example, to have a margin of error $2\sqrt{\text{Var}(\hat{p})}$ bounded by 0.005, l has to be at least 40,000. Note that $\hat{\lambda}_{\text{GML}}$ is computed for each sample. Therefore, this approach is computationally intensive.

Our simulation results suggest that the null distribution of $-(n - m) \log t_{\text{GML}}$ can be well approximated by a mixture of χ_1^2 and χ_0^2 , denoted by $r\chi_0^2 + (1 - r)\chi_1^2$. However, the ratio r is not fixed. It depends on the order m , sample size n and the design points t_i . Thus we propose an alternative method that estimates the ratio r first and then calculates the p -value based on the mixture of χ_1^2 and χ_0^2 with the estimated r . The motivation behind this approach is that a relatively small sample size k is required to estimate r .

We now compare sample sizes required by these two approaches. For the alternative approach, let x'_1, \dots, x'_k be k random samples of $-(n - m) \log t_{\text{GML}}$ under the null hypothesis. We estimate r by

$$\hat{r} = \frac{1}{k} \sum_{i=1}^k I(x'_i = 0).$$

Then $E\hat{r} = r$ and $\text{Var}(\hat{r}) = r(1 - r)/k$. Note that χ_0^2 represents a point mass at zero. Thus the p -value is estimated by

$$\tilde{p} = (1 - \hat{r})P(\chi_1^2 > x_0).$$

Assuming that the null distribution of nonzero $-(n - m) \log t_{\text{GML}}$ is exactly χ_1^2 , we have $E\tilde{p} = p$ and $\text{Var}(\tilde{p}) = rp^2/(k(1 - r))$.

For $\text{Var}(\hat{p}) = \text{Var}(\tilde{p})$, we need $k = rpl/((1 - r)(1 - p))$. Based on our simulations with $m = 2, n = 100$ and a uniform design in $[0, 1]$, r is usually around 0.7. It is easy to check that for $p = 0.05$ and $r = 0.7$, we have $k \approx 0.12l$. Thus about $k = 5000$ samples are needed for the alternative method if $l = 40,000$.

Simulation results in Section 5 indicate that approximations based on the alternative approach are accurate when the sample size is large.

3.3 F-type Tests

For the hypothesis (7) under model (1), the usual F test statistic will not follow an F distribution because the hat matrix $A(\lambda)$ is not idempotent. Two F -type tests were mentioned by Raz (1990) for the following hypothesis

$$H_0: f = \text{constant}, \quad H_1: f \neq \text{constant} \tag{16}$$

in the context of general nonparametric regressions. They were used to derive the permutation test and their performances were not investigated. In this section, we first extend these two F -type statistics for our hypothesis (7). Then we compare them with the GCV test proposed by Wahba (1990) and the SKL test proposed by Xiang and Wahba (1995).

Let \hat{f}_0 be the maximum likelihood estimate of the regression function under the null model. Then $\hat{\mathbf{f}}_0 = H\mathbf{y}$, where $\hat{\mathbf{f}}_0 = (\hat{f}_0(t_1), \dots, \hat{f}_0(t_n))'$ and $H = T(T'T)^{-1}T'$. Note that H is an idempotent hat matrix and $A(\lambda)H = HA(\lambda) = H$.

Define

$$\begin{aligned} S_1 &= \sum_{i=1}^n (\hat{f}_i(t_i) - \hat{f}_0(t_i))^2, \\ S_2 &= \sum_{i=1}^n (y_i - \hat{f}_i(t_i))^2, \\ S_3 &= \sum_{i=1}^n (y_i - \hat{f}_0(t_i))^2, \end{aligned} \tag{17}$$

where S_1 measures the difference between \hat{f}_0 and \hat{f}_i , S_2 is the residual sum of squares under H_1 and S_3 is the residual sum of squares under H_0 .

In terms of the hat matrices, Eq. (17) can be re-expressed as $S_1 = \mathbf{y}'(A(\lambda) - H)^2\mathbf{y}$, $S_2 = \mathbf{y}'(I - A(\lambda))^2\mathbf{y}$ and $S_3 = \mathbf{y}'(I - H)\mathbf{y}$. In terms of \mathbf{z} , we have

$$\begin{aligned} S_1 &= \sum_{v=1}^{n-m} \left(\frac{\lambda_{vn}/n\lambda}{1 + \lambda_{vn}/n\lambda} \right)^2 z_v^2 = S_2 + S_3 - 2 \sum_{v=1}^{n-m} \frac{z_v^2}{1 + \lambda_{vn}/n\lambda}, \\ S_2 &= \sum_{v=1}^{n-m} \frac{z_v^2}{(1 + \lambda_{vn}/n\lambda)^2}, \\ S_3 &= \sum_{v=1}^{n-m} z_v^2. \end{aligned} \tag{18}$$

Contrary to the parametric case, the equality $S_1 + S_2 = S_3$ usually does not hold. Similar to Raz (1990), we consider two generalizations of the standard F test statistic

$$F_1 = \frac{(n - g_1)S_1}{(g_1 - m)(S_3 - S_1)} \quad \text{with} \quad g_1 = \text{tr}(A^2(\lambda)), \tag{19}$$

and

$$F_2 = \frac{g_1^*(S_3 - S_2)}{(n - g_1^* - m)S_2} \quad \text{with} \quad g_1^* = \text{tr}((I - A(\lambda))^2). \tag{20}$$

When λ is fixed, the permutation test statistic in Raz (1990), $(n - 1)S_1/S_3$, is equivalent to F_1 . Cantoni and Hastie (2002) considered a different hypothesis where λ (or b) was fixed under the alternative. Their F test statistic is equivalent to F_2 when σ^2 is estimated under their alternative hypothesis. They used approximation methods for linear combinations of χ^2 variables to compute p -values (Davies, 1980). Contrary to the LMP test, this method cannot be used here because the smoothing parameters are not fixed under the alternative hypothesis. In Section 5, we investigate the performances of F_1 and F_2 tests with a data-based choice of λ .

3.3.1 The SKL Test

For nonGaussian data, Xiang and Wahba (1995) proposed the SKL test based on the SKL distance between \hat{f}_λ and \hat{f}_0 :

$$t_{\text{SKL}} = \frac{1}{n} \left[E_{\hat{f}_0} \left(\log \left(\frac{\hat{f}_0}{\hat{f}_\lambda} \right) \right) + E_{\hat{f}_\lambda} \left(\log \left(\frac{\hat{f}_\lambda}{\hat{f}_0} \right) \right) \right].$$

For Gaussian data, it reduces to

$$t_{\text{SKL}} = \frac{1}{n\sigma^2} \|\hat{f}_\lambda - \hat{f}_0\|^2 = \frac{1}{n\sigma^2} S_1. \tag{21}$$

When σ^2 is estimated by $S_3/(n - m)$, $t_{\text{SKL}} = (n - m)S_1/nS_3$. Thus t_{SKL} is equivalent to F_1 for fixed λ . The performance of the SKL test is compared with F -type tests and other tests in Section 5 with λ estimated from data.

3.3.2 GCV Test

The GCV test is based on the following GCV score (Wahba, 1990):

$$V(\lambda) = \frac{n\|(I - A(\lambda))\mathbf{y}\|^2}{(\text{tr}(I - A(\lambda)))^2}.$$

The GCV estimate of λ , $\hat{\lambda}_{\text{GCV}}$, is the minimizer of $V(\lambda)$.

Wahba (1990) defined the GCV test statistic as

$$t_{\text{GCV}} = \frac{V(\hat{\lambda}_{\text{GCV}})}{V(\infty)} = (n - m)^2 \frac{\sum_{v=1}^{n-m} z_v^2 / (1 + \lambda_{vn}/n\hat{\lambda}_{\text{GCV}})^2}{[\sum_{v=1}^{n-m} 1 / (1 + \lambda_{vn}/n\hat{\lambda}_{\text{GCV}})]^2} \frac{1}{\sum_{v=1}^{n-m} z_v^2}.$$

H_0 is rejected when t_{GCV} is too small. It is easily seen that t_{GCV} is equivalent to F_2 if the smoothing parameter is fixed instead of being estimated from the GCV score. Again, the performance of the GCV test is compared with F -type tests and the other tests in Section 5 with λ estimated from data.

4 AN OVERALL COMPARISON

All tests except F_1 and F_2 can be written in the form $\sum_{v=1}^{n-m} a_v z_v^2 / \sum_{v=1}^{n-m} z_v^2$, where coefficients for the approximate LMP, GML, GCV and SKL tests are $a_v^{\text{LMP}} = \lambda_{vn}$, $a_v^{\text{GML}} = \prod_{v=1}^{n-m} (\lambda_{vn} + n\hat{\lambda}_{\text{GML}})^{1/(n-m)} / (\lambda_{vn} + n\hat{\lambda}_{\text{GML}})$, $a_v^{\text{GCV}} = (n-m)^2 / [(1 + \lambda_{vn}/n\hat{\lambda}_{\text{GCV}})^2 \sum_{v=1}^{n-m} 1/(1 + \lambda_{vn}/n\hat{\lambda}_{\text{GCV}})^2]$ and $a_v^{\text{SKL}} = (n-m)(\lambda_{vn}/n\hat{\lambda})^2 / [n(1 + \lambda_{vn}/n\hat{\lambda})^2]$ with $\hat{\lambda}$ being an estimate of λ .

Note that a_v^{LMP} and a_v^{SKL} decrease while a_v^{GML} and a_v^{GCV} increase. This is because the rejection regions for the approximate LMP and SKL tests are on the right hand side while the rejection regions of the GML and GCV tests are on the left hand side. Notice that $t_{\text{GML}} \leq 1$ and $t_{\text{GCV}} \leq 1$. For comparison, we use the equivalent test statistics $1 - t_{\text{GML}}$ and $1 - t_{\text{GCV}}$ as the GML and GCV test statistics in this section.

The differences between the approximate LMP, GML, GCV and SKL lie in the differences between weights. The weights depend on the smoothing kernel matrix Σ , the design matrix T and the smoothing parameter λ except for the approximate LMP test. We now compare these weights for a cubic smoothing spline with $n = 100$ and a uniform design in $[0, 1]$. Note that a_v s are not directly comparable because their scales are different and the corresponding statistics have different distributions. In the following, we scale the four null test statistics so that they have the same 95% quantiles. We first generate 40,000 sets of \mathbf{z} under the null hypothesis. Note that all the test statistics are transformation invariant with respect to σ^2 , so it is taken as 1 in the simulation. For each set of \mathbf{z} , the smoothing parameters $\hat{\lambda}_{\text{GML}}$ and $\hat{\lambda}_{\text{GCV}}$ are calculated and $\hat{\lambda}$ in a_v^{SKL} is replaced by $\hat{\lambda}_{\text{GCV}}$. Based on the 40,000 null test statistics of t_{appLMP} , $1 - t_{\text{GML}}$, $1 - t_{\text{GCV}}$ and t_{SKL} , we find that t_{appLMP} , $0.464(1 - t_{\text{GML}})$, $0.199(1 - t_{\text{GCV}})$ and $0.083t_{\text{SKL}}$ have approximately the same 95% quantiles. Therefore we define $w_v^{\text{LMP}} = a_v^{\text{LMP}}$, $w_v^{\text{GML}} = 0.464(1 - a_v^{\text{GML}})$, $w_v^{\text{GCV}} = 0.199(1 - a_v^{\text{GCV}})$ and $w_v^{\text{SKL}} = 0.083a_v^{\text{SKL}}$. All the four tests are equivalent to $\sum_{v=1}^{n-m} w_v z_v^2 / \sum_{v=1}^{n-m} z_v^2$, where w_v denotes one of the w_v^{LMP} , w_v^{GML} , w_v^{SKL} and w_v^{GCV} . Let \bar{w}_v be the average of the 40,000 realizations of w_v . Note that $\bar{w}_v^{\text{LMP}} = w_v^{\text{LMP}}$ since w_v^{LMP} does not depend on the smoothing parameter λ .

In Figure 1 we show the comparison of \bar{w}_v for $v = 1, \dots, 10$. Except for the SKL test, the weights for all the tests decrease very quickly. The SKL test puts almost equal weights on all z_v s. Although not shown in Figure 1, the weights of the SKL test are larger than those of other tests when $v \geq 15$. We observe that on average, the approximate LMP test has the largest weight on z_1 . Thus the LMP is more powerful in the direction of the first column of U , which is the first eigenvector of $Q_2' \Sigma Q_2$. The localness is by no means defined in terms of the distance in W_m or the L_2 distance. It is easy to find two directions, such as $\sin(2\pi t)$ and $\cos(2\pi t)$, such that the powers are very different even when they have the same distances to the null space. The GCV test has the largest weights on z_v , $2 \leq v \leq 10$. Thus it is more sensitive to changes in these directions. The GML is a compromise between the LMP and the GCV tests.

Figure 2 plots the first four columns of U . If we consider the number of modes as the frequency of a function, then the columns of U represent functions with increasing frequencies. Thus we can expect that the approximate LMP test is the most powerful when the true function has frequency 1 and the GCV test is more powerful for higher frequency functions. These observations are confirmed by our simulation results in Section 5. In theory, one can construct new tests in the form $\sum_{v=1}^{n-m} w_v z_v^2 / \sum_{v=1}^{n-m} z_v^2$ with weights chosen to achieve specific

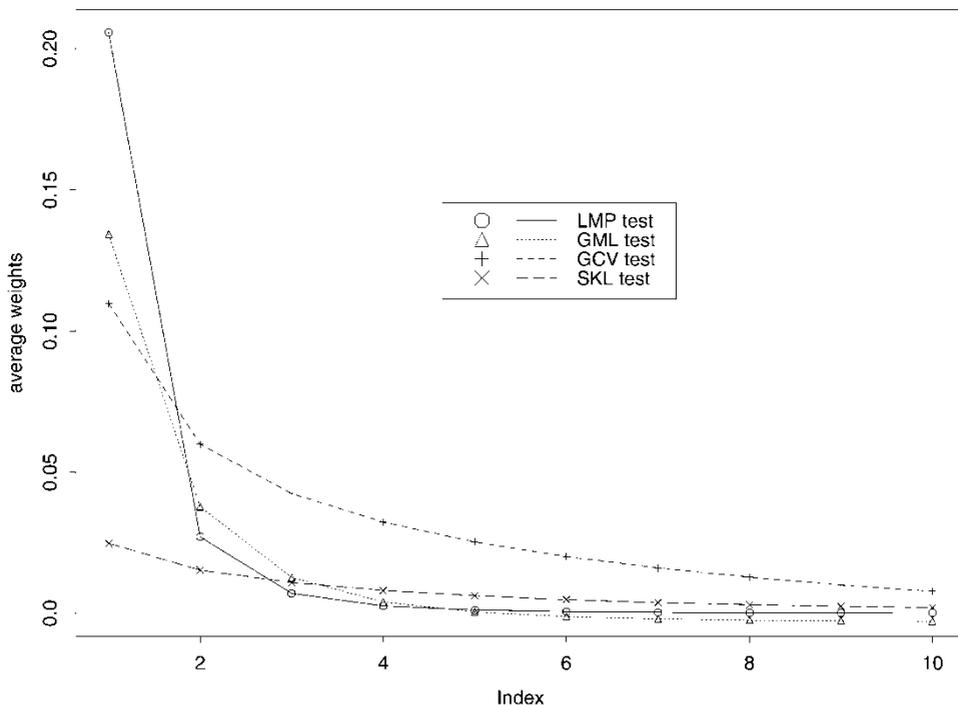


FIGURE 1 Weight comparisons among the approximate LMP, GML, SKL and GCV tests.

purposes. We have also studied weights for other smoothing spline models such as the linear spline and the periodic spline. The results remain the same.

5 SIMULATIONS

Wahba (1990) conducted a small scale simulation to compare the LMP, GML and GCV tests. Since data were generated from the stochastic Bayesian model (5), it was not clear if these results hold when the data are generated from the deterministic model (1). In this section, we conduct simulations to evaluate and compare the relative powers of the LMP, GML, GCV, F_1 , F_2 and SKL tests. Cubic splines ($m = 2$) are used throughout this section.

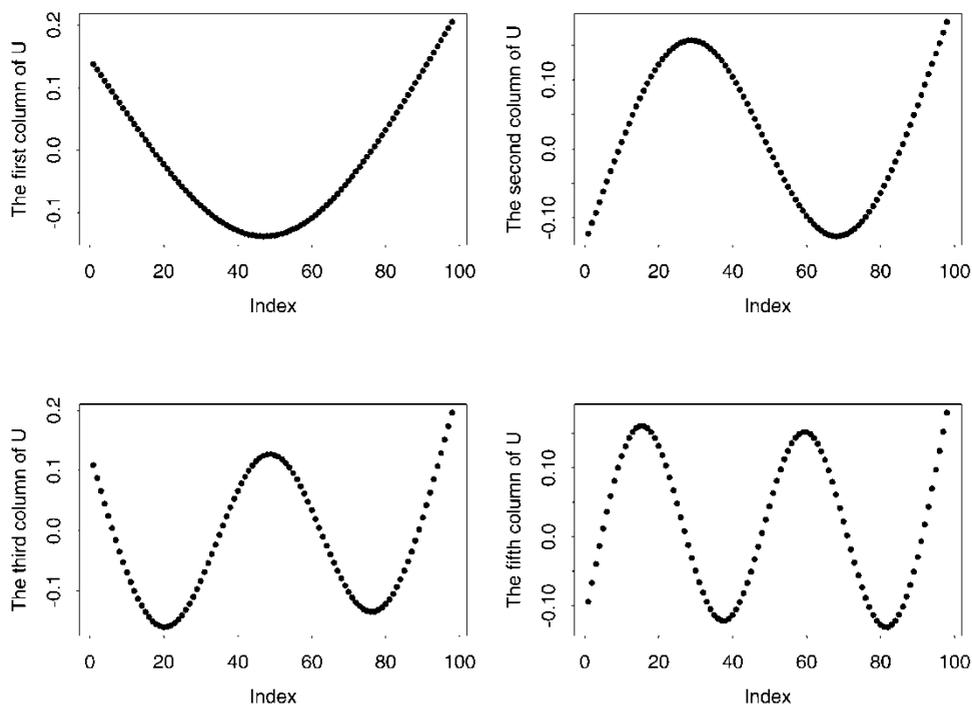
5.1 Power Comparisons

We conduct three simulations to compare powers of these tests. 100 observations were generated from model (1) with the following three f functions:

$$f(t) = 1 + t + at^2, \tag{22}$$

$$f(t) = 1 + t + 3a(t - 0.5)^3, \tag{23}$$

$$f(t) = 1 + t + \sqrt{2}a \cos(6\pi t). \tag{24}$$

FIGURE 2 The first four columns of the U matrix.

The first function is close to the first eigenvector, the second function is close to the second eigenvector and the third function is a high frequency function. The design points are $t_i = (i - 1)/99$, $i = 1, \dots, 100$. We use $\varepsilon_i \stackrel{iid}{\sim} N(0, 0.2^2)$ for the first two models and $\varepsilon_i \stackrel{iid}{\sim} N(0, 1)$ for the last one. For Eq. (22), a takes one of the following five values: 0, 0.2, 0.5, 0.7 or 1. For Eq. (23), a takes one of the following four values: 0, 0.5, 1 or 1.3. For Eq. (24), a takes one of the following four values: 0, 0.3, 0.5 or 1. For all three models, we test hypothesis (7) with $m = 2$.

We repeat each setting 1000 times. The null distributions are generated by the Monte Carlo method described in Section 4 with a simulation sample size of 40,000. The smoothing parameters in the test statistics are estimated for each simulation sample. The proportion of rejections with significance level 0.05 for each test is obtained by counting the percentage of rejection in the 1000 repetitions. Results are shown in Tables I–III. Figure 3 shows the proportion of rejections for combinations of four tests and three models.

Generally speaking, all tests hold their levels properly. As expected from discussions in Section 4, the approximate LMP test is the best under model (22) but the worst under models

TABLE I Proportion of Rejections in 1000 Replications under Model (22).

	$a = 0$	$a = 0.2$	$a = 0.5$	$a = 0.7$	$a = 1$
LMP test	0.055	0.103	0.483	0.759	0.978
GML test	0.051	0.093	0.454	0.734	0.969
GCV test	0.051	0.083	0.325	0.582	0.9
F_1 test	0.052	0.102	0.441	0.702	0.931
F_2 test	0.049	0.1	0.436	0.7	0.934
SKL test	0.049	0.047	0.104	0.245	0.554

TABLE II Proportion of Rejections in 1000 Replications under Model (23).

	$a = 0$	$a = 0.5$	$a = 1.0$	$a = 1.3$
LMP test	0.049	0.070	0.096	0.136
GML test	0.048	0.165	0.538	0.826
GCV test	0.048	0.161	0.553	0.840
F_1 test	0.048	0.118	0.425	0.720
F_2 test	0.044	0.155	0.507	0.806
SKL test	0.041	0.080	0.261	0.525

(23) and (24). This confirms that the approximate LMP test is the most powerful only in the direction of the first eigenvector. The GML test performs well for low frequency functions [models (22) and (23)]. The lack of power under model (24) when $a = 0.3$ and 0.5 is caused by a combination of bad choices of smoothing parameters (the GML method tends to over-smooth in these cases) and small weights of the GML test for higher frequency functions. The GCV method has similar performance as the GML test under model (23) and is the best under model (24), again as expected. The F_1 and F_2 tests perform similarly as the GML test. The SKL test lacks the power to detect lower frequency functions. None of these tests perform consistently well for all simulation settings. The best method to use in practice depends on the shape of the true function. To detect departure in the form of the first eigenvector of U , the approximate LMP method is recommended. To detect low frequency departure, the GML method is recommended and to detect departure of higher frequencies, the GCV method is recommended.

Simulations are also conducted for other functions and spline models. The results obtained are similar.

5.2 Approximation to Null Distribution of the GML Test

Comparing with the Monte Carlo null distribution, we now investigate the accuracy of the mixture of the Chi-square approximation proposed in Section 3.2. We consider two sample sizes, 100 and 200, with design points evenly spaced in $[0, 1]$. Monte Carlo null distributions are generated as described in Section 4 with a simulation size of 40,000.

The results are shown in Figure 4. The estimates of r, \hat{r} are obtained by simulating 5000 of $-(n - m)\log t_{\text{GML}}$ under the null hypothesis and counting the proportion of zeros out of the 5000. Approximations are better for sample size 200.

To further assess the accuracy of the approximation, we repeat the power calculations for the test under model (22) in Section 5.1 using the approximate null distribution. Table IV

TABLE III Proportion of Rejections in 1000 Replications under Model (24).

	$a = 0$	$a = 0.3$	$a = 0.5$	$a = 1.0$
LMP test	0.053	0.054	0.058	0.047
GML test	0.054	0.094	0.229	0.985
GCV test	0.055	0.362	0.872	1.000
F_1 test	0.052	0.076	0.483	1.000
F_2 test	0.053	0.161	0.699	1.000
SKL test	0.045	0.327	0.849	1.000

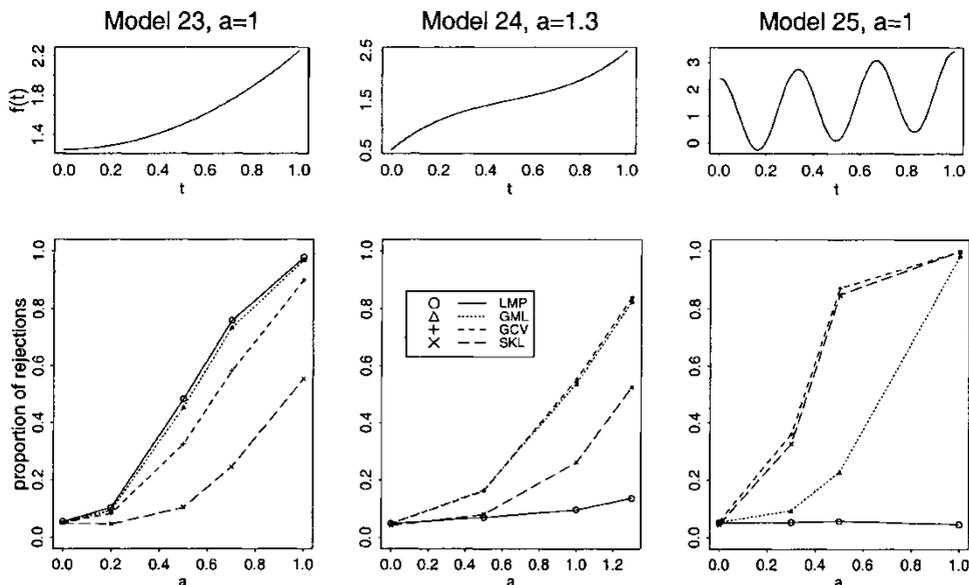


FIGURE 3 The left column of panels corresponds to model (22), the center column to model (23), and the right column to model (24). Top row: $f(t)$ for each model for one of the simulation a values. Bottom row: Proportion of rejections in each of 1000 replications for each of the models versus a . A line is drawn for each of the tests. Crosses indicate results using the simulated null distributions.

shows the results. For the LMP test under the same model, Table IV also lists power computations based on the simulation and Davies’s methods. We see that for the LMP test, the powers based on the simulation method and Davies’s method are very close; for the GML test, the Chi-square mixture approximation works reasonably well for sample size 200. For sample size 100, the approximated level is slightly lower than the nominal level 0.05. Nevertheless, the difference is smaller than twice of the standard deviation $2\sqrt{0.05 \times 0.95/1000} = 0.014$.

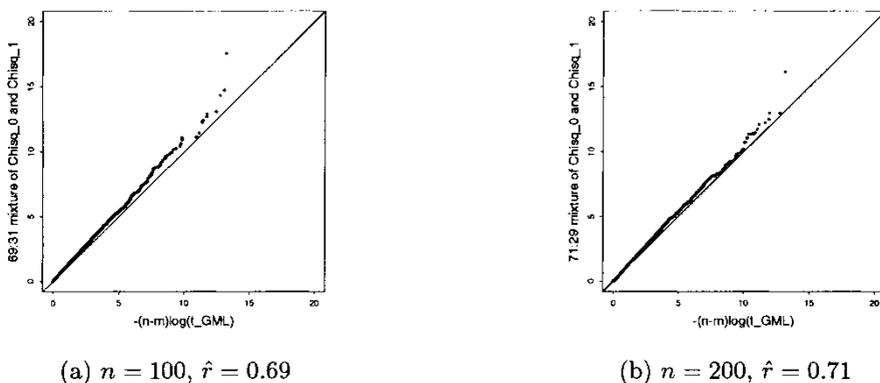


FIGURE 4 QQ plot of quantiles of the $r\chi^2_0 + (1-r)\chi^2_1$ distribution against quantiles of the Monte Carlo null distribution of $-(n-m)\log t_{\text{GML}}$.

TABLE IV Comparison of Powers Obtained via the Simulated Null Distribution (Denoted as ‘Simu’) and Davies’s Method for the LMP Test; the Simulated and the Approximated Distribution (Denoted as ‘Approx’) for the GML Test, Based on Model (22).

<i>Size</i>	<i>Null</i>	$a = 0$	$a = 0.2$	$a = 0.5$	$a = 0.7$	$a = 1$
LMP						
100	Simu	0.055	0.103	0.483	0.759	0.978
	Davies’s	0.050	0.106	0.459	0.756	0.960
200	Simu	0.047	0.392	0.991	1	1
	Davies’s	0.047	0.383	0.987	1	1
GML						
100	Simu	0.051	0.093	0.454	0.734	0.969
	Approx	0.041	0.087	0.395	0.673	0.94
200	Simu	0.044	0.156	0.704	0.946	1.000
	Approx	0.040	0.146	0.697	0.938	1.000

5.3 Robustness of the Tests

The approximate LMP, GML and SKL tests are derived based on the assumption that observations follow *iid* normal distributions. To check the robustness of the tests to the *iid* normal assumption, we generate observations from

$$y_i = 1 + t_i + at_i \exp(-2t_i) + \varepsilon_i, \quad i = 1, \dots, 100, \tag{25}$$

where $t_i = (i - 1)/99$. We consider four choices of a : 0, 0.1, 0.3 and 0.5, and five choices of random errors: $\varepsilon_i \stackrel{iid}{\sim} N(0, 0.2^2)$, $\varepsilon_i \stackrel{iid}{\sim} t_3$ (t distribution with three degrees of freedom; samples of random errors in this case were multiplied by a factor of 0.14 to match their 5% percentile with that of the normal random errors), $\varepsilon_i \stackrel{iid}{\sim}$ 50:50 mixture of $N(-2, 2^2)$ and $N(2, 1)$, $\varepsilon_i \sim$ AR(1) with autoregression coefficient 0.25, and $\varepsilon \sim N(0, W)$ where W is a diagonal matrix with diagonal elements evenly spaced between 0.04 and 0.22. Data are scaled to match variances except for the case with t_3 random errors which has a larger variance (0.059). For each setting, we repeat the simulation 1000 times. Null distributions are calculated by the Monte Carlo method.

Table V shows that all tests hold the levels except when random errors are correlated. The powers are similar when random errors are from *iid* distributions. When the random errors are from the t_3 distribution, the powers are a little bit lower than those from $N(0, 0.2^2)$, partly due to the larger variance. The mixture distribution is asymmetric with a slightly heavier tail on the left. It can be seen that the powers for the mixture distribution are similar to those of $N(0, 0.2^2)$. Results with other asymmetric distributions such as Gamma remain the same. When random errors are from $N(0, W)$, the powers are significantly lower than those in the *iid* cases. Comparative results of the different tests remain the same. We conclude that the tests are relatively robust to the violation of the normality assumption, but not to the independence assumption. For independent data with heterogeneous variances, the tests can approximately maintain their levels but lack power. We also conducted simulations under settings as in Section 5.1. The results remain the same.

5.4 Summary of Simulation Results

In summary, the LMP, GML and GCV tests outperform other tests. Which one to use in practice depends on the shape of the true function: the LMP and GML are more powerful for detecting departure in forms of low frequency functions while the GCV test is more powerful

TABLE V Proportion of Rejections in 1000 Replications.

	<i>LMP</i>	<i>GML</i>	<i>GCV</i>	F_1	F_2	<i>SKL</i>
<i>a</i> = 0						
$N(0, 0.2^2)$	0.049	0.047	0.048	0.047	0.046	0.052
t_3	0.043	0.043	0.039	0.034	0.038	0.033
Mixture	0.047	0.049	0.048	0.047	0.050	0.042
AR(1)	0.128	0.176	0.540	0.089	0.147	0.574
$N(0, W)$	0.055	0.055	0.051	0.056	0.056	0.037
<i>a</i> = 0.1						
$N(0, 0.2^2)$	0.063	0.064	0.061	0.075	0.071	0.045
t_3	0.060	0.058	0.050	0.061	0.061	0.041
Mixture	0.073	0.083	0.085	0.082	0.087	0.065
AR(1)	0.110	0.173	0.498	0.090	0.149	0.545
$N(0, W)$	0.041	0.047	0.062	0.062	0.046	0.063
<i>a</i> = 0.3						
$N(0, 0.2^2)$	0.134	0.128	0.102	0.120	0.115	0.049
t_3	0.112	0.114	0.098	0.101	0.108	0.056
Mixture	0.138	0.135	0.108	0.132	0.131	0.066
AR(1)	0.168	0.219	0.550	0.092	0.175	0.595
$N(0, W)$	0.069	0.067	0.064	0.062	0.063	0.054
<i>a</i> = 0.5						
$N(0, 0.2^2)$	0.417	0.398	0.289	0.403	0.393	0.095
t_3	0.317	0.310	0.232	0.306	0.299	0.082
Mixture	0.414	0.403	0.300	0.385	0.378	0.107
AR(1)	0.212	0.262	0.553	0.130	0.185	0.587
$N(0, W)$	0.104	0.103	0.090	0.110	0.106	0.069

Note: Random errors are generated from independent $N(0, 0.2^2)$, t_3 distribution, mixture of normal distributions, AR(1) process and $N(0, W)$.

for high frequency functions. Overall, the GML test is recommended. These tests are applicable even when the normality assumption is violated. However, they should not be used when observations are correlated.

Null distributions of all the tests can be obtained through Monte Carlo simulations. Since the LMP test does not involve smoothing parameters, its null distribution simulation is easy and fast. Davies's (1980) method works equally well in this case. The null distribution simulation for other tests require the calculation of smoothing parameters for all simulation samples, which is computationally intensive. For the GML test, a null distribution approximation method using a mixture of Chi-square distributions is proposed and evaluated. It is relatively fast and provides satisfactory approximations when the sample size is large.

6 DISCUSSION

The connection between smoothing spline models and the Bayesian models (the mixed effects models) transfers the hypothesis on parametric regression to a much simpler hypothesis on a variance component. The approximate LMP, GML and GCV tests derived from the Bayesian model (or the mixed effect model) work well under the deterministic models. The good properties of the tests make them desirable for more complicated models. The hypothesis (7) can be written more generally as

$$H_0: f \in \mathcal{M}_0, \quad H_1: f \in \mathcal{M}_1 \quad \text{and} \quad f \notin \mathcal{M}_0,$$

where \mathcal{M}_0 is the model space under the null hypothesis, and \mathcal{M}_1 is a bigger model space which contains a substantially large family of plausible functions. \mathcal{M}_0 could be a linear or nonlinear parametric model, or a simple nonparametric model. For example, to test a nonlinear regression model \mathcal{M}_0 , one can use nonlinear partial splines (Wahba, 1990) or nonlinear nonparametric regression models (Ke and Wang, 2002) as \mathcal{M}_1 . To test an additive model \mathcal{M}_0 (Hastie and Tibshirani, 1990), one can use smoothing spline analysis of variance (SS ANOVA) models (Wahba, 1990) as \mathcal{M}_1 and the test interaction components equal zero. This approach can also be employed to test the functional form of fixed effects in a mixed effects model. For example, to test LMM, one may use the semiparametric mixed effects models in Wang (1998b) as \mathcal{M}_1 . To test nonlinear mixed effects models, one may use the semi-parametric nonlinear mixed effects models in Ke and Wang (2001) as \mathcal{M}_1 . Another direction is to extend the current test methods for nonGaussian data, which will allow us to test the generalized linear models (McCullagh and Nelder, 1989; Liu *et al.*, 2003), the generalized additive models (Hastie and Tibshirani, 1990) and the generalized LMM (Breslow and Clayton, 1993). All current methods are sensitive to the independence assumption. Thus new methods need to be developed for correlated data. Some research has been done in these directions. Guo (2002) generalized the GML test to the mixed effect SS ANOVA models. Zhang and Lin (2003) generalized the score test, which is equivalent to the approximate LMP test, to the semiparametric additive mixed models with nonGaussian data. The SKL test was initially developed for smoothing spline models with nonGaussian data by Xiang and Wahba (1995). These generalizations all showed good performances. We are currently working on extensions of the approximate LMP, GML and GCV tests for SS ANOVA model with nonGaussian data. Preliminary results are encouraging. Tests for more complicated models as described above will be pursued in the future.

Acknowledgements

This work was supported by NIH grants R01 GM58533. The authors thank the associate editor and a referee for their useful comments that have greatly improved this article.

References

- Azzalini, A. and Bowman, A. W. (1993). On the use of nonparametric regression for checking linear relationships. *J. Roy. Stat. Soc. B*, **55**, 549–557.
- Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Am. Stat. Assoc.*, **88**, 9–25.
- Cantoni, E. and Hastie, T. (2002). Degrees of freedom tests for smoothing splines. *Biometrika*, **89**, 251–263.
- Cleveland, W. S. and Devlin, S. J. (1988). Locally-weighted regression: an approach to regression analysis by local fitting. *J. Am. Stat. Assoc.*, **83**, 597–610.
- Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.
- Cox, D., Koh, E., Wahba, G. and Yandell, B. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Stat.*, **16**, 113–119.
- Crainiceanu, C. M. and Ruppert, D. (2003). Asymptotic distribution of likelihood ratio tests in linear mixed models. *J. Roy. Stat. Soc. B* (to appear).
- Crainiceanu, C. M., Ruppert, D. and Vogelsang, T. J. (2003). Some properties of likelihood ratio tests in linear mixed models (submitted).
- Davies, R. B. (1980). The distribution of a linear combination of x^2 random variables. *Appl. Stat.*, **29**, 323–333.
- Guo, W. (2002). Inference in smoothing spline analysis of variance. *J. Roy. Stat. Soc. B*, **64**, 887–898.
- Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London.
- Ke, C. and Wang, Y. (2001). Semi-parametric nonlinear mixed effects models and their applications. *J. Am. Stat. Assoc.*, **96**, 1272–1298 (with discussion).
- Ke, C. and Wang, Y. (2002). Nonparametric nonlinear regression models (Unpublished).
- Kimeldorf, G. and Wahba, G. (1971). Some results on teheycheffian spline functions. *J. Math. Anal. Appl.*, **33**, 82–94.

- Liu, A., Meiring, W. and Wang, Y. (2003). Testing generalized linear models using smoothing spline methods. *Technical Report 388*, Department of Statistics and Applied Probability, University of California, Santa Barbara.
- McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. Chapman and Hall, London.
- Raz, J. (1990). Testing for no effect when estimating a smooth function by nonparametric regression: A randomization approach. *J. Am. Stat. Assoc.*, **85**, 132–138.
- Self, S. and Liang, K. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Stat. Assoc.*, **82**, 605–610.
- Wahba, G. (1990). Spline models for observational data. *CBMS-NSF Regional Conference Series in Applied Mathematics*, SIAM.
- Wang, Y. (1998a). Smoothing spline models with correlated random errors. *J. Am. Stat. Assoc.*, **93**, 341–348.
- Wang, Y. (1998b). Mixed-effects smoothing spline ANOVA. *J. Roy. Stat. Soc. B*, **60**, 159–174.
- Xiang, D. and Wahba, G. (1995). Testing the generalized linear model null hypothesis versus ‘smooth’ alternatives. *Technical Report*. Dept. Statistics, Univ., Wisconsin-Madison.
- Zhang, D. and Lin, X. (2003). Hypothesis testing in semiparametric additive mixed models. *Biostatistics*, **4**, 57–74.

Copyright of Journal of Statistical Computation & Simulation is the property of Taylor & Francis Ltd and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.