

## A NOTE ON A NONPARAMETRIC REGRESSION TEST THROUGH PENALIZED SPLINES

Huaihou Chen<sup>1</sup>, Yuanjia Wang<sup>2</sup>, Runze Li<sup>3</sup> and Katherine Shear<sup>2</sup>

<sup>1</sup>New York University, <sup>2</sup>Columbia University and <sup>3</sup>Pennsylvania State University

*Abstract:* We examine a test of a nonparametric regression function based on penalized spline smoothing. We show that, similarly to a penalized spline estimator, the asymptotic power of the penalized spline test falls into a small- $K$  or a large- $K$  scenarios characterized by the number of knots  $K$  and the smoothing parameter. However, the optimal rate of  $K$  and the smoothing parameter maximizing power for testing is different from the optimal rate minimizing the mean squared error for estimation. Our investigation reveals that compared to estimation, some under-smoothing may be desirable for the testing problems. Furthermore, we compare the proposed test with the likelihood ratio test (LRT). We show that when the true function is more complicated, containing multiple modes, the test proposed here may have greater power than LRT. Finally, we investigate the properties of the test through simulations and apply it to two data examples.

*Key words and phrases:* Goodness of fit, likelihood ratio test, nonparametric regression, partial linear model, spectral decomposition.

### 1. Introduction

Penalized splines have become a popular nonparametric smoothing technique (Eilers and Marx (1996); Ruppert, Wand, and Carroll (2009)). In contrast, testing nonparametric functions through penalized splines is less explored, especially the cases that do not rely on a linear mixed effects model (LME). In this work, we consider testing a nonparametric function relating a covariate  $u_i \in [a, b]$  to an outcome  $y_i$ ,

$$y_i = f(u_i) + \varepsilon_i, \quad E(\varepsilon_i) = 0, \quad \text{var}(\varepsilon_i) = \sigma^2, \quad i = 1, \dots, n, \quad (1.1)$$

where  $f(\cdot)$  is an unspecified smooth function (extension to a partial linear model is discussed in Section 2.4). A first problem is to test for the significance of the regression function,

$$H_0 : f(u) = 0, \quad \text{for all } u \in [a, b]. \quad (1.2)$$

A second problem is to test the nonparametric deviation of  $f(\cdot)$  from a polynomial model, or goodness-of-fit of a polynomial model, where the null hypothesis is

$$H_0 : f(\cdot) \in \mathcal{M}^p[a, b] = \{\theta_0 + \theta_1 u + \dots + \theta_p u^p : (\theta_0, \theta_1, \dots, \theta_p) \in \mathcal{R}^{p+1}, u \in [a, b]\}. \quad (1.3)$$

To accommodate a flexible class of functions, a number of works have constructed test statistics through the smoothing spline estimator of  $f(\cdot)$ . These include Cox et al. (1988), Cox and Koh (1989), Eubank and Spiegelman (1990), Raz (1990), Chen (1994), Jayasuriya (1996), Ramil-Novo and GonzKalez-Manteiga (2000), Cantoni and Hastie (2002), and Liu and Wang (2004). Cantoni and Hastie (2002) considered a test statistic based on a mixed-effects model with a fixed smoothing parameter. Liu and Wang (2004) compared such smoothing spline-based tests as the locally most powerful test in Cox et al. (1988), the generalized maximum likelihood ratio test, and the generalized cross validation test (GCV test, Wahba (1990)). Another line of work on testing the mean function in a nonparametric regression has used local polynomial smoothing under the alternative. For example, Cai, Fan, and Li (2000) proposed a likelihood ratio test for the coefficient functions in varying-coefficient models. Fan, Zhang, and Zhang (2001) introduced generalized likelihood ratio statistics for testing nonparametric functions. Li and Nie (2008) proposed various generalized likelihood ratio tests and generalized  $F$  tests. Zhang (2004) assessed the equivalence of nonparametric tests based on smoothing splines and local polynomials, and reported their equivalent asymptotic distributions under the null and the equivalent rate of smoothing parameters under the alternative.

The hypothesis (1.2) can also be examined by a likelihood ratio test (LRT) through the use of penalized splines and a linear mixed effects model representation (Wand (2003)). Specifically, under the alternative, one uses a mixed effects model to represent  $f(\cdot)$  and tests for several fixed effects and a variance component in an LME. Crainiceanu and Ruppert (2004) and Crainiceanu et al. (2005) reported that the asymptotic distribution of the LRT or restricted likelihood ratio test (RLRT) involving a variance component in an LME does not have the typical chi-square mixture distribution. These tests are based on the likelihood assuming normality of the random effects and the residual errors. The smoothing parameter is taken as the ratio of two variance components and estimated through a restricted maximum likelihood (REML). There is no literature on the optimal rate of the smoothing parameter or the optimal number of knots  $K$  to maximize power in a testing setting.

We present a test of a nonparametric function and a test of a higher order nonparametric deviation from a polynomial model based on penalized splines. Our proposed test differs from others that have been advanced. Unlike the test in Cantoni and Hastie (2002), we do not assume a fixed smoothing parameter under the alternative hypothesis, since a reasonable smoothing parameter may not be available in practice. The proposed test is different from the tests in Crainiceanu and Ruppert (2004) and Crainiceanu et al. (2005) in that it does not rely on mixed-effects model representation, and thus relaxes the normality assumption.

Most of the test statistics in the literature are based on either smoothing spline or local polynomial smoothing, while our proposed test is based on penalized splines.

We examine the asymptotic properties of the proposed test under the null and the alternative. We show that the asymptotic distribution of the penalized spline test falls into two categories characterized by the number of knots  $K$  and the smoothing parameter: a small- $K$  scenario and a large- $K$  scenario. Unlike penalized spline estimation, the optimal rate for a testing problem to maximize power is different from an estimation problem to minimize the average mean squared error. Our investigation reveals that, compared to estimation some under-smoothing may be desirable for testing problems. We compare the proposed test with LRT and RLRT and provide heuristics on why the latter may have better power to detect simpler functions and worse power for more complicated functions. We investigate numerical properties of the proposed test through simulations and apply it to two studies: the Framingham Heart Study data (Cupples et al. (2003)) which examines the association between cholesterol level and BMI; the Complicated Grief Study (Shear et al. (2005)) to examine the association between a subject's work and social functioning impairment, and the severity of complicated grief disorder.

## 2. Test Statistic and Its Asymptotic Distribution

### 2.1. Testing an unspecified function

Denote by  $\mathbf{N}(u)$  a vector of  $p$ th order B-spline basis functions with  $K$  knots,  $\tau_1, \dots, \tau_K$ , and by  $\mathbf{N} = (\mathbf{N}(u_1), \dots, \mathbf{N}(u_n))^T$  the matrix of basis functions. The penalized spline estimator of  $f(u)$  is  $\hat{f}_n(u) = \mathbf{N}^T(u)\hat{\boldsymbol{\beta}}$ , where  $\hat{\boldsymbol{\beta}}$  minimizes

$$(\mathbf{Y} - \mathbf{N}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{N}\boldsymbol{\beta}) + \lambda\boldsymbol{\beta}^T\mathbf{D}_q\boldsymbol{\beta}, \quad (2.1)$$

$\lambda$  a smoothing parameter, and  $\mathbf{D}_q = \int_a^b \mathbf{N}^{(q)}(x)^T \mathbf{N}^{(q)}(x) dx$  a  $q$ th order derivative-based penalty matrix (Wand and Ormerod (2008); Claeskens, Kivobokova, and Opsomer (2009)). As  $\hat{\boldsymbol{\beta}} = (\mathbf{N}^T\mathbf{N} + \lambda\mathbf{D}_q)^{-1}\mathbf{N}^T\mathbf{Y}$ ,

$$\hat{f}_n(u) = \mathbf{N}^T(u)(\mathbf{N}^T\mathbf{N} + \lambda\mathbf{D}_q)^{-1}\mathbf{N}^T\mathbf{Y}.$$

Let  $\widehat{\mathbf{f}}_n = (\hat{f}_n(u_1), \dots, \hat{f}_n(u_n))^T$  and  $\mathbf{Y} = (y_1, \dots, y_n)^T$ . To test the null hypothesis (1.2) in model (1.1), we propose a simple test statistic based on the sum of squared distances of the fitted values,

$$T_n = \widehat{\mathbf{f}}_n^T \widehat{\mathbf{f}}_n = \mathbf{Y}^T \mathbf{N}(\mathbf{N}^T \mathbf{N} + \lambda \mathbf{D}_q)^{-1} \mathbf{N}^T \mathbf{N} (\mathbf{N}^T \mathbf{N} + \lambda \mathbf{D}_q)^{-1} \mathbf{N}^T \mathbf{Y}. \quad (2.2)$$

A similar test based on the smoothing spline estimator was proposed in Eubank and Spiegelman (1990), Chen (1994), and Jayasuriya (1996).

It is useful to introduce a singular value decomposition used in Claeskens, Kivobokova, and Opsomer (2009),

$$(\mathbf{N}^T \mathbf{N})^{-1/2} \mathbf{D}_q (\mathbf{N}^T \mathbf{N})^{-1/2} = \mathbf{U} \mathbf{S} \mathbf{U}^T,$$

where  $\mathbf{U}$  is the matrix of eigenvectors, and  $\mathbf{S} = \text{diag}(s_1, \dots, s_{K+p+1})$  is the diagonal matrix of the eigenvalues. Let  $\mathbf{A} = \mathbf{N}(\mathbf{N}^T \mathbf{N})^{-1/2} \mathbf{U}$ , so  $\mathbf{A}^T \mathbf{A} = \mathbf{I}_{K+p+1}$  and  $\mathbf{A} \mathbf{A}^T = \mathbf{N}(\mathbf{N}^T \mathbf{N})^{-1} \mathbf{N}^T$ . It is easy to show that

$$\widehat{\mathbf{f}}_n = \mathbf{N}(\mathbf{N}^T \mathbf{N} + \lambda \mathbf{D}_q)^{-1} \mathbf{N}^T \mathbf{Y} = \mathbf{A}(\mathbf{I}_n + \lambda \mathbf{S})^{-1} \mathbf{A}^T \mathbf{Y}.$$

If we take  $\mathbf{H}_n = \mathbf{A}(\mathbf{I}_{K+p+1} + \lambda \mathbf{S})^{-1} \mathbf{A}^T$ , the test statistic is

$$T_n = \mathbf{Y}^T \mathbf{H}_n^2 \mathbf{Y}.$$

Under the null hypothesis (1.2) and the assumption  $\epsilon_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$ , we have

$$T_n \stackrel{d}{=} \sigma^2 \sum_{i=1}^{K+p+1} \frac{\omega_i^2}{(1 + \lambda s_i)^2},$$

where  $\omega_i$  are i.i.d.  $N(0, 1)$ . Under the alternative hypothesis and the assumption  $\epsilon_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$ ,

$$T_n = (\mathbf{A}^T \mathbf{Y})^T (\mathbf{I}_{K+p+1} + \lambda \mathbf{S})^{-2} \mathbf{A}^T \mathbf{Y},$$

where  $\mathbf{A}^T \mathbf{Y} \sim N(\mathbf{A}^T \mathbf{f}_n, \sigma^2 \mathbf{I}_{K+p+1})$ , and  $\mathbf{f}_n = E\mathbf{Y} = (f(u_1), \dots, f(u_n))^T$ . Then  $T_n$  is noncentral mixture  $\chi^2$

$$T_n \stackrel{d}{=} \sigma^2 \sum_{i=1}^{K+p+1} \frac{(\omega_i + \delta_i)^2}{(1 + \lambda s_i)^2},$$

where  $\delta_i$  is the  $i$ th component of  $\mathbf{A}^T \mathbf{f}_n$ .

## 2.2. Asymptotic null distribution

We look to the asymptotic null distribution of  $T_n$ , first considering the  $p$ th order B-spline basis with  $K$  knots. Similar results with a truncated polynomial basis can be obtained by a suitable transformation.

**Theorem 1.** *If assumptions A1–A3 in the Appendix hold,  $\epsilon_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$ ,  $K \rightarrow \infty$ , and  $\lambda/n \rightarrow 0$ , then the null distribution of  $T_n$  as  $n \rightarrow \infty$  is*

$$\frac{T_n - \sigma^2 \text{trace}(\mathbf{H}_n^2)}{\sigma^2 \{2 \text{trace}(\mathbf{H}_n^4)\}^{1/2}} \rightarrow N(0, 1).$$

Chen (1994) and Jayasuriya (1996) proved a similar theorem using a smoothing spline based estimator. In practice,  $\sigma^2$  is unknown and is estimated from data. Substituting a suitable consistent estimator  $\hat{\sigma}_n^2$  for  $\sigma^2$  has ignorable impact on the asymptotic null distribution of  $T_n$  (Eubank and LaRiccia (1993); Jayasuriya (1996)).

The normality assumption on the  $\varepsilon_i$ 's is related in the following theorem.

**Theorem 2.** *Suppose assumptions A1–A3 in the Appendix hold. If the  $\varepsilon_i$  are i.i.d. with  $E(\varepsilon_1) = 0$ ,  $\text{var}(\varepsilon_1) = \sigma^2$ ,  $0 < E(\varepsilon_1^4) < \infty$ ,  $K \rightarrow \infty$ ,  $K^3 = o(n)$ , and  $\lambda/n \rightarrow 0$ , then the null distribution of  $T_n$  as  $n \rightarrow \infty$  is*

$$\frac{T_n - \sigma^2 \text{trace}(\mathbf{H}_n^2)}{\sigma^2 \{2 \text{trace}(\mathbf{H}_n^4)\}^{1/2}} \rightarrow N(0, 1).$$

### 2.3. Power considerations and two asymptotic scenarios

To study the asymptotic distribution of  $T_n$  under the alternative, let  $K_q = (\lambda/n)^{1/2q}K$ , where  $q$  is the order of the derivative-based penalty matrix  $D_q$ , and let  $C^{p+1}[a, b]$  be the set of all  $p+1$  times continuously differentiable functions on  $[a, b]$ .

**Theorem 3.** *The assumptions A1–A3 in the Appendix hold,  $\varepsilon_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$ , and  $0 \leq c < E[f^2(u_1)] = \|f\|_u^2$ , then the following hold.*

(i) *If  $K_q = o(1)$  and  $f(\cdot) \in C^{p+1}[a, b]$ , then*

$$\frac{T_n - \sigma^2 \text{trace}(\mathbf{H}_n^2)}{\sigma^2 \{2 \text{trace}(\mathbf{H}_n^4)\}^{1/2}} \rightarrow^d N\left(c_1 \|f\|_u^2 n K^{-1/2}, 1\right),$$

where  $c_1$  is a constant, and  $T_n$  can detect alternatives of order  $\{nK^{-1/2}\}^{-1/2}$  or slower from the null model. As  $n \rightarrow \infty$ ,

$$P\left[\frac{T_n - \sigma^2 \text{trace}(\mathbf{H}_n^2)}{\sigma^2 \{2 \text{trace}(\mathbf{H}_n^4)\}^{1/2}} \geq z_\alpha\right] \rightarrow 1, \quad (2.3)$$

where  $z_\alpha$  is the 100(1 -  $\alpha$ )th percentile of the standard normal.

(ii) *If  $K_q = O(1)$  and  $f(\cdot) \in C^{p+1}[a, b]$ , then*

$$\frac{T_n - \sigma^2 \text{trace}(\mathbf{H}_n^2)}{\sigma^2 \{2 \text{trace}(\mathbf{H}_n^4)\}^{1/2}} \rightarrow^d N\left(c_2 \|f\|_u^2 n \left(\frac{\lambda}{n}\right)^{1/4q}, 1\right),$$

where  $c_2$  is a constant, and  $T_n$  can detect alternatives of order  $\{n(\lambda/n)^{1/4q}\}^{-1/2}$  or slower from the null model. The power of  $T_n$  is asymptotically one as  $n \rightarrow \infty$ .

**Remark 1.** For an optimal testing procedure, a local alternative can converge to the null at the fastest rate at which the test still maintains consistency. For  $K_q = o(1)$ , the optimal rates of the number of knots and the smoothing parameter for testing are  $K = O(n^{2/(4p+5)})$  and  $\lambda = O(n^\nu)$ , where  $\nu \leq (2p-2q+3)/(4p+5)$ . For  $K_q = O(1)$ , the optimal rates are  $\lambda = O(n^{1/(4q+1)})$  and  $K = O(n^\nu)$  with  $\nu \geq 2q/(4q+1)(p+1)$ .

**Remark 2.** Case (i) in Theorem 2 corresponds to the small- $K$  scenario: the optimal rates are determined by the number of knots as long as the smoothing parameter is sufficiently small. Case (ii) in Theorem 2 corresponds to the large- $K$  scenario: the optimal rates are determined by the smoothing parameter and the order of the penalty as long as the number of knots is sufficiently large.

**Remark 3.** The optimal rates of  $\lambda$  and  $K$  obtained here for testing are different from the optimal rates for estimation in Claeskens, Kivobokova, and Opsomer (2009). Under the small- $K$  scenario, the optimal rate is  $O(n^{-(2p+2)/(2p+3)})$  for estimation and is  $O(n^{-(2p+2)/(4p+5)})$  for testing; under the large- $K$  scenario, the optimal rate is  $O(n^{-2q/(2q+1)})$  for estimation and is  $O(n^{-2q/(4q+1)})$  for testing.

**Remark 4.** For consistency of the large- $K$  scenario with similar results using smoothing splines, note that the smoothing parameter  $\lambda^*$  in Zhang (2004) and the smoothing parameter here have the relationship  $\lambda^* = \lambda/n$ . Under technical conditions, the detectable rate of a local alternative obtained in Zhang (2004) is  $\{n\lambda^{1/4q}\}^{-1/2}$  for testing based on smoothing splines; the rate in our case (ii) is the same as in Theorem 2 of Zhang (2004).

**Remark 5.** In conjunction with Theorem 1, it is possible to relax the normality condition in Theorem 3 with additional assumptions. Specifically, the condition  $\varepsilon_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$  is replaced by  $\text{var}(\varepsilon_i) = \sigma^2$  and  $0 < E(\varepsilon_i^4) < \infty$ , and we require  $K^3 = o(n)$ .

Minimizing mean squared error and maximizing power do not necessarily lead to the same optimal rates for the number of knots and the smoothing parameter. Under the small- $K$  scenario, the optimal rate for testing is  $K = O(n^{2/(4p+5)})$  when  $\lambda/n$  converges to zero sufficiently fast, which is faster than the optimal rate for estimation,  $K = O(n^{1/(2p+3)})$  (Claeskens, Kivobokova, and Opsomer (2009)). This suggests that using a larger number of knots for testing as compared to estimation may be desirable. Under the large- $K$  scenario, the optimal rate for testing is  $\lambda = O(n^{1/(4q+1)})$  for a sufficiently large number of knots, which is slower than the optimal rate for estimation,  $\lambda = O(n^{1/(2q+1)})$ . This suggests that using a smaller smoothing parameter for testing might be desirable.

## 2.4. Extension to a partial linear model

When there are other covariates  $\mathbf{x}_i$  predicting the outcome, we consider testing the association with the covariate of interest,  $u_i$ , through a partial linear model. Thus, we test (1.2) in the model

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + f(u_i) + \varepsilon_i, \quad E(\varepsilon_i) = 0, \quad \text{var}(\varepsilon_i) = \sigma^2, \quad i = 1, \dots, n, \quad (2.4)$$

where  $f(\cdot)$  is an unspecified smooth function. When  $\mathbf{x}_i = (1, u_i, \dots, u_i^p)^T$ , testing (1.2) in this model is equivalent to testing goodness-of-fit of a  $p$ th order polynomial model.

To construct a test statistic for a partial linear model, we use an orthogonal contrast that transforms the model into one without covariates. Let  $\mathbf{X}$  denote the stacked matrix of  $\mathbf{x}_i$  and let  $\mathbf{Q}$  be an orthogonal contrast such that

$$\mathbf{Q}^T \mathbf{X} = \mathbf{0}, \quad \mathbf{Q}^T \mathbf{Q} = \mathbf{I}_{n-p}, \quad \text{and} \quad \mathbf{Q} \mathbf{Q}^T = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T.$$

One way to construct such a  $\mathbf{Q}$  is in the Appendix of Wang and Chen (2012). Applying the transformation  $\mathbf{Q}$  to (2.4), we arrive at

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{f}} + \tilde{\boldsymbol{\varepsilon}}, \quad \text{var}(\tilde{\boldsymbol{\varepsilon}}) = \sigma^2 \mathbf{I}_{n-p},$$

where  $\tilde{\mathbf{Y}} = \mathbf{Q}^T \mathbf{Y}$ ,  $\tilde{\mathbf{f}} = \mathbf{Q}^T \mathbf{f}$ , and  $\tilde{\boldsymbol{\varepsilon}} = \mathbf{Q}^T \boldsymbol{\varepsilon}$ . A test statistic similar to (2.2) is obtained as

$$Z_n = \tilde{\mathbf{Y}}^T \tilde{\mathbf{N}} (\tilde{\mathbf{N}}^T \tilde{\mathbf{N}} + \lambda \tilde{\mathbf{D}}_q)^{-1} \tilde{\mathbf{N}}^T \tilde{\mathbf{N}} (\tilde{\mathbf{N}}^T \tilde{\mathbf{N}} + \lambda \tilde{\mathbf{D}}_q)^{-1} \tilde{\mathbf{N}}^T \tilde{\mathbf{Y}}, \quad (2.5)$$

where  $\tilde{\mathbf{N}} = \mathbf{Q}^T \mathbf{N}$ , and  $\tilde{\mathbf{D}}_q = \mathbf{Q}^T \mathbf{D}_q \mathbf{Q}$ . Since testing the goodness-of-fit of a polynomial model is a special case of testing  $H_0 : f(u) = 0$  in a partial linear model,  $Z_n$  can be used to examine (1.3).

To derive the null and alternative distributions of the test statistic with a truncated polynomial basis, note that  $\mathbf{D}_q = \text{diag}(\mathbf{0}_{p+1}, \mathbf{I}_K)$  and

$$Z_n = \mathbf{Y}^T \mathbf{P}_X \mathbf{N} (\mathbf{N}^T \mathbf{P}_X \mathbf{N} + \lambda \mathbf{D}_q)^{-1} \mathbf{N}^T \mathbf{P}_X \mathbf{N} (\mathbf{N}^T \mathbf{P}_X \mathbf{N} + \lambda \mathbf{D}_q)^{-1} \mathbf{N}^T \mathbf{P}_X \mathbf{Y},$$

where  $\mathbf{P}_X = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ , and  $\mathbf{Q}^T \mathbf{Y}$  is  $N(\mathbf{0}, \sigma^2 \mathbf{I}_{n-p})$  under  $H_0$  and  $N(\mathbf{Q}^T \mathbf{f}_n, \sigma^2 \mathbf{I}_{n-p})$  under  $H_a$ . Thus, under  $H_0$ ,

$$Z_n = {}^d \sigma^2 \sum_{i=p+2}^{n-p} \frac{\mu_i^2 \omega_i^2}{(\lambda + \mu_i)^2} + \sigma^2 \sum_{i=1}^{p+1} \omega_i^2,$$

where  $\mu_i$  is the  $i$ th eigenvalue of  $\mathbf{N}^T \mathbf{P}_X \mathbf{N}$ . Under the alternative,

$$Z_n = {}^d \sigma^2 \sum_{i=p+2}^{n-p} \frac{\mu_i^2 (\omega_i + \delta'_i)^2}{(\lambda + \mu_i)^2} + \sigma^2 \sum_{i=1}^{p+1} (\omega_i + \delta'_i)^2, \quad (2.6)$$

where  $\delta'_i$  is the  $i$ th component of  $\mathbf{Q}^T \mathbf{f}_n$ , and  $\mathbf{f}_n = \{f(u_1), \dots, f(u_n)\}^T$ .

## 2.5. Connection with the RLRT

As well, LRT or RLRT based on an LME can be used to test (1.2) or (1.3). Under the alternative, represent  $f(u)$  using a truncated polynomial basis by an LME,

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b} + \varepsilon_i, \quad \mathbf{b} \sim N(\mathbf{0}, \sigma_b^2 \mathbf{I}_K), \quad \varepsilon_i \sim N(0, \sigma^2),$$

where  $\mathbf{x}_i = (1, u_i, \dots, u_i^p)^T$ ,  $\mathbf{z}_i = ((u_i - \tau_1)_+^p, \dots, (u_i - \tau_K)_+^p)^T$ , and  $\tau_1, \dots, \tau_K$  is a sequence of knots. Under this model, hypothesis (1.2) can be tested as

$$H_0 : \boldsymbol{\beta} = \mathbf{0}, \sigma_b^2 = 0$$

through an LRT; the hypothesis (1.3) can be tested as

$$H_0 : \sigma_b^2 = 0$$

through an RLRT. The smoothing parameter  $\lambda$  in (2.1) corresponds to  $\sigma_\varepsilon^2 / \sigma_b^2$ .

Theorem 1 in Crainiceanu et al. (2005) has that at local alternatives, in distribution,

$$\text{RLRT} = \sup_{d \geq 0} \left\{ \sum_{s=1}^K \frac{d\mu_s(1 + d_0\mu_s)}{1 + d\mu_s} w_s^2 - \sum_{s=1}^K \log(1 + d\mu_s) \right\}, \quad (2.7)$$

where  $\mu_s$  is the limit of the  $s$ th eigenvalue of  $\mathbf{G}_n = \boldsymbol{\Sigma}^{1/2} \mathbf{Z}^T \mathbf{P}_X \mathbf{Z} \boldsymbol{\Sigma}^{1/2}$ ,  $n^{-a} d_0$  is the true variance ratio  $\sigma_b^2 / \sigma^2$  ( $a$  is a positive constant), and the  $w_s$  are independent standard normal random variables. The test statistic in (2.5) with  $p$ th order truncated polynomial basis satisfies (2.6) under the alternative. The  $\mu_s$  are the same as in the expressions (2.7) and (2.6), and  $\lambda = 1/d$ .

We explore the connection between the RLRT and the proposed test. Under the alternative,  $d_0$ ,  $d$  and  $\lambda$  range from small to large, depending on the complexity of the underlying function. When the underlying function is complex, such as a sine function,  $\lambda$  is small while  $d_0$  and  $d$  are large. The weights  $d\mu_s(1 + d_0\mu_s)/(1 + d\mu_s)$  in (2.7) are then approximately  $d_0\mu_s$ , which are proportional to the eigenvalues  $\mu_s$ . In this case, RLRT places larger weights along directions of the first few eigenvectors of  $\mathbf{G}_n$ . However,  $\mathbf{G}_n$  is solely determined by the design matrices of the basis functions,  $\mathbf{X}$  and  $\mathbf{Z}$ , which are not related to the true function  $f(\cdot)$  (also noted in Liu and Wang (2004)). Weighting the test statistic by the directions of eigenvectors of  $\mathbf{G}_n$  may not improve the power of the test. In contrast, the proposed test statistic  $Z_n$  with a small  $\lambda$  is approximately distributed as  $\sigma^2 \sum_{s=1}^{n-p} (\omega_s + \delta'_s)^2$ . Since  $\delta'_s$  is the  $s$ th eigenvalue of  $\mathbf{Q}^T \mathbf{f}_n$ , with  $\mathbf{f}_n = \{f(u_1), \dots, f(u_n)\}^T$ ,  $Z_n$  contains information on the true function  $f(\cdot)$ .



This comparison offers heuristics on a phenomenon observed in our simulation studies (Section 3): for more complicated functions with multiple modes, LRT is less powerful than the proposed test.

### 3. Simulation Studies

In the simulations, we generated the outcome from the model

$$y_i = d \cdot \mu(u_i) + \varepsilon_i, \quad i = 1, \dots, n,$$

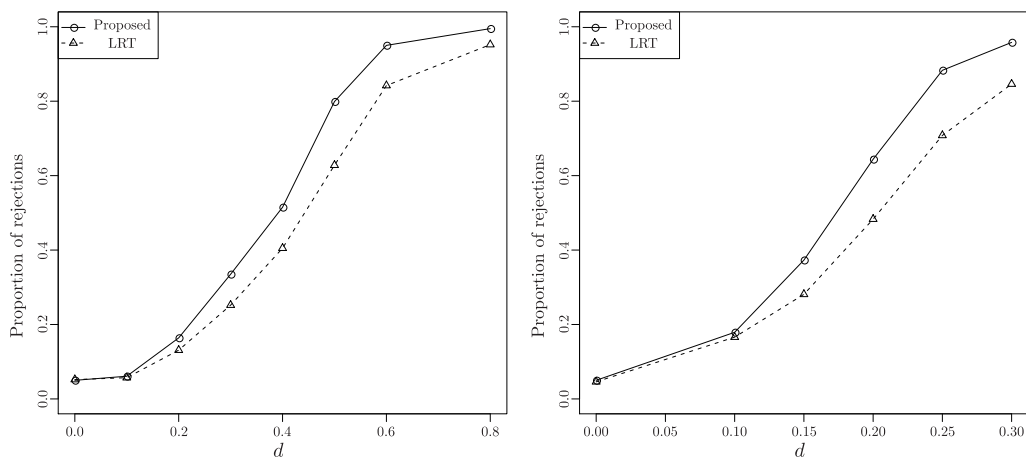
where the  $u_i$  were independently generated from an uniform distribution with support  $(0,1)$ , the underlying mean function was  $f(u) = d \cdot \mu(u)$ , and  $\mu(u)$  was  $\sin(2\pi u)$ ,  $u^3$ , or  $\exp(u)$ . To obtain the power curves, we varied the scalar  $d$  to control the deviation of the true function from the null. Specifically, type I error was computed under  $d = 0$  (the null hypothesis), and the power was computed under  $d > 0$  (the alternative hypothesis). The residual errors  $\varepsilon_i$  were i.i.d.  $N(0, 1)$ ,  $U(-1, 1)$ , or  $Laplace(a = 0, b = 1)$ . Eubank and Spiegelman (1990) and Jayasuriya (1996) observed in their simulation studies that for smoothing spline-based tests, directly applying the normal to approximate the finite sample distribution of  $T_n$  at the tail area may not be satisfactory and the type I error rate may slightly deviate from the nominal level. They used various transformations to improve accuracy of the asymptotic approximation. Here, we applied a square root transformation to the test statistic  $T_n$  in all simulation settings. The type I error rate of normal approximation to the square root transformed test statistic is satisfactory and close to both the nominal level and those based on the exact distribution obtained through permutation.

We compared the proposed test with the LRT. The exact null distribution of LRT was computed using the methods in Crainiceanu and Ruppert (2004) and Scheipl, Greven, and Küchenhoff (2008). Since the LRT selects the smoothing parameter by REML, for a fair comparison we also used a REML-based smoothing parameter to compute  $T_n$  in the normal random error scenario. Since the methods used to compute the null distribution of LRT is an exact approach, in addition to computing the power of  $T_n$  based on critical values obtained from the asymptotic distribution, we also computed power using critical values obtained from the exact null distribution of  $T_n$  through permutation. We considered two sample sizes,  $n = 100$  and  $n = 500$ . For all the scenarios, we carried out 5,000 simulation runs.

Table 1 summarizes the simulation results for the normal residual error case. Both the proposed test and LRT maintain the nominal type I error rate. In terms of power, when the underlying function is more complex, such as  $\sin(2\pi u)$ , the proposed tests (both the exact and asymptotic) are more powerful than LRT for both sample sizes. This is also seen in the plot of the power functions of the two

Table 1. Proportion of rejections in 5,000 repetitions in a nonparametric model with normal measurement error.

$f(u) = d \cdot \sin(2\pi u)$		$n=100$				$n=500$			
$d$		0	0.3	0.5	0.8	0	0.1	0.2	0.3
Exact		0.050	0.335	0.799	0.995	0.050	0.179	0.644	0.958
Asymptotic		0.055	0.354	0.809	0.998	0.051	0.189	0.658	0.961
LRT		0.052	0.252	0.628	0.952	0.046	0.166	0.483	0.846
$f(u) = d \cdot u^3$		$n=100$				$n=500$			
$d$		0	0.5	0.8	1	0	0.1	0.3	0.5
Exact		0.050	0.231	0.837	0.898	0.050	0.085	0.512	0.952
Asymptotic		0.054	0.2415	0.845	0.901	0.052	0.089	0.529	0.955
LRT		0.049	0.231	0.833	0.902	0.053	0.089	0.548	0.962
$f(u) = d \cdot \exp(u)$		$n=100$				$n=500$			
$d$		0	0.1	0.15	0.3	0	0.05	0.08	0.1
Exact		0.050	0.290	0.561	0.995	0.050	0.340	0.775	0.922
Asymptotic		0.054	0.304	0.574	0.996	0.053	0.3555	0.783	0.929
LRT		0.050	0.305	0.570	0.995	0.051	0.377	0.790	0.933

Figure 1. Proportion of rejections based on 5,000 simulations with  $f(u) = d \cdot \sin(2\pi u)$  and normal measurement error,  $n = 100$  (left panel),  $n = 500$  (right panel).

tests in Figure 1. At various effect sizes, the proposed test is more powerful than LRT. From the second and third panels of Table 1, when the underlying function is relatively simple, such as  $\mu(u) = u^3$  and  $\mu(u) = \exp(u)$ , the powers of the proposed test and LRT are similar. Figure 2 presents the power of the two tests as a function of  $d$  when  $\mu(u) = \exp(u)$  and the  $\varepsilon_i$  are normal. The two power curves are very close. The differences based on the asymptotic null distribution and exact null distribution are ignorable.

Table 2 summarizes the simulation results when the residual errors are non-

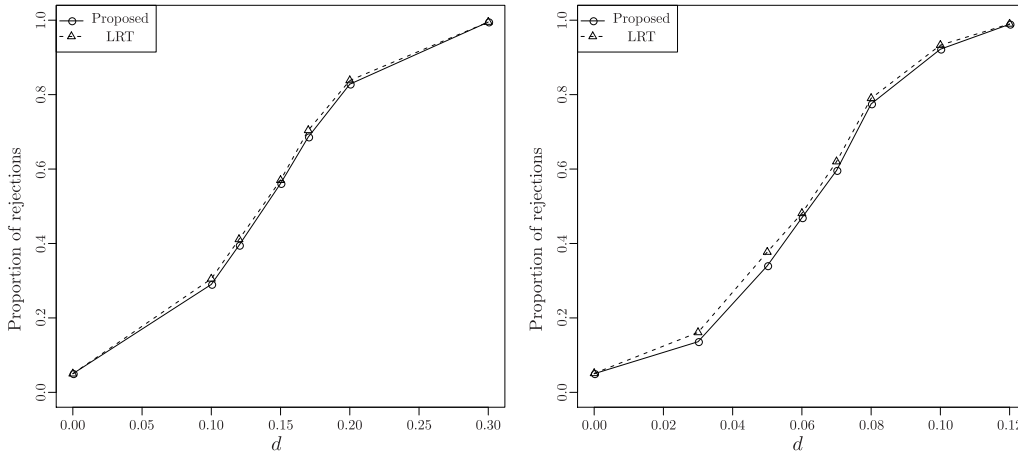


Figure 2. Proportion of rejections based on 5,000 simulations with  $f(u) = d \cdot \exp(u)$  and normal measurement error,  $n = 100$  (left panel),  $n = 500$  (right panel).

Table 2. Proportion of rejections in 5,000 repetitions with  $f(u) = d \cdot \sin(2\pi u)$  and uniform or Laplace measurement error.

U(-1,1)	n=100				n=500			
d	0	0.2	0.3	0.4	0	0.05	0.1	0.2
Asymptotic	0.059	0.445	0.791	0.975	0.050	0.148	0.536	0.995
Laplace	n=100				n=500			
d	0	0.5	0.8	1	0	0.2	0.3	0.5
Asymptotic	0.057	0.439	0.846	0.989	0.049	0.346	0.747	0.993

normal. For these cases, we used generalized cross-validation (GCV) to select the smoothing parameter. The proposed test maintained the correct type I error rate. We also computed the power under different  $d$  and report results in Table 2. To reach similar power, the required effect size  $d$  is greater for the Laplace residual errors than for the uniform residual errors.

To assess sensitivity of the test to the choice of the smoothing parameter, we computed the size and power of  $T_n$  under different  $d$  with  $\lambda$  ranging from  $10^{-4}$  to  $10^5$ . From Table 3, the size of the test was not sensitive to the values of  $\lambda$ , especially when the sample size was large. In terms of power, in all the cases it increases with increasing  $\lambda$  before reaching its highest value and then starts to decrease or becomes flat. When  $\lambda$  is large enough, for example greater than or equal to 100, there is no difference among the different choices of  $\lambda$ . As expected, these analyses suggest that a good choice of  $\lambda$  may increase power of a test. Theorem 3 and its remarks justify these observations from a theoretical point of view.

Table 3. Sensitivity of type I error and power to choice of  $\lambda$  in a nonparametric model with normal measurement error and 5,000 repetitions.

Type I error rate										
$\lambda$	$10^{-4}$	$10^{-3}$	0.01	0.1	1	10	100	$10^3$	$10^4$	$10^5$
$n = 100$	0.054	0.054	0.051	0.045	0.043	0.043	0.043	0.043	0.043	0.043
$n = 500$	0.047	0.044	0.044	0.048	0.046	0.044	0.044	0.044	0.044	0.044
Power, $f(u) = d \cdot \sin(2\pi u)$ , $n=100$										
$\lambda$	$10^{-4}$	$10^{-3}$	0.01	0.1	1	10	100	$10^3$	$10^4$	$10^5$
$d=0.3$	0.298	0.334	0.356	0.330	0.289	0.289	0.289	0.289	0.289	0.289
$d=0.5$	0.669	0.719	0.737	0.694	0.564	0.559	0.559	0.559	0.559	0.559
$d=0.8$	0.989	0.992	0.993	0.991	0.965	0.962	0.962	0.962	0.962	0.962
Power, $f(u) = d \cdot \sin(2\pi u)$ , $n=500$										
$\lambda$	$10^{-4}$	$10^{-3}$	0.01	0.1	1	10	100	$10^3$	$10^4$	$10^5$
$d=0.1$	0.159	0.185	0.202	0.210	0.187	0.174	0.174	0.174	0.174	0.174
$d=0.2$	0.588	0.643	0.673	0.692	0.591	0.530	0.528	0.528	0.528	0.528
$d=0.3$	0.940	0.96	0.973	0.977	0.944	0.873	0.871	0.871	0.871	0.871
Power, $f(u) = d \cdot \exp(u)$ , $n=100$										
$\lambda$	$10^{-4}$	$10^{-3}$	0.01	0.1	1	10	100	$10^3$	$10^4$	$10^5$
$d=0.1$	0.219	0.229	0.248	0.282	0.286	0.286	0.286	0.286	0.286	0.286
$d=0.15$	0.392	0.448	0.491	0.541	0.550	0.552	0.552	0.552	0.552	0.552
$d=0.2$	0.738	0.789	0.819	0.847	0.858	0.858	0.858	0.858	0.858	0.858
Power, $f(u) = d \cdot \exp(u)$ , $n=500$										
$\lambda$	$10^{-4}$	$10^{-3}$	0.01	0.1	1	10	100	$10^3$	$10^4$	$10^5$
$d=0.05$	0.235	0.265	0.284	0.306	0.331	0.335	0.336	0.336	0.336	0.336
$d=0.08$	0.552	0.600	0.668	0.704	0.732	0.730	0.730	0.730	0.730	0.730
$d=0.1$	0.772	0.817	0.851	0.873	0.897	0.894	0.894	0.894	0.894	0.894

For a partial linear model, we conducted several simulation studies to investigate performance of test statistic  $Z_n$  under different scenarios. The simulation model was

$$Y_i = \beta x_i + d \cdot \mu(u_i) + \epsilon_i, \quad i = 1, \dots, n,$$

where the covariate  $x_i$  were  $U(0, 1)$ ,  $\beta = 1$ , and the random errors were standard normal. Table 4 summarizes the simulation results. As before, we computed the critical value of LRT using the methods in Crainiceanu and Ruppert (2004), and computed the critical value of  $Z_n$  based both on the exact distribution through permutation and the asymptotic approximation. We used REML to choose the smoothing parameter for both tests. The proposed asymptotic approximation had a type I error rate close to the nominal level. The power comparison of  $Z_n$  with the LRT in a partial linear model revealed a similar trend to  $T_n$  in a nonparametric model: for more complicated functions,  $Z_n$  has greater power than LRT, for simpler functions,  $Z_n$  has power similar to LRT.

Table 4. Proportion of rejections in 5,000 repetitions in a partial linear model with normal measurement error.

$f(u) = d \cdot \sin(2\pi u)$	$n=100$				$n=500$			
$d$	0	0.8	1.2	2	0	0.3	0.5	0.8
Exact	0.050	0.285	0.549	0.955	0.050	0.166	0.525	0.923
Asymptotic	0.054	0.301	0.564	0.957	0.052	0.174	0.541	0.933
LRT	0.047	0.215	0.410	0.823	0.053	0.121	0.380	0.766
$f(u) = d \cdot u^3$	$n=100$				$n=500$			
$d$	0	2	3	4	0	0.8	1.5	2
Exact	0.050	0.308	0.716	0.927	0.050	0.297	0.807	0.947
Asymptotic	0.052	0.312	0.724	0.934	0.051	0.314	0.816	0.951
LRT	0.049	0.332	0.721	0.931	0.048	0.332	0.824	0.958
$f(u) = d \cdot \exp(u)$	$n=100$				$n=500$			
$d$	0	0.6	1	1.3	0	0.3	0.4	0.6
Exact	0.050	0.282	0.816	0.953	0.050	0.423	0.696	0.951
Asymptotic	0.053	0.302	0.822	0.959	0.051	0.434	0.712	0.952
LRT	0.048	0.302	0.823	0.963	0.053	0.436	0.729	0.962

## 4. Two Data Examples

### 4.1. The Framingham heart study

Our first data example addresses a research question encountered in the Framingham Heart Study (Cupples et al. (2003)). High cholesterol level is known to be one of the risk factors for cardiovascular disease (Boden (2000)). The functional relationship between obesity and cholesterol level is of interest in cardiovascular research. Here we examine the relationship between cholesterol level and body mass index (BMI) in the Framingham Heart Study baseline data. The Framingham Heart Study is a large population-based study of risk factors for cardiovascular disease. Subjects' demographic and clinical information, such as cholesterol and blood sugar level, were collected. We tested the hypothesis that BMI is associated with cholesterol level after adjusting for other predictors of cholesterol, and its linearity.

There were 777 subjects included in the analyses. We tested the significance of association between cholesterol and BMI through model (2.4), where  $y_i$  is the  $i$ th subject's cholesterol level,  $u_i$  is BMI, and  $\mathbf{x}_i$  is a vector of predictors including baseline age, sex, and smoking status. We found that the test was significant with a  $p$  value less than 0.001. We next tested the significance of departure from a linear association. This test also emerged as significant with a  $p$  value of 0.007. We show the estimated association  $\hat{f}(u)$  superimposed on a scatter plot in Figure 3. We see a non-linear trend in Figure 3, which suggests that adjusting for other factors, the relationship between cholesterol level and BMI among obese and extremely obese subjects can be different from the relationship in the normal

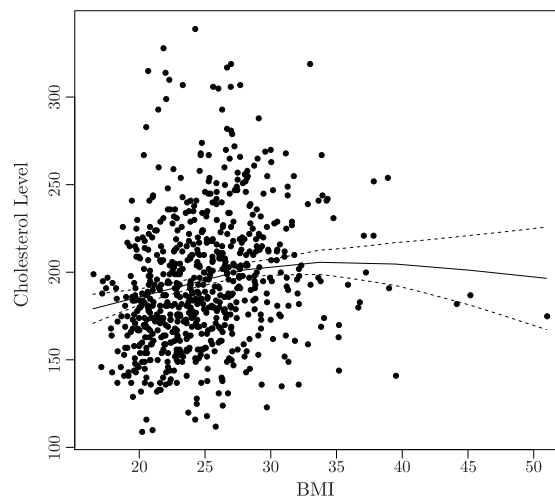


Figure 3. Scatter plot of cholesterol level versus BMI and estimated association adjusting for baseline age, sex and smoking status. The solid line is the estimated association and the dashed lines are the 95% pointwise confidence band.

weight to overweight subjects. There is a clear positive association between cholesterol level and BMI for both normal weight and overweight subjects (BMI between 18 and 30). The association trajectory is flat for obese subjects (BMI between 30 and 40) and extremely obese subjects (BMI greater than 40). This analysis suggests a potentially different pattern for the overweight, obese, and extremely obese subjects which is worth further investigation.

#### 4.2. The complicated grief study

Complicated grief (CG) is a disorder characterized by significant functional impairment lasting more than a month following six months of bereavement (Shear et al. (2005)). Patients' CG symptoms and functioning impairment attributable to CG were measured using several instruments, including the Inventory of Complicated Grief (ICG) scale and the Work and Social Adjustment Scale (WSAS). ICG, a 19-item self-report, provides a continuous measure of severity of CG. WSAS, a 5-item instrument, provides a continuous measure of a subject's degree of interference of work and social activity due to CG. We included 175 subjects (mean age = 47 years), 28 males and 147 females, at the baseline for the analysis. We tested whether there is an association between WSAS and ICG, and its linearity.

We tested the significance of association between WSAS and ICG adjusting for age and gender by model (2.4). The  $p$ -value of the test was less than 0.001. Next, we tested deviation of the association from a linear model and the result was

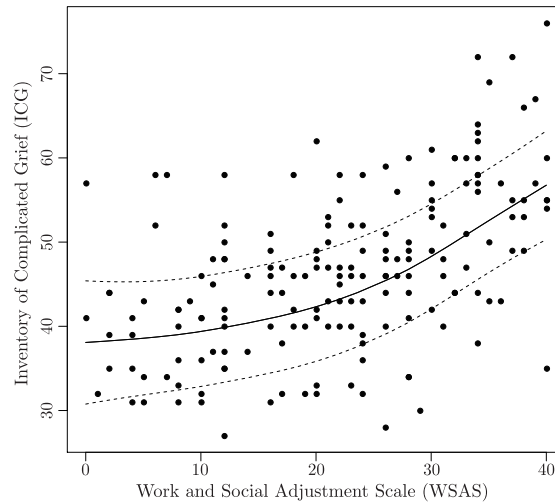


Figure 4. Scatter plot of WSAS versus ICG and estimated association adjusting for baseline age and sex. The solid line is the estimated association and the dashed lines are the 95% pointwise confidence band.

significant ( $p = 0.0012$ ). The two tests suggest a significant non-linear association between WSAS and ICG measured at baseline. Moreover, the association cannot be modeled adequately by a simple linear model. We present the scatter plot of WSAS versus ICG and the estimated association adjusting for baseline age and sex in Figure 4. The solid line is the estimated association and the dashed lines are the 95% confidence bands. From Figure 4, we see that when WSAS is less than 20, interference on work and social activities is mild or moderate, the CG symptoms only increase slightly with the increase in WSAS. When the interference of work and social activity is marked or severe (WSAS between 20 and 40), we observed a considerable positive association between WSAS and ICG. For instance, as WSAS increases from 20 to 40, the ICG changes from 42.3 (95% CI: [35.8, 48.8]) to 56.8 (95% CI: [50.3, 63.2]) and the Pearson correlation between them is 0.466 ( $p < 0.001$ ). However, as WSAS increases from 0 to 20, the ICG only varies from 38.1 (95% CI: [30.8, 45.4]) to 42.3 (95% CI: [35.8, 48.8]), and the Pearson correlation is 0.163 ( $p = 0.161$ ). Therefore, only those with marked or severe interference in work and social activities show a positive association between WSAS and CG symptoms.

In summary, the association between WSAS and ICG is more complicated than a simple linear relationship. A flexible nonparametric approach is desirable for modeling the nonlinear association.

## 5. Discussions

We considered several testing problems of a nonlinear function using penalized splines. Our theoretical investigations revealed that, compared to estimation through penalized splines, improving power for testing problems may require undersmoothing the data. In the literature, how to choose the smoothing parameter in the estimation setting has been well-studied. For example, Reiss and Ogden (2009) and Krivobokova and Kauermann (2007) suggest better performance of the REML-based smoothing parameter compared to other methods, including GCV. In the testing setting, to the best of our knowledge, no work has discussed how to choose the smoothing parameter to maximize power. Based on our results and data analyses, choosing a smoothing parameter slightly smaller than the one chosen by REML may increase power. Additionally, we find that the LRT based on a linear mixed effects model has good power for simpler functions and that the proposed test has good power for more complicated functions with a larger number of modes. Overall, how to choose an optimal smoothing parameter to maximize power in practice is still an open research question.

## Acknowledgement

Wang's research is supported by NIH grant R01NS073670. Li's research is supported by NIH grants, R21 DA024260 and P50 DA-10075, NSF grant and NNSF of China grant 11028103. The Framingham data were obtained from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine (Contract No. N01-HC-25195). The authors wish to thank Dr. Yuliya Yoncheva and Ms. April Myung for editorial assistance.

## Appendix

We state our assumptions (see also Zhou, Shen, and Wolfe (1998) and Claeskens, Kivobokova, and Opsomer (2009)).

**Assumption 1.** Let  $\delta_j = \tau_{j+1} - \tau_j$  and  $\delta = \max_{0 \leq j \leq K} \delta_j$ , where  $\tau_1, \dots, \tau_K$  are the  $K$  knots. There exists a constant  $M > 0$ , such that  $\delta / (\min_{0 \leq j \leq K} \delta_j) \leq M$  and  $\delta \sim K^{-1}$ .

**Assumption 2.** For design points  $u_i \in [a, b]$ ,  $i = 1, \dots, n$ , there exists a distribution function  $Q$  with corresponding positive continuous design density  $\rho$  such that, with  $Q_n$  the empirical distribution of  $u_1, \dots, u_n$ ,  $\sup_{u \in [a, b]} |Q_n(u) - Q(u)| = o(K^{-1})$ .

**Assumption 3.** The number of knots  $K = o(n)$ .

The assumption A1 is a weak restriction on the knot distribution, and assures that  $M^{-1} < K\delta < M$ , which is required for stable numerical computations.



The proofs of all theorems are in the online supplementary material.

## References

- Andrews, D. W. K. (1984). Non-strong mixing autoregressive processes. *J. Appl. Probab.* **21**, 930-934.
- Boden, W. (2000). High-density lipoprotein cholesterol as an independent risk factor in cardiovascular disease: assessing the data from framingham to the veterans affairs high-density lipoprotein intervention trial. *Amer. J. Cardiology* **86**, 19-22.
- Cai, Z., Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *J. Amer. Statist. Assoc.* **95**, 888-902.
- Cantoni, E. and Hastie, T. (2002). Degrees-of-freedom tests for smoothing splines. *Biometrika* **89**, 251-263.
- Chen, J. (1994). Testing goodness of fit of polynomial models via spline smoothing techniques. *Statist. Probab. Lett.* **19**, 65-76.
- Chen, H. and Wang, Y. (2011). A penalized spline approach to functional mixed effects model analysis. *Biometrics* **67**, 861-870.
- Claeskens, G., Kivobokova, T. and Opsomer, J. D. (2009). Asymptotic properties of penalized spline estimators. *Biometrika* **96**, 529-544.
- Cox, D. and 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. and Yandell, B. (1988). Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Statist.* **16**, 113-119.
- Crainiceanu, C. and Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *J. Roy. Statist. Soc. B* **65**, 165-185.
- Crainiceanu, C., Ruppert, D., Claeskens, G., and Wand, P. (2005). Exact likelihood ratio tests for penalised splines. *Biometrika* **92**, 91-103.
- Cupples, L. A., Yang, Q., Demissie, S., Copenhaver, D. and Levy, D. (2003). Description of the Framingham Heart Study data for Genetic Analysis Workshop 13. *BMC genetics*, 4(Suppl 1), S2.
- De Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Rel. Fields* **25**, 261-277.
- Eilers, P. and Marx, B. (1996). Flexible smoothing with B-splines. *Statist. Sci.* **11**, 89-121.
- Eubank, R. L. and LaRiccia, V. N. (1993). Testing for no effect in non-parametric regression. *J. Statist. Plann. Inference* **36**, 1-14.
- Eubank, R. L. and 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., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks Phenomenon. *Ann. Statist.* **29**, 153-193.
- Jayasuriya, B. R. (1996). Testing for polynomial regression using nonparametric regression techniques. *J. Amer. Statist. Assoc.* **91**, 1626-1630.
- Kauermann, G., Krivobokova, T. and Fahrmeir, L. (2009). Some asymptotic results on generalized penalized spline smoothing. *J. Roy. Statist. Soc. Ser. B* **71**, 487-503.
- Krivobokova, T. and Kauermann, G. (2007). A note on penalized splines with correlated errors. *J. Amer. Statist. Assoc.* **102**, 1328-1337.

- Li, R. and Nie, L. (2008). Efficient statistical inference procedures for partially nonlinear models and their applications. *Biometrics* **64**, 904-911.
- Li, Y. and Ruppert, D. (2008). On the asymptotics of penalized splines. *Biometrika* **95**, 415-436.
- Liu, A. and Wang, Y. (2004). Hypothesis testing in smoothing spline models. *J. Statist. Comput. Simulation* **74**, 581-597.
- Ramil-Novo, L. A. and GonzKalez-Manteiga, W.(2000). F-tests and regression ANOVA based on smoothing spline estimators. *Statist. Sinica* **10**, 819-837.
- Raz, J. (1990). Testing for no effect when estimating a smooth function by nonparametric regression: a randomization approach. *J. Amer. Statist. Assoc.* **85**, 132-138.
- Reiss, P. T. and Ogden, T. R. (2009). Smoothing parameter selection for a class of semiparametric linear models. *J. Roy. Statist. Soc. Ser. B*, **71**, 505-523.
- Ruppert, D., Wand, M. P. and Carroll, R. J. (2009). Semiparametric regression during 2003-2007. *Electronic J. Statist.* **3**, 1193-1256.
- Scheipl, F., Greven, S. and Küchenhoff, H. (2008). Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Comput. Statist. Data Anal.* **52**, 3283-3299.
- Shear, K., Frank, E., Houck, P. R. and Reynolds, C. F. (2005). Treatment of complicated grief: a randomized controlled trial. *J. Amer. Med. Assoc.* **293**, 2601-2608.
- Speckman, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist.* **13**, 970-983.
- Wahba, G. (1990). *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, Philadelphia, PA.
- Wand, M. P. (2003). Smoothing and mixed models. *Comput. Statist.* **18**, 223-249.
- Wand, M. and Ormerod, J. (2008). On semiparametric regression with O'Sullivan penalised splines. *Aust. New Zeal. J. Statist.* **50**, 179-198.
- Wang, X., Shen, J. and Ruppert, D. (2011). On the asymptotics of penalized spline smoothing. *Electronic J. Statist.* **5**, 1-17.
- Wang, Y. and Chen, H. (2012). On testing an unspecified function through a linear mixed effects model with multiple variance components. *Biometrics* **68**, 1113-1125.
- Wu, H. and Zhang, J. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis Mixed-Effects Modeling Approaches*. Wiley, New York.
- Zhang, C. (2004). Assessing the equivalence of nonparametric regression tests based on spline and local polynomial smoothers. *J. Statist. Plann. Inference* **126**, 73-95.
- Zhou, S., Shen, X. and Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Ann. Statist.* **26**, 1760-1782.
- Department of Child and Adolescent Psychiatry, New York University School of Medicine, New York, NY 10016, U.S.A.  
E-mail: huaihou.chen@nyumc.org
- Department of Biostatistics, Columbia University, New York, NY 10032, U.S.A.  
E-mail: yw2016@columbia.edu
- Department of Statistics and The Methodology Center, The Pennsylvania State University, University Park, Pennsylvania, 16802 USA.  
E-mail: rzli@psu.edu
- School of Social Work, Columbia University, New York, NY 10027, U.S.A.  
E-mail: ks2394@columbia.edu

(Received July 2012; accepted July 2013)