Statistica Sinica **19** (2009), 325-342

POLYNOMIAL SPLINE CONFIDENCE BANDS FOR REGRESSION CURVES

Jing Wang and Lijian Yang

University of Illinois at Chicago and Michigan State University

Abstract: Asymptotically exact and conservative confidence bands are obtained for a nonparametric regression function, using piecewise constant and piecewise linear spline estimation, respectively. Compared to the pointwise confidence interval of Huang (2003), the confidence bands are inflated by a factor proportional to $\{\log(n)\}^{1/2}$, with the same width order as the Nadaraya-Watson bands of Härdle (1989), and the local polynomial bands of Xia (1998) and Claeskens and Van Keilegom (2003). Simulation experiments corroborate the asymptotic theory. The linear spline band has been used to identify an appropriate polynomial trend for fossil data.

Key words and phrases: Brownian bridge, B spline, knots, nonparametric regression, quantile transformation.

1. Introduction

For two decades, nonparametric regression has been widely applied to biostatistics, econometrics, engineering and geography, due to its flexibility in modelling complex relationships among variables by "letting the data speak for themselves". Two popular nonparametric techniques are local polynomial/kernel and polynomial spline smoothing.

The kernel-type estimators, namely the Nadaraya-Watson and the local polynomial estimator, are based on locally weighted averaging. Polynomial spline estimators are "global" in terms of implementation, as a single least square procedure leads to the ultimate function estimate over the entire data range, see Stone (1994). In terms of pointwise asymptotics, of course, both kernel and spline type estimators are local in nature, see Fan and Gijbels (1996) and Huang (2003).

The fidelity of a nonparametric regressor is measured in terms of its rate of convergence to the unknown regression function. The convergence rate can be pointwise, least squares or uniform. For kernel-type estimators, rates of convergence have been established by Claeskens and Van Keilegom (2003), Fan and Gijbels (1996) and Mack and Silverman (1982), to name a few. For polynomial splines, least squares rates of convergence have been obtained by Stone (1994),

while pointwise convergence rates and asymptotic distributions have been recently established in Huang (2003). Confidence bands for polynomial spline regression, however, are available only under the restriction of homoscedastic normal errors, see Zhou, Shen and Wolfe (1998).

In this paper, we present confidence bands for univariate regression functions based on polynomial spline smoothing. We assume that observations $\{(X_i, Y_i)\}_{i=1}^n$ and unobserved errors $\{\varepsilon_i\}_{i=1}^n$ are i.i.d. copies of (X, Y, ε) satisfying

$$Y = m(X) + \sigma(X)\varepsilon, \tag{1.1}$$

where ε is conditional white noise, namely $E(\varepsilon|X) \equiv 0$, $E(\varepsilon^2|X) \equiv 1$, see Assumption (A4) in Section 2. The unknown mean m and standard deviation σ, defined on $[a, b]$, need not be of any specific form. If the data actually follows a polynomial regression model, m is a polynomial and σ is constant.

Confidence bands have been obtained for kernel-type estimators of m, see Claeskens and Van Keilegom (2003), Hall and Titterington (1988), Härdle (1989) and Xia (1998). These bounds are computationally intensive, as the kernel estimator requires solving an optimization problem at every point. In contrast, it is enough to solve only one such problem to get the polynomial spline estimator. The greatest advantages of polynomial spline estimation are simplicity of implementation and fast computation. Hence it is desirable from a theoretical and practical point of view to have confidence bands for polynomial spline estimators.

We organize our paper as follows. In Section 2 we state our main results on confidence bands constructed from (piecewise) constant/linear splines. In Section 3 we provide further insight into the error structure of spline estimators. Section 4 describes the actual steps to implement the confidence bands. Section 5 reports findings in an extensive simulation study, and the testing of the polynomial trend hypothesis for fossil data using a linear spline band. Section 6 concludes. All technical proofs are contained in Appendices A and B in the Supplement, available at http://www3.stat.sinica.edu.tw/statistica.

2. Main Results

To introduce spline functions, we divide the finite interval $[a, b]$ into $(N+1)$ subintervals $J_j = [t_j, t_{j+1}), j = 0, \ldots, N-1, J_N = [t_N, b]$. A sequence of equally-spaced points $\{t_j\}_{j=1}^N$, called interior knots, is given as

$$t_j = a + jh,\ j = 0, \ldots, N+1, \tag{2.1}$$

where $h = (b-a)/(N+1)$ is the distance between neighboring knots. We denote by $G^{(p-2)} = G^{(p-2)}[a, b]$ the space of functions that are polynomials of degree $(p-1)$ on each J_j and have $p-2$ continuous derivatives. For example,

$G^{(-1)}$ denotes the space of functions that are constant on each J_j, and $G^{(0)}$ the space of functions that are linear on each J_j and continuous on $[a,b]$.

In what follows, $\|\cdot\|_\infty$ denotes the supremum norm of a function w on $[a,b]$ and the modulus of continuity of a continuous function w on $[a,b]$ is denoted by $\omega(w,h) = \max_{x,x' \in [a,b], |x-x'| \le h} |w(x) - w(x')|$. That $\lim_{h \to 0} \omega(w,h) = 0$ follows from the uniform continuity of w on a compact interval $[a,b]$.

An asymptotically exact (conservative) $100(1-\alpha)\%$ confidence band for the unknown m over the interval $[a,b]$ consists of an estimator $\hat{m}(x)$ of $m(x)$, and lower and upper confidence limits $\hat{m}(x) - l_n(x)$ and $\hat{m}(x) + l_n(x)$ at each x in $[a,b]$ such that

$$\lim_{n \to \infty} P\{m(x) \in \hat{m}(x) \pm l_n(x), \forall x \in [a,b]\} = 1 - \alpha, \text{ exact},$$
$$\liminf_{n \to \infty} P\{m(x) \in \hat{m}(x) \pm l_n(x), \forall x \in [a,b]\} \ge 1 - \alpha, \text{ conservative}.$$

Our approach is to use the polynomial spline estimator based on data $\{(X_i, Y_i)\}_{i=1}^n$ drawn from model (1.1) given by

$$\hat{m}_p(x) = \underset{g \in G^{(p-2)}[a,b]}{\operatorname{argmin}} \sum_{i=1}^n \{Y_i - g(X_i)\}^2, p = 1, 2. \tag{2.2}$$

We then construct an error bound function l_n around this spline estimator. The technical assumptions we need are as follows.

(A1) $m(\cdot) \in C^{(p)}[a,b]$, $p = 1, 2$.

(A2) *The density function $f(\cdot)$ of X is continuous and positive on the interval $[a,b]$; the standard deviation function $\sigma(\cdot) \in C[a,b]$ is of bounded variation and has a positive lower bound on $[a,b]$.*

(A3) *The number of interior knots is $N \sim n^{1/(2p+1)}$.*

(A4) *The joint distribution $F(x, \varepsilon)$ of random variables (X, ε) satisfies*

- (a) $E(\varepsilon | X = x) \equiv 0$, $E(\varepsilon^2 | X = x) \equiv 1$;
- (b) *there exists a positive value $\eta > 1/p$ and a finite positive M_η such that $E|\varepsilon|^{2+\eta} < M_\eta$ and $\sup_{x \in [a,b]} E(|\varepsilon|^{2+\eta} | X = x) < M_\eta$.*

Assumptions (A1)−(A3) are the same as in Huang (2003), while (A4) is the same as (C2) (a) of Mack and Silverman (1982). All are typical for nonparametric regression, with (A1), (A2), and (A4) weaker than their counterparts in Härdle (1989).

To define the confidence bands, we introduce some additional notation. For any $x \in [a,b]$, define its location and relative position indices $j(x), r(x)$ as

$$j(x) = j_n(x) = \min\left\{\left[\frac{x-a}{h}\right], N\right\}, r(x) = \frac{\{x - t_{j(x)}\}}{h}. \tag{2.3}$$

Since any x is between two consecutive knots, it is clear that $t_{j_n(x)} \leq x < t_{j_n(x)+1}$, $0 \leq r(x) < 1, \forall x \in [a,b)$, and $r(b) = 1$. Let $\|\phi\|_2^2 = E\left\{\phi^2(X)\right\} = \int_a^b \phi^2(x) f(x)\, dx$, and take $\|\phi\|_{2,n}^2 = n^{-1}\sum_{i=1}^n \phi^2(X_i)$. With standard inner products notation use, note that $E\langle\phi,\varphi\rangle_n = \langle\phi,\varphi\rangle$.

Elementary algebra shows the linear equivalence of the B-spline basis to the truncated power basis introduced in Section 4, see de Boor (2001). Hence the same estimator $\hat{m}_p(x)$ $(p = 1, 2)$ can be expressed as a linear combination of either of the two basis. While the truncated power basis is convenient for implementation, it is easier to work with the B-spline basis for theoretical analysis. The B-spline basis of $G^{(-1)}$, the space of piecewise constant splines, consists of indicator functions of intervals $J_j, b_{j,1}(x) = I_j(x) = I_{J_j}(x), 0 \leq j \leq N$. The B-spline basis of $G^{(0)}$, the space of piecewise linear splines, are $\{b_{j,2}(x)\}_{j=-1}^N$, where

$$b_{j,2}(x) = K\left(\frac{x - t_{j+1}}{h}\right), j = -1, 0, \ldots, N, \text{ for } K(u) = (1 - |u|)_+ .$$

The rescaled B-spline basis $\{B_{j,p}(x)\}_{j=1-p}^N$ for $G^{(1-p)}$, with

$$B_{j,p}(x) \equiv b_{j,p}(x)\|b_{j,p}\|_2^{-1}, 1 - p \leq j \leq N, p = 1, 2,$$

features basis functions with norm 1.

To express the estimator $\hat{m}_p(x)$ based on the standardized basis $\{B_{j,p}(x)\}_{j=1-p}^N$, we introduce the following vectors in R^n for $p = 1, 2$:

$$\mathbf{Y} = (Y_1, \ldots, Y_n)^T, \mathbf{B}_{j,p}(\mathbf{X}) = \{B_{j,p}(X_1), \ldots, B_{j,p}(X_n)\}^T, j = 1 - p, \ldots, N.$$

The definition of $\hat{m}_p(x)$ in (2.2) implies that $\hat{m}_p(x) \equiv \sum_{j=1-p}^N \hat{\lambda}_{j,p} B_{j,p}(x)$, where the coefficients $\left\{\hat{\lambda}_{1-p,p}, \ldots, \hat{\lambda}_{N,p}\right\}^T$ are solutions of the least squares problem

$$\left\{\hat{\lambda}_{1-p,p}, \ldots, \hat{\lambda}_{N,p}\right\}^T = \underset{R^{N+p}}{\operatorname{argmin}} \sum_{i=1}^n \left\{Y_i - \sum_{j=1-p}^N \lambda_{j,p} B_{j,p}(X_i)\right\}^2 . \tag{2.4}$$

Typically these are solutions of the normal equations

$$\left(\left\langle B_{j,p}, B_{j',p}\right\rangle_n\right)_{j,j'=1-p}^N \left(\hat{\lambda}_{j,p}\right)_{j=1-p}^N = \left(n^{-1}\sum_{i=1}^n B_{j,p}(X_i) Y_i\right)_{j=1-p}^N .$$

It is straightforward that $\left\langle B_{j,p}, B_{j',p}\right\rangle \equiv 0, |j - j'| \geq p$, thus the inner product matrix on the left side of the normal equation is diagonal for the constant B spline basis $(p = 1)$, and tridiagonal for the linear B spline basis $(p = 2)$. According to

Lemma 3.1, the latter is approximated by its deterministic version, whose inverse has an explicit formula given in Section 4.

For $p = 2$, this inverse matrix S and its 2×2 diagonal submatrices $\{S_j, 0 \leq j \leq N\}$ are expressed as

$$S = \left(s_{j,j'}\right)_{j,j'=-1}^{N} = \left(\left\langle B_{j,2}, B_{j',2} \right\rangle\right)^{-1}, S_j = \begin{pmatrix} s_{j-1,j-1} & s_{j-1,j} \\ s_{j,j-1} & s_{j,j} \end{pmatrix}. \tag{2.5}$$

The width of the confidence bands depends on the heteroscedastic variance function. Define

$$\sigma_{n,1}^2(x) = \frac{\int_{I_{j(x)}} \sigma^2(v) f(v)\, dv}{n \| b_{j(x),1} \|_2^2}, \sigma_{n,2}^2(x) = \sum_{j,j',l,l'=-1}^{N} \frac{B_{j',2}(x) B_{l',2}(x) s_{jj'} s_{ll'} \sigma_{jl}}{n} \tag{2.6}$$

with $j(x)$ defined in (2.3), and $s_{ll'}$ in (2.5), and

$$\left(\sigma_{jl}\right)_{j,j'=-1}^{N} = \Sigma = \left\{ \int \sigma^2(v) B_{j,2}(v) B_{l,2}(v) f(v)\, dv \right\}_{j,j'=-1}^{N}. \tag{2.7}$$

These $\sigma_{n,p}^2(x)$ are shown in Lemmas A.4 and B.4 in the Supplement to be the pointwise variance functions of $\hat{m}_p(x)$, $p = 1, 2$.

We state our main results in the next two theorems.

Theorem 1. *Under Assumptions* (A1)−(A4), *if* $p = 1$, *then an asymptotic* $100(1-\alpha)\%$ *exact confidence band for* $m(x)$ *over the interval* $[a, b]$ *is*

$$\hat{m}_1(x) \pm \sigma_{n,1}(x) \left\{2 \log(N+1)\right\}^{\frac{1}{2}} d_n, \tag{2.8}$$

in which $\sigma_{n,1}(x)$ *is given in* (2.6), *and*

$$d_n = 1 - \left\{2 \log(N+1)\right\}^{-1} \left[\log\left\{ -\frac{\log(1-\alpha)}{2} \right\} + \frac{\log\log(N+1) + \log 4\pi}{2} \right]. \tag{2.9}$$

Note that $\sigma_{n,1}$ can be replaced by $\sigma(x) \{f(x) nh\}^{-1/2}$, according to (A.7) in Lemma A.4 in the Supplement.

The confidence band in Theorem 1 is superior to the connected error bar of Hall and Titterington (1988) in two aspects: we treat random instead of equally-spaced designs and, by applying the strong approximation theorem of Tusnády (1977), our confidence band is asymptotically exact rather than conservative. The upcrossing result (Theorem 3.4) used in the proof of Theorem 1 is for a sequence of i.i.d. Gaussian variables, while its counterpart used in Bickel and Rosenblatt (1973), Härdle (1989) and Rosenblatt (1976) is for a continuous time Gaussian process.

Theorem 2. *Under Assumptions* (A1)−(A4), *if* $p = 2$, *then an asymptotic* $100(1-\alpha)\%$ *conservative confidence band for* $m(x)$ *over the interval* $[a, b]$ *is*

$$\hat{m}_2(x) \pm \sigma_{n,2}(x) \{2\log(N+1) - 2\log\alpha\}^{\frac{1}{2}}, \tag{2.10}$$

where $\sigma_{n,2}(x)$ *is as in* (2.6).

Note that $\sigma_{n,2}$ is replaceable by $\sigma(x) \{2f(x) nh/3\}^{-1/2} \mathbf{\Delta}^T(x) S_{j(x)} \mathbf{\Delta}(x)$ according to Lemma B.4, and by $\sigma(x) \{2f(x) nh/3\}^{-1/2} \mathbf{\Delta}^T(x) \Xi_{j(x)} \mathbf{\Delta}(x)$ according to Lemma B.3.

Theorem 2 on linear confidence band bears no similarity to the local polynomial bands in Claeskens and Van Keilegom (2003) and Xia (1998), but the width of all these bands is of the order $n^{-1/5} (\log n)^{1/2}$. The asymptotic variance function $\sigma^2_{n,2}(x)$ of $\hat{m}_2(x)$ in (2.6) is a special unconditional version of equation (6.2) in Huang (2003). Thus, the linear band localized at any given point x is a factor of $\{2\log(N+1)\}^{1/2}$ wider than the pointwise confidence interval of Huang (2003).

3. Error Decomposition

In this section, we break the estimation error $\hat{m}_p(x) - m(x)$ into a bias term and a noise term. To understand this decomposition, we begin by discussing the spline space $G^{(p-2)}$ and the representation of the spline estimator $\hat{m}_p(x)$ in (2.2).

We note first the uniform convergence of the empirical inner product to the theoretical counterparts.

Lemma 3.1. *Under Assumptions* (A2) *and* (A3), *as* $n \to \infty$,

$$A_{n,1} = \sup_{0 \le j \le N} \left| \|B_{j,1}\|^2_{2,n} - 1 \right| = O_p\left(\sqrt{n^{-1}h^{-1}\log(n)}\right), \tag{3.1}$$

$$A_{n,2} = \sup_{g_1, g_2 \in G^{(0)}} \left| \frac{\langle g_1, g_2\rangle_n - \langle g_1, g_2\rangle}{\|g_1\|_2 \|g_2\|_2} \right| = O_p\left(\sqrt{n^{-1}h^{-1}\log(n)}\right). \tag{3.2}$$

We write $\mathbf{Y}$ as the sum of a signal vector $\mathbf{m}$ and a noise vector $\mathbf{E}$ as

$$\mathbf{Y} = \mathbf{m} + \mathbf{E}, \mathbf{m} = \{m(X_1), \ldots, m(X_n)\}^T, \mathbf{E} = \{\sigma(X_1)\varepsilon_1, \ldots, \sigma(X_n)\varepsilon_n\}^T.$$

Projecting the response $\mathbf{Y}$ onto the linear space $G_n^{(p-2)}$ spanned by $\{\mathbf{B}_{j,p}(\mathbf{X})\}_{j=1-p}^N$, one gets

$$\hat{\mathbf{m}}_p = \{\hat{m}_p(X_1), \ldots, \hat{m}_p(X_n)\}^T = \mathrm{Proj}_{G_n^{(p-2)}} \mathbf{Y} = \mathrm{Proj}_{G_n^{(p-2)}} \mathbf{m} + \mathrm{Proj}_{G_n^{(p-2)}} \mathbf{E}.$$

Correspondingly in the space $G^{(p-2)}$ of spline functions, one has

$$\hat{m}_p(x) = \tilde{m}_p(x) + \tilde{\varepsilon}_p(x) \tag{3.3}$$

$$\tilde{m}_p(x) = \sum_{j=1-p}^{N} \tilde{\lambda}_{j,p} B_{j,p}(x), \tilde{\varepsilon}_p(x) = \sum_{j=1-p}^{N} \tilde{a}_{j,p} B_{j,p}(x). \tag{3.4}$$

The vectors $\left\{\tilde{\lambda}_{1-p,p},\ldots,\tilde{\lambda}_{N,p}\right\}^T$ and $\{\tilde{a}_{1-p,p},\ldots,\tilde{a}_{N,p}\}^T$ are solutions to (2.4) with Y_i replaced by $m(X_i)$ and $\sigma(X_i)\varepsilon_i$ respectively.

We next cite two important results from de Boor (2001) and Huang (2003).

Theorem 3.1. *There exists a constant $C_p > 0$, for $p \geq 1$ such that for every $m \in C^{(p)}[a,b]$, there exists a function $g \in G^{(p-2)}[a,b]$ with*

$$\|g-m\|_\infty \leq C_p \left\|\omega\left(m^{(p-1)},h\right)\right\|_\infty h^{p-1} \leq C_p \left\|m^{(p)}\right\|_\infty h^p.$$

Theorem 3.2. *There exists a constant $C_p > 0$, for $p \geq 1$ such that for any $m \in C^{(p)}[a,b]$ and the function $\tilde{m}_p(x)$ defined in* (3.4),

$$\|\tilde{m}_p(x) - m(x)\|_\infty \leq C_p \inf_{g\in G^{(p-2)}} \|g-m\|_\infty = O_p\left(h^p\right). \tag{3.5}$$

According to Theorem 3.2, the bias term $\tilde{m}_p(x) - m(x)$ is of order $O_p(h^p)$. Hence the main hurdle for a proof of Theorems 1 and 2 is the noise term $\tilde{\varepsilon}_p(x)$. This is handled by the next two propositions.

Proposition 3.1. *With $\sigma_{n,1}(x)$ given in (2.6), the process $\sigma_{n,1}^{-1}(x)\tilde{\varepsilon}_1(x), x \in [a,b]$, is almost surely uniformly approximated by a Gaussian process $U(x), x \in [a,b]$, with covariance structure*

$$EU(x)U(y) = \sum_{j=0}^{N} I_j(x)\cdot I_j(y) = \delta_{j(x),j(y)}, \forall x,y \in [a,b],$$

where $\delta_{j,l} = 1$ if $j = l$ and 0 otherwise.

Proposition 3.2. *For a given $0 < \alpha < 1$,and $\sigma_{n,2}(x)$ as given in* (2.6),

$$\liminf_{n\to\infty} P\left[\sup_{x\in[a,b]}\left|\sigma_{n,2}^{-1}(x)\tilde{\varepsilon}_2(x)\right| \leq \{2\log(N+1) - 2\log\alpha\}^{\frac{1}{2}}\right] \geq 1-\alpha. \tag{3.6}$$

We next state the Strong Approximation Theorem of Tusnády (1977). It will be used later in the proof of Lemmas A.6 and B.6 in the Supplement, key steps to the proof of Propositions 3.1 and 3.2. Let $U_1,\ldots,U_n$ be i.i.d. r.v.'s on the 2 -dimensional unit square with $P(U_i < \mathbf{t}) = \lambda(\mathbf{t}), \mathbf{0} \leq \mathbf{t} \leq \mathbf{1}$, where $\mathbf{t} = (t_1,t_2)$, $\mathbf{1} = (1,1)$, and $\lambda(\mathbf{t}) = t_1t_2$. The empirical distribution function $F_n^u(\mathbf{t})$ based on sample $(U_1,\ldots,U_n)$ is defined as $F_n^u(\mathbf{t}) = n^{-1}\sum_{i=1}^n I_{\{U_i<\mathbf{t}\}}$ for $\mathbf{0} \leq \mathbf{t} \leq \mathbf{1}$. The 2-dimensional Brownian bridge $B(\mathbf{t})$ is defined by $B(\mathbf{t}) = W(\mathbf{t}) - \lambda(\mathbf{t})W(\mathbf{1})$ for $\mathbf{0} \leq \mathbf{t} \leq \mathbf{1}$, where $W(\mathbf{t})$ is a 2-dimensional Wiener process.

Theorem 3.3. *There is a version of $F_n^u(\mathbf{t})$ and $B(\mathbf{t})$ such that*

$$P\left[\sup_{\mathbf{0}\leq\mathbf{t}\leq\mathbf{1}}\left|n^{\frac{1}{2}}\{F_n^u(\mathbf{t}) - \lambda(\mathbf{t})\} - B(\mathbf{t})\right| > (C\log n + x)\frac{\log n}{n^{\frac{1}{2}}}\right] < Ke^{-\lambda x} \tag{3.7}$$

holds for all x, where C, K, λ are positive constants.

The well-known Rosenblatt quantile transformation is

$$\left(X', \varepsilon'\right) = M\left(X, \varepsilon\right) = \left\{F_X\left(X\right), F_{\varepsilon|X}\left(\varepsilon|X\right)\right\}. \tag{3.8}$$

It produces random variables X' and ε' with independent and identical uniform distributions on the interval $[0, 1]$. This transformation has been used, for instance, in Bickel and Rosenblatt (1973) and Härdle (1989). Replacing $\mathbf{t} = (t_1, t_2)$ in Theorem 3.3 with (X', ε'), and the stochastic process $n^{1/2}\left\{F_n^u\left(\mathbf{t}\right) - \lambda\left(\mathbf{t}\right)\right\}$ with

$$Z_n\left\{M^{-1}\left(x', \varepsilon'\right)\right\} = Z_n\left(x, \varepsilon\right) = \sqrt{n}\left\{F_n\left(x, \varepsilon\right) - F\left(x, \varepsilon\right)\right\}, \tag{3.9}$$

where $F_n\left(x, \varepsilon\right)$ denotes the empirical distribution of (X, ε), (3.7) implies that there exists a version of the 2-dimensional Brownian bridge B such that

$$\sup_{x,\varepsilon}\left|Z_n\left(x, \varepsilon\right) - B\left\{M\left(x, \varepsilon\right)\right\}\right| = O\left(n^{-\frac{1}{2}}\log^2 n\right), \text{w.p.1.} \tag{3.10}$$

The next result on upcrossing probabilities is from Leadbetter, Lindgren and Rootzén (1983), Theorem 1.5.3. It plays the role of Theorem A1 in Bickel and Rosenblatt (1973), or of Theorem C in Rosenblatt (1976).

Theorem 3.4. *If $\xi_1, \ldots, \xi_n$ are i.i.d. standard normal r.v.'s, then for $M_n = \max\left\{\xi_1, \ldots, \xi_n\right\}, \tau \in R$, as $n \to \infty$,*

$$P\left\{a_n\left(M_n - b_n\right) \leq \tau\right\} \to \exp\left(-e^{-\tau}\right), P\left\{\left|M_n\right| \leq \frac{\tau}{a_n} + b_n\right\} \to \exp\left(-2e^{-\tau}\right),$$

where $a_n = \left(2\log n\right)^{1/2}, b_n = \left(2\log n\right)^{1/2} - (1/2)\left(2\log n\right)^{-1/2}\left(\log\log n + \log 4\pi\right)$.

4. Implementation

In this section, we describe procedures to implement the confidence bands in Theorems 1 and 2. Our codes are written in XploRe in order to use kernel smoothing. See Härdle, Hlávka and Klinke (2000).

Given any sample $\{(X_i, Y_i)\}_{i=1}^n$ from model (1.1), we take $a = \min(X_1, \ldots, X_n)$ and $b = \max\left(X_1, \ldots, X_n\right)$. When outliers are present, the 2.5%- and 97.5%-tiles can be used as endpoints a, b. The number of interior knots is taken to be $N = \left[c_1 n^{1/(2p+1)}\right] + c_2$, where c_1 and c_2 are positive integers. The knots are taken to be equally spaced, as in (2.1). Since an explicit formula for coverage probability does not exist for the bands, there is no optimal method to select (c_1, c_2). In simulation, the simple choice of 5 for c_1 and 1 for c_2 seems to work well, so these are set as default values.

The least squares problem in (2.2) can be solved via the truncated power basis $\left\{1, x, \ldots, x^{p-1},\ (x-t_j)_+^{p-1}, j=1,\ldots,N\right\}$. In other words

$$\hat{m}_p(x) = \sum_{k=0}^{p-1} \hat{\gamma}_k x^k + \sum_{j=1}^{N} \hat{\gamma}_{j,p} (x-t_j)_+^{p-1}, p=1,2, \tag{4.1}$$

where the coefficients $\{\hat{\gamma}_0, \ldots, \hat{\gamma}_{p-1}, \hat{\gamma}_{1,p}, \ldots, \hat{\gamma}_{N,p}\}^T$ are solutions to the least squares problem

$$\{\hat{\gamma}_0, \ldots, \hat{\gamma}_{N,p}\}^T = \underset{R^{N+p}}{\operatorname{argmin}} \sum_{i=1}^{n} \left\{ Y_i - \sum_{k=0}^{p-1} \gamma_k X_i^k - \sum_{j=1}^{N} \gamma_{j,p} (X_i - t_j)_+^{p-1} \right\}^2.$$

When constructing the confidence bands, one needs to evaluate the functions $\sigma_{n,p}^2(x)$ in (2.6) differently for the exact and conservative bands, and the description is separated into two subsections. For both cases, following Lemmas A.4 and B.4 in the Supplement, one estimates the unknown functions f and σ^2 and then plugs in these estimates, the same approach taken in Hall and Titterington (1988), Härdle (1989) and Xia (1998). This is analogous to using $\overline{X} \pm 1.96 \times s_n/\sqrt{n}$ instead of $\overline{X} \pm 1.96 \times \sigma/\sqrt{n}$ as a large sample 95% confidence interval for a normal population mean μ, where the sample standard deviation s_n is a plugin substitute for the unknown population standard deviation σ.

Let $\widetilde{K}(u) = 15\left(1-u^2\right)^2 I\{|u| \le 1\}/16$ be the quartic kernel, s_n =the sample standard deviation of $(X_i)_{i=1}^n$, and

$$\hat{f}(x) = n^{-1} \sum_{i=1}^{n} h_{\mathrm{rot},f}^{-1} \widetilde{K}\left(\frac{X_i - x}{h_{\mathrm{rot},f}}\right), h_{\mathrm{rot},f} = (4\pi)^{\frac{1}{10}} \left(\frac{140}{3}\right)^{\frac{1}{5}} n^{-\frac{1}{5}} s_n, \tag{4.2}$$

where $h_{\mathrm{rot},f}$ is the rule-of-thumb bandwidth in Silverman (1986). Define $\mathbf{Z}_p = \{Z_{1,p}, .., Z_{n,p}\}^T$, $p=1,2$, with $Z_{i,p} = \{Y_i - \hat{m}_p(X_i)\}^2$ and

$$\mathbf{X} = \mathbf{X}(\mathbf{x}) = \begin{pmatrix} 1 & \ldots, & 1 \\ X_1 - x & \ldots, & X_n - x \end{pmatrix}^T, \mathbf{W} = \mathbf{W}(x) = \operatorname{diag}\left\{ \widetilde{K}\left(\frac{X_i - x}{h_{\mathrm{rot},\sigma}}\right) \right\}_{i=1}^n,$$

where $h_{\mathrm{rot},\sigma}$ is the rule-of-thumb bandwidth of Fan and Gijbels (1996) based on data $(X_i, Z_{i,p})_{i=1}^n$. Take

$$\hat{\sigma}_p^2(x) = \left(1, 0\right)\left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}\mathbf{Z}_p, p=1,2. \tag{4.3}$$

The following uniform consistency result is given in Bickel and Rosenblatt (1973) and Fan and Gijbels (1996):

$$\max_{p=1,2} \sup_{x\in[a,b]} |\hat{\sigma}_p(x) - \sigma(x)| + \sup_{x\in[a,b]} \left|\hat{f}(x) - f(x)\right| = o_p(1). \tag{4.4}$$

As two referees pointed out, the estimation of $\sigma^2(x)$ can be done via spline instead of kernel smoothing as well. We decided to use the kernel smoothing technique only because it is a built-in procedure in the XploRe language in which we wrote our codes. Since the local linear smoothing is done on the data $(X_i, Z_{i,p})_{i=1}^n$ with all $Z_{i,p}$ positive, the estimator $\hat{\sigma}_p^2(x)$ always takes positive values.

4.1. Implementing the exact band

The function $\sigma_{n,1}(x)$ is approximated by the following, with $\hat{f}(x)$ and $\hat{\sigma}_1(x)$ defined in (4.2) and (4.3), $j(x)$ defined in (2.3):

$$\hat{\sigma}_{n,1}(x) = \hat{\sigma}_1(x)\hat{f}^{-\frac{1}{2}}(x)n^{-\frac{1}{2}}h^{-\frac{1}{2}}.$$

Then (4.4) implies that, as $n \to \infty$, the band below is asymptotically exact, with $\hat{m}_1(x)$ given in (4.1) and d_n in (2.9):

$$\hat{m}_1(x) \pm \hat{\sigma}_{n,1}(x)\{2\log(N+1)\}^{\frac{1}{2}} d_n. \tag{4.5}$$

4.2. Implementing the conservative band

According to Lemma B.3 in the Supplement, for $0 \le j \le N$, the matrix Ξ_j approximates matrix S_j uniformly. Hence the band below is asymptotically conservative, with $\hat{m}_2(x)$ given in (4.1):

$$\hat{m}_2(x) \pm \hat{\sigma}_{n,2}(x)\{2\log(N+1) - 2\log\alpha\}^{\frac{1}{2}}, \tag{4.6}$$

where the function $\sigma_{n,2}(x)$ in (2.6) for the linear band is estimated consistently by

$$\hat{\sigma}_{n,2}(x) = \left\{\mathbf{\Delta}^T(x)\Xi_{j(x)}\mathbf{\Delta}(x)\right\}^{\frac{1}{2}}\hat{\sigma}_2(x)\left\{\frac{2}{3}\hat{f}(x)nh\right\}^{-\frac{1}{2}},$$

with $j(x)$ defined in (2.3), and $\hat{f}(x)$ and $\hat{\sigma}_2(x)$ defined in (4.2) and (4.3). Here $\mathbf{\Delta}(x)$ and Ξ_j are defined as follows:

$$\mathbf{\Delta}(x) = \begin{pmatrix} c_{j(x)-1}\{1 - r(x)\} \\ c_{j(x)}r(x) \end{pmatrix}, c_j = \begin{cases} \sqrt{2} & j = -1, N \\ 1 & j = 0, \ldots, N-1 \end{cases}, \tag{4.7}$$

$$\Xi_j = \begin{pmatrix} l_{j+1,j+1} & l_{j+1,j+2} \\ l_{j+2,j+1} & l_{j+2,j+2} \end{pmatrix}, j = 0, \ldots, N, \tag{4.8}$$

with the $l_{ik}, |i-k| \le 1$, defined through the matrix inversion

$$M_{N+2} = \begin{pmatrix} 1 & \frac{\sqrt{2}}{4} & & & & 0 \\ \frac{\sqrt{2}}{4} & 1 & \frac{1}{4} & & & \\ & \frac{1}{4} & 1 & \ddots & & \\ & & \ddots & \ddots & \frac{1}{4} & \\ & & & \frac{1}{4} & 1 & \frac{\sqrt{2}}{4} \\ 0 & & & & \frac{\sqrt{2}}{4} & 1 \end{pmatrix}_{(N+2)\times(N+2)} = (l_{ik})^{-1}_{(N+2)\times(N+2)}, \quad (4.9)$$

and computed via (4.12), (4.13), and (4.14) given below.

To calculate the matrix M^{-1}_{N+2}, which is needed for (4.8), we use equation (43) in Gantmacher and Krein (1960), and Theorem 4.5 in Zhang (1999).

Theorem 4.1. *For the symmetric Jacobi matrix*

$$J = \begin{pmatrix} a_1 & b_1 & & & 0 \\ b_1 & \ddots & \ddots & & \\ & \ddots & \ddots & b_{N+1} & \\ 0 & & b_{N+1} & a_{N+2} & \end{pmatrix}_{(N+2)\times(N+2)},$$

$J^{-1} = (l_{ik})_{(N+2)\times(N+2)}$ *satisfies*

$$l_{i,k} = \psi_i \chi_k, i \le k, l_{i,k} = \psi_k \chi_i, k \le i, \quad (4.10)$$

where

$$\psi_i = \frac{(-1)^i \det\left(J_{(1,\ldots,i-1)}\right) b_i b_{i+1} \cdots b_{N+1}}{\det(J)}, \chi_k = \frac{(-1)^k \det\left(J_{(k+1,\ldots,N+2)}\right)}{b_k b_{k+1} \cdots b_{N+1}}, \quad (4.11)$$

and $J_{(1,\ldots,i-1)}$ is defined as the upper left $(i-1)\times(i-1)$ submatrix of J, $\det(J)$ is the determinant of matrix J, and $J_{(k+1,\ldots,N+2)}$ is the corresponding lower right $(N+2-k)\times(N+2-k)$ submatrix.

Theorem 4.2. *For the tridiagonal*

$$T_N = \begin{pmatrix} a & b & & 0 \\ c & a & \ddots & \\ & \ddots & \ddots & b \\ 0 & & c & a \end{pmatrix}_{N\times N}, N \ge 1,$$

if $a^2 \ne 4bc$,

$$\det T_N = \frac{\alpha^{N+1} - \beta^{N+1}}{\alpha - \beta}, \alpha = \frac{a + \sqrt{a^2 - 4bc}}{2}, \beta = \frac{a - \sqrt{a^2 - 4bc}}{2}.$$

Now we let

$$z_1 = \frac{2+\sqrt{3}}{4}, z_2 = \frac{2-\sqrt{3}}{4}, \theta = \frac{z_2}{z_1} = \left(2-\sqrt{3}\right)^2 = 7-4\sqrt{3}, \tag{4.12}$$

and apply Theorems 4.1 and 4.2 to obtain

$$\begin{aligned} l_{11} &= l_{N+2,N+2} = \frac{8z_1^2\left(1-\theta^{N+1}\right) - z_1\left(1-\theta^N\right)}{8z_1^2\left(1-\theta^{N+1}\right) - 2z_1\left(1-\theta^N\right) + 8\left(1-\theta^{N-1}\right)}, \\ l_{i,i} &= \frac{\left\{8z_1\left(1-\theta^{N+2-i}\right) - \left(1-\theta^{N+1-i}\right)\right\}\left\{8z_1\left(1-\theta^{i-1}\right) - \left(1-\theta^{i-2}\right)\right\}}{(z_1-z_2)\left\{64z_1^2\left(1-\theta^{N+1}\right) - 16z_1\left(1-\theta^N\right) + 64\left(1-\theta^{N-1}\right)\right\}} \end{aligned} \tag{4.13}$$

for $2 \le i \le N+1$, and

$$\begin{aligned} l_{12} &= l_{N+1,N+2} = \frac{\left(-2\sqrt{2}\right)z_1\left(1-\theta^N\right) - \left(1-\theta^{N-1}\right)}{8z_1^2\left(1-\theta^{N+1}\right) - 2z_1\left(1-\theta^N\right) + 8\left(1-\theta^{N-1}\right)}, \\ l_{i,i+1} &= \frac{\left\{8z_1\left(1-\theta^{N+1-i}\right) - \left(1-\theta^{N-i}\right)\right\}\left\{8z_1\left(1-\theta^{i-1}\right) - \left(1-\theta^{i-2}\right)\right\}}{(-4)(z_1-z_2)\left\{64z_1^2\left(1-\theta^{N+1}\right) - 16z_1\left(1-\theta^N\right) + 64\left(1-\theta^{N-1}\right)\right\}} \end{aligned} \tag{4.14}$$

for $2 \le i \le N$. By the symmetry of the matrix M_{N+2}, the lower diagonal entries are $l_{i+1,i} = l_{i,i+1}, \forall i = 1, \ldots, N+1$.

5. Examples

5.1. Simulation example

To illustrate the finite-sample behavior of our confidence bands, we simulate data from model (1.1), with $X \sim U[-1/2, 1/2]$, and

$$m(x) = \sin(2\pi x), \sigma(x) = \sigma_0 \frac{100 - \exp(x)}{100 + \exp(x)}, \varepsilon \sim N(0,1).$$

The noise levels are $\sigma_0 = 0.2, 0.5$, while sample sizes are taken to be $n = 100, 200, 500, 10,000$, and confidence levels are $1-\alpha = 0.99, 0.95$. Table 1 contains the coverage probabilities as the percentage of coverage of the true curve at all data points by the confidence bands in (4.5) and (4.6), over 500 replications of sample size n. Also listed in the table, in brackets are averages over the 500 replications of the confidence bands' enclosed areas which, as one referee pointed out, measure the confidence bands' widths.

At all noise levels, the constant bands become much closer with sample sizes increasing. The coverage percentages for linear bands show positive confirmation of Theorem 2. At sample size 200, regardless of noise level, the candidate linear band in (4.6) achieves at least 95.6% and 90% for the confidence levels $1-\alpha = 0.99, 0.95$, respectively.

Table 1. Coverage probabilities and mean areas (in brackets) from 500 replications.

σ_0	n	$1-\alpha$	Constant Band	Linear Band
0.2	100	0.99	0.458 (0.617)	0.896 (0.417)
		0.95	0.246 (0.501)	0.814 (0.363)
	200	0.99	0.708 (0.472)	0.962 (0.314)
		0.95	0.456 (0.387)	0.904 (0.274)
	500	0.99	0.834 (0.336)	0.988 (0.223)
		0.95	0.456 (0.279)	0.958 (0.195)
0.5	100	0.99	0.618 (1.384)	0.904 (1.039)
		0.95	0.504 (1.124)	0.814 (0.902)
	200	0.99	0.860 (1.096)	0.960 (0.784)
		0.95	0.716 (0.899)	0.902 (0.683)
	500	0.99	0.932 (0.805)	0.988 (0.557)
		0.95	0.802 (0.668)	0.960 (0.488)

It is clear that larger sample sizes guarantee improved coverage, with reasonable coverage achieved at moderate sample sizes. The linear band outperforms the constant band, corroborating the theory. The noise level has more influence on the constant band than on the linear one.

According to Theorems 1 and 2, the constant band has larger enclosed area than the linear band for the same data. In addition, the enclosed area decreases with increasing sample size and with lower confidence level. All these three phenomena can be observed from the numbers in brackets in Table 1. In addition, Table 1 shows that the enclosed area is larger for noise level $\sigma_0 = 0.5$ than for $\sigma_0 = 0.2$, consistent with Theorems 1 and 2 as well.

For the linear bands, we have also carried out 500 simulations for sample size $n = 10,000$. Regardless of the noise level, the coverages are both 99.4% for $\alpha = 0.01$, and 97.6% for $\alpha = 0.05$. Both are higher than the nominal coverages of 99% and 95%, and consistent with their conservative definitions. Remarkably, it takes only 88 minutes to run 500 simulations with a sample size as large as $10,000$ on a Pentium 4 PC. This is extremely fast considering that nonparametric regression is done without WARPing, see Härdle, Hlávka and Klinke (2000).

The graphs in Figure 1 were created based on two samples of size 100 and 500, respectively, each with four types of symbols: points (data), center thin solid line (true curve), center dashed line (the estimated curve), upper and lower thick solid line (confidence band). In all figures, the confidence bands for $n = 500$ are thinner and fit better than those for $n = 100$.

5.2. Fossil data example

The fossil data reflects global climate millions of years ago through ratios of strontium isotopes found in fossil shells. These were studied by Chaudhuri and Marron (1999) to detect structure, via kernel smoothing. The corresponding

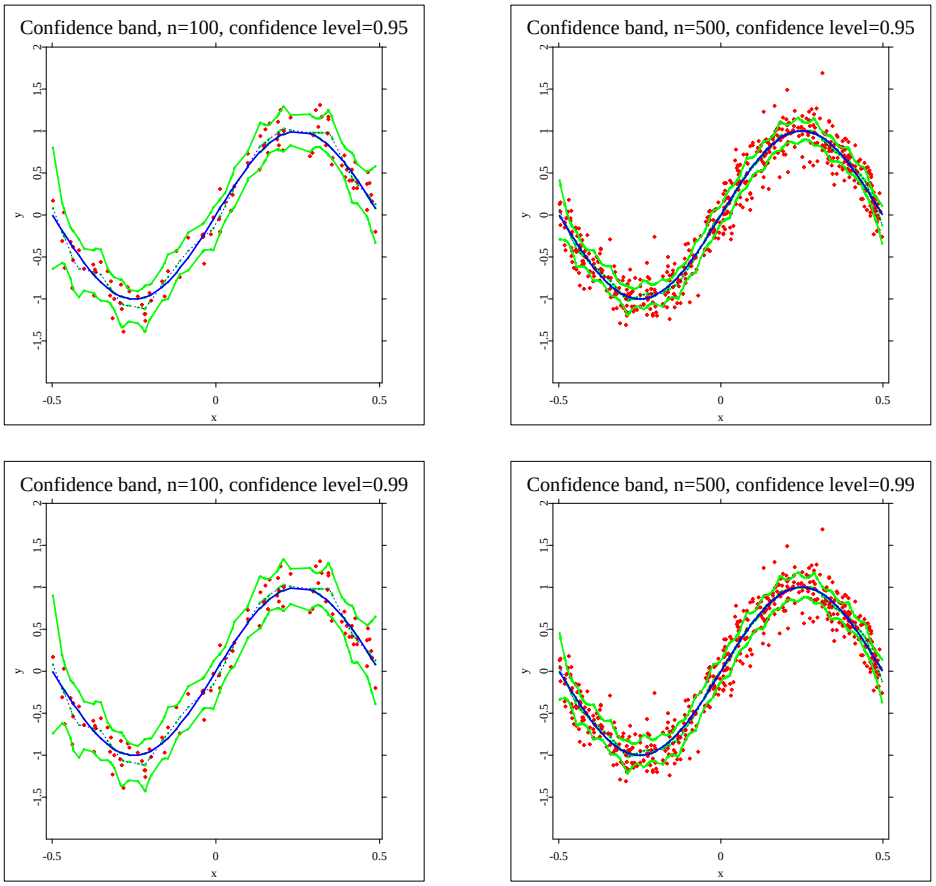

Figure 1. Plots of confidence bands (thick solid curves), the linear spline estimator $\hat{m}_2(x)$ (dashed curve), the true function $m(x) = \sin(2\pi x)$ (thin solid curve), and the data scatter plots. The bands are computed from (4.6).

penalized spline fit was provided in Ruppert, Wand and Carroll (2003). In this section we test the polynomial form of the fossil data regression curve. The null hypothesis is $H_0 : m(x) = \sum_{k=1}^{d} a_k x^k$, with polynomial degree $d = 2, 3, 5, 6$. The response Y is the strontium isotopes ratio after linear transformation, $Y = 0.70715 + \text{ratio} * 10^{-5}$, since all the values are very close to 0.707, while the predictor X is the fossil shell age in millions of years.

In Figure 2, the center dotted line is the linear spline fit. The upper/lower thin lines represent a linear band implemented according to (4.6). The solid line is the least squares polynomial fit with degree d. Clearly, the oversmoothed quadratic null curve ($d = 2$) is rejected at significance level 0.01, since it is not totally covered by the confidence band with confidence level 0.99. When $d = 3, 5$

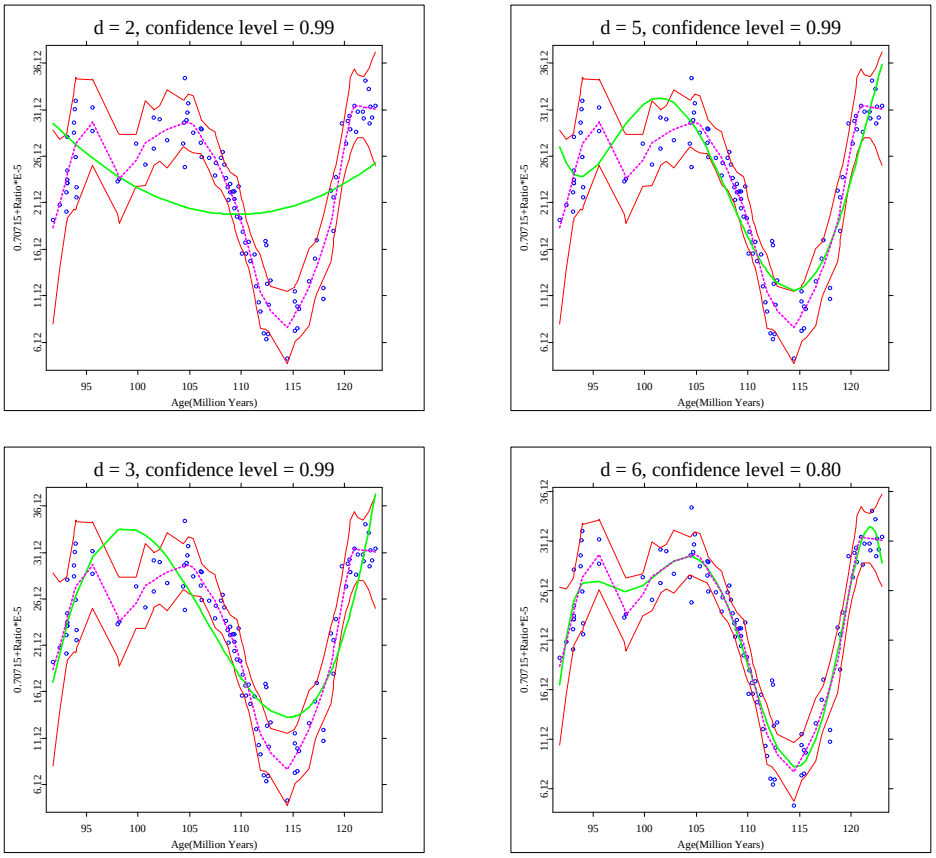

Figure 2. Plots of null hypothesis curves of $H_0 : m(x) = \sum_{k=1}^{d} a_k x^k, d = 2, 3, 5, 6$ (solid line), linear confidence bands (upper and lower thin lines), the linear spline estimator (dotted line) and the data (circle).

the null solid curves can capture the dip in the range of $110 - 115$ million years, but still do not fit well. Thus both null parametric models H_0 are rejected at the level 0.01. The same conclusion is reached for $d = 4$, although the graph is not included. In the case of $d = 6$, all significant features are shown in the null polynomial curve, the relative high ratio before 105 million years, the substantial dip around 115 million years, and the relative flat stage between 95 million and 105 million. Given an 80% confidence band the entire null curve falls between the upper and lower limits, even though the band is narrower than that with confidence 99%. In other words, a p-value greater than 0.20 indicates that the null hypotheses of a degree 6 polynomial is not rejected. The shape of the polynomial curve with $d = 6$ is consistent with the findings in Chaudhuri and Marron (1999) and Ruppert, Wand and Carroll (2003).

6. Conclusions

We provide closed forms of confidence bands constructed from polynomial spline regression. Asymptotic properties are established for equally spaced, non-adaptive selection of knots. Extension to adaptive design is infeasible, as Härdle, Marron and Yang (1997) have shown that adaptive knots selection can lead to L_∞ inconsistency.

It is possible, however, to extend the constant band in Theorem 1 to unequally spaced deterministic knots subject to mesh constraints, as in Huang (2003). The linear band in Theorem 2 does not allow such direct extension. This is one of the reasons that the constant band remains viable despite the fact that the linear band has much better theoretical property and practical performance. The constant band is also kept for its simplicity. When implemented according to (4.5) with estimation on equally-spaced knots, the confidence limits at x are the same as those at the nearest knot $t_{j(x)}$, so the constant band is in fact $(N+1)$ independently inflated confidence intervals. In contrast, the linear band has to be calibrated at each new point x, and the confidence limits at x and $t_{j(x)}$ are different.

One referee pointed out the advantages of the linear band over the constant band, and conjectured that further improvement can be made by using the more popular cubic spline. While in principle our method can be extended to cubic or other higher order splines, the main complication is not theoretical, but computational. Closed form solutions for the inverse of the inner product matrix of a B-spline basis exist only for constant and linear splines, with the aid of (4.10) and (4.11), but not for higher order splines due to the multi-diagonal shape of the inner product matrix. This creates substantial difficulty when computing the width of the confidence band. Further research is needed to implement a cubic spline confidence band.

Extension to multivariate regression is difficult for lack of a sharp approximation, as in (3.7). This limitation is also in Claeskens and Van Keilegom (2003) and Xia (1998). However, the univariate bands in this paper are still valuable for multivariate regression for the following reason: semiparametric dimension reduction models such as the additive model, the partial linear model and the single index model, provide ways to reduce multivariate nonparametric regression to some form of univariate smoothing. For instance, the components of the additive model, the nonparametric component of the partially linear model, and the nonparametric link function of the single index model are all estimable via univariate smoothing.

Acknowledgements

This research is part of the first author's dissertation work at Michigan State University under the supervision of the second author, and has been supported

in part by NSF awards DMS 0405330, 0706518, BCS 0308420 and SES 0127722. The helpful comments by two referees are gratefully acknowledged.

References

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

Chaudhuri, P. and Marron, J. S. (1999). SiZer for exploration of structures in curves. *J. Amer. Statist. Assoc.* **94**, 807-823.

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

de Boor, C. (2001). *A Practical Guide to Splines*. Springer-Verlag, New York.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.

Gantmacher, F. R. and Krein, M. G. (1960). *Oszillationsmatrizen, Oszillationskerne und kleine Schwingungen mechanischer Systeme*. Akademie-Verlag, Berlin.

Hall, P. and Titterington, D. M. (1988). On confidence bands in nonparametric density estimation and regression. *J. Multivariate Anal.* **27**, 228-254.

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

Härdle, W., Hlávka, Z. and Klinke, S. (2000). *XploRe Application Guide*. Springer-Verlag, Berlin.

Härdle, W., Marron, J. S. and Yang, L. (1997). Discussion of "Polynomial splines and their tensor products in extended linear modeling" by Stone et al. *Ann. Statist.* **25**, 1443-1450.

Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *Ann. Statist.* **31**, 1600-1635.

Johnson, R. A. and Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis*. Prentice-Hall, New Jersey.

Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York.

Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. Verw. Gebiete* **61**, 405-415.

Rosenblatt, M. (1976). On the maximal deviation of k-dimensional density estimates. *Ann. Prob.* **4**, 1009-1015.

Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press, Cambridge.

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

Stone, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Ann. Statist.* **22**, 118-184.

Tusnády, G. (1977). A remark on the approximation of the sample df in the multidimensional case. *Periodica Mathematica Hungarica* **8**, 53-55.

Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *J. Roy. Statist. Assoc. Ser. B* **60**, 797-811.

Wang, J. and Yang, L. (2006). Polynomial Spline Confidence Bands for Regression Curves. http://www.msu.edu/~yangli/bandfull.pdf.

Zhang, F. (1999). *Matrix Theory. Basic Results and Techniques*. Springer-Verlag, New York.

Zhou, S., Shen, X. and Wolfe, D. A. (1998). Local asymptotics of regression splines and confidence regions. *Ann. Statist.* **26**, 1760-1782.

Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL 60607, U.S.A.

E-mail: wangjing@math.uic.edu

Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, U.S.A.

E-mail: yang@stt.msu.edu

(Received February 2007; accepted July 2007)