POLYNOMIAL SPLINE CONFIDENCE BANDS FOR REGRESSION CURVES

by Jing Wang and Lijian Yang

Michigan State University

Asymptotically exact and conservative confidence bands are obtained for 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 of \( \{ \log (n) \}^{1/2} \), with the same width order as the Nadaraya-Watson bands of Härde (1989), the local polynomial bands of Xia (1998) and Claeskens and Van Keilegom (2003). Simulation experiments provide strong evidence that corroborates with the asymptotic theory. Application of the linear band to motorcycle data rejects the polynomial regression model.

1 Introduction

For the past two decades, nonparametric regression has been widely used in many statistical applications, from biostatistics to econometrics, from engineering to the forecasting of time series. This is due to its flexibility in modelling complex relationships among variables by “letting the data speak for themselves”. Two popular nonparametric smoothing techniques are local polynomial/kernel and polynomial spline. The kernel type estimators are “local”, treated comprehensively in Fan and Gijbels (1996). The polynomial spline estimators, on the other hand, are global, see Stone (1985, 1994) 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 of all three types have been established by Mack and Silverman (1982), Fan and Gijbels (1996), Claeskens and Van Keilegom (2003), to name a few. For polynomial splines, least squares rates of convergence have been obtained by Stone (1985, 1994), while pointwise convergence rates and asymptotic distribution have been recently established in Huang (2003). Confidence band for polynomial spline regression, however, remains unavailable except under the strong restriction of homoscedastic normal errors, see Zhou, Shen and Wolfe (1998).

In this paper, we present confidence bands of univariate regression function 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 \( (X, Y, \varepsilon) \) satisfying the regression model

\[
Y = m(X) + \sigma(X) \varepsilon, \tag{1}
\]

where the joint distribution of \( (X, \varepsilon) \) satisfies Assumption (A4) in Section 2. The unknown mean and standard deviation functions \( m(x) \) and \( \sigma(x) \), defined on interval \([a, b]\), need not to be of

*Supported in part by NSF grants DMS 0405330, BCS 0308420 and SES 0127722.


Key words and phrases: Brownian bridge, B spline, knots, nonparametric regression, quantile transformation
any specific form. If the data actually follows a polynomial regression model, \( m(x) \) would be a polynomial and \( \sigma(x) \), a constant.

Confidence band has been obtained for kernel type estimators of \( m(x) \) by Hall and Titterington (1988), Härdle (1989), Xia (1998), Claeskens and Van Keilegom (2003). These are computationally intensive since a least squares estimation has to be done at every point. In contrast, it is enough to solve only one least squares problem to get the polynomial spline estimator. The greatest advantages of polynomial spline estimation are its simplicity of implementation and fast computation. Hence, it is desirable from a theoretical as well as a practical point of view to have a confidence band 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 insights 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 application to the testing of polynomial trend hypothesis for the well-known motorcycle data. Section 6 concludes. All technical proofs are contained in Appendices A and B.

### 2 Main results

To introduce the spline functions, 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, are given as

\[
t_0 = a < t_1 < \cdots < t_N < b = t_{N+1}, t_j = a + jh, \quad j = 0, 1, \ldots, N + 1,
\]

in which \( h = (b-a)/(N+1) \) 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 has continuous \((p-2)\)th derivative. For example, \( G^{(-1)} \) denotes the space of functions that are constant on each \( J_j \), and \( G^{(0)} \) denotes the space of functions that are linear on each \( J_j \) and continuous on \([a, b]\).

We denote by \( \| r \|_\infty \) the supremum norm of a function \( r \) on \([a, b]\), i.e. \( \| r \|_\infty = \sup \{ r(x) \mid x \in [a, b] \} \), the moduli of continuity of a continuous function \( r \) on \([a, b]\) by \( \omega(r, h) = \max_{x, x' \in [a, b], |x-x'| \leq h} | r(x) - r(x') | \).

By the uniform continuity of \( r \) on a compact interval \([a, b]\), \( \lim_{h \to 0} \omega(r, h) = 0 \).

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

\[
\begin{align*}
\lim_{n \to \infty} \mathbb{P} \{ m(x) \in \hat{m}(x) \pm l_n(x), \forall x \in [a, b] \} &= 1 - \alpha, \text{ exact,} \\
\lim_{n \to \infty} \mathbb{P} \{ m(x) \in \hat{m}(x) \pm l_n(x), \forall x \in [a, b] \} &\geq 1 - \alpha, \text{ conservative.}
\end{align*}
\]

Our approach is to construct error bound function \( l_n(x) \) around the following polynomial spline estimator based on data \( \{(X_i, Y_i)\}_{i=1}^n \) drawn from model (1)

\[
\hat{m}_p(x) = \arg\min_{g \in G^{(p-2)}[a, b]} \sum_{i=1}^n \{ Y_i - g(X_i) \}^2, \quad p = 1, 2.
\]  

(2.1)

The technical assumptions we need are as follows:

(A1) The regression function \( m(x) \in C^{(p)}[a, b], p = 1, 2 \).
of piecewise constant splines, are indicator functions of intervals

It is straightforward to see that

The subinterval length $h \sim n^{-1/(2p+1)}$. I.e., the number of interior knots $N \sim n^{1/(2p+1)}$.

The joint distribution $F(x, \varepsilon)$ of random variables $(X, \varepsilon)$ satisfies the following:

(a) The error is a white noise: $E(\varepsilon | X = x) = 0$, $E(\varepsilon^2 | X = x) = 1$.

(b) There exists a positive value $\delta > 1/p$ and finite positive $M_\delta$ such that $E|\varepsilon|^{2+\delta} < M_\delta$ and

$\sup_{x \in [a,b]} E \left( |\varepsilon|^{2+\delta} | X = x \right) < M_\delta$.

Assumptions (A1)-(A3) are the same as in Huang (2003), while Assumption (A4) is the same as

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

(A3) The subinterval length $h \sim n^{-1/(2p+1)}$. I.e., the number of interior knots $N \sim n^{1/(2p+1)}$.

(A4) The joint distribution $F(x, \varepsilon)$ of random variables $(X, \varepsilon)$ satisfies the following:

(a) The error is a white noise: $E(\varepsilon | X = x) = 0$, $E(\varepsilon^2 | X = x) = 1$.

(b) There exists a positive value $\delta > 1/p$ and finite positive $M_\delta$ such that $E|\varepsilon|^{2+\delta} < M_\delta$ and

$\sup_{x \in [a,b]} E \left( |\varepsilon|^{2+\delta} | X = x \right) < M_\delta$.

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

$$j(x) = j_n(x) = \min \left\{ \left[ \frac{x-a}{h} \right], N \right\}, \delta(x) = \frac{x-t_{j_n(x)} h}{h}. \tag{2.2}$$

It is clear that $t_{j_n(x)} \leq x < t_{j_n(x)+1}, 0 \leq \delta(x) < 1, \forall x \in [a,b],$ and $\delta(b) = 1$. Denote by $\| \phi \|_2$ the theoretical $L^2$ norm of a function $\phi$ on $[a,b]$, $\| \phi \|_2^2 = E \left\{ \phi^2(X) \right\} = \int_a^b \phi^2(x) f(x) dx$, and the empirical $L^2$ norm as $\| \phi \|_{2,n}^2 = n^{-1} \sum_{i=1}^n \phi^2(X_i)$.

Corresponding inner products are defined by

$$\langle \phi, \varphi \rangle = \int_a^b \phi(x) \varphi(x) f(x) dx = E \left\{ \phi(X) \varphi(X) \right\}, \langle \phi, \varphi \rangle_n = \frac{1}{n} \sum_{i=1}^n \phi(X_i) \varphi(X_i).$$

for any $L^2$-integrable functions $\phi, \varphi$ on $[a,b]$. Clearly $E \langle \phi, \varphi \rangle_n = \langle \phi, \varphi \rangle$.

Although the truncated power basis is used in implementation (see Section 4), it is more convenient to work with the B-spline basis for theoretical analysis. The B-spline basis of $G^{(-1)}$, the space of piecewise constant splines, are indicator functions of intervals $J_j, b_{j,1}(x) = I_j(x) = I_{j_n^+}(x), j = 0, 1, ..., N$. The B-spline basis of $G^{(0)}$, the space of piecewise linear splines, are $\{b_{j,2}(x)\}_{j=-1}^N$

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

Define next their theoretical norms

$$c_{j,n} = \| b_{j,1} \|_{2,n}^2 = \int_a^b I_j(x) f(x) dx, d_{j,n} = \| b_{j,2} \|_{2,n}^2 = \int_a^b K^2 \left( \frac{x-t_{j+1}}{h} \right) f(x) dx. \tag{2.3}$$

We introduce the rescaled B-spline basis $\{B_{j,1}(x)\}_{j=0}^N$ and $\{B_{j,2}(x)\}_{j=-1}^N$ for $G^{(-1)}$ and $G^{(0)}$

$$B_{j,1}(x) \equiv b_{j,1}(x) (c_{j,n})^{-1/2}, j = 0, ..., N,$$

$$B_{j,2}(x) \equiv b_{j,2}(x) (d_{j,n})^{-1/2}, j = -1, ..., N. \tag{2.4}$$

It is straightforward to see that

$$\| B_{j,1} \|_2^2 = 1, j = 0, 1, ..., N, \langle B_{j,1}, B_{j',1} \rangle \equiv 0, j \neq j'.$$
The inner product matrix $V$ of the B-spline basis $\{B_{j,2}(x)\}_{j=-1}^{N}$ is denoted as

$$V = (v_{j'}^{(j)})_{j,j'= -1}^{N} = \left(\langle B_{j',2}, B_{j,2}\rangle\right)_{j,j'= -1}^{N},$$

whose inverse $S$ and $2 \times 2$ diagonal submatrices of $S$ are expressed as

$$S = (s_{j'}^{(j)})_{j,j'= -1}^{N} = V^{-1}, S_{j} = \begin{pmatrix} s_{j-1,j-1} & s_{j-1,j} \\ s_{j,j-1} & s_{j,j} \end{pmatrix}, j = 0, ..., N.$$  (2.7)

Next define matrices $\Sigma, \Delta (x)$ and $\Xi_{j}$ as

$$\Sigma = (\sigma_{jl})_{j,j'= -1}^{N} = \left\{ \int \sigma^{2} (v) B_{j,2} (v) B_{l,2} (v) f (v) dv \right\}_{j,j'= -1}^{N}.$$  (2.8)

$$\Delta (x) = \begin{pmatrix} c_{j}(x) \end{pmatrix}_{j = -1}^{1} \begin{pmatrix} 1 - \delta (x) \end{pmatrix}, c_{j} = \begin{pmatrix} \sqrt{2} & j = -1, N \\ 1 & j = 0, ..., N - 1 \end{pmatrix}$$

$$\Xi_{j} = \begin{pmatrix} l_{j+1,j+1} & l_{j+1,j+2} \\ l_{j+2,j+1} & l_{j+2,j+2} \end{pmatrix}, j = 0, 1, ..., N, $$

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

$$M_{N+2} = \begin{pmatrix} 1 & \sqrt{2}/4 & 0 \\ \sqrt{2}/4 & 1 & 1/4 \\ 0 & 1/4 & \ddots \\ & \ddots & \ddots & 1/4 \\ & & \ddots & \sqrt{2}/4 \\ & & & 1/4 \end{pmatrix}_{(N+2) \times (N+2)} = (l_{ik})_{(N+2) \times (N+2)}.$$  (2.10)

and computed via (4.14), (4.17), and (4.18).

We define now

$$\sigma_{n,1}^{2} (x) = \frac{\int I_{j(x)} (v) \sigma^{2} (v) f (v) dv}{nc_{j(x),n}}, \sigma_{n,2}^{2} (x) = \frac{1}{n} \sum_{j,j',l,l'= -1}^{N} B_{j',2} (x) B_{l,2} (x) s_{j,j'} s_{l,l'} \sigma_{jl},$$

with $j (x)$ defined in (2.2), $c_{j,n}$ in (2.3), $B_{j,2} (x)$ in (2.4), and $s_{ll'}$ and $\sigma_{jl}$ in (2.7), (2.8). These $\sigma_{n,p}^{2} (x)$ are shown in Lemmas A.5, B.4 to be the pointwise variance functions of $\hat{m}_{p} (x), p = 1, 2$.

We now 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 interval $[a, b]$ is

$$\hat{m}_{1} (x) \pm \sigma_{n,1} (x) \left\{ 2 \log (N + 1) \right\}^{1/2} d_{n},$$

in which $\sigma_{n,1} (x)$ is given in (2.11) and can be replaced by $\sigma (x) \left\{ f (x) nh \right\}^{-1/2}$, according to (A.8) in Lemma A.5, and

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

(2.13)
Theorem 2 Under Assumptions (A1)-(A4), if \( p = 2 \), then an asymptotic 100 \((1 - \alpha)\)\% conservative confidence band for \( m(x) \) over interval \([a, b]\) is
\[
\hat{m}_2(x) = \hat{m}_2(x) \pm \sigma_{n,2}(x) \{ 2 \log (N + 1) - 2 \log \alpha \}^{1/2}, \tag{2.14}
\]
in which \( \sigma_{n,2}(x) \) is given in (2.11) and can be replaced by \( \sigma(x) \{ 2f(x) nh / 3 \}^{-1/2} \Delta^T(x) S_{j(x)} \Delta(x) \), according to Lemma B.4, and by \( \sigma(x) \{ 2f(x) nh / 3 \}^{-1/2} \Delta^T(x) \Xi_{j(x)} \Delta(x) \) according to Lemma B.3.

The construction in Theorem 1 is similar to the connected error bar of Hall and Titterington (1988). Ours is superior in two aspects: first, we treat not only equally-spaced designs, but random designs; second, by applying the strong approximation theorem of Tusnády (1977), our confidence band is asymptotically exact rather than conservative. The error bars of Hall and Titterington (1988) are based on a kernel estimator while ours regressogram. The upcrossing results (Theorem 3.4) used in the proof of Theorem 1 is also different from that used in Bickel and Rosenblatt (1973), Rosenblatt (1976) and Härdle (1989). Theorem 2 on linear confidence band, however, bears no similarity to the local polynomial bands in Xia (1998), Claeskens and Van Keilegom (2003), except the width of the band being of the same order \( n^{-1/5} (\log n)^{1/2} \). The asymptotic variance function \( \sigma_{n,2}(x) \) of \( \hat{m}_2(x) \) in (2.11) is a special unconditional version of equation (6.2), in Huang (2003), Remark 6.1, page 1624. Thus, the linear band localized at any given point \( x \), is only a factor of \((\log n)^{1/2}\) 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 and the width of the band being of the same order \( n^{-1/5} (\log n)^{1/2} \). The upcrossing results (Theorem 3.4) used in the proof of Theorem 1 is also different from that used in Bickel and Rosenblatt (1973), Rosenblatt (1976) and Härdle (1989). Theorem 2 on linear confidence band, however, bears no similarity to the local polynomial bands in Xia (1998), Claeskens and Van Keilegom (2003), except the width of the band being of the same order \( n^{-1/5} (\log n)^{1/2} \). The asymptotic variance function \( \sigma_{n,2}(x) \) of \( \hat{m}_2(x) \) in (2.11) is a special unconditional version of equation (6.2), in Huang (2003), Remark 6.1, page 1624. Thus, the linear band localized at any given point \( x \), is only a factor of \((\log n)^{1/2}\) wider than the pointwise confidence interval of Huang (2003).

\[
\hat{m}_2(x) = \hat{m}_2(x) \pm \sigma_{n,2}(x) \{ 2 \log (N + 1) - 2 \log \alpha \}^{1/2}, \tag{2.14}
\]
in which \( \sigma_{n,2}(x) \) is given in (2.11) and can be replaced by \( \sigma(x) \{ 2f(x) nh / 3 \}^{-1/2} \Delta^T(x) S_{j(x)} \Delta(x) \), according to Lemma B.4, and by \( \sigma(x) \{ 2f(x) nh / 3 \}^{-1/2} \Delta^T(x) \Xi_{j(x)} \Delta(x) \) according to Lemma B.3.

The construction in Theorem 1 is similar to the connected error bar of Hall and Titterington (1988). Ours is superior in two aspects: first, we treat not only equally-spaced designs, but random designs; second, by applying the strong approximation theorem of Tusnády (1977), our confidence band is asymptotically exact rather than conservative. The error bars of Hall and Titterington (1988) are based on a kernel estimator while ours regressogram. The upcrossing results (Theorem 3.4) used in the proof of Theorem 1 is also different from that used in Bickel and Rosenblatt (1973), Rosenblatt (1976) and Härdle (1989). Theorem 2 on linear confidence band, however, bears no similarity to the local polynomial bands in Xia (1998), Claeskens and Van Keilegom (2003), except the width of the band being of the same order \( n^{-1/5} (\log n)^{1/2} \). The asymptotic variance function \( \sigma_{n,2}(x) \) of \( \hat{m}_2(x) \) in (2.11) is a special unconditional version of equation (6.2), in Huang (2003), Remark 6.1, page 1624. Thus, the linear band localized at any given point \( x \), is only a factor of \((\log n)^{1/2}\) wider than the pointwise confidence interval of Huang (2003).
We write $Y$ as the sum of a signal vector $m$ and a noise vector $E$

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

Projecting this relationship into the linear space spanned by $G_n^{(p-2)} = \{B_{j,p}(X)\}_{j=1-p}^N$, a subspace of $R^n$, one gets

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

It entails that in the space $G^{(p-2)}$ of spline functions

$$\hat{m}_p(x) = \tilde{m}_p(x) + \tilde{\varepsilon}_p(x), \quad (3.4)$$

where

$$\tilde{m}_p(x) = \sum_{j=1-p}^N \tilde{\lambda}_{j,p} B_{j,p}(x), \ \tilde{\varepsilon}_p(x) = \sum_{j=1-p}^N \tilde{a}_{j,p} B_{j,p}(x). \quad (3.5)$$

The vectors $\{\tilde{\lambda}_{1-p,p}, ..., \tilde{\lambda}_{N,p}\}^T$ and $\{\tilde{a}_{1-p,p}, ..., \tilde{a}_{N,p}\}^T$ are solutions to (3.3) with $Y_i$ replaced by $m(X_i)$ and $\sigma(X_i)\varepsilon_i$ respectively.

We cite next two important results, the first from de Boor (2001), p. 149, the second is Theorem 5.1 of Huang (2003).

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

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

**Theorem 3.2** There is an absolute constant $C_p > 0, p \geq 1$ such that for any $m \in C^{(p)}[a, b]$ and the function $\tilde{m}_p(x)$ defined in (3.5), with probability approaching 1

$$\|\tilde{m}_p(x) - m(x)\|_{\infty} \leq C_p \inf_{g \in G^{(p-2)}} \|g - m\|_{\infty} = O_p(h^p). \quad (3.6)$$

According to (3.4), the estimation error $\hat{m}_p(x) - m(x) = \{\tilde{m}_p(x) - m(x)\} + \tilde{\varepsilon}_p(x)$ where 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 of proving 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 (2.11), the process $\sigma_{n,1}(x)^{-1}\varepsilon_1(x), x \in [a, b]$ is almost surely uniformly approximated by a Gaussian process $U(x), x \in [a, b]$ with covariance structure

$$EU(x)U(y) = \sum_{j=0}^N I_j(x) \cdot I_j(y) = \delta_1(x,y), \forall x, y \in [a, b],$$

where $\delta_{j,l}$ is the Kronecker symbol, i.e., $\delta_{j,l} = 1$ if $j = l$ and 0 otherwise.

**Proposition 3.2** For a given $0 < \alpha < 1$, and $\sigma_{n,2}(x)$ as given in (2.11)

$$\liminf_{n \to \infty} P\left[\sup_{x \in [a, b]} \left|\sigma_{n,2}^{-1}(x)\tilde{\varepsilon}_2(x)\right| \leq \{2\log(N + 1) - 2\log\alpha\}^{1/2}\right] \geq 1 - \alpha. \quad (3.7)$$
We state next the strong approximation theorem of Tusnády (1977), which will be used later in the proof of Lemmas A.7 and B.6, key steps in proving Propositions 3.1 and 3.2 respectively.

**Theorem 3.3** Let $U_1, \ldots, U_n$ be i.i.d. r.v.’s on the 2-dimensional unit square with

$$P(U_i < t) = \lambda(t), 0 \leq t \leq 1,$$

where $t = (t_1, t_2)$ and 1 = (1, 1) are 2-dimensional vectors, $\lambda(t) = t_1t_2$. The empirical distribution function $F_n(t)$ based on sample $(U_1, \ldots, U_n)$ is $F_n(t) = n^{-1} \sum_{i=1}^n 1_{\{U_i < t\}}$ for $0 \leq t \leq 1$. The 2-dimensional Brownian bridge $B(t)$ is defined by $B(t) = W(t) - \lambda(t) W(1)$ for $0 \leq t \leq 1$, where $W(t)$ is a 2-dimensional Wiener process. Then there is a version of $\lambda(t)$ and $B(t)$ such that

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

holds for all $x$, where $C, K, \lambda$ are positive constants.

For the rest of the paper, we denote the well-known Rosenblatt quantile transformation as

$$(X', \epsilon') = M(X, \epsilon) = \{F_X(x), F_{\epsilon|X}(\epsilon|x)\},$$

which produces random variables $X'$ and $\epsilon'$ with independent and identical uniform distribution on the interval $[0, 1]$. This transformation had been used in, for instance, Bickel and Rosenblatt (1973), Hárdle (1989). Substituting the vector $t = (t_1, t_2)$ in Theorem 3.3 with $(X', \epsilon')$, and the stochastic process $n^{1/2} \{F_n(t) - \lambda(t)\}$ with

$$Z_n \{M^{-1}(x', \epsilon')\} = Z_n(x, \epsilon) = \sqrt{n} \{F_n(x, \epsilon) - F(x, \epsilon)\},$$

where $F_n(x, \epsilon)$ denotes the empirical distribution of $(X, \epsilon)$, then (3.8) implies that there exists a version of 2-dimensional Brownian bridge $B$ such that

$$\sup_{x, \epsilon} \left| Z_n(x, \epsilon) - B \{M(x, \epsilon)\} \right| = O \left(n^{-1/2} \log^2 n\right), \text{w.p.1.} \quad (3.11)$$

The next result on upcrossing probability is from Leadbetter, Lindgren and Rootzén (1983), Theorem 1.5.3, page 14. In our proof of Theorem 1, it plays the role of Theorem A1 in Bickel and Rosenblatt (1973) or 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 $M_n = \max \{\xi_1, \ldots, \xi_n\}$, $\tau \in R$

$$P\{a_n(M_n - b_n) \leq \tau\} \to \exp(-e^{-\tau}), P\{|M_n| \leq \tau/a_n + b_n\} \to \exp(-2e^{-\tau}), \text{ as } n \to \infty$$

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

4 Implementation

In this section, we describe the procedures to implement the confidence bands in Theorems 1 and 2. We have written our codes in XploRe due to the convenience of using certain kernel type estimators. Information on XploRe is in Hárdle, Hlávka and Klinke (2000).

Given any sample $\{(X_i, Y_i)\}_{i=1}^n$ from model (1), we use $\min (X_1, \ldots, X_n)$ and $\max (X_1, \ldots, X_n)$ respectively as the endpoints of interval $[a, b]$. Minor adjustments could be made for outliers. The
number of interior knots is taken to be \( N = [c_1 n^{1/(2p+1)}] + c_2 \), where \( c_1 \) and \( c_2 \) are positive integers. As with previous works on confidence bands (Härdle 1989, Xia 1998, Claeskens and Van Keilegom 2003), explicit formula of coverage probability for the bands do not exist, hence 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 (2.1) can be solved via the truncated power basis \( 1, x, \ldots, x^{p-1}, (x - t_j)^{p-1}, j = 1, \ldots, N \). In other words
\[
\hat{m}_p(x) = \sum_{k=0}^{p-1} \hat{\gamma}_k x^k + \sum_{j=1}^{N} \hat{\gamma}_{j,p} (x - t_j)^{p-1}, \tag{4.1}
\]
where the coefficients \{\( \hat{\gamma}_0, \ldots, \hat{\gamma}_{p-1}, \hat{\gamma}_{1,p}, \ldots, \hat{\gamma}_{N,p} \)\} are solutions of the following least squares problem
\[
\{\hat{\gamma}_0, \ldots, \hat{\gamma}_{p-1}, \hat{\gamma}_{1,p}, \ldots, \hat{\gamma}_{N,p} \}^T = \underset{\{\gamma_0, \ldots, \gamma_{p-1}, \gamma_{1,p}, \ldots, \gamma_{N,p}\} \in \mathbb{R}^{N+p}}{\text{argmin}} \sum_{i=1}^{n} \left\{ Y_i - \sum_{k=0}^{p-1} \gamma_k x_i^k + \sum_{j=1}^{N} \gamma_{j,p} (x_i - t_j)^{p-1} \right\}^2. \tag{4.2}
\]

When constructing the confidence bands, one needs to evaluate the functions \( \sigma^2_{n,p}(x) \) in (2.11). This is done differently for the exact and conservative bands, and the description is separated into two subsections. For both constant and linear bands, according to Lemmas A.5, B.4, one needs the unknown functions \( f(x) \) and \( \sigma^2(x) \). Let \( \tilde{K}(u) = 15 \left(1 - u^2 \right)^2 I \{ |u| \leq 1 \} / 16 \) be the quartic kernel, \( s_n \) = the sample standard deviation of \((X_i)_{i=1}^n\) and
\[
\hat{f}(x) = n^{-1} \sum_{i=1}^{n} h_{rot,f}^{-1} \tilde{K} \left( \frac{X_i - x}{h_{rot,f}} \right), \quad h_{rot,f} = (4\pi)^{1/10} (140/3)^{1/5} n^{-1/5} s_n, \tag{4.2}
\]
where \( h_{rot,f} \) is the rule-of-thumb bandwidth of Silverman (1986). Define next matrices \( Z_p = \{Z_{1,p}, \ldots, Z_{n,p}\}^T, p = 1, 2 \) with \( Z_{i,p} = \{Y_i - \hat{m}_p(X_i)\}^2 \) and
\[
X = \begin{pmatrix} 1 & \cdots & 1 \\ X_1 - x & \cdots & X_n - x \end{pmatrix}^T, \quad W = \text{diag} \left\{ \tilde{K} \left( \frac{X_i - x}{h_{rot,\sigma}} \right) \right\}_{i=1}^{n}, \tag{4.2}
\]
where \( h_{rot,\sigma} \) = the rule-of-thumb bandwidth of Fan and Gijbels (1996) based on data \((X_i, Z_{i,p})_{i=1}^n\). Then one defines the following estimators of \( \sigma^2(x) \)
\[
\hat{\sigma}^2_p(x) = (1, 0) \left( X^T W X \right)^{-1} X^T W Z_p, p = 1, 2, \tag{4.3}
\]
Bickel and Rosenblatt (1973), Fan and Gijbels (1996) provide the following uniform consistency results
\[
\frac{\max_{p=1,2} \sup_{x \in [a,b]} |\hat{\sigma}_p(x) - \sigma(x)| + \sup_{x \in [a,b]} \left| \hat{f}(x) - f(x) \right|}{\sigma(x)} = o_p(1). \tag{4.4}
\]

4.1 Implementing the exact band

The function \( \sigma_{n,1}(x) \) is approximated by either one of the following, with \( \hat{f}(x) \) and \( \hat{\sigma}_1(x) \) defined in (4.2) and (4.3), \( j(x) \) defined in (2.2)
\[
\hat{\sigma}_{n,1}(x, 1) = \hat{\sigma}_1 \left( t_{j(x)} \right) \hat{f}^{-1/2}(t_{j(x)}) n^{-1/2} h^{-1/2}, \tag{4.5}
\]
\[
\hat{\sigma}_{n,1}(x, 2) = \hat{\sigma}_1 \left( x \right) \hat{f}^{-1/2}(x) n^{-1/2} h^{-1/2}, \tag{4.6}
\]
where the additional parameter value 1 or 2 indicating the estimation at each value \( x \) or at the nearest left knot. Since \( \sup_{x \in [a,b]} |x - t_j(x)| \leq h \to 0 \), as \( n \to \infty \), (4.4) entails that both of the bands below are asymptotically exact with \( \hat{\sigma}_1(x) \) given in (4.1) and \( d_n \) in (2.13)

\[
\hat{\sigma}_1(x) = \sigma_{n,1}(x, \text{opt}) \{2 \log (N + 1)\}^{1/2} d_n, \text{opt} = 1, 2. \tag{4.7}
\]

### 4.2 Implementing the conservative band

According to Lemma B.3, for \( 0 \leq j \leq N \), the matrix \( \Xi_j \) approximates matrix \( S_j \) uniformly. Hence both of the bands below are asymptotically conservative, with \( \hat{\sigma}_2(x) \) given in (4.1)

\[
\hat{\sigma}_2(x) = \sigma_{n,2}(x, \text{opt}) \{2 \log (N + 1) - 2 \log \alpha\}^{1/2}, \text{opt} = 1, 2, \tag{4.8}
\]

where the function \( \sigma_{n,2}(x) \) in (2.11) for the linear band is estimated consistently by either one of the next two formulæ

\[
\hat{\sigma}_2(x, 1) = \left\{\Delta^T(x) \Xi_{j(x)} \Delta(x)\right\}^{1/2} \sqrt{3/2} \hat{\sigma}_2(t_j(x)) \hat{f}^{-1/2}(t_j(x)) n^{-1/2} h^{-1/2}, \tag{4.9}
\]

\[
\hat{\sigma}_2(x, 2) = \left\{\Delta^T(x) \Xi_{j(x)} \Delta(x)\right\}^{1/2} \sqrt{3/2} \hat{\sigma}_2(x) \hat{f}^{-1/2}(x) n^{-1/2} h^{-1/2}, \tag{4.10}
\]

with \( \Delta(x) \) and \( \Xi_j \) defined in (2.9), \( j(x) \) defined in (2.2), and \( \hat{f}(x) \) and \( \hat{\sigma}_2(x) \) defined in (4.2) and (4.3).

To calculate the matrix \( M_{N+2}^{-1} \) needed for (2.9), we need two theorems from matrix theory.

**Theorem 4.1** [Gantmacher and Krein (1960), page 95, equation (43)] For a symmetric Jacobi matrix \( J \) given as follows

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

its inverse matrix \( J^{-1} = (l_{ik})_{(N+2) \times (N+2)} \) satisfies

\[
l_{i,k} = \psi_i \chi_k, i \leq k, l_{i,i} = \psi_i \chi_i, k \leq i, \tag{4.11}
\]

where

\[
\psi_i = \frac{(-1)^i \det(J_{(1,\ldots,i-1)}) b_i b_{i+1} \cdots b_{N+1}}{\det(J)}, \chi_k = \frac{(-1)^k \det(J_{(k+1,\ldots,N+2)})}{b_kb_{k+1} \cdots b_{N+1}}, \tag{4.12}
\]

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 \( J \), while \( J_{(k+1,\ldots,N+2)} \) is the corresponding lower right \( (N+2-k) \times (N+2-k) \) submatrix.

**Theorem 4.2** [Zhang (1999), page 101, Theorem 4.5] For a tridiagonal matrix given as

\[
T_N = \begin{pmatrix}
a & b & 0 & & \\
c & a & \ddots & & \\
& \ddots & \ddots & b & \\
0 & c & a & \ddots & \\
& & & \ddots & \ddots
\end{pmatrix}_{N \times N}, \tag{4.13}
\]
if \( a^2 \neq 4bc \), then the determinant of \( T_N \) is

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

To apply Theorem 4.1 and Theorem 4.2, we let

\[
z_1 = \frac{2 + \sqrt{3}}{4}, \quad z_2 = \frac{2 - \sqrt{3}}{4}, \quad \theta = \frac{z_2}{z_1} = \left(2 - \sqrt{3}\right)^2 = 7 - 4\sqrt{3}. \tag{4.14}
\]

For any \( N \geq 1 \), Theorem 4.2 entails that

\[
\det (T_N) = \frac{\left(z_1^{N+1} - z_2^{N+1}\right)}{(z_1 - z_2)},
\]

if one takes \( a = 1, b = c = 1/4 \) in (4.13). Next, denote for any \( N \geq 1 \)

\[
\tilde{M}_{N+1} = \left( \begin{array}{cc} T_N & \tilde{T}_N \\ \tilde{T}_N & 1 \end{array} \right) \quad \text{and} \quad \tilde{T}_N = \left( \begin{array}{c} 0, \ldots, 0, \sqrt{2}/4 \end{array} \right)_{1 \times N}
\]

with the convention that \( \tilde{M}_1 \equiv 1 \). By the expansion of determinant of matrix \( \tilde{M}_i \) along the last row and then the last column,

\[
\det (\tilde{M}_i) = \det (T_{i-1}) - 8^{-1} \det (T_{i-2}) = \frac{8z_i^{-1}\{z_i(1 - \theta^i) - (1 - \theta^{i-1})\}}{8(z_1 - z_2)}, \forall i = 1, \ldots, N + 1.
\]

The determinant of matrix \( M_{N+2} \) can be expanded along the first row and then the first column:

\[
\det (M_{N+2}) = \det (\tilde{M}_{N+1}) - 8^{-1} \det (\tilde{M}_N)
\]

\[
= z_i^{N-1} \left\{64z_i^2 (1 - \theta^{N+1}) - 16z_i (1 - \theta^N) + (1 - \theta^{N-1})\right\} / \{64 (z_1 - z_2)\}.
\]

Applying (4.12) to matrix \( M_{N+2} \) yields

\[
\psi_k = \begin{cases} 
(1) \{(1/4)^{N-1}(\sqrt{2}/4)^2\} / \det (M_{N+2}), & i = 1, \\
(1)^i \{(1/4)^{N+1-i}(\sqrt{2}/4)^2\} / \det (\tilde{M}_{i-1}) / \det (M_{N+2}), & 2 \leq i \leq N,
\end{cases} \tag{4.15}
\]

\[
\chi_k = \begin{cases} 
(1) \{(1/4)^{N-1}(\sqrt{2}/4)^2\}^{-1} \det (\tilde{M}_{N+1}), & k = 1, \\
(1)^k \{(1/4)^{N+1-k}(\sqrt{2}/4)^2\}^{-1} \det (\tilde{M}_{(N+2)-k}), & 2 \leq k \leq N.
\end{cases} \tag{4.16}
\]

Next, we apply (4.11) from Theorem 4.1 together with (4.15) and (4.16), for all \( i, k = 1, \ldots, N + 2 \). Then the principle diagonal entries are

\[
l_{k,k} = \begin{cases} 
\det (\tilde{M}_{N+1}) / \det (M_{N+2}), & k = 1, N + 2 \\
\det (\tilde{M}_{(N+2)-k}) \det (\tilde{M}_{k-1}) / \det (M_{N+2}), & k = 2, \ldots, N + 1
\end{cases}
\]

which, after some algebra, becomes

\[
l_{11} = l_{N+2,N+2} = \frac{8z_i^2 (1 - \theta^{N+1}) - z_1 (1 - \theta^N)}{8z_i^2 (1 - \theta^{N+1}) - 2z_1 (1 - \theta^N) + 8 (1 - \theta^{N-1})},
\]

10
\[ l_{k,k} = \left\{ \frac{8z_1 (1 - \theta^{N+2-k}) - (1 - \theta^{N+1-k})}{(z_1 - z_2) \left( 64z_1^2 (1 - \theta^{N+1}) - 16z_1 (1 - \theta^N) + 64 (1 - \theta^{N-1}) \right)} \right\}, 2 \leq k \leq N + 1. \] (4.17)

Similarly, the upper diagonal entries are

\[ l_{i,i+1} = l_{i+1,i} = \left\{ \begin{array}{ll}
(\sqrt{2}/4) \det \left( \bar{M}_N \right) / \det (M_{N+2}), & i = 1, N + 1 \\
(-1/4) \det (M_{N+1-i})/ \det (M_{N+2}), & i = 2, \ldots, N
\end{array} \right. \]

which, by applying again (4.11), (4.15) and (4.16), becomes

\[ l_{12} = l_{N+1,N+2} = \frac{(-2\sqrt{2}) z_1 (1 - \theta^N) - (1 - \theta^{N-1})}{8z_1^2 (1 - \theta^{N+1}) - 2z_1 (1 - \theta^N) + 8 (1 - \theta^{N-1})}, \]

\[ l_{i,i+1} = \left\{ \frac{8z_1 (1 - \theta^{N+1-i}) - (1 - \theta^{N-i})}{(-4) (z_1 - z_2) \left( 64z_1^2 (1 - \theta^{N+1}) - 16z_1 (1 - \theta^N) + 64 (1 - \theta^{N-1}) \right)} \right\}, 2 \leq i \leq N. \] (4.18)

By the symmetry of 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 present some simulation results. The data set is generated from model (1), with

\[ m(x) = \sin(2\pi x), \sigma(x) = \sigma_0 \frac{100 - \exp(x)}{100 + \exp(x)}, X \sim U [-1/2, 1/2], \varepsilon \sim N(0,1). \] (5.1)

The noise level \( \sigma_0 = 0.2, 0.5 \) while sample size \( n = 100, 200, 500, 10000. \) Confidence level \( 1 - \alpha = 0.99, 0.95. \) Tables 1 and 2 contain the coverage probabilities as the percentage of coverage of the true curve at all data points by the confidence bands in (4.7) and (4.8), over 500 replications of sample size \( n. \) We have also computed the coverage probabilities of the confidence bands in (2.12) by plugging in the true value of density function \( f(x) = f_{[-1/2,1/2]}(x) \) and the variance function \( \sigma(x) \) in (5.1). These bands are called “oracle bands” as they use quantities that are unknown but for “oracles”; whereas the bands in (4.7) and (4.8) are called “estimated bands”, as they are computed directly from the data.

In Table 1 the surprising outcome is that all four bands have the same coverage with noise level 0.5. At noise level 0.2, the performance of all four bands becomes much closer with sample sizes increasing, whereas for small sample sizes the oracle bands are slightly better. In Table 2, the coverage percentages show very positive confirmation of Theorem 2. At sample size 200, regardless of noise level, both of the two candidate bands in (4.8) achieve at least 95.6% and 90% for confidence level \( 1 - \alpha = 0.99, 0.95 \) respectively.

From both tables, it is obvious that larger sample size guarantees improved coverage, with reasonable coverage achieved at moderate sample sizes. Under the same circumstances, the linear band performs much better than the constant band, which corroborates with the theory. The noise level has more influence to the constant bands than the linear ones.

For the linear bands, we have also carried out simulation for sample size \( n = 10000 \) and opt = 1. Regardless of the noise level, the coverage is always 99.4% for \( \alpha = 0.01 \) and 97.6% for \( \alpha = 0.05, \) both higher than the nominal coverage of 99% and 95%, consistent with their conservative definitions.
Remarkably, it takes merely 88 minutes to run 500 simulations with sample size as large as 10000 on a Pentium 4 PC. This is extremely fast considering that nonparametric regression is done without WARPing [Härdle, Hlávka and Klinke (2000)].

Four figures are created based on the first replicated 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 bands). In all figures, the confidence bands of \( n = 500 \) are thinner and fits better than those of \( n = 100 \). In Figures 1 and 2, the widths of the confidence bands based on are nearly the same, which is also true for Figures 3 and 4. Overall, linear bands are superior to constant ones in terms of smoothness and narrowness.

Since estimation of \( \sigma_{n,2}(x) \) by \( \hat{\sigma}_{n,2}(x,1) \) at knots as in (4.9) or by \( \hat{\sigma}_{n,2}(x,2) \) at all observations as in (4.10) does not seem to have any noticeable impact on the performance (i.e., the widths and the coverage probabilities of the confidence bands), we recommend always using estimation by \( \hat{\sigma}_{n,2}(x,1) \) at knots for simpler and faster implementation.

<table>
<thead>
<tr>
<th>noise level</th>
<th>sample size ( n )</th>
<th>confidence level ( 1 - \alpha )</th>
<th>estimated bands</th>
<th>oracle bands</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
<td>1 - 0.01</td>
<td>0.476 (0.458)</td>
<td>0.606 (0.606)</td>
</tr>
<tr>
<td>0.2</td>
<td></td>
<td>1 - 0.05</td>
<td>0.256 (0.246)</td>
<td>0.438 (0.436)</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>1 - 0.01</td>
<td>0.704 (0.708)</td>
<td>0.802 (0.802)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 - 0.05</td>
<td>0.454 (0.456)</td>
<td>0.532 (0.532)</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1 - 0.01</td>
<td>0.826 (0.834)</td>
<td>0.832 (0.832)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 - 0.05</td>
<td>0.462 (0.456)</td>
<td>0.468 (0.468)</td>
</tr>
<tr>
<td>0.5</td>
<td>100</td>
<td>1 - 0.01</td>
<td>0.618 (0.618)</td>
<td>0.618 (0.618)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 - 0.05</td>
<td>0.504 (0.504)</td>
<td>0.504 (0.504)</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>1 - 0.01</td>
<td>0.860 (0.860)</td>
<td>0.860 (0.860)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 - 0.05</td>
<td>0.716 (0.716)</td>
<td>0.716 (0.716)</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1 - 0.01</td>
<td>0.932 (0.932)</td>
<td>0.932 (0.932)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 - 0.05</td>
<td>0.802 (0.802)</td>
<td>0.802 (0.802)</td>
</tr>
</tbody>
</table>

Table 1: Constant bands coverage probabilities from 500 replications. The numbers outside/inside of the parentheses are based on estimating \( \sigma(X_i), i = 1, \ldots, n \) from (4.5)/(4.6).

<table>
<thead>
<tr>
<th>noise level</th>
<th>sample size ( n )</th>
<th>confidence level ( 1 - \alpha )</th>
<th>confidence level ( 1 - 0.01 )</th>
<th>confidence level ( 1 - 0.05 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>100</td>
<td>0.900 (0.896)</td>
<td>0.816 (0.814)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.956 (0.962)</td>
<td>0.902 (0.904)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.990 (0.988)</td>
<td>0.954 (0.958)</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>100</td>
<td>0.904 (0.904)</td>
<td>0.822 (0.814)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.956 (0.960)</td>
<td>0.900 (0.902)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.990 (0.988)</td>
<td>0.956 (0.960)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Linear spline bands coverage probabilities, in 500 replications. The numbers outside/inside of the parentheses are based on estimating \( \sigma(X_i), i = 1, \ldots, n \) from (4.9)/(4.10).
5.2 Motorcycle data

Although many authors such as H¨ ardle (1990) have used the motorcycle data to illustrate the need for nonparametric smoothing, no formal testing was done to establish the inadequacy of parametric model for this data. In this section we test the polynomial form of the motorcycle data regression curve. The null hypothesis is \( H_0 : m(x) = \sum_{k=1}^{d} a_k x^k \), with polynomial degree \( d = 4, 6, 8, 12 \). The variable \( Y \) is the accelerometer readings of a PTMO (post mortem human test object) over time in an experiment on the crash helmets, while the predictor \( X \) is the time in milliseconds after a simulated impact with motorcycles. Values of \( X \) and \( Y \) were recorded from 133 experiments. Applying the procedure to polynomials of other degrees has produced similar results.

In Figure 5, the center dotted line is the linear spline fit. The upper/lower thin lines represent linear bands based on Theorem 2, implemented according to (4.8). The solid line is the polynomial fit with degrees 4, 6, 8, 12. Clearly, the higher degree polynomial fits are better. The polynomial fit of degree 12 is nearly identical to the linear spline fit except for the first 15 milliseconds. Since all polynomial fits fall outside of the 0.999999 confidence bands, all polynomial null hypotheses have been rejected at significance level 0.000001.

6 Conclusions

We provide exact forms of two confidence bands constructed from polynomial spline regression. Asymptotic properties have been established for equally spaced, nonadaptive selection of knots. Extension to adaptive design is infeasible, as H¨ ardle, Marron and Yang (1997) had shown that adaptive knots selection could lead to inconsistency in \( L_{\infty} \) norm.

It is possible, however, to extend the constant spline 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 two 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 kept also for its simplicity. When implemented according to (4.7) with estimation on equally-spaced knots, the confidence limits at point \( x \) is the exact 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 piecewise linear band has to be calibrated at each new point \( x \). That is, the confidence limits at \( x \) and the ones at \( t_j(x) \) are different.

Extension to multivariate regression is difficult for lack of sharp approximation of the kind in (3.8). This limitation is also in Xia (1998), Claeskens and Van Keilegom (2003). The main hurdle of generalizing our method to higher order splines is the inversion of the inner product matrix of B-spline basis, for which close form solutions exist in the case of linear spline with the aid of (4.11) and (4.12). The inner product matrices of the two basis in (2.4) are diagonal and tridiagonal respectively, while for higher order splines it becomes multi-diagonal.

APPENDIX A: PROOF OF THEOREM 1

A.1 Preliminaries

Throughout Appendices A and B, we denote by the same letters \( c, C \), any positive constants, without distinction in each case.
**Lemma A.1** Under Assumptions (A3) and (A4), there exists a sequence \( \{D_n\} = \{n^{\alpha_0}\} \) for some \( \alpha_0 > 0 \), such that the following conditions are fulfilled

\[
\frac{D_n \log^2 n}{\sqrt{nh}} \to 0, \quad \sum_{n=1}^{\infty} D_n^{-2(2+\delta)} < \infty, \quad \frac{\sqrt{nh}}{D_n^{1+\delta}} \to 0, \quad \frac{1}{D_n^{2+\delta}} \to 0. \tag{A.1}
\]

**Proof.** By Assumption (A3), the four conditions can be expressed in terms of \( \alpha_0 \) as follows

\[
\alpha_0 < \frac{p}{2p+1}, \quad \alpha_0 > \frac{1}{2+\delta}, \quad \alpha_0 > \frac{p}{(2p+1)(1+\delta)}, \quad \alpha_0 > \frac{1}{2\delta (2p+1)}. \tag{A.2}
\]

Since Assumption (A4) (b) entails that \( \delta > 1/p \), one has

\[
\frac{1}{2+\delta} < \frac{p}{2p+1}, \quad \frac{p}{(2p+1)(1+\delta)} < \frac{p}{2p+1}. \quad \frac{1}{2\delta (2p+1)} < \frac{p}{2p+1}.
\]

There exists a \( \alpha_0 \) satisfying (A.2). Hence the lemma follows. \( \square \)

**Lemma A.2** For the sequence \( \{D_n\} \) that satisfies all four conditions in (A.1)

\[
P\{ \omega | \exists N(\omega), |\varepsilon_i| \leq D_n, i = 1, \ldots, n, n > N(\omega) \} = 1.
\]

**Proof.** Based on Lemma A.1

\[
\sum_{n=1}^{\infty} P(|\varepsilon_n| \geq D_n) \leq \sum_{n=1}^{\infty} \frac{E|\varepsilon|^{2+\delta}}{D_n^{1+\delta}} \leq M_{\delta} \sum_{n=1}^{\infty} D_n^{-(2+\delta)} < \infty.
\]

By Borel-Cantelli Lemma, one has

\[
P\{ \omega | \exists N_1(\omega), |\varepsilon_n| \leq D_n, \forall n > N_1(\omega) \} = 1.
\]

Since \( \{D_n\} = \{n^{\alpha_0}\} \) is increasing, the lemma follows. \( \square \)

**Lemma A.3** As \( n \to \infty \), for \( c_{j,n} \) and \( d_{j,n} \) defined in (2.3)

\[
c_{j,n} = f(t_j) h (1 + r_{j,n,1}), \quad \langle b_{j,1}, b_{j',1} \rangle = 0, j \neq j' \tag{A.3}
\]

\[
d_{j,n} = \frac{2}{3} f(t_{j+1}) h \begin{cases} 1 + r_{j,n,2} & j = 0, \ldots, N - 1, \\ 1/2 + r_{j,n,2} & j = -1, N, \end{cases} \tag{A.4}
\]

\[
\langle b_{j,2}, b_{j',2} \rangle = \frac{1}{6} f(t_{j+1}) h \begin{cases} 1 + \tilde{r}_{j,n,2} & |j' - j| = 1, \\ 0 & |j' - j| > 1, \end{cases} \tag{A.5}
\]

where

\[
\max_{0 \leq j \leq N} |r_{j,n,1}| + \max_{-1 \leq j \leq N} |r_{j,n,2}| + \max_{-1 \leq j \leq N-1} |\tilde{r}_{j,n,2}| \leq C \omega(f, h). \tag{A.6}
\]

In particular,

\[
\frac{1}{3} f(t_{j+1}) h \{1 - C \omega(f, h)\} \leq d_{j,n} \leq \frac{2}{3} f(t_{j+1}) h \{1 + C \omega(f, h)\}. \tag{A.7}
\]
PROOF. By the definition of $c_{j,n}$ in (2.3)

$$c_{j,n} = \int_{[t_j,t_{j+1}]} I_j(x) f(x) dx = \int_{[t_j,t_{j+1}]} f(x) dx = f(t_j) h + \int_{[t_j,t_{j+1}]} \{ f(x) - f(t_j) \} dx$$

hence for all $j = 0, \ldots, N$

$$|c_{j,n} - f(t_j) h| \leq \int_{[t_j,t_{j+1}]} |f(x) - f(t_j)| dx \leq \omega(f, h) h$$
or

$$|r_{j,n,1}| = |c_{j,n} - f(t_j) h| \{ f(t_j) h \}^{-1} \leq C \omega(f, h), j = 0, \ldots, N.$$  
Likewise, by the definition of $d_{j,n}$ in (2.3),

$$\langle b_{j,2}, b_{j',2} \rangle = \int_a^b K \left( \frac{x - t_{j+1}}{h} \right) K \left( \frac{x - t_{j'+1}}{h} \right) f(x) dx$$
leads to similar conclusions for $r_{j,n,2}, j = -1, \ldots, N$ and $r_{j,n,2}, j = -1, \ldots, N - 1.$ \hfill \square

PROOF OF LEMMA 3.1. For brevity, we give only the proof of (3.1) for $A_{n,1}$. Take any $j = 0, 1, \ldots, N$

$$\|B_{j,1}\|_{2,n}^2 - 1 = \left| \sum_{i=1}^n \xi_i |, \xi_i = \{ B_{j,1}^2(X_i) - 1 \} \right|^{n-1}$$
with $E\xi_i = 0$ and for any $k \geq 2$, Minkowski’s inequality implies that

$$E|\xi_i|^k = n^{-k} E|B_{j,1}^2(X_i) - 1|^k \leq (2/n)^k 2^{-1} E \left[ B_{j,1}^2(X_i) + 1 \right]$$

$$= (2/n)^k 2^{-1} \left\{ \int_{t_j}^{t_{j+1}} f(x) dx \right\} \{c_{j,n}\}^{-k} + 1 = (2/n)^k 2^{-1} \left\{ \{c_{j,n}\}^{-k} + 1 \right\} \leq \left\{ \frac{2}{nh} \right\}^k C_0 h,$$
while (A.3) entails that

$$E\xi_i^2 = n^{-2} E\|B_{j,1}^2(X_i) - 1\|^2 \geq n^{-2} E \left[ B_{j,1}^2(X_i) - 1 \right] = n^{-2} \left[ \frac{c_{j,n}}{2} - 1 \right] \geq \left\{ \frac{2}{nh} \right\}^2 C_1 h.$$  
It is then clear that one can find a constant $c > 0$ such that for all $k \geq 2, E|\xi_i|^k \leq (cn^{-1}h^{-1})^{k-2} k! E|\xi_i|^2$. Applying Bernstein’s inequality to $\sum_{i=1}^n \xi_i$, for any large enough $\delta > 0$

$$P \left\{ \sum_{i=1}^n \xi_i \geq \delta \sqrt{(nh)^{-1} \log(n)} \right\} \leq 2 \exp \left\{ \frac{-\delta^2 (nh)^{-1} \log(n)}{4n E\xi_i^2 + 2c (nh)^{-1} \delta \sqrt{(nh)^{-1} \log(n)}} \right\}$$

$$\leq 2 \exp \left\{ \frac{-\delta^2 (nh)^{-1} \log(n)}{4n \left(2(\log(n))^{-1} \right)^2 C_0 h + 2c (nh)^{-1} \delta \sqrt{(nh)^{-1} \log(n)}} \right\}$$

$$= 2 \exp \left\{ \frac{-\delta^2 \log(n)}{16C_0 + 2c \delta \sqrt{(nh)^{-1} \log(n)}} \right\} \leq 2n^{-3}.$$  
Hence

$$\sum_{n=1}^\infty P \left\{ \sup_{0 \leq j \leq N} \left| \|B_{j,1}\|_{2,n}^2 - 1 \right| \geq \delta \sqrt{(nh)^{-1} \log(n)} \right\} \leq \sum_{n=1}^\infty 2n^{-3} N \leq \sum_{n=1}^\infty 2n^{-2} < \infty$$
for such $\delta > 0$, then (3.1) follows. \hfill \square
A.2 Proof of Theorem 1

In this section, we will investigate the asymptotic behavior of \( \hat{\varepsilon}_1 (x) \) defined in (3.5). Since \( \langle B_{j,1} (x) , B_{j,1} (X) \rangle_n = 0 \) unless \( j = j' \), \( \hat{\varepsilon}_1 (x) \) can be written as

\[
\hat{\varepsilon}_1 (x) = \sum_{j=0}^{N} \varepsilon_j^* B_{j,1} (x) \| B_{j,1} \|_{2,n}^{-2}
\]

in which

\[
\varepsilon_j^* = \langle E , B_{j,1} (X) \rangle_n = \frac{1}{n} \sum_{i=1}^{n} B_{j,1} (X_i) \sigma (X_i) \varepsilon_i.
\]

**Lemma A.4** Let \( \hat{\varepsilon}_1 (x) = \sum_{j=0}^{N} \varepsilon_j^* B_{j,1} (x) , x \in [a, b] \) then

\[
| \hat{\varepsilon}_1 (x) - \hat{\varepsilon}_1 (x) | \leq A_{n,1} (1 - A_{n,1})^{-1} | \hat{\varepsilon}_1 (x) | , x \in [a, b],
\]

where \( A_{n,1} \) is defined in (3.1).

**Proof.** For any \( x \in [a, b] \),

\[
| \hat{\varepsilon}_1 (x) - \hat{\varepsilon}_1 (x) | \leq | \hat{\varepsilon}_1 (x) | \left\{ \sup_{0 \leq j \leq N} \left| \| B_{j,1} \|_{2,n}^{-2} - 1 \right| \sup_{0 \leq j \leq N} \left| \| B_{j,1} \|_{2,n}^{-2} \right| \right\}.
\]

Meanwhile (3.1) implies that

\[
\sup_{0 \leq j \leq N} \left| \| B_{j,1} \|_{2,n}^{-2} - 1 \right| \leq A_{n,1} (1 + A_{n,1})^{-1} \leq \sup_{0 \leq j \leq N} \| B_{j,1} \|_{2,n}^{-2} \leq (1 - A_{n,1})^{-1}
\]

hence the lemma follows. \( \square \)

The asymptotic behavior of \( \sup_{x \in [a,b]} | \hat{\varepsilon}_1 (x) | \) therefore is the same as that of \( \sup_{x \in [a,b]} | \hat{\varepsilon}_1 (x) | \).

**Lemma A.5** The pointwise variance of \( \hat{\varepsilon}_1 (x) \) is the function \( \sigma_{n,1}^2 (x) \) defined in (2.11) which satisfies

\[
E \{ \hat{\varepsilon}_1 (x) \}^2 \equiv \sigma_{n,1}^2 (x) = \frac{\sigma^2 (x)}{f (x) nh} \{ 1 + r_{n,1} (x) \} , x \in [a, b]
\]

with \( \sup_{x \in [a,b]} | r_{n,1} (x) | \rightarrow 0. \)

**Proof.** The term \( E \{ \hat{\varepsilon}_1 (x) \}^2 \) is

\[
E \left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=0}^{N} B_{j,1} (x) B_{j,1} (X_i) \sigma (X_i) \varepsilon_i \right\}^2 = \frac{1}{n} \int \left\{ B_{j(x),1} (x) B_{j(x),1} (v) \sigma (v) \right\}^2 f (v) dv
\]

\[
= \frac{1}{n} \int \left\{ \frac{1}{c_{j(x),n}} I_{j(x)} (v) \sigma (v) \right\}^2 f (v) dv = \frac{\int I_{j(x)} (v) \sigma^2 (v) f (v) dv}{n c_{j(x),n}^2}
\]

which is the expression for \( \sigma_{n,1}^2 (x) \) in (2.11). By (A.6) and the continuity of functions \( \sigma^2 (x) \) and \( f (x) \),

\[
\sigma_{n,1}^2 (x) = \frac{\sigma^2 (x) f (x) h + \int_{j(x)} \{ \sigma^2 (v) f (v) - \sigma^2 (x) f (x) \} dv}{n \left\{ f \left( j(x) \right) h + r_{j(x),n,1} \right\}^2}
\]

\[
= \frac{\sigma^2 (x) \{ 1 + r_{n,1} (x) \}}{n f (x) h}
\]

with \( \sup_{x \in [a,b]} | r_{n,1} (x) | \rightarrow 0, \) establishing (A.8). \( \square \)
Lemma A.6 Let the sequence \( \{D_n\} \) satisfy (A.1) and define for \( x \in [a, b] \)

\[
\hat{\xi}_{n,1} (x) = \sigma_{n,1} (x)^{-1} \sum_{j=0}^{N} B_{j,1} (x) \xi_{j}^* = \sigma_{n,1} (x)^{-1} \sum_{j=0}^{N} B_{j,1} (x) (\xi_{j}^* - E\xi_{j}^*) ,
\]

\[
\hat{\xi}_{n,1}^D (x) = \sigma_{n,1} (x)^{-1} \sum_{j=0}^{N} B_{j,1} (x) (\xi_{j}^* - E\xi_{j}^*) I_{||x| < D_n} \tag{A.9}
\]

then with probability 1

\[
\|\hat{\xi}_{n,1} (x) - \hat{\xi}_{n,1}^D (x)\|_\infty = O \left( D_n^{-(1+\delta)} \sqrt{nh} \right) = o (1) .
\]

Proof. Notice that \( E\xi_{j}^* = E \{ \frac{1}{n} \sum_{i=1}^{n} B_{j,1} (X_i) \sigma (X_i) \xi_i \} = 0 \) since \( E (\xi_i | X_i) = 0 \). So

\[
\hat{\xi}_{n,1} (x) = \left\{ \sigma_{n,1} (x)^{-1} B_{j,1} (x) \right\} \int \int B_{j,1} (v) \sigma (v) \varepsilon I_{|v| < D_n} \{ F_n (v, \varepsilon) - F (v, \varepsilon) \}
\]

according to the definition of \( Z_n (v, \varepsilon) \) in (3.10). The process \( \hat{\xi}_{n,1} (x) \) is separated into two parts \( \hat{\xi}_{n,1} (x) = \hat{\xi}_{n,1}^D (x) + \left\{ \hat{\xi}_{n,1} (x) - \hat{\xi}_{n,1}^D (x) \right\} \). The truncated part \( \hat{\xi}_{n,1}^D (x) \) is defined in (A.9). The tail part \( \hat{\xi}_{n,1} (x) - \hat{\xi}_{n,1}^D (x) \) is bounded uniformly over \( [a, b] \) by

\[
\sup_{x \in [a, b]} \left\| \sigma_{n,1} (x)^{-1} \int I_j (x) (v) \sigma (v) \varepsilon I_{|v| < D_n} \right\|_\infty \leq \sup_{x \in [a, b]} \left\| \sigma_{n,1} (x) c_j (x, n)^{-1} \int I_j (x) (v) \sigma (v) I_{|v| \geq D_n} \right\|_\infty \tag{A.10}
\]

\[
+ \sup_{x \in [a, b]} \left\| \sigma_{n,1} (x) c_j (x, n)^{-1} \int I_j (x) (v) \sigma (v) I_{|v| \geq D_n} \right\|_\infty \cdot \tag{A.11}
\]

By Lemma A.2, the term in (A.10) is 0 almost surely. The term in (A.11) is bounded by

\[
\sup_{x \in [a, b]} \left\| \sigma_{n,1} (x) c_j (x, n)^{-1} \int I_j (x) (v) \sigma (v) f (v) \right\|_\infty \int \left| I_j (v) \right| I_{|v| \geq D_n} (x) dF (v) \leq C \left( \frac{\sqrt{nh}}{h} \right) \frac{M_\delta}{D_n^{1+\delta}} \leq C \frac{\sqrt{nh}}{D_n^{1+\delta}}.
\]

The lemma follows immediately by the third condition in (A.1).

Lemma A.7 Define for \( x \in [a, b] \)

\[
\hat{\xi}_{n,1}^{(0)} (x) = \left\{ \sigma_{n,1} (x)^{-1} \int I_j (x) (v) \sigma (v) \varepsilon I_{|v| < D_n} dB \{ M (v, \varepsilon) \} \right\}
\]

then with probability 1

\[
\sup_{x \in [a, b]} \left| \hat{\xi}_{n,1}^{(0)} (x) - \hat{\xi}_{n,1}^D (x) \right| = O \left( h^{-1/2} n^{-1/2} D_n \log^2 n \right) = o (1) .
\]

17
Proof. First, $\sup_{x \in [a, b]} \| \hat{\varepsilon}^{(0)}_{n, 1}(x) - \hat{\varepsilon}^{D}_{n, 1}(x) \|$ can be written as

$$\sup_{x \in [a, b]} \left\{ \sigma_{n, 1}(x) \sqrt{n} c_{j}(x, n) \right\}^{-1} \int \int I_{j}(x) \sigma(v) \varepsilon I_{\{ |v| < D_n \}} d[Z_n(v, \varepsilon) - B\{ M(v, \varepsilon) \}]$$

which becomes the following via integration by parts

$$\sup_{x \in [a, b]} \left\{ \sigma_{n, 1}(x) \sqrt{n} c_{j}(x, n) \right\}^{-1} \int \int [Z_n(v, \varepsilon) - B\{ M(v, \varepsilon) \}] d\left\{ I_{j}(x) \sigma(v) \varepsilon I_{\{ |v| < D_n \}} \right\}$$

$$\leq \sup_{x \in [a, b]} \left\{ \sigma_{n, 1}(x) \sqrt{n} c_{j}(x, n) \right\}^{-1} \int \int [Z_n(v, \varepsilon) - B\{ M(v, \varepsilon) \}] d\left\{ \varepsilon I_{\{ |v| < D_n \}} \right\} d\left\{ I_{j}(x) \sigma(v) \right\}.$$
Lemma A.9 The process $\hat{\xi}_{n,1}^{(1)}(x)$ has the same probability structure as the process

$$\hat{\xi}_{n,1}^{(2)}(x) = \left\{ \sigma_{n,1}(x) \sqrt{n} c_{j(x),n} \right\}^{-1} \int I_{j(x)}(v) \sigma(v) s_{n}(v) f^{\frac{1}{2}}(v) dW(v), \ x \in [a,b]$$

where

$$s_{n}^{2}(v) = \int \varepsilon^{2} I_{[|\varepsilon|<D_{n}]} f_{\varepsilon|v}(\varepsilon) d\varepsilon.$$ (A.13)

Proof. By the definition of $\hat{\xi}_{n,1}^{(1)}(x)$ and $\hat{\xi}_{n,1}^{(2)}(x)$, applying Itô’s Isometry Theorem for any $x \in [a,b]$,

$$\begin{align*}
\text{var} \left\{ \hat{\xi}_{n,1}^{(1)}(x) \right\} & = \left\{ \sigma_{n,1}(x) \sqrt{n} c_{j(x),n} \right\}^{-2} E \left[ \int \int I_{j(x)}(v) \sigma(v) \varepsilon I_{[|\varepsilon|<D_{n}]} dW(M,v) \right]^{2} \\
& = \left\{ \sigma_{n,1}(x) \sqrt{n} c_{j(x),n} \right\}^{-2} \int \int I_{j(x)}(v) \sigma^{2}(v) \varepsilon^{2} I_{[|\varepsilon|<D_{n}]} f(v, \varepsilon) dv d\varepsilon \\
& = \left\{ \sigma_{n,1}(x) \sqrt{n} c_{j(x),n} \right\}^{-2} \int I_{j(x)}(v) \sigma^{2}(v) s_{n}^{2}(v) f(v) dv,
\end{align*}$$

$$\begin{align*}
\text{var} \left\{ \hat{\xi}_{n,1}^{(2)}(x) \right\} & = \left\{ \sigma_{n,1}(x) \sqrt{n} c_{j(x),n} \right\}^{-1} E \left\{ \int I_{j(x)}(v) \sigma(v) s_{n}(v) f^{\frac{1}{2}}(v) dW(v) \right\}^{2} \\
& = \left\{ \sigma_{n,1}(x) \sqrt{n} c_{j(x),n} \right\}^{-2} \int I_{j(x)}(v) \sigma^{2}(v) s_{n}^{2}(v) f(v) dv.
\end{align*}$$

Since the above two variances are equal for any $x \in [a,b]$, the two Gaussian processes $\hat{\xi}_{n}^{(1)}(x)$ and $\hat{\xi}_{n}^{(2)}(x)$ have the same probability structure.

Lemma A.10 Define for any $x \in [a,b]$,

$$\hat{\xi}_{n,1}^{(3)}(x) = \left\{ \sigma_{n,1}(x) \sqrt{n} c_{j(x),n} \right\}^{-1} \int I_{j(x)}(v) \sigma(v) f^{\frac{1}{2}}(v) dW(v)$$

then

$$\| \hat{\xi}_{n,1}^{(2)}(x) - \hat{\xi}_{n,1}^{(3)}(x) \|_{\infty} = O \left( \frac{1}{D_{n} \sqrt{h}} \right) = o(1) \ w. \ p. \ 1.$$ (A.1)

Proof. By the fourth condition in (A.1),

$$\begin{align*}
\sup_{x \in [a,b]} \left| \hat{\xi}_{n,1}^{(2)}(x) - \hat{\xi}_{n,1}^{(3)}(x) \right| &= \sup_{x \in [a,b]} \left\{ \sigma_{n,1}(x) \sqrt{n} c_{j(x),n} \right\}^{-1} \left| \int I_{j(x)}(v) \sigma(v) \{ s_{n}(v) - 1 \} f^{\frac{1}{2}}(v) dW(v) \right| \\
& \leq \sup_{x \in [a,b]} \left| s_{n}^{2}(v) - 1 \right| \left\{ \sigma_{n,1}(x) \sqrt{n} c_{j(x),n} \right\}^{-1} \sup_{x \in [a,b]} \left| \int I_{j(x)}(v) \sigma(v) f^{\frac{1}{2}}(v) dW(v) \right| \\
& \leq \frac{M_{\delta}}{D_{n} \sqrt{h}} \left| \int_{a}^{b} \sigma(v) f^{\frac{1}{2}}(v) dW(v) \right| = O \left( \frac{1}{D_{n} \sqrt{h}} \right) = o(1) \ w. \ p. \ 1.
\end{align*}$$

Lemma A.11 The process $\hat{\xi}_{n,1}^{(3)}(x)$ is a Gaussian process with mean 0, variance 1, and covariance

$$\text{cov} \left\{ \hat{\xi}_{n,1}^{(3)}(x), \hat{\xi}_{n,1}^{(3)}(y) \right\} = \delta_{j(x),j(y)}, \forall x, y \in [a,b].$$
PROOF. The variance and covariance are given by Itô’s Isometry Theorem
\[
\text{var}\left\{\tilde{\varepsilon}_{n,1}^{(3)}(x)\right\} = \left\{\sigma_{n,1}(x)\sqrt{n}c_{j(x),n}\right\}^{-2} E\left\{\int_{J_{j(x)}} I_{j(x)}(v) \sigma(v) f^{1/2}(v) dW(v)\right\}^2
\]
\[
= \left\{\sigma_{n,1}(x)\sqrt{n}c_{j(x),n}\right\}^{-2} \int_{J_{j(x)}} I_{j(x)}(v) \sigma^2(v) f(v) dv = 1
\]
according to (A.8). Likewise the covariance \(\text{cov}\left\{\tilde{\varepsilon}_{n,1}^{(3)}(x),\tilde{\varepsilon}_{n,1}^{(3)}(y)\right\}\) is
\[
\left\{\sigma_{n,1}(x)\sigma_{n,1}(y) n c_{j(x),n} c_{j(y),n}\right\}^{-1} E\left\{\int_{J_{j(x)}} \sigma(v) f^{1/2}(v) dW(v) \int_{J_{j(y)}} \sigma(v) f^{1/2}(v) dW(v)\right\}
\]
\[
= \left\{\sigma_{n,1}(x)\sigma_{n,1}(y) n c_{j(x),n} c_{j(y),n}\right\}^{-1} \int_{J_{j(x)} \cap J_{j(y)}} \sigma^2(v) f(v) dv = \delta_{j(x),j(y)},
\]
according to the definitions of \(\sigma_{n,1}(x)\) in (2.11) and of \(c_{i,n}\) in (2.3), thus completing the proof. □

PROOF OF PROPOSITION 3.1. The proof follows immediately from Lemmas A.4, A.6, A.7, A.8, A.9, A.10 and A.11. □

PROOF OF THEOREM 1. It is clear from Proposition 3.1 that the Gaussian process \(U(x)\) consists of \((N+1)\) i.i.d. standard normal variables \(U(t_0),\ldots,U(t_N)\), hence Theorem 3.4 implies that as \(n \to \infty\)
\[
P\left\{\sup_{x \in [a,b]} |U(x)| \leq \tau/a_{N+1} + b_{N+1}\right\} \to \exp\left(-2e^{-\tau}\right).
\]
By letting \(\tau = -\log\left(-\frac{1}{2} \log (1 - \alpha)\right)\), and using the definition of \(a_{N+1}\) and \(b_{N+1}\), we obtain
\[
\lim_{n \to \infty} P\left\{\sup_{x \in [a,b]} |U(x)| \leq -\sqrt{-\frac{1}{2} \log (1 - \alpha)} \{2 \log(N+1)\}^{-1/2}
\]
\[
+ \{2 \log(N+1)\}^{1/2} \frac{1}{2} \{2 \log(N+1)\}^{-1/2} \{\log log(N+1)+\log 4\pi\} = 1 - \alpha.
\]
Replacing \(U(x)\) with \(\sigma_{n,1}(x)^{-1} \tilde{\varepsilon}_1(x)\) (Proposition 3.1), and the definition of \(d_n\) in (2.13) entail that
\[
\lim_{n \to \infty} P\left\{\sup_{x \in [a,b]} |\sigma_{n,1}(x)^{-1} \tilde{\varepsilon}_1(x)| \leq \left\{2 \log(N+1)\right\}^{1/2} d_n\right\} = 1 - \alpha.
\]
According to (3.6), \(\|\tilde{m}_1(x) - m(x)\|_{\infty} = O_p(h)\), which implies that
\[
(nh)^{-1/2} \sqrt{\log(N+1)} \|\tilde{m}_1(x) - m(x)\|_{\infty} = O_p\left\{(nh)^{-1/2} \sqrt{\log(N+1)h}\right\} = o_p(1).
\]
Thus according to (3.4)
\[
\lim_{n \to \infty} P\left\{m(x) \in \tilde{m}_1(x) \pm \sigma_{n,1}(x) \left\{2 \log(N+1)\right\}^{1/2} d_n, \forall x \in [a,b]\right\}
\]
\[
= \lim_{n \to \infty} P\left\{\left\{2 \log(N+1)\right\}^{-1/2} d_n^{-1} \sup_{x \in [a,b]} |\sigma_{n,1}(x)^{-1} \tilde{\varepsilon}_1(x) + \tilde{m}_1(x) - m(x)| \leq 1\right\}
\]
\[
= \lim_{n \to \infty} P\left\{\left\{2 \log(N+1)\right\}^{-1/2} d_n^{-1} \sup_{x \in [a,b]} |\sigma_{n,1}(x)^{-1} \tilde{\varepsilon}_1(x)| \leq 1\right\} = 1 - \alpha. \quad \Box
\]
APPENDIX B: PROOF OF THEOREM 2

B.1 Preliminaries

In this subsection we examine some matrices used in the construction of confidence band in (2.14) and in the proof of Theorem 2.

The next lemma corresponds to (2.5) for piecewise constant basis. In what follows, we use $|T|$ to denote the maximal absolute value of any matrix $T$, and $M_{N+2}$ is the tridiagonal matrix as defined in (2.6).

**Lemma B.1** The inner product matrix $V$ of the B-spline basis $\{B_{j,2}(x)\}_{j=-1}^N$ defined in (2.6) has the following decomposition

$$V = M_{N+2} + (\tilde{v}_{j,j'})_{j,j'=-1}^N = M_{N+2} + \tilde{V}$$

where $\tilde{v}_{j,j'} \equiv 0$ if $|j - j'| > 1$, and

$$|\tilde{V}| \leq C\omega(f, h).$$  \hspace{1cm} (B.1)

**Proof.** By (A.4), (A.5) and (A.6), the inner product of $\langle b_{j',2}, b_{j,2} \rangle$ can be replaced by $\frac{1}{3} f(t_{j+1}) h$ if $|j' - j| = 1$, and $\frac{1}{3} f(t_{j+1}) h$ or $\frac{2}{3} f(t_{j+1}) h$ when $j' = j$. Plus some uniformly infinitesimal differences dominated by the moduli of continuity $\omega(f, h)$. Then based on the definition $B_{j,2}(x) = b_{j,2}(x) ||b_{j,2}||_2^{-1}$, the lemma follows immediately. \hfill \Box

The next lemma shows that multiplication by $M_{N+2}$ behaves similarly to multiplication by a constant.

**Lemma B.2** Given matrix $\Omega = M_{N+2} + \Gamma$, in which $\Gamma = (\gamma_{j,j'})_{j,j'=-1}^N$ satisfies $\gamma_{j,j'} \equiv 0$ if $|j - j'| > 1$ and $|\Gamma| \overset{P}{\to} 0$. Then there exist constants $c, C > 0$ independent of $n$ and $\Gamma$, such that with probability approaching one

$$c |\xi| \leq |\Omega\xi| \leq C |\xi|, C^{-1} |\xi| \leq |\Omega^{-1}\xi| \leq c^{-1} |\xi|, \forall \xi \in R^{N+2}. \hspace{1cm} (B.3)$$

**Proof.** Since each row of $M_{N+2}$ has diagonal element equal to 1, and one or two nonzero off-diagonal terms whose total absolute values do not exceed $2\sqrt{2}/4 = 1/\sqrt{2}$, hence

$$\left(1 - 1/\sqrt{2} - 3|\Gamma|\right) |\xi| \leq |\Omega\xi| \leq 3 (1 + |\Gamma|)|\xi|,$$

which entails the first half of (B.3) and the second half follows by switching the roles of $\xi$ and $\Omega\xi$. \hfill \Box

As an application of Lemma B.2, consider the matrix $S = V^{-1}$ defined in (2.7). Let $\tilde{\xi}_{j'} = \{\text{sgn}(s_{j,j'})\}_{j=-1}^N$, then there exists a positive $C_s$ such that

$$\sum_{j=-1}^N |s_{j,j'}| \leq |S\tilde{\xi}_{j'}| \leq C_s |\tilde{\xi}_{j'}| = C_s, \forall j' = -1, 0, ..., N. \hspace{1cm} (B.4)$$

The matrix $S$ appears in the construction of the confidence band, but it can not be computed exactly as it involves the unknown density $f(x)$. We approximate $S$ with the inverse of $M_{N+2}$, with a simpler, distribution-free form in (2.10). This approximation is uniform for $S_j$ in (2.7) and $\Xi_j$ (2.9) as well.
Lemma B.3 As \( n \to \infty, |M_{N+2}^{-1} - S| \to 0 \) and \( \max_{0 \leq j \leq N} |\Xi_j - S_j| \to 0 \).

Proof. By definition,
\[
M_{N+2}^{-1} = I = VS = \left(M_{N+2} + \tilde{V}\right)S.
\]
Denote by \( e_i \) the unit vector with \( i \)-th element 1, then applying Lemma B.2 with \( \Omega = M_{N+2} \), one derives
\[
c |M_{N+2}^{-1} - S| = c \max_{i=1}^{N+2} |(M_{N+2}^{-1} - S) e_i| \leq \max_{i=1}^{N+2} |M_{N+2} (M_{N+2}^{-1} - S) e_i| = |M_{N+2} (M_{N+2}^{-1} - S)|
\]
\[
= \|\tilde{V} S\| \leq \|\tilde{V}\| |S| \leq \|\tilde{V}\| \left(|M_{N+2}^{-1} - S| + |M_{N+2}^{-1}|\right).
\]
Since (B.2) makes \( |\tilde{V}| \leq C \omega(f, h) \), as \( n \to \infty \)
\[
|M_{N+2}^{-1} - S| \leq \frac{C \omega(f, h)}{c - C \omega(f, h)} |M_{N+2}^{-1}| = O \{ \omega(f, h) \} \to 0.
\]
Now by definition of submatrices \( S_j \) and \( \Xi_j \), \( \max_{0 \leq j \leq N} |\Xi_j - S_j| \leq |M_{N+2}^{-1} - S| \), the lemma follows. \( \square \)

B.2 Variance calculation

We now examine the asymptotic behavior of \( \text{Proj}_{G_n^{(0)}} E \), which is
\[
\tilde{e}_2(x) = \text{Proj}_{G_n^{(0)}} E = \sum_{j=-1}^{N} \tilde{a}_j B_{j,2}(x), x \in [a,b] \tag{B.5}
\]
where the spline coefficient vector \( \tilde{a} = (\tilde{a}_{-1}, ... , \tilde{a}_{N})^T \) are solutions to the normal equations
\[
\left(\langle B_{j,2}, B_{j',2}\rangle_n\right)_{j,j'=-1}^{N} \begin{pmatrix} \tilde{a}_{-1} \\ \vdots \\ \tilde{a}_{N} \end{pmatrix} = \left( \frac{1}{n} \sum_{i=1}^{n} B_{j,2}(X_{i}) \sigma(X_{i}) \epsilon_{i} \right)_{j=-1}^{N}.
\]
In other words
\[
\tilde{a} = \begin{pmatrix} \tilde{a}_{-1} \\ \vdots \\ \tilde{a}_{N} \end{pmatrix} = \left(V + \tilde{B}\right)^{-1} \left( \frac{1}{n} \sum_{i=1}^{n} B_{j,2}(X_{i}) \sigma(X_{i}) \epsilon_{i}\right)_{j=-1}^{N}, \tag{B.6}
\]
where \( \|\tilde{B}\| \leq \tilde{A}_{n,2} = O_p \left( \sqrt{n^{-1}h^{-1} \log(n)} \right) \) by (3.2).

Now define \( \hat{a}_j \)'s by replacing \( \left(V + \tilde{B}\right)^{-1} \) with \( V^{-1} = S \) in above formula, i.e.
\[
\hat{a} = \begin{pmatrix} \hat{a}_{-1} \\ \vdots \\ \hat{a}_{N} \end{pmatrix} = \left( \sum_{j=-1}^{N} s_{j,j} \frac{1}{n} \sum_{i=1}^{n} B_{j,2}(X_{i}) \sigma(X_{i}) \epsilon_{i}\right)_{j'=-1,..,N}. \tag{B.7}
\]
and define for \( x \in [a, b] \)

\[
\hat{\varepsilon}_2(x) = \sum_{j=-1}^{N} \hat{a}_j B_{j,2}(x) = \sum_{j,j'=-1}^{N} s_{j,j'} \frac{1}{n} \sum_{i=1}^{n} B_{j,2}(X_i) \sigma(X_i) \varepsilon_i B_{j',2}(x). \tag{B.8}
\]

In order to calculate the variance of \( \hat{\varepsilon}_2(x) \), we express the matrix \( \Sigma \) defined in (2.8) as

\[
\Sigma = \Theta_n V \Theta_n + (\tilde{\sigma}_j)_{j,j'=-1}^{N} = \Theta_n V \Theta_n + \tilde{\Sigma}, \Theta_n = \text{diag} \{ \sigma(t_0), \ldots, \sigma(t_{N+1}) \}, \tag{B.9}
\]

where

\[
\tilde{\sigma}_j \equiv 0 \text{ if } |j - j'| > 1, \quad \sup_{j,j'=-1}^{N} |\tilde{\sigma}_{jl}| \leq C \{ \omega(f,h) + \omega(f \sigma^2, h) \}. \tag{B.10}
\]

The next lemma is a special case of the unconditional version of Huang (2003), Remark 6.1, equation (6.2), page 1624.

**Lemma B.4** The pointwise variance of \( \hat{\varepsilon}_2(x) \) is the function \( \sigma^2_{n,2}(x) \) defined in (2.11), which satisfies

\[
E \left\{ \varepsilon^2_2(x) \right\} = \sigma^2_{n,2}(x) = \frac{3\sigma^2(x)}{2f(x)\Delta(x)} \sum_{j} S_j(x) \Delta(x) \{ 1 + r_{n,2}(x) \} \tag{B.11}
\]

with \( \sup_{x \in [a,b]} |r_{n,2}(x)| \to 0 \), \( j(x) \) is as defined in (2.2), \( \Delta(x) \) as defined in (2.9) and matrix \( S_j \) in (2.7). Consequently, there exist positive constants \( c_\sigma \) and \( C_\sigma \) such that for large enough \( n \)

\[
c_\sigma (nh)^{-1/2} \leq \sigma_{n,2}(x) \leq C_\sigma (nh)^{-1/2}, \forall x \in [a, b]. \tag{B.12}
\]

**Proof.** The first equality of (B.11) follows immediately from (B.8) and (2.8). Using (B.9), we have the decomposition

\[
\sigma^2_{n,2}(x) = I_n(x) + I_n(x) \tag{B.13}
\]

where

\[
I_n(x) = \frac{1}{n} \sum_{j,j'=-1}^{N} B_{j,2}(x) B_{j',2}(x) s_{j,j'} s_{ll} e_j^T \Theta_n V \Theta_n e_l
\]

\[
= \frac{1}{n} \sum_{j,j',l,l'=-1}^{N} B_{j,2}(x) B_{j',2}(x) s_{j,j'} s_{ll} v_{jl} \sigma(t_{l+1}) \sigma(t_{l+1}) \tag{B.14}
\]

\[
I_n(x) = \frac{1}{n} \sum_{j,j',l,l'=-1}^{N} B_{j,2}(x) B_{j',2}(x) s_{j,j'} s_{ll} \tilde{\sigma}_{jl}. \tag{B.15}
\]

According to (A.7), (B.15) is bounded by

\[
\frac{1}{n} \sum_{j,j'=j(x)-1}^{j(x)} \frac{\left| \sum_{j,l=-1}^{N} s_{j,j'} s_{ll} \tilde{\sigma}_{jl} \right|}{\sqrt{\left\{ \frac{1}{n} f(t_{l+1}) - C \omega(f,h) \right\} h \left\{ \frac{1}{n} f(t_{l'+1}) - C \omega(f,h) \right\} h}}.
\]

Hence, by applying (B.4) and (B.10), as \( n \to \infty \), for any \( x \in [a, b] \)

\[
|I_n(x)| \leq \sum_{j,j'=j(x)-1}^{j(x)} \left( \frac{1}{n} f(t_{j'+1}) - C \omega(f,h) \right)^{1/2} \left( \frac{1}{n} f(t_{l'+1}) - C \omega(f,h) \right)^{1/2} \tag{23}
\]

\[
\leq \frac{(nh)^{-1} C_\sigma^2 \| \Sigma \|}{\sqrt{\left\{ \frac{1}{n} f(t_{l+1}) - C \omega(f,h) \right\} h \left\{ \frac{1}{n} f(t_{l'+1}) - C \omega(f,h) \right\} h}}.
\]
\[ \leq C \left| \sum (nh)^{-1} = o \left( n^{-1}h^{-1} \right) \right. \tag{B.16} \]

Replacing \( \sigma (t_{j+1}) \sigma (t_{l+1}) \) with \( \sigma^2 (x) \) in (B.14) and define

\[ I_n^{(1)} (x) = \frac{1}{n} \sum_{j,j',l,l'=1}^N B_{j',2} (x) B_{l',2} (x) s_{jj'} s_{ll'} v_{jl} \sigma^2 (x) \tag{B.17} \]

then \( |I_n^{(1)} (x) - I_n (x)| \) can be bounded by

\[ \left| \frac{1}{2n} \sum_{j,j',l,l'=1}^N B_{j',2} (x) B_{l',2} (x) s_{jj'} s_{ll'} v_{jl} \{ \sigma (t_{j+1}) - \sigma (t_{l+1}) \} \right|^2 \tag{B.18} \]

\[ + \left| \frac{1}{n} \sum_{j,j',l,l'=1}^N B_{j',2} (x) B_{l',2} (x) s_{jj'} s_{ll'} v_{jl} \left\{ \frac{\sigma^2 (t_{j+1}) + \sigma^2 (t_{l+1})}{2} - \sigma^2 (x) \right\} \right|^2 . \tag{B.19} \]

Based on (B.4), we have

\[ \left| \sum_{j,l=1}^N s_{jj'} s_{ll'} v_{jl} \{ \sigma (t_{j+1}) - \sigma (t_{l+1}) \} \right|^2 \leq C_n^2 \omega^2 (\sigma, h), \forall x \in [a, b]. \tag{B.20} \]

Notice that

\[ \sum_{j,l=1}^N s_{jj'} s_{ll'} v_{jl} \sigma^2 (t_{j+1}) = \sum_{j,l=1}^N \sigma^2 (t_{j+1}) s_{jj'} s_{ll'} v_{jl} = \sum_{j,l=1}^N \sigma^2 (t_{j+1}) \delta_{j,l} \tag{B.21} \]

where \( \delta_{j,l} = 1 \), if \( j = l' \); 0 otherwise. By using (B.21), for any \( x \in [a, b] \), as \( n \to \infty \), term (B.19) can be represented as

\[ \leq \frac{1}{n} \sum_{j,j'=1}^N B_{j',2} (x) s_{jj'} \left| \frac{\sigma^2 (t_{j'+1}) - \sigma^2 (t_{j+1})}{2} + \sigma^2 (x) \right| \]

\[ \leq \frac{1}{n} \sum_{j,j'=1}^N B_{j',2} (x) s_{jj'} B_{l',2} (x) \omega (\sigma^2, h) \leq \frac{c}{nh} \omega (\sigma^2, h). \tag{B.22} \]

Hence by (B.20) and (B.22), it follows that as \( n \to \infty \)

\[ \left| I_n^{(1)} (x) - I_n (x) \right| = \frac{C}{nh} \omega^2 (\sigma, h) + \frac{c}{nh} \omega (\sigma^2, h), \forall x \in [a, b]. \tag{B.23} \]

Now \( I_n^{(1)} (x) \) is expanded specifically based on (B.21), the definition of \( b_{j,2} (x) \) and \( B_{j,2} (x) \) in Section 3,
Lemma B.5

\[ \frac{1}{n} \sigma^2(x) \left[ d_{j(x)-1,n}^{-1} \{1 - \delta(x)\}^2 s_{j(x)-1,j(x)-1} + d_{j(x),n}^{-1} \delta(x)^2 s_{j(x),j(x)} \right] + 2 \left\{ d_{j(x)-1,n} d_{j(x),n}(x) \right\}^{-1/2} \{1 - \delta(x)\} \delta(x) s_{j(x)-1,j(x)} \]

where \( \delta(x) \) is as defined in (2.2) and \( S_j \) is defined in (2.7). Define

\[ \eta_n(x) = \Delta^T(x) S_j(x) \Delta(x), \]

where \( \Delta(x) \) is defined in (2.9). Then by (A.4), one has

\[ \sup_{x \in [a,b]} \left\{ \frac{\sigma^2(x)}{2f(x) n h \eta_n(x)} \right\} \left\{ \frac{3 \sigma^2(x)}{2f(x) n h \eta_n(x)} \right\}^{-1} \to 0, \]

Together with (B.13), (B.16) and (B.23), it implies that

\[ \sup_{x \in [a,b]} \left\{ \frac{\sigma^2(x)}{2f(x) n h \eta_n(x)} \right\} \left\{ \frac{3 \sigma^2(x)}{2f(x) n h \eta_n(x)} \right\}^{-1} = \sup_{x \in [a,b]} |r_{n,2}(x)| \to 0, \]

hence we have finished the proof of equation (B.11). Together with (B.4) and the restriction on \( f(x) \) and \( \sigma^2(x) \) in Assumption (A2), (B.12) follows. \( \square \)

B.3 Proof of Theorem 2

Several lemmas will be given below for the proof of Proposition 3.2.

**Lemma B.5** Define for \( x \in [a,b] \)

\[ \check{\varepsilon}_{n,2}(x) = \sigma_{n,2}^{-1}(x) \check{\varepsilon}_2(x) = \sigma_{n,2}^{-1}(x) \sum_{j=-1}^{N} \hat{a}_{j} B_{j,2}(x), \]

\[ \check{\varepsilon}_{n,2}^{D}(x) = \sigma_{n,2}^{-1}(x) \sum_{j'=-1}^{N} \hat{a}_{j'} B_{j',2}(x) I_{|\varepsilon| < D_n}. \]  

(B.24)

where \( D_n \) satisfies (A.1). Then with probability 1

\[ \|\check{\varepsilon}_{n,2}(x) - \check{\varepsilon}_{n,2}^{D}(x)\|_\infty = O \left(n^{1/2} h^{1/2} D_n^{-1(1+\delta)}\right) = o(1). \]

**Proof.** Since obviously \( E\check{\varepsilon}_{n,2}(x) = 0, \forall x \in [a,b], \)

\[ \check{\varepsilon}_{n,2}(x) = \sigma_{n,2}^{-1}(x) n^{-1/2} \sum_{j'=-1}^{j(x)} B_{j',2}(x) \sum_{j=-1}^{N} s_{j'j} \int \int B_{j,2}(v) \sigma(v) \varepsilon dZ_n(v, \varepsilon) \]

where \( Z_n(x, \varepsilon) \) is defined in (3.10), hence \( \|\check{\varepsilon}_{n,2}(x) - \check{\varepsilon}_{n,2}^{D}(x)\|_\infty \) is bounded by

\[ \sup_{x \in [a,b]} \left\| \sum_{j'=-1}^{j(x)} \sigma_{n,2}^{-1}(x) B_{j',2}(x) \sum_{j=-1}^{N} s_{j'j} \frac{1}{n} \sum_{i=1}^{n} B_{j,2}(X_i) \sigma(X_i) \varepsilon_i I_{|\varepsilon_i| \geq D_n} \right\| \]  

(B.25)

\[ + \sup_{x \in [a,b]} \left\| \sum_{j'=-1}^{j(x)} \sigma_{n,2}^{-1}(x) B_{j',2}(x) \sum_{j=-1}^{N} s_{j'j} \int \int B_{j,2}(v) \sigma(v) \varepsilon I_{|\varepsilon| \geq D_n} dF(v, \varepsilon) \right\|. \]  

(B.26)
By Lemma A.2, the term in (B.25) is 0 with probability 1. The term in (B.26) is bounded by

\[
\sup_{x \in [a,b]} \left| \sum_{j'=j(x)-1}^{j(x)} \sigma_{n,2}^{-1} (x) B_{j',2} (x) \sum_{j=-1}^{N} s_{j'} s_{j} \int B_{j,2} (v) \sigma (v) \, dv \left( \int |\varepsilon| I_{|\varepsilon| \geq D_n} \right) dF (\varepsilon | v) \right|
\]

\[
\leq \sup_{x \in [a,b]} \sum_{j'=j(x)-1}^{j(x)} \left\{ \sigma_{n,2}^{-1} (x) B_{j',2} (x) \sum_{j=-1}^{N} |s_{j'} s_{j}| \int B_{j,2} (v) \sigma (v) \, dv \right\} D_n^{-(1+\delta)} M_4
\]

\[
\leq D_n^{-(1+\delta)} M_4 c_\sigma^{-1} (\varepsilon) \sup_{x \in [a,b]} \max_{j'=-1}^{j(x)-1} \sum_{j=-1}^{N} |s_{j'} s_{j}| \int B_{j,2} (v) \sigma (v) \, dv
\]

\[
\leq CD_n^{-(1+\delta)} (nh)^{1/2} h^{-1} \sup_{x \in [a,b]} \max_{j'=-1}^{N} \sum_{j=-1}^{N} |s_{j'} s_{j}| \int b_{j,2} (v) \sigma (v) \, dv.
\]

According to (B.4), the above term is bounded with probability 1 by

\[
CD_n^{-(1+\delta)} n^{1/2} C_s \max_{-1 \leq j \leq N} \left\{ \int b_{j,2} (v) \sigma (v) \, dv \right\} = O (\sqrt{nhD_n^{-(1+\delta)}}) = o (1). \tag*{\Box}
\]

**Lemma B.6** Let \( M \) be the Rosenblatt transformation given in (3.9) and define for \( x \in [a,b] \)

\[
\hat{\varepsilon}_{n,2}^{(0)} (x) = \left\{ \sqrt{n} \sigma_{n,2} (x) \right\}^{-1} \sum_{j'=-1}^{N} B_{j',2} (x) s_{j'} s_{j} \int \int B_{j,2} (v) \sigma (v) \varepsilon I_{|\varepsilon| < D_n} dB \{ M (v, \varepsilon) \}.
\]

Then with probability 1

\[
\sup_{x \in [a,b]} \left| \hat{\varepsilon}_{n,2}^{(0)} (x) - \hat{\varepsilon}_{n,2}^{D} (x) \right| = O (n^{-1/2} h^{-1/2} D_n \log^2 n) = o (1).
\]

**Proof.** Using integration by parts, and definition (B.24), \( \sup_{x \in [a,b]} \left| \hat{\varepsilon}_{n,2}^{(0)} (x) - \hat{\varepsilon}_{n,2}^{D} (x) \right| \) is

\[
\sup_{x \in [a,b]} \left| \frac{\sigma^{-1}_{n,2} (x)}{\sqrt{n}} \sum_{j'=-1}^{j(x)} B_{j',2} (x) s_{j'} s_{j} \int \int B_{j,2} (v) \sigma (v) \varepsilon I_{|\varepsilon| < D_n} d [Z_n (v, \varepsilon) - B \{ M (v, \varepsilon) \}] \right|
\]

\[
= \sup_{x \in [a,b]} \left| \frac{\sigma^{-1}_{n,2} (x)}{\sqrt{n}} \sum_{j'=-1}^{j(x)} B_{j',2} (x) \int \int [Z_n (v, \varepsilon) - B \{ M (v, \varepsilon) \}] \right|
\]

\[
\times | d \left\{ \sum_{j=-1}^{N} s_{j'} s_{j} B_{j,2} (v) \sigma (v) \right\} | d [\varepsilon I_{|\varepsilon| < D_n}] \right|
\]

\[
\leq Ch^{-1/2} \sup_{x, \varepsilon} |Z_n (x, \varepsilon) - B \{ M (x, \varepsilon) \}| \left| d \left\{ \sum_{j=-1}^{N} s_{j'} s_{j} b_{j,2} (v) \sigma (v) \right\} \right| d [\varepsilon I_{|\varepsilon| < D_n}] \right|
\]

which, by the bounded variation of \( \sigma (v) \) in Assumption (A2), and using (B.4), (3.11), with probability 1

\[
\leq CD_n h^{-1/2} \sup_{x, \varepsilon} |Z_n (x, \varepsilon) - B \{ M (x, \varepsilon) \}| = O \left( D_n n^{-1/2} h^{-1/2} \log^2 n \right) = o (1).
\]
Lemma B.7 Define for \( x \in [a, b] \)

\[
\hat{\xi}_{n,2}^{(1)} (x) = \frac{\sigma_{n,2}^{-1}(x)}{\sqrt{n}} \sum_{j' = -1}^{N} B_{j',2} (x) s_{j'} \int \int B_{j,2} (v) \sigma (v) \varepsilon I_{|\varepsilon| < D_n} dW \{ M (v, \varepsilon) \},
\]

then with probability 1

\[
\sup_{x \in [a, b]} \left| \xi_{n,2}^{(1)} (x) - \xi_{n,2}^{(0)} (x) \right| = O \left( h^{1/2} D_n^{-1/\delta} \right) \Rightarrow o (1).
\]

Proof. The left-hand side in the above equation is

\[
\left| \frac{\sigma_{n,2}^{-1}(x)}{\sqrt{n}} \sum_{j' = -1}^{N} B_{j',2} (x) s_{j'} \int \int B_{j,2} (v) \sigma (v) \varepsilon I_{|\varepsilon| < D_n} dM \{ M (v, \varepsilon) \} \right|
\]

\[
\leq \sup_{x \in [a, b]} \left| \frac{\sigma_{n,2}^{-1}(x)}{\sqrt{n}} \sum_{j' = -1}^{N} B_{j',2} (x) s_{j'} \int \int B_{j,2} (v) \sigma (v) \varepsilon \left| I_{|\varepsilon| < D_n} \right| dM (v, \varepsilon) W (1, 1) \right|
\]

\[
\leq \frac{C}{D_n^{1/\delta} h^{1/2}} \max_{x \in [a, b]} \sup_{j' = j(x) - 1} \left| \sum_{j' = -1}^{N} B_{j',2} (x) s_{j'} \int \int B_{j,2} (v) \sigma (v) f (v) dv \right| W (1, 1)
\]

\[
\leq \frac{C}{D_n^{1/\delta} h^{1/2}} \sup_{j' = j(x) - 1} \left| \sum_{j' = -1}^{N} B_{j',2} (x) s_{j'} \int \int B_{j,2} (v) \sigma (v) f (v) dv \right| W (1, 1)
\]

by using (B.4) and (A.1). \( \Box \)

Lemma B.8 The process \( \xi_{n,2}^{(1)} (x), x \in [a, b] \) has the same probability structure as

\[
\hat{\xi}_{n,2}^{(2)} (x) = \frac{\sigma_{n,2}^{-1}(x)}{\sqrt{n}} \sum_{j' = -1}^{N} B_{j',2} (x) s_{j'} \int \int B_{j,2} (v) \sigma (v) s_n (v) f^{1/2} (v) dW (v), x \in [a, b]
\]

where \( s_n^2 (v) \) is as defined in (A.13).

Proof. The variances of the two processes are

\[
\text{var} \left\{ \hat{\xi}_{n,2}^{(1)} (x) \right\} = \frac{\sigma_{n,2}^{-2}(x)}{n} \left( \sum_{j' = -1}^{N} B_{j',2} (x) s_{j'} \int \int B_{j,2} (v) \sigma (v) \varepsilon I_{|\varepsilon| < D_n} dW \{ M (v, \varepsilon) \} \right)^2
\]

\[
= \frac{\sigma_{n,2}^{-2}(x)}{n} \left( \sum_{j' = -1}^{N} B_{j',2} (x) B_{j,2} (x) s_{j'} s_{j'} \int \int B_{j,2} (v) B_{j,2} (v) \sigma^2 (v) \varepsilon^2 I_{|\varepsilon| < D_n} f (v, \varepsilon) dv d\varepsilon \right)
\]

\[
= \frac{\sigma_{n,2}^{-2}(x)}{n} \left( \sum_{j' = -1}^{N} B_{j',2} (x) B_{j,2} (x) s_{j'} s_{j'} \int \int B_{j,2} (v) B_{j,2} (v) \sigma^2 (v) s_n^2 (v) f (v) dv \right).
\]
\begin{align*}
\text{var}\left\{ \xi_{n,2}^{(2)}(x) \right\} &= \frac{\sigma_{n,2}^{-2}(x)}{n} \mathbb{E} \left\{ \sum_{j' = \argmax_{j < x}}^{j} B_{j'2}(x) \left[ \sum_{j = -1}^{N} s_{j'j} \int B_{j2}(v) \sigma(v) s_n(v) f^2(v) \, dv \right] \right\}^2 \\
&= \frac{\sigma_{n,2}^{-2}(x)}{n} \left\{ \sum_{j' = -1}^{N} \sum_{j = -1}^{N} B_{j'2}(x) B_{j2}(x) s_{j'j} s_{nl} \int B_{j2}(v) B_{l2}(v) \sigma^2(v) s_n^2(v) f(v) \, dv \right\}.
\end{align*}
\]

Since the two Gaussian processes have the same variances at any \( x \in [a, b] \), Itô’s Isometry Theorem ensures that they have the same probability structure.

**Lemma B.9** Define for any \( x \in [a, b] \)
\[ \xi_{n,2}^{(3)}(x) = \frac{\sigma_{n,2}^{-1}(x)}{\sqrt{n}} \sum_{j' = -1}^{N} B_{j'2}(x) s_{j'j} \int B_{j2}(v) \sigma(v) f^2(v) \, dv \]
then \( \text{var}\left\{ \xi_{n,2}^{(3)}(x) \right\} \equiv 1, \forall x \in [a, b], \) and with probability 1
\[ \left\| \xi_{n,2}^{(2)}(x) - \xi_{n,2}^{(3)}(x) \right\|_\infty = O \left( h^{-1/2} D_n^{-\delta} \right) = o(1). \]

**Proof.** Using (A.1) in the last step, the term \( \sup_{x \in [a, b]} \left| \xi_{n,2}^{(3)}(x) - \xi_{n,2}^{(3)}(x) \right| \) is bounded by
\[
\sup_{x \in [a, b]} \left( \frac{\sigma_{n,2}^{-1}(x)}{\sqrt{n}} \sum_{j' = -1}^{N} B_{j'2}(x) s_{j'j} \int B_{j2}(v) \sigma(v) |1 - s_n(v)| f^2(v) \, dv \right) \\
\leq \sup_{x \in [a, b]} \left( 1 - s_n^2(x) \right) \sup_{x \in [a, b]} \left( \frac{\sigma_{n,2}^{-1}(x)}{\sqrt{n}} \sum_{j' = -1}^{N} B_{j'2}(x) s_{j'j} \int B_{j2}(v) \sigma(v) f^2(v) \, dv \right) \\
\leq \sup_{x \in [a, b]} \left( \frac{\sigma_{n,2}^{-1}(x)}{\sqrt{n}} \sum_{j' = -1}^{N} B_{j'2}(x) s_{j'j} \int B_{j2}(v) \sigma(v) f^2(v) \, dv \right) \\
\leq M_\delta D_n^{-\delta} h^{1/2} \left| \int \sigma(v) f^2(v) \, dv \right| = O \left( h^{-1/2} D_n^{-\delta} \right) = o(1) \text{ w. p. 1.}
\]

Meanwhile, for any \( x \in [a, b] \)
\[
\text{var}\left\{ \xi_{n,2}^{(3)}(x) \right\} = \mathbb{E} \left( \frac{\sigma_{n,2}^{-1}(x)}{\sqrt{n}} \sum_{j' = -1}^{N} B_{j'2}(x) s_{j'j} \int B_{j2}(v) \sigma(v) f^2(v) \, dv \right)^2 \\
= \sigma_{n,2}^{-2}(x) \left\{ \frac{1}{n} \sum_{j, j' = -1}^{N} B_{j'2}(x) B_{j2}(x) s_{j'j} s_{ll} \int B_{j2}(v) B_{l2}(v) \sigma^2(v) f(v) \, dv \right\} \\
= \sigma_{n,2}^{-2}(x) \left\{ \frac{1}{n} \sum_{j, j' = -1}^{N} B_{j'2}(x) B_{j2}(x) s_{j'j} s_{ll} \sigma_{jl} \right\} = 1
\]
directly from (2.8) and (2.11).

Now define for any \( j' = -1, \ldots, N \) and \( x \in [a, b] \), the functions
\[ \zeta_{j'}(x) = n^{-1/2} \sigma_{n,2}^{-1}(x) B_{j'2}(x), \zeta_j(x) = (\zeta_{j(x)-1}(x), \zeta_j(x))^T \]
and the random vector $\mathbf{A} = (\Lambda_{-1}, \Lambda_0, \ldots, \Lambda_N)^T$ where

$$\Lambda_{j'} = \sum_{j=-1}^{N} s_{j'j} \int \int B_{j,2}(v) \sigma(v) f^{1/2}(v) dW(v).$$

Then $\mathbf{A} \sim \mathbf{N}(0, S \sum S)$ as $E \Lambda_{j'} = 0, \forall j' = -1, \ldots, N$, and the covariance is

$$E \Lambda_{j} \Lambda_{l'} = \sum_{j,l=-1}^{N} s_{j'j} s_{l'l} \int \int B_{j,2}(v) B_{l,2}(v) \sigma^2(v) f(v) dv = \sum_{j,l=-1}^{N} s_{j'j} s_{l'l}, \forall j', l' = -1, \ldots, N,$

in which $s_{j'l'}$ is defined in (2.8). Notice that

$$\hat{\varepsilon}_{n,2}^{(3)}(x) \equiv \sum_{j'=j(x)-1}^{j(x)} \zeta_{j'}(x) \Lambda_{j'} = \tilde{\zeta}(x)^T \mathbf{A}_{j(x)}, \mathbf{A}_j = (\Lambda_{j-1}, \Lambda_j)^T, j = 0, \ldots, N$$

and since Lemma B.9 states that the term $\hat{\varepsilon}_{n,2}^{(3)}(x)$ always has variance 1, it means that

$$\hat{\varepsilon}_{n,2}^{(3)}(x) = \frac{\tilde{\zeta}(x)^T \mathbf{A}_{j(x)}}{\sqrt{\tilde{\zeta}(x)^T \{\text{cov}(\mathbf{A}_{j(x)})\} \tilde{\zeta}(x)}}. \quad (B.27)$$

**Lemma B.10** For any given $0 < \alpha < 1$, one has

$$\liminf_{n \to \infty} P \left( \sup_{x \in [a,b]} |\hat{\varepsilon}_{n,2}^{(3)}(x)| \leq \left[ 2 \{ \log (N + 1) - \log \alpha \} \right]^{1/2} \right) \geq 1 - \alpha. \quad (B.28)$$

**Proof.** Define for any $j = 0, \ldots, N$

$$Q_j = \mathbf{A}_j^T \{\text{cov}(\mathbf{A}_j)\}^{-1} \mathbf{A}_j.$$ 

Result 4.7 (a), page 140 of Johnson and Wichern (1992) ensures that $Q_j$ is distributed as $\chi_2^2$ for any $j = 0, \ldots, N$, hence

$$P \left[ Q_j > \{ \log (N + 1) - \log \alpha \} \right] = \frac{\alpha}{N + 1}, \forall 0 \leq j \leq N.$$ 

Then (B.27) and the Maximization Lemma of Johnson and Wichern (1992), page 66 ensure that for any $x \in [a,b]$

$$\left\{ \hat{\varepsilon}_{n,2}^{(3)}(x) \right\}^2 = \frac{\left| \tilde{\zeta}(x)^T \mathbf{A}_{j(x)} \right|^2}{\tilde{\zeta}(x)^T \{\text{cov}(\mathbf{A}_{j(x)})\} \tilde{\zeta}(x)} \leq \mathbf{A}_{j(x)}^T \{\text{cov}(\mathbf{A}_{j(x)})\}^{-1} \mathbf{A}_{j(x)} = Q_{j(x)}.$$ 

One has therefore $\sup_{x \in [a,b]} |\hat{\varepsilon}_{n,2}^{(3)}(x)| \leq \max_{0 \leq j \leq N} \{Q_j\}$ and

$$P \left[ \sup_{x \in [a,b]} \left| \hat{\varepsilon}_{n,2}^{(3)}(x) \right| \leq \{ 2 \log (N + 1) - 2 \log \alpha \} \right] \geq 1 - (N + 1) \times \frac{\alpha}{N + 1} = 1 - \alpha. \quad (B.29)$$

Now (B.28) follows from Lemmas B.5, B.6, B.7, B.8, B.9, and (B.29).
Lemma B.11

\[
\left| \sup_{x \in [a,b]} \frac{\hat{\epsilon}_2(x)}{\sigma_{n,2}(x)} - \sup_{x \in [a,b]} \frac{\hat{\epsilon}_2(x)}{\sigma_{n,2}(x)} \right| = O_p \left( \sqrt{\frac{\log n}{nh}} \right) = o_p(1).
\]

Proof. Recall the definition for \( \hat{a} = (\hat{a}_{-1}, \hat{a}_0, ..., \hat{a}_N)^T \) and \( \hat{A} = (\hat{a}_{-1}, \hat{a}_0, ..., \hat{a}_N)^T \) in (B.6) and (B.7), one has \( V + \hat{B} = \hat{V} \hat{a} \). Based on Lemma B.2 and (3.2), there exists a constant \( c \) such that

\[
c | \hat{a} - \hat{a} | \leq | V (\hat{a} - \hat{a}) | = | B \hat{a} | \leq A_{n,2} (| \hat{a} - \hat{a} | + | \hat{a} |).
\]

or

\[
| \hat{a} - \hat{a} | \leq \frac{A_{n,2}}{c - A_{n,2}} | \hat{a} | . \tag{B.30}
\]

From the definitions of \( \hat{\epsilon}_2(x) \) in (B.5) and \( \hat{\epsilon}_2(x) \) in (B.8), plus (B.12), (B.30) and (A.7), as \( n \to \infty \)

\[
\sup_{x \in [a,b]} \left| \frac{\hat{\epsilon}_2(x)}{\sigma_{n,2}(x)} - \frac{\hat{\epsilon}_2(x)}{\sigma_{n,2}(x)} \right| \leq \sup_{x \in [a,b]} \sigma_{n,2}^{-1}(x) \sup_{x \in [a,b]} \left\{ \sum_{j=1}^{N} | \hat{a} - \hat{a} | B_{j,2}(x) \right\} \leq C_{n/2} \frac{A_{n,2}}{c - A_{n,2}} | \hat{a} | . \tag{B.31}
\]

Use (A.7) again, it implies that as \( n \to \infty \)

\[
\sup_{x \in [a,b]} \left| \frac{\hat{\epsilon}_2(x)}{\sigma_{n,2}(x)} \right| \geq C_{n/2} \sqrt{nh} \sup_{x \in [a,b]} \left\{ \sum_{j=1}^{N} \hat{a}_j B_{j,2}(x) \right\} = C_{n/2} \sqrt{nh} \sup_{x \in [a,b]} | \hat{a} \hat{B}_2^T(x) | \geq C_{n/2} \sqrt{nh} C_{n/2}^{-1} h^{-1/2} | \hat{a} | \geq C \sqrt{n} | \hat{a} | . \tag{B.32}
\]

where \( \hat{B}_2(x) = \{ B_{-1,2}(x), ..., B_{N,2}(x) \}^T \) and \( \hat{B}_2(x) = \{ b_{-1,2}(x), ..., b_{N,2}(x) \}^T \).

Then the desired result follows from (B.31) and (B.32), i.e.

\[
\sup_{x \in [a,b]} \left| \frac{\hat{\epsilon}_2(x)}{\sigma_{n,2}(x)} - \frac{\hat{\epsilon}_2(x)}{\sigma_{n,2}(x)} \right| \leq C_{n/2} \frac{A_{n,2}}{c - A_{n,2}} \left( \sqrt{n} \right)^{-1} \sup_{x \in [a,b]} \left| \frac{\hat{\epsilon}_2(x)}{\sigma_{n,2}(x)} \right| = O_p \left( \sqrt{\frac{\log n}{nh}} \right) = o_p(1). \qed
\]

Proof of Proposition 3.2. The desired result follows from Lemma B.10 and Lemma B.11 automatically. \qed

Proof of Theorem 2. Now (3.6) implies that \( \| \hat{m}_n(x) - m(x) \|_\infty = O_p \left( h^2 \right) \), and hence

\[
(nh)^{-1/2} \sqrt{\log (N + 1)} \| \hat{m}_n(x) - m(x) \|_\infty = O_p \left\{ (nh)^{-1/2} \sqrt{\log (N + 1) h^2} \right\} = o_p(1).
\]

Applying (3.7) in Proposition 3.2

\[
\liminf_{n \to \infty} P \left[ m(x) \in \hat{m}_n(x) \pm \sigma_{n,2}(x) \left\{ 2 \log (N + 1) - 2 \log \alpha \right\}^{1/2}, \forall x \in [a,b] \right] = \liminf_{n \to \infty} P \left[ \sup_{x \in [a,b]} \sigma_{n,2}^{-1}(x) | \hat{\epsilon}_2(x) + \hat{m}_n(x) - m(x) | \leq \left\{ 2 \log (N + 1) - 2 \log \alpha \right\}^{1/2} \right] \geq 1 - \alpha. \qed
\]
Acknowledgments. The authors thank three anonymous referees, an Associate Editor and the Editor for helpful comments.

REFERENCES


**JING WANG**  
**Department of Statistics & Probability**  
**Michigan State University**  
**East Lansing, MI 48824**  
**E-mail:** wangjin3@stt.msu.edu

**LIJIAN YANG**  
**Department of Statistics & Probability**  
**Michigan State University**  
**East Lansing, MI 48824**
Figure 1: Plots of confidence bands (thick solid curves), the piecewise constant estimator $\hat{m}_1(x)$ (dashed curve), the true function $m(x) = \sin(2\pi x)$ (thin solid curve), and the data scatter plots. The bands are computed from (4.7) with opt = 1.
Figure 2: Plots of confidence bands (thick solid curves), the piecewise constant estimator $\hat{m}_1(x)$ (dashed curve), the true function $m(x) = \sin(2\pi x)$ (thin solid curve), and the data scatter plots. The bands are computed from (4.7) with opt = 2.
Figure 3: Plots of confidence bands (thick solid curves), the piecewise linear 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 (4.8) with opt = 1.
Figure 4: Plots of confidence bands (thick solid curves), the piecewise linear 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 (4.8) with opt = 2.
Figure 5: Testing $H_0 : m(x) = \sum_{k=1}^{d} a_k x^k$, $d = 4, 6, 8, 12$ by linear confidence bands, the significance level is $\alpha = 0.000001$. 