Journal of Nonparametric Statistics
Vol. 21, No. 5, July 2009, 589–609

Spline confidence bands for variance functions

Qiongxia Song and Lijian Yang*

Michigan State University, East Lansing, MI 48824, USA

(*Received 24 March 2008; final version received 8 February 2009*)

Asymptotically exact and conservative confidence bands are obtained for possibly heteroscedastic variance functions, using piecewise constant and piecewise linear spline estimation, respectively. The variance estimation is as efficient as an infeasible estimator when the conditional mean function is known, and the widths of the confidence bands are of optimal order. Simulation experiments provide strong evidence that corroborates the asymptotic theory while the computing is extremely fast. A slower bootstrap band is also proposed, with much higher accuracy. As illustrations, the bootstrap spline band has been applied to test for heteroscedasticity in fossil data and in motorcycle data.

Keywords: B-spline; heteroscedasticity; infeasible estimator; knots; nonparametric regression; variance function

AMS Subject Classification: 62G08; 62G10; 62G15

1. Introduction

Quantification of local variability of regression data is an indispensable ingredient for many scientific investigations. The most intuitive measure of such is the conditional variance function, whose estimation has been the subject of Müller and Stadtmüller [1], Hall and Carroll [2], Ruppert et al. [3] and Fan and Yao [4], which employed kernel-type smoothing methods for the nonparametric variance function. Similar smoothing methods have also been used to estimate signal-to-noise ratio in [5], with applications to time-series volatility estimation. These existing works estimate the conditional variance function via kernel smoothing of the squares of residuals from an initial kernel smoothing of the regression data. Such a two-stage smoothing technique has also been used in estimating homoscedastic variance in [6]. More recently, a new approach to variance estimation based on differencing has been proposed, which can successfully handle serially correlated errors, see [7–9].

What has been lacking is a uniform confidence band for the whole variance curve over an entire bounded range and an explicit formula for the estimated variance function. The former is useful for making inference on the shape of the variance function, such as testing of homoscedasticity, while the latter is appealing to practitioners without much statistics expertise but wish to implement nonparametric procedures. Uniform confidence bands have been constructed for the conditional

*Corresponding author. Email: yang@stt.msu.edu

ISSN 1048-5252 print/ISSN 1029-0311 online

DOI: 10.1080/10485250902811151
http://www.informaworld.com

mean function in [10–13] and for probability density function in [14]. All these and other related works, such as that of Mack and Silverman [15], are based on kernel smoothing and make use of the 'Hungarian embedding' type results such as in [16,17]. More recently, Zhou et al. [18] and Wang and Yang [19] constructed confidence bands for the conditional mean function using the polynomial spline method with explicit formulae for both the estimated conditional mean function and the confidence band. In particular, Wang and Yang [19] allow for heteroscedastic and nonnormal errors and is useful for testing hypothesis on the shape of regression curve.

In this paper, we propose polynomial spline confidence bands for heteroscedastic variance function in a nonparametric regression model. The greatest advantages of polynomial spline estimation are its simplicity of implementation and fast computation, see for instance, [20,21] for the basic theory of polynomial spline smoothing and [22] for computing speed comparison of spline vs. kernel smoothing. Hence, it is desirable from a theoretical as well as a practical point of view to have confidence bands for polynomial spline estimators. 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 the regression model

$$Y = m(X) + \varepsilon, \tag{1}$$

where the error ε is conditional noise, with $E(\varepsilon|X) \equiv 0,\ E(\varepsilon^2|X) \equiv \sigma^2(X)$, see Assumption (A4) in Section 2. The conditional mean and conditional variance functions $m(x)$ and $\sigma^2(x)$, defined on interval $[a, b]$, need not be of any known form. Our goal is to construct a simultaneous confidence band for $\sigma^2(x)$ over $[a, b]$. In addition, the proposed variance estimator is asymptotically as efficient as the infeasible estimator, i.e., the asymptotic mean squared error is as small as if the conditional mean function $m(x)$ is given (equivalently, as if the unobservable error ε is actually observed). As an example, consider the motor cycle data. Figure 4 shows that with a p-value as small as 0.008, one rejects the null hypotheses that the conditional variance function of the data is a constant as no horizontal line can be squeezed into the 99.2% variance function confidence band. For other methods of testing the heteroscedasticity or the lack-of-fit of regression function, see [23,24] and Section 5 for simulation comparison of our method with that of Dette and Munk [23].

The paper is organised as follows. In Section 2, we state our main results on variance confidence bands using constant/linear splines. In Section 3, we investigate the error structure of spline variance estimators leading to insights of proof. We give the actual steps to implement the confidence band in Section 4, and in Section 5, we report simulation results and applications to a fossil data and the well-known motorcycle data. The Appendix contains all the technical proofs needed for the main results.

2. Main results

An asymptotic exact (conservative) $100(1-\alpha)\%$ confidence band for the unknown $\sigma^2(x)$ over the interval $[a, b]$ consists of an estimator $\hat{\sigma}^2(x)$ of $\sigma^2(x)$, lower and upper confidence limits $\hat{\sigma}^2(x) - l_{n,L}(x)$, $\hat{\sigma}^2(x) + l_{n,U}(x)$ at every $x \in [a, b]$ such that

$$\lim_{n\to\infty} P\{\sigma^2(x) \in [\hat{\sigma}^2(x) - l_{n,L}(x), \hat{\sigma}^2(x) + l_{n,U}(x)], \forall x \in [a, b]\} = 1 - \alpha, \quad \text{exact,}$$

$$\liminf_{n\to\infty} P\{\sigma^2(x) \in [\hat{\sigma}^2(x) - l_{n,L}(x), \hat{\sigma}^2(x) + l_{n,U}(x)], \forall x \in [a, b]\} \geq 1 - \alpha, \quad \text{conservative.}$$

If the mean function $m(x)$ were known, one could compute the errors $\varepsilon_i = Y_i - m(X_i)$, $1 \leq i \leq n$, and make use of the fact that $E(\varepsilon_i^2|X_i = x) \equiv \sigma^2(x)$ to carry out polynomial spline regression of the data $\{(X_i, Z_i)\}_{i=1}^n$, in which $Z_i = \varepsilon_i^2$ are the squared errors. Specifically, one could define the 'infeasible estimator' of the variance function as

$$\widetilde{\sigma}^2_{p_2}(x) = \underset{g \in G^{(p_2-2)}_{N_2}[a,b]}{\text{argmin}} \sum_{i=1}^{n} \{Z_i - g(X_i)\}^2,$$

in which $G^{(p_2-2)}_{N_2} = G^{(p_2-2)}_{N_2}[a, b]$ is the space of functions that are piecewise polynomials of degree $(p_2 - 1)$ on interval $[a, b]$, defined precisely below, for some positive integer p_2. To mimic the above unattainable spline smoother, we define

$$\hat{\sigma}^2_{p_1,p_2}(x) = \underset{g \in G^{(p_2-2)}_{N_2}[a,b]}{\text{argmin}} \sum_{i=1}^{n} \{\hat{Z}_{i,p_1} - g(X_i)\}^2, \quad (2)$$

where $\hat{Z}_{i,p_1} = \hat{\varepsilon}^2_{i,p_1}$ are the squares of residuals $\hat{\varepsilon}_{i,p_1}$ obtained from spline regression,

$$\hat{\varepsilon}_{i,p_1} = Y_i - \hat{m}_{p_1}(X_i), \quad 1 \le i \le n, \quad (3)$$

for some positive integer p_1, in which

$$\hat{m}_{p_1}(x) = \underset{g \in G^{(p_1-2)}_{N_1}[a,b]}{\text{argmin}} \sum_{i=1}^{n} \{Y_i - g(X_i)\}^2. \quad (4)$$

To introduce spline functions, for the two steps $\nu = 1, 2$, we divide the finite interval $[a, b]$ into $(N_\nu + 1)$ subintervals $J_j = [t_j, t_{j+1}),\ \ j = 0, \ldots, N_\nu - 1,\ \ J_{N_\nu} = [t_{N_\nu}, b]$. A sequence of equally spaced interior knots $\{t_j\}_{j=1}^{N_\nu}$ are given as

$$t_0 = a < t_1 < \cdots < t_{N_\nu} < b = t_{N_\nu+1}, \quad t_j = a + jh_\nu, \quad j = 0, 1, \ldots, N_\nu + 1,$$

in which $h_\nu = (b - a)/(N_\nu + 1)$ is the distance between neighboring knots. We denote by $G^{(p_\nu-2)}_{N_\nu} = G^{(p_\nu-2)}_{N_\nu}[a, b]$ the space of functions that are polynomials of degree $(p_\nu - 1)$ on each J_j and have continuous $(p_\nu - 2)$ th derivative. For example, $G^{-1}_{N_\nu}$ denotes the space of functions that are constant on each J_j, and $G^0_{N_\nu}$ 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 $[a, b]$, i.e., $\|w\|_\infty = \sup_{x\in[a,b]} |w(x)|$, and the moduli of continuity of a continuous function w on $[a, b]$ is denoted by $\omega(w, h_\nu) = \max_{x,x'\in[a,b],|x-x'|\le h_\nu} |w(x) - w(x')|$. That $\lim_{h_\nu\to 0} \omega(w, h_\nu) = 0$ follows from the uniform continuity of w on compact $[a, b]$.

Our approach is to construct the error bound function $l_n(x)$ around the spline estimators $\hat{\sigma}^2_{p_1,p_2}(x)$. The technical assumptions we need are as follows:

(A1) The regression function $m(\cdot) \in C^{(p_1)}[a, b]$.
(A2) The density function $f(\cdot)$ of X is continuous and positive on the interval $[a, b]$.
(A3) The subinterval length $h_\nu \sim n^{-1/(2p_\nu+1)}$, i.e., the number of interior knots $N_\nu \sim n^{1/(2p_\nu+1)}$, $\nu = 1, 2$.
(A4) The joint distribution $F(x, \varepsilon)$ of random variables (X, ε) satisfies:
  (a) There exists a positive value $\eta > 1/p_2$ and a finite positive M_η such that

$$\sup_{x\in[a,b]} E(|\varepsilon|^{4+2\eta}|X = x) < M_\eta.$$

(b) The error is conditional noise: $E(\varepsilon|X=x)\equiv 0$, $E(\varepsilon^2|X=x)\equiv\sigma^2(x)$ with $E(\varepsilon^4|X=x)\equiv\mu_4(x)$, which is a positive function on $[a,b]$ with bounded variation. The variance function $\sigma^2(\cdot)\in C^{(p_2)}[a,b]$ and has a positive lower bound on $[a,b]$.

Assumptions (A1)–(A4) are adapted from Wang and Yang [19] for sample $\{(X_i,Z_i)\}_{i=1}^n$. In particular, Assumption (A4) (a) implies that $\mathrm{var}(\varepsilon^2|X=x)\equiv\mu_4(x)-\sigma^4(x)$ is the conditional variance of $Z=\varepsilon^2$, denoted as $v_Z^2(x)$. We denote also $p_*=\min(p_1,p_2)$, $p^*=\max(p_1,p_2)$, $N_*=\min(N_1,N_2)$, $N^*=\max(N_1,N_2)$. The idea of allowing different degrees of smoothness for m and σ comes from one referee.

To properly define the confidence bands, we denote for any $x\in[a,b]$, define its location and relative position indices $j_\nu(x), r_\nu(x)$ as

$$j_\nu(x)=j_{n\nu}(x)=\min\left\{\left[\frac{(x-a)}{h_\nu}\right],N_\nu\right\},\quad r_\nu(x)=\frac{\{x-t_{j_\nu(x)}\}}{h_\nu}. \tag{5}$$

Since any x is between two consecutive knots, it is clear that $t_{j_{n\nu}(x)}\le x<t_{j_{n\nu}(x)+1}$, $0\le r_\nu(x)<1$, $\forall x\in[a,b)$, and $r_\nu(b)=1$. We denote by $\|\phi\|_2$ the theoretical L^2 norm of a function ϕ on $[a,b]$, i.e., $\|\phi\|_2^2=E\{\phi^2(X)\}=\int_a^b\phi^2(x)f(x)\mathrm{d}x$, 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

$$\langle\phi,\varphi\rangle=\int_a^b\phi(x)\varphi(x)f(x)\mathrm{d}x=E\{\phi(X)\varphi(X)\},$$

$$\langle\phi,\varphi\rangle_n=n^{-1}\sum_{i=1}^n\phi(X_i)\varphi(X_i)$$

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

Algebra shows that the space $G_{N_\nu}^{(p_\nu-2)}$ can be spanned linearly by the B-spline basis introduced below or the truncated power basis introduced in Section 4, see [25]. Hence, the same estimator $\hat m_{p_1}(x)$ can be expressed as a linear combination of either of the two bases. 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 $G_{N_\nu}^{(-1)}$, the space of piecewise constant splines, are indicator functions of intervals J_j, $b_{j,1}(x)=I_j(x)=I_{J_j}(x)$, $0\le j\le N_\nu$. The B-spline basis of $G_{N_\nu}^0$, the space of piecewise linear splines, are $\{b_{j,2}(x)\}_{j=-1}^{N_\nu}$, where

$$b_{j,2}(x)=K\left(\frac{x-t_{j+1}}{h_1}\right),\quad j=-1,0,\ldots,N_\nu,\quad \text{for } K(u)=(1-|u|)_+.$$

Define the rescaled B-spline basis $\{B_{j,p_\nu}(x)\}_{j=1-p_\nu}^{N_\nu}$ for $G_{N_\nu}^{(p_\nu-2)}$

$$B_{j,p_\nu}(x)\equiv b_{j,p_\nu}(x)\|b_{j,p_\nu}\|_2^{-1},\quad 1-p_\nu\le j\le N_\nu.$$

Obviously, all the rescaled basis functions will have theoretical norm 1.

To express the estimator $\hat m_{p_1}(x)$ based on the basis $\{B_{j,p_1}(x)\}_{j=1-p_1}^{N_1}$, we introduce the following vectors in R^n: $\mathbf{Y}=(Y_1,\ldots,Y_n)^{\mathrm{T}}$,

$$\mathbf{B}_{j,p_1}(\mathbf{X})=\{B_{j,p_1}(X_1),\ldots,B_{j,p_1}(X_n)\}^{\mathrm{T}},\quad j=1-p_1,\ldots,N_1,$$

and let the design matrix for spline regression be

$$\mathbf{B}_{p_1}=\mathbf{B}_{p_1}(\mathbf{X})=\{\mathbf{B}_{1-p_1,p_1}(\mathbf{X}),\ldots,\mathbf{B}_{N_1,p_1}(\mathbf{X})\}.$$

Then the estimator $\hat{m}_{p_1}(x)$ in Equation (4) is expressed as

$$\hat{m}_{p_1}(x) = \{B_{1-p_1,p_1}(x), \ldots, B_{N_1,p_1}(x)\}(\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{B}_{p_1})^{-1}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{Y} = \sum_{j=1-p_1}^{N_1} \hat{\lambda}_{j,p_1} B_{j,p_1}(x),$$

where the coefficients $\{\hat{\lambda}_{1-p_1,p_1}, \ldots, \hat{\lambda}_{N_1,p_1}\}^{\mathrm{T}}$ are solutions of the following least-squares problem

$$\{\hat{\lambda}_{1-p_1,p_1}, \ldots, \hat{\lambda}_{N_1,p_1}\}^{\mathrm{T}} = \underset{R^{N_1+p_1}}{\operatorname{argmin}} \sum_{i=1}^{n} \left\{ Y_i - \sum_{j=1-p_1}^{N_1} \lambda_{j,p_1} B_{j,p_1}(X_i) \right\}^2,$$

or equivalently, of the normal equation

$$(\langle B_{j,p_1}, B_{j',p_1}\rangle_n)_{j,j'=1-p_1}^{N_1} (\hat{\lambda}_{j,p_1})_{j=1-p_1}^{N_1} = \left(n^{-1} \sum_{i=1}^{n} B_{j,p_1}(X_i) Y_i \right)_{j=1-p_1}^{N_1}.$$

It is straightforward that $\langle B_{j,p_1}, B_{j',p_1}\rangle \equiv 0$, $|j - j'| \geq p_1$, thus the inner product matrix on the left side of the normal equation is diagonal for the constant B-spline basis ($p_1 = 1$), and tridiagonal for the linear B-spline basis ($p_1 = 2$). According to Lemma A.2, it is approximated by its deterministic version, whose inverse has an explicit formula given in [19].

For $p_2 = 2$, define the inverse of inner product matrix as S with its 2×2 diagonal submatrices $\{S_j, 0 \leq j \leq N_2\}$

$$S = (s_{j,j'})_{j,j'=-1}^{N_2} = (\langle B_{j,2}, B_{j',2}\rangle)^{-1}, \quad S_j = \begin{pmatrix} s_{j-1,j-1} & s_{j-1,j} \\ s_{j,j-1} & s_{j,j} \end{pmatrix}. \tag{6}$$

The widths of the confidence bands depend on the variance function:

$$v_{n,1}^2(x) = \frac{\int_{I_{j(x)}} v_Z^2(v) f(v)\,\mathrm{d}v}{n\|b_{j(x),1}\|_2^2}, \quad v_{n,2}^2(x) = \sum_{j,j',l,l'=-1}^{N_2} \frac{B_{j',2}(x) B_{l',2}(x) s_{jj'} s_{ll'} v_{jl}}{n}, \tag{7}$$

with $j(x)$ defined in Equation (5), and $s_{ll'}$ in Equation (6), and

$$(v_{jl})_{j,j'=-1}^{N_2} = \Sigma = \left\{ \int v_Z^2(v) B_{j,2}(v) B_{l,2}(v) f(v)\,\mathrm{d}v \right\}_{j,j'=-1}^{N_2}.$$

Under Assumptions (A1)–(A4), applying Theorems 1 and 2 from Wang and Yang [19] to the unobserved sample $\{(X_i, Z_i)\}_{i=1}^n$, an asymptotic $100(1-\alpha)\%$ exact confidence band for $\sigma^2(x)$ over $[a, b]$ is

$$\widetilde{\sigma}_1^2(x) \pm v_{n,1}(x)\{2\log(N_2+1)\}^{1/2} d_n(\alpha),$$

where $v_{n,1}(x)$ is given in Equation (7) and replaceable by $\hat{v}_Z(x)\{\hat{f}(x)nh_1\}^{-1/2}$ and

$$d_n = 1 - \{2\log(N_2+1)\}^{-1} \left[\log\left\{ -\frac{\log(1-\alpha)}{2} \right\} + \frac{\log\log(N_2+1) + \log 4\pi}{2} \right], \tag{8}$$

and an asymptotic $100(1-\alpha)\%$ conservative confidence band for $\sigma^2(x)$ over $[a, b]$ is

$$\widetilde{\sigma}_2^2(x) \pm v_{n,2}(x)\{2\log(N_2+1) - 2\log\alpha\}^{1/2},$$

where $v_{n,2}(x)$ is as in Equation (7), replaceable by $\hat{v}_{n,2}(x)$ in Equation (16).

We state our main results in the next theorems.

THEOREM 2.1 *Under Assumptions* (A1)–(A4)*, as* $n \to \infty$*, the spline estimator* $\hat{\sigma}^2_{p_1,p_2}$ *of* σ^2 *is asymptotically as efficient as 'infeasible estimator', i.e.,*

$$\|\hat{\sigma}^2_{p_1,p_2} - \widetilde{\sigma}^2_{p_2}\|_\infty = \sup_{x\in[a,b]} |\hat{\sigma}^2_{p_1,p_2}(x) - \widetilde{\sigma}^2_{p_2}(x)| = o_p(n^{-p_1/(2p_1+1)}).$$

Theorem 2.1 and the aforementioned properties of $\hat{\sigma}^2_{p_1,p_2}$ imply the following:

THEOREM 2.2 *Under Assumptions* (A1)–(A4)*, for* $p_2 = 1$ *or* 2*, an asymptotic* $100(1-\alpha)\%$ *exact or a conservative confidence band for* $\sigma^2(x)$ *over the interval* $[a, b]$ *is*

$$\hat{\sigma}^2_{1,1}(x) \pm v_{n,1}(x)\{2\log(N_2+1)\}^{1/2} d_n(\alpha),$$
$$\hat{\sigma}^2_{2,2}(x) \pm v_{n,2}(x)\{2\log(N_2+1) - 2\log\alpha\}^{1/2},$$

respectively. That is,

$$\lim_{n\to\infty} P\{\sigma^2(x) \in \hat{\sigma}^2_{1,1}(x) \pm v_{n,1}(x)\{2\log(N_2+1)\}^{1/2} d_n(\alpha), \forall x \in [a,b]\} = 1-\alpha,$$
$$\liminf_{n\to\infty} P\{\sigma^2(x) \in \hat{\sigma}^2_{2,2}(x) \pm v_{n,2}(x)\{2\log(N_2+1) - 2\log\alpha\}^{1/2}, \forall x \in [a,b]\} \geq 1-\alpha.$$

The proof of Theorem 2.1 and therefore also of Theorem 2.2 depend on Propositions A.1, A.2, and A.3 in the next section, whose proofs are given in the Appendix.

3. Error decomposition

In this section, we break the estimation error $\hat{\sigma}^2_{p_2,p_1}(x) - \widetilde{\sigma}^2_{p_2}(x)$ into three parts. To understand this decomposition, we begin by discussing the spline space $G^{(p_1-2)}$ and the representation of the spline estimators $\hat{m}_{p_1}(x)$ in Equation (4) and $\hat{\sigma}^2_{p_2,p_1}(x)$ in Equation (2).

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

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

Projecting the response $\mathbf{Y}$ onto the linear space $G_n^{(p_1-2)}$ spanned by $\{\mathbf{B}_{j,p_1}(\mathbf{X})\}_{j=1-p_1}^{N_1}$, one gets

$$\hat{\mathbf{m}}_{p_1} = \{\hat{m}_{p_1}(X_1), \ldots, \hat{m}_{p_1}(X_n)\}^{\mathrm{T}} = \operatorname*{Proj}_{G_n^{(p_1-2)}} \mathbf{Y} = \operatorname*{Proj}_{G_n^{(p_1-2)}} \mathbf{m} + \operatorname*{Proj}_{G_n^{(p_1-2)}} \mathbf{E}.$$

Correspondingly, in the space $G^{(p_1-2)}$ of spline functions, one has $\hat{m}_{p_1}(x) = \tilde{m}_{p_1}(x) + \tilde{\varepsilon}_{p_1}(x)$, where,

$$\tilde{m}_{p_1}(x) = [\{B_{j,p_1}(x)\}_{j=1-p_1}^{N_1}]^{\mathrm{T}} (\mathbf{B}_{p_1}^{\mathrm{T}} \mathbf{B}_{p_1})^{-1} \mathbf{B}_{p_1}^{\mathrm{T}} \mathbf{m},$$
$$\tilde{\varepsilon}_{p_1}(x) = [\{B_{j,p_1}(x)\}_{j=1-p_1}^{N_1}]^{\mathrm{T}} (\mathbf{B}_{p_1}^{\mathrm{T}} \mathbf{B}_{p_1})^{-1} \mathbf{B}_{p_1}^{\mathrm{T}} \mathbf{E}. \tag{9}$$

Regarding variance, we define $\mathbf{Z} = \{\varepsilon_1^2, \ldots, \varepsilon_n^2\}^{\mathrm{T}}$, $\hat{\mathbf{Z}}_{p_1} = \{\hat{\varepsilon}^2_{1,p_1}, \ldots, \hat{\varepsilon}^2_{n,p_1}\}^{\mathrm{T}}$, then

$$\widetilde{\sigma}^2_{p_2}(x) = [\{B_{j,p_2}(x)\}_{j=1-p_2}^{N_2}]^{\mathrm{T}} (\mathbf{B}_{p_2}^{\mathrm{T}} \mathbf{B}_{p_2})^{-1} \mathbf{B}_{p_2}^{\mathrm{T}} \mathbf{Z},$$
$$\hat{\sigma}^2_{p_2,p_1}(x) = [\{B_{j,p_2}(x)\}_{j=1-p_2}^{N_2}]^{\mathrm{T}} (\mathbf{B}_{p_2}^{\mathrm{T}} \mathbf{B}_{p_2})^{-1} \mathbf{B}_{p_2}^{\mathrm{T}} \hat{\mathbf{Z}}_{p_1}.$$

Taking difference,

$$\hat{\sigma}^2_{p_2,p_1}(x) - \widetilde{\sigma}^2_{p_2}(x) = [\{B_{j,p_2}(x)\}_{j=1-p_2}^{N_1}]^{\mathrm{T}}(\mathbf{B}_{p_2}^{\mathrm{T}}\mathbf{B}_{p_2})^{-1}\mathbf{B}_{p_2}^{\mathrm{T}}(\hat{\mathbf{Z}}_{p_1} - \mathbf{Z})$$
$$= [\{B_{j,p_2}(x)\}_{j=1-p_2}^{N_1}]^{\mathrm{T}}(\mathbf{B}_{p_2}^{\mathrm{T}}\mathbf{B}_{p_2})^{-1}\mathbf{B}_{p_2}^{\mathrm{T}} \times (\{Y_1 - \hat{m}_{p_1}(X_1)\}^2 - \varepsilon_1^2, \ldots, \{Y_n - \hat{m}_{p_1}(X_n)\}^2 - \varepsilon_n^2)^{\mathrm{T}}.$$

Then one writes

$$\hat{\sigma}^2_{p_2,p_1}(x) - \widetilde{\sigma}^2_{p_2}(x) = \mathrm{I}_{p_2,p_1}(x) + \mathrm{II}_{p_2,p_1}(x) + \mathrm{III}_{p_2,p_1}(x), \tag{10}$$

in which

$$\mathrm{I}_{p_2,p_1} = \mathrm{I}_{p_2,p_1}(x) = [\{B_{j,p_2}(x)\}_{j=1-p_2}^{N_1}]^{\mathrm{T}}(\mathbf{B}_{p_2}^{\mathrm{T}}\mathbf{B}_{p_2})^{-1}\mathbf{B}_{p_2}^{\mathrm{T}}(\mathrm{I}_{1,p_1}, \ldots, \mathrm{I}_{n,p_1})^{\mathrm{T}},$$
$$\mathrm{II}_{p_2,p_1} = \mathrm{II}_{p_2,p_1}(x) = [\{B_{j,p_2}(x)\}_{j=1-p_2}^{N_1}]^{\mathrm{T}}(\mathbf{B}_{p_2}^{\mathrm{T}}\mathbf{B}_{p_2})^{-1}\mathbf{B}_{p_2}^{\mathrm{T}}(\mathrm{II}_{1,p_1}, \ldots, \mathrm{II}_{n,p_1})^{\mathrm{T}},$$
$$\mathrm{III}_{p_2,p_1} = \mathrm{III}_{p_2,p_1}(x) = [\{B_{j,p_2}(x)\}_{j=1-p_2}^{N_1}]^{\mathrm{T}}(\mathbf{B}_{p_2}^{\mathrm{T}}\mathbf{B}_{p_2})^{-1}\mathbf{B}_{p_2}^{\mathrm{T}}(\mathrm{III}_{1,p_1}, \ldots, \mathrm{III}_{n,p_1})^{\mathrm{T}},$$
$$\mathrm{I}_{i,p_1} = \{m(X_i) - \tilde{m}_{p_1}(X_i)\}^2 + \tilde{\varepsilon}^2_{p_1}(X_i) + 2\{m(X_i) - \tilde{m}_{p_1}(X_i)\}\tilde{\varepsilon}_{p_1}(X_i),$$
$$\mathrm{II}_{i,p_1} = 2\tilde{\varepsilon}_{p_1}(X_i)\varepsilon_i, \mathrm{III}_{i,p_1} = 2\{m(X_i) - \tilde{m}_{p_1}(X_i)\}\varepsilon_i.$$

4. Implementation

In this section, we describe procedures to implement the confidence bands in Theorem 2.2. Our codes are written in XploRe for convenience in order to use kernel smoothing, see [26].

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 $[a, b]$. Motivated by the comment of one referee, we select the number of interior knots N_ν using a BIC criteria. For knot location, we use equally spaced knots. According to Assumption (A3), the optimal order of N_ν is $n^{1/(2p_\nu+1)}$. Thus we propose selecting the 'optimal' N_ν, denoted by $\hat{N}_\nu^{\mathrm{opt}}$, from $[0.5N_{r\nu},\ \min(5N_{r\nu}, Tb)]$, with $N_{r\nu} = n^{1/(2p_\nu+1)}$ and $Tb = n/4 - 1$ to ensure that the total number of parameters in the least-square estimation is less than $n/4$.

To be specific, let $q_n = (1 + N_n)$ be the total number of parameters. Then $\hat{N}_\nu^{\mathrm{opt}}$ is the one minimising the BIC value

$$\hat{N}_\nu^{\mathrm{opt}} = \underset{N_n \in [0.5N_{r\nu},\ \min(5N_{r\nu}, Tb)]}{\operatorname{argmin}} \mathrm{BIC}(N_n),$$

where $\mathrm{BIC} = \log(\mathrm{MSE}) + q_n \log(n)/n$, with $\mathrm{MSE} = \sum_{i=1}^n \{Y_i - \hat{Y}_i\}^2/n$.The least-squares problem in Equation (4) can be solved via the truncated power basis $\{1, x, \ldots, x^{p_1-1}, (x - t_j)_+^{p_1-1}, j = 1, \ldots, N_1\}$. In other words

$$\hat{m}_{p_1}(x) = \sum_{k=0}^{p_1-1} \hat{\gamma}_k x^k + \sum_{j=1}^{N_1} \hat{\gamma}_{j,p_1}(x - t_j)_+^{p_1-1},$$

where the coefficients $\{\hat{\gamma}_0, \ldots, \hat{\gamma}_{p_1-1}, \hat{\gamma}_{1,p_1}, \ldots, \hat{\gamma}_{N_1,p_1}\}^{\mathrm{T}}$ are solutions to the following least-squares problem

$$\{\hat{\gamma}_0, \ldots, \hat{\gamma}_{N_1,p_1}\}^{\mathrm{T}} = \underset{R^{N_1+p_1}}{\operatorname{argmin}} \sum_{i=1}^n \left\{ Y_i - \sum_{k=0}^{p_1-1} \gamma_k X_i^k - \sum_{j=1}^{N_1} \gamma_{j,p_1}(X_i - t_j)_+^{p_1-1} \right\}^2.$$

The variance estimators $\hat{\sigma}^2_{p_1,p_2}(x)$ are computed likewise.

When constructing the confidence bands, one needs to evaluate the functions $v_{n,p_2}^2(x)$ in Equation (7) differently for the exact and conservative bands, and the description is separated into two subsections. For both cases, one estimates the unknown functions $f(x)$ and $v_Z^2(x)$ and then plugs in these estimates, as in [19]. 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 $\widetilde{K}(u) = 15(1-u^2)^2 I\{|u| \le 1\}/16$ be the quadric kernel and s_n the sample standard deviation of $(X_i)_{i=1}^n$ and

$$\hat{f}(x) = n^{-1} \sum_{i=1}^{n} h_{2\,\mathrm{rot},f}^{-1} \widetilde{K}\left(\frac{X_i - x}{h_{2\,\mathrm{rot},f}}\right), \quad h_{2\,\mathrm{rot},f} = (4\pi)^{1/10} \left(\frac{140}{3}\right)^{1/5} n^{-1/5} s_n, \tag{11}$$

where $h_{2\,\mathrm{rot},f}$ is the rule-of-thumb bandwidth in [27]. Define $\Xi_{p_2} = \{\Xi_{i,p_2}, 1 \le i \le n\}^{\mathrm{T}}$, $\Xi_{i,p_2} = \{\hat{Z}_{i,p_1} - \hat{\sigma}_{p_1,p_2}^2(X_i)\}^2$, and

$$\mathbf{X} = \mathbf{X}(x) = \begin{pmatrix} 1 & ,\ldots, & 1 \\ X_1 - x & ,\ldots, & X_n - x \end{pmatrix}^{\mathrm{T}}, \quad \mathbf{W} = \mathbf{W}(x) = \mathrm{diag}\left\{\widetilde{K}\left(\frac{X_i - x}{h_{2\,\mathrm{rot},\sigma}}\right)\right\}_{i=1}^{n},$$

where $h_{2\,\mathrm{rot},\sigma}$ is the rule-of-thumb bandwidth of Fan and Gijbels [28] based on data $(X_i, \Xi_{i,p_2})_{i=1}^n$. Define the following estimators of $v_Z^2(x)$,

$$\hat{v}_{Z,p_2}^2(x) = (1,\ 0)(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}\boldsymbol{\Xi}_{p_2}. \tag{12}$$

The following uniform consistency results are provided in [14,28]

$$\max_{p_2} \sup_{x\in[a,b]} |\hat{v}_{Z,p_2}^2(x) - v_Z^2(x)| + \sup_{x\in[a,b]} |\hat{f}(x) - f(x)| = o_p(1). \tag{13}$$

4.1. *Implementing the exact band*

The function $v_{n,1}(x)$ is approximated by the following, with $\hat{f}(x)$ and $\hat{v}_{Z,1}(x)$ defined in Equations (11) and (12), $j(x)$ defined in Equation (5)

$$\hat{v}_{n,1}(x) = \hat{v}_{Z,1}(x)\hat{f}^{-1/2}(x)n^{-1/2}h_2^{-1/2}.$$

Then Equations (13) and (8) imply that as $n \to \infty$, the band below is asymptotically exact

$$\hat{\sigma}_{1,1}^2(x) \pm \hat{v}_{n,1}(x)\{2\log(N_2+1)\}^{1/2} d_n. \tag{14}$$

4.2. *Implementing the conservative band*

The band below is asymptotically conservative

$$\hat{\sigma}_{2,2}^2(x) \pm \hat{v}_{n,2}(x)\{2\log(N_2+1) - 2\log\alpha\}^{1/2}, \tag{15}$$

where the function $v_{n,2}(x)$ in Equation (7) for the linear band is estimated consistently by

$$\hat{v}_{n,2}(x) = \{\boldsymbol{\Delta}^{\mathrm{T}}(x)L_{j_2(x)}\boldsymbol{\Delta}(x)\}^{1/2}\hat{v}_{Z,2}(x)\left\{\frac{2}{3}\hat{f}(x)nh_2\right\}^{-1/2}, \tag{16}$$

with $j_2(x)$ defined in Equation (5), and $\hat{f}(x)$ and $\hat{v}^2_{Z,2}(x)$ defined in Equations (11) and (12), $\mathbf{\Delta}(x)$ and L_j defined as follows:

$$\mathbf{\Delta}(x) = \begin{pmatrix} c_{j_2(x)-1}\{1-r_2(x)\} \\ c_{j_2(x)} r_2(x) \end{pmatrix}, \quad c_j = \begin{cases} \sqrt{2} & j=-1, N_2 \\ 1 & j=0,\ldots,N_2-1 \end{cases},$$

$$L_j = \begin{pmatrix} l_{j+1,j+1} & l_{j+1,j+2} \\ l_{j+2,j+1} & l_{j+2,j+2} \end{pmatrix}, \quad j=0,1,\ldots,N_2. \tag{17}$$

The terms l_{ik}, $|i-k| \le 1$ are defined through the following matrix inversion

$$M_{N_2+2} = \begin{pmatrix} 1 & \sqrt{2}/4 & & & & 0 \\ \sqrt{2}/4 & 1 & 1/4 & & & \\ & 1/4 & 1 & \ddots & & \\ & & \ddots & \ddots & 1/4 & \\ & & & 1/4 & 1 & \sqrt{2}/4 \\ 0 & & & & \sqrt{2}/4 & 1 \end{pmatrix}_{(N_2+2)\times(N_2+2)} = (l_{ik})^{-1}_{(N_2+2)\times(N_2+2)},$$

and computed via Equations (18), (19), and (20) given below, which are needed for Equation (17). Letting

$$z_1 = \frac{2+\sqrt{3}}{4}, \quad z_2 = \frac{2-\sqrt{3}}{4}, \quad \theta = \frac{z_2}{z_1} = \left(2-\sqrt{3}\right)^2 = 7-4\sqrt{3}, \tag{18}$$

and applying the matrix theory from [29,30], we have

$$l_{11} = l_{N_2+2,N_2+2} = \frac{8z_1^2(1-\theta^{N_2+1}) - z_1(1-\theta^{N_2})}{8z_1^2(1-\theta^{N_2+1}) - 2z_1(1-\theta^{N_2}) + (1-\theta^{N_2-1})/8},$$

$$l_{i,i} = \frac{\{8z_1(1-\theta^{N_2+2-i}) - (1-\theta^{N_2+1-i})\}\{8z_1(1-\theta^{i-1}) - (1-\theta^{i-2})\}}{(z_1-z_2)\{64z_1^2(1-\theta^{N_2+1}) - 16z_1(1-\theta^{N_2}) + (1-\theta^{N_2-1})\}} \tag{19}$$

for $2 \le i \le N_2+1$ and

$$l_{12} = l_{N_2+1,N_2+2} = \frac{(-2\sqrt{2})z_1(1-\theta^{N_2}) - (1-\theta^{N_2-1})/8}{8z_1^2(1-\theta^{N_2+1}) - 2z_1(1-\theta^{N_2}) + 8(1-\theta^{N_2-1})/8},$$

$$l_{i,i+1} = -\frac{\{8z_1(1-\theta^{N_2+1-i}) - (1-\theta^{N_2-i})\}\{8z_1(1-\theta^{i-1}) - (1-\theta^{i-2})\}}{4z_1(z_1-z_2)\{64z_1^2(1-\theta^{N_2+1}) - 16z_1(1-\theta^{N_2}) + (1-\theta^{N_2-1})\}} \tag{20}$$

for $2 \le i \le N_2$. By the symmetry of the matrix M_{N_2+2}, the lower diagonal entries are $l_{i+1,i} = l_{i,i+1}$, $\forall i = 1,\ldots,N_2+1$; see [19] for details.

4.3. *Implementing the bootstrap band*

In this subsection, we use wild bootstrap for improved performance following the suggestion of one referee. We define the residuals $\hat{\xi}_{i,p_1,p_2} = \hat{\varepsilon}^2_{i,p_1} - \hat{\sigma}^2_{p_1,p_2}(X_i)$, where $\hat{\varepsilon}_{i,p_1}$ are defined in Equation (3), and denote a predetermined integer by n_B, whose default value is 500. The steps to compute the bootstrap band, similar to that of Yang [31], are described in the following.

Step 1. Let $\{\delta_{i,k}\}_{1\le k\le n_B}$, $1 \le i \le n$, be i.i.d. samples of the following discrete distribution $\delta_{i,k} = \pm 1$ with probability $1/2$. It is easily verified that $E(\delta_{i,k}) = 0$, $\mathrm{Var}(\delta_{i,k}) = 1$.

Step 2. For any $1 \le k \le n_B$, define the kth wild bootstrap sample $\hat{\varepsilon}^2_{i,p_{1,k}} = \hat{\sigma}^2_{p_1,p_2}(X_i) + \hat{\hat{\xi}}_{i,p_1,p_2}\delta_{i,k}$, $1 \le i \le n$. Taking $\mathbf{E}_{p_1,k} = \{\hat{\varepsilon}^2_{i,p_{1,k}}\}_{i=1}^n$, we apply the linear spline on $\mathbf{E}_{p_1,k}$ to get the spline estimate

$$\hat{\sigma}^2_{p_1,p_2,k}(x) = [\{B_{j,p_1}(x)\}_{j=1-p_1}^{N_1}]^{\mathrm{T}}(\mathbf{B}^{\mathrm{T}}_{p_1}\mathbf{B}_{p_1})^{-1}\mathbf{B}^{\mathrm{T}}_{p_1}\mathbf{E}_{p_1,k} \tag{21}$$

Step 3. The wild bootstrap $(1-\alpha)$ pointwise confidence interval for function value $\sigma^2(x)$ at one point x is $[\widehat{\sigma}^2_{L,\alpha/2}(x), \widehat{\sigma}^2_{U,\alpha/2}(x)]$, where $\widehat{\sigma}^2_{L,\alpha/2}(x)$ and $\widehat{\sigma}^2_{U,\alpha/2}(x)$ are the lower and upper $100(\alpha/2)\%$ quantiles of the set $\hat{\sigma}^2_{p_1,p_2,k}(x)_{1\le k\le n_B}$ obtained from Equation (21) for each of the bootstrap sample generated in Step 2.

Step 4. According to Wang and Yang [19], the uniform confidence band is wider than the pointwise confidence interval by a common inflation factor of $K_n = z^{-1}_{1-\alpha/2}\sqrt{2\{\log(N_2+1)-\log(\alpha/2)\}}$ when localised at any point x. Hence we define the wild bootstrap $(1-\alpha)$ confidence band for the function $\sigma^2(x)$ over $[a,b]$ as $[\widehat{\sigma}^2_{L,\alpha/2}(x), \widehat{\sigma}^2_{U,\alpha/2}(x)]$, $x \in [a,b]$, where

$$\hat{\sigma}^2_{L,\alpha/2}(x) = \hat{\sigma}^2_{p_1,p_2}(x) + (\widehat{\sigma}^2_{L,\alpha/2}(x) - \hat{\sigma}^2_{p_1,p_2}(x))z^{-1}_{1-\alpha/2}\sqrt{2\{\log(N_2+1)-\log(\alpha/2)\}},$$

$$\hat{\sigma}^2_{U,\alpha/2}(x) = \hat{\sigma}^2_{p_1,p_2}(x) + (\widehat{\sigma}^2_{U,\alpha/2}(x) - \hat{\sigma}^2_{p_1,p_2}(x))z^{-1}_{1-\alpha/2}\sqrt{2\{\log(N_2+1)-\log(\alpha/2)\}}.$$

As one referee pointed out, instead of resampling at each point x and then inflate by a universal factor K_n, it is also possible to resample the maximal deviation distribution, as was done in [32], and obtain bootstrap lower and upper $100(\alpha/2)\%$ quantiles of $\sup_{x\in[a,b]} |\hat{\sigma}^2_{p_1,p_2}(x) - \sigma^2(x)|v^{-1}_{n,p_2}(x)$. Our approach, however, has the advantage of adaptivity since the confidence band is locally calibrated at each point x, without the constraint of symmetry.

5. Examples

5.1. *Simulation example*

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

$$m(x) = \sin(2\pi x), \quad \sigma(x) = \sigma_0\frac{c-\exp(x)}{c+\exp(x)}, \quad \varepsilon|x \sim N_1\{0, \sigma^2(x)\}. \tag{22}$$

Table 1. Coverage probabilities for $c = 100$ from 500 replications.

σ_0	n	$1-\alpha$	Constant band	Linear band	Bootstrap band
0.2	100	0.99	0.882	0.886	0.944
		0.95	0.806	0.858	0.858
	200	0.99	0.940	0.970	0.996
		0.95	0.874	0.958	0.968
	500	0.99	0.984	0.994	1
		0.95	0.942	0.992	0.984
0.5	100	0.99	0.764	0.892	0.956
		0.95	0.690	0.870	0.886
	200	0.99	0.896	0.970	0.992
		0.95	0.830	0.962	0.960
	500	0.99	0.974	0.996	0.998
		0.95	0.926	0.994	0.984

Table 2. Coverage probabilities for $c = 5$ from 500 replications.

σ_0	n	$1-\alpha$	Constant band	Linear band	Bootstrap band
0.2	100	0.99	0.824	0.858	0.944
		0.95	0.764	0.834	0.874
	200	0.99	0.912	0.896	0.986
		0.95	0.832	0.884	0.954
	500	0.99	0.978	0.970	1
		0.95	0.916	0.964	0.992
0.5	100	0.99	0.886	0.856	0.946
		0.95	0.648	0.828	0.878
	200	0.99	0.916	0.918	0.992
		0.95	0.688	0.904	0.958
	500	0.99	0.958	0.966	1
		0.95	0.726	0.964	0.986

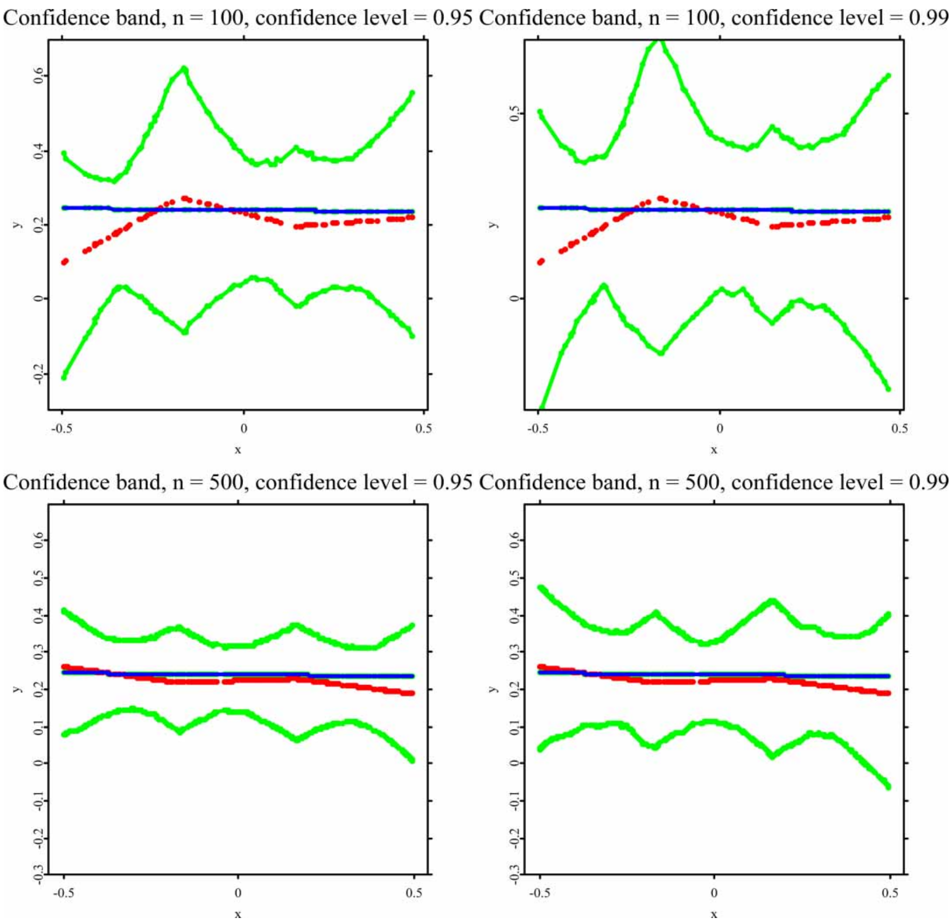

Figure 1. For data generated from model (22) (with $\sigma_0 = 0.5$, $c = 100$) of different sample size n and confidence level $1-\alpha$, plots of confidence bands for variance (thick solid), the linear spline estimator $\hat{\sigma}^2_{2,2}(x)$ (dotted), and the true function $\sigma^2(x)$ (solid). The bands are computed from the bootstrap method.

The noise levels are $\sigma_0 = 0.2, 0.5$, while sample sizes are taken to be $n = 100, 200, 500$. The confidence level $1 - \alpha = 0.99, 0.95$. For $c = 100$ and $c = 5$, Tables 1 and 2 contain the coverage probabilities as the percentage of coverage of the true curve $\sigma(x)$ at all data points $\{X_i\}_{i=1}^n$ by the confidence bands in Equations (14), (15) and using the bootstrap method, over 500 replications of sample size n. Following the suggestion of one referee, we have included variance functions $\sigma^2(x)$ that are strongly heteroscedastic ($c = 5$) and nearly homoscedastic ($c = 100$).

In all cases, the performance of the constant band is worse than the linear band in terms of coverage, whereas the bootstrap band has the best coverage. In all cases, the coverage improves with increase in sample sizes, showing positive confirmation of Theorem 2.2. The bootstrap band achieves reasonable coverage for moderate sample size as low as 100, while for the nearly homoscedastic case of $c = 100$, the asymptotic linear band has good coverage for sample size as low as $n = 200$. For the strongly heteroscedastic case $c = 5$, it seems that the bootstrap band is the only satisfactory one. We therefore recommend using the bootstrap band for analyzing real data.

The graphs in Figures 1 and 2 are created based on two samples of size 100 and 500, respectively, for $c = 100$ and 5, respectively, each with three types of symbols: center thin solid line (true curve),

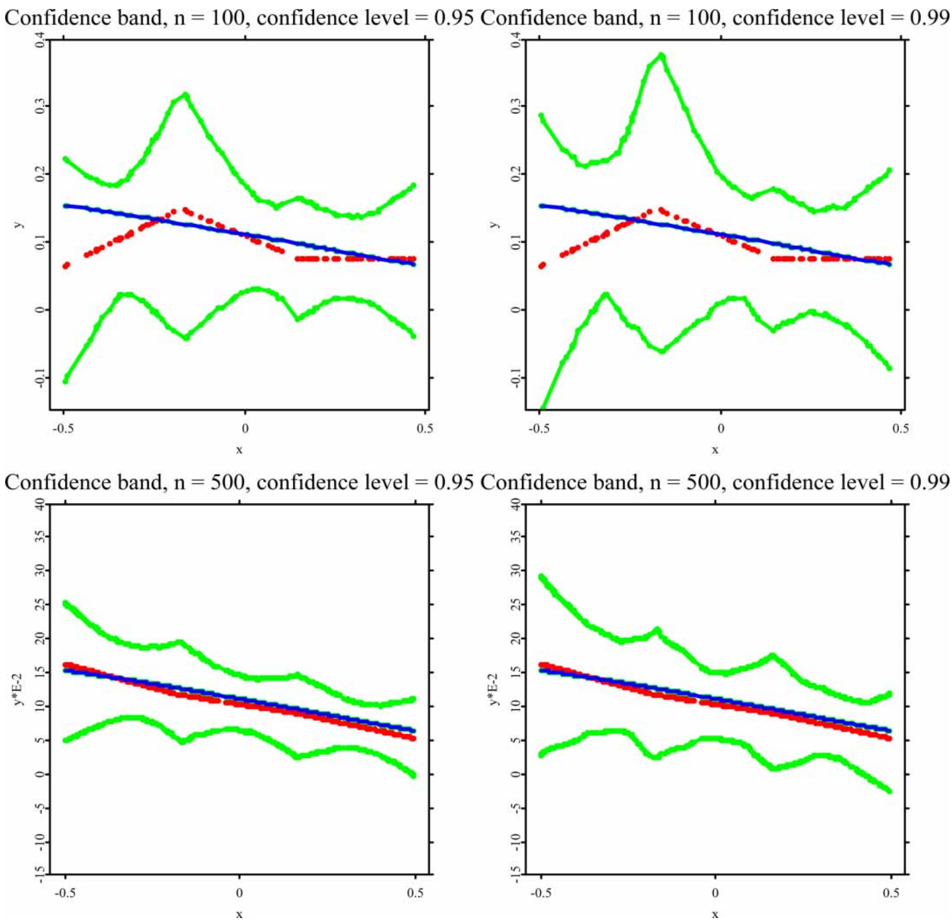

Figure 2. For data generated from model (22) (with $\sigma_0 = 0.5$, $c = 5$) of different sample size n and confidence level $1 - \alpha$, plots of confidence bands for variance (thick solid), the linear spline estimator $\hat{\sigma}^2_{2,2}(x)$ (dotted), and the true function $\sigma^2(x)$ (solid). The bands are computed from the bootstrap method.

Table 3. Simulated rejection probabilities of test homoscedasticity from 500 replications.

	$n=50$			$n=100$			$n=200$		
α	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
c					Model I				
0	0.004	0.004	0.012	0	0	0.002	0	0	0
	(0.038)	(0.056)	(0.101)	(0.028)	(0.057)	(0.093)	(0.037)	(0.059)	(0.105)
0.5	0.014	0.020	0.030	0.002	0.006	0.018	0	0.004	0.034
	(0.055)	(0.084)	(0.132)	(0.064)	(0.097)	(0.151)	(0.086)	(0.134)	(0.200)
1.0	0.038	0.058	0.110	0.024	0.072	0.254	0.150	0.362	0.690
	(0.095)	(0.148)	(0.223)	(0.153)	(0.215)	(0.313)	(0.249)	(0.337)	(0.458)
c					Model II				
0	0.004	0.004	0.012	0	0	0.002	0	0	0
	(0.031)	(0.053)	(0.100)	(0.026)	(0.049)	(0.089)	(0.032)	(0.056)	(0.100)
0.5	0.082	0.106	0.158	0.296	0.484	0.766	0.694	0.918	0.992
	(0.197)	(0.276)	(0.390)	(0.333)	(0.433)	(0.568)	(0.527)	(0.637)	(0.761)
1.0	0.316	0.422	0.612	0.356	0.512	0.734	0.656	0.884	0.984
	(0.272)	(0.365)	(0.481)	(0.477)	(0.557)	(0.674)	(0.693)	(0.790)	(0.884)
c					Model III				
0	0.004	0.004	0.012	0	0	0.002	0	0	0
	(0.034)	(0.054)	(0.097)	(0.028)	(0.053)	(0.100)	(0.031)	(0.053)	(0.094)
0.5	0.02	0.034	0.066	0.010	0.030	0.110	0.032	0.142	0.394
	(0.073)	(0.113)	(0.185)	(0.105)	(0.158)	(0.233)	(0.175)	(0.239)	(0.342)
1.0	0.078	0.112	0.216	0.122	0.312	0.642	0.668	0.984	0.978
	(0.136)	(0.198)	(0.291)	(0.221)	(0.304)	(0.412)	(0.378)	(0.476)	(0.598)

center dotted line (the estimated curve), and upper and lower thick solid line (bootstrap confidence band). In all figures, the confidence bands for $n = 500$ are thinner and fit better than those for $n = 100$.

We next compare by simulation the testing of heteroscedasticity based on the proposed bootstrap confidence band to the results of Dette and Munk [23] for the following three models:

$$\begin{aligned} &m(x) = 1 + \sin(x), \quad \sigma(x) = \sigma \exp(cx) \ \text{(monotone, model I)}, \\ &m(x) = 1 + x, \quad \sigma(x) = \sigma\{1 + c\sin(10x)\}^2 \ \text{(high frequency, model II)}, \\ &m(x) = 1 + x, \quad \sigma(x) = \sigma(1 + cx)^2 \ \text{(unimodal, model III)}, \end{aligned} \tag{23}$$

for $c = 0, 0.5, 1.0$ and $\sigma^2 = 0.25$ with standard normal errors. The design points X were generated uniformly from $[0, 1]$ and the sample sizes were $n = 50, 100, 200$. Table 3 shows the relative proportion of rejections for the various situations using both our method and the results from [23, Table 1, p. 700 (in brackets)]. Our method performs poorly when heteroscedasticity is weak ($c = 0.5$) for models I and III, so the type II error is larger than that of Dette and Munk [23]. For strongly heteroscedastic models ($c = 1$), however, our method achieves higher rejection power for models II and III, and comparable rejection power for model I, so the type II error is either comparable to Dette and Munk [23] or lower. For homoscedastic model ($c = 0$), our rejection rate is always lower, hence the bootstrap confidence band based test has smaller type I error than that of Dette and Munk [23]. Based on the above simulation, our method is better than that of Dette and Munk [23] at detecting strong heteroscedasticity and retaining homoscedasticity, whereas theirs is better than ours at discovering weak heteroscedasticity.

5.2. *Fossil data and motorcycle data*

In this subsection, we apply the bootstrap band to two real data sets, both of which have a sample size below 200.

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 [33] to detect the structure via kernel smoothing. The corresponding penalised spline fit was provided in Ruppert et al. [34]. In this section, we test the heteroscedasticity of the fossil data variance. The null hypothesis is $H_0 : \sigma^2(x) = \sigma_0^2 > 0$. 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 million years.

In Figure 3, the center dotted line is the linear spline fit $\hat{\sigma}^2_{2,2}(x)$ for the variance function $\sigma^2(x)$. The upper/lower thick solid lines represent the bootstrap confidence band. The constant

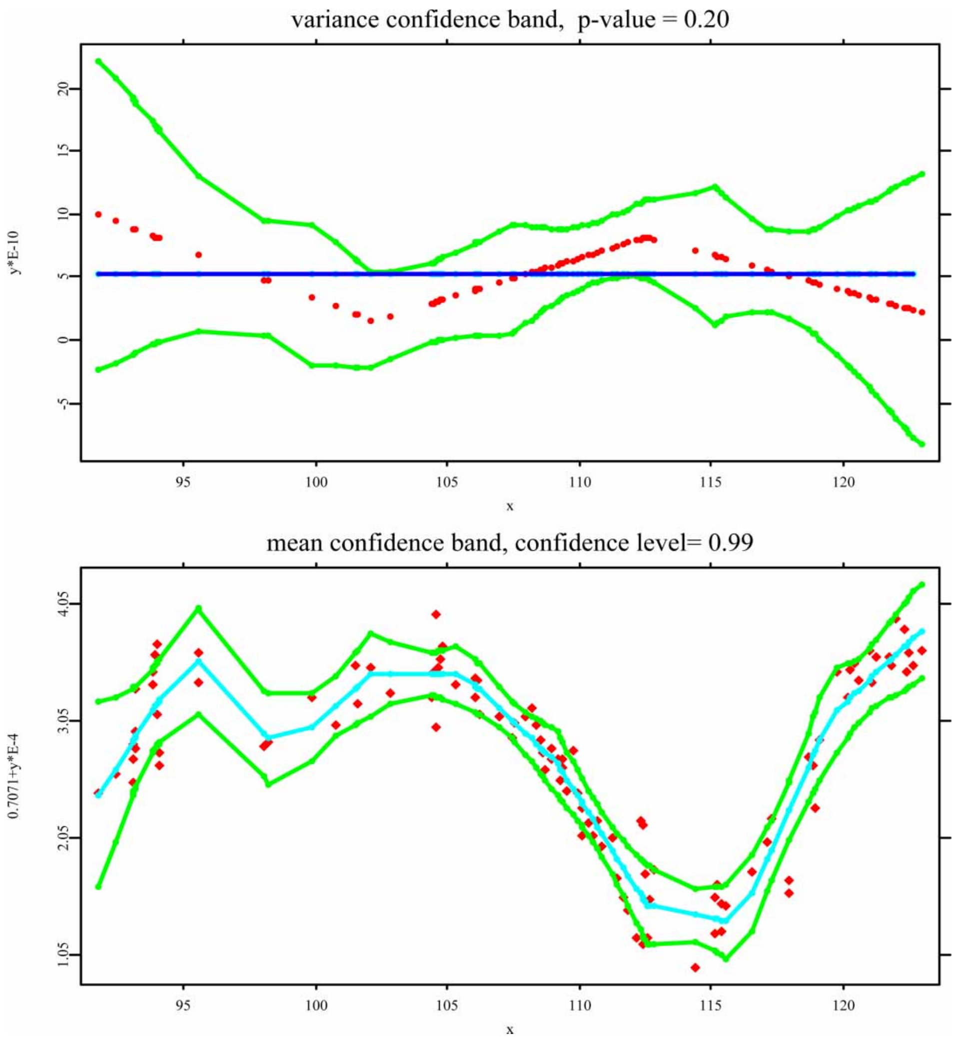

Figure 3. For the fossil data, plots of variance confidence bands (thick solid) computed by the bootstrap method, the linear spline estimator $\hat{\sigma}^2_{2,2}(x)$ (dotted), and a constant variance function that fits in the confidence band (solid). The lower picture is the data scatter plot and the confidence band for mean (thin solid).

horizontal line between the upper/lower thick lines represents the average of the minimum of the upper line and the maximum of the lower line, which indicates if one can fit a constant line into the confidence band. Since the variance band of high confidence level $100(1-0.20)\%$ contains the fitted constant line entirely, we have failed to reject the null hypothesis of homoscedasticity with p-value 0.20.

A second data used to illustrate our technique is the well-known motorcycle data. The X-values denote time (in milliseconds) after a simulated impact with motorcycles. The response variable Y is the head acceleration of a PTMO (post mortem human test object).

In Figure 4, the center dotted line is the linear spline fit $\hat{\sigma}^2_{2,2}(x)$ for $\sigma^2(x)$. The upper/lower thick solid lines represent the bootstrap confidence band. The constant line between the upper/lower thick lines represents the average of the minimum of the upper line and the maximum of the lower line. Since the variance band of an extremely high confidence level $100(1-0.008)\%$ does not

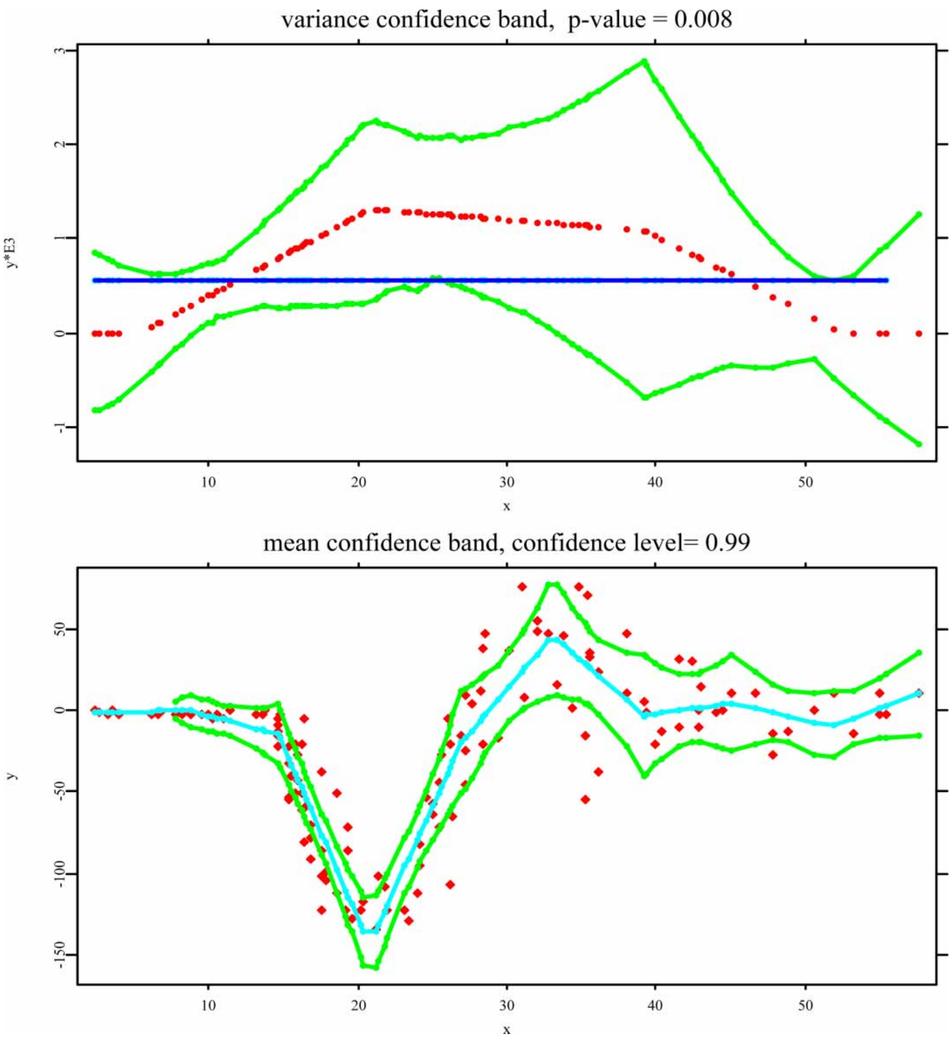

Figure 4. For the motorcycle data, plots of variance confidence bands (thick solid) computed by the bootstrap method, the linear spline estimator $\hat{\sigma}^2_{2,2}(x)$ (dotted), and a constant variance function that fits in the confidence band (solid). The lower picture is the data scatter plot and the confidence band for mean (thin solid).

contain the fitted constant line entirely, we reject the null hypothesis of homoscedasticity with $p \leq 0.008$.

In both Figures 3 and 4, there exists an exact correspondence of high $\hat{\sigma}^2_{2,2}(x)$ value in the upper plot to greater width of the confidence band for the conditional mean function in the lower plot, throughout the entire data range.

Acknowledgements

This research is supported in part by NSF grant DMS 0706518, and is part of the first author's dissertation work under the supervision of the second author. The authors appreciate the helpful comments made by two referees and the Editor Michael Akritas, which have led to significant improvements of the paper.

References

[1] H.G. Müller and U. Stadtmüller, *Variable bandwidth kernel estimators of regression curves*, Ann. Statist. 15 (1997), pp. 182–201.
[2] P. Hall and R.J. Carroll, *Variance function estimation in regression: the effect of estimating the mean*, J. Roy. Statist. Soc. Ser. B 51 (1989), pp. 3–14.
[3] D. Ruppert, M.P. Wand, U. Holst, and O. Hössjer, *Local polynomial variance-function estimation*, Technometrics 39 (1997), pp. 262–273.
[4] J. Fan and Q. Yao, *Efficient estimation of conditional variance functions in stochastic regression*, Biometrika 85 (1998), pp. 645–660.
[5] Q. Yao and H. Tong, *Nonparametric estimation of ratios of noise to signal in stochastic regression*, Statist. Sinica 10 (2000), pp. 751–770.
[6] P. Hall and J.S. Marron, *On variance estimation in nonparametric regression*, Biometrika 77 (1990), pp. 415–419.
[7] C.M. Dahl and M. Levine *Nonparametric estimation of volatility models with serially dependent innovations*, Statist. Probab. Lett. 76 (2006), pp. 2007–2016.
[8] M. Levine, *Bandwidth selection for a class of difference-based variance estimators in the nonparametric regression: a possible approach*, Comput. Statist. Data Anal. 50 (2006), pp. 3405–3431.
[9] L.D. Brown and M. Levine, *Variance estimation in nonparametric regression via the difference sequence method*, Ann. Statist. 35 (2007), pp. 2219–2232.
[10] P. Hall and D.M. Titterington, *On confidence bands in nonparametric density estimation and regression*, J. Multi. Anal. 27 (1988), pp. 228–254.
[11] W. Härdle, *Asymptotic maximal deviation of M-smoothers*, J. Multi. Anal. 29 (1989), pp. 163–179.
[12] Y. Xia, *Bias-corrected confidence bands in nonparametric regression*, J. Roy. Statist. Assoc. B 60 (1998), pp. 797–811.
[13] G. Claeskens and I. Van Keilegom, *Bootstrap confidence bands for regression curves and their derivatives*, Ann. Statist. 31 (2003), pp. 1852–1884.
[14] P.J. Bickel and M. Rosenblatt, *On some global measures of the deviations of density function estimates*, Ann. Statist. 1 (1973), pp. 1071–1095.
[15] Y.P. Mack and B.W. Silverman, *Weak and strong uniform consistency of kernel regression estimates*, Z. Wahrscheinlichkeitstheorie verm. Gebiete 61 (1982), pp. 405–415.
[16] M. Rosenblatt, *On the maximal deviation of k-dimensional density estimates*, Ann. Prob. 4 (1976), pp. 1009–1015.
[17] G. Tusnády, *A remark on the approximation of the sample df in the multidimensional case*, Period. Math. Hungar. 8 (1977), pp. 53–55.
[18] S. Zhou, X. Shen, and D.A. Wolfe, *Local asymptotics of regression splines and confidence regions*, Ann. Statist. 26 (1998), pp. 1760–1782.
[19] J. Wang and L. Yang, *Polynomial spline confidence bands for regression curves*, Statist. Sinica 19 (2009), pp. 325–342.
[20] C.J. Stone, *The use of polynomial splines and their tensor products in multivariate function estimation*, Ann. Statist. 22 (1994), pp. 118–184.
[21] J.Z. Huang, *Local asymptotics for polynomial spline regression*, Ann. Statist. 31 (2003), pp. 1600–1635.
[22] L. Xue and L. Yang, *Additive coefficient modeling via polynomial spline*, Statist. Sinica 16 (2006), pp. 1423–1446.
[23] H. Dette and A. Munk, *Testing heteroscedasticity in nonparametric regression*, J. Royal Statist. Society B. 60 (1998), pp. 693–708.
[24] N. Bissantz, G. Claeskens, H. Holzmann, and A. Munk, *Testing for lack of fit in inverse regression – with applications to biophotonic imaging*, J. Royal Statist. Soc. Ser. B. 71 (2009), pp. 25–48.
[25] C. de Boor, *A Practical Guide to Splines*, Springer-Verlag, New York, 2001.
[26] W. Härdle, Z. Hlávka, and S. Klinke, *XploRe Application Guide*, Springer-Verlag, Berlin, 2000.
[27] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 1986.
[28] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London, 1996.
[29] F.R. Gantmacher and M.G. Krein, *Oszillationsmatrizen, Oszillationskerne und kleine Schwingungen mechanischer Systeme*. Akademie-Verlag, Berlin, 1960.
[30] F. Zhang, *Matrix Theory. Basic Results and Techniques*, Springer-Verlag, New York, 1999.

[31] L. Yang, *Confidence band for additive regression model*, J. Data Sci. 6 (2008), pp. 207–217.
[32] M.H. Neumann and J.P. Kreiss, *Regression-type inference in nonparametric autoregression*, Ann. Statist. 26 (1998), pp. 1570–1613.
[33] P. Chaudhuri and J.S. Marron, *SiZer for exploration of structures in curves*, J. Amer. Statist. Assoc. 94 (1999), pp. 807–823.
[34] D. Ruppert, M.P. Wand, and R.J. Carroll, *Semiparametric Regression*, Cambridge University Press, Cambridge, 2003.
[35] M.R. Leadbetter, G. Lindgren, and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York, 1983.

Appendix

The goals of this appendix are to prove Propositions A.1, A.2, and A.3. These clearly establish Theorem 2.1 and Theorem 2.2. In what follows, we denote by $\|\xi\|$ the Euclidean norm and by $|\xi|$ the largest absolute value of the elements of any vector ξ. We use c, C to denote positive constants in the generic sense.

The following result is based on [19, Theorem 3.2 and Propositions 3.1, 3.2], see also [21,35].

LEMMA A.1 *Under Assumptions* (A1)–(A4), *there exists a constant* $C_{p_1} > 0$, $p_1 \geq 1$, *such that for any* $m \in C^{(p_1)}[a,b]$ *and the function* $\tilde{m}_{p_1}(x)$ *given in Equation* (9),

$$\|\tilde{m}_{p_1}(x) - m(x)\|_\infty \leq C_{p_1} \inf_{g \in G^{(p_1-2)}} \|g - m\|_\infty = O_p(h_1^{p_1}). \tag{A1}$$

Moreover, for the function $\tilde{\varepsilon}_{p_1}(x)$ *given in Equation* (9),

$$\|\tilde{\varepsilon}_{p_1}(x)\|_\infty = O_p(h_1^{p_1}\sqrt{\log n}). \tag{A2}$$

According to Lemma A.1, the bias term $\tilde{m}_{p_1}(x) - m(x)$ is uniformly of order $O_p(h_1^{p_1}) = O_p(n^{-p_1/(2p_1+1)})$, while the noise term $\tilde{\varepsilon}_{p_1}(x)$ is uniformly of order $O_p(h_1^{p_1}\sqrt{\log n}) = O_p(n^{-p_1/(2p_1+1)}\sqrt{\log n})$.

The following lemma on uniform convergence of the empirical inner product to the theoretical counterparts is from [19, Lemma 3.1].

LEMMA A.2 *Under Assumptions* (A2) *and* (A3), *as* $n \to \infty$,

$$A_{n,p_1} = \sup_{g_1,g_2 \in G^{(p_1-2)}} \left| \frac{\langle g_1, g_2\rangle_n - \langle g_1, g_2\rangle}{\|g_1\|_2 \|g_2\|_2} \right| = O_p\left(\sqrt{n^{-1}h_1^{-1}\log(n)}\right). \tag{A3}$$

The next result on the empirical inner product matrix is based on [12, Lemma A.5; 19, Lemma B.2].

LEMMA A.3 *Under Assumptions* (A2) *and* (A3), *there exist constants* $c(f)$, $C(f) > 0$ *independent of* n *but dependent on* f, *such that as* $n \to \infty$, *with probability approaching* 1

$$c(f)|\xi| \leq |(n^{-1}\mathbf{B}_{p_\nu}^{\mathrm{T}}\mathbf{B}_{p_\nu})^{-1}\xi| \leq C(f)|\xi|, \tag{A4}$$

$$c(f)\|\xi\|^2 \leq \xi^{\mathrm{T}}(n^{-1}\mathbf{B}_{p_\nu}^{\mathrm{T}}\mathbf{B}_{p_\nu})^{-1}\xi \leq C(f)\|\xi\|^2, \quad \forall \xi \in R^{N_\nu+p_\nu}, \ \ \nu = 1, 2. \tag{A5}$$

Using the above three results, we establish two additional technical lemmas to be used in proving Propositions A.1, A.2, and A.3.

LEMMA A.4 *Under Assumptions* (A2) *and* (A3), *as* $n \to \infty$,

$$\sup_{x\in[a,b]} \{|\{B_{j,p_\nu}(x)\}_{j=1-p_\nu}^{N_\nu}| + \|\{B_{j,p_\nu}(x)\}_{j=1-p_\nu}^{N_\nu}\|\} = O(h_\nu^{-1/2}), \tag{A6}$$

$$\max_{j=1-p_\nu}^{N_\nu} \{\langle B_{j,p_\nu}, 1\rangle\} = O(h_\nu^{1/2}), \qquad \max_{j=1-p_\nu}^{N_\nu} \{\langle B_{j,p_\nu}, 1\rangle_n\} = O_p(h_\nu^{1/2} + \sqrt{n^{-1}h_\nu^{-1}\log n}). \tag{A7}$$

Proof For each $x \in [a,b]$, at most p_ν of the $B_{j,p_\nu}(x)$'s are nonzero, (A6) follows directly from the definition of $\{B_{j,p_\nu}(x)\}_{j=1-p_\nu}^{N_\nu}$ and the simple fact that $\|b_{j,p_\nu}\|_2 \geq ch_\nu^{1/2}$, $1 - p_\nu \leq j \leq N_\nu$. The same definition and fact also imply that

$$\max_{j=1-p_\nu}^{N_\nu} \{\langle B_{j,p_\nu}, 1\rangle\} \leq p_\nu \times h_\nu \times (ch_\nu^{1/2})^{-1} = O(h_\nu^{1/2}).$$

As all $\{B_{j,p_\nu}(x)\}_{j=1-p_\nu}^{N_\nu}$ are standardised, the definition and rate of A_{n,p_ν} in (A3) imply the second half of (A7). ■

LEMMA A.5 *Under Assumptions* (A2) *and* (A3), *as* $n \to \infty$,

$$\sum_{j=1-p_2}^{N_2} \sum_{k=1-p_1}^{N_1} \left\{ n^{-1} \sum_{i=1}^{n} B_{j,p_2}(X_i)\varepsilon_i B_{k,p_1}(X_i) \right\}^2 = O_p(n^{5/2(2p_*+1)-1/2(2p^*+1)-1}), \tag{A8}$$

while for any continuous function r defined on $[a, b]$,

$$E \sum_{j=1-p_1}^{N_1} \left[n^{-1} \sum_{i=1}^{n} B_{j,p_1}(X_i) r(X_i)\varepsilon_i \right]^2 \leq \|\sigma\|_\infty^2 \|r\|_\infty^2 (N_1 + p_1) n^{-1}. \tag{A9}$$

Proof

$$\begin{aligned} &E \sum_{j=1-p_2}^{N_2} \sum_{k=1-p_1}^{N_1} \left\{ n^{-1} \sum_{i=1}^{n} B_{j,p_2}(X_i)\varepsilon_i B_{k,p_1}(X_i) \right\}^2 \\ &= \sum_{j=1-p_2}^{N_2} \sum_{k=1-p_1}^{N_1} n^{-2} \sum_{i=1}^{n} E\{B_{j,p_2}(X_i)^2 B_{k,p_1}(X_i)^2 \sigma^2(X_i)\} \\ &\leq n^{-1} \max(N_1 + p_1, N_2 + p_2) N_*^{-1} N^* \max_{|k-j|\leq p_1} E\{B_{j,p_1}(X_1)^2 B_{k,p_1}(X_1)^2 \sigma^2(X_1)\}. \end{aligned}$$

With the definition of $B_{j,p_1}(x) \equiv b_{j,p_1}(x)\|b_{j,p_1}\|_2^{-1},\ 1 - p_1 \leq j \leq N_1$, we have

$$\max_{|k-j|\leq p_1} E\{B_{j,p_2}(X_1)^2 B_{k,p_1}(X_1)^2 \sigma^2(X_1)\} \leq c(\sigma) \frac{c(f)\sqrt{h_1 h_2}}{C(f) h_1 h_2} = \frac{C(f,\sigma)}{\sqrt{h_1 h_2}}.$$

Thus Equation (A8) follows from

$$\begin{aligned} E \sum_{j=1-p_2}^{N_2} \sum_{k=1-p_1}^{N_1} \left\{ n^{-1} \sum_{i=1}^{n} B_{j,p_2}(X_i)\varepsilon_i B_{k,p_1}(X_i) \right\}^2 &\leq n^{-1} \max(N_1 + p_1, N_2 + p_2) N_*^{-1} N^* \times \frac{C(f,\sigma)}{\sqrt{h_1 h_2}} \\ &= O(n^{5/2(2p_*+1)-1/2(2p^*+1)-1}). \end{aligned}$$

To prove (A9), we argue that

$$\begin{aligned} E \sum_{j=1-p_1}^{N_1} \left[n^{-1} \sum_{i=1}^{n} B_{j,p_1}(X_i) r(X_i)\varepsilon_i \right]^2 &= \sum_{j=1-p_1}^{N_1} n^{-1} E\{B_{j,p_1}^2(X_1) r^2(X_1) \sigma^2(X_1)\} \\ &\leq \|\sigma\|_\infty^2 \|r\|_\infty^2 n^{-1} \sum_{j=1-p_1}^{N_1} E\{B_{j,p_1}(X_1)^2\} = \|\sigma\|_\infty^2 \|r\|_\infty^2 (N_1 + p_1) n^{-1}. \end{aligned}$$

The next three propositions show the asymptotical property of the three terms, I_{p_2,p_1}, II_{p_2,p_1}, and III_{p_2,p_1} in Equation (10), decomposed from Section 3, and then establish Theorem 2.1. ■

PROPOSITION A.1 *Under Assumptions* (A1)–(A4), *as* $n \to \infty$, $\|\mathrm{I}_{p_2,p_1}\|_\infty = \sup_{x\in[a,b]} |\mathrm{I}_{p_2,p_1}(x)|$ *is of order*

$$O_p(h_1^{2p_1} \log n) = O_p(n^{-2p_1/(2p_1+1)} \log n) = o_p(n^{-p_2/(2p_2+1)}).$$

Proof By the Cauchy–Schwarz inequality, $|\mathrm{I}_{i,p_1}| \leq 2\{m(X_i) - \tilde{m}_{p_1}(X_i)\}^2 + 2\tilde{\varepsilon}_{p_1}^2(X_i)$, thus $\max_{i=1}^{n} |\mathrm{I}_{i,p_1}|$ is bounded by

$$2\{\max_{i=1}^{n}\{m(X_i) - \tilde{m}_{p_1}(X_i)\}^2 + \max_{i=1}^{n} \tilde{\varepsilon}_{p_1}^2(X_i)\} \leq 2\{\|m - \tilde{m}_{p_1}\|_\infty^2 + \|\tilde{\varepsilon}_{p_1}\|_\infty^2\}.$$

It follows that

$$\|I_{p_1,p_2}\|_\infty = \sup_{x\in[a,b]} |\{B_{j,p_2}(x)\}_{j=1-p_2}^{N_2} (\mathbf{B}_{p_2}^{\mathrm{T}}\mathbf{B}_{p_2})^{-1}\mathbf{B}_{p_2}^{\mathrm{T}}(\mathrm{I}_{i,p_1}, 1 \le i \le n)^{\mathrm{T}}|,$$

which, as for each $x \in [a, b]$, $B_{j,p_2}(x) \neq 0$ for at most p_2 values of j, is bounded by

$$p_2 \max_{j=1-p_2}^{N_2} B_{j,p_2}(x)|(n^{-1}\mathbf{B}_{p_2}^{\mathrm{T}}\mathbf{B}_{p_2})^{-1} \times n^{-1}\mathbf{B}_{p_2}^{\mathrm{T}}(|\mathrm{I}_{i,p_1}|, 1 \le i \le n)^{\mathrm{T}}|.$$

Using Equation (A6) in Lemma A.4 and (A4) in Lemma A.3, the above is bounded by

$$p_2 c(f) h_2^{-1/2} \times C(f) \times |n^{-1}\mathbf{B}_{p_2}^{\mathrm{T}}(|\mathrm{I}_{i,p_1}|, 1 \le i \le n)^{\mathrm{T}}|.$$

Using the bound on $\max_{i=1}^{n} |\mathrm{I}_{i,p_1}|$,

$$\|\mathrm{I}_{p_1,p_2}\|_\infty \le C(f) h_2^{-1/2} \times \{\|m - \tilde{m}_{p_1}\|_\infty^2 + \|\tilde{\varepsilon}_{p_1}\|_\infty^2\} \times \max_{j=1-p_2}^{N_2} \{\langle B_{j,p_2}, 1\rangle_n\},$$

which, applying Equations (A1) and (A2) in Lemma A.1, and (A3), (A7) in Lemma A.4, is bounded by

$$O_p\left(h_2^{-1/2} \times \left(h_1^{2p_1} + h_1^{2p_1}\log n\right) \times \left(h_2^{1/2} + \sqrt{n^{-1}h_2^{-1}\log n}\right)\right) = O_p\left(h_1^{2p_1}\log n\right).$$

■

Proposition A.2 *Under Assumptions* (A1)–(A4), *as* $n \to \infty$,

$$\|\mathrm{II}_{p_2,p_1}\|_\infty = \sup_{x\in[a,b]} |\mathrm{II}_{p_2,p_1}(x)| = O_p(n^{3/(2p_*+1)-3/2}) = o_p(n^{-p_2/(2p_2+1)}).$$

Proof By definition

$$\begin{aligned}
\mathrm{II}_{p_1,p_2}(x) &= \{B_{1-p_2,p_2}(x), \ldots, B_{N_2,p_2}(x)\}(\mathbf{B}_{p_2}^{\mathrm{T}}\mathbf{B}_{p_2})^{-1}\mathbf{B}_{p_2}^{\mathrm{T}}(\mathrm{II}_{1,p_1}, \ldots, \mathrm{II}_{n,p_1})^{\mathrm{T}} \\
&= 2\{B_{1-p_2,p_2}(x), \ldots, B_{N_2,p_2}(x)\}(n^{-1}\mathbf{B}_{p_2}^{\mathrm{T}}\mathbf{B}_{p_2})^{-1}\left\{n^{-1}\sum_{i=1}^{n} B_{j,p_2}(X_i)\tilde{\varepsilon}_{p_1}(X_i)\varepsilon_i\right\}_{j=1-p_2}^{N_2}.
\end{aligned}$$

Applying (A4) in Lemma A.3, $|\mathrm{II}_{p_1,p_2}(x)|$, with probability approaching 1, is bounded by

$$C\|\{B_{1-p_2,p_2}(x), \ldots, B_{N_2,p_2}(x)\}\| C(f) \left\| \left\{n^{-1}\sum_{i=1}^{n} B_{j,p_2}(X_i)\tilde{\varepsilon}_{p_1}(X_i)\varepsilon_i\right\}_{j=1-p_2}^{N_2} \right\|,$$

applying Equation (A6) in Lemma A.4,

$$\sup_{x\in[a,b]} |\mathrm{II}_{p_1,p_2}(x)| \le C(f) h_2^{-1/2} \left\| \left\{n^{-1}\sum_{i=1}^{n} B_{j,p_2}(X_i)\tilde{\varepsilon}_{p_1}(X_i)\varepsilon_i\right\}_{j=1-p_2}^{N_2} \right\|.$$

Next, one can write for any $1 - p_2 \le j \le N_2$,

$$\begin{aligned}
&n^{-1}\sum_{i=1}^{n} B_{j,p_2}(X_i)\tilde{\varepsilon}_{p_1}(X_i)\varepsilon_i \\
&\quad = n^{-1}\sum_{i=1}^{n} B_{j,p_2}(X_i)\varepsilon_i\{B_{1-p_2,p_2}(X_i), \ldots, B_{N_2,p_2}(X_i)\}(n^{-1}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{B}_{p_1})^{-1}n^{-1}\mathbf{B}_{p1}^{\mathrm{T}}\mathbf{E} \\
&\quad = \left\{n^{-1}\sum_{i=1}^{n} B_{j,p_2}(X_i)\varepsilon_i B_{k,p_1}(X_i)\right\}_{k=1-p_1}^{N_1} (n^{-1}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{B}_{p_1})^{-1}n^{-1}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{E},
\end{aligned}$$

hence, $\sup_{x\in[a,b]} |\mathrm{II}_{p_1,p_2}(x)|$ is bounded by

$$C(f)h_2^{-1/2}\sqrt{\sum_{j=1-p_2}^{N_2}\left[\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\varepsilon_i B_{k,p_1}(X_i)\right\}_{k=1-p_1}^{N_1}\left(\frac{1}{n}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{B}_{p_1}\right)^{-1}n^{-1}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{E}\right]^2}$$

$$\leq C(f)h_2^{-1/2}\sqrt{\sum_{j=1-p_2}^{N_2}\left\|\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\varepsilon_i B_{k,p_1}(X_i)\right\}_{k=1-p_1}^{N_1}\right\|^2}\left\|\left(\frac{1}{n}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{B}_{p_1}\right)^{-1}\frac{1}{n}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{E}\right\|$$

$$= C(f)h_2^{-1/2}\sqrt{\sum_{j=1-p_2}^{N_2}\sum_{k=1-p_1}^{N_1}\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\varepsilon_i B_{k,p_1}(X_i)\right\}^2}\left\|\left(\frac{1}{n}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{B}_{p_1}\right)^{-1}\frac{1}{n}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{E}\right\|$$

by the Cauchy–Schwarz inequality. Note that with probability approaching 1,

$$\|(n^{-1}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{B}_{p_1})^{-1}n^{-1}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{E}\|^2 \leq C(f)^2\{(n^{-1}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{E})^{\mathrm{T}}n^{-1}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{E}\}$$

$$= C(f)^2\sum_{k=1-p_1}^{N_1}\left\{n^{-1}\sum_{i=1}^{n}B_{j,p_1}(X_i)\varepsilon_i\right\}^2 = O_p\{(N_1+p_1)n^{-1}\} = O_p(N_1 n^{-1})$$

according to Equation (A9) of Lemma A.5 with function $r(x) \equiv 1$. Meanwhile, according to Equation (A8) in Lemma A.5, we have

$$\sup_{x\in[a,b]}|\mathrm{II}_{p_1,p_2}(x)| = O_p(h_2^{-1/2}\times n^{5/2(2p_*+1)-1/2(2p^*+1)-1}\times\sqrt{N_1/n})$$

$$= O_p(n^{1/2(2p_2+1)}\times n^{5/2(2p_*+1)-1/2(2p^*+1)-1}\times n^{1/2(2p_1+1)}n^{-1/2}) = O_p(n^{3/(2p_*+1)-3/2}).$$

■

Proposition A.3 *Under Assumptions* (A1)–(A4), *as* $n\to\infty$,

$$\|\mathrm{III}_{p_2,p_1}\|_\infty = \sup_{x\in[a,b]}|\mathrm{III}_{p_2,p_1}(x)| = O_p(n^{3/(2p_*+1)-1}) = o_p(n^{-p_{21}/(2p_2+1)}).$$

Proof

$$\mathrm{III}_{p_1,p_2}(x) = \{B_{1-p_2,p_2}(x),\ldots,B_{N_2,p_2}(x)\}\left(\frac{1}{n}\mathbf{B}_{p_2}^{\mathrm{T}}\mathbf{B}_{p_2}\right)^{-1}\frac{1}{n}\mathbf{B}_{p_2}^{\mathrm{T}}(\mathrm{III}_{1,p_1},\ldots,\mathrm{III}_{n,p_1})^{\mathrm{T}}$$

$$= 2\{B_{j,p_2}(x)\}_{j=1-p_2}^{N_2}\left(\frac{1}{n}\mathbf{B}_{p_2}^{\mathrm{T}}\mathbf{B}_{p_2}\right)^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\{m(X_i)-\tilde{m}_{p_1}(X_i)\}\varepsilon_i\right\}_{j=1-p_2}^{N_2}$$

$$= 2\{B_{j,p_2}(x)\}_{j=1-p_2}^{N_2}\left(\frac{1}{n}\mathbf{B}_{p_2}^{\mathrm{T}}\mathbf{B}_{p_2}\right)^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\{m(X_i)-g_{p_1}(X_i)\}\varepsilon_i\right\}_{j=1-p_2}^{N_2}$$

$$+ 2\{B_{j,p_2}(x)\}_{j=1-p_2}^{N_2}\left(\frac{1}{n}\mathbf{B}_{p_2}^{\mathrm{T}}\mathbf{B}_{p_2}\right)^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\{g_{p_1}(X_i)-\tilde{m}_{p_1}(X_i)\}\varepsilon_i\right\}_{j=1-p_2}^{N_2}$$

$$= 2\{B_{j,p_2}(x)\}_{j=1-p_2}^{N_2}\left(\frac{1}{n}\mathbf{B}_{p_2}^{\mathrm{T}}\mathbf{B}_{p_2}\right)^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\{m(X_i)-g_{p_1}(X_i)\}\varepsilon_i\right\}_{j=1-p_2}^{N_2}$$

$$+ 2\{B_{j,p_2}(x)\}_{j=1-p_2}^{N_2}\left(\frac{1}{n}\mathbf{B}_{p_2}^{\mathrm{T}}\mathbf{B}_{p_2}\right)^{-1}\left\{\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\varepsilon_i B_{k,p_1}(X_i)\right\}_{k=1-p_1}^{N_1}\right.$$

$$\left.\times\left(\frac{1}{n}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{B}_{p_1}\right)^{-1}\frac{1}{n}\mathbf{B}_{p_1}^{\mathrm{T}}(\mathbf{g}_{p_1}-\mathbf{m})\right\}_{j=1-p_2}^{N_2},$$

in which the spline function $g_{p_1}\in G^{(p_1-1)}$ satisfies $\| m-g_{p_1}\|_\infty \leq Ch_1^{p_1}$, and $\mathbf{g}_{p_1} = \{g_{p_1}(X_1),\ldots,g_{p_1}(X_n)\}^{\mathrm{T}}$.

With probability approaching 1, according to Equation (A6) in Lemma A.4, the first term in the above is bounded by

$$\left|2\{B_{1-p_2,p_2}(x)\}_{j=1-p_2}^{N_2}\left(\frac{1}{n}\mathbf{B}_{p_2}^{\mathrm{T}}\mathbf{B}_{p_2}\right)^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\{m(X_i)-g_{p_1}(X_i)\}\varepsilon_i\right\}_{j=1-p_2}^{N_2}\right|$$

$$\leq C(f)h_2^{-1/2}\left\|\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\{m(X_i)-g_{p_1}(X_i)\}\varepsilon_i\right\}_{j=1-p_2}^{N_2}\right\|.$$

By (A9) in Lemma A.5 with $r(x)=m(x)-g_{p_1}(x)$, the above has order

$$O_p\left(h_2^{-1/2}\sqrt{\|\sigma\|_\infty^2\|m-g_{p_1}\|_\infty^2(N_1+p_1)n^{-1}}\right)=O_p(N_2^{1/2}N_1/n).$$

For the second term,

$$\left|2\{B_{1-p_2,p_2}(x),\ldots,B_{N_2,p_2}(x)\}\left(\frac{1}{n}\mathbf{B}_{p_2}^{\mathrm{T}}\mathbf{B}_{p_2}\right)^{-1}\right.$$

$$\left.\times\left\{\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\varepsilon_iB_{k,p_1}(X_i)\right\}_{k=1-p_1}^{N_1}\left(\frac{1}{n}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{B}_{p_1}\right)^{-1}\frac{1}{n}\mathbf{B}_{p_1}^{\mathrm{T}}(\mathbf{g}_{p_1}-\mathbf{m})\right\}_{j=1-p_2}^{N_2}\right|$$

$$\leq\frac{C(f)}{h_2^{1/2}}\left\|\left\{\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\varepsilon_iB_{k,p_1}(X_i)\right\}_{k=1-p_1}^{N_1}\left(\frac{1}{n}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{B}_{p_1}\right)^{-1}\frac{1}{n}\mathbf{B}_{p_1}^{\mathrm{T}}(\mathbf{g}_{p_1}-\mathbf{m})\right\}_{j=1-p_2}^{N_2}\right\|$$

$$=\frac{C(f)}{h_2^{1/2}}\sqrt{\sum_{j=1-p_2}^{N_2}\left[\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\varepsilon_iB_{k,p_1}(X_i)\right\}_{k=1-p_1}^{N_1}\left(\frac{1}{n}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{B}_{p_1}\right)^{-1}\frac{1}{n}\mathbf{B}_{p_1}^{\mathrm{T}}(\mathbf{g}_{p_1}-\mathbf{m})\right]^2}$$

$$\leq\frac{C(f)}{h_2^{1/2}}\sqrt{\sum_{j=1-p_2}^{N_2}\left\|\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_2}(X_i)\varepsilon_iB_{k,p_1}(X_i)\right\}_{k=1-p_1}^{N_1}\right\|^2}\left\|\left(\frac{1}{n}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{B}_{p_1}\right)^{-1}\frac{1}{n}\mathbf{B}_{p_1}^{\mathrm{T}}(\mathbf{g}_{p_1}-\mathbf{m})\right\|.$$

The order of $\sqrt{\sum_{j=1-p_2}^{N_2}\|\{n^{-1}\sum_{i=1}^{n}B_{j,p_2}(X_i)\varepsilon_iB_{k,p_1}(X_i)\}_{k=1-p_1}^{N_1}\|^2}$ is $O_p(n^{5/2(2p_*+1)-1/2(2p^*+1)-1})$ according to Equation (A8) in Lemma A.5. And with probability approaching 1, (A5) of Lemma A.3 implies that

$$\|(n^{-1}\mathbf{B}_{p_1}^{\mathrm{T}}\mathbf{B}_{p_1})^{-1}n^{-1}\mathbf{B}_{p_1}^{\mathrm{T}}(\mathbf{g}_{p_1}-\mathbf{m})\|\leq C(f)\|n^{-1}\mathbf{B}_{p_1}^{\mathrm{T}}(\mathbf{g}_{p_1}-\mathbf{m})\|,$$

while $\|n^{-1}\mathbf{B}_{p_1}^{\mathrm{T}}(\mathbf{g}_{p_1}-\mathbf{m})\|$ is bounded by

$$\sqrt{\sum_{j=1-p_1}^{N_1}\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_1}(X_i)|g_{p_1}-m|(X_i)\right\}^2}\leq\|\mathbf{g}_{p_1}-\mathbf{m}\|_\infty\sqrt{\sum_{j=1-p_1}^{N_1}\left\{\frac{1}{n}\sum_{i=1}^{n}B_{j,p_1}(X_i)\right\}^2}$$

$$=O_p(h_1^{p_1})\sqrt{\sum_{j=1-p_1}^{N_1}\max_{j=1-p_1}^{N_1}\{\langle B_{j,p_1},1\rangle_n\}^2},$$

which is of order $O_p\left\{h_1^{p_1}\times\sqrt{N_1}\times(h_1^{1/2}+\sqrt{n^{-1}h_1^{-1}\log n})\right\}=O_p(h_1^{p_1})$ by (A7) in Lemma A.4. Combining them, the second term is of order $O_p(h_1^{-1/2}\times n^{5/2(2p_*+1)-1/2(2p^*+1)-1}\times h_1^{p_1})=O_p(n^{5/2(2p_*+1)-1/2(2p^*+1)-3/2})=o_p(n^{3/2(2p_*+1)-1})$. Putting the first and the second term together, we have established that $\sup_{x\in[a,b]}|\mathrm{III}_{p_1}(x)|=O_p(n^{3/2(2p_*+1)-1})$. ■