Journal of Nonparametric Statistics
Vol. 22, No. 8, November 2010, 999–1018

Simultaneous confidence bands for time-series prediction function

Li Wang[a] and Lijian Yang[b]*

[a]*Department of Statistics, University of Georgia, Athens, GA 30602, USA;* [b]*Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA*

(*Received 14 September 2009; final version received 17 December 2009*)

Although many types of confidence bands exist for nonparametric regression with i.i.d. data, theoretical properties of such bands have never been established under dependence. We propose simultaneous confidence bands for nonparametric prediction function of time-series data using spline estimation. Asymptotic properties are established under the assumption of strong mixing, and simulation experiments have provided strong evidence that corroborates with the asymptotic theory. As an application, after removing the environmental Kuznets curve trend effects, the impact of the economic intervention on environmental quality change is quantified for the USA and Japan, with different conclusions.

Keywords: B-spline; Berry–Esseen bound; environmental Kuznets curve; geometric mixing; knots; nonparametric regression

AMS Subject Classification: 62G15; 62G08; 62M10

1. Introduction

Classic regression and time series tools such as generalised linear model and linear autoregression are known to be inadequate for complex data that exhibit nonlinearity. Nonlinear time series analysis offers an approach to detect structure which sometimes remains undetected by traditional parametric estimation techniques. Nonparametric models have become more and more popular over the last two decades. This recognition has motivated the development of non- and semiparametric regression techniques, with far reaching applications (see, for example, Fan and Gijbels 1996; Bosq 1998; Fan and Yao 2003). Two very popular forms of nonparametric regression are kernel/local polynomial type and spline type smoothing.

Theoretical properties of nonparametric smoothers are typically examined in terms of mean square, pointwise, or uniform rate of convergence, while practical consideration favours methods that are easy to implement and interpret. In addition, fast computing is appealing for users of smoothers. For kernel smoothing of independent data, satisfactory results on rates of convergence have been obtained (see Fan and Gijbels 1996 for pointwise and mean square convergence

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

ISSN 1048-5252 print/ISSN 1029-0311 online

DOI: 10.1080/10485251003592575
http://www.informaworld.com

rates, Hall and Titterington 1988; Härdle 1989; Xia 1998 for uniform confidence bands). Spline smoothers of independent data have been investigated in parallel (see, for example, Stone 1985 for mean square convergence and Huang 2003 for pointwise convergence).

Nonparametric smoothing of weakly dependent data has been vigorously pursued in many directions due to its superiority for the modelling and forecasting of nonlinear time series (see, for instance, Beran and Feng 2002; Fan and Yao 2003; Dahl and Levine 2006; Su and Ullah 2006 for kernel-type autoregression smoothing, and Huang and Yang 2004; Xue and Yang 2006; Wang and Yang 2007 for spline-type autoregressive smoothing). For two decades, a lot of research effort has been devoted to the development of statistical tests for structural change in the trend function of a time series (see Juhl and Xiao 2005). Simultaneous confidence bands are the most natural tools for testing hypotheses on the entirety of unknown curves/functions as they contain a whole curve with all of its features by some predetermined confidence level (see McKeague and Zhao 2002, 2006; Peng and Qi 2006). Visually examining the bands summarises the structural change in deterministic trend regressors (see Huang, Wang, Yang, and Kravchenko 2008; Wang and Yang 2009; Song and Yang 2009 for more applications of confidence band methods).

To put the discussion in perspective, consider the question of how the adjustment of GDP autonomously influence the change of the environmental quality in Japan. The logarithm of GDP per capita and the emission per capita of Japan are decomposed as $u(t) + X_t$ and $v(t) + Y_t$, $t = 1, \dots, n$, respectively, where the quadratic trends $u(t)$ and $v(t)$ are given in Equation (22), and $\{(X_t, Y_t)\}_{t=1}^n$ are zero mean stationary time series of residuals. The aforementioned question can be formulated in terms of various hypotheses about the prediction function $m(x) = E(Y_t | X_t = x)$. In Figure 5(b), a 99% conservative simultaneous confidence band of $m(x)$ is plotted together with the linear regression line, clearly showing the nonlinear dependence of Y_t on X_t. The corresponding Figure 4(b) for the USA, however, shows a linear $m(x)$. Making such inference about the global shape of the prediction function $m(x)$ depends crucially on the construction of simultaneous confidence bands for m using the time-series observations $\{(X_i, Y_i)\}_{i=1}^n$, which is the main contribution of this paper.

We formulate the problem broadly for any sequence of strictly stationary bivariate time series data $\{(X_i, Y_i)\}_{i=1}^n$, satisfying

$$Y_i = m(X_i) + \sigma(X_i)\varepsilon_i, \quad i = 1, \dots, n, \tag{1}$$

with m, σ denoting the unknown conditional mean and conditional standard deviation functions on a compact interval $[a, b]$, i.e. $m(x) = E(Y_i|X_i = x)$, $\sigma^2(x) = \mathrm{Var}(Y_i|X_i = x)$, $x \in [a, b]$, where the errors $\{\varepsilon_i\}_{i=1}^n$ are conditional white noise, i.e. $E(\varepsilon_i|X_i) = 0$, $\mathrm{Var}(\varepsilon_i|X_i) = 1$ and ε_i is a martingale difference with respect to the σ-field $\mathcal{F}_i = \sigma\{X_j, \varepsilon_{j-1}, 1 \le j \le i\}$ for $i = 1, \dots, n$. If the data followed a linear regression model, $m(x)$ would be linear in x.

Zhou, Shen, and Wolfe (1998) obtained the confidence band for $m(\cdot)$ based on the spline estimation when the data $\{(X_i, Y_i)\}_{i=1}^n$ forms an i.i.d. sample. However, confidence bands remain unavailable for all nonparametric smoothers based on dependent observations, due to the fact that Hungarian embedding or strong approximation results for dependent random variables are not as sharp as that established by Tusnády (1977) for independent random variables. Existing results on nonparametric confidence bands for i.i.d. data rely on sharpness of such strong approximation (see Bickel and Rosenblatt 1973; Rosenblatt 1976).

In this paper, the piecewise constant and linear confidence bands are obtained for the unknown function $m(x)$ based on the polynomial spline estimation, while the observations $\{(X_i, Y_i)\}_{i=1}^n$ are only assumed to have α-mixing coefficient $\alpha(k)$ decaying geometrically; see Assumption (A4) in Section 2. Instead of applying the usual Hungarian embedding technique in most existing works, we make use of the Berry–Esseen bound in Sunklodas (1984) for sequences of mixing random variables to establish that constructed confidence bands are conservative. The resulting confidence bands are comparable in terms of formula and narrowness to those constructed for i.i.d. sample.

We organise our paper as follows. In Section 2, we state our main findings of spline confidence bands. In Section 3, we provide further insights into the error structure of spline estimators from which we are able to obtain the asymptotic simultaneous confidence bands. This is accomplished by establishing simultaneous Berry–Esseen bound for the estimation noise. Section 4 describes the actual steps to implement the bands. Sections 5 reports our findings in an extensive simulation study. The proposed method is then applied in Section 6 to the environmental Kuznets curves. All technical proofs are contained in the Appendix.

2. Estimation and main results

In general, the regression function $m(x)$ in model (1) is assumed to belong to $C^{(p)}[a,b]$, the space of functions that have pth order continuous derivatives for some integer $p > 0$, on the interval $[a,b]$.

Polynomial spline estimator of $m(x)$ is known for its simple implementation and fast computation, once a knot sequence is determined. To be specific, we divide $[a,b]$ into $(N+1)$ subintervals $J_j = [t_j, t_{j+1}),\ j = 0,\ldots,N-1$, $J_N = [t_N, b]$, where $\{t_j\}_{j=1}^N$ is a sequence of equally spaced points, called interior knots, given as $t_{-(p-1)} = \cdots = t_0 = a < t_1 < \cdots < t_N < b = t_{N+1} = \cdots = t_{N+p}, t_j = a + jh, j = 0, 1, \ldots, N+1$, in which $h = (b-a)/(N+1)$ is the distance between neighbouring knots. For any $x \in [a,b]$, we define its location index $j(x)$ and relative location index $\delta(x)$ as

$$j(x) = j_n(x) = \min\{[h^{-1}(x-a)], N\}, \delta(x) = h^{-1}\{x - t_{j(x)}\}. \tag{2}$$

It is clear that $t_{j(x)} \le x < t_{j(x)+1}$, $0 \le \delta(x) < 1$, $\forall x \in [a,b]$, $j(b) = N$, $\delta(b) = 1$.

Denote $G^{(p-2)} = G^{(p-2)}[a,b]$ the space of all $C^{(p-2)}[a,b]$ functions that are polynomials of degree $p-1$ on each interval. For example, $G^{(-1)}$ is the space of functions that are constant on each J_j, and $G^{(0)}$ is the space of functions that are linear on each J_j and continuous on $[a,b]$. The B-spline basis of $G^{(-1)}$ are indicator functions of intervals J_j, $b_{j,1}(x) = I_j(x) = I_{J_j}(x)$, $j = 0, 1, \ldots, N$. The B-spline basis of $G^{(0)}$ are $b_{j,2}(x) = K\{(x - t_{j+1})h^{-1}\}$, $j = -1, 0, 1, \ldots, N$, where $K(u) = (1-|u|)_+$.

For any functions ϕ, φ on $[a,b]$, define their empirical inner products $\langle \phi, \varphi \rangle_n = n^{-1}\sum_{i=1}^n \{\phi(X_i)\varphi(X_i)\}$ with the norm $\|\phi\|_{2,n}^2 = n^{-1}\sum_{i=1}^n \phi^2(X_i)$. If ϕ and φ are L^2-integrable, we define their theoretical inner product $\langle \phi, \varphi \rangle = E\{\phi(X_i)\varphi(X_i)\}$ with norm $\|\phi\|_2^2 = E\{\phi^2(X)\}$. For theoretical analysis, we use the rescaled B-spline basis: $B_{j,p}(x) \equiv \|b_{j,p}(x)\|^{-1} b_{j,p}(x)$, $p = 1, 2$. Let $\mathbf{V}_{n,p}$ and $\mathbf{V}_p$ be the empirical and theoretical inner product matrices of the B-spline basis $\{B_{j,p}(x)\}_{j=-1}^N$, i.e.

$$\mathbf{V}_{n,p} = (\langle B_{j',p}, B_{j,p} \rangle_n)_{j,j'=1-p}^N, \quad \mathbf{V}_p = (\langle B_{j',p}, B_{j,p} \rangle)_{j,j'=1-p}^N, \qquad p = 1, 2. \tag{3}$$

With a slight abuse of notation, we introduce a function $\mathbf{Y}(X_i) \equiv Y_i$, $1 \le i \le n$. Then the polynomial spline estimator is

$$\begin{aligned} \hat{m}_p(\cdot) &= \arg \min_{g(\cdot) \in G^{(p-2)}[a,b]} \sum_{i=1}^n \{Y_i - g(X_i)\}^2 \\ &= \{B_{j,p}(\cdot)\}_{1-p \le j \le N}^{\mathrm{T}} \mathbf{V}_{n,p}^{-1} \{\langle \mathbf{Y}, B_{j,p} \rangle_n\}_{j=1-p}^N. \end{aligned} \tag{4}$$

Remark 1 It is straightforward that $\langle B_{j',p}, B_{j,p} \rangle_n \equiv 0$ for $|j - j'| > p$, thus $\mathbf{V}_{n,1}$ is a diagonal matrix and $\mathbf{V}_{n,2}$ is a tridiagonal matrix. By Lemma A4, the random matrix $\mathbf{V}_{n,p}$ can be approximated by its deterministic version $\mathbf{V}_p$. Clearly, $\mathbf{V}_1 = \mathbf{I}$, and according to Lemma A2, $\mathbf{V}_2$ is decomposed as a simpler matrix $\mathfrak{J}$ with a distribution-free form given in Equation (19) plus a negligible term.

The width of the bands depends on the pointwise variances of $\hat{m}_p(x)$. Define

$$\sigma_{n,1}^2(x) = \{n\|b_{j(x),1}\|_2^2\}^{-1} \int I_{j(x)}(v)\sigma^2(v)f(v)\,\mathrm{d}v, \tag{5}$$

$$\sigma_{n,2}^2(x) = \frac{1}{n}\sum_{j,j',l,l'=-1}^{N} B_{j',2}(x)B_{l',2}(x)s_{jj'}s_{ll'}\sigma_{jl}, \tag{6}$$

where $j(x)$ is in Equation (2), $s_{jj'}$ is the jj'th entry of the matrix $\mathbf{V}_2^{-1}$ defined in Equation (3), and for $j, l = -1, \ldots, N$, $\sigma_{jl} = \int \sigma^2(v)B_{j,2}(v)B_{l,2}(v)f(v)\,\mathrm{d}v$.

Before giving the form of the spline confidence bands, we first cite some technical assumptions used in the construction of the bands.

(A1) The regression function $m \in C^{(p)}[a, b]$, $p = 1, 2$.
(A2) The density $f(x)$ of X is continuous and positive on its compact support $[a, b]$. The standard deviation $\sigma(x)$ is continuous and positive on $[a, b]$.
(A3) The number of interior knots N satisfies: $(n/\log n)^{1/(2p+1)} \ll N \ll n^{1/3}$, hence for $p = 2$, one can take $N \sim n^{1/5}$, while for $p = 1$, one can take $N \sim n^{1/3}(\log n)^{-1/6}$.
(A4) There exist positive constants K_0 and λ_0 such that $\alpha(k) \le K_0 \mathrm{e}^{-\lambda_0 k}$ holds for all k, where the strong mixing coefficient of order k is defined as

$$\alpha(k) = \sup_{B\in\sigma\{X_s,Y_s,s\le t\},C\in\sigma\{X_s,Y_s,s\ge t+k\}} |P(B\cap C) - P(B)P(C)|, \quad k \ge 1.$$

(A5) The joint distribution of random variables (X, ε) satisfies the following:
  (a) The error is a white noise, $E(\varepsilon|X = x) = 0$, $E(\varepsilon^2|X = x) = 1$.
  (b) There exists $M_0 > 0$ such that $\sup_{x\in[a,b]} E(|\varepsilon|^3|X = x) < M_0$.

Remark 2 Assumptions (A1)–(A5) are typical in the nonparametric smoothing literature (see for instance, Fan and Yao 2003; Huang and Yang 2004; Xue and Yang 2006; Wang and Yang 2007).

THEOREM 2.1 *Under Assumptions* (A1)–(A5), *for any* $\alpha \in (0, 1)$, *an asymptotic* $100(1-\alpha)\%$ *conservative confidence band for* $m(x)$ *over interval* $[a, b]$ *is*

$$\hat{m}_p(x) \pm \sigma_{n,p}(x)\{2p\log(N+1)\}^{1/2}d_n(\alpha/p), \quad p = 1, 2, \tag{7}$$

where $\sigma_{n,p}(x)$ *is given in Equations* (5) *and* (6), *and* $d_n(\alpha)$ *is an inflation correction factor*

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

In other words, for $p = 1, 2$

$$\liminf_{n\to\infty} P\left[m(x) \in \hat{m}_p(x) \pm \sigma_{n,p}(x)\{2p\log(N+1)\}^{1/2}d_n\left(\frac{\alpha}{p}\right), \forall x \in [a, b]\right] \ge 1 - \alpha,$$

Remark 3 The standard deviation function $\sigma_{n,1}(x)$ is replaceable by $\sigma(x)\{f(x)nh\}^{-1/2}$ according to Lemma A5, and function $\sigma_{n,2}(x)$ is replaceable by $\sigma(x)\{2f(x)nh/3\}^{-1/2}\{\mathbf{\Delta}^{\mathrm{T}}(x)\mathbf{L}_{j(x)}\mathbf{\Delta}(x)\}^{1/2}$ according to Lemmas A3 and A8, with matrices $\mathbf{\Delta}(x)$, $\mathbf{L}_j$ defined in Equations (15)–(19).

3. Error decomposition

In this section, we break the polynomial spline estimation error $\hat{m}_p(x) - m(x)$ into a bias term and a noise term, with $\hat{m}_p(x)$ given in Equation (4).

We define a function $\boldsymbol{\epsilon}$ as $\boldsymbol{\epsilon}(X_i) \equiv \sigma(X_i)\varepsilon_i$, $1 \le i \le n$, then the responses $\mathbf{Y} = \mathbf{m} + \boldsymbol{\epsilon}$ with $\mathbf{m} = \{m(X_1), \ldots, m(X_n)\}^{\mathrm{T}}$. Thus $\hat{m}_p(x) = \tilde{m}_p(x) + \tilde{\varepsilon}_p(x)$, where

$$\tilde{m}_p(x) = \{B_{j,p}(x)\}^{\mathrm{T}}_{1-p \le j \le N} \mathbf{V}^{-1}_{n,p} \{\langle \mathbf{m}, B_{j,p}\rangle_n\}^N_{j=1-p}, \tag{9}$$

$$\tilde{\varepsilon}_p(x) = \{B_{j,p}(x)\}^{\mathrm{T}}_{1-p \le j \le N} \mathbf{V}^{-1}_{n,p} \{\langle \boldsymbol{\epsilon}, B_{j,p}\rangle_n\}^N_{j=1-p}. \tag{10}$$

Thus, the estimation error $\hat{m}_p(x) - m(x)$ consists of a bias term $\tilde{m}_p(x) - m(x)$ and a noise term $\tilde{\varepsilon}_p(x)$, such that

$$\hat{m}_p(x) - m(x) = \{\tilde{m}_p(x) - m(x)\} + \tilde{\varepsilon}_p(x). \tag{11}$$

According to the result of de Boor (2001, p. 149) and Huang (2003, Theorem 5.1), the bias term is of the order $\mathrm{O}_p(h^p)$ uniformly over $x \in [a, b]$. Hence the main hurdle of proving Theorem 2.1 is the noise term $\tilde{\varepsilon}_p(x)$ in Equation (10). This is handled by Proposition 3.1.

PROPOSITION 3.1 *Under Assumptions* (A2)–(A5), *for* $\sigma_{n,p}(x)$, $p = 1, 2$, *given in Equations* (5) *and* (6), *and for any* $0 < \alpha < 1$, *one has*

$$\liminf_{n\to\infty} P\left[\sup_{x\in[a,b]} |\sigma^{-1}_{n,p}(x)\tilde{\varepsilon}_p(x)| \le \{2p\log(N+1)\}^{1/2} d_n\left(\frac{\alpha}{p}\right)\right] \ge 1-\alpha. \tag{12}$$

To prove Proposition 3.1, we make use of the following strong approximation theorem of Sunklodas (1984).

LEMMA 3.2 *Let* $\{\xi_i\}^n_{i=1}$ *be an* α*-mixing sequence with mean zero. Denote* $d = \max_{1\le i\le n} \{E|\xi_i|^{2+\delta}\}$, $0 < \delta \le 1$, $S_n = \sum^n_{i=1} \xi_i$, $\sigma^2_n = ES^2_n \ge c_0 n$ *for some* $c_0 \in (0, +\infty)$. *If* $\alpha(n) \le K_0 \mathrm{e}^{-\lambda_0 n}$, $\lambda_0 > 0$, $K_0 > 0$, *then there exist* $c_1 = c_1(K, \delta)$, $c_2 = c_2(K, \delta)$, *such that*

$$\Delta_n = \sup_z |P\{\sigma^{-1}_n S_n < z\} - \Phi(z)| \le c_1 \frac{d}{c_0\sigma^\delta_n}\left\{\frac{\log(\sigma_n/c_0^{1/2})}{\lambda}\right\}^{1+\delta}$$

for any λ *with* $\lambda_1 \le \lambda \le \lambda_2$, *where* $\lambda_1 = c_2\{\log(\sigma_n/c_0^{1/2})\}^b/n$, $b > 2(1+\delta)/\delta$, $\lambda_2 = 4\delta^{-1}(2+\delta)\log(\sigma_n/c_0^{1/2})$.

The next result from Leadbetter, Lindgren, and Rootzén (1983) plays an important role in obtaining the inflation correction factor $d_n(\alpha)$ in Equation (8).

LEMMA 3.3 *As* $N \to \infty$, $[\Phi(\tau/a_N + b_N)]^N \to \exp(-\mathrm{e}^{-\tau})$, *where* $a_N = (2\log N)^{1/2}$, $b_N = (2\log N)^{1/2} - (2\log N)^{-1/2}(\log\log N + \log 4\pi)/2$.

The details of the proof of Proposition 3.1 are given in the Appendix.

4. Implementation

In this section, we describe in detail the procedures to construct the confidence bands in Theorem 2.1. All of our codes have been written in R.

While the B-spline basis is convenient for theoretical analysis, it is easier to work with the truncated power basis for implementation. Given any sample $\{(X_i, Y_i)\}_{i=1}^n$, we use the minimum and maximum values of $\{X_i\}_{i=1}^n$ as the endpoints of interval $[a, b]$. The number of knots N is taken to be $[c_p n^{1/3}(\log n)^{-1/6}]$ for $p = 1$ and $[c_p n^{1/5}]$ for $p = 2$, where c_p $(p = 1, 2)$ are positive integers. As discussed in the previous works on confidence bands (Härdle 1989; Xia 1998), the explicit formula of coverage probability for the bands does not exist, hence there is no optimal method to select c_p $(p = 1, 2)$. Therefore, we have not attempted an adaptive knot selection, as Härdle, Marron, and Yang (1997) had illustrated that it could lead to uniform inconsistency. We have set $c_1 = 6$, $c_2 = 3$ for constant and linear bands, respectively, which works well in all simulations.

When constructing the confidence bands, one needs to estimate the unknown functions $f(x)$ and $\sigma^2(x)$ for the evaluation of the functions $\sigma_{n,1}(x)$ and $\sigma_{n,2}(x)$ in Equations (5) and (6). Let $\hat{f}(x)$ be the Nadaraya–Watson density estimator with the quartic kernel $\tilde{K}(u) = 15(1 - u^2)^2 I\{|u| \leq 1\}/16$ and the rule-of-thumb bandwidth of Silverman (1986) $h_{\text{rot},f} = (4\pi)^{1/10}(140/3)^{1/5} n^{-1/5} s_n$. Define vectors $\mathbf{Z}_p = \{Z_{1,p}, \ldots, Z_{n,p}\}^{\text{T}}$, $p = 1, 2$ with $Z_{i,p} = \{Y_i - \hat{m}_p(X_i)\}^2$, then the spline estimation of $\sigma^2(x)$, $\hat{\sigma}_p^2(x)$, $p = 1, 2$, can be obtained based on the data $\{X_i, Z_{i,p}\}_{i=1}^n$.

The standard deviation function $\sigma_{n,1}(x)$ in Equation (5) is approximated by

$$\hat{\sigma}_{n,1}(x, 1) = (nh)^{-1/2} \hat{\sigma}_1(t_{j(x)}) \hat{f}^{-1/2}(t_{j(x)}), \tag{13}$$

$$\hat{\sigma}_{n,1}(x, 2) = (nh)^{-1/2} \hat{\sigma}_1(x) \hat{f}^{-1/2}(x), \tag{14}$$

where $j(x)$ is the nearest left knot defined in Equation (2). To approximate the standard deviation function $\sigma_{n,2}(x)$ in Equation (6), let

$$\mathbf{\Delta}(x) = \{c_{j(x)-1}\{1 - \delta(x)\}, c_{j(x)}\delta(x)\}^{\text{T}}, \quad c_j = \begin{cases} 1, & 0 \leq j \leq N-1, \\ \sqrt{2}, & j = -1, N. \end{cases} \tag{15}$$

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

$$\hat{\sigma}_{n,2}(x, 1) = \{\mathbf{\Delta}^{\text{T}}(x) \mathbf{L}_{j(x)} \mathbf{\Delta}(x)\}^{1/2} \{nh \hat{f}(t_{j(x)})\}^{-1/2} \sqrt{3/2}\, \hat{\sigma}_2(t_{j(x)}), \tag{16}$$

$$\hat{\sigma}_{n,2}(x, 2) = \{\mathbf{\Delta}^{\text{T}}(x) \mathbf{L}_{j(x)} \mathbf{\Delta}(x)\}^{1/2} \{nh \hat{f}(x)\}^{-1/2} \sqrt{3/2}\, \hat{\sigma}_2(x), \tag{17}$$

where

$$\mathbf{L}_j = \begin{pmatrix} l_{j-1,j-1} & l_{j-1,j} \\ l_{j,j-1} & l_{j,j} \end{pmatrix}, \quad j = 0, 1, \ldots, N \tag{18}$$

with $\{l_{ik}\}_{|i-k| \leq 1}$ being the $(i+2, k+2)$th entry of the inverse of matrix

$$\mathfrak{J} = \begin{pmatrix} 1 & \sqrt{2}/4 & & & & 0 \\ \sqrt{2}/4 & 1 & 1/4 & & & \\ & 1/4 & 1 & \ddots & & \\ & & \ddots & \ddots & 1/4 & \\ & & & 1/4 & 1 & \sqrt{2}/4 \\ 0 & & & & \sqrt{2}/4 & 1 \end{pmatrix}. \tag{19}$$

Remark 4 According to Lemma A2, $\mathfrak{J}$ is a simpler and distribution-free approximation of the inner product matrix $\mathbf{V}_2$ defined in Equation (3). To obtain the matrix $\mathbf{L}_j$ in Equation (18), we need to find an easy way to calculate the tridiagonal terms of the matrix $\mathfrak{J}^{-1}$. Since $\mathfrak{J}$ is a symmetric and tridiagonal Jacobi matrix, these terms can be easily and quickly computed through Lemma 4.1, which is a direct result of Zhang (1999, Theorem 4.5).

LEMMA 4.1 *Let* $z_1 = (2+\sqrt{3})/4$, $z_2 = (2-\sqrt{3})/4$, $\theta = z_2/z_1 = 7-4\sqrt{3}$, *one can compute the terms* $l_{i,k} = l_{k,i}$, $|i-k| \le 1$ *in Equation* (18) *by the following formulae*

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

$$l_{k,k} = \frac{\{8z_1(1-\theta^{N+2-k}) - (1-\theta^{N+1-k})\}\{8z_1(1-\theta^{k-1}) - (1-\theta^{k-2})\}}{(z_1-z_2)\{64z_1^2(1-\theta^{N+1}) - 16z_1(1-\theta^N) + (1-\theta^{N-1})\}},$$

for $2 \le k \le N+1$*;*

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

$$l_{k,k+1} = -\frac{\{8z_1(1-\theta^{N+1-k}) - (1-\theta^{N-k})\}\{8z_1(1-\theta^{k-1}) - (1-\theta^{k-2})\}}{4z_1(z_1-z_2)\{64z_1^2(1-\theta^{N+1}) - 16z_1(1-\theta^N) + (1-\theta^{N-1})\}},$$

for $2 \le k \le N$. *In particular, there exists a constant* $c_l > 0$ *such that* $\max_{|i-k|\le 1} |l_{ik}| \le c_l$.

Now one can compute the following confidence bands

$$\hat{m}_p(x) \pm \hat{\sigma}_{n,p}(x, \mathrm{opt})\{2p\log(N+1)\}^{1/2} d_n(\alpha/p), \quad p = 1, 2, \ \ \mathrm{opt} = 1, 2, \tag{20}$$

where $\hat{m}_p(x)$ is given in Equation (4), the additional parameter opt $= 1, 2$ indicating the estimation being at each value x or at the nearest left knot.

5. Simulation

To illustrate the finite-sample behaviour of the confidence bands, we present some simulation results. The data are generated from the heteroscedastic regression model (1), with

$$m(x) = \sin(2\pi x), \quad \sigma(x) = \sigma_0 \frac{100 - \exp(x)}{100 + \exp(x)}, \quad \varepsilon \sim N(0,1), \quad \sigma_0 = 0.2, 0.5.$$

We simulate $\{T_i\}_{i=1}^n$ from a moving average sequence of the order q, i.e.,

$$T_i = \frac{(\xi_i + \theta_1\xi_{i-1} + \theta_2\xi_{i-2} + \cdots + \theta_q\xi_{i-q})}{\sqrt{1+\theta_1^2+\cdots+\theta_q^2}},$$

where in our simulation, q is taken to be 4, $\theta_1 = \cdots = \theta_q = 0.2$ and ξ_i's are i.i.d. r.v.'s $\sim N(0,1)$. We then define $X_i = \Phi(T_i)$, where Φ is the standard normal distribution function, so X_i is uniformly distributed on [0, 1].

We choose the sample size n to be 100, 200, 500 and 10, 000, confidence level $1-\alpha = 0.99, 0.95$ as usual. Table 1 contains the coverage probabilities as the percentage of coverage of the true curve at all data points by the confidence bands in Equation (20) with 10, 000 replications

Table 1. Piecewise-constant spline bands coverage probabilities for 10, 000 replications.[a]

Noise level	Sample size	Confidence level	Constant bands	Linear bands
0.2	100	0.99	0.5954 (0.6609)	0.9878 (0.9985)
		0.95	0.3417 (0.3854)	0.9583(0.9938)
	200	0.99	0.6749 (0.7356)	0.9935(0.9999)
		0.95	0.4175 (0.4632)	0.9754 (0.9990)
	500	0.99	0.8062 (0.8334)	0.9939(1.0000)
		0.95	0.5325 (0.5656)	0.9728 (0.9993)
0.5	100	0.99	0.6053 (0.7462)	0.9856 (0.9991)
		0.95	0.3728 (0.5200)	0.9581 (0.9944)
	200	0.99	0.7221 (0.8014)	0.9939(1.0000)
		0.95	0.5022 (0.5904)	0.9790 (0.9987)
	500	0.99	0.8752 (0.9171)	0.9960(1.0000)
		0.95	0.6961 (0.7549)	0.9853 (0.9996)

[a]The numbers outside/inside of the parentheses are based on estimating $\{\sigma(X_i)\}_{i=1}^n$ from Equation (13)/(14) for constant bands and from Equation (16)/(17) for linear bands.

of sample size n. The coverage percentages show very positive confirmation of Theorem 2.1. From Table 1, it is obvious that larger sample size guarantees improved coverage for both constant and linear confidence bands, while reasonable coverage has also been achieved at moderate sample sizes for linear bands. Under the same circumstances, the band by the linear spline performs much

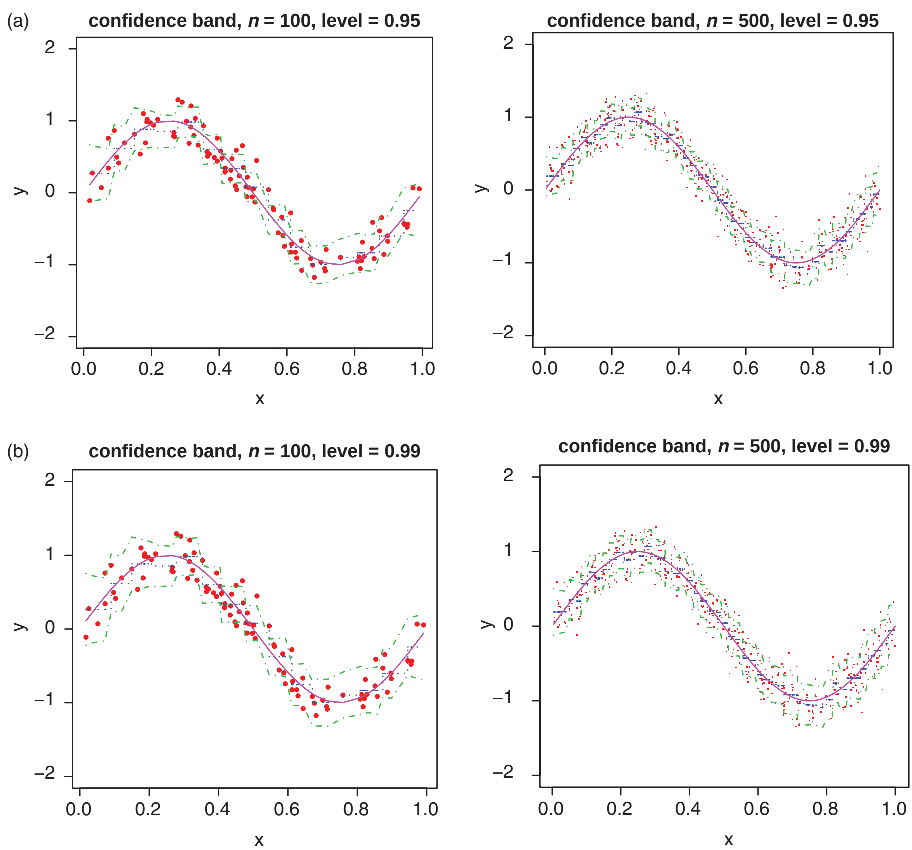

Figure 1. Plots of the piecewise constant confidence bands (upper and lower dashed curves), the estimator $\hat{m}_1(x)$ (centre dotted curve), the true function $m(x) = \sin(2\pi x)$ (centre smooth solid curve), and the data (dots). The bands are computed from Equation (20) with opt = 2: (a) confidence level = 95% (b) confidence level = 99%.

better than the band by the constant spline. We also observe that the noise level has more influence on the constant bands coverage, and very little on the linear bands. At sample size 100, regardless of the noise level, both of the two piecewise linear bands in Equation (20) achieve at least 0.9856 and 0.9581 for confidence level $1-\alpha = 0.99$ and 0.95, respectively. Therefore, in practice, we would recommend to using the linear bands.

Piecewise constant bands (Figure 1) and piecewise linear bands (Figure 2) are created for graphical comparison at the noise level 0.2, each with four types of symbols: dots (data), centre smooth solid line (true curve), centre dotted line (the spline estimated curve), upper and lower thick solid line (confidence bands). In both figures, the confidence bands of $n = 500$ are thinner and fits better than those of $n = 100$, and the smaller the significance level, the wider the confidence band. Overall, the linear bands are superior to the constant ones in terms of smoothness and narrowness. In addition, we find that the estimation of $\sigma_{n,p}(x)$ $(p = 1, 2)$ by $\hat{\sigma}_{n,p}(x, 1)$ at knots as in Equation (13) $(p = 1)$ and Equation (16) $(p = 2)$ or by $\hat{\sigma}_{n,p}(x, 2)$ at all observations as in Equation (14) $(p = 1)$ and Equation (17) $(p = 2)$ does not seem to have much noticeable impact on the widths of the confidence bands.

For the linear bands, we have also carried out the simulation at noise level 0.2, for $n = 10,000$ and opt $= 1$ (estimation on knots). The coverage is always 99.6% 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

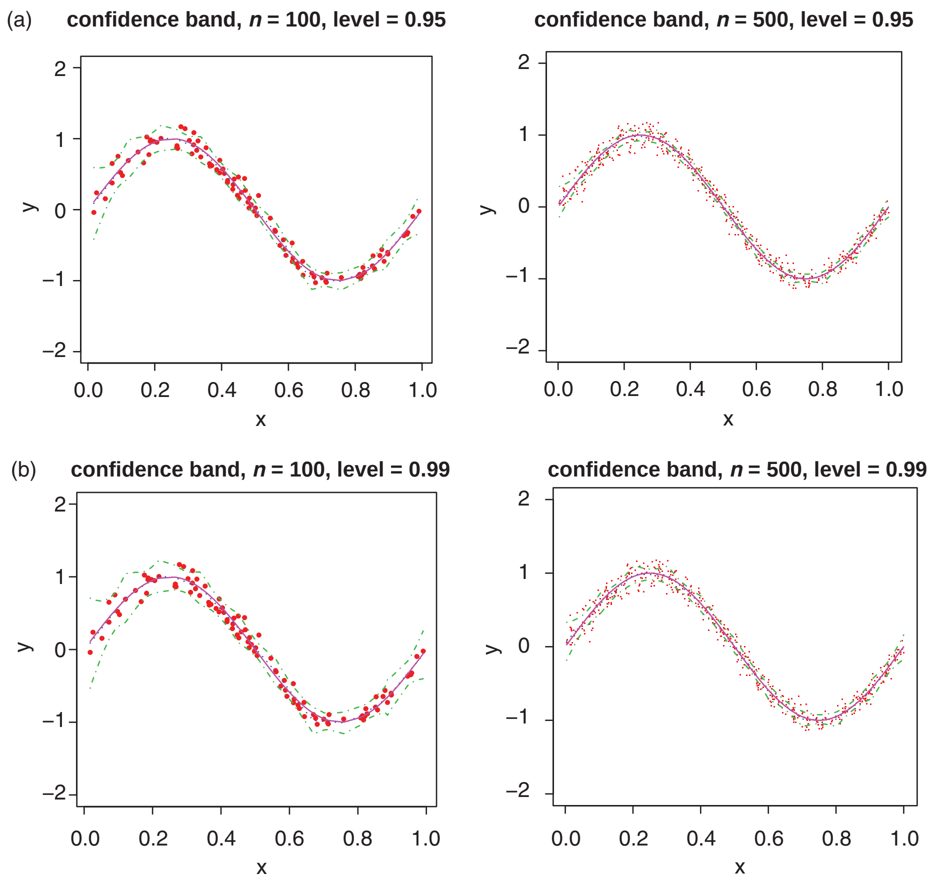

Figure 2. Plots of the piecewise linear confidence bands (upper and lower dashed curves), the estimator $\hat{m}_2(x)$ (centre dotted curve), the true function $m(x) = \sin(2\pi x)$ (centre smooth solid curve), and the data (dots). The bands are computed from Equation (20) with opt $= 2$: (a) confidence level $= 95\%$ (b) confidence level $= 99\%$.

definitions. Remarkably, it takes merely 0.73 s to run 1 replication with n as large as 10, 000 on a regular PC. This is extremely fast considering that the nonparametric regression is done without WARPing (see Härdle, Hlávka and Klinke 2000).

6. Application

The environmental Kuznets curve (EKC), an inverted-U relationship between pollution and income, is an influential generalisation about the way environmental quality changes as a country makes the transition from poverty to relative affluence. The EKC predicts that pollution will first increase, but subsequently decline if income growth proceeds far enough. The shape of the relationship between the rate of environmental degradation and GDP per capita has been the subject of much empirical examination. Several studies have attempted to test the EKC hypothesis empirically. The majority of these studies use panel data in conjunction with a static fixed and/or random effects panel estimator. In this paper, we examine whether or not countries (here we select the USA and Japan) actually behave like the EKC, and we further look at the nonparametric time-series nature of the data set after elimination of the trend.

One key variable of this study, the environment index, is the emission of sulphur from 1850 to 1990 (Lefohn, Husar and Husar 1999). The other key variable is GDP per capita, which can be obtained in Maddison (2003). To gain an insight into the model structure, we decompose the logarithm of GDP and emission series into their trend parts and noise parts, i.e.

$$\{\log(\text{GDP per capita})\}_t = u(t) + X_t, \{\log(\text{Emission per capita})\}_t = v(t) + Y_t,$$

for $t = 1, \ldots, n$. We are interested in two sets of hypotheses, given here separately in terms of the relationship between the trends $u(t)$ and $v(t)$, and between the stationary noise $\{X_t\}_{t=1}^n$ and $\{Y_t\}_{t=1}^n$.

EKC hypothesis: There exists an inverted-U shape relationship between $u(t)$ and $v(t)$ (Figure 3).

Residual/noise hypothesis: There exists a linear relationship between $\{X_t\}_{t=1}^n$ and $\{Y_t\}_{t=1}^n$.

The EKC hypothesis can be tested by performing a routine trend analysis. After detrending, $\{X_t\}_{t=1}^n$ and $\{Y_t\}_{t=1}^n$ are obtained, then one can estimate the regression relationship and construct a piecewise linear spline confidence band for the testing.

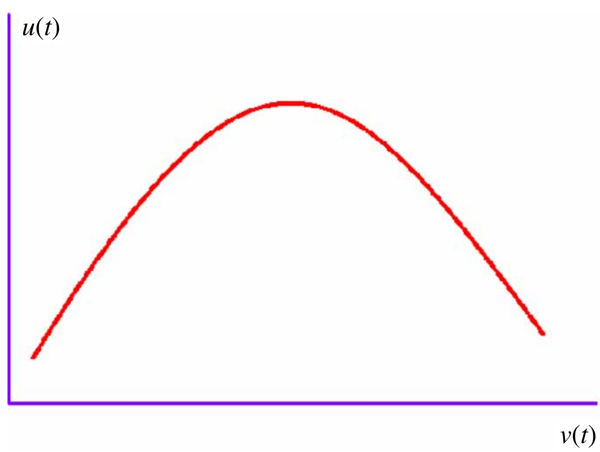

Figure 3. Plot of the EKC in terms of $u(t)$ and $v(t)$.

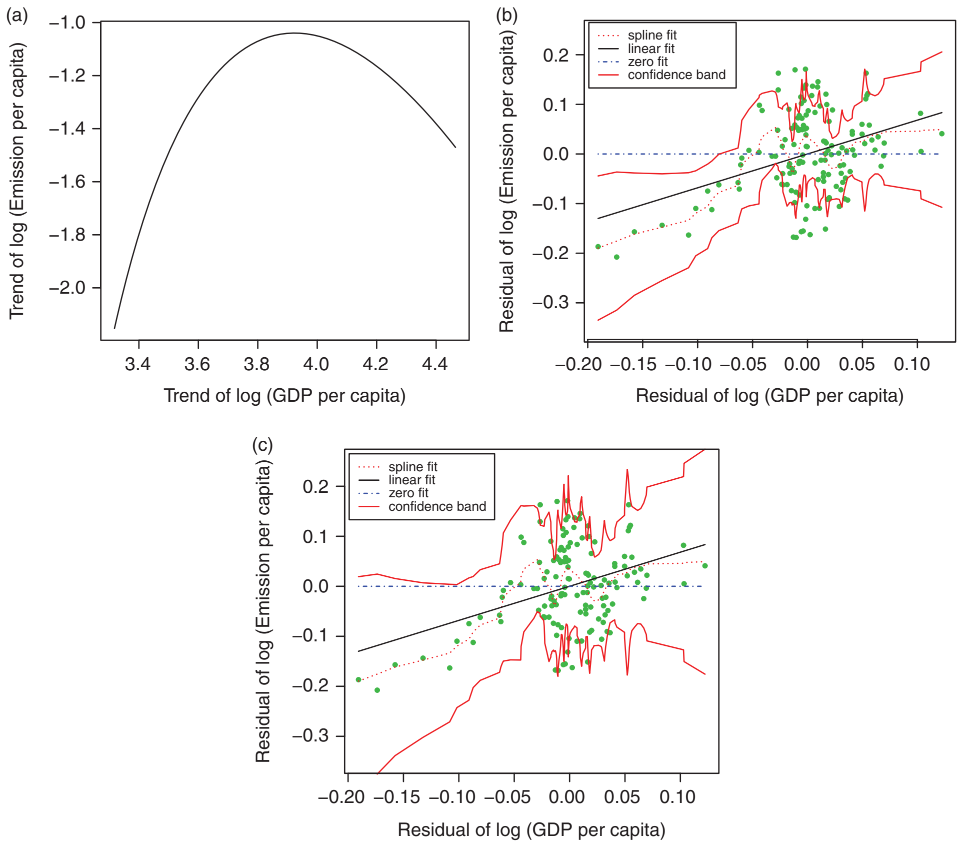

Figure 4. Plots of the USA: (a) trend of $\{(u(t), v(t)), t = 1, \ldots, n\}$ with $u(t)$, $v(t)$ given in Equation (21); (b) residuals of the EKC with 87% confidence band; (c) residuals with 99% confidence band.

We get the trends, $u(t)$ and $v(t)$, by fitting a polynomial regression on time t

$$u(t) = 0.0051t + 3.3127, \quad v(t) = -0.0001t^2 + 0.0261t - 2.1788, \tag{21}$$

where the corresponding $R^2 = 0.9814$ and 0.9256. Therefore, for the USA, the EKC hypothesis is retained by the trend analysis, see Figure 4(a). After the elimination of the trend, $\{X_t\}_{t=1}^n$, $\{Y_t\}_{t=1}^n$ appear to be stationary. For the residual hypothesis, Figure 4(b) shows that when the confidence level is as small as 87%, the linear regression line is still covered by the confidence band. This phenomenon implies that the residual hypothesis is retained and in fact the p-value is 0.1390. However, as we have seen in Figure 4(b) that the confidence bands cannot cover the horizontal line $E(Y_t|X_t) \equiv 0$ at the 87% level. Therefore, we increase the confidence level to construct a wider band. At the 99% level, the horizontal line $E(Y_t|X_t) \equiv 0$ is then completely covered by the bands, see Figure 4(c). Furthermore, the p-value is found to be 0.0129. Therefore, we reject the hypothesis that Y_t is unpredictable from X_t at significance level 0.05. This phenomenon implies that the intervention of emission is not immune to the intervention of economy, and the adjustment of GDP has autonomous linear influence on the change of environmental quality.

The quadratic trends $u(t)$, $v(t)$ for Japan data are given as

$$u(t) = 0.0003t^2 - 0.0019t + 6.7308, \quad v(t) = -0.0005t^2 + 0.0952t - 9.0772 \tag{22}$$

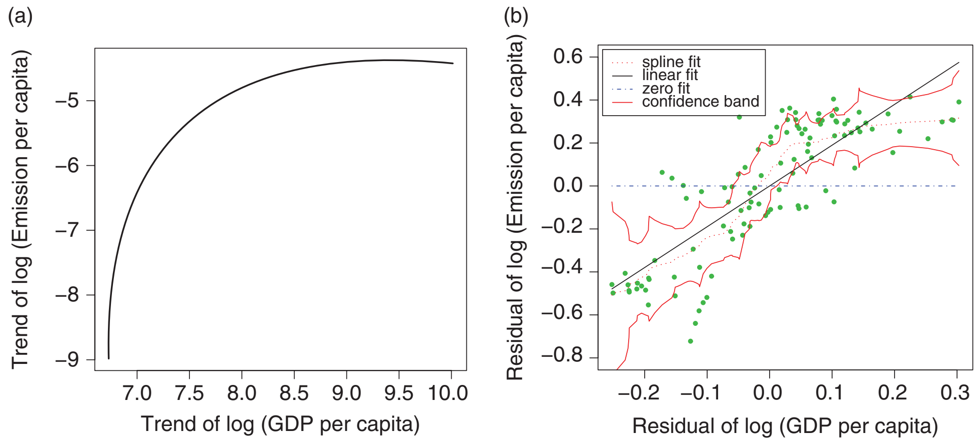

Figure 5. Plots of Japan: (a) trend of $\{(u(t), v(t)), t = 1, \ldots, n\}$ with $u(t), v(t)$ given in Equation (22); (b) residuals of the EKC with 99% confidence band.

with $R^2 = 0.9829, 0.9544$. From the trend relationship curve, Figure 5(a), one sees that it is not a U-shaped curve as EKC predicted. However, we are not sure whether it would succeed to decouple environmental pollution and resource use from economic growth, which will make this a tuning point and U shape later. To test the residual hypothesis, Figure 5(b) shows that neither the linear regression line nor the horizontal line $E(Y_t|X_t) \equiv 0$ can be covered by the confidence bands even when the confidence level reaches 99%, and the p-value is found to be 0.0001. Hence at the significance level 0.05, we reject both the linear hypothesis and the nonpredictable hypothesis. This implies that the adjustment of GDP has autonomous influence on the change of environmental quality, but not in a linear way.

Acknowledgement(s)

This work was partially supported by NSF awards DMS 0405330, DMS 0706518 and DMS 0905730. The constructive comments by two referees are gratefully acknowledged.

References

Beran, J., and Feng, Y. (2002), ‘Local Polynomial Fitting with Long-memory, Short-memory and Antipersistent Errors’, *Annals of the Institute of Statistical Mathematics*, 54, 291–311.
Bickel, P.J., and Rosenblatt, M. (1973), ‘On Some Global Measures of the Deviations of Density Function Estimates’, *The Annals of Statistics*, 1, 1071–1095.
de Boor, C. (2001), *A Practical Guide to Splines*, New York: Springer.
Bosq, D. (1998), *Nonparametric Statistics for Stochastic Processes*, New York: Springer.
Dahl, C., and Levine, M. (2006), ‘Nonparametric Estimation of Volatility Models with Serially Dependent Innovations’, *Statistics and Probability Letters*, 18, 2007–2016.
Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, London: Chapman and Hall.
Fan, J., and Yao, Q. (2003), *Nonlinear Time Series: Nonparametric and Parametric Methods*, New York: Springer.
Gut, A. (2005) *TProbability: A graduate course*, New York: Springer.
Hall, P., and Titterington, D.M. (1998), ‘On Confidence Bands in Nonparametric Density Estimation and Regression’, *Journal of Multivariate Analysis*, 27, 228–254.
Härdle, W. (1989), ‘Asymptotic Maximal Deviation of M-smoothers’, *Journal of Multivariate Analysis*, 29, 163–179.
Härdle, W., Marron, J.S., and Yang, L. (1997), ‘Discussion of “Polynomial Splines and Their Tensor Products in Extended Linear Modeling” by Stone et al.’, *The Annals of Statistics*, 25, 1443–1450.
Härdle, W., Hlávka, Z., and Klinke, S. (2000), *XploRe Application Guide*, Heidelberg: Springer.
Huang, J.Z. (2003), ‘Local Asymptotics for Polynomial Spline Regression’, *The Annals of Statistics*, 31, 1600–1635.
Huang, J.Z., and Yang, L. (2004), ‘Identification of Nonlinear Additive Autoregressive Models’, *Journal of the Royal Statistical Society Series B*, 66, 463–477.

Huang, X. Wang, L. Yang, L., and Kravchenko, A.N. (2008), 'Management Practice Effects on Relationships of Grain Yields with Topography and Precipitation', *Agronomy Journal*, 100, 1463–1471.
Johnson, R.A., and Wichern, D.W. (1992), *Applied Multivariate Statistical Analysis*, Englewood Cliffs, NJ: Prentice Hall.
Juhl, T. and Xiao, Z. (2005), 'A Nonparametric Test for Changing Trends', *Journal of Econometrics*, 127, 179–199.
Leadbetter, M.R. Lindgren, G., and Rootzén, H. (1983), *Extremes and Related Properties of Random Sequences and Processes*, New York: Springer.
Lefohn, A.S., Husar, J.D., and Husar, R.B. (1999), 'Estimating Historical Anthropogenic Global Sulfur Emission Patterns for the Period 1850–1990', *Atmospheric Environment*, 33, 3435–3444.
Maddison, A. (2003), *The World Economy: Historical Statistics*, Paris: OECD.
McKeague, I.W., and Zhao, Y. (2002), 'Simultaneous Confidence Bands for Ratios of Survival Functions Via Empirical Likelihood', *Statistics and Probability Letters*, 60, 405–415.
McKeague, I.W., and Zhao, Y. (2006), 'Width-scaled Confidence Bands for Survival Functions', *Statistics and Probability Letters*, 76, 327–339.
Peng, L., and Qi, Y. (2006), 'Confidence Regions for High Quantiles of a Heavy Tailed Distribution', *The Annals of Statistics*, 34, 1964–1986.
Rosenblatt, M. (1976), 'On the Maximal Deviation of k-Dimensional Density Estimates', *The Annals of Probability*, 4, 1009–1015.
Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.
Song, Q., and Yang, L. (2009), 'Spline Confidence Bands for Variance Function', *Journal of Nonparametric Statistics*, 21, 589–609.
Stone, C.J. (1985), 'Additive Regression and Other Nonparametric Models', *The Annals of Statistics*, 13, 689–705.
Su, L., and Ullah, A. (2006), 'More Efficient Estimation in Nonparametric Regression with Nonparametric Autocorrelated Errors', *Econometric Theory*, 22, 98–126.
Sunklodas, J. (1984), 'On the Rate of Convergence in the Central Limit Theorem for Strongly Mixing Random Variables', *Lithuanian Mathematical Journal*, 24, 182–190.
Tusnády, G. (1977), 'A Remark on the Approximation of the Sample df in the Multidimensional Case', *Periodica Mathematica Hungarica*, 8, 53–55.
Wang, L., and Yang, L. (2007), 'Spline-Backfitted Kernel Smoothing of Nonlinear Additive Autoregression Model', *The Annals of Statistics*, 35, 2474–2503.
Wang, J., and Yang, L. (2009), 'Polynomial Spline Confidence Bands for Regression Curves', *Statistca Sinica*, 19, 325–342.
Xia, Y. (1998), 'Bias-corrected Confidence Bands in Nonparametric Regression', *Journal of the Royal Statistical Society Series B* 60, 797–811.
Xue, L., and Yang, L. (2006), 'Additive Coefficient Modelling Via Polynomial Spline', *Statistca Sinica*, 16, 1423–1446.
Zhang, F. (1999), *Matrix Theory: Basic Results and Techniques*, New York: Springer.
Zhou, S. Shen, X., and Wolfe, D.A. (1998), 'Local Asymptotics of Regression Splines and Confidence Regions', *The Annals of Statistics*, 26, 1760–1782.

Appendix 1. Proofs

In this section, we give the proofs of Proposition 3.1 and Theorem 2.1. For notation simplicity, we denote $\|\cdot\|_\infty$ the supremum norm of a function r on $[a,b]$, i.e. $\|r\|_\infty = \sup_{x\in[a,b]} |r(x)|$. The moduli of continuity of a continuous function r on $[a,b]$ is denoted as $\omega(r,h) = \max_{x,x'\in[a,b],|x-x'|\le h} |r(x) - r(x')|$. By the uniform continuity of r on an interval $[a,b]$, one has $\lim_{h\to 0} \omega(r,h) = 0$. Let $|\mathbf{T}|$ be the maximal absolute value of any matrix $\mathbf{T}$.

A.1. *Preliminaries*

The following lemma shows the properties of the norms and inner products of the original B-spline basis and the proof consists of direct algebraic verifications.

LEMMA A.1 *As $n \to \infty$, one has*

$$\|b_{j,1}\|_2^2 = f(t_j)h(1 + r_{j,n,1}), \quad \|b_{j,2}\|_2^2 = \frac{2f(t_{j+1})h}{3}\begin{cases} 1 + r_{j,n,2}, & 0 \le j \le N-1, \\ \frac{1}{2} + r_{j,n,2}, & j = -1, N, \end{cases}$$

$$\langle b_{j,1}, b_{j',1}\rangle = \begin{cases} 1, & j = j', \\ 0, & j \ne j', \end{cases} \quad \langle b_{j,2}, b_{j',2}\rangle = \frac{1}{6}f(t_{j+1})h\begin{cases} 1 + \tilde r_{j,n,2}, & |j'-j| = 1, \\ 0, & |j'-j| > 1, \end{cases}$$

where $\max_{0\le j\le N} |r_{j,n,1}| + \max_{-1\le j\le N} |r_{j,n,2}| + \max_{-1\le j\le N-1} |\tilde r_{j,n,2}| \le C\omega(f,h)$ *and*

$$\frac{1}{3}f(t_{j+1})h\{1 - C\omega(f,h)\} \le d_{j,n} \le \frac{2}{3}f(t_{j+1})h\{1 + C\omega(f,h)\}.$$

LEMMA A.2 *For the theoretical inner product matrices in Equation* (3)*, one has* $\mathbf{V}_1 = \mathbf{I}$*, and* $\mathbf{V}_2$ *has the following decomposition*

$$\mathbf{V}_2 = \mathfrak{J} + (\tilde{v}_{2,jj'})_{jj'=-1}^{N} = \mathfrak{J} + \tilde{\mathbf{V}}_2,$$

where $\mathfrak{J}$ *is given in Equation* (19), $\tilde{v}_{2,jj'} \equiv 0$ *if* $|j - j'| \geq 1$ *and* $\tilde{\mathbf{V}}_2 \leq C\omega(f, h)$.

The lemma follows directly from Lemma A1, thus omitted. Define the inverse matrix of $\mathbf{V}_2$ and its 2×2 diagonal submatrices as

$$\mathbf{S} = (s_{j'j})_{j,j'=-1}^{N} = \mathbf{V}_2^{-1}, \quad \mathbf{S}_j = \begin{pmatrix} s_{j-1,j-1} & s_{j-1,j} \\ s_{j,j-1} & s_{j,j} \end{pmatrix}, \quad j = 0, \ldots, N. \tag{A1}$$

The next lemma ensures that one can approximate $\mathbf{S}$ with the inverse of $\mathfrak{J}$, with a simpler distribution-free form in Equation (19). This approximation is uniform for $\mathbf{S}_j$ in Equation (A1) and $\mathbf{L}_j$ in Equation (18) as well.

LEMMA A.3 *As* $n \to \infty$, $|\mathfrak{J}^{-1} - \mathbf{S}| \to 0$ *and* $\max_{0\leq j\leq N} |\mathbf{L}_j - \mathbf{S}_j| \to 0$.

Proof As given in Equation (19), $\mathfrak{J}$ has diagonal elements 1, and the sum of the absolute values of off-diagonal elements in each row does not exceed $1/\sqrt{2}$. Thus, there exist constants $c, C > 0$, independent of n, such that with probability approaching 1,

$$c|\boldsymbol{\xi}| \leq |\mathfrak{J}\boldsymbol{\xi}| \leq C|\boldsymbol{\xi}|, \quad C^{-1}|\boldsymbol{\xi}| \leq |\mathfrak{J}^{-1}\boldsymbol{\xi}| \leq c^{-1}|\boldsymbol{\xi}|, \quad \forall \boldsymbol{\xi} \in R^{N+2}. \tag{A2}$$

By definition, $\mathfrak{J}\mathfrak{J}^{-1} = \mathbf{I} = \mathbf{V}_2\mathbf{S} = (\mathfrak{J} + \tilde{\mathbf{V}}_2)\mathbf{S}$. Denote $\mathbf{e}_i$ the unit vector with ith element 1, then by Equation (A2), we have

$$\begin{aligned} c|\mathfrak{J}^{-1} - \mathbf{S}| &= c \max_{i=1,\ldots,N+2} |(\mathfrak{J}^{-1} - \mathbf{S})\mathbf{e}_i| \leq \max_{i=1,\ldots,N+2} |\mathfrak{J}(\mathfrak{J}^{-1} - \mathbf{S})\, \mathbf{e}_i| \\ &= |\mathfrak{J}(\mathfrak{J}^{-1} - \mathbf{S})| = |\tilde{\mathbf{V}}\mathbf{S}| \leq |\tilde{\mathbf{V}}_2||\mathbf{S}| \leq |\tilde{\mathbf{V}}_2|(|\mathfrak{J}^{-1} - \mathbf{S}| + |\mathfrak{J}^{-1}|). \end{aligned}$$

According to Lemma A2, as $n \to \infty$,

$$|\mathfrak{J}^{-1} - \mathbf{S}| \leq \frac{C\omega(f, h)}{c - C\omega(f, h)} |\mathfrak{J}^{-1}| = O\{\omega(f, h)\} \to 0.$$

Now by the definition of submatrices $\mathbf{S}_j$ and $\mathbf{L}_j$, $\max_{0\leq j\leq N} |\mathbf{L}_j - \mathbf{S}_j| \leq |\mathfrak{J}^{-1} - \mathbf{S}|$, the lemma follows. ∎

Next, we establish the uniform rate at which the empirical inner product approximates the theoretical inner product for the rescaled B-splines.

LEMMA A.4 *Under Assumptions* (A3) *and* (A5)*, we have*

$$\begin{aligned} A_{n,1} &= \sup_{0\leq j\leq N} |\|B_{j,1}\|_{2,n}^2 - 1| = \mathrm{O}_p\{(nh)^{-1/2} \log n\}, \\ A_{n,2} &= \sup_{g_1,g_2\in G^{(0)}} \left| \frac{\langle g_1, g_2\rangle_n - \langle g_1, g_2\rangle}{\|g_1\|_2\|g_2\|_2} \right| = \mathrm{O}_p\{(nh)^{-1/2} \log n\}. \end{aligned} \tag{A3}$$

Proof For brevity, we only give the proof for $A_{n,1}$. For any $0 \leq j \leq N$, let $\eta_{i,j} = B_{j,1}^2(X_i) - 1$, then $\|B_{j,1}\|_{2,n}^2 - 1 = n^{-1} \sum_{i=1}^{n} \eta_{i,j}$, with $E\eta_{i,j} = 0$ while for any integer $r \geq 2$, C_r inequality (Gut 2005, p. 127, Theorem 2.2) implies that

$$E|\eta_{i,j}|^r = E|B_{j,1}^2(X_i) - 1|^r \leq 2^{r-1}E[B_{j,1}^{2r}(X_i) + 1] \leq C_0\{2h^{-1}\}^{r-1}.$$

On the other hand $E(\eta_{i,j}^2) \geq E[B_{j,1}^4(X_i) - 1] = \{2\|b_{j,1}\|_2^2\}^{-1} - 1 \geq C_1h^{-1}$. Therefore, there is a constant c, such that for all $k > 2$, $E|\eta_{i,j}|^k \leq (ch)^{k-2}k!E\eta_{i,j}^2$. Thus, Cramer's condition is satisfied with Cramer's constant equal to ch. Applying the Bernstein inequality to $n^{-1} \sum_{i=1}^{n} \eta_{i,j}$, for any $\delta > 0$, $q \in [1, n/2]$, one has for $k = 3$

$$P\left\{\frac{1}{n}|\sum_{i=1}^{n} \eta_{i,j}| > \delta_n\right\} \leq a_1 \exp\left(\frac{-q\delta_n^2}{25m_2^2 + 5c\delta_n}\right) + a_2(3)\alpha\left(\left[\frac{n}{q+1}\right]\right)^{6/7},$$

where $\delta_n = \delta \log n/\sqrt{nh}$, $a_1 = 2n/q + 2\{1 + \delta_n^2/(25m_2^2 + 5c\delta_n)\}$, $m_2^2 = E\eta_{i,j}^2 \sim h^{-1}$, $a_2(3) = 11n(1 + 5m_3^{6/7}/\delta_n)$, $m_3 = \max_{1\leq i\leq n} \|\eta_{i,j}\|_3 \leq \{C_0(2h^{-1})^2\}^{1/3}$. Observe that $\delta_n = \mathrm{o}(1)$, then by taking q such that $[n/(q+1)] \geq c_0 \log n$,

$q \geq c_1 n/\log n$ for some constants $c_0, c_1, a_1 = \mathrm{O}(n/q) = \mathrm{O}(\log n), a_2(3) = \mathrm{o}(n^2)$. Assumption (A4) yields that

$$\alpha([n(q+1)^{-1}])^{6/7} \leq \{K_0 \exp(-\lambda_0[n(q+1)^{-1}])\}^{6/7} \leq Cn^{-6\lambda_0 c_0/7}.$$

Thus, for n large enough,

$$P\left\{\frac{1}{n}\left|\sum_{i=1}^{n}\eta_{i,j}\right| > \frac{\delta \log n}{\sqrt{nh}}\right\} \leq c \log n \exp\{-c_2\delta^2 \log n\} + Cn^{2-6\lambda_0 c_0/7}.$$

Taking c_0, δ large enough, for large n, $P\{(1/n)|\sum_{i=1}^{n}\eta_{i,j}| > (nh)^{-1/2}\delta \log n\} \leq n^{-3}$. Hence Equation (A3) holds because

$$\sum_{n=1}^{\infty} P\left\{\sup_{0\leq j\leq N} |\|B_{j,1}\|_{2,n}^2 - 1| > \frac{\delta \log n}{\sqrt{nh}}\right\} \leq \sum_{n=1}^{\infty} n^{-3}N \leq \sum_{n=1}^{\infty} 2n^{-2} < \infty.$$ ■

For the empirical and theoretical inner product matrices $\mathbf{V}_{n,p}$ and $\mathbf{V}_p$ in Equation (3), it follows immediately from Lemma A4 that

$$|\mathbf{V}_{n,p} - \mathbf{V}_p| \leq A_{n,p}, \quad p = 1, 2 \tag{A4}$$

A.2. *Proofs of Proposition* 3.1 *and Theorem* 2.1 *for constant spline*

Note that $\tilde{\varepsilon}_1(x)$ in Equation (10) can be rewritten as $\tilde{\varepsilon}_1(x) = \sum_{j=0}^{N} \varepsilon_j^* B_{j,1}(x)\|B_{j,1}\|_{2,n}^{-2}$, with $\varepsilon_j^* = \langle \boldsymbol{\epsilon}, B_{j,1}\rangle_n = (1/n)\sum_{i=1}^{n} B_{j,1}(X_i)\sigma(X_i)\varepsilon_i$. Now we define

$$\hat{\varepsilon}_1(x) = \sum_{j=0}^{N} \varepsilon_j^* B_{j,1}(x), \quad x \in [a,b]. \tag{A5}$$

The next lemma gives the pointwise variance of $\hat{\varepsilon}_1(x)$.

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

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

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

Proof Note that $E[B_{j,1}(X_i)B_{j,1}(X_k)\sigma(X_i)\sigma(X_k)\varepsilon_i\varepsilon_k] = 0, \forall i \neq k$ according to the martingale difference property of the ε_i's, the rest of the proof follows from Lemma A1 and the continuity of functions $\sigma(x)$ and $f(x)$. ■

The next lemma implies that the difference between $\tilde{\varepsilon}_1(x)$ and $\hat{\varepsilon}_1(x)$ is negligible uniformly over $x \in [a,b]$.

Lemma A.6 *For any* $x \in [a,b]$, $|\tilde{\varepsilon}_1(x) - \hat{\varepsilon}_1(x)| \leq A_{n,1}\,(1 - A_{n,1})^{-1}|\hat{\varepsilon}_1(x)|$.

Proof Note that

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

for any $x \in [a,b]$. Then Equation (A3) implies that

$$\sup_{0\leq j\leq N} |\|B_{j,1}\|_{2,n}^2 - 1| \leq A_{n,1}, \quad (1 + A_{n,1})^{-1} \leq \sup_{0\leq j\leq N} \|B_{j,1}\|_{2,n}^{-2} \leq (1 - A_{n,1})^{-1},$$

hence the lemma follows. ■

Since the stochastic function $\hat{\varepsilon}_1(x)$ given in Equation (A5) takes constant value on each interval I_j, one only has to bound each of the $N+1$ rescaled noise terms simultaneously by the Berry–Esseen bound for weakly dependent data. First, we verify the conditions in Lemma 3.2 for $\xi_{i,j} \equiv B_{j,1}(X_i)\sigma(X_i)\varepsilon_i$, $1 \leq i \leq n$, $j = 0, \ldots, N$.

Lemma A.7 *There exist constants $c_0(f,\sigma), C_0(f,\sigma) > 0$, such that*

$$\sigma_{n,j}^2 \equiv E\left(\sum_{i=1}^n \xi_{i,j}\right)^2 = nE\{B_{j,1}(X_i)\sigma(X_i)\varepsilon_i\}^2 = nc_{0,j}, \tag{A6}$$

for each $j = 0,\ldots,N$, where $c_{0,j} = \|b_{j,1}\|_2^{-2}\int_{I_j}\sigma^2(u)f(u)du \geq c_0(f,\sigma) > 0$ and

$$d_j \equiv E|\xi_{1,j}|^3 = E\{B_{j,1}^3(X_i)\sigma^3(X_i)|\varepsilon_i|^3\} \leq C_0(f,\sigma)h^{-1/2}. \tag{A7}$$

Proof Using the definition of $\sigma_{n,1}^2(x)$ in Equation (5)

$$\sigma_{n,j}^2 = E\left(\sum_{i=1}^n \xi_{i,j}\right)^2 = n^2E\left\{\frac{1}{n}\sum_{i=1}^n \|b_{j,1}\|_2^{-1}b_{j,1}(X_i)\sigma(X_i)\varepsilon_i\right\}^2$$
$$= n\|b_{j,1}\|_2^{-2}\int_{I_j}\sigma^2(u)f(u)\,\mathrm{d}u = nc_{0,j} \geq nc_0(f,\sigma) > 0.$$

Next, by Lemma A1 and the continuity of functions $\sigma^2(x)$ and $f(x)$, one has

$$d_j = E|\xi_{1,j}|^3 \leq \|b_{j,1}\|_2^{-3/2}\int_{I_j}\sigma^3(u)f(u)\,\mathrm{d}u \leq C_0(f,\sigma)h^{-1/2}.$$ ■

Proof of Proposition 3.1 $(p = 1)$ Note that for any $j = 0,\ldots,N$ and $x \in I_j$,

$$\sigma_{n,1}^{-1}(x)\frac{1}{n}\sum_{i=1}^n B_{j,1}(x)B_{j,1}(X_i)\sigma(X_i)\varepsilon_i = \sigma_{n,j}^{-1}\sum_{i=1}^n \xi_{i,j}, \tag{A8}$$

where $\sigma_{n,j}^2 = nc_{0,j} \geq c_0(f,\sigma) > 0$ as in Equation (A6). Define

$$\Delta_n \equiv \max_{0\leq j\leq N}\sup_{z\in R}|P\{\sigma_{n,1}^{-1}(x)\tilde{\varepsilon}_1(x) \leq z, x\in I_j\} - \Phi(z)|. \tag{A9}$$

Observing that $\{\xi_{i,j}\}_{i=1}^n$ forms a stationary α-mixing sequence, with $E\xi_{i,j} = 0$, then Equations (A7)–(A9) and Lemmas 3.2 and A7 imply

$$\Delta_n = \max_{0\leq j\leq N}\sup_z\left|P\left\{\frac{\sum_{i=1}^n \xi_{i,j}}{\sigma_{n,j}} \leq z, x\in I_j\right\} - \Phi(z)\right| \leq \frac{c_1C_0(f,\sigma)}{h^{1/2}c_0(f,\sigma)\sigma_{n,j}} = \mathrm{o}(N^{-1}),$$

where the last step follows from Assumption (A3). Using the above approximation, for a_N and b_N given in Lemma 3.3 and each $j = 1,\ldots,N$, one can show that

$$P\left\{\left|\sigma_{n,j}^{-1}\sum_{i=1}^n \xi_{i,j}\right| \leq -\frac{\log(\alpha/2)}{a_N} + b_N, x\in I_j\right\}$$
$$= P\left\{\frac{\sum_{i=1}^n \xi_{i,j}}{\sigma_{n,j}} \leq -\frac{\log(\alpha)/2}{a_N} + b_N, x\in I_j\right\} - P\left\{\frac{\sum_{i=1}^n \xi_{i,j}}{\sigma_{n,j}} \leq \frac{\log\,\alpha/2}{a_N} - b_N, x\in I_j\right\}$$
$$= \Phi\left(-\frac{\log(\alpha/2)}{a_N} + b_N\right) - \Phi\left(\frac{\log(\alpha/2)}{a_N} - b_N\right) + \mathrm{o}(N^{-1}).$$

Applying Lemma 3.3 with $2\mathrm{e}^{-\tau} = \alpha$ or $\tau = -\log(\alpha/2)$, one has uniformly in j,

$$P\left\{\left|\sigma_{n,j}^{-1}\sum_{i=1}^n \xi_{i,j}\right| \leq -\frac{\log(\alpha/2)}{a_{N+1}} + b_{N+1}, x\in I_j\right\} = 1 - \frac{\alpha}{1+N} + \mathrm{o}(N^{-1}).$$

Thus,

$$\liminf_{n\to\infty} P\left[\sup_{x\in[a,b]}|\sigma_{n,1}^{-1}(x)\hat{\varepsilon}_1(x)| \leq \{2\log(N+1)\}^{1/2}d_n(\alpha)\right]$$
$$= \liminf_{n\to\infty} P\left[\left|\sigma_{n,j}^{-1}\sum_{i=1}^n \xi_{i,j}\right| \leq -\frac{\log(\alpha/2)}{a_{N+1}} + b_{N+1}, x\in I_j, 0\leq j\leq N\right] \geq 1-\alpha.$$

Therefore, using Lemma A6, one has proved Equation (12) for $p = 1$. ■

Proof of Theorem 2.1 ($p = 1$) By Huang (2003, Theorem 5.1) and Assumption (A3), the uniform bias is $\|\tilde{m}_1(x) - m(x)\|_\infty = \mathrm{o}_p\{n^{-1/2}h^{-1/2}(\log(N+1))^{1/2}\}$, where $\tilde{m}_1(x)$ is the smoothed signal given in Equation (10). Thus the bias order is negligible compared with $(nh)^{-1/2}\{\log(N+1)\}^{1/2}$, which is the uniform noise order of

$$\sigma_{n,1}(x)\left\{-\frac{\log(\alpha/2)}{a_{N+1}} + b_{N+1}\right\} = \sigma_{n,1}(x)\{2\log(N+1)\}^{1/2} d_n(\alpha).$$

Now Equation (11) and Proposition 3.1 yield the conservativity of the band in Equation (7) for $p = 1$ as

$$\begin{aligned}
&\liminf_{n\to\infty} P[m(x) \in \hat{m}_1(x) \pm \sigma_{n,1}(x)\{2\log(N+1)\}^{1/2} d_n(\alpha), \forall x \in [a,b]] \\
&= \liminf_{n\to\infty} P\left[\sup_{x\in[a,b]} \sigma_{n,1}^{-1}(x)|\tilde{\varepsilon}_1(x) + \tilde{m}_1(x) - m(x)| \le \{2\log(N+1)\}^{1/2} d_n(\alpha)\right] \\
&= \liminf_{n\to\infty} P\left[\sup_{x\in[a,b]} |\sigma_{n,1}^{-1}(x)\tilde{\varepsilon}_1(x)| \le \{2\log(N+1)\}^{1/2} d_n(\alpha)\right] \ge 1 - \alpha.
\end{aligned}$$

Therefore, Theorem 2.1 has been proved for the case of $p = 1$. ■

A.3. *Noise term of the linear spline estimator*

For the noise term $\tilde{\varepsilon}_2(x)$ in Equation (10), one can rewrite it as $\tilde{\varepsilon}_2(x) = \sum_{j=-1}^{N} \tilde{a}_j B_{j,2}(x)$, for any $x \in [a,b]$, where according to Equation (10),

$$\tilde{\mathbf{a}} = (\tilde{a}_{-1}, \ldots, \tilde{a}_N)^{\mathrm{T}} = \mathbf{V}_{n,2}^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n} B_{j,2}(X_i)\sigma(X_i)\varepsilon_i\right\}_{j=-1}^{N}.$$

Define $\hat{\mathbf{a}} = (\hat{a}_{-1}, \ldots, \hat{a}_N)^{\mathrm{T}}$ by replacing $\mathbf{V}_{n,2}^{-1}$ with $\mathbf{V}_2^{-1} = \mathbf{S}$ in the above, i.e.

$$\hat{\mathbf{a}} = \mathbf{S}\left\{\frac{1}{n}\sum_{i=1}^{n} B_{j,2}(X_i)\sigma(X_i)\varepsilon_i\right\}_{j=-1}^{N} = \left\{\sum_{j=-1}^{N} s_{j'j}\frac{1}{n}\sum_{i=1}^{n} B_{j,2}(X_i)\sigma(X_i)\varepsilon_i\right\}_{j'=-1}^{N},$$

and define for any $x \in [a,b]$,

$$\begin{aligned}
\hat{\varepsilon}_2(x) &= \sum_{j=-1}^{N} \hat{a}_j B_{j,2}(x) = \sum_{j,j'=-1}^{N} s_{j'j}\frac{1}{n}\sum_{i=1}^{n} B_{j,2}(X_i)\sigma(X_i)\varepsilon_i B_{j',2}(x) \\
&= \sum_{j'=j(x)-1,j(x)} B_{j',2}(x)\sum_{j=-1}^{N} s_{j',j}\frac{1}{n}\sum_{i=1}^{n} B_{j,2}(X_i)\sigma(X_i)\varepsilon_i. \qquad \text{(A10)}
\end{aligned}$$

Next we calculate the variance function of $\hat{\varepsilon}_2(x)$. We first introduce two vectors $\mathbf{U} = \{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} B_{j,2}(X_i)\sigma(X_i)\varepsilon_i\}_{j=-1}^{N}$ and

$$\mathbf{\Lambda}_j = \begin{pmatrix}\Lambda_{j1}\\ \Lambda_{j2}\end{pmatrix} \equiv \tilde{\mathbf{S}}_j\mathbf{U} = \frac{1}{\sqrt{n}}\begin{pmatrix}\sum_{i=1}^{n}\sum_{j'=-1}^{N} s_{j-1,j'}B_{j',2}(X_i)\sigma(X_i)\varepsilon_i\\ \sum_{i=1}^{n}\sum_{j'=-1}^{N} s_{j,j'}B_{j',2}(X_i)\sigma(X_i)\varepsilon_i\end{pmatrix}, \qquad \text{(A11)}$$

in which the $(j-1)$th and jth rows of the matrix $\mathbf{S}$ is denoted as a $2 \times (N+2)$ matrix

$$\tilde{\mathbf{S}}_j = \begin{pmatrix} s_{j-1,-1} & s_{j-1,0} & \cdots & s_{j-1,N}\\ s_{j,-1} & s_{j,0} & \cdots & s_{j,N}\end{pmatrix}, \quad 0 \le j \le N. \qquad \text{(A12)}$$

Then, one can write $\hat{\varepsilon}_2(x)$ in the following matrix form

$$\hat{\varepsilon}_2(x) = \mathbf{D}^{\mathrm{T}}(x)\mathbf{\Lambda}_{j(x)}, \quad x \in [a,b], \qquad \text{(A13)}$$

in which the function $\mathbf{D}(x)$ is a 2-vector matrix such that

$$\mathbf{D}(x) \equiv \{D_{j(x)-1}(x), D_{j(x)}(x)\}^{\mathrm{T}}, \quad D_j(x) \equiv n^{-1/2}B_{j,2}(x), \quad -1 \le j \le N. \qquad \text{(A14)}$$

The next lemma provides the pointwise variance of $\hat{\varepsilon}_2(x)$. Let

$$\mathbf{\Sigma} = (\sigma_{jl})_{j,l=-1}^{N} = \left\{\int \sigma^2(v)B_{j,2}(v)B_{l,2}(v)f(v)\,\mathrm{d}v\right\}_{j,l=-1}^{N}. \qquad \text{(A15)}$$

Lemma A.8 *The pointwise variance of* $\hat{\varepsilon}_2(x)$ *is* $\sigma^2_{n,2}(x)$ *in Equation* (6), *which satisfies*

$$\sigma^2_{n,2}(x) = \frac{3\sigma^2(x)}{2f(x)nh}\mathbf{\Delta}^{\mathrm{T}}(x)\mathbf{S}_{j(x)}\mathbf{\Delta}(x)\{1 + r_{n,2}(x)\},$$

with $\sup_{x\in[a,b]}|r_{n,2}(x)| \to 0$, $j(x)$ *is as defined in Equation* (2), $\mathbf{\Delta}(x)$ *as defined in Equation* (15) *and matrix* $\mathbf{S}_j$ *in Equation* (A1). *Consequently, there exist positive constants* c_σ, C_σ *such that for large enough* n, $c_\sigma(nh)^{-1/2} \le \sigma_{n,2}(x) \le C_\sigma(nh)^{-1/2}$, $\forall x \in [a,b]$.

Proof From Equations (A11) and (A13), $E\{\hat{\varepsilon}^2_2(x)\}\mathbf{D}^{\mathrm{T}}(x)\tilde{\mathbf{S}}_{j(x)}\mathrm{Cov}(\mathbf{U})\tilde{\mathbf{S}}^{\mathrm{T}}_{j(x)}\mathbf{D}(x)$. Note that $E[B_{j,2}(X_i)B_{l,2}(X_k)\sigma(X_i)\sigma(X_k)\varepsilon_i\varepsilon_k] = 0$, $\forall i \ne k$ according to the martingale difference property of the ε_i's, the jlth entry of the covariance matrix of $\mathbf{U}$ is

$$\frac{1}{n}\sum_{i=1}^{n} E\{B_{j,2}(X_i)B_{l,2}(X_i)\sigma^2(X_i)\} = \int \sigma^2(v)B_{j,2}(v)B_{l,2}(v)f(v)\,\mathrm{d}v = \sigma_{jl},$$

which is the jlth entry of the matrix $\mathbf{\Sigma}$ defined in (A15), i.e. $\mathrm{Cov}(\mathbf{U}) = \mathbf{\Sigma}$. The rest of the proof is simple algebra. ∎

Lemma A.9 *Under Assumptions* (A3) *and* (A5),

$$\left|\sup_{x\in[a,b]}|\sigma^{-1}_{n,2}(x)\hat{\varepsilon}_2(x)| - \sup_{x\in[a,b]}|\sigma^{-1}_{n,2}(x)\tilde{\varepsilon}_2(x)|\right| = \mathrm{O}_p\{(nh)^{-1/2}\log n\} = \mathrm{o}_p(1).$$

Proof Note that $\mathbf{V}_{n,2}\tilde{\mathbf{a}} = \mathbf{V}_2\hat{\mathbf{a}}$. Based on Lemma A4 and Equation (A4), there exists a constant c such that $|\hat{\mathbf{a}} - \tilde{\mathbf{a}}| \le A_{n,2}|\hat{\mathbf{a}}|/(c - A_{n,2})$. From the definitions of $\tilde{\varepsilon}_2(x)$ in Equation (10) and $\hat{\varepsilon}_2(x)$ in Equation (A10), plus Lemmas A1 and A8, as $n \to \infty$

$$\sup_{x\in[a,b]}\left|\frac{\hat{\varepsilon}_2(x)}{\sigma_{n,2}(x)} - \frac{\tilde{\varepsilon}_2(x)}{\sigma_{n,2}(x)}\right| \le \sup_{x\in[a,b]}\frac{1}{\sigma_{n,2}(x)}\left|\sum_{j=-1}^{N}|\hat{\mathbf{a}} - \tilde{\mathbf{a}}|B_{j,2}(x)\right| \le \frac{Cn^{1/2}A_{n,2}}{c - A_{n,2}}|\hat{\mathbf{a}}|.$$

Using Lemma A1 again, we have as $n \to \infty$,

$$\sup_{x\in[a,b]}\left|\frac{\hat{\varepsilon}_2(x)}{\sigma_{n,2}(x)}\right| \ge C_\sigma^{-1}\sqrt{nh}C^{-1/2}h^{-1/2}\sup_{x\in[a,b]}\left|\sum_{j=-1}^{N}\hat{a}_j b_{j,2}(x)\right| \ge C\sqrt{n}|\hat{\mathbf{a}}|.$$

Then the desired result holds. ∎

A.4. *Proofs of Proposition* 3.1 *and Theorem* 2.1 *for linear spline*

Prior to the proof, we define

$$\mathbf{Z}_j \equiv (Z_{j1}, Z_{j2})^{\mathrm{T}} = \mathbf{\Lambda}_j^{\mathrm{T}}\{\mathrm{Cov}(\mathbf{\Lambda}_j)\}^{-1/2} = \begin{pmatrix}\beta^{(j)}_{11}\Lambda_{j1} + \beta^{(j)}_{12}\Lambda_{j2}\\ \beta^{(j)}_{12}\Lambda_{j1} + \beta^{(j)}_{22}\Lambda_{j2}\end{pmatrix}, \tag{A16}$$

where $\mathbf{\Lambda}_j$ is defined in Equation (A11). It is clear that $\mathrm{Var}(\mathbf{Z}_j) = \mathbf{I}$, $\mathrm{Var}(Z_{j\gamma}) = 1$, $\gamma = 1, 2$, for any $j = 0, \ldots, N$. The next lemma shows that the covariance matrix of $\mathbf{\Lambda}_j$ approximates $\sigma^2(t_{j+1})\mathbf{S}_j$ defined in Equation (A1) uniformly.

Lemma A.10 *For vector* $\mathbf{\Lambda}_j$ *and matrix* $\mathbf{S}_j$ *in Equations* (A11) *and* (A1),

$$\mathrm{Cov}(\mathbf{\Lambda}_j) = \sigma^2(t_{j+1})\mathbf{S}_j + \tilde{\mathbf{R}}_j,$$

for any $0 \le j \le N$ *and* $\lim_{n\to\infty}\max_{0\le j\le N}|\tilde{\mathbf{R}}_j| = 0$.

Proof Since $\mathbf{\Lambda}_j = \tilde{\mathbf{S}}_j\mathbf{U}$ with $\tilde{\mathbf{S}}_j$ defined in Equation (A12) and $\mathrm{Cov}(\mathbf{U}) = \mathbf{\Sigma}$ as in the proof of Lemma A8, the covariance matrix of $\mathbf{\Lambda}_j$ is

$$\mathrm{Cov}(\mathbf{\Lambda}_j) = \tilde{\mathbf{S}}_j\mathbf{\Sigma}\tilde{\mathbf{S}}_j^{\mathrm{T}} = \begin{pmatrix} \sum_{k,l=-1}^{N} s_{j-1,k}s_{j-1,l}\sigma_{kl} & \sum_{k,l=-1}^{N} s_{j,k}s_{j-1,l}\sigma_{kl} \\ \sum_{k,l=-1}^{N} s_{j-1,k}s_{j,l}\sigma_{kl} & \sum_{k,l=-1}^{N} s_{j,k}s_{j,l}\sigma_{kl} \end{pmatrix}.$$

By Assumption (A2) and Equations (A12) and (A15)

$$\sigma_{kl} = \int \sigma^2(v)B_{k,2}(v)B_{l,2}(v)f(v)dv = \sigma^2(t_{k+1})v_{kl} + cw(f\sigma^2, h).$$

Similarly, one also has $\sigma_{kl} = \sigma^2(t_{l+1})v_{kl} + cw(f\sigma^2, h)$. Thus,

$$\mathrm{Cov}(\mathbf{\Lambda}_j) = \sum_{k,l=-1}^{N} \begin{pmatrix} s_{j-1,k}s_{j-1,l}v_{kl}\sigma^2(t_{l+1}) & s_{j,k}s_{j-1,l}v_{kl}\sigma^2(t_{l+1}) \\ s_{j-1,k}s_{j,l}v_{kl}\sigma^2(t_{k+1}) & s_{j,k}s_{j,l}v_{kl}\sigma^2(t_{k+1}) \end{pmatrix} + \tilde{\mathbf{R}}_j^*,$$

where

$$\tilde{\mathbf{R}}_j^* = cw(f\sigma^2, h)\begin{pmatrix} \sum_{k,l=-1}^{N} s_{j-1,k}s_{j-1,l} & \sum_{k,l=-1}^{N} s_{j,k}s_{j-1,l} \\ \sum_{k,l=-1}^{N} s_{j-1,k}s_{j,l} & \sum_{k,l=-1}^{N} s_{j,k}s_{j,l} \end{pmatrix}.$$

Note that $\sum_{k=-1}^{N} s_{j,k}v_{kl} = 0$ if $l \neq j$ and $\sum_{k=-1}^{N} s_{j,k}v_{kl} = 1$ if $l = j$, thus

$$\mathrm{Cov}(\mathbf{\Lambda}_j) = \begin{pmatrix} s_{j-1,j}\sigma^2(t_j) & s_{j-1,j}\sigma^2(t_{j+1}) \\ s_{j-1,j}\sigma^2(t_{j+1}) & s_{j,j}\sigma^2(t_{j+1}) \end{pmatrix} + \tilde{\mathbf{R}}_j^* = \sigma^2(t_{j+1})\mathbf{S}_j + \tilde{\mathbf{R}}_j.$$

Using Equation (A2) and letting $\tilde{\boldsymbol{\xi}}_{j'} = \{\mathrm{sgn}(s_{j'j})\}_{j=-1}^{N}$, one can show that there exists a positive C_s such that

$$\sum_{j=-1}^{N} |s_{j'j}| \leq |\mathbf{S}\tilde{\boldsymbol{\xi}}_{j'}| \leq C_s|\tilde{\boldsymbol{\xi}}_{j'}| = C_s, \quad \forall j' = -1, 0, \ldots, N.$$

Thus $\max_{0\leq j\leq N} |\tilde{\mathbf{R}}_j| \leq Cw(f\sigma^2, h) \to 0$, as $n \to \infty$. ■

Lemma A.11 *For the matrices* $\boldsymbol{L}_j$, $j = 0, \ldots, N$, *defined in Equation* (18),

$$\lim_{n\to\infty} \max_{0\leq j\leq N} |\boldsymbol{L}_j^{-1/2} - \sigma(t_{j+1})\{\mathrm{Cov}(\mathbf{\Lambda}_j)\}^{-1/2}| = 0.$$

Proof Note that $\mathbf{L}_j^{-1/2}$, $\{\mathrm{Cov}(\mathbf{\Lambda}_j)\}^{-1/2}$ are symmetric matrices. Using the following fact for symmetric matrices $\mathbf{A}$ and $\mathbf{B}$: $c|\mathbf{A}^{-1/2} - \mathbf{B}^{-1/2}| \leq |\mathbf{B} - \mathbf{A}|$, together with Lemma A10,

$$\begin{aligned} c|\mathbf{L}_j^{-1/2} - \sigma(t_{j+1})\{\mathrm{Cov}(\mathbf{\Lambda}_j)\}^{-1/2}| &\leq |\sigma^{-2}(t_{j+1})\mathrm{Cov}(\mathbf{\Lambda}_j) - \mathbf{L}_j| \\ &\leq |\mathbf{S}_j - \mathbf{L}_j| + |\sigma^{-2}(t_{j+1})\mathrm{Cov}(\mathbf{\Lambda}_j) - \mathbf{S}_j| = |\mathbf{S}_j - \mathbf{L}_j| + \sigma^{-2}(t_{j+1})\tilde{\mathbf{R}}_j. \end{aligned}$$

The desired result follows from Lemma A3. ■

Lemma A.12 *Under Assumptions* (A1)–(A5), *for the variables* $Z_{j\gamma}$, $\gamma = 1, 2$, $0 \leq j \leq N$, *defined in Equation* (A16), *one has*

$$\max_{\gamma=1,2} \limsup_{n\to\infty} P\left[\max_{0\leq j\leq N} Z_{j\gamma}^2 > 2\{\log(N+1)\}\left\{d_n\left(\frac{\alpha}{2}\right)\right\}^2\right] \leq \frac{\alpha}{2}.$$

Proof Without loss of generality, we prove this lemma only for $\gamma = 1$. By Equation (A16),

$$Z_{j1} = \beta_{11}^{(j)}\Lambda_{j1} + \beta_{12}^{(j)}\Lambda_{j2} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k=-1}^{N}(\beta_{11}^{(j)}s_{j-1,k} + \beta_{12}^{(j)}s_{j,k})B_{k,2}(X_i)\sigma(X_i)\varepsilon_i.$$

Let $\zeta_{i,j} = \sum_{k=-1}^{N}(\beta_{11}^{(j)}s_{j-1,k} + \beta_{12}^{(j)}s_{j,k})B_{k,2}(X_i)\sigma(X_i)\varepsilon_i$, $j = 0, \ldots, N$, $i = 1, \ldots, n$, then $\sqrt{n}Z_{j1} = S_n = \sum_{i=1}^{n}\zeta_{i,j}$, and $E(\sqrt{n}Z_{j1})^2 = nEZ_{j1}^2 = n$. One only needs to find a bound for $E|\zeta_{1,j}|^3$ to apply Lemma 3.2to S_n. By the boundedness of $\max_{0\le j\le N}|\mathbf{L}_j|$, Lemmas A1 and A11,

$$E|\zeta_{1,j}|^3 = E\left[\left|\sum_{k=-1}^{N}(\beta_{11}^{(j)}s_{j-1,k}+\beta_{12}^{(j)}s_{j,k})B_{k,2}(X_1)\right|^3\sigma^3(X_1)|\varepsilon_1^3|\right] \le C(f,\sigma)h^{-1/2}.$$

Lemma 3.2 entails that $\Delta_n = \mathrm{o}(n^{-1/2}h^{-1/2}) = \mathrm{o}(N^{-2})$, in which Δ_n is

$$\max_{0\le j\le N}\sup_z|P\{Z_{j1}\le z\}-\Phi(z)| = \max_{0\le j\le N}\sup_z\left|P\left\{n^{-1/2}\sum_{i=1}^{n}\zeta_{i,j}\le z\right\}-\Phi(z)\right|.$$

By Lemma 3.3, one has uniformly in j,

$$P\left[|Z_{j1}|\le\{2\log(N+1)\}^{1/2}d_n\left(\frac{\alpha}{2}\right)\right] = 1-\frac{\alpha}{2(N+1)}+\mathrm{o}(N^{-1}).$$

Hence $\limsup_{n\to\infty}P[\max_{0\le j\le N}\{Z_{j1}^2\}>2\{\log(N+1)\}\{d_n(\alpha/2)\}^2] = \alpha/2$. ■

Proof of Proposition 3.1 $(p=2)$ According to Lemma A9, we only need to show that

$$\liminf_{n\to\infty}P\left[\sup_{x\in[a,b]}|\sigma_{n,2}^{-1}(x)\hat{\varepsilon}_2(x)|\le 2\{\log(N+1)\}^{1/2}d_n\left(\frac{\alpha}{2}\right)\right]\ge 1-\alpha. \tag{A17}$$

Note that $\hat{\varepsilon}_2(x) = \mathbf{D}^{\mathrm{T}}(x)\mathbf{\Lambda}_{j(x)}$, where $\mathbf{D}(x)$ and $\mathbf{\Lambda}_{j(x)}$ are defined in Equations (A14) and (A11). Thus, standardisation leads to

$$\{\sigma_{n,2}^{-1}(x)\hat{\varepsilon}_2(x)\}^2 = \frac{\{\sigma_{n,2}^{-1}(x)\mathbf{D}(x)\}^{\mathrm{T}}\mathbf{\Lambda}_{j(x)}\mathbf{\Lambda}_{j(x)}^{\mathrm{T}}\{\sigma_{n,2}^{-1}(x)\mathbf{D}(x)\}}{\{\sigma_{n,2}^{-1}(x)\mathbf{D}(x)\}^{\mathrm{T}}\mathrm{cov}(\mathbf{\Lambda}_{j(x)})\{\sigma_{n,2}^{-1}(x)\mathbf{D}(x)\}}. \tag{A18}$$

Define for any $j=0,\ldots,N$, $\mathbf{Q}_j = \mathbf{\Lambda}_j^{\mathrm{T}}\{\mathrm{Cov}(\mathbf{\Lambda}_j)\}^{-1}\mathbf{\Lambda}_j = \mathbf{Z}_j\mathbf{Z}_j^{\mathrm{T}} = \sum_{\gamma=1,2}Z_{j\gamma}^2$. The maximisation lemma of Johnson and Wichern (1992, p. 166) ensures that for any $x\in[a,b]$

$$\frac{\{\mathbf{D}(x)/\sigma_{n,2}(x)\}^{\mathrm{T}}\mathbf{\Lambda}_{j(x)}\mathbf{\Lambda}_{j(x)}^{\mathrm{T}}\{\mathbf{D}(x)/\sigma_{n,2}(x)\}}{\{\mathbf{D}(x)/\sigma_{n,2}(x)\}^{\mathrm{T}}\mathrm{Cov}(\mathbf{\Lambda}_{j(x)})\{\mathbf{D}(x)/\sigma_{n,2}(x)\}}\le\mathbf{\Lambda}_{j(x)}^{\mathrm{T}}\{\mathrm{Cov}(\mathbf{\Lambda}_{j(x)})\}^{-1}\mathbf{\Lambda}_{j(x)} = \mathbf{Q}_{j(x)},$$

which together with Equation (A18) entails that $\sup_{x\in[a,b]}|\sigma_{n,2}^{-1}(x)\hat{\varepsilon}_2(x)|\le\max_{0\le j\le N}\mathbf{Q}_j$.

Thus, Lemma A12 implies

$$\begin{aligned}
&\liminf_{n\to\infty}P\left[\sup_{x\in[a,b]}|\sigma_{n,2}^{-1}(x)\hat{\varepsilon}_2(x)|\le 2\{\log(N+1)\}^{1/2}d_n\left(\frac{\alpha}{2}\right)\right]\\
&\ge\liminf_{n\to\infty}P\left[\max_{0\le j\le N}\mathbf{Q}_j\le 4\{\log(N+1)\}\left\{d_n\left(\frac{\alpha}{2}\right)\right\}^2\right]\\
&\ge 1-\sum_{\gamma=1,2}\limsup_{n\to\infty}P\left[\max_{0\le j\le N}\{Z_{j\gamma}^2\}>2\{\log(N+1)\}\left\{d_n\left(\frac{\alpha}{2}\right)\right\}^2\right]\ge 1-\alpha.
\end{aligned}$$

The desired result follows from Lemma A9 automatically. ■

Proof of Theorem 2.1 $(p=2)$ Huang (2003, Theorem 5.1) implies that

$$\left\{\frac{nh}{\log(N+1)}\right\}^{1/2}\|\tilde{m}_2(x)-m(x)\|_\infty = \mathrm{O}_p\left[\left\{\frac{nh^5}{\log(N+1)}\right\}^{-1/2}\right] = \mathrm{o}_p(1),$$

which implies that the bias order is negligible compared with the noise order. Applying Equation (12) with $p=2$ in Proposition 3.1

$$\begin{aligned}
&\liminf_{n\to\infty}P[m(x)\in\hat{m}_2(x)\pm 2\sigma_{n,2}(x)\{\log(N+1)\}^{1/2}d_n(\alpha/2),\forall x\in[a,b]]\\
&=\liminf_{n\to\infty}P\left[\sup_{x\in[a,b]}\sigma_{n,2}^{-1}(x)|\tilde{\varepsilon}_2(x)+\tilde{m}_2(x)-m(x)|\le 2\{\log(N+1)\}^{1/2}d_n(\alpha/2)\right]\\
&=\liminf_{n\to\infty}P\left[\sup_{x\in[a,b]}|\sigma_{n,2}^{-1}(x)\tilde{\varepsilon}_2(x)|\le 2\{\log(N+1)\}^{1/2}d_n(\alpha/2)\right]\ge 1-\alpha.
\end{aligned}$$

■