Journal of Nonparametric Statistics
Vol. 23, No. 1, March 2011, 67–81

A jump-detecting procedure based on spline estimation

Shujie Ma and Lijian Yang*

Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, USA

(*Received 13 April 2009; final version received 16 December 2009*)

In a random-design nonparametric regression model, procedures for detecting jumps in the regression function via constant and linear spline estimation method are proposed based on the maximal differences of the spline estimators among neighbouring knots, the limiting distributions of which are obtained when the regression function is smooth. Simulation experiments provide strong evidence that corroborates with the asymptotic theory, while the computing is extremely fast. The detecting procedure is illustrated by analysing the thickness of pennies data set.

Keywords: B spline; knots; jump points; nonparametric regression; asymptotic Gaussian process; asymptotic p-value, BIC; upcrossing probability

AMS Subject Classification: 62G08; 62G10; 62G20

1. Introduction

In application of regression methods, ignoring possible jump points may result in a serious error in drawing inference about the process under study. Whenever there is no appropriate parametric method available, one may start from nonparametric regression. Two popular nonparametric techniques are kernel and spline smoothing. For properties of kernel estimators in the absence of jump points, see Mack and Silverman (1982), Fan and Gijbels (1996), Xia (1998) and Claeskens and Van Keilegom (2003), and for spline estimators, see Zhou, Shen, and Wolfe (1998), Huang (2003) and Wang and Yang (2009).

One is often interested in detecting jump points and estimating regression function with jumps. 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

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

where the joint distribution of (X, ε) satisfies Assumptions (A3) and (A4) in Section 2. The unknown mean function $m(x)$, defined on interval $[a, b]$, may have a finite number of jump points.

*Corresponding author. Emails: yang@stt.msu.edu, yangli@stt.msu.edu

ISSN 1048-5252 print/ISSN 1029-0311 online

DOI: 10.1080/10485250903571978
http://www.informaworld.com

Jump regression analysis started in the early 1990s and has become an important research topic in statistics. See, for instance, Qiu, Asano, and Li (1991), Müller (1992), Wu and Chu (1993), Qiu (1994) and Qiu and Yandell (1998) for procedures that detect the jumps explicitly before estimating the regression curve, Kang, Koo, and Park (2000) for comparing two estimators of the regression curve after the jump points are detected, Qiu (2003) and Gijbels, Lambert, and Qiu (2007) for jump-preserving curve estimators and Joo and Qiu (2009) for jump detection in not only the regression curve but also its derivatives. For a comprehensive view on jump regression, see Qiu (2005).

Jump detection has been tackled with many techniques, including local polynomial smoothing (Qiu 2003; Gijbels et al. 2007), smoothing spline (Shiau 1987) and wavelet methods (Hall and Patil 1995; Wang 1995; Park and Kim 2004, 2006). For two-dimensional cases, see Qiu (2007). We propose a spline smoothing method to detect jumps by solving one optimisation problem over the range of x instead of each point, which is computationally more expedient than the kernel-type method in Müller (1992). The spline method was also discussed in Koo (1997), which proposed estimating discontinuous regression function without providing theoretical justifications. In contrast, asymptotic distributions in Theorem 2.1 are established by making use of the strong approximation results in Wang and Yang (2009), normal comparison lemma in Leadbetter, Lindgren, and Rootzén (1983) and a convenient formula from Kılıç (2008) for inverting a tridiagonal matrix. The automatic procedures proposed for detecting jumps are based on implementing the asymptotics of Theorem 2.1.

The paper is organised as follows. Section 2 states main theoretical results based on (piecewise) constant and linear splines. Section 3 provides steps to implement the procedure based on the asymptotic result. Section 4 reports findings in both simulation and real data studies. All technical proofs are contained in the Appendices.

2. Main results

We denote the space of the pth order smooth functions as $C^{(p)}[a,b] = \{\varphi | \varphi^{(p)} \in C[a,b]\}$, for $p = 1, 2$. Without loss of generality, we take the range of X, $[a,b]$ to be $[0,1]$. To introduce the spline functions, divide the finite interval $[0,1]$ into $(N+1)$ equal subintervals $J_j = [t_j, t_{j+1})$, $j = 0, \dots, N-1$, $J_N = [t_N, 1]$. A sequence of equally spaced points $\{t_j\}_{j=1}^N$, called interior knots, are given as

$$t_0 = 0 < t_1 < \cdots < t_N < 1 = t_{N+1}, \quad t_j = jh, \quad j = 0, \dots, N+1,$$

in which $h = 1/(N+1)$ is the distance between neighbouring knots. We denote by $G^{(p-2)} = G^{(p-2)}[0,1]$ the space of functions that are polynomials of degree $p-1$ on each J_j and have a continuous $(p-2)$th derivative. For example, $G^{(-1)}$ denotes the space of functions that are constant on each J_j, and $G^{(0)}$ denotes the space of functions that are linear on each J_j and continuous on $[0,1]$. Define the spline estimator based on data $\{(X_i, Y_i)\}_{i=1}^n$ drawn from model (1) as

$$\hat{m}_p(x) = \operatorname{argmin}_{g \in G^{(p-2)}[0,1]} \sum_{i=1}^n \{Y_i - g(X_i)\}^2, \quad p = 1, 2. \tag{2}$$

The unknown function $m(x)$ in Equation (1) may be smooth or have jump points $\{\tau_i\}_{i=1}^k$, for $0 = \tau_0 < \tau_1 < \cdots < \tau_k < \tau_{k+1} = 1$. Technical assumptions are listed as follows.

(A1) There exist a function $m_0(x) \in C^{(p)}[0,1]$ and a vector $\mathbf{c} = (c_1, \dots, c_k)$ of jump magnitudes such that the regression function $m(x) = c_l + m_0(x), x \in [\tau_l, \tau_{l+1})$, for $l = 1, \dots, k-1, m(x) = m_0(x), x \in [\tau_0, \tau_1), m(x) = c_k + m_0(x), x \in [\tau_k, \tau_{k+1}]$.

(A2) The number of interior knots $N = o(n^{1/(2p+1)+\vartheta})$ for any $\vartheta > 0$ while $N^{-1} = o(n^{-1/(2p+1)}/\log n)$.
(A3) X is uniformly distributed on interval [0, 1], i.e. the density function of X is $f(x) = I(0 \le x \le 1)$.
(A4) The joint distribution $F(x, \varepsilon)$ 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 exist a positive value $\eta > 1$ and a finite positive M_η such that $E|\varepsilon|^{2+\eta} < M_\eta$ and $\sup_{x\in[0,1]} E(|\varepsilon|^{2+\eta}|X = x) < M_\eta$.

Assumption (A1) is similar to that in Müller and Song (1997). Assumption (A2) is similar to the undersmoothing condition in Claeskens and Van Keilegom (2003); thus, the subinterval length $h = o(n^{-1/(2p+1)}/\log n)$ while $h^{-1} = o(n^{1/(2p+1)+\vartheta})$ for any $\vartheta > 0$. A uniform distribution of X in Assumption (A3) is for the simplicity of proofs, which can be relaxed to any distribution with continuous and positive density function on [0, 1]. Assumption (A4) is identical with Mack and Silverman (1982, (C2)(a)). All are typical assumptions for a nonparametric regression, with Assumption (A4) weaker than the corresponding assumption in Härdle (1989).

Denote by $\|\phi\|_2$ the theoretical L^2 norm of a function ϕ on $[0, 1]$, $\|\phi\|_2^2 = E\{\phi^2(X)\} = \int_0^1 \phi^2(x) f(x)\,\mathrm{d}x = \int_0^1 \phi^2(x)\,\mathrm{d}x$, 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 $\langle\phi, \varphi\rangle = \int_0^1 \phi(x)\varphi(x) f(x)\,\mathrm{d}x = \int_0^1 \phi(x)\varphi(x)\,\mathrm{d}x = E\{\phi(X)\varphi(X)\}$, $\langle\phi, \varphi\rangle_n = n^{-1}\sum_{i=1}^n \phi(X_i)\varphi(X_i)$, for any L^2-integrable functions ϕ, φ on [0, 1]. Clearly, $E\langle\phi, \varphi\rangle_n = \langle\phi, \varphi\rangle$. We now introduce the B-spline basis for theoretical analysis. The B-spline basis of $G^{(-1)}$, the space of piecewise constant splines, are indicator functions of intervals J_j, $b_{j,1}(x) = I_j(x) = I_{J_j}(x)$, $0 \le j \le N$. The B-spline basis of $G^{(0)}$, the space of piecewise linear splines, are $\{b_{j,2}(x)\}_{j=-1}^N$

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

Next define their theoretical norms

$$c_{j,n} = \|b_{j,1}\|_2^2 = \int_0^1 I_j^2(x)\mathrm{d}x = \int_0^1 I_j(x)\mathrm{d}x = h, \quad 0 \le j \le N,$$

$$d_{j,n} = \|b_{j,2}\|_2^2 = \int_0^1 K^2\left(\frac{x - t_{j+1}}{h}\right)\mathrm{d}x = \begin{cases} 2h/3, & 0 \le j \le N - 1, \\ h/3, & j = -1, N, \end{cases}$$

$$\langle b_{j,2}, b_{j',2}\rangle = \int_0^1 K\left(\frac{x - t_{j+1}}{h}\right) K\left(\frac{x - t_{j'+1}}{h}\right)\mathrm{d}x = \begin{cases} h/6, & |j - j'| = 1, \\ 0, & |j - j'| > 1. \end{cases} \quad (3)$$

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

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

$$B_{j,2}(x) \equiv b_{j,2}(x)\{d_{j,n}\}^{-1/2}, \quad j = -1, \ldots, N. \quad (4)$$

The inner product matrix V of the B-spline basis $\left\{B_{j,2}(x)\right\}_{j=-1}^{N}$ is denoted as

$$\mathbf{V} = \left(v_{j'j}\right)_{j,j'=-1}^{N} = \left(\left\langle B_{j',2}, B_{j,2}\right\rangle\right)_{j,j'=-1}^{N}$$
$$= \begin{pmatrix} 1 & \sqrt{2}/4 & & & & 0 \\ \sqrt{2}/4 & 1 & 1/4 & & & \\ & 1/4 & 1 & \ddots & & \\ & & \ddots & \ddots & 1/4 & \\ & & & 1/4 & 1 & \sqrt{2}/4 \\ 0 & & & & \sqrt{2}/4 & 1 \end{pmatrix}_{(N+2)\times(N+2)} = (l_{ik})^{-1}_{(N+2)\times(N+2)}, \quad (5)$$

which computed via Equation (3). Denote the inverse of $\mathbf{V}$ by $\mathbf{S}$ and 3×3 diagonal submatrices of $\mathbf{S}$ are expressed as

$$\mathbf{S} = (s_{j'j})_{j,j'=-1}^{N} = V^{-1}, \quad S_j = \begin{pmatrix} s_{(j-2),(j-2)} & s_{(j-2),(j-1)} & s_{(j-2),j} \\ s_{(j-1),(j-2)} & s_{(j-1),(j-1)} & s_{(j-1),j} \\ s_{j,(j-2)} & s_{j,(j-1)} & s_{jj} \end{pmatrix}, \quad j = 1, \ldots, N. \quad (6)$$

To detect jumps in m, one tests the hypothesis $\mathcal{H}_0$: $m \in C^{(p)}[0, 1]$ *vs.* $\mathcal{H}_1$:$m \notin C[0, 1]$. Denote by $\|\mathbf{c}\|_2 = (c_1^2 + \cdots + c_k^2)^{1/2}$, the Euclidean norm of the vector $\mathbf{c}$ of all the k jump magnitudes, then under Assumption (A1), one can write alternatively $\mathcal{H}_0$: $\|\mathbf{c}\|_2 = 0$ *vs.* $\mathcal{H}_1$: $\|\mathbf{c}\|_2 > 0$. For $\hat{m}_p(x)$ given in Equation (2), $p = 1, 2$, define the test statistics

$$T_{1n} = \max_{0\le j\le N-1} \hat{\delta}_{1j}, \hat{\delta}_{1j} = \frac{|\hat{m}_1(t_{j+1}) - \hat{m}_1(t_j)|}{\sigma_{n,1}},$$
$$T_{2n} = \max_{1\le j\le N} \hat{\delta}_{2j}, \hat{\delta}_{2j} = \frac{|\{\hat{m}_2(t_{j+1}) + \hat{m}_2(t_{j-1})\}/2 - \hat{m}_2(t_j)|}{\sigma_{n,2,j}}, \quad (7)$$

where

$$\sigma_{n,1}^2 = \frac{2\sigma^2}{nh}, \quad \sigma_{n,2,j}^2 = \sigma^2 \left(\frac{8nh}{3}\right)^{-1} \zeta^{\mathrm{T}} S_j \zeta, \quad \zeta = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \quad (8)$$

with S_j defined in Equation (6). To state our main results, denote

$$d_N(\alpha) = 1 - \{2\log N\}^{-1}\left[\log\left\{-\frac{1}{2}\log(1-\alpha)\right\} + \frac{1}{2}\{\log\log(N) + \log 4\pi\}\right]. \quad (9)$$

THEOREM 2.1 *Under Assumptions* (A1)–(A4) *and* $\mathcal{H}_0$,

$$\lim_{n\to\infty} P[T_{pn} > \{2\log(N-2p+2)\}^{1/2} d_{N-2p+2}(\alpha)] = \alpha, \quad p = 1, 2.$$

A similar result by kernel smoothing with a fixed-design regression model exists in Theorem 3 of Wu and Chu (1993). The proof of that result, however, does not contain sufficient details for us to further comment. It is feasible to derive a similar asymptotic result for T_{pn} under $\mathcal{H}_1$ but that is beyond the scope of this paper and so we leave it to a future work.

3. Implementation

In this section, we describe how to implement in XploRe (Härdle, Hlávka, and Klinke 2000) the jump points detection procedure by using the results in Theorem 2.1.

Given any sample $\{(X_i, Y_i)\}_{i=1}^n$ from model (1), denote $X_{\min} = \min(X_1, \ldots, X_n)$ and $X_{\max} = \max(X_1, \ldots, X_n)$. Then we transform $\{X_i\}_{i=1}^n$ onto interval [0, 1] by subtracting each X_i from $X_{\min}$ and then dividing by $X_{\max} - X_{\min}$. The definition of $\hat{m}_p(x)$ in Equation (2) entails

$$\hat{m}_p(x) \equiv \sum_{j=1-p}^{N} \hat{\lambda}'_{j,p} b_{j,p}(x), \quad p = 1, 2, \tag{10}$$

where coefficients $\{\hat{\lambda}'_{1-p,p}, \ldots, \hat{\lambda}'_{N,p}\}^{\mathrm{T}}$ are solutions of the following least squares problem

$$\{\hat{\lambda}'_{1-p,p}, \ldots, \hat{\lambda}'_{N,p}\}^{\mathrm{T}} = \underset{\{\lambda_{1-p,p},\ldots,\lambda_{N,p}\}\in R^{N+p}}{\operatorname{argmin}} \sum_{i=1}^{n} \left\{ Y_i - \sum_{j=1-p}^{N} \lambda_{j,p} b_{j,p}(X_i) \right\}^2.$$

By Assumption (A2), the number of interior knots N is taken to be $N = [n^{1/3}(\log n)^2/5]$ for a constant spline ($p = 1$), and $N = [n^{1/5}(\log n)^2/5]$ for a linear spline ($p = 2$), in which $[a]$ denotes the integer part of a. Denote the estimator for Y_i by $\hat{Y}_{i,p} = \hat{m}_p(X_i)$, for $i = 1, \ldots, n$, with $\hat{m}_p$ given in Equation (10), and define the estimator of σ as $\hat{\sigma}_p = \{\sum_{i=1}^n (Y_i - \hat{Y}_{i,p})^2/(n - N - p)\}^{1/2}$. Basic spline smoothing theory as in Wang and Yang (2009) ensures that $\hat{\sigma}_p^2 \to_{\mathrm{p}} \sigma^2$, as $n \to \infty$; hence, Theorem 2.1 holds if one replaces σ by $\hat{\sigma}_p$. The asymptotic p-value $p_{\mathrm{value},p}$ is obtained by solving the equation $T_{pn} = \{2\log(N - 2p + 2)\}^{1/2} d_{N-2p+2}(p_{\mathrm{value},p})$, $p = 1, 2$ with T_{pn} defined in Equation (7) and estimated by replacing σ^2 with $\hat{\sigma}_p^2$, then

$$\begin{aligned} p_{\mathrm{value},p} = 1 - \exp[-2\exp[2\log(N - 2p + 2)\{1 - \{2\log(N - 2p + 2)\}^{-1/2} T_{pn}\} \\ - 2^{-1}\{\log\log(N - 2p + 2) + \log 4\pi\}]]. \end{aligned} \tag{11}$$

When the p-value is below a pre-determined α, one concludes that there exist jump points in m. The jump locations and magnitudes are estimated as follows. We use the BIC criteria proposed in Xue and Yang (2006) to select the 'optimal' N, denoted $\hat{N}^{\mathrm{opt}}$, from $[[4n^{1/3}] + 4, \min([10n^{1/3}], [n/2] - 1)]$, which minimises the BIC value $\mathrm{BIC}(N) = \log(\hat{\sigma}_1^2) + (N + 1) \times \log(n)/n$. By letting $p = 1$ and replacing T_{1n} with $\hat{\delta}_{1j}$, for $0 \le j \le N - 1$ in Equation (11), we obtain the p-value $p_{\mathrm{value},1,j}$ for each $\hat{\delta}_{1j}$. The jump locations τ_i, $1 \le i \le k$, are estimated by $\hat{\tau}_i = (t_{l_i} + t_{l_i+1})/2$, for $p_{\mathrm{value},1,l_i} < \alpha$, with $\hat{c}_i = \hat{m}_1(t_{l_i+1}) - \hat{m}_1(t_{l_i})$ as the estimated jump magnitudes, for $0 \le l_1, \ldots, l_k \le N - 1$. It is apparent that for $\tau_i \in [t_{l_i}, t_{l_i+1}]$, $\hat{\tau}_i \to \tau_i$, $1 \le i \le k$, as $n \to \infty$.

4. Examples

4.1. *Simulation example*

Here, we examine the finite-sample performance of the procedure described in Section 3 where $m(x)$ has at most one jump. The data set is generated from model (1) with

Table 1. Powers calculated from the test statistic T_{pn} defined in Equation (7) by constant and linear splines, respectively, over $n_s = 500$ replications.

c	σ	Sample size n	$\hat{\beta}_2(c)$ $\alpha = 0.05$	$\hat{\beta}_2(c)$ $\alpha = 0.01$	$\hat{\beta}_1(c)$ $\alpha = 0.05$	$\hat{\beta}_1(c)$ $\alpha = 0.01$
0	0.2	200	0.100	0.032	0.640	0.280
		600	0.062	0.014	0.390	0.140
		1000	0.046	0.010	0.220	0.050
	0.5	200	0.058	0.012	0.220	0.070
		600	0.054	0.006	0.180	0.040
		1000	0.050	0.010	0.120	0.030
2	0.2	200	1.000	0.998	1.000	1.000
		600	1.000	1.000	1.000	1.000
		1000	1.000	1.000	1.000	1.000
	0.5	200	0.942	0.776	0.890	0.680
		600	1.000	0.980	1.000	0.970
		1000	1.000	1.000	1.000	1.000

Table 2. Computing time (in seconds) per replication over $n_s = 500$ replications of generating data and detecting jump by constant and linear spline methods.

Sample size n	Constant	Linear
200	0.04	0.06
600	0.21	0.30
1000	0.55	0.60

$X \sim U[-1/2, 1/2]$, $\varepsilon \sim N(0, 1)$, and with $m(x) = \sin(2\pi x) + c \times I(\tau_1 \leq x \leq 1/2)$, for $\tau_1 = \sqrt{2}/4$. The noise level $\sigma = 0.2, 0.5$, sample size $n = 200, 600, 1000$ and significant level $\alpha = 0.05, 0.01$. Let n_s be the number of replications. Denote the asymptotic powers based on constant and linear splines by $\hat{\beta}_p(c)$, $p = 1, 2$, calculated from $\hat{\beta}_p(c) = \sum_{q=1}^{n_s} I[T_{n,p,q} > \{2\log(N - 2p + 2)\}^{1/2} d_{N-2p+2}(\alpha)]/n_s$, where $T_{n,p,q}$ is the qth replication of T_{pn}, with T_{pn} given in Equation (7), and $d_N(\alpha)$ given in Equation (9), for $p = 1, 2$. Table 1 shows values of $\hat{\beta}_p(c)$ for $c = 0$ and $c = 2$.

In Table 1, $\hat{\beta}_p(2)$, $p = 1, 2$ approach to 1 rapidly. Meanwhile $\hat{\beta}_2(0)$ approaches α as the sample size increases, which shows a very positive confirmation of Theorem 2.1, in contrast to $\hat{\beta}_1(0)$, the convergent rate of which is much slower, indicating that the linear spline method outperforms the constant spline method. Table 1 also shows that the noise level has more influence on the constant spline method than the linear spline method. Table 2 shows the average computing time (in seconds) of generating data and detecting jumps with constant and linear spline methods, which are comparable.

There are 500 replications for $n = 200, 600$ satisfying $p_{\text{value},2} < \alpha = 0.05$, with $p_{\text{value},2}$ given in Equation (11), when $c = 2$, $n_s = 500$. Figure 1 shows the kernel estimators of the densities of $\hat{\tau}_1$ and $\hat{c}_1$ given in Section 3 with sample size $n = 200$ (thick lines) and $n = 600$ (median lines) at $\sigma = 0.2$. The vertical lines at $\sqrt{2}/4$ and 2 are the standard lines for comparing $\hat{\tau}_1$ with τ_1 and $\hat{c}_1$ with c_1, respectively. One clearly sees that both of the centres of the density plots move towards the standard lines with a much narrower spread when the sample size n is increased. The frequencies over 500 replications for τ_1 falling between t_{l_1} and t_{l_1+1} described in Section 3 are 0.994 and 1 for $n = 200$ and 600, respectively.

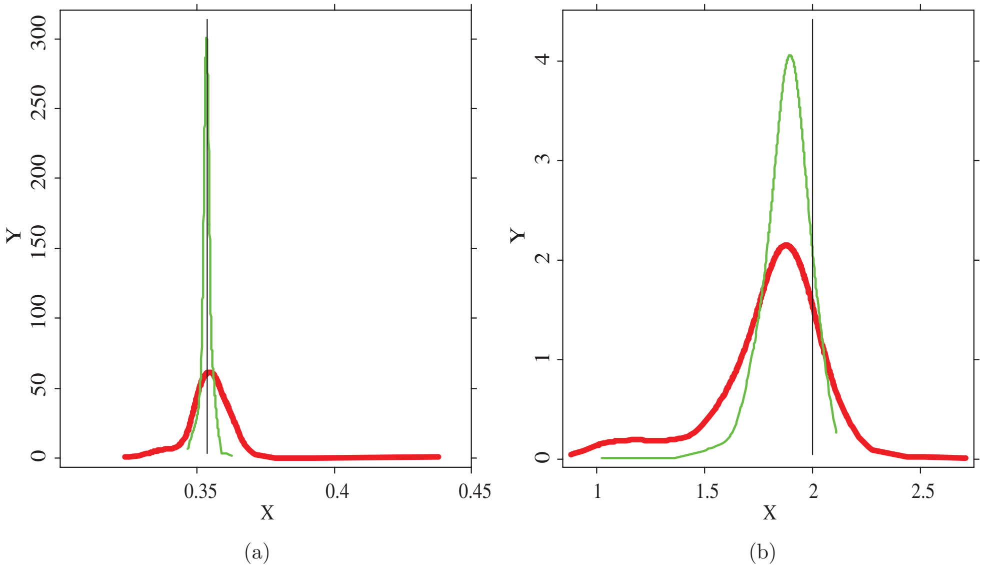

Figure 1. Kernel density plots of $\hat{\tau}_1$ in (a) and $\hat{c}_1$ in (b) over 500 replications of sample size $n = 200$ (thick solid) and $n = 600$ (solid) for which $\mathcal{H}_0$ is rejected.

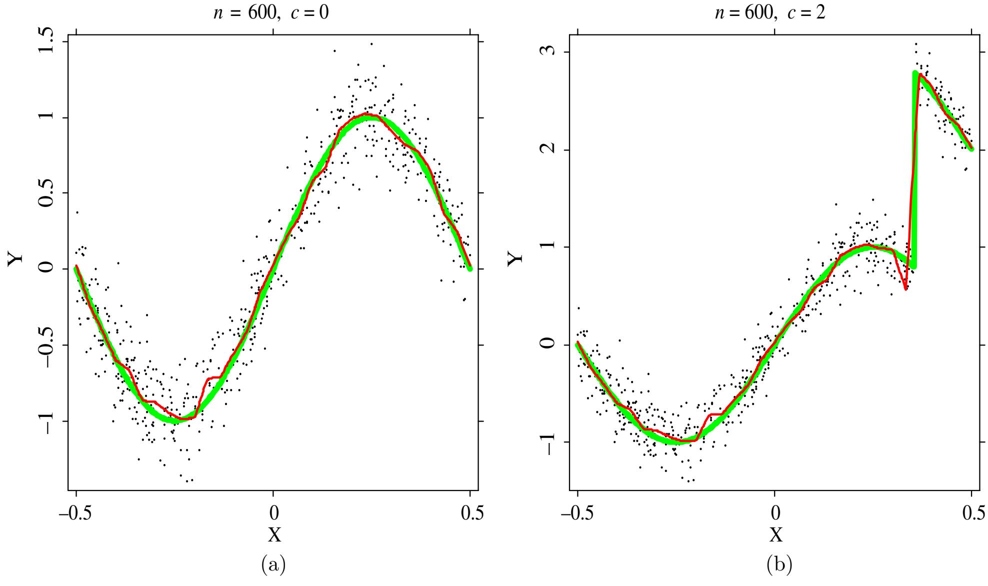

Figure 2. Plots of the true function $m(x)$ (thick solid curve), spline estimator $\hat{m}_2(x)$ (solid curve) and the data scatter plots at $\sigma = 0.2$.

For the visual impression of the actual function estimates, at noise level $\sigma = 0.2$ with sample size $n = 600$, we plot the spline estimator $\hat{m}_2(x)$ (solid curves) for the true functions $m(x)$ (thick solid curves) in Figure 2. The spline estimators seem rather satisfactory.

4.2. *Real-data analysis*

We apply the jump detection procedures in Section 3 to the thickness of pennies data set given in Scott (1992), which consists of measurements in mm of the thickness of 90 US Lincoln pennies.

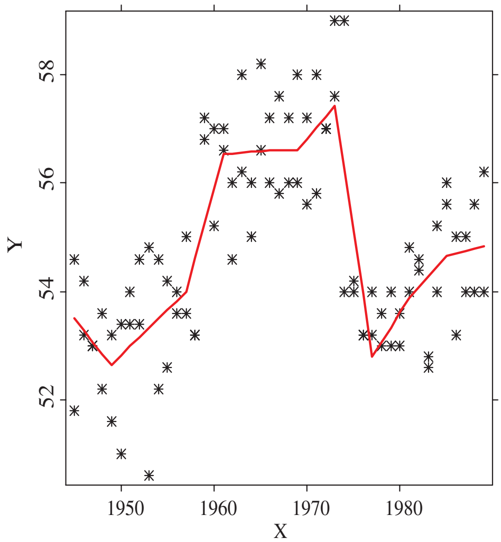

Figure 3. The thickness of pennies data (points) and the spline estimator $\hat{m}_2(x)$.

There are two measurements taken as the response variable Y each year, from 1945 through 1989. Penny thickness was reduced in World War II, restored to its original thickness sometime around 1960 and reduced again in the 1970s. The asymptotic p-value $p_{\text{value},2} < 10^{-20}$. Two jump points are detected with the p-values 0.014468 and 0.00077337, located around the year 1958 with increased magnitude 2.80 and around the year 1974 with decreased magnitude 3.75, respectively, which confirms the result in Gijbels et al. (2007). Figure 3 depicts the data points and the spline estimator $\hat{m}_2(x)$ (solid line), which visually confirm these findings. Findings from both simulated and real data demonstrate the effectiveness of our approach in detecting the existence of jumps. The plots of $\hat{m}_2(x)$ in Figures 2 and 3 give an outline of the true function, without breaking the curve at the jumps. Obtaining a jump-preserving spline estimator of the true non-smooth function is beyond the scope of this paper, but makes an interesting topic for further research.

Acknowledgements

This work was supported in part by an NSF grant DMS 0706518 and is part of the first author's dissertation under the supervision of the second author. The detailed and insightful comments from Editor Suojin Wang, the Associate Editor and two referees are gratefully acknowledged.

References

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

Claeskens, G., and Van Keilegom, I. (2003), 'Bootstrap Confidence Bands for Regression Curves and Their Derivatives', *The Annals of Statistics*, 31, 1852–1884.

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

Gijbels, I., Lambert, A., and Qiu, P.H. (2007), 'Jump-preserving Regression and Smoothing Using Local Linear Fitting: A Compromise', *Annals of the Institute Statistical Mathematics*, 59, 235–272.

Hall, P., and Patil, P. (1995), 'Formulae for Mean Integrated Squared Error of Nonlinear Wavelet-based Density Estimators', *The Annals of Statistics*, 23, 905–928.

Härdle, W. (1989), 'Asymptotic Maximal Deviation of M-smoothers', *Journal of Multivariate Analysis*, 29, 163–179.

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

Huang, J.Z. (2003), 'Local Asymptotics for Polynomial Spline Regression', *The Annals of Statistics*, 31, 1600–1635.

Joo, J., and Qiu, P. (2009), 'Jump Detection in a Regression Curve and Its Derivative', *Technometrics*, 51, 289–305.
Kang, K.H., Koo, J.Y., and Park, C.W. (2000), 'Kernel Estimation of Discontinuous Regression Function', *Statistics and Probability Letters*, 47, 277–285.
Kılıç, E. (2008), 'Explicit Formula for the Inverse of a Tridiagonal Matrix by Backward Continued Fractions', *Applied Mathematics and Computation*, 197, 345–357.
Koo, J.Y. (1997), 'Spline Estimation of Discontinuous Regression Functions', *Journal of Computational and Graphical Statistics*, 6, 266–284.
Leadbetter, M.R., Lindgren, G., and Rootzén, H. (1983), *Extremes and Related Properties of Random Sequences and Processes*, New York: Springer-Verlag.
Mack, Y.P., and Silverman, B.W. (1982), 'Weak and Strong Uniform Consistency of Kernel Regression Estimates', *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 61, 405–415.
Müller, H.G. (1992), 'Change-points in Nonparametric Regression Analysis', *The Annals of Statistics*, 20, 737–761.
Müller, H.G., and Song, K.S. (1997), 'Two-stage Change-point Estimators in Smooth Regression Models', *Statistics and Probability Letters*, 34, 323–335.
Park, C., and Kim, W. (2004), 'Estimation of a Regression Function with a Sharp Change Point Using Boundary Wavelets', *Statistics and Probability Letters*, 66, 435–448.
Park, C., and Kim, W. (2006), 'Wavelet Estimation of a Regression Function with a Sharp Change Point in a Random Design', *Journal of Statistical Planning Inference*, 136, 2381–2394.
Qiu, P.H. (1994), 'Estimation of the Number of Jumps of the Jump Regression Functions', *Communications in Statistics – Theory and Methods*, 23, 2141–2155.
Qiu, P.H. (2003), 'A Jump-preserving Curve Fitting Procedure Based on Local Piecewise-linear Kernel Estimation', *Journal of Nonparametric Statistics*, 15, 437–453.
Qiu, P.H. (2005), '*Image Processing and Jump Regression Analysis*, New York: Wiley.
Qiu, P.H. (2007), 'Jump Surface Estimation, Edge Detection, and Image Restoration', *Journal of the American Statistical Association*, 102, 745–756.
Qiu, P.H., and Yandell, B. (1998), 'A Local Polynomial Jump Detection Algorithm in Nonparametric Regression', *Technometrics*, 40, 141–152.
Qiu, P.H., Asano, C., and Li, X. (1991), 'Estimation of Jump Regression Function', *Bulletin of Informatics and Cybernetics*, 24, 197–212.
Scott, D.W. (1992), *Multivariate Density Estimation. Theory, Practice, and Visualization*, New York: Wiley.
Shiau, J. (1987), 'A Note on MSE Coverage Intervals in a Partial Spline Model', *Communications in Statistics-Theory and Methods*, 16, 1851–1866.
Wang, Y. (1995), 'Jump and Sharp Cusp Detection by Wavelets', *Biometrika*, 82, 385–397.
Wang, J., and Yang, L.J. (2009), 'Polynomial Spline Confidence Bands for Regression Curves', *Statistica Sinica*, 19, 325–342.
Wu, J.S., and Chu, C.K. (1993), 'Kernel-type Estimators of Jump Points and Values of a Regression Function', *The Annals of Statistics*, 3, 1545–1566.
Xia, Y.C. (1998), 'Bias-corrected Confidence Bands in Nonparametric Regression', *Journal of the Royal Statistical Society Series B*, 60, 797–811.
Xue, L., and Yang, L.J. (2006), 'Additive Coefficient Modeling Via Polynomial Spline', *Statistica Sinica*, 16, 1423–1446.
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

A.1. *Preliminaries*

Denote by $\|\cdot\|_\infty$ the supremum norm of a function r on [0, 1], i.e. $\|r\|_\infty = \sup_{x\in[0,1]} |r(x)|$. We denote by the same letters c, C, any positive constants without distinction. The following extension of Leadbetter et al. (1983, Theorem 6.2.1) is a key result on the absolute maximum of discrete-time Gaussian processes.

LEMMA A1 *Let $\xi_1^{(n)}, \ldots, \xi_n^{(n)}$ have jointly normal distribution with $E\xi_i^{(n)} \equiv 0$, $E(\xi_i^{(n)})^2 \equiv 1$, $1 \le i \le n$ and there exist constants $C > 0, a > 1, r \in (0, 1)$ such that the correlations $r_{ij} = r_{ij}^{(n)} = E\xi_i^{(n)}\xi_j^{(n)}$, $1 \le i \ne j \le n$ satisfy $|r_{ij}| \le \min(r, Ca^{-|i-j|})$ for $1 \le i \ne j \le n$. Then the absolute maximum $M_{n,\xi} = \max\{|\xi_1^{(n)}|, \ldots, |\xi_n^{(n)}|\}$ satisfies for any $\tau \in R$, $P(M_{n,\xi} \le \tau/a_n + b_n) \to \exp(-2e^{-\tau})$, as $n \to \infty$, where $a_n = (2\log n)^{1/2}$, $b_n = a_n - (1/2)a_n^{-1}(\log\log n + \log 4\pi)$.*

Proof Take $\varepsilon > 0$ such that $(2-\varepsilon)(1+r)^{-1} = 1+\delta$, for some $\delta > 0$. Let $\tau_n = \tau/a_n + b_n$, then $\tau_n^2/(2\log n) \to 1$, as $n \to \infty$, so for large n, $\tau_n^2 > (2-\varepsilon)\log n$. By the condition $|r_{ij}| \le \min(r, Ca^{-|i-j|}) < 1$ for $i \ne j$, one has $|r_{ij}|(1-|r_{ij}|^2)^{-1/2} \le Ca^{-|i-j|}(1-r^2)^{-1/2}$. Let $M_{n,\eta} = \max\{|\eta_1|, \ldots, |\eta_n|\}$, where $\eta_1, \ldots, \eta_n$ are i.i.d. copies of $N(0, 1)$. By Leadbetter et al. (1983, Theorem 1.5.3), $P(M_{n,\eta} \le \tau_n) \to \exp(-2e^{-\tau})$, as $n \to \infty$. The normal comparison lemma

(Leadbetter et al. (1983, Lemma 11.1.2) entails that

$$|P(-\tau_n < \xi_j^{(n)} \le \tau_n \text{ for } j=1,\dots,n) - P(-\tau_n < \eta_j \le \tau_n \text{ for } j=1,\dots,n)|$$
$$\le \left(\frac{4}{2\pi}\right) \sum_{1\le j<i\le n} |r_{ij}|(1-|r_{ij}|^2)^{-1/2} \exp\left\{\frac{-\tau_n^2}{1+r_{ij}}\right\}.$$

$$|P(M_{n,\xi} \le \tau_n) - P(M_{n,\eta} \le \tau_n)| \le \frac{4}{2\pi} \sum_{1\le i<j\le n} Ca^{-|i-j|}(1-r^2)^{-1/2} \exp\left(\frac{-\tau_n^2}{1+r}\right)$$
$$\le \left(\frac{4}{2\pi}\right) C(1-r^2)^{-1/2} \sum_{1\le j<i\le n} a^{-(i-j)} \exp\{-(2-\varepsilon)(1+r)^{-1}\log n\}$$
$$= \left(\frac{4}{2\pi}\right) C(1-r^2)^{-1/2} \sum_{k=1}^{n-1} (n-k)a^{-k}n^{-1-\delta} \le Cn^{-\delta} \to 0, \text{ as } n\to\infty,$$

for large n, hence $P(M_{n,\xi} \le \tau_n) \to \exp(-2\mathrm{e}^{-\tau})$, as $n \to \infty$. ■

We break the estimation error $\hat{m}_p(x) - m(x)$ into bias and noise. $\hat{m}_p(x)$ defined in Equation (2) can be written as $\hat{m}_p(x) \equiv \sum_{j=1-p}^{N} \hat{\lambda}_{j,p} B_{j,p}(x)$, where the coefficients $\{\hat{\lambda}_{1-p,p},\dots,\hat{\lambda}_{N,p}\}^{\mathrm{T}}$ are solutions of the following least squares problem

$$\{\hat{\lambda}_{1-p,p},\dots,\hat{\lambda}_{N,p}\}^{\mathrm{T}} = \underset{\{\lambda_{1-p,p},\dots,\lambda_{N,p}\}\in R^{N+p}}{\operatorname{argmin}} \sum_{i=1}^{n} \left\{ Y_i - \sum_{j=1-p}^{N} \lambda_{j,p} B_{j,p}(X_i) \right\}^2. \quad \text{(A1)}$$

Projecting the relationship in model (1) leads to the following decomposition in $G^{(p-2)}$

$$\hat{m}_p(x) = \tilde{m}_p(x) + \tilde{\varepsilon}_p(x), \quad \text{(A2)}$$

$$\tilde{m}_p(x) = \sum_{j=1-p}^{N} \tilde{\lambda}_{j,p} B_{j,p}(x), \quad \tilde{\varepsilon}_p(x) = \sum_{j=1-p}^{N} \tilde{a}_{j,p} B_{j,p}(x). \quad \text{(A3)}$$

The vectors $\{\tilde{\lambda}_{1-p,p},\dots,\tilde{\lambda}_{N,p}\}^{\mathrm{T}}$ and $\{\tilde{a}_{1-p,p},\dots,\tilde{a}_{N,p}\}^{\mathrm{T}}$ are solutions to Equation (A1) with Y_i replaced by $m(X_i)$ and $\sigma\varepsilon_i$, respectively.

Next lemma is from de Boor (2001, p. 149) and Huang (2003, Theorem 5.1).

LEMMA A2 *There are constants $C_p > 0$, $p \ge 1$ such that for any $m \in C^{(p)}[0,1]$, there exists a function $g \in G^{(p-2)}[0,1]$ such that $\|g - m\|_\infty \le C_p \|m^{(p)}\|_\infty h^p$ and $\tilde{m}(x)$ defined in Equation* (A3), *with probability approaching* 1, *satisfies* $\|\tilde{m}_p(x) - m(x)\|_\infty = O(h^p)$.

A.2. *Proof of Theorem 2.1 for $p = 1$*

For $x \in [0,1]$, define its location and relative position indices $j(x), \delta(x)$ as $j(x) = j_n(x) = \min\{[x/h], N\}, \delta(x) = \{x - t_{j(x)}\}h^{-1}$. It is clear that $t_{j_n(x)} \le x < t_{j_n(x)+1}$, $0 \le \delta(x) < 1, \forall x \in [0,1)$ and $\delta(1) = 1$. Since $\langle B_{j',1}, B_{j,1}\rangle_n = 0$ unless $j = j'$, for $B_{j,1}(x)$ given in Equation (4). $\tilde{\varepsilon}_1(x)$ in Equation (A3) can be written as

$$\tilde{\varepsilon}_1(x) = \sum_{j=0}^{N} \varepsilon_{j,1}^* B_{j,1}(x)\|B_j\|_{2,n}^{-2}, \quad \varepsilon_{j,1}^* = n^{-1}\sum_{i=1}^{n} B_{j,1}(X_i)\sigma\varepsilon_i.$$

Let $\hat{\varepsilon}_1(x) = \sum_{j=0}^{N} \varepsilon_{j,1}^* B_{j,1}(x)$, it is easy to prove that $E\{\hat{\varepsilon}_1(x)\}^2 = \sigma^2/(nh)$ and for $0 \le j \le N-1$, $E\{\hat{\varepsilon}_1(t_{j+1}) - \hat{\varepsilon}_1(t_j)\}^2 = 2\sigma^2/(nh)$, which is $\sigma_{n,1}^2$ defined in Equation (8). Define for $0 \le j \le N-1$, $\tilde{\xi}_{n,1,j} = \sigma_{n,1}^{-1}\{\tilde{\varepsilon}_1(t_{j+1}) - \tilde{\varepsilon}_1(t_j)\}$, $\hat{\xi}_{n,1,j} = \sigma_{n,1}^{-1}\{\hat{\varepsilon}_1(t_{j+1}) - \hat{\varepsilon}_1(t_j)\}$.

LEMMA A3 *Under Assumptions* (A2)–(A4), *as* $n \to \infty$, $\sup_{0\le j\le N-1} |\tilde{\xi}_{n,1,j} - \hat{\xi}_{n,1,j}| = O_{\mathrm{a.s}}(n^{-1/2}h^{-1/2}\log n) = o_{\mathrm{a.s}}(1)$.

LEMMA A4 *Under Assumptions* (A2)–(A4), *there exist* $\{\hat{\xi}_{n,1,j}^{(k)}\}_{j=0}^{N-1}$, $k = 1, 2, 3$ *such that as* $n \to \infty$, $\sup_{0\le j\le N-1} |\hat{\xi}_{n,1,j} - \hat{\xi}_{n,1,j}^{(1)}| + \sup_{0\le j\le N-1} |\hat{\xi}_{n,1,j}^{(2)} - \hat{\xi}_{n,1,j}^{(3)}| = o_{a.s}(1)$. $\{\hat{\xi}_{n,1,j}^{(1)}\}_{j=0}^{N-1}$ *has the same probability distribution as* $\{\hat{\xi}_{n,1,j}^{(2)}\}_{j=0}^{N-1}$, *and* $\{\hat{\xi}_{n,1,j}^{(3)}\}_{j=0}^{N-1}$ *is a Gaussian process with mean* 0, *variance* 1 *and covariance*

$$\mathrm{cov}\{\hat{\xi}_{n,1,j}^{(3)}, \hat{\xi}_{n,1,j'}^{(3)}\} = \begin{cases} -1/2 & \text{for } |j - j'| = 1, \\ 0 & \text{for } |j - j'| > 1. \end{cases}$$

Lemmas A3 and A4 follow from Appendix A of Wang and Yang (2009).

Proof of Theorem 2.1 *for* $p = 1$ It is clear from Lemma A4 that the Gaussian process $\{\hat{\xi}_{n,1,j}^{(3)}\}_{j=0}^{N-1}$ satisfies the conditions of Lemma A1, hence as $n \to \infty$, $P\{(\sup_{0\le j\le N-1} |\hat{\xi}_{n,1,j}^{(3)}| \le \tau/a_N + b_N)\} \to \exp(-2\mathrm{e}^{-\tau})$. By letting $\tau = -\log\{-(1/2)\log(1-\alpha)\}$ and using the definition of a_N, b_N and $d_N(\alpha)$, we obtain

$$\lim_{n\to\infty} P\left\{\sup_{0\le j\le N-1} |\hat{\xi}_{n,1,j}^{(3)}| \le -\log\left\{-\left(\frac{1}{2}\right)\log(1-\alpha)\right\}(2\log N)^{-1/2} + (2\log N)^{1/2} - (1/2)(2\log N)^{-1/2}(\log\log N + \log 4\pi)\right\} = 1-\alpha$$

$$\lim_{n\to\infty} P\left\{\sup_{0\le j\le N-1} |\hat{\xi}_{n,1,j}^{(3)}| \le (2\log N)^{1/2} d_N(\alpha)\right\} = 1-\alpha.$$

By Lemmas A3 and A4, we have $\lim_{n\to\infty} P\{\sup_{0\le j\le N-1} |\tilde{\xi}_{n,1,j}| \le (2\log N)^{1/2} d_N(\alpha)\} = 1-\alpha$, which implies for $0 \le j \le N-1$

$$\lim_{n\to\infty} P\left\{d_N(\alpha)^{-1}(2\log N)^{-1/2}\sigma_{n,1}^{-1} \sup_{0\le j\le N-1} |\tilde{\varepsilon}_1(t_{j+1}) - \tilde{\varepsilon}_1(t_j)| \le 1\right\} = 1-\alpha.$$

Lemma A2 entails that under $\mathcal{H}_0$ $\sup_{0\le j\le N-1} |\tilde{m}_1(t_j) - m(t_j)| = \mathrm{O}_p(h)$ and $\sup_{0\le j\le N-1} |m(t_{j+1}) - m(t_j)| = \mathrm{O}_p(h)$, which imply that

$$\sigma_{n,1}^{-1}(\log N)^{-1/2} \sup_{0\le j\le N-1} |m(t_{j+1}) - m(t_j)| = \mathrm{O}_p\{(nh)^{1/2}(\log N)^{-1/2}h\} = \mathrm{o}_p\{(\log n)^{-2}\}.$$

Thus, by Equation (A2), $\hat{m}_1(t_{j+1}) - \hat{m}_1(t_j) = \{\tilde{m}_1(t_{j+1}) - m(t_{j+1})\} - \{\tilde{m}_1(t_j) - m\{t_j\}\} + \{m(t_{j+1}) - m(t_j)\} + \{\tilde{\varepsilon}_1(t_{j+1}) - \tilde{\varepsilon}_1(t_j)\}$, then for $d_N(\alpha)$ defined in Equation (9),

$$\lim_{n\to\infty} P\left\{\sup_{0\le j\le N-1} |\hat{m}_1(t_{j+1}) - \hat{m}_1(t_j)| \le \sigma_{n,1}(2\log N)^{1/2} d_N(\alpha)\right\} = \lim_{n\to\infty} P\left\{\sup_{0\le j\le N-1} |\tilde{\varepsilon}_1(t_{j+1}) - \tilde{\varepsilon}_1(t_j)| \le \sigma_{n,1}(2\log N)^{1/2} d_N(\alpha)\right\} = 1-\alpha.$$

■

Appendix 2

B.1. *Preliminaries*

The following lemma from Wang and Yang (2009) shows that multiplication by **V** defined in Equation (5) behaves similarly to multiplication by a constant. We use $|T|$ to denote the maximal absolute value of any matrix T.

LEMMA B1 *Given matrix* $\Omega = \mathbf{V} + \Gamma$, *in which* $\Gamma = (\gamma_{jj'})_{j,j'=-1}^{N}$ *satisfies* $\gamma_{jj'} \equiv 0$ *if* $|j - j'| > 1$ *and* $|\Gamma| \xrightarrow{p} 0$. *Then there exist constants* $c, C > 0$ *independent of* n *and* Γ, *such that with probability approaching* 1

$$c|\boldsymbol{\xi}| \le |\Omega\boldsymbol{\xi}| \le C|\boldsymbol{\xi}|, \quad C^{-1}|\boldsymbol{\xi}| \le |\Omega^{-1}\boldsymbol{\xi}| \le c^{-1}|\boldsymbol{\xi}|, \quad \forall \boldsymbol{\xi} \in R^{N+2}. \tag{B1}$$

To prove Theorem 2.1 for $p = 2$, we need the below result (Kılıç 2008, Corollary 16), which gives an explicit formula for the inverse of symmetric tridiagonal matrix.

Lemma B2 *For any symmetric tridiagonal matrix* $G_n = \begin{pmatrix} x_1 & y_1 & & \\ y_1 & x_2 & \ddots & \\ & \ddots & \ddots & y_{n-1} \\ & & y_{n-1} & x_n \end{pmatrix}$, *the inverse of the matrix* G_n, $G_n^{-1} = [w_{ij}]$ *is given by*

$$w_{ij} = \begin{cases} (C_i^b)^{-1} + \sum_{k=i+1}^{n} (C_k^b)^{-1} \prod_{t=i}^{k-1} (C_t^b)^{-2} y_t^2, & i = j \\ (-1)^{i+j} \left\{ \prod_{t=j}^{i-1} (C_t^b)^{-1} y_t \right\} w_{ii}, & i > j \end{cases}$$

in which $C_1^b = x_1$, $C_n^b = x_n - (C_{n-1}^b)^{-1} y_{n-1}^2$, $n = 2, 3, \ldots$.

Lemma B3 *There exists a constant* $C_s > 0$, *such that* $\sum_{j=-1}^{N} |s_{j'j}| \le C_s$, *and* $17/16 \le s_{jj} \le 5/4$, *where* $s_{j'j}$, $0 \le j, j' \le N-1$, *is the element of* **S** *defined in Equation* (6).

Proof By Equation (B1), let $\tilde{\boldsymbol{\xi}}_{j'} = \{\mathrm{sgn}(s_{j'j})\}_{j=-1}^{N}$, then $\sum_{j=-1}^{N} |s_{j'j}| \le |\mathbf{S}\tilde{\boldsymbol{\xi}}_{j'}| \le C_s |\tilde{\boldsymbol{\xi}}_{j'}| = C_s, \forall j' = -1, 0, \ldots, N$. Applying Lemma B2 to **V** with $x_{-1} = \cdots = x_N = 1$, $y_{-1} = y_{N-1} = \sqrt{2}/4$, $y_0 = \cdots = y_{N-1} = 1/4$, we have $s_{jj} = (C_j^b)^{-1} + \sum_{k=j+1}^{N} (C_k^b)^{-1} \prod_{t=j}^{k-1} (C_t^b)^{-2} y_t^2$. By mathematical induction, one obtains that $9/10 \le C_j^b \le 1$, for $0 \le j \le N-1$. Therefore, $1 \le (C_j^b)^{-1} \le 10/9$, and for $0 \le j \le N-1$,

$$s_{jj} \ge 1 + \sum_{k=j+1}^{N} \prod_{t=j}^{k-1} y_t^2 \ge 1 + \sum_{k=j+1}^{N} (16)^{-(k-j)} \ge \frac{17}{16},$$

$$s_{jj} \le \frac{10}{9} + \left(\frac{10}{9}\right) \sum_{k=j+1}^{N} \left(\frac{1}{9}\right)^{k-j} \le \frac{5}{4}.$$

■

B.2. *Variance calculation*

Vector $\tilde{\mathbf{a}}_2 = (\tilde{a}_{-1,2}, \ldots, \tilde{a}_{N,2})^{\mathrm{T}}$ given in Equation (A3) solves the normal equations,

$$(\langle B_{j,2}, B_{j',2}\rangle_n)_{j,j'=-1}^{N} \tilde{\mathbf{a}}_2 = \left(n^{-1} \sum_{i=1}^{n} B_{j,2}(X_i)\sigma\varepsilon_i \right)_{j=-1}^{N},$$

for $B_{j,2}(x)$ given in Equation (4). In other words, $\tilde{\mathbf{a}}_2 = (V + \tilde{B})^{-1} (n^{-1} \sum_{i=1}^{n} B_{j,2}(X_i)\sigma\varepsilon_i)_{j=-1}^{N}$, where $\tilde{B} = (\langle B_{j,2}, B_{j',2}\rangle_n)_{j,j'=-1}^{N} - V$ satisfies $|\tilde{B}| = \mathrm{O}_p(\sqrt{n^{-1}h^{-1}\log(n)})$ according to Section B.2 of the supplement to Wang and Yang (2009).

Now define $\hat{\mathbf{a}}_2 = (\hat{a}_{-1,2}, \ldots, \hat{a}_{N,2})^{\mathrm{T}}$ by replacing $(\mathbf{V} + \tilde{B})^{-1}$ with $\mathbf{V}^{-1} = \mathbf{S}$ in the above formula, i.e. $\hat{\mathbf{a}}_2 = (\sum_{j=-1}^{N} s_{j'j} n^{-1} \sum_{i=1}^{n} B_{j,2}(X_i)\sigma\varepsilon_i)_{j'=-1,\ldots,N}$, and define for $x \in [0, 1]$

$$\hat{\varepsilon}_2(x) = \sum_{j=-1}^{N} \hat{a}_{j,2} B_{j,2}(x) = \sum_{j,j'=-1}^{N} s_{j'j} n^{-1} \sum_{i=1}^{n} B_{j,2}(X_i)\sigma\varepsilon_i B_{j',2}(x),$$

$$\hat{\xi}_{2,j} = \{\hat{\varepsilon}_2(t_{j+1}) + \hat{\varepsilon}_2(t_{j-1})\}/2 - \hat{\varepsilon}_2(t_j), 2 \le j \le N-1, \tag{B2}$$

$$\mathbf{D} = \begin{pmatrix} 0 & 1 & -2 & 1 & & & 0 & 0 \\ \vdots & & \ddots & \ddots & \ddots & & & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & & \vdots \\ 0 & 0 & & & 1 & -2 & 1 & 0 \end{pmatrix}_{(N-2)\times(N+2)}. \tag{B3}$$

LEMMA B4 *With* **S** *and* **D** *given in Equation* (6) *and* (B3), $\{\hat{\xi}_{2,j}\}_{j=2}^{N-1}$ *has a covariance matrix*

$$[\text{cov}(\hat{\xi}_{2,j}, \hat{\xi}_{2,j'})]_{j,j'=2}^{N-1} = \sigma^2 \left(\frac{8nh}{3}\right)^{-1} \mathbf{DSD}^{\mathrm{T}}. \tag{B4}$$

Proof For $0 \le j, j' \le N+1$, $\hat{\varepsilon}_2(t_j) = \sum_{k,k'=-1}^{N} s_{k'k} n^{-1} \sum_{i=1}^{n} B_{k,2}(X_i)\sigma\varepsilon_i B_{k',2}(t_j)$

$$
\begin{aligned}
&= \sigma \sum_{k=-1}^{N} n^{-1} \sum_{i=1}^{n} B_{k,2}(X_i)\varepsilon_i s_{(j-1),k} B_{(j-1),2}(t_j). \\
E[\hat{\varepsilon}_2(t_j)\hat{\varepsilon}_2(t_{j'})] &= \sigma^2 E\left[\sum_{k=-1}^{N} n^{-1} \sum_{i=1}^{n} B_{k,2}(X_i)\varepsilon_i s_{j-1,k} B_{j-1,2}(t_j)\right] \\
&\quad \times \left[\sum_{k'=-1}^{N} n^{-1} \sum_{i=1}^{n} B_{k',2}(X_i)\varepsilon_i s_{j'-1,k'} B_{j'-1,2}\left(t_{j'}\right)\right] \\
&= \sigma^2 n^{-1} \sum_{k,k'=-1}^{N} B_{j-1,2}(t_j) B_{j'-1,2}\left(t_{j'}\right) s_{j-1,k} s_{j'-1,k'} E B_{k,2}(X) B_{k',2}(X) \\
&= \sigma^2 n^{-1} \sum_{k,k'=-1}^{N} B_{j-1,2}(t_j) B_{j'-1,2}\left(t_{j'}\right) s_{j-1,k} s_{j'-1,k'} v_{k,k'} \\
&= \sigma^2 n^{-1} B_{j-1,2}(t_j) B_{j'-1,2}\left(t_{j'}\right) \sum_{k'=-1}^{N} s_{j'-1,k'} \sum_{k=-1}^{N} s_{j-1,k} v_{k,k'} \\
&= \sigma^2 n^{-1} B_{j-1,2}(t_j) B_{j'-1,2}\left(t_{j'}\right) \sum_{k'=-1}^{N} s_{j'-1,k'} \delta_{j-1,k'} \\
&= \sigma^2 n^{-1} B_{j-1,2}(t_j) B_{j'-1,2}\left(t_{j'}\right) s_{j'-1,j-1} = \sigma^2 n^{-1} d_{j-1,n}^{-1/2} d_{j'-1,n}^{-1/2} s_{j'-1,j-1}.
\end{aligned}
$$

By definitions of $\hat{\xi}_{2,j}$ and $d_{j,n}$ in Equations (B2) and (3), for $2 \le j, j' \le N-1$, $E(\hat{\xi}_{2,j}\hat{\xi}_{2,j'})$ is

$$
\begin{aligned}
&\sigma^2 \left(\frac{8nh}{3}\right)^{-1} (s_{j',j} + s_{j'-2,j} - 2s_{j'-1,j} + s_{j',j-2} + s_{j'-2,j-2} - 2s_{j'-1,j-2} - 2s_{j',j-1} - 2s_{j'-2,j-1} + 4s_{j'-1,j-1}) \\
&\quad = \sigma^2 \left(\frac{8nh}{3}\right)^{-1} (1\ \ -2\ \ 1) \begin{pmatrix} s_{j'-2,j-2} & s_{j'-2,j-1} & s_{j'-2,j} \\ s_{j'-1,j-2} & s_{j'-1,j-1} & s_{j'-1,j} \\ s_{j',j-2} & s_{j',j-1} & s_{j',j} \end{pmatrix} (1\ \ -2\ \ 1)^{\mathrm{T}}.
\end{aligned}
$$

Therefore, for $2 \le j, j' \le N-1$, $[\text{cov}(\hat{\xi}_{2,j}, \hat{\xi}_{2,j'})]_{j,j'=2}^{N-1} = \sigma^2(8nh/3)^{-1}\mathbf{DSD}^{\mathrm{T}}$. ■

LEMMA B5 *For* $2 \le j \le N-1$, $\sigma_{n,2,j}^2$ *defined in Equation* (8) *satisfies that* $c_\sigma(8nh/3)^{-1}\sigma^2 \le \sigma_{n,2,j}^2 \le C_\sigma(8nh/3)^{-1}\sigma^2$, *for* $c_\sigma = (65/8)(17/16)$, $C_\sigma = 100/9$.

Proof It follows from Equation (B4) that $\sigma_{n,2,j}^2 = E\hat{\xi}_{2,j}^2$. Then by Lemmas B2 and B4, for $2 \le j \le N-1$, $\{\sigma^2(8nh/3)^{-1}\}^{-1}\sigma_{n,2,j}^2$ is

$$
\begin{aligned}
&s_{j,j} - 4s_{j,j-1} + 2s_{j,j-2} + 4s_{j-1,j-1} - 4s_{j-1,j-2} + s_{j-2,j-2} \\
&\quad = s_{j-2,j-2} + 4\{(C_{j-2}^b)^{-1} y_{j-2} + 1\} s_{j-1,j-1} + \{2(C_{j-2}^b C_{j-1}^b)^{-1} y_{j-2} y_{j-1} + 4(C_{j-1}^b)^{-1} y_{j-1} + 1\} s_{j,j},
\end{aligned}
$$

thus,

$$\sigma^2_{n,2,j} \le \left\{1 + 4\left(\frac{1}{3}+1\right) + \left(\frac{2}{9}+\frac{4}{3}+1\right)\right\}\left(\frac{5}{4}\right)\sigma^2\left(\frac{8nh}{3}\right)^{-1}$$
$$= \left(\frac{100}{9}\right)\left(\frac{8nh}{3}\right)^{-1}\sigma^2,$$
$$\sigma^2_{n,2,j} \ge \left\{1 + 4\left(\frac{1}{4}+1\right) + \left(\frac{2}{16}+1+1\right)\right\}\left(\frac{17}{16}\right)\sigma^2\left(\frac{8nh}{3}\right)^{-1}$$
$$= \left(\frac{65}{8}\right)\left(\frac{17}{16}\right)\left(\frac{8nh}{3}\right)^{-1}\sigma^2.$$

■

B.3. *Proof of Theorem* 2.1 *for* $p = 2$

Several lemmas will be given below for proving Theorem 2.1 for $p = 2$. With $\tilde\varepsilon_2(x)$, $\hat\xi_{2,j}$ and $\sigma_{n,2,j}$ defined in Equations (A3), (B2) and (8), define for $2 \le j \le N-1$

$$\tilde\xi_{n,2,j} = \sigma^{-1}_{n,2,j}\left[\frac{\{\tilde\varepsilon_2(t_{j+1}) + \tilde\varepsilon_2(t_{j-1})\}}{2 - \tilde\varepsilon_2(t_j)}\right], \quad \hat\xi_{n,2,j} = \sigma^{-1}_{n,2,j}\hat\xi_{2,j}, \tag{B5}$$

Lemma B6 *Under Assumptions* (A2)–(A4), *as* $n \to \infty$, $\sup_{2\le j\le N-1}|\hat\xi_{n,2,j} - \tilde\xi_{n,2,j}| = O_{a.s}(\sqrt{\log n/(nh)}) = o_{a.s}(1)$.

Lemma B7 *Under Assumptions* (A2)–(A4), *there exist* $\{\hat\xi^{(k)}_{n,2,j}\}_{j=2}^{N-1}$, $k = 1, 2, 3$, *such that as* $n \to \infty$, $\sup_{2\le j\le N-1}|\hat\xi_{n,2,j} - \hat\xi^{(1)}_{n,2,j}| + \sup_{2\le j\le N-1}|\hat\xi^{(2)}_{n,2,j} - \hat\xi^{(3)}_{n,2,j}| = o_{a.s}(1)$. $\hat\xi^{(1)}_{n,2,j}$ *has the same probability distribution as* $\hat\xi^{(2)}_{n,2,j}$. $\{\hat\xi^{(3)}_{n,2,j}\}$ *is a Gaussian process with mean* 0, *variance* 1, *covariance* $r^\xi_{j,j'} = \mathrm{cov}(\hat\xi^{(3)}_{n,2,j}, \hat\xi^{(3)}_{n,2,j'}) = \sigma^{-1}_{n,2,j}\sigma^{-1}_{n,2,j'}E(\hat\xi_{2,j}\hat\xi_{2,j'})$ *for which there exist constants* $0 < C$, $0 < r < 1$ *such that for large* n, $|r^\xi_{j,j'}| \le \min(r, C3^{-|j-j'|})$, $2 \le j, j' \le N-1$.

Proof We only prove $|r^\xi_{j,j'}| \le \min(r, C3^{-|j-j'|})$. Lemma B6 and the rest of Lemma B7 follow from Appendix 2 of the supplement to Wang and Yang (2009). By Equation (B4),

$$\left\{\sigma^2\left(\frac{8nh}{3}\right)^{-1}\right\}^{-1} E(\hat\xi_{2,j}\hat\xi_{2,j'}) = s_{j',j} + s_{j'-2,j} - 2s_{j'-1,j} + s_{j',j-2} + s_{j'-2,j-2}$$
$$- 2s_{j'-1,j-2} - 2s_{j',j-1} - 2s_{j'-2,j-1} + 4s_{j'-1,j-1}.$$

By Lemma B2, for $-1 \le j' < j \le N$, $s_{j,j'} = (-1)^{j+j'}\prod_{t=j'}^{j-1}(C^b_t)^{-1}y_t s_{jj}$, then for $2 \le j, j' \le N-1$ and $j - j' > 2$, by Lemma B3,

$$\left\{\sigma^2\left(\frac{8nh}{3}\right)^{-1}\right\}^{-1}|E[\hat\xi_{2,j}\hat\xi_{2,j'}]|$$
$$= \left|(-1)^{j+j'}\left\{(C^b_{j'-2})^{-1}y_{j'-2} + 2(C^b_{j'-1})^{-1}y_{j'-1} + 1\right\}\right.$$
$$\left.\times\left\{s_{j-2,j-2} + 2(C^b_{j-2})^{-1}y_{j-2}s_{j-1,j-1} + (C^b_{j-2}C^b_{j-1})^{-1}y_{j-2}y_{j-1}s_{j,j}\right\}\prod_{t=j'}^{j-3}(C^b_t)^{-1}y_t\right|$$
$$\le \left(\frac{5}{4}\right)\left(\frac{1}{3}+\frac{2}{3}+1\right)\left\{1+\frac{2}{3}+\left(\frac{1}{3}\right)^2\right\}3^{-(j-j'-2)} \le 40\times 3^{-(j-j')}.$$

By Lemma B5, $\{\sigma^2(8nh/3)^{-1}\}^{-1}\sigma^2_{n,2,j} \ge (65/8)(17/16)$, for $2 \le j \le N-1$. Therefore, for $2 \le j, j' \le N-1$ and $j - j' > 2$, $|r^\xi_{j,j'}| \le C3^{-(j-j')} \le r < 1$, with $C = 40(8/65)(16/17)$ and $r = 40(8/65)(16/17)/3^3 < 1$. For $j - j' = 1, 2$, the result can be proved following the same procedure above. ■

Proof of Theorem 2.1 *for* $p = 2$ It is clear from Lemma B7 that the Gaussian process $\{\hat{\xi}^{(3)}_{n,2,j}\}_{j=2}^{N-1}$ satisfies the conditions of Lemma A1, hence as $n \to \infty$, $P(\sup_{2\le j\le N-1} |\hat{\xi}^{(3)}_{n,2,j}| \le \tau/a_N + b_N) \to \exp(-2e^{-\tau})$. By Lemmas B6 and B7, with $\tau = -\log\{-(1/2)\log(1-\alpha)\}$ and using the definitions of a_N and b_N, we obtain

$$\lim_{n\to\infty} P\left(\sup_{2\le j\le N-1} |\tilde{\xi}_{n,2,j}| \le \{2\log(N-2)\}^{1/2} d_{N-2}(\alpha)\right) = 1-\alpha,$$

for any $0 < \alpha < 1$, $\hat{\xi}_{n,2,j}$ and $d_N(\alpha)$ defined in Equations (B5) and (9). By Equations (A2) and (B2), $\{\hat{m}_2(t_{j+1}) + \hat{m}_2(t_{j-1})\}/2 - \hat{m}_2(t_j)$ is

$$\frac{\{\tilde{m}_2(t_{j+1}) - m(t_{j+1})\}}{2} + \frac{\{\tilde{m}_2(t_{j-1}) - m(t_{j-1})\}}{2} - \{\tilde{m}_2(t_j) - m(t_j)\} + \left[\frac{\{m(t_{j+1}) + m(t_{j-1})\}}{2} - m(t_j)\right] + \hat{\xi}_{2,j+1}.$$

Now Lemma A2 implies that under $\mathcal{H}_0$, $\|\tilde{m}_2 - m\|_\infty = O_p(h^2)$, hence

$$(nh)^{1/2}\{\log(N-2)\}^{-1/2} \sup_{2\le j\le N-1}\left|\frac{\{\tilde{m}_2(t_{j+1}) - m(t_{j+1})\}}{2} + \frac{\{\tilde{m}_2(t_{j-1}) - m(t_{j-1})\}}{2} - \{\tilde{m}_2(t_j) - m(t_j)\}\right|$$
$$= O_p[(nh)^{1/2}\{\log(N-2)\}^{-1/2}h^2] = o_p\{(\log n)^{-3}\}.$$

By Taylor expansion, $\sup_{2\le j\le N-1} |\{m(t_{j+1}) + m(t_{j-1})\}/2 - m(t_j)| = O_p(h^2)$ under $\mathcal{H}_0$, as $n \to \infty$. Hence, $(nh)^{1/2}\{\log(N-2)\}^{-1/2} \sup_{2\le j\le N-1} |\{m(t_{j+1}) + m(t_{j-1})\}/2 - m(t_j)| = O_p[\sqrt{nh}\{\log(N-2)\}^{-1/2}h^2] = o_p\{(\log n)^{-3}\}$. By the above results, for T_{2n} defined in Equation (7), $\lim_{n\to\infty} P\{T_{2n} \le \{2\log(N-2)\}^{1/2} d_{N-2}(\alpha)\} = 1-\alpha$. ■