Statistica Sinica **29** (2019), 000-000
doi:https://doi.org/10.5705/ss.202017.0033

SIMULTANEOUS CONFIDENCE BANDS FOR MEAN AND VARIANCE FUNCTIONS BASED ON DETERMINISTIC DESIGN

Li Cai[1,2], Rong Liu[3] , Suojin Wang[4] and Lijian Yang[2]

[1]*Soochow University,* [2]*Tsinghua University,* [3]*University of Toledo and* [4]*Texas A&M University*

Abstract: Asymptotically correct simultaneous confidence bands (SCBs) are proposed for the mean and variance functions of a nonparametric regression model based on deterministic designs. The variance estimation is as efficient, up to order $n^{-1/2}$, as an infeasible estimator if the mean function were known. Simulation experiments provide strong evidence that corroborates the asymptotic theory. The proposed SCBs are used to analyze two sets of strata pressure data from the Bullianta Coal Mine in Erdos City, Inner Mongolia, China.

Key words and phrases: Brownian motion, heteroscedasticity, kernel, oracle efficiency, spline, strata pressure.

1. Introduction

Simultaneous confidence intervals (SCIs) have long been recognized as vital tools for inference on the global shape of curves; see, for instance, Stapleton (2009) Section 5.2 for the Scheffé SCIs of a simple linear regression function, and Section 5.3 for Tukey SCIs of a surface of contrasts. In the more complicated context of nonparametric function estimation, SCIs are generally known as simultaneous confidence bands (SCBs), and were first constructed in Bickel and Rosenblatt (1973) for a probability density function, and extended by Johnston (1982) and Härdle (1989) to univariate kernel regression. Xia (1998) proposed bias-corrected SCBs based on local polynomial fitting under the assumption of homoscedasticity, while Härdle and Marron (1991) and Keilegom and Claeskens (2003) studied bootstrap kernel SCBs.

More recently, nonparametric SCB methodology has diversified in both techniques and scope. For instance, Wang and Yang (2009) proposed SCBs for nonparametric regression function based on polynomial splines, which was extended in Song and Yang (2009) to oracally efficient spline SCBs for the conditional variance function. Cai and Yang (2015) improved Song and Yang (2009) by a spline-kernel oracally efficient two-step estimator for the variance function with SCBs. Wang, Cheng and Yang (2013) and Wang et al. (2014) proposed smooth SCBs for cumulative distribution functions. Degras (2011), Cao, Yang and Todem (2012), Ma, Yang and Carroll (2012) and Cao et al. (2016) constructed various SCBs for functional data, while Gu and Yang (2015) established oracle efficiency of an SCB for the single-index link function.

Existing literature on SCBs for nonparametric regression is mostly concerned with the random design model $Y_i = m(X_i) + \sigma(X_i)\varepsilon_i$ with independent and identically distributed (i.i.d.) points $\{(X_i, Y_i)\}_{i=1}^n$ and errors $\{\varepsilon_i\}_{i=1}^n$. Often encountered in applications (e.g., the strata pressure data discussed in Subsection 5.2) is the deterministic design nonparametric regression model

$$Y_i = m\left(\frac{i}{n}\right) + \sigma\left(\frac{i}{n}\right)\varepsilon_i \tag{1.1}$$

in which the Y_i's are responses at equally spaced design points $i/n, 1 \leq i \leq n$, and $\{\varepsilon_i\}_{i=1}^n$ are unobserved i.i.d. random errors with $\mathrm{E}(\varepsilon_1) = 0$, $\mathrm{var}(\varepsilon_1) = 1$. Assume that there are smooth but unknown mean and variance functions $m(x)$ and $\sigma^2(x)$ that satisfy model (1.1) for all n. In this paper, we aim to construct asymptotically correct SCBs for both the mean function $m(x)$ and variance function $\sigma^2(x)$ in model (1.1) without restrictive assumptions. As an illustration, the SCBs for the mean and variance functions are applied to two strata pressure data sets collected from the Bulianta Coal Mine located in Ordos City, Inner Mongolia, China. Figures 4 and 5 depict the SCBs for one set of the data, and the SCBs for the second data set are given in Figures S.5 and S.6 in the online supplement. For both data sets, the null hypothesis of the mean function being $m(x) = a_0 + \sum_{k=1}^5 \{a_k \sin(k\omega x) + b_k \cos(k\omega x)\}$ for some constants a_0, a_k and b_k ($k = 1, \ldots, 5$) cannot be rejected with the p-values as high as 0.847 and 0.545 respectively. Meanwhile, the SCB for the variance function is used to test the homoscedastic null hypothesis for the two data sets. The conclusions are (i) strong rejection for one with the p-value = 0.0024 and (ii) no rejection for the other with the p-value = 0.545; see Subsection 5.2 for details.

Based on design model (1.1), Donoho and Johnstone (1996) and Angelini, De Canditiis and Frédérique (2003) studied nonparametric estimation for the mean function. SCBs for the mean function in model (1.1) were studied in Hall and Titterington (1988) and Cai, Low and Ma (2014). These SCBs are adaptive for $m(x)$ belonging to some function class but, as a result, asymptotically conservative instead of asymptotically correct. A more serious limitation of these adaptive

SCBs is their reliance on assumptions that the $\{\varepsilon_i\}_{i=1}^n$ are i.i.d. $N(0,1)$ and the variance function $\sigma^2(x)$ is constant. Alternatively, Eubank and Speckman (1993) obtained SCBs for the mean function $m(x)$ based on kernel smoothing, but under the restrictive assumption of homoscedasticity ($\sigma^2(x) \equiv \sigma^2$) and the mean function $m(x)$ being periodic. None of these works can handle clearly heteroscedastic data, such as the strata pressure data. Wang (2012) constructed a spline SCB for the mean function $m(x)$ based on deterministic designs and $\{\varepsilon_i\}_{i=1}^n$ being α-mixing, but its asymptotically conservative coverage limits its usefulness for testing hypotheses. For variance function estimation, Brown and Levine (2007) and Levine (2006) proposed difference-based kernel estimators and an approach of bandwidth selection, respectively, and Cai, Levine and Wang (2009) extended them to the multivariate situation, establishing the minimax convergence rate in the i.i.d. Gaussian case. Meanwhile, Wang et al. (2008) studied the effect of the unknown mean on the variance function estimation function in nonparametric regression. However, there are no SCBs for the variance function in these works.

The rest of the paper is organized as follows. Section 2 establishes the main asymptotic theoretical results. Section 3 provides insights of proofs and Section 4 gives concrete steps to implement the SCBs. Section 5 reports some simulation results and data analyses. The proofs are given in the online supplement.

2. Main Results

2.1. SCB for the mean function

We first formulate an SCB for the mean function $m(x)$ in model (1.1) by smoothing the data set $\{(i/n, Y_i)\}_{i=1}^n$ to approximate $m(x)$. The basic idea is to find a locally weighted least squares estimate $\hat{m}(x)$ which solves the minimization problem

$$\min_\theta n^{-1} \sum_{i=1}^n (Y_i - \theta)^2 K_h\left(\frac{i}{n} - x\right) = n^{-1} \sum_{i=1}^n \{Y_i - \hat{m}(x)\}^2 K_h\left(\frac{i}{n} - x\right),$$

in which $K(u)$ is a kernel function, $h = h_n > 0$ is a sequence of smoothing parameters called bandwidth, and $K_h(u) = h^{-1}K(u/h)$ is the kernel function rescaled by h. Clearly,

$$\hat{m}(x) = \frac{n^{-1}\sum_{i=1}^n K_h(i/n - x) Y_i}{\hat{f}(x)}, \tag{2.1}$$

where $\hat{f}(x) = n^{-1}\sum_{i=1}^n K_h(i/n - x)$.

We denote by $\psi^{(s)}(x)$ the s-th order derivative of a function $\psi(x)$. For

$\theta \in (0,1]$ and integer $p \geq 0$, let $C^{p,\theta}[0,1]$ be the space of functions with $\theta-$Hölder continuous p-th-order derivatives on $[0,1]$,

$$C^{p,\theta}[0,1] = \left\{ \phi(x) : \|\phi\|_{p,\theta} = \sup_{x \neq x', x, x' \in [0,1]} \frac{\left|\phi^{(p)}(x) - \phi^{(p)}(x')\right|}{|x - x'|^{\theta}} < +\infty \right\},$$

and denote by $C^{(p)}[0,1]$ the space of p-times continuously differentiable functions. For sequences of positive real numbers c_n and d_n, $c_n \ll d_n$ means $c_n/d_n \to 0$ as $n \to \infty$.

We need the following assumptions to construct SCBs for $m(x)$.

(M1) *The function* $m(\cdot) \in C^{p-1,\theta}[0,1]$ *for integer* $p > 1$ *and* $\theta \in (0,1]$.

(M2) *The error* ε *satisfies* $\mathrm{E}(\varepsilon) = 0$, $\mathrm{E}(\varepsilon^2) = 1$ *and* $\sigma^2(x) \in C^{(1)}[0,1]$ *with* $0 < c_\sigma \leq \sigma^2(x) \leq C_\sigma < +\infty$ *for any* $x \in [0,1]$.

(M3) *There exist* $\beta \in (0, 1/2 - 1/(4\theta + 4p - 2)), C_0 \in (0, +\infty), \gamma \in (1, +\infty)$, *and i.i.d.* $N(0,1)$ *variables* $\{Z_{in}\}_{i=1}^n$ *such that*

$$P\left\{ \max_{1 \leq l \leq n} \left| \sum_{i=1}^{l} \varepsilon_i - \sum_{i=1}^{l} Z_{in} \right| > n^{\beta} \right\} < C_0 n^{-\gamma}.$$

(M4) *The kernel function* $K \in C^{(1)}(\mathbb{R})$, *is of order* p, *and is supported on* $[-1,1]$.

(M5) *The bandwidth* $h = h_n$ *satisfies* $\log h_n / (-\log n) \to t > 0$ *as* $n \to \infty$ *and*

$$\max\left(n^{-1/2}\log^{1/2} n, n^{2\beta-1}\log n\right) \ll h_n \ll (n \log n)^{-1/(2\theta+2p-1)}.$$

Hence $1/(2\theta + 2p - 1) \leq t \leq \min(1/2, 1 - 2\beta)$.

Assumptions (M1), (M2) and (M4) are typical for kernel smoothing, adapted from Härdle (1989) and Eubank and Speckman (1993), while (M5) is the general condition on the choice of bandwidth h. It is more convenient to make the inequalities on t strict in (M5). The same holds true for $\tilde{t}$ and τ in Assumptions (E5) and (E7) below. Assumption (M3) provides the Gaussian approximation of the error process. According to Lemma S.2 in the supplement, Assumption (M3) is ensured by an elementary Assumption (M3'):

(M3') *There exists* $\eta > 2/\beta - 2$, $\beta \in (0, 1/2 - 1/(4\theta + 4p - 2))$ *such that* $\mathrm{E}\left|\varepsilon_1\right|^{2+\eta} < +\infty$.

Let $\mathcal{I}_n = [h_n, 1 - h_n]$. Our SCB for $m(x)$ is a direct corollary of Propositions 1–4 in Section 3.

Theorem 1. *If* (M1)–(M5) *hold, as* $n \to \infty$,

$$P\left(a_h\left[\sup_{x\in\mathcal{I}_n}\left|\frac{\{m(x)-\hat{m}(x)\}}{v(x)}\right|-b_h\right]\leq z\right)\to\exp\left(-2\exp\left(-z\right)\right), z\in\mathbb{R},$$

where $a_h=\left(2\log h^{-1}\right)^{1/2}$, $b_h=a_h+a_h^{-1}\left\{2^{-1}\log\left(C_K/\left(4\pi^2\right)\right)\right\}$,

$$C_K=\frac{\int_{-1}^{1}K^{(1)}\left(v\right)^2dv}{\int_{-1}^{1}K\left(v\right)^2dv, v(x)}=(nh)^{-1/2}\sigma(x)\left\{\int_{-1}^{1}K^2\left(u\right)du\right\}^{1/2}.$$

Then, for any $\alpha\in(0,1)$,

$$P\left(m(x)\in\hat{m}(x)\pm v(x)\left[a_h+a_h^{-1}\left\{q_\alpha+\frac{1}{2}\log\left(\frac{C_K}{(4\pi^2)}\right)\right\}\right], \forall x\in\mathcal{I}_n\right)\to 1-\alpha,$$

where $q_\alpha=-\log\left\{-1/2\log\left(1-\alpha\right)\right\}$.

Theorem 1 implies that the SCB contracts to zero at the rate $n^{-1/2}h^{-1/2}\log^{1/2}n$. In the special case $p=2,\theta=1$, as in Subsection 4.1, the implemented order of h satisfying (M5) is $n^{-1/5}\log^{-1/5-\delta_1}n$ for any $\delta_1>0$. Thus, the optimal bandwidth order of $n^{-1/5}$ is under-smoothed by $\log^{-1/5-\delta_1}n$, and the contraction rate of SCB is $n^{-2/5}\log^{3/5+\delta_1/2}n$.

2.2. SCBs for the variance function

The variance function $\sigma^2(x)$ measures the heteroscedastic variation of the errors $e_i=Y_i-m\left(i/n\right), 1\leq i\leq n$ in model (1.1). Following Cai and Yang (2015), if $m(x)$ were known by 'oracle', one could compute the squared errors $\left\{e_i^2\right\}_{i=1}^n$, and then by smoothing the data $\left\{\left(i/n, e_i^2\right)\right\}_{i=1}^n$ obtain a would-be kernel estimator of $\sigma^2(x)$:

$$\tilde{\sigma}_{\mathrm{K}}^2(x)=\frac{n^{-1}\sum_{i=1}^n\tilde{K}_{\tilde{h}}\left(i/n-x\right)e_i^2}{n^{-1}\sum_{i=1}^n\tilde{K}_{\tilde{h}}\left(i/n-x\right)},$$

where $\tilde{K}\left(u\right)$ is a kernel function and $\tilde{h}=\tilde{h}_n>0$ a bandwidth. However, $\tilde{\sigma}_{\mathrm{K}}^2\left(x\right)$ is infeasible as the errors $\left\{e_i^2\right\}_{i=1}^n$ are unobservable. To mimic $\tilde{\sigma}_{\mathrm{K}}^2(x)$, a spline-kernel estimator $\hat{\sigma}_{\mathrm{SK}}^2(x)$ is proposed

$$\hat{\sigma}_{\mathrm{SK}}^2(x)=\frac{n^{-1}\sum_{i=1}^n\tilde{K}_{\tilde{h}}\left(i/n-x\right)\hat{e}_i^2}{n^{-1}\sum_{i=1}^n\tilde{K}_{\tilde{h}}\left(i/n-x\right)},\tag{2.2}$$

where $\hat{e}_i=Y_i-\hat{m}_p\left(i/n\right)$ and $\hat{m}_p(x)$ is the p-th order spline estimator for $m\left(x\right)$ with integer $p>0$,

$$\hat{m}_p(x)=\arg\min_{g\in\mathcal{H}_N^{(p-2)}}\sum\nolimits_{i=1}^n\left\{Y_i-g\left(\frac{i}{n}\right)\right\}^2,\tag{2.3}$$

in which $\mathcal{H}_N^{(p-2)}=\mathcal{H}_N^{(p-2)}\left[0,1\right]$ is the space of spline functions on interval $[0,1]$

defined below.

Divide the interval $[0,1]$ into $(N+1)$ subintervals $J_j = \left[\chi_j, \chi_{j+1}\right)$, $j = 0,1,2,\ldots,N$ by equally spaced points $\left\{\chi_j\right\}_{j=1}^{N}$ called interior knots,

$$0=\chi_0<\chi_1<\cdots<\chi_{N+1}=1, \chi_j=\frac{j}{(N+1)}, j=0,1,\ldots,N+1.$$

$\mathcal{H}_N^{(p-2)}$ is the space of functions that are polynomials of degree $(p-1)$ on each J_j with continuous $(p-2)$-th derivative on $[0,1]$. For instance, $\mathcal{H}_N^{(-1)}$ consists of functions that are constant on each J_j, and $\mathcal{H}_N^{(0)}$ the space of functions that are linear on each J_j and continuous on $[0,1]$.

Let $\mathrm{E}\left(\varepsilon_i^4\right) = \mu_4$ and $\eta_i = \left(\varepsilon_i^2 - 1\right)(\mu_4 - 1)^{-1/2}$. We need the following assumptions to construct SCBs for $\sigma^2(x)$.

(E1) *The function* $m(\cdot) \in C^p[0,1]$ *for integer* $p>1$.

(E2) *The error* ε *satisfies* $\mathrm{E}(\varepsilon)=0$, $\mathrm{E}(\varepsilon^2)=1$ *and* $\sigma^2(x) \in C^{p_0-1,\theta_0}[0,1]$ *for integer* $p_0>1$, $\theta_0 \in (0,1]$ *with* $0<c_\sigma \le \sigma^2(x) \le C_\sigma < +\infty$ *for any* $x \in [0,1]$.

(E3) *There exist* $\beta' \in (0, 1/2 - 1/(4\theta_0 + 4p_0 - 2)), C_0' \in (0,+\infty), \gamma' \in (1,+\infty)$, *and i.i.d.* $N(0,1)$ *variables* $\{Z'_{in}\}_{i=1}^{n}$ *such that*

$$P\left(\max_{1\le l\le n}\left|\sum\nolimits_{i=1}^{l}\eta_i-\sum\nolimits_{i=1}^{l}Z'_{in}\right|>n^{\beta'}\right)<C_0'n^{-\gamma'}.$$

(E4) *The kernel function* $\tilde{K} \in C^{(1)}(\mathbb{R})$, *is of order* p_0, *and is supported on* $[-1,1]$.

(E5) *The bandwidth* $\tilde{h}=\tilde{h}_n$ *satisfies* $\log\tilde{h}_n/(-\log n) \to \tilde{t} > 0$ *as* $n\to\infty$ *and*

$$\max(n^{-1/2}\log^{1/2}n, n^{2\beta'-1}\log n, n^{-2(p-1)/(2p+1)}) \ll \tilde{h} \ll (n\log n)^{-1/(2\theta_0+2p_0-1)}.$$

Consequently, $1/(2\theta_0+2p_0-1) \le \tilde{t} \le \min\left(1/2, 1-2\beta', 2(p-1)/(2p+1)\right)$.

(E6) *There exist* $C_0 \in (0,+\infty), \gamma \in (1,+\infty), \beta \in (0,b]$ *and i.i.d.* $N(0,1)$ *variables* $\{Z_{in}\}_{i=1}^{n}$ *such that*

$$P\left(\max_{1\le l\le n}\left|\sum\nolimits_{i=1}^{l}\varepsilon_i-\sum\nolimits_{i=1}^{l}Z_{in}\right|>n^{\beta}\right)<C_0n^{-\gamma},$$

where $b=\min\big(1-3/2\,(2p+1)-\tilde{t}, 1-5/2\,(2p+1)-5\tilde{t}/2\,(2p+3), 1/2-1/(4\theta_0+4p_0-2)\big)$.

(E7) *The number of interior knots* N *satisfies* $\log N/\log n \to \tau$ for some $\tau > 0$ *and*

$$\max\left(n^{1/4p}, \tilde{h}^{-1/(p-1)} n^{(\beta-1/2)/(p-1)}, \tilde{h}^{-1/2(p-1)} \log^{1/2(p-1)} n\right) \ll N$$
$$\ll \min\left(\tilde{h}^{2/3} n^{2(1-\beta)/3}, n^{2(1-\beta)/5}, n^{1/3}\tilde{h}^{1/3} \log^{-1/3} n\right).$$

Consequently, $\max(1/4p, \left(2\tilde{t}+2\beta-1\right)/2\left(p-1\right), \tilde{t}/2\left(p-1\right)) \leq \tau \leq \min(2\left(1-\beta\right)/3 - 2\tilde{t}/3, 2\left(1-\beta\right)/5, 1/3-\tilde{t}/3)$.

Assumptions (E2)–(E5) are adapted from Assumptions (M1)–(M5) of Subsection 2.1. Here (E1) is a general condition for spline regression of the mean function in model (1.1), while (E7), on the choice of knots number N, ensures the oracle efficiency in Theorem 2 and the extreme distribution result in (2.4) below. Lemma S.2 in the supplement implies that (E3) and (E6) are ensured by Assumption (E3').

(E3') *There exists* $\eta' > 2/\beta - 2$, $\beta \in (0, b\,]$ *as in (E6) such that* $\mathrm{E}\left|\varepsilon_1\right|^{4+2\eta'} < +\infty$.

Under (E2)–(E5), applying Theorem 1 to unobservable sample $\left\{(i/n, e_i^2)\right\}_{i=1}^n$ and letting $\tilde{\mathcal{I}}_n = \left[\tilde{h}_n, 1-\tilde{h}_n\right]$, for each $z \in \mathbb{R}$ one has

$$P\left(a_{\tilde{h}}\left(\sup_{x\in\tilde{\mathcal{I}}_n}\left|\frac{\left\{\sigma^2(x) - \tilde{\sigma}^2_{\mathrm{K}}(x)\right\}}{v_0(x)}\right| - b_{\tilde{h}}\right) \leq z\right) \to \exp\left(-2\exp\left(-z\right)\right), \tag{2.4}$$

where

$$a_{\tilde{h}} = \left(2\log\tilde{h}^{-1}\right)^{1/2}, b_{\tilde{h}} = a_{\tilde{h}} + a_{\tilde{h}}^{-1}\left\{2^{-1}\log\left(\frac{C_{\tilde{K}}}{(4\pi^2)}\right)\right\}, \tag{2.5}$$

$$C_{\tilde{K}} = \frac{\int_{-1}^{1}\tilde{K}^{(1)}\left(v\right)^2 dv}{\int_{-1}^{1}\tilde{K}\left(v\right)^2 dv}, \tag{2.6}$$

$$v_0(x) = \left\{n^{-1}\tilde{h}^{-1}\sigma_0^2(x)\int_{-1}^{1}\tilde{K}^2\left(u\right)du\right\}^{1/2}. \tag{2.7}$$

Here $\sigma_0^2(x) = \sigma^4(x)\left(\mu_4 - 1\right)$ so that $\mathrm{var}(e_i^2) = \mathrm{E}\left(e_i^4\right) - \left\{\mathrm{E}\left(e_i^2\right)\right\}^2 = \sigma_0^2\left(i/n\right)$, as the second and fourth moments of e_i are $\mathrm{E}\left(e_i^2\right) = \sigma^2\left(i/n\right), \mathrm{E}\left(e_i^4\right) = \sigma^4\left(i/n\right)\mu_4$.

According to (2.4), it is obvious that an asymptotic $100(1-\alpha)\%$ 'infeasible' SCB for $\sigma^2(x)$ over $\tilde{\mathcal{I}}_n = \left[\tilde{h}_n, 1-\tilde{h}_n\right]$ is

$$\tilde{\sigma}^2_{\mathrm{K}}(x) \pm v_0(x)\left[a_{\tilde{h}} + a_{\tilde{h}}^{-1}\left\{q_\alpha + \frac{1}{2}\log\left(\frac{C_{\tilde{K}}}{(4\pi^2)}\right)\right\}\right]. \tag{2.8}$$

Theorem 2. *If* (E1)–(E7) *hold, as* $n \to \infty$ *the spline-kernel estimator* $\hat{\sigma}^2_{SK}(x)$ *is asymptotically as efficient as the 'infeasible' estimator* $\tilde{\sigma}^2_K(x)$ *with*

$$\sup_{x\in[0,1]}\left|\hat{\sigma}^2_{SK}(x) - \tilde{\sigma}^2_K(x)\right| = o_p\left(n^{-1/2}\right).$$

The proof of Theorem 2 depends on Propositions 5–7 given in Subsection 3.2. Theorem 1, Theorem 2 and Slutsky's Theorem together imply the following.

Theorem 3. *If* (E1)–(E7) *hold, as* $n \to \infty$ *an asymptotic* $100\ (1-\alpha)\%$ *SCB for* $\sigma^2(x)$ *over* $\tilde{\mathcal{I}}_n = \left[\tilde{h}_n, 1-\tilde{h}_n\right]$ *is*

$$\hat{\sigma}^2_{SK}(x) \pm v_0(x)\left[a_{\tilde{h}} + a_{\tilde{h}}^{-1}\left\{q_\alpha + \frac{1}{2}\log\left(\frac{C_{\tilde{K}}}{(4\pi^2)}\right)\right\}\right],$$

with $a_{\tilde{h}}$, $C_{\tilde{K}}$, $v_0(x)$ *and* q_α *given in* (2.5), (2.6), (2.7) *and Theorem* 1, *respectively.*

Theorem 3 implies that the SCB contracts to zero at the rate $n^{-1/2}\tilde{h}^{-1/2}$ $\log^{1/2} n$. In the special case $p = 4, p_0 = 2, \theta_0 = 1$ as in Subsection 4.2, the implemented order of $\tilde{h}$ satisfying (E5) is $n^{-1/5}\log^{-1/5-\delta_2}$ for any $\delta_2 > 0$. Thus the optimal bandwidth order of $n^{-1/5}$ is under-smoothed by $\log^{-1/5-\delta_2} n$, and the contraction rate of the SCB is $n^{-2/5}\log^{3/5+\delta_2/2} n$.

3. Error Decomposition

3.1. Case of the mean function

An asymptotic SCB for $m(x)$ is constructed by examining $\sup_{x\in\mathcal{I}_n}|\hat{m}(x)-m(x)|$. We find that

$$\begin{aligned}
\hat{m}(x) - m(x) &= n^{-1}\hat{f}(x)^{-1}\sum_{i=1}^{n} K_h\left(\frac{i}{n}-x\right)Y_i - m(x) \\
&= n^{-1}\hat{f}(x)^{-1}\sum_{i=1}^{n} K_h\left(\frac{i}{n}-x\right)\left\{m\left(\frac{i}{n}\right) - m(x) + \sigma\left(\frac{i}{n}\right)\varepsilon_i\right\} \\
&= \hat{f}(x)^{-1}\left\{A_n(x) + B_n(x)\right\},
\end{aligned}$$

in which

$$A_n(x) = n^{-1}\sum_{i=1}^{n} K_h\left(\frac{i}{n}-x\right)\left\{m\left(\frac{i}{n}\right) - m(x)\right\}, \tag{3.1}$$

$$B_n(x) = n^{-1}\sum_{i=1}^{n} K_h\left(\frac{i}{n}-x\right)\sigma\left(\frac{i}{n}\right)\varepsilon_i. \tag{3.2}$$

The following stochastic processes approximate $B_n(x)$:

$$B_{n1}(x) = n^{-1}\sum_{i=1}^{n} K_h\left(\frac{i}{n}-x\right)\sigma\left(\frac{i}{n}\right)Z_{in}, \tag{3.3}$$

$$B_{n2}(x) = n^{-1}\sum_{i=1}^{n} K_h\left(\frac{i}{n}-x\right)\sigma(x)Z_{in}, \tag{3.4}$$

$$B_{n3}(x) = n^{-1/2} \int K_h(u-x)\,\sigma(x)\,dW_n(u), \tag{3.5}$$

where $\{Z_{in}\}_{i=1}^{n}$ are i.i.d. $N(0,1)$ variables satisfying (M3) and $W_n(u)$ is a two-sided Brownian motion on $(-\infty,+\infty)$ satisfying $Z_{in} = \sqrt{n}\{W_n(i/n) - W_n((i-1)/n)\}$.

Proposition 1. *Under Assumption* (M4)*, as* $n \to \infty$*,*

$$\sup_{x\in\mathcal{I}_n} \left|\hat{f}(x) - 1\right| = \mathcal{O}\left(n^{-1}h^{-2}\right).$$

Proposition 2. *Under Assumptions* (M1)*,* (M4) *and* (M5)*, as* $n \to \infty$*,*

$$\sup_{x\in\mathcal{I}_n} |A_n(x)| = \mathcal{O}\left(h^{\theta+p-1} + n^{-1}h^{-1}\right).$$

Proposition 3. *Under Assumptions* (M2)–(M4)*, as* $n \to \infty$*,*

$$(a) \quad \sup_{x\in[0,1]} |B_n(x) - B_{n1}(x)| = \mathcal{O}_p\left(n^{\beta-1}h^{-1}\right),$$

$$(b) \quad \sup_{x\in[0,1]} |B_{n1}(x) - B_{n2}(x)| = \mathcal{O}_p\left(n^{-1/2}h^{1/2}\log^{1/2} n\right),$$

$$(c) \quad \sup_{x\in\mathcal{I}_n} |B_{n2}(x) - B_{n3}(x)| = \mathcal{O}_p\left(n^{-3/2}h^{-2}\log^{1/2} n\right),$$

$$(d) \quad \sup_{x\in[0,1]} |B_{n3}(x)| = \mathcal{O}_p\left(n^{-1/2}h^{-1/2}\log^{1/2} n\right).$$

The proofs of these propositions are given in the online supplement.

As $\mathrm{E}\{B_{n3}^2(x)\} = n^{-1}h^{-1}\sigma^2(x)\int_{-1}^{1} K^2(u)\,du$, standardizing the process $B_{n3}(x)$ for $x \in [0,1]$, one obtains the standard Gaussian process

$$\frac{\int K_h(x-u)\,dW_n(u)}{\left\{h^{-1}\int_{-1}^{1} K^2(u)\,du\right\}^{1/2}}, x \in [0,1],$$

whose absolute maximum follows the same probability law as

$$\mathcal{L}\left(\frac{h^{-1}\int K(s-u/h)\,dW_n(u)}{\left\{h^{-1}\int_{-1}^{1} K^2(u)\,du\right\}^{1/2}}, s \in \left[0,h^{-1}\right]\right)$$

$$= \mathcal{L}\left(\frac{\int K(s-r)\,dW_n(r)}{\left\{\int_{-1}^{1} K^2(t)\,dt\right\}^{1/2}}, s \in \left[0,h^{-1}\right]\right).$$

Let

$$\zeta(s) = \frac{\int K(s-r)\,dW_n(r)}{\left\{\int_{-1}^{1} K^2(t)\,dt\right\}^{1/2}}, s \in \left[0,h^{-1}\right].$$

By Equation (2.5) in Bickel and Rosenblatt (1973), we have the following.

Proposition 4. *Under Assumptions* (M2) *and* (M4), *as* $n \to \infty$,

$$P\left(a_h\left\{\sup_{s\in[0,h^{-1}]}\left|\zeta\left(s\right)\right| - b_h\right\} < z\right) \to \exp\left(-2\exp\left(-z\right)\right), z \in \mathbb{R},$$

in which a_h *and* b_h *are given in Theorem* 1.

3.2. Case of the variance function

To prove Theorem 2, the estimation error $\hat{\sigma}^2_{\mathrm{SK}}(x) - \tilde{\sigma}^2_{\mathrm{K}}(x)$ is broken into three parts. We begin by describing the spline space $\mathcal{H}_N^{(p-2)}$ and the representation of the spline estimators $\hat{m}_p(x)$ in Equation (2.1).

The space $\mathcal{H}_N^{(p-2)}$ is spanned linearly by B-spline basis $\{b_{j,p}\}_{j=1-p}^N$ introduced in de Boor (2001). Denote by $\|\phi\|_2$ the theoretical L^2 norm of a function ϕ on $[0,1]$, $\|\phi\|_2^2 = \int_0^1 \phi^2(x)dx$, and the empirical L^2 norm as $\|\phi\|_{2,n}^2 = n^{-1}\sum_{i=1}^n \phi^2\left(i/n\right)$. The rescaled B-spline basis for $\mathcal{H}_N^{(p-2)}$ is $\{B_{j,p}\}_{j=1-p}^N$, where $B_{j,p}(x) = b_{j,p}(x)\left\|b_{j,p}\right\|_2^{-1}$ with theoretical norm equal to 1, $1-p \le j \le N$.

The estimator $\hat{m}_p(x)$ in (2.3) can then be expressed as

$$\hat{m}_p(x) = \sum\nolimits_{j=1-p}^{N} \hat{\lambda}_{j,p} B_{j,p}(x),$$

where the vector $\left(\hat{\lambda}_{1-p,p}, \ldots, \hat{\lambda}_{N,p}\right)^T$ is the solution of the least-squares problem

$$\left(\hat{\lambda}_{1-p,p}, \ldots, \hat{\lambda}_{N,p}\right)^T = \operatorname*{argmin}_{\mathbb{R}^{N+p}} \sum_{i=1}^{n}\left\{Y_i - \sum\nolimits_{j=1-p}^{N} \lambda_{j,p} B_{j,p}\left(\frac{i}{n}\right)\right\}^2. \tag{3.6}$$

Write $\mathbf{Y}$ as the sum of signal vector $\mathbf{m}$ and error term $\mathbf{E}$, $\mathbf{Y} = \mathbf{m} + \mathbf{E}$, where $\mathbf{Y} = \{Y_1, \ldots, Y_n\}^T$, $\mathbf{m} = \{m(1/n), \ldots, m(n/n)\}^T$ and $\mathbf{E} = \{\sigma(1/n)\varepsilon_1, \ldots, \sigma(n/n)\varepsilon_n\}^T$. Projecting this relationship into the space $\mathcal{H}_N^{(p-2)}$, one obtains that

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

where

$$\tilde{m}_p(x) = \sum\nolimits_{j=1-p}^{N} \tilde{\lambda}_{j,p} B_{j,p}(x), \tilde{\varepsilon}_p(x) = \sum\nolimits_{j=1-p}^{N} \tilde{a}_{j,p} B_{j,p}(x), \tag{3.7}$$

and the vectors $\{\tilde{\lambda}_{1-p,p}, \ldots, \tilde{\lambda}_{N,p}\}^T$ and $\{\tilde{a}_{1-p,p}, \ldots, \tilde{a}_{N,p}\}^T$ in (3.7) are solutions of (3.6) with Y_i replaced by $m\left(i/n\right)$ and $\sigma\left(i/n\right)\varepsilon_i$, respectively. One then obtains that

$$\tilde{m}_p(x) = \{B_{1-p,p}(x), \ldots, B_{N,p}(x)\}\left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{m}, \tag{3.8}$$

$$\tilde{\varepsilon}_p(x) = \{B_{1-p,p}(x), \ldots, B_{N,p}(x)\}\left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{E}, \tag{3.9}$$

where the design matrix $\mathbf{B}$ is

$$\mathbf{B} = \begin{pmatrix} B_{1-p,p}\left(\frac{1}{n}\right) & \cdots & B_{N,p}\left(\frac{1}{n}\right) \\ \vdots & \ddots & \vdots \\ B_{1-p,p}\left(1\right) & \cdots & B_{N,p}\left(1\right) \end{pmatrix}_{n\times(N+p)}.$$

It is obvious that $\hat{\sigma}^2_{\mathrm{SK}}(x) - \tilde{\sigma}^2_{\mathrm{K}}(x)$ can be decomposed as

$$\begin{aligned}\hat{\sigma}^2_{\mathrm{SK}}(x) - \tilde{\sigma}^2_{\mathrm{K}}(x) &= \frac{\sum_{i=1}^n \tilde{K}_{\tilde{h}}\left(i/n - x\right)\left(I_{i,p} + II_{i,p} + III_{i,p}\right)}{\sum_{i=1}^n \tilde{K}_{\tilde{h}}\left(i/n - x\right)} \\ &= \frac{I_1 + I_2 + I_3}{n^{-1}\sum_{i=1}^n \tilde{K}_{\tilde{h}}\left(i/n - x\right)}\end{aligned}$$

in which

$$I_1 = I_1(x) = n^{-1}\sum\nolimits_{i=1}^n \tilde{K}_{\tilde{h}}\left(\frac{i}{n} - x\right) I_{i,p}, \tag{3.10}$$

$$I_2 = I_2(x) = n^{-1}\sum\nolimits_{i=1}^n \tilde{K}_{\tilde{h}}\left(\frac{i}{n} - x\right) II_{i,p}, \tag{3.11}$$

$$I_3 = I_3(x) = n^{-1}\sum\nolimits_{i=1}^n \tilde{K}_{\tilde{h}}\left(\frac{i}{n} - x\right) III_{i,p}, \tag{3.12}$$

$$I_{i,p} = \left\{m\left(\frac{i}{n}\right) - \tilde{m}_p\left(\frac{i}{n}\right)\right\}^2 + \tilde{\varepsilon}^2_p\left(\frac{i}{n}\right) + 2\left\{\tilde{m}_p\left(\frac{i}{n}\right) - m\left(\frac{i}{n}\right)\right\}\tilde{\varepsilon}_p\left(\frac{i}{n}\right),$$

$$II_{i,p} = -2\sigma\left(\frac{i}{n}\right)\varepsilon_i\tilde{\varepsilon}_p\left(\frac{i}{n}\right), III_{i,p} = \left\{m\left(\frac{i}{n}\right) - \tilde{m}_p\left(\frac{i}{n}\right)\right\}\sigma\left(\frac{i}{n}\right)\varepsilon_i.$$

Theorem 2 follows from the next three propositions whose proofs are in the supplement.

Proposition 5. *Under Assumptions* (E1)–(E7)*, as* $n \to \infty$*,*

$$\sup_{x\in[0,1]} |I_1(x)| = \mathcal{O}_p\left(N^{-2p} + n^{-1}N\right).$$

Proposition 6. *Under Assumptions* (E2)–(E7)*, as* $n \to \infty$*,*

$$\sup_{x\in[0,1]} |I_2(x)| = \mathcal{O}_p\left(n^{-1}\tilde{h}^{-1/2}N^{3/2}\log^{1/2} n + \tilde{h}^{-1}n^{\beta-3/2}N^{3/2} + n^{\beta-3/2}N^{5/2}\right).$$

Proposition 7. *Under Assumptions* (E1)–(E7)*, as* $n \to \infty$*,*

$$\sup_{x\in[0,1]} |I_3(x)| = \mathcal{O}_p\left(n^{-1/2}\tilde{h}^{-1/2}N^{1-p}\log^{1/2} n + \tilde{h}^{-1}n^{\beta-1}N^{1-p} + n^{\beta-1}N^{2-p}\right).$$

4. Implementation

In this section we describe detailed procedures for implementing the SCBs in Theorems 1 and 3 based on a data set $\{(i/n, Y_i)\}_{i=1}^{n}$ that follows model (1.1). This is used throughout Section 5 for simulations and data examples. The implementation codes are written in R 3.03 and posted on the website: `https://github.com`.

4.1. Implementing mean function SCB

As the default, we set $p = 2, \theta = 1$ in (M1). When constructing the SCB for the mean function $m(x)$ in model (1.1) according to Theorem 1, one chooses a kernel function K and bandwidth h for computing $\hat{m}(x)$ and estimating the variance function $\sigma^2(x)$, and then plugs in these estimates, as in Eubank and Speckman (1993), Hall and Titterington (1988), Härdle (1989) and Xia (1998).

We choose the quartic kernel $K(u) = 15\left(1-u^2\right)^2 I\{|u| \le 1\}/16$ to satisfy (M4), and the bandwidths $h = h_{\mathrm{rot}} \times \log^{-1/5-\delta_1} n$ $(\delta_1 > 0)$ to satisfy (M5), where the rule-of-thumb bandwidth h_{rot} is from Equation (4.3) of Fan and Gijbels (1996):

$$h_{\mathrm{rot}} = \left[\frac{35\sum_{i=1}^{n}\left\{Y_i - \sum_{k=0}^{4}\widehat{a}_k\,(i/n)^k\right\}^2}{n\sum_{i=1}^{n}\left\{2\widehat{a}_2 + 6\widehat{a}_3\,(i/n) + 12\widehat{a}_4\,(i/n)^2\right\}^2}\right]^{1/5}, \tag{4.1}$$

in which $(\widehat{a}_k)_{k=0}^{4} = \operatorname{argmin}_{(a_k)_{k=0}^{4}\in\mathbb{R}^5}\sum_{i=1}^{n}\left\{Y_i - \sum_{k=0}^{4} a_k\,(i/n)^k\right\}^2$. Here h_{rot} has order $n^{-1/5}$ and h order $n^{-1/5}\log^{-1/5-\delta_1} n$, satisfying (M5). We have found in extensive simulations that $h = h_{\mathrm{rot}}\log^{-1/2} n$ works quite well and is what we recommend.

The two step spline-kernel estimator $\hat{\sigma}^2_{\mathrm{SK}}(x)$ in (2.2) is used for the variance function $\sigma^2(x)$, with detailed procedures introduced in Section 4.2.

The asymptotic $100(1-\alpha)\%$ SCB for the mean function is

$$\hat{m}(x) \pm \hat{v}(x)\left[a_h + a_h^{-1}\left\{q_\alpha + \frac{1}{2}\log\left(\frac{C_K}{4\pi^2}\right)\right\}\right], x \in \mathcal{I}_n, \tag{4.2}$$

with $\hat{v}(x) = \left\{n^{-1}h^{-1}\hat{\sigma}^2_{\mathrm{SK}}(x)\int_{-1}^{1}K^2(v)\,dv\right\}^{1/2}$.

4.2. Implementing variance function SCB

To construct the SCB for $\sigma^2(x)$, we set the default values $p = 4, p_0 = 2, \theta_0 = 1$ in (E1) and (E2) and take the kernel $\tilde{K}(u) = 15\left(1-u^2\right)^2 I\{|u| \le 1\}/16$, satisfying (E4), with bandwidth $\tilde{h} = h_{\mathrm{rot},\sigma}\log^{-1/5-\delta_2} n$ satisfying (E5), in which $\delta_2 > 0, h_{\mathrm{rot},\sigma}$ is as in (4.1) but with Y_i replaced by $\hat{e}_i^2 = \{Y_i - \hat{m}_p(i/n)\}^2$. Extensive simulation experiments show that $\tilde{h} = h_{\mathrm{rot},\sigma} \log^{-1/2} n$ works quite well and is what we recommend.

According to Theorem 1 of Xue and Yang (2006), for any $m(x) \in C^p[0,1]$, $p \geq 2$, the optimal order of knots number N for $m(x)$ is $n^{1/(2p+1)}$, $n^{1/9}$ with $p = 4$. Denote the 'optimal' N by $\hat{N}^{\mathrm{opt}}$, the minimizer of the AIC defined below over integers in $[0.5N_r, \min\{5N_r, Tb\}]$, where $N_r = n^{1/9}$ and $Tb = n/4 - 1$ to ensure that $\hat{N}^{\mathrm{opt}}$ is of order $n^{1/9}$ and the total parameters in the least square estimation is less than $n/4$. This particular $\hat{N}^{\mathrm{opt}}$ satisfies (E7), but is of course not the only one. Let $\hat{Y}_i = \hat{m}_p(i/n)$ be the predictor of the i-th response Y_i and $q_n = (4 + N)$ represent the number of parameters in (3.6). The AIC value corresponding to N is

$$\mathrm{AIC}(N) = \log \mathrm{MSE} + \frac{2q_n}{n}, \quad \mathrm{MSE} = n^{-1} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2. \tag{4.3}$$

To estimate the variance function $\sigma_0^2(x)$ of $\left\{(i/n, e_i^2)\right\}_{i=1}^n$, one uses the spline-kernel method as described. Specifically, let $\hat{\sigma}_{\mathrm{S}}^2(x)$ be the spline estimator based on data set $\left\{(i/n, \hat{e}_i^2)\right\}_{i=1}^n$:

$$\hat{\sigma}_{\mathrm{S}}^2(x) = \{B_{1-p,p}(x), \ldots, B_{N,p}(x)\} \left(\mathbf{B}^T \mathbf{B}\right)^{-1} \mathbf{B}^T \left\{\hat{e}_1^2, \ldots, \hat{e}_n^2\right\}^T.$$

Let $\nabla_i = \left\{\hat{e}_i^2 - \hat{\sigma}_{\mathrm{S}}^2(i/n)\right\}^2$. Then we find

$$\hat{\sigma}_0^2(x) = \frac{n^{-1} \sum_{i=1}^n \mathcal{K}_{h_{\sigma_0}}(i/n - x) \nabla_i}{n^{-1} \sum_{i=1}^n \mathcal{K}_{h_{\sigma_0}}(i/n - x)},$$

where h_{σ_0} is an under smoothing bandwidth, $h_{\sigma_0} = h_{\mathrm{rot},\sigma_0} \log^{-1/2} n$, in which $h_{\mathrm{rot},\sigma_0}$ is the rule-of-thumb bandwidth as in Subsection 4.1 with Y_i replaced by ∇_i. According to Fan and Gijbels (1996), one has $\sup_{x \in [0,1]} \left|\hat{\sigma}_0^2(x) - \sigma_0^2(x)\right| = o_p(1)$. The asymptotic $100(1-\alpha)\%$ SCB for variance function is:

$$\hat{\sigma}_{\mathrm{SK}}^2(x) \pm \hat{v}_0(x) \left[a_{\tilde{h}} + a_{\tilde{h}}^{-1} \left\{q_\alpha + \frac{1}{2} \log\left(\frac{C_{\tilde{K}}}{(4\pi^2)}\right)\right\}\right], x \in \tilde{\mathcal{I}}_n, \tag{4.4}$$

with $\hat{v}_0(x) = \left\{n^{-1} h^{-1} \hat{\sigma}_0^2(x) \int_{-1}^1 \tilde{K}^2(v) dv\right\}^{1/2}$.

5. Empirical Studies

5.1. Monte Carlo examples

To investigate the finite-sample behavior of the proposed SCBs in Section 2, the four cases in Table 1 were examined, where

Table 1. Four cases of study.

Case 1	Case 2	Case 3	Case 4
$m_1(x), \sigma_1(x)$	$m_1(x), \sigma_2(x)$	$m_2(x), \sigma_1(x)$	$m_2(x), \sigma_2(x)$

$$m_1(x) = \cos(3\pi x), m_2(x) = \exp\left(-32x^2\right),$$
$$\sigma_1(x) = 0.1\sin(2\pi x) + 0.2, \sigma_2(x) = \frac{\exp(x) - 0.9}{\exp(x) + 0.9},$$

and ε was either $N(0,1)$ or the standardized t-distribution with freedom 10, $\varepsilon \sim 0.8^{1/2} * t_{10}$. The mean functions $m_1(x), m_2(x)$ resemble those in Eubank and Speckman (1993), but without periodicity. The sample sizes were $n = 300, 600, 900$, while for the SCBs the confidence level was $1 - \alpha = 0.95, 0.99$.

The coverage frequencies by SCBs defined in (4.2) for $m(x)$ are reported in Table 2; these are relative frequencies in 2,000 replications of coverage of the true curve at equally spaced points $\{x_j, j = 1, 2, \ldots, 400\}$ on $\mathcal{I}_n$. For comparison, the coverage frequencies from Eubank and Speckman (1993) are also listed in Table 2 and denoted as SCB-ES. In all cases with $\varepsilon \sim N(0,1)$ (the left side of the parentheses) and $\varepsilon \sim 0.8^{1/2} * t_{10}$ (inside the parentheses) the coverage frequencies improve and approach the nominal level as the sample size n increases, which supports Theorem 1. It is also evident that the SCBs in (4.2) perform far better than those in Eubank and Speckman (1993).

The coverage frequencies at equally spaced points $\{x_j, j = 1, 2, \ldots, 400\}$ on $\tilde{\mathcal{I}}_n$ by the SCB in (4.4) and the 'infeasible' SCB in (2.8) for $\sigma^2(x)$ with $\varepsilon \sim N(0,1)$ (the left side of the parentheses) and $\varepsilon \sim 0.8^{1/2} * t_{10}$ (inside the parentheses) are also shown in Table 2. The coverage frequencies improve and approach the nominal levels as the sample size n increases for all cases, which supports Theorem 3. Meanwhile, the coverage frequencies by the SCB and the 'infeasible' SCB are very close, as anticipated in Theorem 2.

Figures 1 depicts the boxplots of $\triangle_n = \sqrt{n} \max_j \left|\tilde{\sigma}^2_{\mathrm{K}}(x_j) - \hat{\sigma}^2_{\mathrm{SK}}(x_j)\right|$ based on $\varepsilon \sim N(0,1)$ over 2,000 replications, where $\{x_j, j = 1, 2, \ldots, 400\}$ are equally-spaced points on $\tilde{\mathcal{I}}_n$. The boxplot of $\triangle_n$ becomes narrower as the sample size n increases, so the difference between $\hat{\sigma}^2_{\mathrm{SK}}(x)$ and $\tilde{\sigma}^2_{\mathrm{K}}(x)$ is asymptotically of an order smaller than $n^{-1/2}$, which aggrees with Theorem 2. The scenario with $\varepsilon \sim 0.8^{1/2} * t_{10}$ is shown in Figure S.1 in the supplement.

To visualize the SCBs for the mean and variance functions, Figures 2 and 3 were created based on two samples of size 300 and 600 in Case 1 with $\varepsilon \sim N(0,1)$ and confidence level 95%. The scenario with confidence level 99% was shown in

Table 2. Empirical coverage frequencies of the SCB in (4.2) and in Eubank and Speckman (1993) for $m(x)$, and of oracle SCB in (4.4) and 'infeasible' SCB in (2.4) for $\sigma^2(x)$ using 2,000 replications with $\varepsilon \sim N(0,1)$ (the left side of the parentheses) and $\varepsilon \sim 0.8^{1/2} * t_{10}$ (inside the parentheses) respectively.

			$\varepsilon \sim N(0,1)$ ($\varepsilon \sim 0.8^{1/2} * t_{10}$)			
Case	n	$1-\alpha$	SCB	SCB-ES	oracle	infeasible
1	300	0.95	0.940(0.945)	0.832(0.828)	0.889(0.846)	0.907(0.862)
		0.99	0.995(0.995)	0.950(0.953)	0.959(0.939)	0.969(0.953)
	600	0.95	0.962(0.947)	0.810(0.828)	0.929(0.906)	0.942(0.904)
		0.99	0.995(0.993)	0.941(0.955)	0.982(0.973)	0.986(0.972)
	900	0.95	0.952(0.959)	0.818(0.815)	0.939(0.928)	0.945(0.931)
		0.99	0.996(0.996)	0.952(0.953)	0.990(0.983)	0.990(0.985)
2	300	0.95	0.951(0.959)	0.863(0.852)	0.889(0.852)	0.905(0.864)
		0.99	0.995(0.997)	0.961(0.954)	0.952(0.945)	0.964(0.947)
	600	0.95	0.961(0.960)	0.840(0.836)	0.943(0.907)	0.955(0.906)
		0.99	0.997(0.995)	0.953(0.951)	0.984(0.970)	0.988(0.966)
	900	0.95	0.962(0.959)	0.817(0.828)	0.958(0.925)	0.955(0.933)
		0.99	0.998(0.997)	0.949(0.952)	0.991(0.981)	0.992(0.983)
3	300	0.95	0.964(0.966)	0.858(0.866)	0.893(0.855)	0.907(0.862)
		0.99	0.996(0.998)	0.963(0.968)	0.960(0.945)	0.969(0.953)
	600	0.95	0.966(0.968)	0.846(0.861)	0.928(0.906)	0.942(0.904)
		0.99	0.997(0.997)	0.962(0.965)	0.983(0.975)	0.986(0.972)
	900	0.95	0.970(0.966)	0.850(0.851)	0.941(0.926)	0.945(0.931)
		0.99	0.997(0.999)	0.963(0.963)	0.989(0.981)	0.990(0.985)
4	300	0.95	0.960(0.960)	0.916(0.914)	0.895(0.869)	0.905(0.864)
		0.99	0.995(0.996)	0.983(0.983)	0.958(0.953)	0.964(0.947)
	600	0.95	0.956(0.950)	0.905(0.903)	0.943(0.914)	0.955(0.906)
		0.99	0.995(0.997)	0.981(0.979)	0.986(0.973)	0.988(0.966)
	900	0.95	0.957(0.956)	0.891(0.900)	0.933(0.917)	0.955(0.933)
		0.99	0.995(0.995)	0.970(0.981)	0.989(0.983)	0.992(0.983)

Figure S.2 and S.3 in the supplement. Each has the center solid line as the true curve, center dashed line the estimated curve and the upper and lower thick solid lines the SCB. As expected, the SCBs for $n = 600$ are thinner and fit better than those for $n = 300$. Figures S.4 and S.5 in the supplement show the SCBs for the mean and variance functions with $\varepsilon \sim 0.8^{1/2} * t_{10}$.

5.2. Data examples

Using our SCBs, we have analyzed two data sets provided by Professor Jiang Yaodong's research group at China University of Mining and Technology, which are available from us upon request. The data are strata pressure records in May 2013, from the Bulianta Coal Mine located in Ordos City, Inner Mongolia,

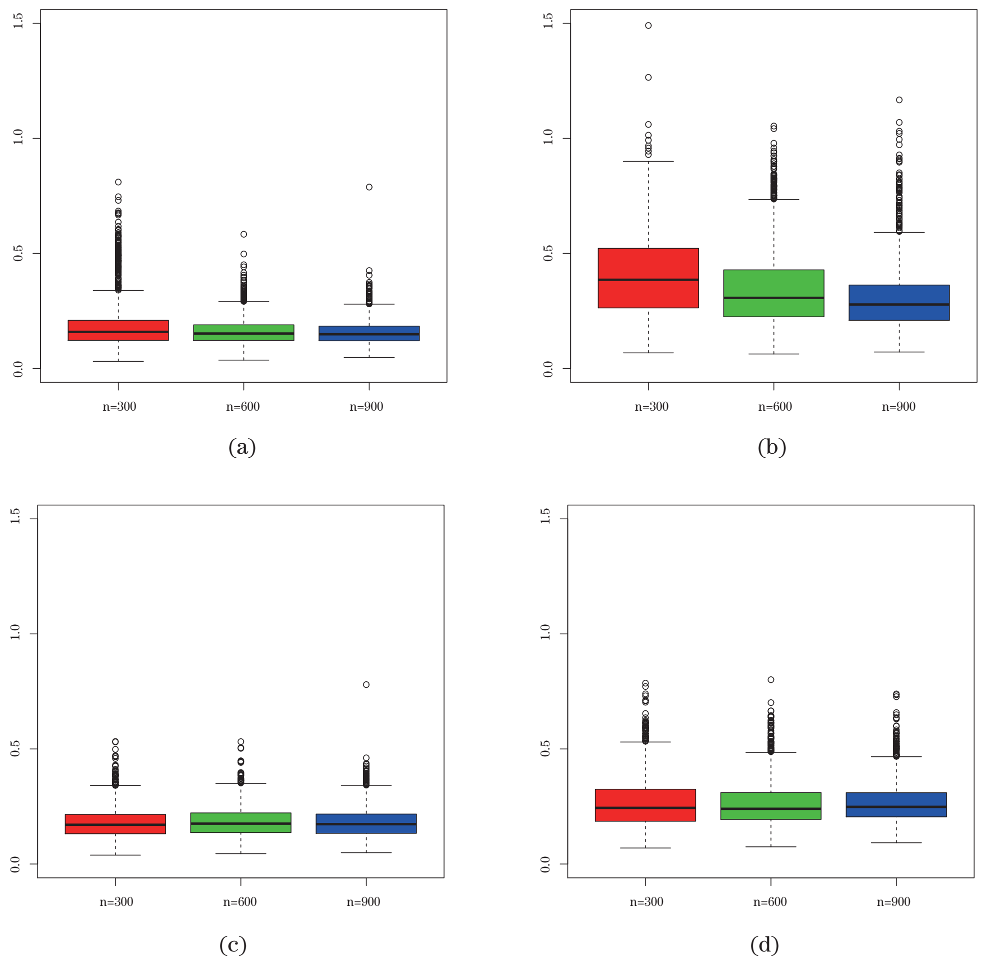
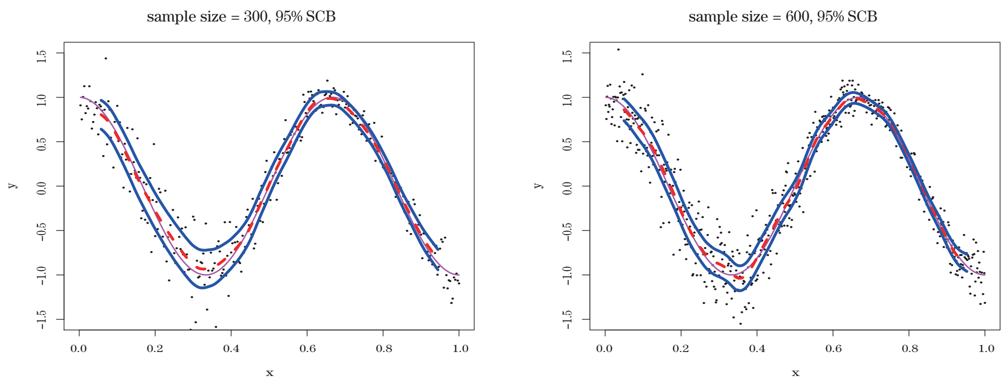


Figure 1. Boxplots of $\triangle_n = \sqrt{n}\max_j |\tilde{\sigma}^2_{\mathrm{K}}(x_j) - \hat{\sigma}^2_{\mathrm{SK}}(x_j)|$ in which $\{x_j, j = 1, 2, \ldots, n_{\mathrm{grid}}\}$ are the points on $\tilde{\mathcal{I}}_n$ with $n_{\mathrm{grid}} = 400$ over 2,000 replications with $\varepsilon \sim N(0,1)$: (a) Case 1; (b) Case 2; (c) Case 3; (d) Case 4.

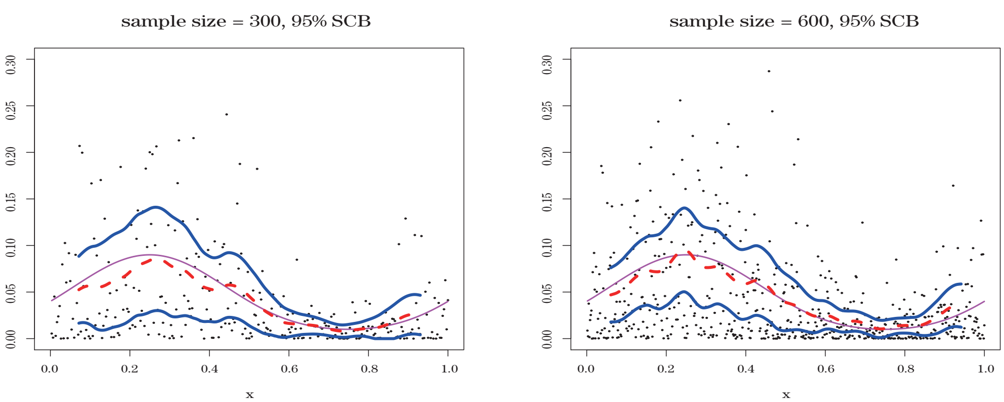
China. Information on strata pressure behavior, range and pressure periodicity in front of a working face is important for the coal mine industry to improve underground mining safety and precision, by preparing the roof support design to prevent accidents caused by sudden increase of strata pressure, see Ju and Xu (2013) and Qian, Shi and Xu (2010).

Strata pressure is the vertical stress on the coal seam roof in front of the working face with unit KN/m^2 (working face is the underground location where miners peel coal from the coal wall mechanically). The pressure sensors are placed at the top of hydraulic supports in front of the working face, and collect data



(a) (b)

Figure 2. Plots of 95% SCB (thick solid) for $m(x)$ (solid) and the estimator $\hat{m}(x)$ (dashed) in Case 1 with $\varepsilon \sim N(0, 1)$ and $n = 300, 600$ respectively.

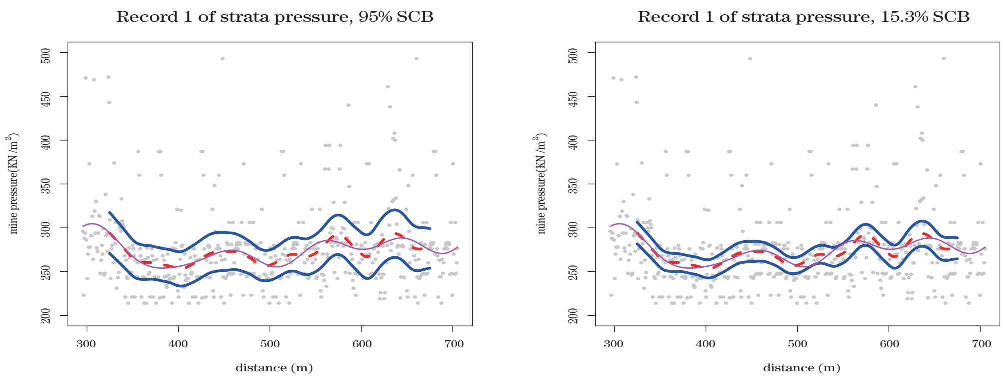


(a) (b)

Figure 3. Plots of 95% SCB (thick solid) for $\sigma^2(x)$ (solid) and the estimator $\hat{\sigma}^2_{\mathrm{SK}}(x)$ (dashed) in Case 1 with $\varepsilon \sim N(0, 1)$ and $n = 300, 600$ respectively.

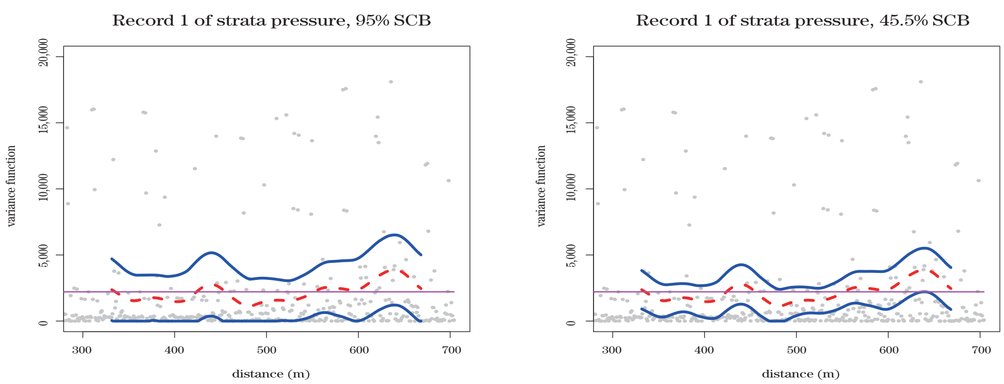
with a record interval of 0.80m: during the mining process, once the hydraulic support has moved forward 0.80m, a pressure sensor records a mine pressure. The propulsion range of the hydraulic support is from 295.5m to 705.1m, so the sample size n is 513. We have chosen from more than 20 pressure records two representative sets for analysis, referred to as records 1 and 2.

A potential concern is whether the independence assumption on errors $\varepsilon_i, 1 \leq i \leq n$, is satisfied in applications. Although it is impossible to "prove" such independence for any data, coal mine experts generally believe that measurement errors in strata pressure are caused by random geological conditions and systematic



(a) (b)

Figure 4. For record 1, plots of the null hypothesis curve of $m(x) = a_0 + \sum_{k=1}^{5} \{a_k \sin(k\omega x) + b_k \cos(k\omega x)\}$ (solid), kernel estimator $\hat{m}(x)$ (dashed), SCB (thick solid) for $m(x)$ with (a) $\alpha = 0.05$ and (b) $\alpha = 0.847$.



(a) (b)

Figure 5. For record 1, plots of the null hypothesis curve of $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} \hat{e}_i^2$ (solid), SCB (thick solid) for $\sigma^2(x)$ and the spline-kernel estimator $\hat{\sigma}^2_{\mathrm{SK}}(x)$ (dashed) with (a) $\alpha = 0.05$ and (b) $\alpha = 0.545$.

errors of the sensors and therefore independent. We plotted for record 1 the autocorrelation function (acf) of the residuals $\hat{\varepsilon}_i = \{y_i - \hat{m}(X_i)\}/\hat{\sigma}_{\mathrm{SK}}(X_i), 1 \le i \le n$, and of $\{|\hat{\varepsilon}_i|\}_{i=1}^n, \{\hat{\varepsilon}_i^2\}_{i=1}^n, \{\hat{\varepsilon}_i^4\}_{i=1}^n$ in Figure S.6 in the supplement. These plots show that the percentage of acfs exceeding the 95% confidence limits was either $1/40 = 0.025$ or $2/40 = 0.05$, hence the null hypothesis of zero autocorrelation was not rejected, for $\hat{\varepsilon}_i, \{|\hat{\varepsilon}_i|\}_{i=1}^n, \{\hat{\varepsilon}_i^2\}_{i=1}^n, \{\hat{\varepsilon}_i^4\}_{i=1}^n$. For the other records, the same acf pattern was observed.

Figure 4 shows the plots of the SCB (thick solid) for record 1 computed according to (4.2) for the mean function $m(x)$, kernel estimate $\hat{m}(x)$ (dashed) with confidence level 95% and 15.3%, respectively. According to the theory of strata pressure, the pressure behavior is periodic; see Qian, Shi and Xu (2010). We therefore proposed the null hypothesis $m\,(x) = a_0 + \sum_{k=1}^{5}\{a_k \sin(k\omega x) + b_k \cos(k\omega x)\}$ to be tested by the SCB for the mean function $m\,(x)$. Since the lowest confidence level of SCB containing the null curve was 15.3%, one retained the null hypothesis with the p-value = 0.847. Likewise, Figure S.7 in the supplement shows the plots of the SCB (thick solid) for record 2 computed according to (4.2) for the mean function, the null hypothesis curve $m\,(x) = a_0 + \sum_{k=1}^{5}\{a_k \sin(k\omega x) + b_k \cos(k\omega x)\}$ (solid) and kernel estimate $\hat{m}\,(x)$ (dashed) with confidence level 95% and 50.8%, respectively. As for record 1, one retained the null hypothesis with the p-value = 0.492. The estimated periodicity for records 1 and 2 were $\omega = 0.0144$ and $\omega = 0.01566$, respectively. Further investigation may lead to conclusive evidence for a general periodicity model for the mean function $m(x)$ with a common ω. Lastly, one observes with 95% confidence that the strata pressure range is [233.5KN/m^2, 320.3KN/m^2] (the lowest and highest values of the 95% SCB curves) for record 1 and [230.8KN/m^2, 370.1KN/m^2] for record 2, also quite close.

In Figure 5 the center dashed line is the spline kernel estimator $\hat{\sigma}^2_{\mathrm{SK}}\,(x)$ for $\sigma^2\,(x)$ and the upper/lower thick solid lines represent the SCB for $\sigma^2\,(x)$. The SCB was used to detect heteroscedasticity in the data with the null hypothesis H_0: $\sigma^2\,(x) \equiv \sigma^2$. Since the lowest confidence level of SCB containing the horizontal line $\hat{\sigma}^2 = 1/n \sum_{i=1}^{n} \hat{e}_i^2$ was 45.5%, where $\hat{\sigma}^2$ is a consistent estimate of σ^2 under H_0, one retained the null hypothesis of homoscedasticity with the p-value = 0.545. In contrast, Figure S.8 in the supplement shows that for record 2 the null hypothesis of homoscedasticity is rejected with the p-value = 0.0024.

Supplementary Materials

The online supplement contains the proofs for the main results and some figures for the simulation and data examples.

Acknowledgment

This research was supported in part by National Natural Science Foundation of China awards 11371272 and 11771240, Research Fund for the Doctoral Program of Higher Education of China award 20133201110002, Jiangsu Key-Discipline Program (Statistics) ZY107992, and the financial support from the Center for Statistical Science at Tsinghua University. We are grateful to Professor Yaodong Jiang and his research group for making the strata pressure data available to us. The authors thank an associate editor and one referee for their helpful comments and suggestions which have led to substantial improvement of this work.

References

Angelini, C., De Canditiis, D. and Frédérique, L. (2003). Wavelet regression estimation in nonparametric mixed effect models. *J. Multivariate Anal.* **85**, 267–291.

Bickel, P. and Rosenblatt, M. (1973). On some global measures of deviations of density function estimates. *Ann. Statist.* **31**, 1852–1884.

Brown, L. and Levine, M. (2007). Variance estimation in nonparametric regression via the difference sequence method. *Ann. Statist.* **35**, 2219–2232.

Cai, T., Levine, M. and Wang, L. (2009). Variance function estimation in multivariate nonparametric regression with fixed design. *J. Multivariate Anal.* **1**, 126–136.

Cai, T., Low, M. and Ma, Z. (2014). Adaptive confidence bands for nonparametric regression functions. *J. Amer. Statist. Assoc.* **109**, 1054–1070.

Cai, L. and Yang, L. (2015). A smooth simultaneous confidence band for conditional variance function. *TEST* **24**, 632–655.

Cao, G., Wang, L., Li, Y. and Yang, L. (2016). Oracle-efficient confidence envelopes for covariance functions in dense functional data. *Statist. Sinica* **26**, 359–383.

Cao, G., Yang, L. and Todem, D. (2012). Simultaneous inference for the mean function based on dense functional data. *J. Nonparametr. Stat.* **24**, 359–377.

Degras, D. (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statist. Sinica* **21**, 1735–1765.

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

Donoho, D. and Johnstone, I. (1996). Neo-classical minimax problems, thresholding and adaptive function estimation. *Bernoulli* **13**, 7–9.

Eubank, R. and Speckman, P. (1993). Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.* **88**, 1287–1301.

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

Gu, L. and Yang, L. (2015). Oracally efficient estimation for single-index link function with simultaneous confidence band. *Electron. J. Stat.* **9**, 1540–1561.

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

Härdle, W. and Marron, J. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19**, 778–796.

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

Johnston, G. (1982). Probabilities of maximal deviations for nonparametric regression function estimates. *J. Multivariate Anal.* **12**, 402–414.

Ju, J. and Xu, J. (2013). Structural characteristics of key strata and strata behaviour of a fully mechanized longwall face with 7.0 m height chocks. *Int. J. of Rock Mech. Min. Sci.* **58**, 46–54.

Komlós, J., Major, P. and Tusnády, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* **34**, 33–58.

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

Levine, M. (2006). Bandwidth selection for a class of difference-based variance estimators in the nonparametric regression: A possible approach. *Comput. Statist. Data Anal.* **50**, 3405–3431.

Ma, S., Yang, L. and Carroll, R. (2012). A simultaneous confidence band for sparse longitudinal regression. *Statist. Sinica* **22**, 95–122.

Qian, M., Shi, P. and Xu, J. (2010). *Mining Pressure and Strata Control.* China University of Mining and Technology Press.

Song, Q. and Yang, L. (2009). Spline confidence bands for variance functions. *J. Nonparametr. Stat.* **5**, 589–609.

Stapleton, J. (2009). *Linear Statistical Models.* 2nd Edition. John Wiley & Sons, Hoboken, NJ.

Wang, J. (2012). Modelling time trend via spline confidence band. *Ann. Inst. Stat. Math.* **64**, 275–301.

Wang, L., Brown, L., Cai, T. and Levine, M. (2008). Effect of mean on variance function estimation in nonparametric regression. *Ann. Statist.* **36**, 646–664.

Wang, J., Cheng, F. and Yang, L. (2013). Smooth simultaneous confidence bands for cumulative distribution functions. *J. Nonparametr. Stat.* **25**, 395–407.

Wang, J., Liu, R., Cheng, F. and Yang, L. (2014). Oracally efficient estimation of autoregressive error distribution with simultaneous confidence band. *Ann. Statist.* **42**, 654–668.

Wang, J. and Yang, L. (2009). Polynomial spline confidence bands for regression curves. *Statist. Sinica* **19**, 325–342.

Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol* **60**, 797–811.

Xue, L and Yang, L. (2006). Additive coefficient modeling via polynomial spline. *Statist. Sinica* **16**, 1423–1446.

Center for Advanced Statistics and Econometrics Research, Soochow University, Suzhou, 215006, China.

Center for Statistical Science and Department of Industrial Engineering, Tsinghua University, Beijing 100084, China.

E-mail: caili16@126.com

Department of Mathematics and Statistics, University of Toledo, Toledo, OH 43006, USA.

E-mail: rong.liu@utoledo.edu

Department of Statistics, Texas A&M University, Texas, TX 77843, USA.

E-mail: sjwang@stat.tamu.edu

Center for Statistical Science and Department of Industrial Engineering, Tsinghua University, Beijing 100084, China.

E-mail: yanglijian@mail.tsinghua.edu.cn

(Received January 2017; accepted August 2017)