TEST
DOI 10.1007/s11749-017-0543-5

ORIGINAL PAPER

A smooth simultaneous confidence band for correlation curve

Yuanyuan Zhang[1] · **Lijian Yang**[1]

Received: 21 December 2016 / Accepted: 2 May 2017

Abstract A plug-in estimator is proposed for a local measure of variance explained by regression, termed correlation curve in Doksum et al. (J Am Stat Assoc 89:571–582, 1994), consisting of a two-step spline–kernel estimator of the conditional variance function and local quadratic estimator of first derivative of the mean function. The estimator is oracally efficient in the sense that it is as efficient as an infeasible correlation estimator with the variance function known. As a consequence of the oracle efficiency, a smooth simultaneous confidence band (SCB) is constructed around the proposed correlation curve estimator and shown to be asymptotically correct. Simulated examples illustrate the versatility of the proposed oracle SCB which confirms the asymptotic theory. Application to a 1995 British Family Expenditure Survey data has found marginally significant evidence for a local version of Engel's law, i.e., food budget share and household real income are inversely related (Hamilton in Am Econ Rev 91:619–630, 2001).

Keywords Confidence band · Correlation curve · Heteroscedasticity · Infeasible estimator · Local quadratic estimator

Mathematics Subject Classification 62G05 · 62G08 · 62G10 · 62G15 · 62G20 · 62P20

Electronic supplementary material The online version of this article (doi:10.1007/s11749-017-0543-5) contains supplementary material, which is available to authorized users.

✉ Lijian Yang
yanglijian@tsinghua.edu.cn

[1] Center for Statistical Science and Department of Industrial Engineering, Tsinghua University, Beijing 100084, China

Published online: 15 May 2017

1 Introduction

Correlation coefficient is used to measure the strength of linear relationship in many contexts of regression and multivariate analysis, see for instance Stapleton (2009). For a pair of random variables (X, Y) that satisfies a homoscedastic linear regression model

$$Y = \beta_0 + \beta_1 X + \sigma_0 \varepsilon, \tag{1}$$

where $\mathsf{E}(\varepsilon | X) = 0, \mathsf{E}\left(\varepsilon^2 | X\right) = 1$, the Galton–Pearson correlation coefficient between X and Y is

$$\rho = \frac{\sigma_1 \beta_1}{\left(\sigma_1^2 \beta_1^2 + \sigma_0^2\right)^{1/2}}, \rho \in [-1, 1] \tag{2}$$

in which $\sigma_1^2 = \mathrm{var}(X)$.

This classic definition of correlation does not provide useful information about the relationship between X and Y, however, when there is strong nonlinear or heteroscedastic dependence. To be precise, suppose that (X, Y) satisfy a nonparametric regression model

$$Y = \mu(X) + \sigma(X)\varepsilon, \tag{3}$$

with nonparametric conditional mean function $\mu(x)$ and conditional variance function $\sigma^2(x)$. Doksum et al. (1994) had creatively defined a localized "correlation curve" at $X = x$ as

$$\rho(x) \equiv \frac{\sigma_1 \beta(x)}{\left\{\sigma_1^2 \beta^2(x) + \sigma^2(x)\right\}^{1/2}}, \tag{4}$$

in which $\beta(x) = \mu'(x)$ and $\sigma^2(x)$ are, respectively, local analogs of slope β_1 and variance σ_0^2 in (2), since for smooth functions $\mu(\cdot)$ and $\sigma(\cdot)$, $Y \approx \mu(x) + \beta(x)(X - x) + \sigma(x)\varepsilon$ when X takes value in a small neighborhood of x. If in fact $\mu(x) \equiv \beta_0 + \beta_1 x$ and $\sigma^2(x) \equiv \sigma_0^2$, then $\rho(x)$ reduces to the Galton–Pearson correlation coefficient in (2).

The correlation curve $\rho(\cdot)$ measures the strength of the relationship between Y and X in heterocorrelatious experiments, in terms of the variance explained by regression at every covariate value of X. It is scale and location invariant, always between -1 and 1, and equals ± 1 when Y is a nonconstant function of X (i.e., $\sigma^2(x) \equiv 0, \beta(x) \neq 0$).

The problem we study concerns a random sample $\{(X_i, Y_i)\}_{i=1}^n$ independently and identically distributed as (X, Y), satisfying (3) as follows

$$Y_i = \mu(X_i) + \sigma(X_i)\varepsilon_i, \tag{5}$$
$$\mathsf{E}(\varepsilon_i | X) = 0, \mathsf{E}\left(\varepsilon_i^2 | X\right) = 1,$$

in which functions $\mu(\cdot)$ and $\sigma(\cdot)$ are defined over a compact interval $S = [a, b]$, and the distribution of X is supported on S. Since the $\rho(x)$ in (4) is computed from three quantities $\beta(x), \sigma^2(x)$ and σ_1^2, Doksum et al. (1994) proposed a plug-in estimate of

$\rho(\cdot)$ by utilizing local estimates of $\beta(x)$ and $\sigma^2(x)$, and standard estimate of σ_1^2

$$\hat{\sigma}_1^2 = (n-1)^{-1} \sum\nolimits_{i=1}^{n} \left(X_i - \bar{X}\right)^2. \tag{6}$$

An asymptotic pointwise confidence interval was also formulated for $\rho(x)$ at any fixed x based on the plug-in estimate.

What Doksum et al. (1994) did not supply was a working simultaneous confidence band (SCB) for $\rho(x)$ over a range of x values. SCBs are versatile tools for statistical inference about global properties of unknown functions, and they play the same role as confidence intervals for parameters, see Bickel and Rosenblatt (1973), Hall and Titterington (1988), Härdle (1989), Eubank and Speckman (1993), Xia (1998), and Claeskens and Van Keilegom (2003) for early statistical literature on SCB. Although SCBs pose many theoretical challenges, they have wide applications in numerous areas such as dimension reduction (Gu and Yang 2015; Zheng et al. 2016), functional data analysis (Ma et al. 2012; Cao et al. 2012; Gu et al. 2014; Song et al. 2014; Zheng et al. 2014; Cao et al. 2016), sample survey (Wang et al. 2016), time series analysis (Wu and Zhao 2007; Zhao and Wu 2008; Wang et al. 2014), distribution estimation (Wang et al. 2013), and variance estimation (Song and Yang 2009; Cai and Yang 2015.

To provide an SCB for $\rho(x)$, we have modified the estimator of Doksum et al. (1994) in several crucial aspects. First, the derivative $\beta(x) = \mu'(x)$ is calculated by local quadratic procedure instead of by the less efficient Gasser and Müller kernel estimator. While the study of local polynomial regression was as early as Stone (1977, 1980) and Cleveland (1979), it became widely popular in the 1990s, with desirable properties such as minimax efficiency, automatic boundary correction, and design adaptivity established in works such as Fan (1993), Ruppert and Wand (1994), and Fan and Gijbels (1996).

The direct estimator $\hat{\sigma}^2(x) = \hat{\mu}_2(x) - \hat{\mu}^2(x)$ of $\sigma^2(x) = \mathsf{E}\left(Y^2 \mid X = x\right) - \mu^2(x)$ is dropped, where $\hat{\mu}(x)$ and $\hat{\mu}_2(x)$ are, respectively, kernel estimators (Gasser et al. 1984) for $\mu(x)$ and $\mu_2(x) \equiv \mathsf{E}\left(Y^2 \mid X = x\right)$. Two serious reasons have motivated this second innovation: this simple plug-in estimator can be negative, while the true variance is positive, and its bias is much larger than that of a two-step variance estimator in Fan and Yao (1998). We have replaced $\hat{\sigma}^2(x)$ by the spline–kernel variance estimator $\hat{\sigma}^2_{\mathrm{SK}}(x)$, which is as efficient as an infeasible estimator $\tilde{\sigma}^2_{\mathrm{K}}(x)$, see Cai and Yang (2015) for details.

This paper is organized as follows. Section 2 defines the local quadratic estimator of the correlation curves and establishes its oracle property with asymptotic SCB. Section 3 provides concrete steps to implement the SCB. Sections 4 and 5 illustrate the usefulness of the proposed SCB via simulation study and application to a well-known cross-sectional data from the British Family Expenditure Survey. Section 6 concludes, while all technical proofs are in "Appendix."

2 Main result

The population derivative $\beta(x)$ is approximated by the local quadratic estimator $\hat{\beta}(x)$ which solves a kernel-weighted least-squares problem. The kernel weights $K_{h_1}(X_i -$

$x)$, $i = 1, \ldots, n$ weed out observations (X_i, Y_i) from the sum of squares if $|X_i - x| > h_1$, in which $K_{h_1}(\cdot) = K(\cdot/h_1)/h_1$, $K(\cdot)$ a symmetric function supported on $[-1, 1]$, $h_1 = h_{1,n} > 0$ a sequence of smoothing parameters called bandwidth. More precisely, according to Fan and Gijbels (1996), $\hat{\beta}(x) = \hat{c}_1$, in which

$$\left(\hat{c}_0, \hat{c}_1, \hat{c}_2\right) = \underset{(c_0,c_1,c_2)\in\mathbb{R}^3}{\operatorname{argmin}} \sum\nolimits_{i=1}^{n} \left\{Y_i - \sum\nolimits_{\alpha=0}^{2} c_\alpha (X_i - x)^\alpha\right\}^2 K_{h_1}(X_i - x).$$

Alternatively, one writes in matrix form

$$\hat{\beta}(x) = e_1^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Y}, \tag{7}$$

$$e_k^T = (\delta_{0k}, \delta_{1k}, \delta_{2k}), \text{ with } \delta_{kk'} = 1 \text{ if } k = k', 0 \text{ otherwise}, \tag{8}$$

in which

$$\mathbf{X}^T = \mathbf{X}(x) = \begin{pmatrix} 1 & ,\ldots, & 1 \\ X_1 - x & ,\ldots, & X_n - x \\ (X_1 - x)^2 & ,\ldots, & (X_n - x)^2 \end{pmatrix}, \tag{9}$$

$$\mathbf{W} = \mathbf{W}(x) = n^{-1} \operatorname{diag}\left\{K_{h_1}(X_i - x)\right\}_{i=1}^{n}. \tag{10}$$

As mentioned in Sect. 1, we have adopted the two-step estimator of Cai and Yang (2015) for $\sigma^2(x)$. The idea is as follows. If the mean function $\mu(\cdot)$ were known a priori, one could compute $Z_i = \{Y_i - \mu(X_i)\}^2$, $1 \le i \le n$ and obtain a pseudo-data set $\{(X_i, Z_i)\}_{i=1}^n$, which satisfies $\mathsf{E}(Z_i | X = x) = \sigma^2(x)$; hence, an "infeasible kernel estimator" of the variance function is

$$\tilde{\sigma}_{\mathrm{K}}^2(x) = \frac{\sum_{i=1}^n K_{h_2}(X_i - x) Z_i}{\sum_{i=1}^n K_{h_2}(X_i - x)}, \tag{11}$$

where $h_2 = h_{2,n}$ is the bandwidth, and $K_{h_2}(\cdot) = K(\cdot/h_2)/h_2$. A spline–kernel estimator $\hat{\sigma}_{\mathrm{SK}}(x)$ of $\sigma^2(x)$ is

$$\hat{\sigma}_{\mathrm{SK}}^2(x) = \frac{\sum_{i=1}^n K_{h_2}(X_i - x) \hat{Z}_i}{\sum_{i=1}^n K_{h_2}(X_i - x)} \tag{12}$$

where $\hat{Z}_i = \left\{Y_i - \hat{\mu}_p(X_i)\right\}^2$ are the residual squares of spline estimator $\hat{\mu}_p(\cdot)$, defined in (13). Cai and Yang (2015) proved that $\hat{\sigma}_{\mathrm{SK}}^2(x)$ not only resembles $\tilde{\sigma}_{\mathrm{K}}^2(x)$ in form, but also is asymptotically equivalent, uniformly for $x \in [a + h_2, b - h_2]$.

To introduce the spline estimator $\hat{\mu}_p(\cdot)$, the interval $[a, b]$ is divided into $(N + 1)$ subintervals $J_j = [t_j, t_{j+1})$, $j = 0, \ldots, N-1$, $J_N = [t_N, 1]$ by a sequence of equally spaced points $\left\{t_j\right\}_{j=1}^N$, called interior knots, given as

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

in which $H = (b-a)/(N+1)$ is the distance between neighboring knots. For an integer $p > 0$, the spline estimator $\hat{\mu}_p(\cdot)$ is

$$\hat{\mu}_p(\cdot) = \underset{g \in G_N^{(p-2)}}{\text{argmin}} \sum_{i=1}^{n} \{Y_i - g(X_i)\}^2, \tag{13}$$

in which the space of p-th order splines on interval $[a, b]$, $G_N^{(p-2)} = G_N^{(p-2)}[a, b]$, is defined as the space of functions that are polynomials degree $(p-1)$ on each J_j, $j = 0, 1, \ldots, N$ and have continuous $(p-2)$-th derivative on $[a, b]$. This space $G_N^{(p-2)}$ has dimension $N + p$ with B spline basis $\left\{b_{j,p}(\cdot)\right\}_{j=1-p}^{N}$ defined in de Boor (2001), p.87, i.e., $G_N^{(p-2)} = \left\{\sum_{j=1-p}^{N} \lambda_j b_{j,p}(\cdot) \,\middle|\, \lambda_j \in \mathbb{R}, 1-p \le j \le N\right\}$.

If one makes use of (6), (7), and (12), the local quadratic correlation curve (LQCC) is a plug-in estimate of $\rho(x)$ in (4)

$$\hat{\rho}_{\text{LQ}}(x) = \frac{\hat{\sigma}_1 \hat{\beta}(x)}{\left\{\hat{\sigma}_1^2 \hat{\beta}^2(x) + \hat{\sigma}_{\text{SK}}^2(x)\right\}^{1/2}}, x \in [a, b]. \tag{14}$$

For benchmarking, one denotes

$$\tilde{\rho}_{\text{LQ}}(x) = \frac{\sigma_1 \hat{\beta}(x)}{\left\{\sigma_1^2 \hat{\beta}^2(x) + \sigma^2(x)\right\}^{1/2}}, x \in [a, b] \tag{15}$$

as an intermediate infeasible estimator of $\rho(x)$.

To formulate the necessary technical assumptions, for sequences of real numbers c_n and d_n, one writes $c_n \ll d_n$ to mean $c_n/d_n \to 0$, as $n \to \infty$, and $c_n \sim d_n$ to mean for any n, $|c_n/d_n| + |d_n/c_n| \le M < \infty$.

(A1) *The function* $\mu(\cdot) \in C^{(3)}[a, b]$.

(A2) *The joint distribution of* (X, ε) *is bivariate continuous with* $\mathsf{E}(\varepsilon | X) = 0$, $\mathsf{E}\left(\varepsilon^2 | X\right) = 1$, *and there exists a* $\eta_2 > 1/2$ *such that*

$$\sup_{x \in [a,b]} \mathsf{E}\left(|\varepsilon|^{4+2\eta_2} | X = x\right) = M_{\eta_2} \in (0, +\infty)$$

and consequently for a $\eta_1 > 1/3$,

$$\sup_{x \in [a,b]} \mathsf{E}\left(|\varepsilon|^{2+\eta_1} | X = x\right) = M_{\eta_1} \in (0, +\infty)$$

as well.

(A3) *The density function of predictor* X, $f(\cdot) \in C^{(1)}[a, b]$, *the variance function* $\sigma^2(\cdot) \in C^{(2)}[a, b]$, $f(x) \in \left[c_f, C_f\right]$, $\sigma(x) \in [c_\sigma, C_\sigma]$, $\forall x \in [a, b]$, *for constants* $0 < c_f < C_f < +\infty$, $0 < c_\sigma < C_\sigma < +\infty$.

(A4) *The kernel function* $K(\cdot) \in C^{(2)}(\mathbb{R})$ *is a symmetric probability density function supported on* $[-1, 1]$.
(A5) *The bandwidth* h_1 *satisfies* $n^{2\alpha_1 - 1}\log^4 n \ll h_1 \ll n^{-1/7}\log^{-1/7} n$, *for some* α_1 *such that* $\alpha_1 \in (2/5, 3/7)$, $\alpha_1(2+\eta_1) > 1$, $\alpha_1(1+\eta_1) > 2/7$. *In particular, one may take* $h_1 \sim n^{-1/7}(\log n)^{-1/7-\delta_1}$ *for any* $\delta_1 > 0$.
(A6) *The bandwidth* h_2 *satisfies* $n^{2\alpha_2 - 1}\log n \ll h_2 \ll n^{-1/5}\log^{-1/5} n$, *for some* $\alpha_2 \in (2/7, 2/5)$ *such that*, $\alpha_2(2+\eta_2) > 1$, $\alpha_2(1+\eta_2) > 2/5$. *In particular, one may take* $h_2 \sim n^{-1/5}(\log n)^{-1/5-\delta_2}$ *for any* $\delta_2 > 0$.
(A7) *The number of interior knots* $N = N_n$ *satisfies*

$$\max\left\{\left(nh_2^{-2}\right)^{1/8}, \frac{\log^{1/2} n}{h_2^{1/2}}\right\} \ll N \ll \min\left\{n^{1/2}h_2, \left(\frac{nh_2}{\log n}\right)^{1/3}, \left(nh_2^{-1}\right)^{1/5}\right\}.$$

Assumptions (A1)–(A3) are adopted from Song and Yang (2009), Assumption (A4) is standard for kernel regression, and Assumption (A5) is a general condition on bandwidth of h_1 leading to the asymptotic Gumbel distribution for $\tilde{\rho}_{\text{LQ}}(\cdot)$. Assumptions (A6) and (A7) are general conditions adopted from Cai and Yang (2015) on the choice of bandwidth h_2 for (12) and number of knots N to guarantee oracle efficiency between $\hat{\rho}_{\text{LQ}}(\cdot)$ and $\tilde{\rho}_{\text{LQ}}(\cdot)$. It is worth noting that Assumptions (A5)–(A6) imply that $h_2 \ll n^{-1/5}\log^{-1/5} n \ll n^{2\alpha_1-1}\log^4 n \ll h_1$ as $\alpha_1 > 2/5$, while $h_2^{-1/2} \ll \left(n^{2\alpha_2-1}\log n\right)^{-1/2} = n^{1/2-\alpha_2}\log^{-1/2} n \ll n^{3/14}\log^{-11/14} n$ as $\alpha_2 > 2/7$, and $n^{3/14}\log^{-11/14} n \ll h_1^{-3/2}\log^{-1} n$ because $h_1 \ll n^{-1/7}\log^{-1/7} n$. In summary, the following asymptotic relations hold between h_1 and h_2, as $n \to \infty$

$$h_2 \ll h_1, h_2^{-1/2} \ll h_1^{-3/2}\log^{-1} n, \tag{16}$$

and in particular, for large enough n,

$$\mathcal{I}_n = [a + h_1, b - h_1] \subset [a + h_2, b - h_2], \tag{17}$$

where $\mathcal{I}_n$, the interval over which all SCBs are constructed, grows with n to (a, b). Data-driven implementation of h_1, h_2 and N that satisfies all the requirements in Assumptions (A5)–(A7), respectively, is given in Sect. 3, aided by explicit formula (26) and Eq. (15) in Cai and Yang (2015) for rule-of-thumb bandwidths.

An SCB based on infeasible estimator $\tilde{\rho}_{\text{LQ}}(x)$ is constructed by the delta method, similar to Carroll and Ruppert (1988). Undersmoothing is performed as in Hall (1991, 1992), Claeskens and Van Keilegom (2003) to handle the nonparametric regression bias in $\hat{\beta}(x)$, more efficiently than explicit bias correction.

Following Aerts and Claeskens (1997), an asymptotic standard deviation of $\tilde{\rho}_{\text{LQ}}(x)$ is obtained, namely

$$V_n(x) = \sigma_1\left\{1 - \rho^2(x)\right\}^{3/2}\left\{n^{-1}h_1^{-3}f^{-1}(x)C_{K^*}\right\}^{1/2}, \tag{18}$$

where one denotes

$$C_{K^*} = \int K^* (u)^2 \, \mathrm{d}u, C_{K^{*\prime}} = \int K^{*\prime} (u)^2 \, \mathrm{d}u, C \left(K^*\right) = C_{K^{*\prime}}^{1/2} / C_{K^*}^{1/2}, \tag{19}$$

in which K^* is the equivalent kernel of order 3,

$$K^*(u) = \left(\sum\nolimits_{\lambda=0}^{2} s_{1\lambda} u^\lambda\right) K(u), \; (s_{\lambda\delta})_{\lambda,\delta=0}^2 \times (\mu_{\lambda+\delta}(K))_{\lambda,\delta=0}^2 = \mathbf{I}_3, \tag{20}$$

and the moments of K, $\mu_j(K) = \int v^j K(v) \, \mathrm{d}v$, see Fan and Gijbels (1996) and Gasser et al. (1984) for details. Denote also

$$a_{h_1} = \sqrt{2 \log \left\{h_1^{-1} (b-a)\right\}}, b_{h_1} = a_{h_1} + a_{h_1}^{-1} \left\{\sqrt{C(K^*)}/2\pi\right\}. \tag{21}$$

The next proposition follows from standard SCB theory.

Proposition 1 *Under Assumptions (A1)–(A5), as $n \to \infty$,*

$$\mathbb{P}\left[a_{h_1} \left\{\sup_{x \in \mathcal{I}_n} \left|\tilde{\rho}_{\mathrm{LQ}}(x) - \rho(x)\right| / V_n(x) - b_{h_1}\right\} \le z\right] \to e^{-2e^{-z}}, z \in \mathbb{R}. \tag{22}$$

Equation (22) yields an infeasible $100(1-\alpha)\,\%$ SCB for $\rho(x)$ over $\mathcal{I}_n$

$$\tilde{\rho}_{\mathrm{LQ}}(x) \pm V_n(x) \left[2 \log \left\{h_1^{-1}(b-a)\right\}\right]^{1/2} Q_n(\alpha), \tag{23}$$

where

$$Q_n(\alpha) = 1 + \frac{\log\{C(K^*)/2\pi\} - \log\{-1/2 \log(1-\alpha)\}}{2 \log \left\{h_1^{-1}(b-a)\right\}}. \tag{24}$$

As pointed out by one reviewer, alternative estimators for $\beta(x)$ such as B spline estimator are also feasible, we have chosen the local quadratic approach for the ease of deriving maximal deviation result (22) in Proposition 1, and the SCB in (23). The adjective "infeasible" highlights the fact that $\tilde{\rho}_{\mathrm{LQ}}(x)$ contains unknown variance function $\sigma^2(x)$; thus, it is not a proper statistic.

Theorem 1 *Under Assumptions (A1)–(A7), as $n \to \infty$,*

$$\begin{aligned}\sup_{x \in \mathcal{I}_n} \left|\tilde{\rho}_{\mathrm{LQ}}(x) - \hat{\rho}_{\mathrm{LQ}}(x)\right| &= \mathcal{O}_p\left(n^{-1/2} h_2^{-1/2} \log^{1/2} n\right) \\ &= o_p\left(n^{-1/2} h_1^{-3/2} \log^{-1/2} n\right),\end{aligned}$$

and consequently, since $a_{h_1} \sim \log^{1/2} n$, $V_n^{-1}(x) = U_p\left(n^{1/2} h_1^{3/2}\right)$,

$$a_{h_1} \sup_{x \in \mathcal{I}_n} \left|\tilde{\rho}_{\mathrm{LQ}}(x) - \hat{\rho}_{\mathrm{LQ}}(x)\right| / V_n(x) = o_p(1), \tag{25}$$

i.e., the estimator $\hat{\rho}_{\mathrm{LQ}}(x)$ *is asymptotically as efficient as "the infeasible estimator"* $\tilde{\rho}_{\mathrm{LQ}}(x)$*, up to uniform order* $n^{-1/2}h_1^{-3/2}\log^{-1/2} n$ *over* $\mathcal{I}_n$.

Proof of Theorem 1 in "Appendix" depends in part on the uniform bound $\sup_{x\in[a+h_2,b-h_2]}\left|\hat{\sigma}^2_{\mathrm{SK}}(x)-\sigma^2(x)\right|$ being of order $\mathcal{O}_p\left(n^{-1/2}h_2^{-1/2}\log^{1/2} n\right)$, which is the reason spline instead of kernel smoothing is used for step one of variance function estimation, both in Cai and Yang (2015) and this work. Regardless linear or higher-order spline is used in step one, the squares of residuals will undergo kernel smoothing in step two; hence, the spline–kernel estimator $\hat{\sigma}^2_{\mathrm{SK}}(x)$ is always smooth.

Putting together Theorem 1 and Proposition 1, one obtains the main result.

Theorem 2 *Under Assumptions (A1)–(A7), an asymptotic* $100(1-\alpha)\%$ *oracle SCB for* $\rho(x)$ *over* $\mathcal{I}_n$ *is*

$$\hat{\rho}_{\mathrm{LQ}}(x) \pm a_{h_1} V_n(x)\, Q_n(\alpha),$$

with $V_n(x)$ *in (18) and* $Q_n(\alpha)$ *in (24). In other words,*

$$\lim_{n\to\infty}\mathbb{P}\left\{\rho(x)\in\hat{\rho}_{\mathrm{LQ}}(x)\pm a_{h_1}V_n(x)\,Q_n(\alpha),\, x\in\mathcal{I}_n\right\}=1-\alpha.$$

The infeasible SCB in Proposition 1 and the oracle SCB in Theorem 2 both shrink to zero at the rate of $n^{-1/2}h_1^{-3/2}\log^{1/2} n$, which is $n^{-2/7}\log^{1.34} n$ for the implemented bandwidth $\hat{h}_1 \sim n^{-1/7}\log^{-0.56} n$ in Sect. 3. This rate is slightly slower than the mean square optimal convergence rate for estimating derivative $\beta(\cdot)$, which is $n^{-2/7}$. The proofs of Proposition 1, and Theorems 1 and 2 are based on Lemmas 1–8, all of which are in "Appendix."

3 Implementation

To choose an appropriate plug-in bandwidth $h_1 = h_{1,n}$ for computation $\hat{\beta}(x)$, one makes use of the following rule-of-thumb (ROT) bandwidth of Fan and Gijbels (1996),

$$h_{1,\mathrm{rot}}=\left\{\frac{8505/11\sum_{i=1}^{n}\left(Y_i-\sum_{k=0}^{5}\widehat{a}_kX_i^k\right)^2}{n\sum_{i=1}^{n}\left(6\widehat{a}_3+24\widehat{a}_4X_i+60\widehat{a}_5X_i^2\right)^2}\right\}^{1/7}, \tag{26}$$

in which $\{\widehat{a}_k\}_{k=0}^5=\mathrm{argmin}_{\{a_k\}_{k=0}^5\in\mathbb{R}^6}\sum_{i=1}^n\left(Y_i-\sum_{k=0}^5 a_kX_i^k\right)^2$. One then sets $\hat{h}_1=\hat{h}_{1,n}=h_{1,\mathrm{rot}}\log^{-0.56} n\sim n^{-1/7}\log^{-0.56} n$, which clearly satisfies Assumption (A5), especially the undersmoothing condition $h_1 \ll n^{-1/7}\log^{-1/7} n$.

For constructing SCB, the unknown function $f(x)$ is evaluated and then plugged in, the same approach taken in Hall and Titterington (1988), Härdle (1989), Xia (1998), Wang and Yang (2009), Song and Yang (2009). Without loss of generality, let the kernel $K(u) = 15\left(1-u^2\right)^2 I\{|u| \le 1\}/16$ be the quartic kernel in (20) and

$$\hat{f}(x) = n^{-1}\sum_{i=1}^{n} h_{1,\mathrm{rot,f}}^{-1} K\left(\frac{X_i - x}{h_{1,\mathrm{rot,f}}}\right),$$
$$h_{1,\mathrm{rot,f}} = (4\pi)^{1/10}\left(\frac{140}{3}\right) n^{-1/5}\hat{\sigma}_1,$$

where $h_{1,\mathrm{rot,f}}$ is the rule-of-thumb bandwidth in Silverman (1986).

To satisfy Assumption (A4), the kernel K is chosen to be the quartic kernel. To satisfy Assumption (A6), the bandwidth $\hat{h}_2 = \hat{h}_{2,n} = h_{2,\mathrm{rot}}$ is used for the computing of $\hat{\sigma}^2_{\mathrm{SK}}(x)$, where the ROT bandwidth $h_{2,\mathrm{rot}}$ is from Cai and Yang (2015).

Although splines of any order can be employed, we have used linear splines (with $p = 2$). To select the number of interior knots N, let $\hat{N}$ be the minimizer of Bayesian Information Criterion (BIC) defined below, over integers from $[0.5N_r, \min(5N_r, Tb)]$, with $N_r = n^{-1/5}$ and $Tb = n/4 - 1$. This ensures that $\hat{N}$ is order of $n^{-1/5}$ and the total number of parameters in the spline least-squares regression is no more than $n/4$. The chosen $\hat{N}$ obviously satisfies Assumption (A7), but other choices of N remain open possibility.

For any candidate integer $N_n \in [0.5N_r, \min(5N_r, Tb)]$, denote the predictor for the i-th response Y_i by $\hat{Y}_i = \hat{\mu}_2(X_i)$. Let $q_n = (1 + N_n)$ be the total number of parameters in (13). The BIC value corresponding to N_n is,

$$\mathrm{BIC}(N_n) = \log(\mathrm{MSE}) + q_n \log n/n,\ \mathrm{MSE} = n^{-1}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2,$$

and $\hat{N} = \mathrm{argmin}_{N_n \in [0.5N_r, \min(5N_r, Tb)]} \mathrm{BIC}(N_n)$.

Algebra shows that the least-squares problem in Eq. (13) can be also solved via the truncated power basis $\left\{1, x, \left(x - t_j\right)_+, j = 1, 2, \ldots, \hat{N}\right\}$, see de Boor (2001), which is regularly used in implementation. In other words,

$$\hat{\mu}_2(x) = \hat{r}_0 + \hat{r}_1 x + \sum_{j=1}^{\hat{N}} \hat{r}_{j,2}\left(x - t_j\right)_+,$$

where the coefficients $\left(\hat{r}_0, \hat{r}_1, \hat{r}_{1,2}, \ldots, \hat{r}_{\hat{N},2}\right)^T$ are solutions to the least-squares problem

$$\left(\hat{r}_0, \ldots, \hat{r}_{\hat{N},2}\right)^T = \underset{\mathbb{R}^{\hat{N}+2}}{\mathrm{argmin}} \sum_{i=1}^{n}\left\{Y_i - r_0 - r_1 X_i - \sum_{j=1}^{\hat{N}} r_{j,2}\left(X_i - t_j\right)_+\right\}^2.$$

All above together can contribute to the plug-in estimator $\hat{\rho}_{\mathrm{LQ}}(x)$. Then the function $V_n(x)$ is approximated by the following,

$$\hat{V}_n(x) = \hat{\sigma}_1\left\{1 - \hat{\rho}^2_{\mathrm{LQ}}(x)\right\}^{3/2}\left\{n^{-1}\hat{h}_1^{-3}\hat{f}^{-1}(x)\int K^*(v)^2\,\mathrm{d}v\right\}^{1/2}. \quad (27)$$

Hence, the oracle SCB in Theorem 2 is computed as the following asymptotically $100\,(1-\alpha)\,\%$ SCB

$$\hat{\rho}_{\mathrm{LQ}}(x) \pm \hat{V}_n\,(x)\left[2\log\left\{\hat{h}_1^{-1}\,(b-a)\right\}\right]^{1/2} Q_n\,(\alpha)\,,\, x \in \left[a+\hat{h}_1, b-\hat{h}_1\right]. \quad (28)$$

All computing in the next two sections is carried out according to the above specifications, using the open-access environment R for statistical computing and graphics, developed by the R Core Team (2013).

4 Simulation

To illustrate the finite-sample behavior of the oracle SCB in (28), data sets are generated from models (1) and (3) with independent X and ε, and $X \sim U\,(0.8, 1.6)\,,\, \varepsilon \sim N\,(0, 1)$, with the following mean and variance functions

$$\text{Case1: } \mu\,(x) = 0.8 - 0.14x,\, \sigma\,(x) = 0.09,$$
$$\text{Case2: } \mu\,(x) = 0.2\sin\,(4\pi x)\,,\, \sigma\,(x) = 3 - x^2.$$

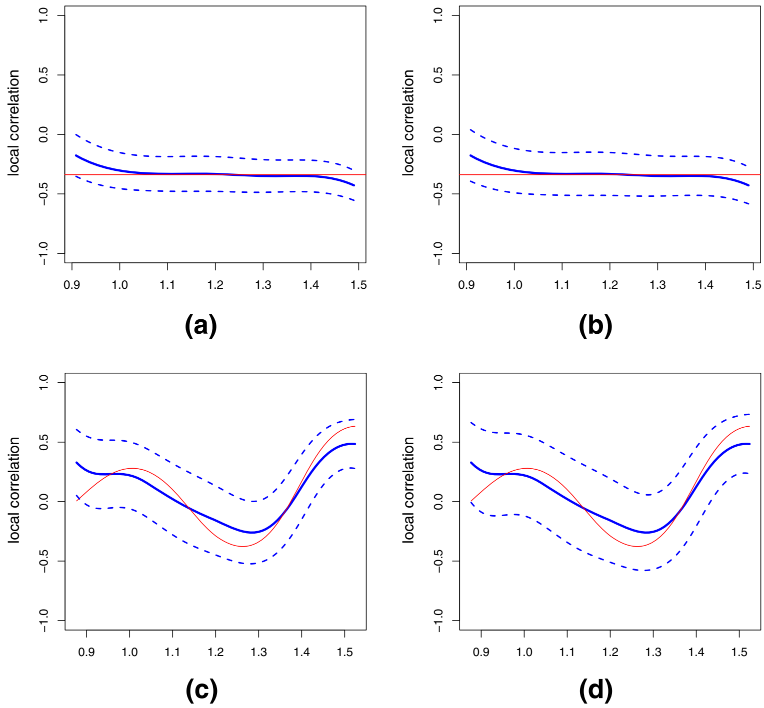

Fig. 1 Plots of oracle SCB for the correlation curve (*dashed*) computed according to (28) in Case 1 and Case 2, the estimator $\hat{\rho}_{\mathrm{LQ}}(x)$ (*thick*), and the true function $\rho\,(x)$ (*thin*). **a** Case 1: $n = 1000$, 95% SCB; **b** Case 1: $n = 1000$, 99% SCB; **c** Case 2: $n = 1000$, 95% SCB; **d** Case 2: $n = 1000$, 99% SCB

Table 1 Coverage frequency of the oracle SCB in (28) and infeasible SCB in (23) from 1000 replications

Case	n	$1-\alpha$	Oracle SCB	Infeasible SCB
Case 1	500	0.950	0.890	0.901
		0.990	0.914	0.927
	1000	0.950	0.940	0.948
		0.990	0.982	0.979
	2000	0.950	0.949	0.952
		0.990	0.990	0.992
Case 2	500	0.950	0.875	0.899
		0.990	0.906	0.919
	1000	0.950	0.939	0.946
		0.990	0.981	0.988
	2000	0.950	0.950	0.952
		0.990	0.989	0.990

Following suggestions of one reviewer, we have set the correlation curve $\rho(x)$ for Case 1 to be a constant, while for Case 2 with more features such as local minima and maxima, see the *thin* curve in each plot of Fig. 1. The sample sizes are taken to be $n = 500, 1000, 2000$ and the confidence levels $1 - \alpha = 0.95, 0.99$. Table 1 contains the coverage frequency from 1000 replications of sample size n of the true correlation curve $\rho(x)$, over 401 equally spaced points from 0.8 to 1.6, by the oracle SCB in (28). Coverage frequency over the same sets of points is also listed in the table for the infeasible SCB in (23). In all cases, the oracle SCB is close to the infeasible SCB in terms of coverage frequency, showing positive confirmation of Theorem 1. The coverage of oracle SCB improves with increasing sample size, approaching the nominal level for sample size as low as $n = 500$, which confirms Theorem 2.

For visual impression, Fig. 1 overlays for a representative sample of size 1000, the 95 and 99% oracle SCBs (*dashed*) computed according to (28), and the estimator $\hat{\rho}_{\mathrm{LQ}}(x)$ (*thick*), together with the true correlation curve $\rho(x)$ (*thin*). One sees clearly that the oracle SCBs are narrow and accurate around the true curve for all cases.

5 Empirical example

In this section, the oracle SCB of (28) is applied to a cross-sectional random sample from the 1995 British Family Expenditure Survey, consisting of the food budget share Y and the logarithm of total expenditure X in 1995 for $n = 1655$ married couples from UK with an employed head-of-household between the ages of 25 and 55 years. The data have been studied in Blundell et al. (2007) to identify the economically meaningful "structural" Engel curve relationship, via the sieve semiparametric IV approach, see https://github.com/JeffreyRacine/R-Package-np/ for detailed information on the British Family Expenditure Survey, and Fig. 2a for a scatter plot of the data $\{(X_i, Y_i)\}_{i=1}^{1655}$.

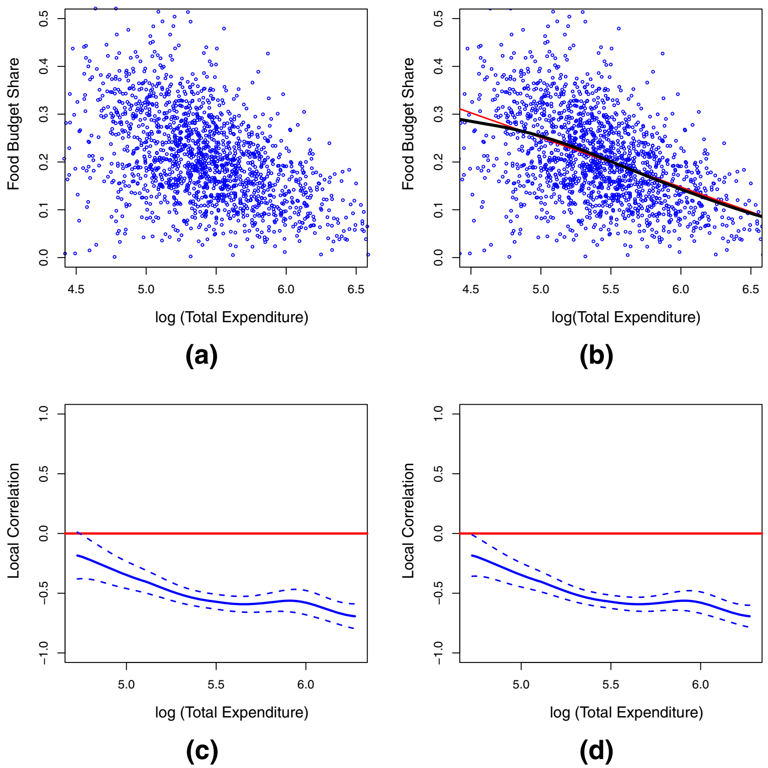

Fig. 2 For the British cross-sectional data, **a** scatterplot consists of $n = 1655$ household-level observations taken from the British Family Expenditure Survey; **b** linear (*solid*) and local quadratic (*thick*) estimators of the local mean function $\mu(x)$; plots of oracle SCB (*dashed*) computed according to (28), the local quadratic estimator $\hat{\rho}_{LQ}(x)$ (*thick*), a constant correlation coefficient (*solid*) which equals 0, **c** 95% SCB; **d** 87.04% SCB

Sample estimate of the Galton–Pearson correlation ρ between X and Y is $\hat{\rho} = -0.4793$, and the null hypothesis $H_0 : \rho = 0$ is rejected in favor of the alternative $H_1 : \rho < 0$ with p value 2×10^{-16}. This supports the Engel's Law (Hamilton 2001) of negative association between food budget share and household real income, see Fig. 2b for linear (*solid*) and local quadratic (*thick*) estimators of the local mean function $\mu(x)$, both demonstrating such significant negative association.

Equipped with the advanced new tool of SCB for local correlation, it would be interesting to examine whether this theory holds locally as well. For this purpose, the null hypothesis is $H_0 : \rho(x) \equiv 0$ vs the alternative $H_1 : \rho(x) < 0$. Figure 2c, d overlays the local quadratic estimator $\hat{\rho}_{LQ}(x)$ (*thick*), the oracle SCBs (*dashed*) for the correlation curve, and a constant correlation 0 (*solid*). Figure 2c shows that the $100(1-0.05)\%$ oracle SCB is not completely below the horizontal zero line, while Fig. 2d depicts that the horizontal zero line is exactly above the maximum of upper dashed line of the $100(1-0.1296)\%$ oracle SCB. One therefore rejects the hypothesis of zero local correlation in favor of negative local correlation with p value

$1 - 0.8704 = 0.1296$. Thus, the local version of Engel's law holds with much less significant evidence.

6 Conclusions

A plug-in estimator is proposed for correlation curve to quantify the hetero-correlaticity in nonparametric regression model, which is shown to be oracally efficient, that is, it uniformly approximates an infeasible estimator at the rate of $o_p\left(n^{-1/2}h_1^{-3/2}\log^{-1/2}n\right)$, much faster than the rate at which the infeasible converges to the true correlation function. A data-driven procedure implements an asymptotically oracle SCB centered around the local correlation estimator, with limiting converge probability which equals to that of the infeasible SCB. As illustrated by a cross-sectional data, the theoretically justified oracle SCB is a useful tool to check the local correlation between a response variable and a covariate and is expected to find wide applications in many scientific disciplines.

Acknowledgements This work has been supported in part by Jiangsu Key-Discipline Program ZY107992, National Natural Science Foundation of China award 11371272, and Research Fund for the Doctoral Program of Higher Education of China award 20133201110002. The authors thank two Reviewers, Editor-in-Chief Ana Militino, Prof. Qin Shao, and participants at the First PKU-Tsinghua Colloquium On Statistics for helpful comments.

Appendix

Throughout this section, for any function $g(u)$, define $\|g\|_\infty = \sup_{u\in\mathcal{I}_n}|g(u)|$. For any vector ξ, one denotes by $\|\xi\|$ the Euclidean norm and $\|\xi\|_\infty$ means the largest absolute value of the elements. We use C to denote any positive constants in the generic sense. A random sequence $\{X_n\}$ "bounded in probability" is denoted as $X_n = \mathcal{O}_p(1)$, while $X_n = o_p(1)$ denotes convergence to 0 in probability. A sequence of random functions which are o_p or $\mathcal{O}_p$ uniformly over $x \in \mathcal{I}_n$ denoted as u_p or U_p.

Next, we state the strong approximation Theorem of Tusnády (1977). It will be used later in the proof of Lemmas 4 and 5.

Let $U_1, \ldots, U_n$ be i.i.d r.v.'s on the 2-dimensional unit square with $\mathbb{P}(U_i < \mathbf{t}) = \lambda(\mathbf{t}), \mathbf{0} \le \mathbf{t} \le \mathbf{1}$, where $\mathbf{t} = (t_1, t_2)$ and $\mathbf{1} = (1, 1)$ are 2-dimensional vectors, $\lambda(\mathbf{t}) = t_1 t_2$. The empirical distribution function $F_n^u(\mathbf{t}) = n^{-1}\sum_{i=1}^n I_{\{U_i<\mathbf{t}\}}$ for $\mathbf{0} \le \mathbf{t} \le \mathbf{1}$.

Lemma 1 *The 2-dimensional Brownian bridge $B(\mathbf{t})$ is defined by $B(\mathbf{t}) = W(\mathbf{t}) - \lambda(\mathbf{t})W(\mathbf{1})$ for $\mathbf{0} \le \mathbf{t} \le \mathbf{1}$, where $W(\mathbf{t})$ is a 2-dimensional Wiener process. Then there is a version $B_n(\mathbf{t})$ of $B(\mathbf{t})$ such that*

$$\mathbb{P}\left[\sup_{\mathbf{0}\le\mathbf{t}\le\mathbf{1}}\left|n^{1/2}\left\{F_n^u(\mathbf{t}) - \lambda(\mathbf{t})\right\} - B_n(\mathbf{t})\right| > n^{-1/2}(C\log n + x)\log n\right] < Ke^{-\lambda x},$$

holds for all x, where C, K, λ are positive constants.

The Rosenblatt transformation for bivariate continuous (X, ε) is

$$\left(X^{*}, \varepsilon^{*}\right) = M\left(X, \varepsilon\right) = \left(F_{X}\left(x\right), F_{\varepsilon|X}\left(\varepsilon\left|x\right.\right)\right), \tag{29}$$

then (X^{*}, ε^{*}) has uniform distribution on $[a, b]^{2}$; therefore,

$$Z_{n}\left\{M^{-1}\left(x^{*}, \varepsilon^{*}\right)\right\} = Z_{n}\left(x, \varepsilon\right) = n^{1/2}\left\{F_{n}\left(x, \varepsilon\right) - F\left(x, \varepsilon\right)\right\},$$

with $F_{n}\left(x, \varepsilon\right)$ denoting the empirical distribution of (X, ε). Lemma 1 implies that there exists a version B_{n} of 2 -dimensional Brownian bridge such that

$$\sup_{x,\varepsilon}\left|Z_{n}\left(x, \varepsilon\right) - B_{n}\left\{M(x, \varepsilon)\right\}\right| = \mathcal{O}_{a.s.}\left(n^{-1/2}\log^{2} n\right). \tag{30}$$

Lemma 2 *Under Assumptions (A2) and (A5), there exists* $\alpha_{1} > 0$ *such that the sequence* $D_{n} = n^{\alpha_{1}}$ *satisfies*

$$n^{-1/2}h_{1}^{-1/2}D_{n}\log^{2} n \to 0, n^{1/2}h_{1}^{1/2}D_{n}^{-(1+\eta_{1})} \to 0,$$
$$\sum\nolimits_{n=1}^{\infty} D_{n}^{-(2+\eta_{1})} < \infty, D_{n}^{-\eta_{1}}h_{1}^{-1/2} \to 0.$$

For such a sequence $\{D_{n}\}$,

$$\mathbb{P}\left\{\omega \mid \exists N\left(\omega\right), \left|\varepsilon_{i}\left(\omega\right)\right| < D_{n}, 1 \le i \le n, n > N\left(\omega\right)\right\} = 1. \tag{31}$$

Lemma 3 *Under Assumptions (A1)–(A5), as* $n \to \infty$,

$$\begin{aligned}\tilde{\rho}_{\mathrm{LQ}}\left(x\right) - \rho(x) &= \sigma_{1}\left\{1 - \rho^{2}\left(x\right)\right\}^{3/2}\sigma^{-1}\left(x\right)\mu_{3}(K^{*})\beta''\left(x\right)h_{1}^{2}/6\\ &\quad + \sigma_{1}\left\{1 - \rho^{2}\left(x\right)\right\}^{3/2}\sigma^{-1}\left(x\right)n^{-1}h_{1}^{-1}f^{-1}\left(x\right)\\ &\quad \times \sum\nolimits_{i=1}^{n} K_{h_{1}}^{*}(X_{i} - x)\sigma\left(X_{i}\right)\varepsilon_{i}\\ &\quad + u_{p}\left(h_{1}^{2} + n^{-1/2}h_{1}^{-3/2}\log^{-1/2} n\right).\end{aligned} \tag{32}$$

Proof From the definition of $\tilde{\rho}_{\mathrm{LQ}}(x)$ in (15), the Taylor series expansions, and $\hat{\beta}\left(x\right) - \beta\left(x\right) = U_{p}\left(n^{-1/2}h_{1}^{-3/2}\log^{1/2} n + h_{1}^{2}\right)$, one has

$$\begin{aligned}\tilde{\rho}_{\mathrm{LQ}}\left(x\right) - \rho(x) &= \sigma_{1}\left\{1 - \rho^{2}\left(x\right)\right\}^{3/2}\sigma^{-1}\left(x\right)\left\{\hat{\beta}\left(x\right) - \beta\left(x\right)\right\}\\ &\quad + U_{p}\left(n^{-1}h_{1}^{-3}\log n + h_{1}^{4}\right).\end{aligned} \tag{33}$$

Write $\mathbf{Y}$ as the sum of a signal vector $\boldsymbol{\mu} = \{\mu(X_1), \ldots, \mu(X_n)\}^T$ and a noise vector $\mathbf{E} = \{\sigma(X_1)\varepsilon_1, \ldots, \sigma(X_n)\varepsilon_n\}^T$,

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{E}. \tag{34}$$

The local quadratic estimator $\hat{\beta}(x)$ has a noise and bias error decomposition

$$\hat{\beta}(x) - \beta(x) = I(x) + II(x),$$

in which the bias term $I(x)$ and noise term $II(x)$ are

$$\begin{aligned} I(x) &= e_1^T(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W} \\ &\quad\times\left\{\boldsymbol{\mu} - \mu(x)\mathbf{X}e_0 - \beta(x)\mathbf{X}e_1 - \beta'(x)/2\mathbf{X}e_2\right\}, \quad (35) \\ II(x) &= e_1^T(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{E}. \quad (36) \end{aligned}$$

where e_k, $k = 0, 1, 2$, as defined in (8), $\mathbf{X}$ in (9), $\mathbf{W}$ in (10), $\boldsymbol{\mu}$ and $\mathbf{E}$ in (34). Standard arguments from kernel smoothing theory yield that

$$I(x) = \mu_3(K^*)\beta''(x)h_1^2/6 + u_p(h_1^2), \tag{37}$$

in which $\mu_3(K^*) = \int v^3K^*(v)\,\mathrm{d}v$. Likewise,

$$\begin{aligned} II(x) &= n^{-1}h_1^{-1}f^{-1}(x)\sum\nolimits_{i=1}^{n}K_{h_1}^*(X_i - x)\sigma(X_i)\varepsilon_i\left\{1 + u_p\left(\log^{-1}n\right)\right\} \\ &= n^{-1}h_1^{-1}f^{-1}(x)\sum\nolimits_{i=1}^{n}K_{h_1}^*(X_i - x)\sigma(X_i)\varepsilon_i \\ &\quad + u_p\left(n^{-1/2}h_1^{-3/2}\log^{-1/2}n\right). \quad (38) \end{aligned}$$

Putting together (33), (37) and (38) completes the proof of the lemma.

Now from Lemma 3, one can rewrite (32) as

$$\begin{aligned} \tilde{\rho}_{\mathrm{LQ}}(x) - \rho(x) &= \sigma_1\left\{1 - \rho^2(x)\right\}^{3/2}\sigma^{-1}(x)\mu_3(K^*)\beta''(x)h_1^2/6 \\ &\quad + \sigma_1\left\{1 - \rho^2(x)\right\}^{3/2}f^{-1/2}(x)n^{-1/2}h_1^{-3/2}Y(x) \\ &\quad + u_p\left(n^{-1/2}h_1^{-3/2}\log^{-1/2}n + h_1^2\right), \quad (39) \end{aligned}$$

in which the process

$$Y(x) = h_1^{1/2}\sigma^{-1}(x)f^{-1/2}(x)n^{-1/2}\sum\nolimits_{i=1}^{n}K_{h_1}^*(X_i - x)\sigma(X_i)\varepsilon_i, x \in \mathcal{I}_n. \tag{40}$$

Define next four stochastic processes, which approximate each other in probability uniformly over $\mathcal{I}_n$ or have the exact same distributions over $\mathcal{I}_n$. More precisely, with

D_n defined in Lemma 2, and B_n in Lemma 1, let

$$Y_0(x) = h_1^{1/2}\sigma^{-1}(x) f^{-1/2}(x) \int_{\mathbb{R}} K_{h_1}^*(u-x)\varepsilon I_{\{|\varepsilon|\le D_n\}} \mathrm{d}B_n\{M(u,\varepsilon)\}, \quad (41)$$

$$Y_1(x) = h_1^{1/2}\sigma^{-1}(x) f^{-1/2}(x) \int_{\mathbb{R}} K_{h_1}^*(u-x)\varepsilon I_{\{|\varepsilon|\le D_n\}} \mathrm{d}W_n\{M(u,\varepsilon)\}, \quad (42)$$

$$Y_2(x) = h_1^{1/2}\sigma^{-1}(x) f^{-1/2}(x) \int_{\mathbb{R}} K_{h_1}^*(u-x) f^{1/2}(u)\sigma(u) s_n(u)\, \mathrm{d}W_n(u), \quad (43)$$

where

$$s_n^2(u) = \int_{\mathbb{R}} \varepsilon^2 I_{\{|\varepsilon|\le D_n\}} \mathrm{d}F(\varepsilon\,|u),$$

and satisfies that

$$\sup_{u\in\mathcal{I}_n} \left| s_n^2(u) - 1\right| = \sup_{u\in\mathcal{I}_n} \int_{\mathbb{R}} \varepsilon^2 I_{\{|\varepsilon|> D_n\}} \mathrm{d}F(\varepsilon\,|u) \le M_{\eta_1} D_n^{-\eta_1}, \quad (44)$$

$$Y_3(x) = h_1^{1/2} \int_{\mathbb{R}} K_{h_1}^*(u-x)\, \mathrm{d}W_n(u). \quad (45)$$

Lemma 4 *Under Assumptions (A2)–(A5), as $n \to \infty$,*

$$\sup_{x\in\mathcal{I}_n} \left| Y(x) - Y^D(x)\right| = \mathcal{O}_p\left(n^{1/2} h_1^{1/2} D_n^{-1-\eta_1}\right),$$

where, for $x \in \mathcal{I}_n$,

$$Y^D(x) = h_1^{1/2}\sigma^{-1}(x) f^{-1/2}(x)\, n^{-1/2} \times \sum_{i=1}^{n} K_{h_1}^*(X_i - x)\sigma(X_i)\,\varepsilon_i I_{\{|\varepsilon|\le D_n\}}. \quad (46)$$

Proof Using notations from Lemma 1, the processes $Y(x)$ defined in (40) and $Y^D(x)$ can be written as

$$Y(x) = h_1^{1/2}\sigma^{-1}(x) f^{-1/2}(x) \iint K_{h_1}^*(u-x)\sigma(u)\,\varepsilon \mathrm{d}Z_n(u,\varepsilon),$$

$$Y^D(x) = h_1^{1/2}\sigma^{-1}(x) f^{-1/2}(x) \iint K_{h_1}^*(u-x)\sigma(u)\,\varepsilon I_{\{|\varepsilon|\le D_n\}} \mathrm{d}Z_n(u,\varepsilon).$$

The tail part $Y(x) - Y^D(x)$ is bounded uniformly over $\mathcal{I}_n$ by

$$\sup_{x\in\mathcal{I}_n} h_1^{1/2}\sigma^{-1}(x) f^{-1/2}(x) \left|\iint K_{h_1}^*(u-x)\sigma(u)\,\varepsilon I_{\{|\varepsilon|>D_n\}} \mathrm{d}Z_n(u,\varepsilon)\right|$$
$$\le \sup_{x\in\mathcal{I}_n} h_1^{1/2}\sigma^{-1}(x) f^{-1/2}(x)\, n^{-1/2}$$
$$\times \left|\sum\nolimits_{i=1}^{n} K_{h_1}^*(X_i - x)\,\sigma(X_i)\,\varepsilon_i I_{\{|\varepsilon_i|>D_n\}}\right| \tag{47}$$
$$+ \sup_{x\in\mathcal{I}_n} n^{1/2} h_1^{1/2}\sigma^{-1}(x) f^{-1/2}(x)$$
$$\times \left|\iint K_{h_1}^*(u-x)\sigma(u)\,\varepsilon I_{\{|\varepsilon|>D_n\}} \mathrm{d}F(u,\varepsilon)\right|. \tag{48}$$

By (31) in Lemma 2 and Borel–Cantelli Lemma, the first term in Eq. (47) is $\mathcal{O}_{a.s.}\left(n^{-a}\right)$ for any $a > 0$, for instance $a = 100$, and the second term in Eq. (48) is bounded by

$$\sup_{x\in\mathcal{I}_n} n^{1/2} h_1^{1/2} f^{-1/2}(x)\,\sigma^{-1}(x)$$
$$\times \int \left|K_{h_1}^*(u-x)\right| \sigma(u)\, f(u) \left[\int |\varepsilon|\, I_{\{|\varepsilon|>D_n\}} \mathrm{d}F(\varepsilon\,|u)\right] \mathrm{d}u$$
$$\le \sup_{x\in\mathcal{I}_n} n^{1/2} h_1^{1/2} f^{-1/2}(x)\,\sigma^{-1}(x) M_{\eta_1} D_n^{-(1+\eta_1)} \int \left|K_{h_1}^*(u-x)\right| \sigma(u)\, f(u)\,\mathrm{d}u$$
$$\le C n^{1/2} h_1^{1/2} D_n^{-1-\eta_1} = \mathcal{O}\left(n^{1/2} h_1^{1/2} D_n^{-1-\eta_1}\right).$$

Lemma 5 *Under Assumptions (A2)–(A5), as* $n \to \infty$,

$$\sup_{x\in\mathcal{I}_n} \left|Y^D(x) - Y_0(x)\right| = \mathcal{O}_p\left(n^{-1/2} h_1^{-1/2} D_n \log^2 n\right).$$

Proof First, $\left|Y^D(x) - Y_0(x)\right|$ can be written as

$$h_1^{1/2}\sigma^{-1}(x) f^{-1/2}(x) \iint K_{h_1}^*(u-x)\sigma(u)\,\varepsilon I_{\{|\varepsilon|\le D_n\}} d\left[Z_n(u,\varepsilon) - B_n\{M(u,\varepsilon)\}\right],$$

which becomes the following via integration by parts

$$h_1^{1/2}\sigma^{-1}(x) f^{-1/2}(x)$$
$$\times \iint \sigma(u)\left[Z_n(u,\varepsilon) - B_n\{M(u,\varepsilon)\}\right] d\left\{\varepsilon I_{\{|\varepsilon|\le D_n\}}\right\} d\left\{K_{h_1}^*(u-x)\right\}.$$

Next, from the strong approximation result in Eq. (30) and the first condition in Lemma 2, $\sup_{x\in\mathcal{I}_n}\left|Y^D(x)-Y_0(x)\right|$ is bounded by

$$\mathcal{O}_{a.s.}\left(h_1^{1/2}h_1^{-2}n^{-1/2}h_1D_n\log^2 n\right)=\mathcal{O}_{a.s.}\left(n^{-1/2}h_1^{-1/2}D_n\log^2 n\right),$$

thus completing the proof of the lemma.

Lemma 6 *Under Assumptions (A2)–(A5), as $n\to\infty$,*

$$\sup_{x\in\mathcal{I}_n}\left|Y_0(x)-Y_1(x)\right|=\mathcal{O}_p\left(h_1^{1/2}\right).$$

Proof Based on Rosenblatt transformation $M(x,\varepsilon)$ defined in Eq. (29) and according to Lemma 2, the term $|Y_0(x)-Y_1(x)|$ is bounded by

$$\begin{aligned}&\sup_{x\in\mathcal{I}_n}h_1^{1/2}\sigma^{-1}(x)f^{-1/2}(x)\\&\quad\times\left|\iint K_{h_1}^*(u-x)\,\sigma(u)\,|\varepsilon|\,I_{\{|\varepsilon|\le D_n\}}\mathrm{d}M(u,\varepsilon)W_n(1,1)\right|\\&\le\sup_{x\in\mathcal{I}_n}h_1^{1/2}\sigma^{-1}(x)f^{-1/2}(x)\left|W_n(1,1)\right|\\&\quad\times\int\left|K_{h_1}^*(u-x)\right|\sigma(u)\,f(u)\,\mathrm{d}u\left\{\int|\varepsilon|\,I_{\{|\varepsilon|\le D_n\}}\mathrm{d}F(\varepsilon\,|u)\right\}=\mathcal{O}_p\left(h_1^{1/2}\right).\end{aligned}$$

The next lemma expresses the distribution of $Y_1(x)$ in terms of one-dimensional Brownian motion.

Lemma 7 *The process $Y_1(x)$ has the same distribution as $Y_2(x)$ over $x\in\mathcal{I}_n$.*

Proof By definitions, $Y_1(x)$ defined in (42) and $Y_2(x)$ in (43) are Gaussian processes with zero mean and unit variance. They have the same covariance functions as

$$\begin{aligned}\mathrm{cov}\left\{Y_1(x),Y_1\left(x'\right)\right\}&=h_1^{1/2}\sigma^{-1}(x)f^{-1/2}(x)\,h_1^{1/2}\sigma^{-1}(x')f^{-1/2}\left(x'\right)\\&\quad\times\int K_{h_1}^*(u-x)K_{h_1}^*(u-x')f(u)\sigma^2(u)s_n^2(u)\,\mathrm{d}u\\&=\mathrm{cov}\left\{Y_2(x),Y_2\left(x'\right)\right\}.\end{aligned}$$

Hence, according to Itô's Isometry Theorem, they have the same distribution.

Lemma 8 *Under Assumptions (A2)–(A5), as $n\to\infty$,*

$$\sup_{x\in\mathcal{I}_n}\left|Y_2(x)-Y_3(x)\right|=\mathcal{O}_p\left(h_1^{1/2}+h_1^{-1/2}D_n^{-\eta_1}\right).$$

Proof By the aforementioned condition in Lemma 2 and Eq. (44), $\sup_{x\in\mathcal{I}_n}|Y_2(x) - Y_3(x)|$ is almost surely bounded by

$$\begin{aligned}
&\sup_{x\in\mathcal{I}_n}|W_n(u)|\, h_1^{1/2}\left|\int d\left[K^*_{h_1}(u-x)\left[\left\{\frac{f(u)}{f(x)}\right\}^{1/2}\left\{\frac{\sigma(u)}{\sigma(x)}\right\}s_n(u)-1\right]\right]\right|\\
&\quad\le \sup_{x\in\mathcal{I}_n}|W_n(u)|\, h_1^{1/2}h_1^{-1}\\
&\qquad\times\int h_1^{-1}\left|K^{*\prime}\left(\frac{u-x}{h_1}\right)\right|\left[\left\{\frac{f(u)}{f(x)}\right\}^{1/2}\left\{\frac{\sigma(u)}{\sigma(x)}\right\}s_n(u)-1\right]\\
&\qquad+\left|K^*\left(\frac{u-x}{h_1}\right)\right|\left[\left\{\frac{f(u)}{f(x)}\right\}^{1/2}\left\{\frac{\sigma(u)}{\sigma(x)}\right\}s_n(u)-1\right]'\mathrm{d}u\\
&\quad=\mathcal{O}_p\left(h_1^{-1/2}\right)\{III(x)+IV(x)\},
\end{aligned}$$

where the term $III(x)$ is bounded by

$$\begin{aligned}
&\sup_{x\in\mathcal{I}_n} Ch_1^{-1}\left\|K^*\right\|_\infty f^{-1/2}(x)\,\sigma^{-1}(x)\,h_1\\
&\qquad\times\left|\left[h_1/2\left\{f'(x)\right\}^{1/2}\sigma(u)\,s_n(u)+f^{1/2}(x)\left\{\sigma(u)\,s_n(u)-\sigma(x)\right\}\right]\right|\\
&\quad\le C\left\|K^*\right\|_\infty C_f^{-1/2}C_\sigma^{-1}\left\{h_1/2+\left\|s_n^2-1\right\|_\infty\right\}\\
&\quad\le C\left(h_1+D_n^{-\eta_1}\right),
\end{aligned}$$

and the term $IV(x)$ is bounded by

$$\begin{aligned}
&Ch_1ch_1 2^{-1}f'(u)\,f^{-1/2}(u)\,f^{-1/2}(x)\,\sigma(u)\,\sigma^{-1}(x)\,s_n(u)\\
&\qquad+Ch_1f^{1/2}(u)\,f^{-1/2}(x)\left\{\sigma'(u)\,\sigma^{-1}(x)\,s_n(u)+\sigma(u)\,s_n'(u)/\sigma(x)\right\}\\
&\quad\le Ch_1\left(2^{-1}\left\|f'\right\|_\infty C_f\left\|s_n\right\|_\infty+\left\|\sigma'\right\|_\infty C_\sigma^{-1}\left\|s_n\right\|_\infty+\left\|s_n'\right\|_\infty\right)\\
&\quad\le Ch_1.
\end{aligned}$$

Putting together the above, one obtains that

$$\begin{aligned}
\sup_{x\in\mathcal{I}_n}|Y_2(x)-Y_3(x)| &= \mathcal{O}_p\left(h_1^{-1/2}\right)\{III(x)+IV(x)\}\\
&= \mathcal{O}_p\left(h_1^{1/2}+h_1^{-1/2}D_n^{-\eta_1}\right)+\mathcal{O}_p\left(h_1^{1/2}\right),
\end{aligned}$$

completing the proof of this lemma.

Proof of Proposition 1 The absolute maximum of $\{Y_3(x), x \in \mathcal{I}_n\}$ is the same as that of

$$\left\{h_1^{-1/2} \int K^*\left(\frac{u}{h_1} - x\right) \mathrm{d}W_n(u), x \in \left[ah_1^{-1} + 1, bh_1^{-1} - 1\right]\right\}$$
$$= \left\{\int K^*(v - x) \mathrm{d}W_n(v), x \in \left[ah_1^{-1} + 1, bh_1^{-1} - 1\right]\right\}. \tag{49}$$

For process $\xi(x) = \int K^*(v-x)\mathrm{d}W_n(v), x \in \left[ah_1^{-1} + 1, bh_1^{-1} - 1\right]$, the correlation function is

$$r(x-y) = \frac{\mathsf{E}\{\xi(x)\xi(y)\}}{\mathrm{var}^{1/2}\{\xi(x)\}\mathrm{var}^{1/2}\{\xi(y)\}},$$

which implies that

$$r(t) = \frac{\int K^*(v) K^*(v-t) \mathrm{d}v}{\int K^*(v)^2 \mathrm{d}v}.$$

Define next a Gaussian process $\varsigma(t), 0 \le t \le T = T_n = (b-a)/h_1 - 2$,

$$\varsigma(t) = \xi\left(t + ah_1^{-1} + 1\right)\left\{\int K^*(v)^2 \mathrm{d}v\right\}^{-1/2},$$

which is stationary with mean zero and variance one, and covariance function

$$r(t) = \mathsf{E}\varsigma(s)\varsigma(t+s) = 1 - Ct^2 + o\left(|t|^2\right) \text{as} t \to 0,$$

with $C = C_{K^{*\prime}}/2C_{K^*}$. Then applying Theorems 11.1.5 and 12.3.5 of Leadbetter et al. (1983), one has for $h_1 \to 0$ or $T \to \infty$,

$$\mathbb{P}\left[a_T\left\{\sup_{t\in[0,T]}|\varsigma(t)| - b_T\right\} \le z\right] \to e^{-2e^{-z}}, \quad \forall z \in \mathbb{R},$$

where $a_T = (2\log T)^{1/2}$ and $b_T = a_T + a_T^{-1}\left\{\sqrt{C(K^*)}/2\pi\right\}$. Note that for a_{h_1}, b_{h_1} defined in (21), as $n \to \infty$,

$$a_{h_1}a_T^{-1} \to 1, a_T\left(b_T - b_{h_1}\right) = \mathcal{O}\left(\log^{1/2} n \times h_1 \log^{-1/2} n\right) \to 0.$$

Hence, applying Slutsky's Theorem twice, one obtains that

$$a_{h_1}\left\{\sup_{t\in[0,T]}|\varsigma(t)| - b_{h_1}\right\} = a_{h_1}a_T^{-1}\left[a_T\left\{\sup_{t\in[0,T]}|\varsigma(t)| - b_T\right\}\right] + a_{h_1}\left(b_T - b_{h_1}\right)$$

converges in distribution to the same limit as $a_T\left\{\sup_{t\in[0,T]}|\varsigma(t)| - b_T\right\}$. Thus,

$$\mathbb{P}\left(a_{h_1}\left[\frac{\sup_{x\in\mathcal{I}_n}|Y_3(x)|}{\left\{\int K^*(v)^2 \mathrm{d}v\right\}^{1/2}} - b_{h_1}\right] \le z\right) \to e^{-2e^{-z}}, \quad \forall z \in \mathbb{R}.$$

Next applying Lemma 8 and Slutsky's Theorem, $\forall z \in \mathbb{R}$,

$$\mathbb{P}\left(a_{h_1}\left[\frac{\sup_{x\in\mathcal{I}_n}|Y_2(x)|}{\left\{\int K^*(v)^2\,\mathrm{d}v\right\}^{1/2}}-b_{h_1}\right]\leq z\right)\to e^{-2e^{-z}}. \tag{50}$$

Furthermore, applying Lemma 7 and Slutsky's Theorem, the limiting distribution (50) is the same as

$$\mathbb{P}\left(a_{h_1}\left[\frac{\sup_{x\in\mathcal{I}_n}|Y_1(x)|}{\left\{\int K^*(v)^2\,\mathrm{d}v\right\}^{1/2}}-b_{h_1}\right]\leq z\right)\to e^{-2e^{-z}}.$$

Furthermore, applying Lemmas 1 to 6 and Slutsky's Theorem, one obtains

$$\mathbb{P}\left(a_{h_1}\left[\frac{\sup_{x\in\mathcal{I}_n}|Y(x)|}{\left\{\int K^*(v)^2\,\mathrm{d}v\right\}^{1/2}}-b_{h_1}\right]\leq z\right)\to e^{-2e^{-z}}. \tag{51}$$

By taking $1-\alpha=e^{-2e^{-z}}$ for $\alpha\in(0,1)$, the above (51) implies that

$$\lim_{n\to\infty}\mathbb{P}\left\{\rho(x)\in\tilde{\rho}_{\mathrm{LQ}}(x)\pm a_{h_1}V_n(x)Q_n(\alpha),\,x\in\mathcal{I}_n\right\}=1-\alpha.$$

Thus, an infeasible SCB for $\rho(x)$ over $\mathcal{I}_n$ is

$$\tilde{\rho}_{\mathrm{LQ}}(x)\pm a_{h_1}V_n(x)Q_n(\alpha),$$

which establishes Proposition 1.

Proof of Theorem 1 Applying Taylor expansion to $\hat{\rho}_{\mathrm{LQ}}(x)-\tilde{\rho}_{\mathrm{LQ}}(x)$, its asymptotic order is the lower of $\hat{\sigma}_1^2-\sigma_1^2$ and $\hat{\sigma}^2_{\mathrm{SK}}(x)-\sigma^2(x)$. While $\hat{\sigma}_1^2-\sigma_1^2=\mathcal{O}_p\left(n^{-1/2}\right)$, $\sup_{x\in[a+h_2,b-h_2]}\left|\hat{\sigma}^2_{\mathrm{SK}}(x)-\sigma^2(x)\right|$ is of order $\mathcal{O}_p\left(n^{-1/2}h_2^{-1/2}\log^{1/2}n\right)$ according to Cai and Yang (2015), and of order $o_p\left(n^{-1/2}h_1^{-3/2}\log^{-1/2}n\right)$ by applying (16). As (17) entails that $\mathcal{I}_n\subset[a+h_2,b-h_2]$ for large enough n, one has

$$\sup_{x\in\mathcal{I}_n}\left|\hat{\sigma}^2_{\mathrm{SK}}(x)-\sigma^2(x)\right|=o_p\left(n^{-1/2}h_1^{-3/2}\log^{-1/2}n\right), \tag{52}$$

and thus $\sup_{x\in\mathcal{I}_n}\left|\hat{\rho}_{\mathrm{LQ}}(x)-\tilde{\rho}_{\mathrm{LQ}}(x)\right|=o_p\left(n^{-1/2}h_1^{-3/2}\log^{-1/2}n\right)$. Hence, the proof of the theorem is complete.

Proof of Theorem 2 Proposition 1, Theorem 1, and repeated applications of Slutsky's Theorem entail that

$$\lim_{n\to\infty}\mathbb{P}\left\{\rho(x)\in\hat{\rho}_{\mathrm{LQ}}(x)\pm a_{h_1}\hat{V}_n(x)Q_n(\alpha),\,x\in\mathcal{I}_n\right\}=1-\alpha,$$

which yields the oracle SCB for $\rho(x)$ over $\mathcal{I}_n$ in Theorem 2.

References

Aerts M, Claeskens G (1997) Local polynomial estimators in multiparameter likelihood models. J Am Stat Assoc 92:1536–1545
Bickel PJ, Rosenblatt M (1973) On some global measures of deviations of density function estimates. Ann Stat 31:1852–1884
Blundell R, Chen X, Kristensen D (2007) Semi-nonparametric IV estimation of shape invariant Engel curves. Econometrica 75:1613–1670
Cao G, Wang L, Li Y, Yang L (2016) Oracle-efficient confidence envelopes for covariance functions in dense functional data. Stat Sin 26:359–383
Cao G, Yang L, Todem D (2012) Simultaneous inference for the mean function based on dense functional data. J Nonparametr Stat 24:359–377
Cai L, Yang L (2015) A smooth confidence band for variance function. TEST 24:632–655
Carroll RJ, Ruppert D (1988) Transformations and weighting in regression. Champman and Hall, London
Claeskens G, Van Keilegom I (2003) Bootstrap confidence bands for regression curves and their derivatives. Ann Stat 31:1852–1884
Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 74:829–836
de Boor C (2001) A practical guide to splines. Springer, New York
Doksum K, Blyth S, Bradlow E, Meng X, Zhao H (1994) Correlation curves as local measures of variance explained by regression. J Am Stat Assoc 89:571–582
Eubank RL, Speckman PL (1993) Confidence bands in nonparametric regression. J Am Stat Assoc 88:1287–1301
Fan J (1993) Local linear regression smoothers and their minimax efficiency. Ann Stat 21:196–216
Fan J, Gijbels I (1996) Local polynomial modelling and its applications. Champman and Hall, London
Fan J, Yao Q (1998) Efficient estimation of conditional variance functions in stochastic regression. Biometrika 85:645–660
Gasser T, Müller HG, Mammitzsch V (1984) Estimating regression functions and their derivatives with kernel methods. Scand J stat 11:171–185
Gu L, Wang L, Härdle W, Yang L (2014) A simultaneous confidence corridor for varying coefficient regression with sparse functional data. TEST 23:806–843
Gu L, Yang L (2015) Oracally efficient estimation for single-index link function with simultaneous band. Electron J Stat 9:1540–1561
Hall P (1991) Edgeworth expansions for nonparametric density estimators, with applications. Statistics 22:215–232
Hall P (1992) Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. Ann Stat 20:675–694
Hall P, Titterington MD (1988) On confidence bands in nonparametric density estimation and regression. J Multivar Anal 27:228–254
Hamilton BW (2001) Using Engel's law to estimate CPI bias. Am Econ Rev 91:619–630
Härdle W (1989) Asymptotic maximal deviation of M-smoothers. J Multivar Anal 29:163–179
Leadbetter MR, Lindgren G, Rootzén H (1983) Extremes and related properties of random sequences and processes. Springer, New York
Ma S, Yang L, Carroll R (2012) A simultaneous confidence band for sparse longitudinal regression. Stat Sin 22:95–122
R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
Ruppert D, Wand MP (1994) Multivariate weighted least squares regression. Ann Stat 22:1346–1370
Silverman WB (1986) Density estimation for statistics and data analysis. Chapman and Hall, London
Song Q, Liu R, Shao Q, Yang L (2014) A simultaneous confidence band for dense longitudinal regression. Commun Stat Theory Methods 43:5195–5210
Song Q, Yang L (2009) Spline confidence bands for variance functions. J Nonparametr Stat 5:589–609
Stapleton J (2009) Linear statistical models, 2nd edn. Wiley, Hoboken
Stone CJ (1977) Consistent nonparametric regression. Ann Stat 5:595–645
Stone CJ (1980) Optimal rates of convergence for nonparametric estimators. Ann Stat 8:1348–1360
Tusnády G (1977) A remark on the approximation of the sample DF in the multidimensional case. Period Math Hung 8:53–55

Wang J, Cheng F, Yang L (2013) Smooth simultaneous confidence bands for cumulative distribution functions. J Nonparametr Stat 25:395–407
Wang J, Liu R, Cheng F, Yang L (2014) Oracally efficient estimation of autoregressive error distribution with simultaneous confidence band. Ann Stat 42:654–668
Wang J, Wang S, Yang L (2016) Simultaneous confidence bands for the distribution function of a finite population and its superpopulation. TEST 25:692–709
Wang J, Yang L (2009) Polynomial spline confidence bands for regression curves. Stat Sin 19:325–342
Wu W, Zhao Z (2007) Inference of trends in time series. J R Stat Soc Ser B 69:391–410
Xia Y (1998) Bias-corrected confidence bands in nonparametric regression. J R Stat Soc Ser B 60:797–811
Zhao Z, Wu W (2008) Confidence bands in nonparametric time series regression. Ann Stat 36:1854–1878
Zheng S, Liu R, Yang L, Härdle W (2016) Statistical inference for generalized additive models: simultaneous confidence corridors and variable selection. TEST 25:607–626
Zheng S, Yang L, Härdle W (2014) A smooth simultaneous confidence corridor for the mean of sparse functional data. J Am Stat Assoc 109:661–673