SFB 649 Discussion Paper 2011-002

A Confidence Corridor for Sparse Longitudinal Data Curves

Shuzhuan Zheng*
Lijian Yang*
Wolfgang Karl Härdle**

*Michigan State University, USA
** Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

SFB 649 ECONOMIC RISK BERLIN

Submitted to the Annals of Statistics

A CONFIDENCE CORRIDOR FOR SPARSE LONGITUDINAL DATA CURVES*

By Shuzhuan Zheng[1], Lijian Yang[2,1] and Wolfgang K. Härdle[3,4]

[1]*Michigan State University,* [2]*Soochow University,* [3]*Humboldt-Universität zu Berlin and* [4]*National Central University*

Longitudinal data analysis is a central piece of statistics. The data are curves and they are observed at random locations. This makes the construction of a simultaneous confidence corridor (SCC) (confidence band) for the mean function a challenging task on both the theoretical and the practical side. Here we propose a method based on local linear smoothing that is implemented in the sparse (i.e., low number of nonzero coefficients) modelling situation. An SCC is constructed based on recent results obtained in applied probability theory. The precision and performance is demonstrated in a spectrum of simulations and applied to growth curve data. Technically speaking, our paper intensively uses recent insights into extreme value theory that are also employed to construct a shoal of confidence intervals (SCI).

1. Introduction. Longitudinal or functional data analysis (FDA) is a central piece of statistical modelling. A well known application is growth curve analysis in biology, medicine and chemistry, see e.g. Müller (2009), James, Hastie and Sugar (2000), Ferraty and Vieu (2006) and the references there. Groundbreaking theoretical work on functional data analysis has been done among others by

*This research was supported by the Deutsche Forschungsgemeinschaft through the CRC 649 "Economic Risk", NSF Awards DMS 0706518, DMS 1007594, an MSU Dissertation Continuation Fellowship, and a grant from Risk Management Institute, National University of Singapore.

AMS 2000 Classification: Primary 62G08; secondary 62G15

JEL Classification: C14, C33

Keywords and phrases: Longitudinal data, confidence band, Karhunen-Loève L^2 representation, local linear estimator, extreme value, double sum, strong approximation.

Cai and Hall (2006), Cardot, Ferraty and Sarda (2003). Much of this work though is devoted to coefficient estimation, semiparametric analysis or dimension reduction methods. Research on statistical inference on the mean curve for example is rather scarce although it is potentially important for characterization of global properties. To characterize global properties of the unknown function of interest, the simultaneous confidence corridor (SCC) and the shoal of confidence intervals (SCI) are puissant instruments. They can be applied to test the overall trend or shape of the mean function. Such decisions are critical e.g. in ozone analysis, see Lucas and Diggle (1997) for a longitudinal study on Sitka spruce. They have pointed out that, in order to assess the cumulative effect of ozone pollution on spruce, an inference on the mean function of spruce growth during the entire experiment rather than at the end of the growth is required. This is one of the many other motivations to develop a new method and its theory to construct an SCC for the mean function of sparse longitudinal data where the measurements are randomly located with random repetitions.

The SCC methodology has been extensively studied in the literature. For the nonparametric regression, see Fan and Zhang (2000) and references there. In this strand of literature though it is not assumed that for family of curves one needs to take care of dependence structures. Wu and Zhao (2007) recently constructed a confidence band for the non-stationary mean function, and Wang and Yang (2009), Song and Yang (2009) obtained the spline-based analogy for the mean and variance functions. Nonparametric time series with specific dependence structures are considered in Zhao and Wu (2008). An SCC construction for longitudinal data remains however an open problem.

The major difficulty to construct the SCC for longitudinal data is that the observations within subject are dependent. In this situation, the "Hungarian embedding", used to construct confidence bands is no longer applicable. The sparse longitudinal data situation has been considered by Yao et al. (2005a) for individual trajectories instead of the mean function, while Yao (2007) obtained an SCI for the mean and covariance functions. Ma et al. (2010) constructed the first SCC of the mean function for the sparse longitudinal data through piecewise constant spline. The constructed SCC, however, is nonsmooth and its convergence rate to the true mean function has suboptimal rate.

Here we propose to construct the SCC for the mean function of the sparse longitudinal data via local linear smoothing. We tackle with this research a variety of interesting issues. First, the proposed SCC allows for the global rather than pointwise inference. Second, the sparse rather than dense longitudinal data setting requires more sophisticated extreme value theory. Third, compared to the piecewise constant spline method of Ma et al. (2010), different extreme value results are employed for a local linear estimator that leads to higher accuracy, better coverage, smooth mean curve and smooth SCC, all of which are desirable in the application.

We organize our paper as follows. In Section 2, we state our model and local linear smoothing methodology. In Section 3, we investigate the asymptotic distribution of the maximal deviation of the local linear estimator from the true mean function, which is used to construct the SCC. Section 4 outlines the key procedures to implement the SCC. Section 5 illustrates the performance of the SCC through extensive simulations followed by an empirical example in Section 6 which illustrates the SCC application on growth curve data. Technical proofs are presented in the Appendix.

2. Model and Methodology. Longitudinal data has the form of $\{X_{ij}, Y_{ij}\}, 1 \leq j \leq N_i, 1 \leq i \leq n$, in which $X_{ij} \in \mathcal{X} = [a, b]$ is the j-th random time point for the i-th subject and Y_{ij} is the response measured at X_{ij}. For the i-th subject, the sample path is the noisy realization of a continuous time stochastic process $\xi_i(x)$, namely,

$$Y_{ij} = \xi_i(X_{ij}) + \sigma(X_{ij})\varepsilon_{ij}, \tag{2.1}$$

where the errors ε_{ij} are i.i.d. with $\mathsf{E}\,\varepsilon_{ij} = 0, \mathsf{E}\,\varepsilon_{ij}^2 = 1$, and $\{\xi_i(x), x \in \mathcal{X}\}$ are i.i.d. copies of the process $\{\xi(x), x \in \mathcal{X}\}$ with $\mathsf{E}\int_{\mathcal{X}} \xi^2(x)\,dx < +\infty$.

Denote by $m(x) = \mathsf{E}\,\xi(x)$ the regression curve and by $G(x, x') = \mathrm{Cov}\{\xi(x), \xi(x')\}$ the covariance operator with the Karhunen-Loève L^2 representation

$$\xi_i(x) = m(x) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(x), \tag{2.2}$$

one has the random coefficients $\{\xi_{ik}\}_{k=1}^{\infty}$ uncorrelated with mean 0 and variance 1. Here $\phi_k(x) = \sqrt{\lambda_k}\psi_k(x)$, where $\{\lambda_k\}_{k=1}^{\infty}$ and $\{\psi_k(x)\}_{k=1}^{\infty}$ are respectively the eigenvalues and eigenfunctions of $G(x,x')$ such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ and $\{\psi_k\}_{k=1}^{\infty}$ forms an orthonormal basis of $L^2(\mathcal{X})$. Therefore, $G(x,x') = \sum_{k=1}^{\infty} \phi_k(x)\phi_k(x')$ and $\int G(x,x')\phi_k(x')\,dx' = \lambda_k\phi_k(x)$.

In applications, the number of eigenfunctions $\psi_k(x), k = 1, 2, ...$ needs to be chosen by some criterion, see Yao et al. (2005a). In the sparse curve data situation, many practical studies have shown that fitting too many eigenfunctions can heavily degrade the overall fit, see e.g. James, Hastie and Sugar (2000). Hence, in what follows, we assume that $\lambda_k = 0$ if $k > \kappa$, where κ is a positive constant. Equations (2.1) and (2.2) can then be written as:

$$Y_{ij} = m(X_{ij}) + \sum\nolimits_{k=1}^{\kappa} \xi_{ik}\phi_k(X_{ij}) + \sigma(X_{ij})\varepsilon_{ij}. \tag{2.3}$$

For convenience, we denote the conditional variance of Y_{ij} given $X_{ij} = x$ as

$$\sigma_Y^2(x) = G(x,x) + \sigma^2(x) = \mathsf{Var}(Y_{ij}\,|X_{ij} = x). \tag{2.4}$$

We are interested in the sparse situation where the number of measurements N_i within subject are i.i.d. copies of a positive random integer N_1, see Yao et al. (2005a), Yao et al. (2005b), Yao (2007).

To introduce the estimator, denote by K a kernel function, $h = h_n > 0$ a bandwidth and $K_h(x) = h^{-1}K(x/h)$. Let $N_{\mathrm{T}} = \sum_{i=1}^{n} N_i$ be the total sample size and define $\mathbf{Y} = (Y_{ij})_{1\leq j\leq N_i, 1\leq i\leq n}$ the $N_{\mathrm{T}} \times 1$ vector of responses. For any $x \in [0,1]$, let $\mathbf{X} = \mathbf{X}(x) = (1, X_{ij} - x)_{1\leq j\leq N_i, 1\leq i\leq n}$ be the design matrix for linear regression and $\mathbf{W} = \mathbf{W}(x) = N_{\mathrm{T}}^{-1}\,\mathrm{diag}\{K_h(X_{11} - x), \cdots, K_h(X_{nN_n} - x)\}$ the kernel weight diagonal matrix. Following Fan and Gijbels (1996), local linear estimators of $m(x)$ and $m'(x)$ are

$$\begin{aligned}\{\widehat{m}(x), \widehat{m'}(x)\}^{\mathsf{T}} &= \arg\min_{a,b}\{\mathbf{Y} - \mathbf{X}(a,b)^{\mathsf{T}}\}^{\mathsf{T}}\mathbf{W}\{\mathbf{Y} - \mathbf{X}(a,b)^{\mathsf{T}}\} \\ &= \left(\mathbf{X}^{\mathsf{T}}\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{W}\mathbf{Y}.\end{aligned}$$

Consequently, with $e_0^{\mathsf{T}} = (1,0)$, $\widehat{m}(x)$ is written as

$$\widehat{m}(x) = e_0^{\mathsf{T}}\left(\mathbf{X}^{\mathsf{T}}\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{W}\mathbf{Y}, \tag{2.5}$$

where the dispersion matrix

$$\mathbf{X}^{\mathsf{T}}\mathbf{W}\mathbf{Y} = \operatorname{diag}(1,h) \begin{pmatrix} s_{n,0} & s_{n,1} \\ s_{n,1} & s_{n,2} \end{pmatrix} \operatorname{diag}(1,h), \tag{2.6}$$

has for any nonnegative integer l,

$$s_{n,l} = s_{n,l}(x) = N_{\mathrm{T}}^{-1} \sum\nolimits_{i,j} K_h(X_{ij} - x) \{(X_{ij} - x)/h\}^l. \tag{2.7}$$

3. Main Results. Without loss of generality, assume $\mathcal{X} = [0,1]$ and consider the assumptions:

(A1) *The mean function $m(x) \in C^2[0,1]$, i.e. twice continuously differentiable.*

(A2) *$\{X_{ij}\}_{i=1,j=1}^{\infty,\infty}$ are i.i.d. with a probability density $f(x)$. The functions $f(x), \sigma(x)$ and $\phi_k \in C^1[0,1]$ with $f(x) \in [c_f, C_f], \sigma(x) \in [c_\sigma, C_\sigma]$ and all involved constants are finite and positive.*

(A3) *The numbers of observations $N_i, i = 1, 2, \ldots$ are i.i.d. random positive integers with $\mathsf{E}\, N_1^r \le r! c_N^r, r = 2, 3, \ldots$ for some constant $c_N > 0$. $(N_i)_{i=1}^{\infty}, (X_{ij})_{i=1,j=1}^{\infty,\infty}, (\xi_{ik})_{i=1,k=1}^{\infty,\kappa}, (\varepsilon_{ij})_{i=1,j=1}^{\infty,\infty}$ are independent, while $\{\xi_{ik}\}_{i=1,k=1}^{\infty,\kappa}$ are i.i.d. $\mathrm{N}(0,1)$.*

(A4) *There exists $r > 5$, such that $\mathsf{E}\,|\varepsilon_{11}|^r < \infty$.*

(A5) *The bandwidth $h = h_n$ satisfies $nh^4 \to \infty$, $nh^5 \log n \to 0$ and $h < 1/2$.*

(A6) *The kernel function $K(x)$ is a symmetric probability density function supported on $[-1,1]$ and $\in C^3[-1,1]$.*

Assumptions (A1), (A2), (A5) and (A6) haven been postulated in many papers related to kernel smoothing. (A3) has been used in Yao et al. (2005a). (A4) can be found also in Ma et al. (2010).

For a nonnegative integer l and a continuous function $L(x)$, define:

$$\mu_{l,x}(L) = \begin{cases} \int_{-x/h}^{1} v^l L(v)\,dv, & x \in [0,h) \\ \mu_l(L) = \int_{-1}^{1} v^l L(v)\,dv, & x \in [h, 1-h] \\ \int_{-1}^{(1-x)/h} v^l L(v)\,dv, & x \in (1-h, 1] \end{cases} \tag{3.1}$$

$$D_x(L) = \mu_{2,x}(L)\mu_{0,x}(L) - \mu_{1,x}^2(L), \tag{3.2}$$

and the equivalent kernel function, see Fan and Gijbels (1996):

$$K_x^*(u) = K(u)\{\mu_{2,x}(K) - \mu_{1,x}(K)u\} D_x^{-1}(K), K_{x,h}^*(u) = K_x^*(u/h)/h \tag{3.3}$$

where $D_x^{-1}(K)$ exists by Lemma A.5. One may verify:

$$\mu_{0,x}(K_x^*) = 1, \mu_{1,x}(K_x^*) = 0$$

$$D_x(K) = \mu_2(K), K_x^*(u) \equiv K(u), \forall x \in [h, 1-h].$$

The asymptotic variance function is:

$$\sigma_n^2(x) \stackrel{\text{def}}{=} \frac{\|K_x^*\|_2^2 \sigma_Y^2(x)}{nhf(x) \mathsf{E} N_1}\left[1 + \frac{\mathsf{E}\left(N_1^2 - N_1\right)}{\mathsf{E} N_1}\frac{G(x,x) f(x) h}{\sigma_Y^2(x)\|K_x^*\|_2^2} + \frac{\mu_{1,x}\left(K_x^{*2}\right)\left\{\sigma_Y^2(x) f(x)\right\}' h}{\|K_x^*\|_2^2 \sigma_Y^2(x) f(x)}\right]. \tag{3.4}$$

Define $z_{1-\alpha/2} \stackrel{\text{def}}{=} \Phi^{-1}(1-\alpha/2)$ and

$$Q_h(\alpha) \stackrel{\text{def}}{=} a_h + a_h^{-1}[\log\{\sqrt{C(K)}/(2\pi)\} - \log\{-\log\sqrt{1-\alpha}\}] \tag{3.5}$$

with $a_h = \sqrt{-2\log h}$, $C(K) = \{\int_{-1}^{1} K'(x)^2 dx\}\{\int_{-1}^{1} K^2(x) dx\}^{-1}$.

THEOREM 3.1. *Under Assumptions (A1)-(A6), for any* $\alpha \in (0,1)$

$$\lim_{n\to\infty} \mathrm{P}\{\sup_{x\in[0,1]} |\widehat{m}(x) - m(x)| / \sigma_n(x) \le Q_h(\alpha)\} = 1 - \alpha,$$

$$\lim_{n\to\infty} \mathrm{P}\{|\widehat{m}(x) - m(x)| / \sigma_n(x) \le z_{1-\alpha/2}\} = 1 - \alpha, \forall x \in [0,1],$$

with $\sigma_n^2(x)$ *and* $Q_h(\alpha)$ *given in (3.4) and (3.5).*

By Theorem 3.1, we construct the SCC and SCI for $m(x)$ as follows,

COROLLARY 3.1. *Assume (A1)-(A6). A* $100(1-\alpha)\%$ *simultaneous confidence corridor (SCC) for* $m(x)$ *is:*

$$[\widehat{m}(x) \pm \sigma_n(x) Q_h(\alpha)]. \tag{3.6}$$

A shoal of confidence intervals (SCI) is given by:

$$\left[\widehat{m}(x) \pm \sigma_n(x) z_{1-\alpha/2}\right]. \tag{3.7}$$

A simple approximation of $\sigma_n^2(x)$ is given by:

$$\sigma_{n,\mathrm{IID}}^2(x) = \frac{\|K_x^*\|_2^2\,\sigma_Y^2(x)}{nhf(x)\,\mathsf{E}\,N_1}. \tag{3.8}$$

PROPOSITION 3.1. *Given (A2), (A3) and (A6), then* $\sup_{x\in[0,1]}\left|\sigma_n^{-1}(x)\,\sigma_{n,\mathrm{IID}}(x)-1\right| = \mathcal{O}(h)$.

Using $\sigma_{n,\mathrm{IID}}^2(x)$ instead of $\sigma_n^2(x)$ is equivalent to treat $\{X_{ij}, Y_{ij}\}, 1\le j\le N_i, 1\le i\le n$ as i.i.d data, which implies that the longitudinal dependence structure is negligible in case of sparsity. This was also observed by Ma et al. (2010), Wang et al. (2005).

4. Implementation. Now we outline the construction of the SCC and SCI. Recall the definition of $\widehat{m}(x)$. The practical implementation of (3.6) and (3.7) is via estimating $\mathsf{E}\,N_1$, $f(x)$ and $\sigma_Y(x)$, see Wang and Yang (2009) and references therein. The quantity $\mathsf{E}\,N_1$ is estimated by N_{T}/n and the estimator of the density $f(x)$ is

$$\widehat{f}(x) = N_{\mathrm{T}}^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h(X_{ij}-x). \tag{4.1}$$

The local linear estimator $\widehat{\sigma}_Y(x) = \widehat{a}_1$ results from:

$$\left(\widehat{a}_1, \widehat{b}_1\right) = \arg\min_{a_1,b_1}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\left\{\widehat{\varepsilon}_{ij}^2 - a_1 - b_1(X_{ij}-x)\right\}^2 w_{ij},$$

where $\widehat{\varepsilon}_{ij} = Y_{ij} - \widehat{m}(X_{ij})$, $w_{ij} = N_{\mathrm{T}}^{-1}K_h(X_{ij}-x)$ and $h = N_{\mathrm{T}}^{-1/5}(\log n)^{-1}$ satisfying (A5). The consistency of $\widehat{f}(x)$ and $\widehat{\sigma}_Y(x)$ is proved e.g. in Li and Hsing (2010), Yao et al. (2005a). Therefore, the SCC $\widehat{m}(x) \pm \hat{\sigma}_{n,\mathrm{IID}}(x)\,Q_h(\alpha)$ and the SCI $\widehat{m}(x) \pm \hat{\sigma}_{n,\mathrm{IID}}(x)\,z_{1-\alpha/2}$ both have asymptotic confidence level $1-\alpha$.

5. Monte Carlo Studies. This section checks the finite sample performance of the SCC. The data are generated from (2.1) with $\kappa = 2$:

$$Y_{ij} = m(X_{ij}) + \sum_{k=1}^{2}\xi_{ik}\phi_k(X_{ij}) + \sigma(X_{ij})\,\varepsilon_{ij},$$

with $m(x) = \sin\{2\pi(x-1/2)\}$, $\phi_1(x) = -0.2\cos\{\pi(x-1/2)\}$, $\phi_2(x) = 0.1\sin\{\pi(x-1/2)\}$, $\sigma(x) = \exp\{3(x-0.5)^2\}/[1+\exp\{3(x-0.5)^2\}]$ and $X \sim \mathrm{U}[0,1]$, $\xi_k \sim \mathrm{N}(0,1)$, $\varepsilon_{ij} \sim \mathrm{N}(0,1)$, while N_i has a discrete uniform distribution from $5, \ldots, 15$ and n varies: 20,50,100,200. The confidence level is set to: $1 - \alpha = 0.95, 0.99$.

(Insert Figure 1 "Dataplot and trajectories" about here)

Table 1
Empirical coverage from 200 replications

n	$1-\alpha=0.95$	$1-\alpha=0.99$
20	0.925	0.965
50	0.940	0.980
100	0.950	0.995
200	0.955	0.990

The empirical coverage is reported in Table 1. The data are displayed in Figure 1. Clearly, the coverage approaches the nominal confidence levels as n increases, see Theorem 3.1. Coverage frequencies remain stable if the bandwidths' slightly vary. In practice, one can choose bandwidths adaptively to achieve better performance. The theoretical study of this issue would require too much space here. We therefore do not pursue this. Figure 2 plots the SCCs with 95% and 99% confidence levels. The above studies have illustrated the reliability of our method, which actually ensures the application of the SCC including the true curve for the real data in Section 6.

(Insert Figure 2 "The 95% and 99% SCCs of the mean curve" about here)

6. Application. Now we apply the SCC and SCI to a longitudinal study of growth curve data. The data curve analysis is a key in the studies of human skeletal health. These data consist of measurements Y_{ij}, spinal bone mineral density (g/cm^2), for $n = 286$ people. However, N_i, the number of measurements for each individual, is between 2 and 4 (sparsity), and X_{ij}, the time points of measurements (aged 8.8–26.2 yr), varies among individuals.

An earlier study of the growth curve data in James, Hastie and Sugar (2000) developed the pointwise inference of the mean function. Using the bootstrap method, they constructed the confidence intervals to test the mean curve at points of interest, e.g. the fastest growth point at about 15 yr. In our study, this task can be also done via constructing the SCI by (3.7). However, its computation is much faster than the bootstrap procedures. Furthermore, we will use the SCC to examine the global shape of the mean curve on the whole domain, such as the upward or downward trend at different stages, the acceleration or plateau during different periods.

(Insert Figure 3 "Growth curve data and the SCCs & SCIs of its mean curve" about here)

Figure 3 (a) exhibits the scatter plot of the spinal bone density v.s. the age. Figure 3 (b), (c) and (d) depict the SCCs and SCIs of the population mean of the growth curve data, at the confidence levels of 90%, 95% and 99%, respectively. For the pointwise inference, James, Hastie and Sugar (2000) and our method share similar SCIs. However, testing the global shape of the growth curve, the constructed SCCs can indicate that the spinal bone density at mean level increases with age, but the bone growth is accelerated during early adolescence (9-15 yr) whereas it reaches the plateau during late puberty (16-26 yr). An R algorithm of our method has been provided on www.quantlet.org.

APPENDIX

A.1. Preliminaries. We introduce Lemmas (A.1)-(A.4) for the proof of Theorem 3.1 (Appendix A.2). For the details of Lemma A.1, see Cierco-Ayrolles et al (2003), Zheng, Yang and Härdle (2010).

LEMMA A.1. *[Cierco-Ayrolles, Croquette and Delmas (2003)] Let $X(t)$ be a Gaussian process with almost surely $\mathcal{C}^1$ sample paths on $[0,T]$. Then*

$$\begin{aligned} &\mathrm{P}\left\{|X(0)|>u\right\}+\mathsf{E}\left[\left(U_u^X[0,T]+D_{-u}^X[0,T]\right)I_{\{|X(0)|\leqslant u\}}\right] \\ &-\frac{1}{2}\mathsf{E}\left(U_u^X[0,T]+D_{-u}^X[0,T]\right)^{[2]}\leq \mathrm{P}\{\sup\nolimits_{x\in[0,T]}|X(t)|>u\}\leq \\ &\mathrm{P}\left\{|X(0)|>u\right\}+\mathsf{E}\left[\left(U_u^X[0,T]+D_{-u}^X[0,T]\right)I_{\{|X(0)|\leqslant u\}}\right]. \end{aligned} \tag{A.1}$$

LEMMA A.2. *[Theorem 1 of Cierco-Ayrolles, Croquette and Delmas (2003)] Suppose X is a $\mathcal{C}^1$ real-valued Gaussian process defined on an interval I and $\{X(t), X(s), X'(t), X'(s)\}$ is non-degenerate $\forall t\neq s, (t,s)\in I^2$. Then, denoting p_V the probability density of a random vector V:*

$$\mathsf{E}(U_u^X[I]^{[2]})=\int_{I^2}\int_{(0,\infty)^2}|x_1'||x_2'|\,p_{X_t;X_s;X_t';X_s'}(u;u;x_1';x_2')\,dx_1'dx_2'dtds,$$

$$\mathsf{E}\left(U_u^X\left[I\right]D_{-u}^X\left[I\right]\right)=\int_{I^2}\int_0^{+\infty}\int_{-\infty}^0\left|x_1'\right|\left|x_2'\right|p_{X_t;X_s;X_t';X_s'}\left(u;-u;x_1';x_2'\right)dx_1'dx_2'dtds.$$

LEMMA A.3. *[Theorem 2.6.7 of Csőrgő and Révész (1981)] Suppose that* $\xi_i, 1 \leq i \leq n$ *are i.i.d. with* $\mathsf{E}\,\xi_1 = 0, \mathsf{E}\,\xi_1^2 = 1$ *and* $H(x) > 0$ $(x \geq 0)$ *is an increasing continuous function such that* $x^{-2-\gamma}H(x)$ *is increasing for some* $\gamma > 0$ *and* $x^{-1}\log H(x)$ *is decreasing with* $\mathsf{E}\,H(|\xi_1|) < \infty$*. Then there exist constants* $C_1, C_2, a > 0$ *which depend only on the distribution of* ξ_1 *and a sequence of Brownian motions* $\{W_n(t), 0 \leq t < \infty\}_{n=1}^{\infty}$ *such that for any* $\{x_n\}_{n=1}^{\infty}$ *satisfying* $H^{-1}(n) < x_n < C_1(n\log n)^{1/2}$ *and* $S_k = \sum_{i=1}^k \xi_i$

$$\mathrm{P}\left\{\max\nolimits_{1\leq k\leq n}\left|S_k - W_n(k)\right| > x_n\right\} \leq C_2 n\left\{H(ax_n)\right\}^{-1}.$$

LEMMA A.4. *[Theorem 1.2 of Bosq (1996)]Suppose that* $\xi_i, 1 \leq i \leq n$ *are i.i.d. with* $\sigma^2 = \mathsf{E}\,\xi_1^2, \mathsf{E}\,\xi_1 = 0$ *and there exists* $c > 0$ *such that for* $r = 3, 4, ..., \mathsf{E}\,|\xi_1|^r \leq c^{r-2}r!\,\mathsf{E}\,\xi_1^2 < +\infty$*, then for each* $n > 1$*,* $t > 0$*,* $\mathrm{P}\left(|S_n| \geq \sqrt{n}\sigma t\right) \leq 2\exp\{-t^2\left(4 + 2ct/\sqrt{n}\sigma\right)^{-1}\}$.

A.2. Proof of Theorem *3.1.* Throughout this section, for functions $a_n(x)$ and $b_n(x)$, $a_n(x) = \mathcal{U}\{b_n(x)\}$ and $a_n(x) = \mathcal{U}\{b_n(x)\}$ respectively means that, as $n \to \infty$, $\sup_{x\in[0,1]}|a_n(x)/b_n(x)| = \mathcal{O}(1)$ and $\sup_{x\in[0,1]}|a_n(x)/b_n(x)| = \mathcal{O}(1)$. In addition, $a_n(x) = \mathcal{U}_{a.s.}\{b_n(x)\}$ and $a_n(x) = \mathcal{U}_{a.s.}\{b_n(x)\}$ respectively means that, as $n \to \infty$, $a_n(x) = \mathcal{U}\{b_n(x)\}$ and $a_n(x) = \mathcal{U}\{b_n(x)\}$ almost surely, and $\mathcal{O}_{a.s.}$, $\mathcal{O}_p$, $\mathcal{O}_{a.s.}$, $\mathcal{O}_p$ are similarly defined.

We denote $\mathbf{m} = (m(X_{ij}))$, $\varepsilon = (\sigma(X_{ij})\varepsilon_{ij})$, $\xi_k = (\xi_{ik}\varphi_k(X_{ij}))$. The signal and noise decomposition $\mathbf{X}^{\mathsf{T}}\mathbf{W}\mathbf{Y} = \mathbf{X}^{\mathsf{T}}\mathbf{W}\mathbf{m} + \sum_{k=1}^{\kappa}\mathbf{X}^{\mathsf{T}}\mathbf{W}\xi_k + \mathbf{X}^{\mathsf{T}}\mathbf{W}\varepsilon$ implies that,

$$\widehat{m}(x) - m(x) = \widetilde{m}(x) - m(x) + \widetilde{e}(x), \tag{A.2}$$
$$\widetilde{e}(x) = \sum\nolimits_{k=1}^{\kappa}\widetilde{\xi}_k(x) + \widetilde{\varepsilon}(x),$$

where $\widetilde{\xi}_k(x) = e_0^{\mathsf{T}}(\mathbf{X}^{\mathsf{T}}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{W}\xi_k$ and $\widetilde{\varepsilon}(x) = e_0^{\mathsf{T}}(\mathbf{X}^{\mathsf{T}}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{W}\varepsilon$.

The error structure in (A.2) allows one to investigate the asymptotics of $\sup_{x\in[0,1]}|\widetilde{e}(x)/\sigma_n(x)|$ and $\sup_{x\in[0,1]}|\{\widetilde{m}(x) - m(x)\}/\sigma_n(x)|$ separately in Lemmas A.6-A.14, with $\sigma_n(x)$ given in (3.4).

We introduce some more notations, defining

$$\mathbf{D}_x = \begin{pmatrix} \mu_{2,x}(K) & -\mu_{1,x}(K) \\ -\mu_{1,x}(K) & \mu_{0,x}(K) \end{pmatrix}, \tag{A.3}$$

with $\mu_{l,x}(K)$ given in (3.1)

$$\widehat{\varepsilon}(x) = f^{-1}(x) N_{\mathrm{T}}^{-1} \sum\nolimits_{i,j} K^*_{x,h}(X_{ij} - x) \sigma(X_{ij}) \varepsilon_{ij}, \tag{A.4}$$

$$\widehat{\xi}_k(x) = f^{-1}(x) N_{\mathrm{T}}^{-1} \sum\nolimits_{i,j} K^*_{x,h}(X_{ij} - x) \phi_k(X_{ij}) \xi_{ik}, \tag{A.5}$$

with $K^*_{x,h}(u)$ given in (3.3)

$$R_{ij,\varepsilon}(x) = K^*_{x,h}(X_{ij} - x) D_x(K) \sigma(X_{ij}), \tag{A.6}$$

$$R_{ik,\xi_k} = \sum\nolimits_{j=1}^{N_i} K^*_{x,h}(X_{ij} - x) D_x(K) \phi_k(X_{ij}), \tag{A.7}$$

with $D_x(K)$ given in (3.2)

$$\sigma^2_{\varepsilon,n}(x) = f^{-2}(x) N_{\mathrm{T}}^{-2} D_x^{-2}(K) \sum\nolimits_{i,j} R^2_{ij,\varepsilon}(x), \tag{A.8}$$

$$\sigma^2_{\xi_k,n}(x) = f^{-2}(x) N_{\mathrm{T}}^{-2} D_x^{-2}(K) \sum\nolimits_{i=1}^{n} R^2_{ik,\xi_k}(x), \tag{A.9}$$

$$C_x(K) = \frac{\mu_{0,x}\{K_x^{*\prime}(x)^2\}}{\mu_{0,x}\{K_x^*(x)^2\}} - \frac{\mu_{0,x}^2\{K_x^*(x) K_x^{*\prime}(x)\}}{\mu_{0,x}^2\{K_x^*(x)^2\}}, \tag{A.10}$$

where $K_x^{*\prime}(x) = dK_x^*(x)/dx$, $\mu_{l,x}(L)$ given in (3.1). It is easily verified that $C_x(K) = C(K)$, $\forall x \in [h, 1-h]$ with $C(K)$ given in (3.5).

LEMMA A.5. *Under Assumptions (A5)-(A6), for* $x \in [0,1]$

$$0 < D_0(K) \le D_x(K) \le D_{1/2}(K) = \mu_2(K) < +\infty, \tag{A.11}$$

while $\sup_{x\in[0,1]} |C_x(K)| < \infty$.

PROOF. See Appendix B, Zheng, Yang and Härdle (2010). □

LEMMA A.6. *Under Assumptions (A1)-(A6), for $D_x(K)$ given in (3.2) and $\mathbf{D}_x$ in (A.3),*

$$\left(\mathbf{X}^{\mathsf{T}}\mathbf{W}\mathbf{X}\right)^{-1} = f^{-1}(x)\operatorname{diag}\left(1, h^{-1}\right)\left\{D_x^{-1}(K)\mathbf{D}_x + \Delta_{1,n}(x)\right\}\operatorname{diag}\left(1, h^{-1}\right)$$

as $n \to \infty$, where the 2×2 random matrices $\Delta_{1,n}(x) = \mathcal{U}(h) + \mathcal{U}_{a.s.}\{\sqrt{\log n/(nh)}\}$.

PROOF. For notational simplicity, let $x \in [h, 1-h]$, we investigate $s_{n,l}(x), l = 0, 1, 2$, given in (2.7).

$$|s_{n,0}(x) - f(x)| \le \left|n(\mathsf{E}\,N_1)N_{\mathrm{T}}^{-1} - 1\right|\left|(n\,\mathsf{E}\,N_1)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h(X_{ij} - x)\right| + \tag{A.12}$$
$$|\mathsf{E}\,K_h(X_{ij} - x) - f(x)| + \left|(n\,\mathsf{E}\,N_1)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{N_i} K_h(X_{ij} - x) - \mathsf{E}\,K_h(X_{ij} - x)\right|$$
$$= I_1(x) + I_2(x) + I_3(x).$$

Clearly, $I_2(x) = \mathcal{U}\left(h^2\right)$ and $\mathsf{E}\{K_h(X_{ij} - x)\}^r = \mathcal{U}\left(h^{1-r}\right)$ for $r \ge 2$. Define $I_3(x) = (n\,\mathsf{E}\,N_1)^{-1}\left|\sum_{i=1}^{n}\zeta_{i,h}\right|$ with $\zeta_{i,h} = \sum_{j=1}^{N_i} K_h(X_{ij} - x) - \mathsf{E}\,K_h(X_{ij} - x)\,\mathsf{E}\,N_1$. For large n,

$$\mathsf{E}\,|\zeta_{i,h}|^r = \mathsf{E}\left|\sum_{j=1}^{N_i} K_h(X_{ij} - x) - \mathsf{E}\,K_h(X_{ij} - x)\,\mathsf{E}\,N_1\right|^r \le \tag{A.13}$$
$$2^{r-1}[\mathsf{E}\{\sum_{j=1}^{N_i} K_h(X_{ij} - x)\}^r + \{\mathsf{E}\,K_h(X_{ij} - x)\,\mathsf{E}\,N_1\}^r] \le$$
$$2^r\,\mathsf{E}\{\sum_{j=1}^{N_i} K_h(X_{ij} - x)\}^r = 2^r\,\mathsf{E}\left[\sum_{0 \le r_1, \ldots, r_{N_i} \le r}^{r_1 + \ldots + r_{N_i} = r}\binom{r}{r_1 \ldots r_{N_i}}\prod_{i=1}^{N_i}\mathsf{E}\{K_h(X_{ij} - x)\}^{r_i}\right]$$
$$\le 2^r\,\mathsf{E}\left[N_i^r \max_{0 \le r_1, \ldots,\, r_{N_i} \le r}^{r_1 + \ldots + r_{N_i} = r}\prod_{i=1}^{N_i}\mathsf{E}\{K_h(X_{ij} - x)\}^{r_i}\right] \le 2^r\left(\mathsf{E}\,N_1^r\right)C_r h^{1-r} \le C_\zeta r! h^{1-r}.$$

It can be next verified that $\mathsf{E}\,(\zeta_{i,h})^2 = (\mathsf{E}\,N_1)\,h^{-1}f(x)\int K^2(v)\,dv\,\{1 + u(1)\}$. Hence, $\exists\, C'_\zeta > c'_\zeta > 0$ such that $c'_\zeta h^{-1} < \mathsf{E}\,(\zeta_{i,h})^2 < C'_\zeta h^{-1}$, i.e., $\mathsf{E}\,|\zeta_{i,h}|^r \le c_*^{r-2} r!\,\mathsf{E}\,(\zeta_{i,h})^2$ with $c_* = (C_\zeta/c'_\zeta)^{\frac{1}{r-2}}h^{-1}$, see (A.13). In fact, it implies $\{\zeta_{i,h}\}_{i=1}^n$ satisfies Cramér's Condition. Therefore, applying Lemma A.4 to $\sum_{i=1}^{n}\zeta_{i,h}$, for large n and large $\delta > 0$, one shows

$$\mathrm{P}\{I_3(x) > \delta\sqrt{\log n/(nh)}\} \le$$
$$2\exp[-(\mathsf{E}\,N_1)^2\,\delta^2\log n\{4C'_\zeta + 2\delta\,\mathsf{E}\,N_1\left(C_\zeta/c'_\zeta\right)^{1/(r-2)}\sqrt{\log n/(nh)}\}^{-1}] \le 2n^{-C\delta^2} \le 2n^{-8}.$$

Now discretize $h = x_0 < x_1 < \cdots < x_{M_n} = 1-h$ with $M_n = n^4$ and then,

$$\mathrm{P}\{\max_{j=0}^{M_n} I_3(x_j) > \delta\sqrt{\log n/(nh)}\} \leq \sum\nolimits_{j=0}^{M_n} P\{|I_3(x)| > \delta\sqrt{\log n/(nh)}\} \leq 2n^{-4},$$

and hence the Borel-Contelli Lemma implies that $\max_{j=0}^{M_n} I_3(x_j) = \mathcal{O}_{a.s.}\{\sqrt{\log n/(nh)}\}$. It is also clear that,

$$\begin{aligned}\sup\nolimits_{x\in[h,1-h]} I_3(x) &\leq \max\nolimits_{j=0}^{M_n} I_3(x_j) + \max\nolimits_{j=0}^{M_n-1}\sup\nolimits_{x\in[x_j,x_{j+1}]} |I_3(x_j) - I_3(x)| \\ &\leq \mathcal{O}_{a.s.}\{\sqrt{\log n/(nh)}\} + \mathcal{U}\left\{(1-2h)/\left(nh^4\right)\right\} = \mathcal{O}_{a.s.}\{\sqrt{\log n/(nh)}\},\end{aligned}$$

which by the definition of $I_3(x)$ implies that

$$\begin{aligned}(n\,\mathsf{E}\,N_1)^{-1}\sum\nolimits_{i=1}^{n}\sum\nolimits_{j=1}^{N_i} K_h(X_{ij}-x) &= \mathsf{E}\,K_h(X_{ij}-x) + \mathcal{U}_{a.s.}\{\sqrt{\log n/(nh)}\} \\ &= f(x) + U(h^2) + \mathcal{U}_{a.s.}\{\sqrt{\log n/(nh)}\}.\end{aligned} \tag{A.14}$$

Applying Lemma A.4 for N_{T}, one has $|(n\,\mathsf{E}\,N_1)/N_{\mathrm{T}} - 1| = \mathcal{O}_{a.s}\{\sqrt{\log n/n}\}$ and (A.14) also implies that $\sup_{x\in[h,1-h]} I_1(x) = \mathcal{O}_{a.s.}(\sqrt{\log n/n})$. Now, by (A.12), $s_{n,0}(x) = f(x) + \mathcal{U}\left(h^2\right) + \mathcal{U}_{a.s.}\{\sqrt{\log n/(nh)}\}$. Similarly, $s_{n,1}(x) = \mathcal{U}(h) + \mathcal{U}_{a.s.}\{\sqrt{\log n/(nh)}\}$and $s_{n,2}(x) = f(x)\mu_2(K) + \mathcal{U}(h) + \mathcal{U}_{a.s.}\{\sqrt{\log n/(nh)}\}$ which imply that $\mathbf{X}^{\mathsf{T}}\mathbf{W}\mathbf{X}$ can be written as

$$f(x)\,\mathrm{diag}(1,h)[\mathrm{diag}\{1,\mu_2(K)\} + \mathcal{U}(h) + \mathcal{U}_{a.s.}\{\sqrt{\log n/(nh)}\}]\,\mathrm{diag}(1,h).$$

Finally, the inverse of this matrix is concluded as this lemma. □

LEMMA A.7. *Under Assumptions (A1)-(A6), as $n\to\infty$, $\|\widetilde{m}(x) - m(x)\|_\infty = \mathcal{O}_{a.s.}\left(h^2\right)$.*

PROOF. See Proof of Theorem 6.5, page 268 of Fan and Yao (2005). □

LEMMA A.8. *Under Assumptions (A1)-(A6), for $\widehat{\varepsilon}(x)$ and $\widehat{\xi}_k(x)$ given in (A.4) and (A.5),*

$$\widetilde{e}(x) = \{1 + \Delta_{2,n}(x)\}\{\widehat{\varepsilon}(x) + \sum\nolimits_{k=1}^{\kappa}\widehat{\xi}_k(x)\}$$

as $n\to\infty$, where the 2×2 random matrices $\Delta_{2,n}(x) = \mathcal{U}(h) + \mathcal{U}_{a.s.}\{\sqrt{\log n/(nh)}\}$.

PROOF. For notational simplicity, let $x \in [h, 1-h]$, therefore $\widehat{\varepsilon}(x) + \sum_{k=1}^{\kappa} \widehat{\xi}_k(x) = f^{-1}(x)T_0(x)$ with $T_l, l = 0, 1$ defined as

$$T_l(x) = N_{\mathrm{T}}^{-1} \sum_{i,j} K_h(X_{ij} - x) \{(X_{ij} - x)/h\}^l \{\sigma(X_{ij}) \varepsilon_{ij} + \sum_{k=1}^{\kappa} \phi_k(X_{ij}) \xi_{ik}\}.$$

Lemma A.6 shows that for $\Delta_{1,n}(x)$ given in Lemma A.6

$$\widetilde{e}(x) = f^{-1}(x) e_0^{\mathsf{T}} \mathrm{diag}\left(1, h^{-1}\right) \left[\mathrm{diag}\left\{1, \mu_2^{-1}(K)\right\} + \Delta_{1,n}(x)\right] \{T_0(x), T_1(x)\}^{\mathsf{T}},$$

i.e., $\widetilde{e}(x) = \{1 + \Delta_{1,n}(x)\} f^{-1}(x) T_0(x)$. Therefore, this lemma holds. □

Let $X_{ij}, 1 \le i \le n,\ 1 \le j \le N_i$ be descendingly ordered as $X_{(t)},\ 1 \le t \le N_{\mathrm{T}},\ S_q = \sum_{t=1}^{q} \varepsilon_{(t)}$ where $\varepsilon_{(t)}$ is corresponding in index to $X_{(t)}$.

LEMMA A.9. *Given (A1)-(A6), then there exists a sequence of Wiener processes* $\{W_{N_T}(t)\}_{t=1}^{N_T}$ *independent of* $\{N_i,\ X_{ij},\ \xi_i\ 1 \le i \le n,\ 1 \le j \le N_i,\ 1 \le k \le \kappa\}$ *such that as* $n \to \infty$ *and for some* $t' > 2/5$

$$\|\widehat{\varepsilon}(x) - \widehat{\varepsilon}_{N_T}(x)\|_\infty = \mathcal{O}_{a.s.}(n^{-t'}),$$

with $\widehat{\varepsilon}_{N_T}(x) = \{N_T f(x)\}^{-1} \sum_{t=1}^{N_T} K_{x,h}^{*}\left(X_{(t)} - x\right) \sigma\left(X_{(t)}\right) \{W_{N_T}(t) - W_{N_T}(t-1)\}.$

PROOF. Without loss of generality, let $x \in [h, 1-h]$. By Lemma A.3, let $H(x) = x^r, r > 5$ (Assumption A4) and $x_n = n^s, s \in \left(2r^{-1}, 2/5\right)$. It is easy to verify that $\left\{\varepsilon_{(t)}\right\}_{t=1}^{N_{\mathrm{T}}}$ satisfies the conditions of Lemma A.3 and $nH^{-1}(ax_n) = a^{-r} n^{1-rs} = \mathcal{O}(n^{-s'})$ for some $s' > 1$. Therefore, there exists a sequence of Wiener process $\{W_{N_{\mathrm{T}}}(t)\}_{t=1}^{N_{\mathrm{T}}}$ independent of $\{N_i,\ X_{ij},\ \xi_i\ 1 \le i \le n,\ 1 \le j \le N_i,\ 1 \le k \le \kappa\}$ such that $P\{M_{N_{\mathrm{T}}} > n^s\} \le C_2 n^{-s'}$ with $M_{N_{\mathrm{T}}} = \max_{1 \le q \le N_{\mathrm{T}}} |S_q - W_{N_{\mathrm{T}}}(q)|$ and hence Borel-Cantelli Lemma warrants that $M_{N_{\mathrm{T}}} = \mathcal{O}_{a.s}(n^s)$.

The technique of summation by parts implies that

$$\sup_{x\in[h,1-h]}|\widehat{\varepsilon}(x)-\widehat{\varepsilon}_{N_{\mathrm{T}}}(x)|\leq\sup_{x\in[h,1-h]}N_{\mathrm{T}}^{-1}c_f^{-1}|K_h\left(X_{(N_{\mathrm{T}})}-x\right)\sigma\left(X_{(N_{\mathrm{T}})}\right)\{W_{N_{\mathrm{T}}}(N_{\mathrm{T}})-S_{N_{\mathrm{T}}}\} \quad \text{(A.15)}$$

$$+\sum\nolimits_{t=1}^{N_{\mathrm{T}}-1}\{K_h\left(X_{(t)}-x\right)\sigma\left(X_{(t)}\right)-K_h\left(X_{(t+1)}-x\right)\sigma\left(X_{(t+1)}\right)\}\leq h^{-1}M_{N_{\mathrm{T}}}N_{\mathrm{T}}^{-1}c_f^{-1}\times$$

$$\sup_{x\in[h,1-h]}[\,3C_KC_\sigma+\sum\nolimits_{1\leq t\leq N_{\mathrm{T}}-1}^{X_{(t)}\in[x-h,x+h]}|K\{\left(X_{(t)}-x\right)/h\}\sigma\left(X_{(t)}\right)-K\{\left(X_{(t+1)}-x\right)/h\}\sigma\left(X_{(t+1)}\right)|\,].$$

Since $|ab-cd|\leq|a||b-d|+|b||a-c|+|a-c||b-d|$, (A.15) is bounded by

$$h^{-1}M_{N_{\mathrm{T}}}N_{\mathrm{T}}^{-1}c_f^{-1}\sup_{x\in[h,1-h]}[\,3C_KC_\sigma+\sum\nolimits_{1\leq t\leq N_{\mathrm{T}}-1}^{X_{(t)}\in[x-h,x+h]}2C_K\times$$

$$|\sigma\left(X_{(t)}\right)-\sigma\left(X_{(t+1)}\right)|+C_\sigma|K\{\left(X_{(t)}-x\right)/h\}-K\{\left(X_{(t+1)}-x\right)/h\}|\,].$$

Therefore, $\exists$ constants $L_{K,\sigma}^1, L_{K,\sigma}^2, C$ and C' such that (A.15) is bounded by

$$h^{-1}M_{N_{\mathrm{T}}}N_{\mathrm{T}}^{-1}c_f^{-1}\sup_{x\in[h,1-h]}(\,3C_KC_\sigma+L_{K,\sigma}^1\sum\nolimits_{1\leq t\leq N_{\mathrm{T}}-1}^{X_{(t)}\in[x-h,x+h]}|X_{(t)}-X_{(t+1)}|+$$

$$L_{K,\sigma}^2h^{-1}\sum\nolimits_{1\leq t\leq N_{\mathrm{T}}-1}^{X_{(t)}\in[x-h,x+h]}|X_{(t)}-X_{(t+1)}|\,)\leq h^{-1}M_{N_{\mathrm{T}}}N_{\mathrm{T}}^{-1}\left(C+C'h\right).$$

Namely $\sup_{x\in[h,1-h]}|\widehat{\varepsilon}(x)-\widehat{\varepsilon}_{N_{\mathrm{T}}}(x)|=\mathcal{O}_{a.s}(h^{-1}n^{s-1})$ and by assumption (A5), one obtains

$$\sup_{x\in[h,1-h]}|\widehat{\varepsilon}_{N_{\mathrm{T}}}(x)-\widehat{\varepsilon}(x)|=\mathcal{O}_{a.s.}(n^{-t'}),\ t'>2/5.$$

This completes the proof. □

LEMMA A.10. *Under Assumptions (A1)-(A6), as $n\to\infty$,*

$$\left\|N_T^{-1}\sum\nolimits_{i,j}R_{ij,\varepsilon}^2(x)-\mathsf{E}\,R_{11,\varepsilon}^2(x)\right\|_\infty=\mathcal{O}_{a.s.}\{\sqrt{\log n/(nh)}\},$$

$$\left\|N_T^{-1}\sum\nolimits_{i=1}^n\sum\nolimits_{k=1}^\kappa R_{ik,\xi_k}^2(x)-(\mathsf{E}\,N_1)^{-1}\sum\nolimits_{k=1}^\kappa\mathsf{E}\,R_{1k,\xi_k}^2(x)\right\|_\infty=\mathcal{O}_{a.s.}\{\sqrt{\log n/(nh)}\},$$

with $R_{ij,\varepsilon}(x)$ and $R_{ik,\xi_k}(x)$ given in (A.6) and (A.7).

Proof. Without loss of generality, let $x \in [h, 1-h]$. Clearly,

$$\sup_{x\in[h,1-h]}\left|N_{\mathrm{T}}^{-1}\sum_{i=1}^{n}\sum_{k=1}^{\kappa}R_{ik,\xi_k}^2(x)-(\mathsf{E}\,N_1)^{-1}\sum_{k=1}^{\kappa}\mathsf{E}\,R_{ik,\xi_k}^2(x)\right|\leq$$
$$(\mathsf{E}\,N_1)^{-1}\sum_{k=1}^{\kappa}\sup_{x\in[h,1-h]}\left|n^{-1}\sum_{i=1}^{n}R_{ik,\xi_k}^2(x)-\mathsf{E}\,R_{ik,\xi_k}^2(x)\right|+$$
$$(\mathsf{E}\,N_1)^{-1}\sum_{k=1}^{\kappa}\sup_{x\in[h,1-h]}\left|n(\mathsf{E}\,N_1)N_{\mathrm{T}}^{-1}-1\right|\left|n^{-1}\sum_{i=1}^{n}R_{ik,\xi_k}^2(x)\right|.$$

It is next straightforward to verify Cramér's Condition for $R_{ik,\xi_k}^2(x)^* = R_{ik,\xi_k}^2(x) - \mathsf{E}\,R_{ik,\xi_k}^2(x)$, i.e., $\mathsf{E}\{R_{ik,\xi_k}^2(x)^*\}^r \leq c_*^{r-2}r!\,\mathsf{E}\,R_{ik,\xi_k}^2(x)^*$ with $r \geq 2$ and $c_* \sim h^{-1}$. Again, by Lemma A.4, one has $\sup_{x\in[h,1-h]}\left|n^{-1}\sum_{i=1}^{n}R_{ik,\xi_k}^2(x)-\mathsf{E}\,R_{ik,\xi_k}^2(x)\right| = \mathcal{O}_{a.s.}\{\sqrt{\log n/(nh)}\}$, i.e., $n^{-1}\sum_{i=1}^{n}R_{ik,\xi_k}^2(x) = \mathcal{U}\left(h^{-1}\right)+\mathcal{U}_{a.s.}\{\sqrt{\log n/(nh)}\}$. Therefore,

$$\sup_{x\in[h,1-h]}\left|N_{\mathrm{T}}^{-1}\sum_{i=1}^{n}\sum_{k=1}^{\kappa}R_{ik,\xi_k}^2(x)-(\mathsf{E}\,N_1)^{-1}\sum_{k=1}^{\kappa}\mathsf{E}\,R_{ik,\xi_k}^2(x)\right|=\mathcal{O}_{a.s.}\{\sqrt{\log n/(nh)}\}.$$

The proof for $R_{ij,\varepsilon}^2(x)$ is similar. □

Throughout the remainder, define the standardized noise processes as

$$\eta_n(x)=\eta(x)=\{\widehat{\varepsilon}_{N_{\mathrm{T}}}(x)+\sum_{k=1}^{\kappa}\widehat{\xi}_k(x)\}\{\sigma_{\varepsilon,n}^2(x)+\sum_{k=1}^{\kappa}\sigma_{\xi_k,n}^2(x)\}^{-1/2}, x\in[0,1] \tag{A.16}$$

with $\widehat{\varepsilon}_{N_{\mathrm{T}}}(x)$, $\widehat{\xi}_k(x)$, $\sigma_{\varepsilon,n}^2(x)$ and $\sigma_{\xi_k,n}^2(x)$, respectively, given in Lemma A.9, (A.5), (A.8) and (A.9). For any n and fixed x ,

$$\mathcal{L}\{\eta(x)\,|\,(X_{ij},N_i),1\leq j\leq N_i,1\leq i\leq n\}=\mathrm{N}(0,1),$$

and hence $\mathcal{L}\{\eta(x)\} = \mathrm{N}(0,1)$ which implies $\eta(x)$ is a standardized Gaussian process.

To compute the extreme value of $\eta(x)$ by Lemma A.1, one needs to study its correlation function. In the following, denote $xh^{-1} = t \in \left[0, h^{-1}\right]$, $m_t = m(t) = \mathsf{E}\,\eta(t)$, $r(t,s) = \mathsf{E}\,\eta(t)\,\eta(s)$, $r_t = r(t,t)$, $r_{0t} = r(0,t)$, $r_{1,0}(t,s) = \partial r(\alpha,\beta)/\partial\alpha|_{(t,s)}$, $r_{1,1}(t,s) = \partial^2 r(\alpha,\beta)/\partial\alpha\partial\beta\big|_{(t,s)}$, $r_{1,1}(t,s) = \partial\,\mathsf{E}\,\eta(t)\,\eta(s)/\partial t\partial s$, $t,s \in \left[0,h^{-1}\right]$ and $C(t) \stackrel{\mathrm{def}}{=} C_{th}(K)$, $t \in \left[0,h^{-1}\right]$, with $C_{th}(K)$ as in (A.10), so that $C(t) \equiv C(K)$, $\forall t \in \left[1, h^{-1}-1\right]$. Clearly, for any n,

$$m(t)=0, r(t,t)=r_t\equiv 1. \tag{A.17}$$

and it is easy to verify that for $\forall t \in \left[0, h^{-1}\right]$

$$r_{1,0}(t,t) = 0, \tag{A.18}$$

while for $v^2 = \mathsf{Var}\left\{\eta'(t)\,|\eta(0), \eta(t)\right\}$, see (15) in Zheng, Yang and Härdle (2010), $s, t \in \left[0, h^{-1}\right]$ and $|t-s| \geq 2$,

$$r_{st} = r_{1,0}(t,s) = 0,\ v^2 = r_{1,1}(t,t). \tag{A.19}$$

LEMMA A.11. *Under Assumptions (A1)-(A6)*

$$\lim_{n\to\infty} \sup_{t\in[0,h^{-1}]} |r_{1,1}(t,t) - C(t)| = 0. \tag{A.20}$$

There exist constants $0 < c < C < \infty$, $1 > \delta > 0$, *such that for large* n

$$\inf_{t,s\in[0,h^{-1}],|t-s|<2} r(t,s) \geq -1 + c > -1, \quad \sup_{2>|t-s|\geq\delta,\ t,s\in[0,h^{-1}]} r(t,s) \leq 1 - c < 1, \tag{A.21}$$

$$\begin{aligned} &\sup_{0<|t-s|<\delta, t,s\in[0,h^{-1}]} \max[r_{1,0}(t,s)/(t-s), \{1 - r^2(t,s)\}/(t-s)^2] \leq C, \\ &\inf_{0<|t-s|<\delta, t,s\in[0,h^{-1}]} \min[r_{1,0}(t,s)/(t-s), \{1 - r^2(t,s)\}/(t-s)^2] \geq c, \end{aligned} \tag{A.22}$$

$$\begin{aligned} &\sup_{0<|t-s|<\delta, t,s\in[0,h^{-1}]} \frac{r_{1,1}(t,t) - r_{1,0}^2(t,s)/\left(1-r^2\right)}{(t-s)^2} \leq C, \\ &\inf_{0<|t-s|<\delta, t,s\in[0,h^{-1}]} \frac{r_{1,1}(t,t) - r_{1,0}^2(t,s)/\left(1-r^2\right)}{(t-s)^2} \geq c, \end{aligned} \tag{A.23}$$

$$\sup_{|t-s|<2, t,s\in[0,h^{-1}]} |r_{1,0}^2(t,s)/\{1 - r^2(t,s)\}| \leq C \tag{A.24}$$

$$\inf_{|t-s|<2, t,s\in[0,h^{-1}]} \frac{|r_{1,0}(t,s)/(1+r)|}{\sqrt{r_{1,1}(t,t) - r_{1,0}^2(t,s)/(1-r^2)}} \geq c \tag{A.25}$$

PROOF. See Appendix C, Zheng, Yang and Härdle (2010). □

In what follows, the "double sum" method of Piterbarg (1996) will be applied to study the extreme value distribution of the sequence of Gaussian processes $\eta(t)$ over the growing interval $\left[0, h^{-1}\right]$. Partition the interval $\left[1, h^{-1}-1\right]$ as $1 = a_1 < b_1 < a_2 < b_2 < \cdots < a_N < b_N = h^1 - 1$, assuming $I_l = \left[a_l, b_l\right], l = 1, \cdots, N, I_l' = \left[b_l, a_{l+1}\right], l = 1, \cdots, N-1$ and the length of I_l and I_l' are λ_n and 2, respectively, where $\left(\lambda_n + 2\right) N = h^{-1}$ and $\lambda_n \to \infty, N \to \infty$ as $n \to \infty$.

LEMMA A.12. *Under Assumptions (A1)-(A6), for $u = u_n$ satisfying $2\sqrt{C\left(K\right)} N \lambda_n \varphi\left(u_n\right) \varphi\left(0\right) \to -\log\left(1-\alpha\right)$ with $C\left(K\right)$ given in (3.5)*

$$\lim_{n\to\infty} \mathrm{P}\{\sup_{t\in[0,1)\cup_{l=1}^{N-1} I_l' \cup (h^{-1}-1, h^{-1}]} \left|\eta\left(t\right)\right|\} \leq u = 1.$$

PROOF. In Lemma A.1, for $\forall \left[a, b\right] \subseteq \left[0, h^{-1}\right]$, one computes according to Cierco-Ayrolles et al (2003), Zheng, Yang and Härdle (2010)

$$\begin{aligned}
&\mathsf{E}\left[\left(U_u^\eta\left[a,b\right] + D_{-u}^\eta\left[a,b\right]\right) I_{\{|X(a)|\leq u\}}\right] = \qquad \text{(A.26)}\\
&2\varphi\left(u\right)\left\{\varphi\left(0\right)\int_a^b \sqrt{r_{1,1}\left(t,t\right)} dt - \int_a^b \left(\varphi\left(0\right)\sqrt{r_{1,1}\left(t,t\right)}\left[1 - \Phi\left\{\sqrt{r_{1,1}\left(t,t\right)}\sqrt{\frac{1-r_{at}}{1+r_{at}}}\frac{u}{v}\right\}\right]\right.\right.\\
&\left.+\frac{r_{1,0}\left(t,a\right)}{\sqrt{1-r_{at}^2}}\varphi\left(\sqrt{\frac{1-r_{at}}{1+r_{at}}}u\right)\Phi\left\{\frac{r_{1,0}\left(t,a\right)}{1+r_{at}}\frac{u}{v}\right\}\right) dt - \int_a^b \left(\varphi\left(0\right)\sqrt{r_{1,1}\left(t,t\right)}\times\right.\\
&\left.\left.\left[1 - \Phi\left\{\sqrt{r_{1,1}\left(t,t\right)}\sqrt{\frac{1+r_{at}}{1-r_{at}}}\frac{u}{v}\right\}\right] - \frac{r_{1,0}\left(t,a\right)}{\sqrt{1-r_{at}^2}}\varphi\left(\sqrt{\frac{1+r_{at}}{1-r_{at}}}u\right)\Phi\left\{\frac{-r_{1,0}\left(t,a\right)}{1-r_{at}}\frac{u}{v}\right\}\right) dt\right\}.
\end{aligned}$$

According to (A.20) and (A.24), it is clear that as $n \to \infty$,

$$\sup_{1\leq l\leq N-1} \mathsf{E}\left[\left(U_u^\eta\left[b_l, b_l+2\right] + D_{-u}^\eta\left[b_l, b_l+2\right]\right) I_{\{|X(b_l)|\leq u\}}\right] = \mathcal{O}\left\{\varphi\left(u\right)\right\}.$$

Hence, the upper bound of (A.1) shows that, if $2\sqrt{C\left(K\right)} N \lambda_n \varphi\left(u_n\right)\varphi\left(0\right) \to -\log\left(1-\alpha\right)$ as $n \to \infty$,

$$\sum\nolimits_{l=1}^{N-1} \mathrm{P}\{\sup_{I_{l'}'}\left|\eta\left(t\right)\right| > u\} = \mathcal{O}\left[2N\left\{1-\Phi\left(u\right)\right\}\right] + \mathcal{O}\left\{N\varphi\left(u\right)\right\} = \mathcal{o}(1), \qquad \text{(A.27)}$$

Similarly, while $t \in [0,1) \cup \left(h^{-1}-1, h^{-1}\right]$, one can show that

$$\mathrm{P}\{\sup_{t\in[0,1)\cup(h^{-1}-1,h^{-1}]}\left|\eta\left(t\right)\right| > u\} = \mathcal{O}\left\{1-\Phi\left(u\right)\right\} + \mathcal{O}\left\{\varphi\left(u\right)\right\} = \mathcal{o}(1). \qquad \text{(A.28)}$$

Finally, this lemma is proved by

$$\mathrm{P}\{\sup_{t\in[0,1)\cup_{l'=1}^{N} I_{l'}\cup(h^{-1}-1,h^{-1}]} |\eta(t)| > u\} \leq$$
$$\mathrm{P}\{\sup_{t\in[0,1)\cup(h^{-1}-1,h^{-1}]} |\eta(t)| > u\} + \sum_{l=1}^{N-1} \mathrm{P}\{\sup_{I_l'} |\eta(t)| > u\}.$$

□

LEMMA A.13. *Under Assumptions (A1)-(A6), for $u = u_n$ satisfying $2\sqrt{C(K)}N\lambda_n\varphi(u_n)\varphi(0) \to -\log(1-\alpha)$ with $C(K)$ given in (3.5),*

$$\lim_{n\to\infty} \mathrm{P}\{\sup_{\cup_{l=1}^{N} I_l} |\eta(t)| \leq u_n\} = 1-\alpha.$$

PROOF. First, in order to apply Lemma A.1, we rewrite

$$\mathsf{E}\left[\left(U_u^{\eta}[a_l,b_l] + D_{-u}^{\eta}[a_l,b_l]\right) I_{\{|X(a_l)|\leq u\}}\right] = \int_{a_l}^{a_l+2} + \int_{a_l+2}^{b_l} = I_{1l} + I_{2l}.$$

Similar to Lemma A.12, one also can show that as $n \to \infty$,

$$\sup_{1\leq l\leq N} I_{1l} = \mathcal{O}\{\varphi(u)\}. \tag{A.29}$$

Further, since $r_{a_l t} = r_{1,0}(t, a_l) = 0, v^2 = r_{1,1}(t,t)$ for $\forall t \in [a_l+2, b_l]$, see (A.19), one can simplify (A.26) as

$$I_{2l} = 2\varphi(u)\varphi(0)\int_{a_l+2}^{b_l}\sqrt{r_{1,1}(t,t)}dt - 4\varphi(u)\varphi(0)\{1-\Phi(u)\}\int_{a_l+2}^{b_l}\sqrt{r_{1,1}(t,t)}dt,$$

hence if $2\sqrt{C(K)}N\lambda_n\varphi(u_n)\varphi(0) \to -\log(1-\alpha)$, as $n\to\infty$,

$$\sup_{1\leq l\leq N}\left|I_{2l} - 2\lambda_n\varphi(u)\varphi(0)\sqrt{C(K)}\right| = \mathcal{O}\{\varphi(u)\lambda_n\}. \tag{A.30}$$

Therefore, (A.29) and (A.30) show that

$$\sup_{1\leq l\leq N}\left|\mathsf{E}\left[\left(U_u^{\eta}[a_l,b_l] + D_{-u}^{\eta}[a_l,b_l]\right) I_{\{|X(a_l)|\leq u\}}\right] - 2\lambda_n\varphi(u)\varphi(0)\sqrt{C(K)}\right| = \mathcal{O}\{\varphi(u)\lambda_n\}. \tag{A.31}$$

Now consider the second order moment and it is easy to verify that

$$\mathsf{E}\left(U_u^{\eta}[a_l,b_l] + D_{-u}^{\eta}[a_l,b_l]\right)^{[2]} = 2\,\mathsf{E}\,U_u^{\eta}[a_l,b_l]^{[2]} + 2\,\mathsf{E}\left(U_u^{\eta}[a_l,b_l]\,D_{-u}^{\eta}[a_l,b_l]\right).$$

By Lemma A.2 and the Hölder inequality

$$\mathsf{E}\, U_u^{\eta}\left[a_l, b_l\right]^{[2]} \tag{A.32}$$

$$= \int_{s,t\in[a_l,b_l]^2}\int_{(0,\infty)^2} |\eta_1'|\,|\eta_2'|\, p_{\eta_t;\eta_s;\eta_t';\eta_s'}\left(u;u;\eta_1';\eta_2'\right) d\eta_1' d\eta_2' dt ds$$

$$= \int_{s,t\in[a_l,b_l]^2} \mathsf{E}\{\eta'\left(t\right)^{+}\eta'\left(s\right)^{+}\left|\eta\left(t\right)=\eta\left(s\right)=u\right.\} p_{\eta(t),\eta(s)}\left(u,u\right) dt ds$$

$$\leq \int_{s,t\in[a_l,b_l]^2} \mathsf{E}^{1/2}[\{\eta'\left(t\right)^{+}\}^2\left|\eta\left(t\right)=\eta\left(s\right)=u\right.]\, \mathsf{E}^{1/2}[\{\eta'\left(s\right)^{+}\}^2\left|\eta\left(t\right)=\eta\left(s\right)=u\right.] p_{\eta(t),\eta(s)}\left(u,u\right) dt ds$$

$$= \int_{2\leq|s-t|,s,t\in[a_l,b_l]^2} + \int_{\delta\leq|s-t|<2,s,t\in[a_l,b_l]^2} + \int_{|s-t|<\delta,s,t\in[a_l,b_l]^2} = I_{1l} + I_{2l} + I_{3l},$$

where $p_{\eta(t),\eta(s)}\left(u,u\right) = (2\pi\sqrt{1-r^2})^{-1}\exp\left\{-u^2/\left(1+r\right)\right\}$, see Azaïs and Wschebor (2009) p.96, Gaussian Rice Formula, and $\delta \in (0,1)$ which does not depend on n, see Lemma A.11.

For I_{1l}, it is cleat that

$$\mathsf{E}\left[\{\eta'\left(t\right)^{+}\}^2\left|\eta\left(t\right)=\eta\left(s\right)=u\right.\right] \leq \mathsf{E}\left[\{\eta'\left(t\right)\}^2\left|\eta\left(t\right)=\eta\left(s\right)=u\right.\right]$$

$$\leq \mathsf{E}^2\left\{\eta'\left(t\right)\left|\eta\left(t\right)=\eta\left(s\right)=u\right.\right\} + \mathsf{Var}\left\{\eta'\left(t\right)\left|\eta\left(t\right)=\eta\left(s\right)=u\right.\right\},$$

$$\mathsf{E}\left\{\eta'\left(t\right)\left|\eta\left(t\right)=\eta\left(s\right)=u\right.\right\} = r_{1,0}\left(t,s\right)u/\left(1+r\right), \tag{A.33}$$

$$\mathsf{Var}\left\{\eta'\left(t\right)\left|\eta\left(t\right)=\eta\left(s\right)=u\right.\right\} = r_{1,1}\left(t,t\right) - r_{1,0}^2\left(t,s\right)/\left(1-r^2\right), \tag{A.34}$$

see Azaïs and Wschebor (2009) p.96. If $|t-s| \geq 2$, then $r_{st} = r_{1,0}\left(t,s\right) = 0$ which implies that $\mathsf{E}\left\{\eta'\left(t\right)\left|\eta\left(t\right)=\eta\left(s\right)=u\right.\right\} = 0$ and $\mathsf{Var}\left\{\eta'\left(t\right)\left|\eta\left(t\right)=\eta\left(s\right)=u\right.\right\} = r_{1,1}\left(t,t\right)$. Hence

$$I_{1l} \leq \int_{2\leq|s-t|,s,t\in[a_l,b_l]^2} \sqrt{r_{1,1}\left(t,t\right)}\sqrt{r_{1,1}\left(s,s\right)}\frac{1}{2\pi}\exp\left(-u^2\right) dt ds, \tag{A.35}$$

which implies that

$$\sup_{1\leq l\leq N} I_{1l} = \mathcal{O}\{\varphi^2\left(u\right)\lambda_n^2\}. \tag{A.36}$$

For I_{2l}, similarly,

$$I_{2l} \leq \int_{\delta\leq|s-t|<2,s,t\in[a_l,b_l]^2} \{r_{1,0}^2\left(t,s\right)u^2/\left(1+r\right)^2 + r_{1,1}\left(t,t\right)\}^{1/2}\times \tag{A.37}$$

$$\{r_{1,0}^2\left(s,t\right)u^2/\left(1+r\right)^2 + r_{1,1}\left(s,s\right)\}^{1/2}\frac{1}{2\pi\sqrt{1-r^2}}\exp\{-u^2/\left(1+r\right)\} dt ds.$$

By (A.21), for large n, $\exists c>0$ such that $\sup_{|t-s|\geq\delta>0}(1+r)\leq 2-c$ and $\inf_{|t-s|\geq\delta>0}\left|1-r^2\right|\geq c>0$, so $\exists$ constants L_1, $K_1>0$ such that

$$\sup_{1\leq l\leq N} I_{2l}\leq L_1\varphi\left\{(1+K_1)\,u\right\}\lambda_n. \tag{A.38}$$

One can bound I_{3l} using the inequalities (4.10) and (4.11), Azaïs and Wschebor (2009) p.97, i.e., for $Z\sim \mathrm{N}\left(\mu,\sigma^2\right)$, if $\mu>0$, $\mathsf{E}\left(Z^+\right)^2\leq\mu^2+\sigma^2$ and if $\mu<0$, $\mathsf{E}\left(Z^+\right)^2\leq\left(\mu^2+\sigma^2\right)\left\{1-\Phi\left(-\mu/\sigma\right)\right\}+\mu\sigma\varphi\left(\mu/\sigma\right)$. Since $\eta'(t)$, $\eta'(s)$ conditioning on $\eta(t)=\eta(s)=u$ have a joint Gaussian distribution, see Azaïs and Wschebor (2009) p.96, we denote

$$\mu_1=\mathsf{E}\left\{\eta'(s)\,|\eta(t)=\eta(s)=u\right\},\ \mu_2=\mathsf{E}\left\{\eta'(t)\,|\eta(t)=\eta(s)=u\right\}, \tag{A.39}$$

$$\sigma_1^2=\mathsf{Var}\left\{\eta'(s)\,|\eta(t)=\eta(s)=u\right\},\ \sigma_2^2=\mathsf{Var}\left\{\eta'(t)\,|\eta(t)=\eta(s)=u\right\}. \tag{A.40}$$

Next, we claim that while $0<|s-t|<\delta$, μ_1 and μ_2 have opposite signs. In fact, if $0<|s-t|<\delta$, by (A.22), for large n, $r_{1,0}(t,s)\sim(t-s)$ and $r_{1.0}(s,t)\sim(s-t)$ and by (A.21), $\inf_{|t-s|<\delta}(1+r)\geq c>0$, which imply that $\mu_1\mu_2<0$, see (A.33). Further, according to (A.25), (A.33) and (A.34), for large n, $\exists$ constant $L_2>$ such that $\inf_{|t-s|<2,t,s\in[0,h^{-1}]}|\mu_2|\,\sigma_2^{-1}\geq L_2u$. Without loss of generality, by (A.39) and (A.40), let $\mu_1>0>\mu_2$, then

$$\begin{aligned} I_{3l}\leq \int_{|s-t|<\delta,s,t\in[a_l,b_l]^2}&\sqrt{\mu_1^2+\sigma_1^2}\left[\left(\mu_2^2+\sigma_2^2\right)\left\{1-\Phi\left(-\mu_2/\sigma_2\right)\right\}+\right.\\ &\left.\mu_2\sigma_2\varphi\left(\mu_2/\sigma_2\right)\right]^{1/2}\frac{1}{2\pi\sqrt{1-r^2}}\exp\{-u^2/\left(1+r\right)\}dtds. \end{aligned}$$

It follows from (A.22) and (A.23) that for large enough n, $\exists$ constants $L_3, L_4, L_5, K_2>0$ such that

$$\begin{aligned} \sup_{1\leq l\leq N} I_{3l}\leq &\int_{|s-t|<\delta,s,t\in[a_l,b_l]^2} L_3\sqrt{(s-t)^2u^2+(s-t)^2}\times \\ &[\{(s-t)^2u^2+(s-t)^2\}\{1-\Phi(L_2u)\}-(s-t)^2u\varphi(-L_2u)]^{1/2}|s-t|^{-1}\varphi(u)\,dsdt\\ &\leq L_5\delta\varphi\left\{(1+K_2)\,u\right\}\lambda_n. \end{aligned} \tag{A.41}$$

Hence, if $2\sqrt{C(K)}N\lambda_n\varphi(u_n)\varphi(0)\to-\log(1-\alpha)$, as $n\to\infty$, (A.36), (A.38) and (A.41) imply that

$$\sup_{1\leq l\leq N}\mathsf{E}\,U_u^\eta\left[a_l,b_l\right]^{[2]}=\mathcal{O}\{\varphi\left(u\right)\lambda_n\}.$$

Similarly, one has $\mathsf{E}\left(U_u^\eta\left[a_l,b_l\right]D_{-u}^\eta\left[a_l,b_l\right]\right)=o\left\{\varphi\left(u\right)\lambda_n\right\}$ and then

$$\sup_{1\le l\le N}\mathsf{E}\left(U_u^\eta\left[a_l,b_l\right]+D_{-u}^\eta\left[a_l,b_l\right]\right)^{[2]}=\mathcal{O}\{\varphi\left(u\right)\lambda_n\}\,. \tag{A.42}$$

In fact, by Lemma A.1, (A.31) and (A.42) show that, as $n\to\infty$,

$$\mathrm{P}\{\sup_{I_l}\left|\eta\left(t\right)\right|>u\}=2\sqrt{C\left(K\right)}\varphi\left(u\right)\varphi\left(0\right)\lambda_n+\mathcal{O}\{\varphi\left(u\right)\lambda_n\}\,. \tag{A.43}$$

Finally, since $\mathsf{E}\,\eta\left(t\right)\eta\left(s\right)=0$ while $t\in I_l,\ s\in I_m,l\ne m$, then $\eta\left(t\right),\eta\left(s\right)$ for $t\in I_l,\ s\in I_m,l\ne m$ are independent Gaussian processes and hence

$$\begin{aligned}
&\mathrm{P}\{\sup_{\cup_{l=1}^N I_l}\left|\eta\left(t\right)\right|\le u\}=\prod\nolimits_{l=1}^N\mathrm{P}\{\sup_{I_l}\left|\eta\left(t\right)\right|\le u\}\\
&=\prod\nolimits_{l=1}^N\left[1-\mathrm{P}\left\{\sup_{I_l}\left|\eta\left(t\right)\right|>u\right\}\right]=\exp(\sum\nolimits_{l=1}^N\log\left[1-\mathrm{P}\{\sup_{I_l}\left|\eta\left(t\right)\right|>u\}\right])\\
&=\exp(\sum\nolimits_{l=1}^N\log[1-2\sqrt{C\left(K\right)}\varphi\left(u\right)\varphi\left(0\right)\lambda_n+\mathcal{O}\{\varphi\left(u\right)\lambda_n\}])\\
&=\exp[-2N\sqrt{C\left(K\right)}\varphi\left(u\right)\varphi\left(0\right)\lambda_n+\mathcal{O}\{N\varphi\left(u\right)\lambda_n\}].
\end{aligned}$$

Since $2\sqrt{C\left(K\right)}N\lambda_n\varphi\left(u\right)\varphi\left(0\right)\to-\log\left(1-\alpha\right)$ as $n\to\infty$, then it follows from the definitions of N,λ_n,u_n that $\lim_{n\to\infty}\mathrm{P}\{\sup_{\cup_{l=1}^N I_l}\left|\eta\left(t\right)\right|\le u\}=1-\alpha$. □

The quantile $Q_h\left(\alpha\right)$ given in (3.5) satisfies $2\sqrt{C\left(K\right)}N\lambda_n\varphi\left\{Q_h\left(\alpha\right)\right\}\varphi\left(0\right)\to-\log\left(1-\alpha\right)$, as $n\to\infty$, then Lemmas A.12 and A.13 imply that $\lim_{n\to\infty}\mathrm{P}\{\sup_{[0,1]}\left|\eta\left(x\right)\right|\le Q_h\left(\alpha\right)\}=1-\alpha$, i.e.,

$$\lim_{n\to\infty}\mathrm{P}\left[a_h\{\sup_{[0,1]}\left|\eta\left(x\right)\right|-a_h\}-\log\{\sqrt{C\left(K\right)}/\left(2\pi\right)\}\le-\log\{-\log\sqrt{1-\alpha}\}\right]=1-\alpha. \tag{A.44}$$

In particular, $\sup_{[0,1]}\left|\eta\left(x\right)\right|=\mathcal{O}_p(\sqrt{\log n})$.

LEMMA A.14. *Under Assumptions (A1)-(A6), let* $\Delta_{3,n}\left(x\right)=\widetilde{\sigma}_n\left(x\right)\sigma_n^{-1}\left(x\right)-1,x\in[0,1]$, *then* $\Delta_{3,n}\left(x\right)=\mathcal{U}\left(h\right)+\mathcal{U}_{a.s.}\{\sqrt{\log n/(nh^2)}\}$ *and for* $\widehat{\varepsilon}\left(x\right),\ \sigma_n^2\left(x\right)$ *given in (3.4) as* $n\to\infty$

$$\begin{aligned}
\sup_{[0,1]}\left|\sigma_n^{-1}\left(x\right)\left\{\widehat{\varepsilon}_{N_T}\left(x\right)+\sum\nolimits_{k=1}^\kappa\widehat{\xi}_k(x)\right\}-\eta\left(x\right)\right|&=\sup_{[0,1]}\left|\Delta_{3,n}\left(x\right)\right|\left|\eta\left(x\right)\right|\\
&=\mathcal{O}_p\{h\sqrt{\log n}+\sqrt{\log^2 n/(nh^2)}\}.
\end{aligned}$$

PROOF. It follows from the definition of $\eta(x)$ given in (A.16) that $|\Delta_{3,n}(x)| = \left|\widetilde{\sigma}_n(x)\sigma_n^{-1}(x) - 1\right| \leq \left|\widetilde{\sigma}_n^2(x)\sigma_n^{-2}(x) - 1\right|$ in which

$$\widetilde{\sigma}_n^2(x) = f^{-2}(x)D_x^{-2}(K)N_{\mathrm{T}}^{-1}\{N_{\mathrm{T}}^{-1}\sum\nolimits_{i,j}R_{ij,\varepsilon}^2(x) + N_{\mathrm{T}}^{-1}\sum\nolimits_{k,i}R_{ik,\xi_k}^2(x)\}.$$

Lemma A.10 implies $N_{\mathrm{T}}^{-1}\sum_{i,j}R_{ij,\varepsilon}^2(x) = \mathsf{E}\,R_{ij,\varepsilon}^2(x) + \mathcal{U}_{a.s}\{\sqrt{\log n/(nh^2)}\}$ and $N_{\mathrm{T}}^{-1}\sum_{k,i}R_{ik,\xi}^2(x) = (\mathsf{E}\,N_1)^{-1}\sum_{k=1}^{\kappa}\mathsf{E}\,R_{ik,\xi_k}^2(x) + \mathcal{U}_{a.s.}\{\sqrt{\log n/(nh^2)}\}$. Also, one has $N_{\mathrm{T}} = n\,\mathsf{E}\,N_1 + \mathcal{U}_{a.s.}(\sqrt{\log n/n})$ and $\mathsf{E}\,R_{ij,\varepsilon}^2(x) + (\mathsf{E}\,N_1)^{-1}\sum_{k=1}^{\kappa}\mathsf{E}\,R_{ik,\xi_k}^2(x) = \mathcal{U}\left(h^{-1}\right)$. Therefore,

$$\widetilde{\sigma}_n^2(x) = f^{-2}(x)D_x^{-2}(K)\left[(n\,\mathsf{E}\,N_1)^{-1} + \mathcal{U}_{a.s.}(\sqrt{\log n/n})\right]\times$$
$$[\mathsf{E}\,R_{ij,\varepsilon}^2(x) + (\mathsf{E}\,N_1)^{-1}\sum\nolimits_{k=1}^{\kappa}\mathsf{E}\,R_{ik,\xi_k}^2(x) + \mathcal{U}_{a.s.}\{\sqrt{\log n/(nh^2)}\}]$$
$$= f^{-2}(x)D_x^{-2}(K)(n\,\mathsf{E}\,N_1)^{-1}\{\mathsf{E}\,R_{ij,\varepsilon}^2(x) + (\mathsf{E}\,N_1)^{-1}\sum\nolimits_{k=1}^{\kappa}\mathsf{E}\,R_{ik,\xi_k}^2(x)\} +$$
$$\mathcal{U}_{a.s.}(\sqrt{\log n/n})\{\mathsf{E}\,R_{ij,\varepsilon}^2(x) + (\mathsf{E}\,N_1)^{-1}\sum\nolimits_{k=1}^{\kappa}\mathsf{E}\,R_{ik,\xi_k}^2(x)\} = \sigma_n^2(x) + \mathcal{U}_{a.s.}\{\sqrt{\log n/(nh^2)}\},$$

which implies that $\widetilde{\sigma}_n^2(x)\sigma_n^{-2}(x) = 1 + \mathcal{U}_{a.s.}\{\sqrt{\log n/(nh^2)}\}$ and hence this lemma holds. □

PROOF OF PROPOSITION 3.1. The proof is trivial. □

PROOF OF THEOREM 3.1. The decomposition (A.2) implies that

$$\sigma_n^{-1}(x)\{\widehat{m}(x) - m(x)\} = \sigma_n^{-1}(x)\{\widetilde{m}(x) - m(x)\} + \sigma_n^{-1}(x)\widetilde{e}(x). \tag{A.45}$$

As (A.44) implies that $\sup_{[0,1]}|\eta(x)| = \mathcal{O}_p(\sqrt{\log n})$, Lemma A.14 leads to

$$\sup\nolimits_{[0,1]}\sigma_n^{-1}(x)\left|\widehat{\varepsilon}_{N_{\mathrm{T}}}(x) + \sum\nolimits_{k=1}^{\kappa}\widehat{\xi}_k(x)\right| = \mathcal{O}_p(\sqrt{\log n}).$$

and hence by Lemma A.9, $\sup_{[0,1]}\sigma_n^{-1}(x)\left|\widehat{\varepsilon}(x) + \sum_{k=1}^{\kappa}\widehat{\xi}_k(x)\right| = \mathcal{O}_p(\sqrt{\log n})$. Therefore, Lemma A.8 implies that

$$\sup\nolimits_{[0,1]}\sigma_n^{-1}(x)\left|\widetilde{e}(x) - \{\widehat{\varepsilon}(x) + \sum\nolimits_{k=1}^{\kappa}\widehat{\xi}_k(x)\}\right| = \mathcal{O}_p\{h\sqrt{\log n} + \sqrt{\log^2 n/(nh^2)}\}. \tag{A.46}$$

It follows from (A.46), Lemmas A.9 and A.14 that for $t' > 2/5$ (assumption A5),

$$\sup\nolimits_{[0,1]}\left|\sigma_n^{-1}(x)\left|\widetilde{e}(x)\right| - \left|\eta(x)\right|\right| = \mathcal{O}_p\{h\sqrt{\log n} + \sqrt{\log^2 n/(nh^2)} + \sqrt{h}n^{-t'+1/2}\},$$

Further, (A.46) and Lemma (A.7) warrants that

$$\sup_{[0,1]} \left|\sigma_n^{-1}(x)\left|\widehat{m}(x) - m(x)\right| - \left|\eta(x)\right|\right| = \tag{A.47}$$
$$\mathcal{O}_p\{\sqrt{n}h^{5/2} + h\sqrt{\log n} + \sqrt{\log^2 n/(nh^2)} + \sqrt{h}n^{-t'+1/2}\},$$

and therefore

$$a_h \sup_{x\in[0,1]} \left|\sigma_n^{-1}(x)\left|\widehat{m}(x) - m(x)\right| - \left|\eta(x)\right|\right| \tag{A.48}$$
$$= \mathcal{O}_p[\sqrt{\log h^{-1}}\{\sqrt{n}h^{5/2} + h\sqrt{\log n} + \sqrt{\log^2 n/(nh^2)} + \sqrt{h}n^{-t'+1/2}\}] = \mathcal{o}_p(1).$$

Finally, by Slutsky's Theorem, (A.44) and (A.48) show that

$$\lim_{n\to\infty} \mathrm{P}\left[a_h\{\sup_{[0,1]} \sigma_n^{-1}(x)\left|\widehat{m}(x) - m(x)\right| - a_h\} - \log\{\sqrt{C(K)}/(2\pi)\} \le -\log\left\{-\log\sqrt{1-\alpha}\right\}\right]$$
$$= 1-\alpha,$$

which is

$$\lim_{n\to\infty} \mathrm{P}\{\sup_{x\in[0,1]} \sigma_n^{-1}(x)\left|\widehat{m}(x) - m(x)\right| \le Q_h(\alpha)\} = 1-\alpha.$$

□

REFERENCES

Azaïs, J.M. and Wschebor, M. (2009). *Level Sets and Extrema of Random Processes and Fields.* John Wiley & Sons, Inc., Hoboken, New Jersey.

Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction.* Springer-Verlag, New York.

Cai, T. and Hall, P. (2006) Prediction in functional linear regression. *Annals of Statistics* **34**, 2159-2179.

Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica* **13** 571–591.

Cierco-Ayrolles, C., Croquette, A. and Delmas, C. (2003). Computing the distribution of the maximum of Gaussian random processes. *Methodology and Computing in Applied Probability*, **5**, 427-438.

Csőrgő, M. and Révész, P. (1981) *Strong Approximations in Probability and Statistics.* Academic Press, New York-London.

Claeskens, G. and Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Annals of Statistics* **31**, 1852-1884.

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

Fan, J. and Yao, Q. (2005). *Nonlinear Time Series.* Springer-Verlag, New York.

Fan, J. and Zhang, W. Y. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics* **27**, 715-731.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice.* Springer Series in Statistics, Springer: Berlin.

Hall, P., Müller, H. G. and Wang, J. L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Annals of Statistics* **34**, 1493-1517.

James, G. M., Hastie, T. and Sugar, C. (2000). Principal Component Models for Sparse Functional Data. *Biometrika* **87**, 587–602.

Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. To appear in *Annals of Statistics*.

Lucas, P. W. and Diggle, P. J. (1997). The use of longitudinal data analysis to study the multi-seasonal growth responses of Norway and Sitka spruce to summer exposure to ozone: implications for the determination of critical levels. *New Phytol.* **137**, 315–323.

Ma, S., Yang, L. and Carroll, R. J. (2010). Simultaneous confidence band for sparse longitudinal regression. *Statistica Sinica.* **forthcoming**.

Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B* **68**, 179–199.

Müller, H. G. (2009). Functional Modeling of Longitudinal data. *Longitudinal Data Analysis, Handbooks of Modern Statistical Methods.* Wiley, New York, 223–252.

Müller, H. G., Stadtmüller, U. (2005). Generalized functional linear models. *Annals of Statistics*

33, 774-805.

Piterbarg, V.I. (1996). *Asymptotic Methods in the Theory of Gaussian Processes and Fields.* American Mathematical Society.

Song, Q. and Yang, L. (2009). Spline confidence bands for variance function. *Journal of Nonparametric Statistics* **21**, 589–609.

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

Wang, N., Carroll, R. J. and Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *Journal of the American Statistical Association* **100**, 147-157.

Wu, W. and Zhao, Z. (2007). Inference of trends in time series. *Journal of the Royal Statistical Society, Series B* **69**, 391-410.

Xia, Y. (1998). Bias-Corrected Confidence Bands in Nonparametric Regression. *Journal of the Royal Statistical Society. Series B* **60**, 797-811.

Yao, F., Müller, H. G. and Wang, J. L. (2005a). Functional linear regression analysis for longitudinal data. *Annals of Statistics* **33**, 2873–2903.

Yao, F., Müller, H. G. and Wang, J. L. (2005b). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577–590.

Yao. F. (2007). Asymptotic distributions of nonparametric regression estimators for longitudinal or functional data. *Journal of Multivariate Analysis* **98**, 40–56.

Zhang, J. T. and Chen, J. (2007). Statistical inferences for functional data. *Annals of Statistics* **35**, 1052-1079.

Zheng, S., Yang, L. and Härdle, W. (2010). Supplement to "A Confidence Corridor of Sparse Longitudinal Date Curves". Manuscript.

Zhao, Z. and Wu, W. (2008). Confidence bands in nonparametric time series regression. *Annals of Statistics* **36**, 1854-1878.

Shuzhuan Zheng
Department of Statistics and Probability,
Michigan State University,
East Lansing, MI 48824
E-mail: zheng@stt.msu.edu

Lijian Yang
Center for Advanced Statistics and Econometrics Research,
Soochow University,
Suzhou 215006,
People's Republic of China
and
Department of Statistics and Probability,
Michigan State University,
East Lansing, MI 48824
E-mail: yang@stt.msu.edu

Wolfgang K. Härdle
Center for Applied Statistics and Economics,
Humboldt-Universität zu Berlin,
Unter den Linden 6,
10099 Berlin, Germany
and
Graduate Institute of Statistcis,
National Central University,
Jhongli 32049, Taiwan
E-mail: stat@wiwi.hu-berlin.de

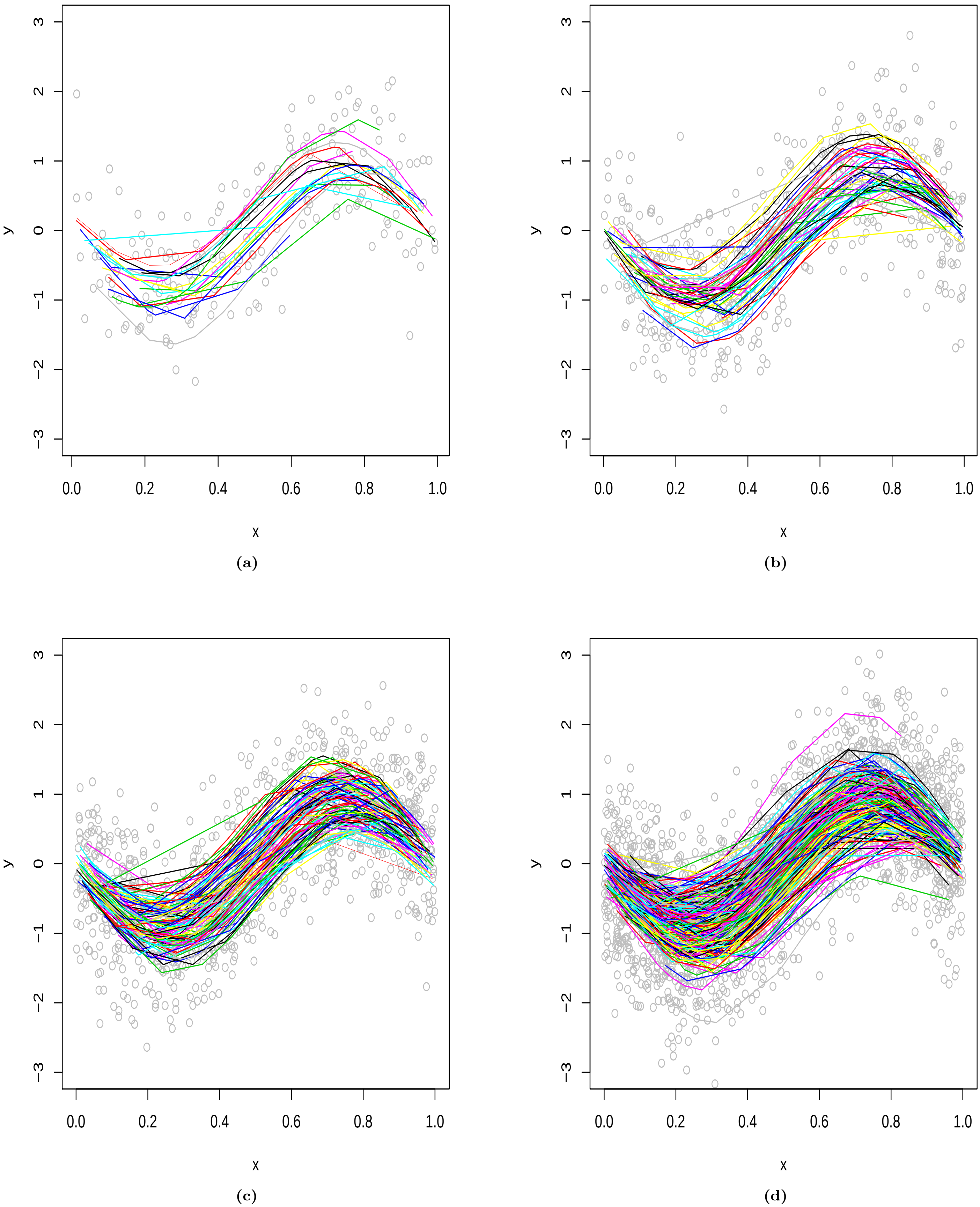

Fig 1. *Plots of simulated data (circles) and trajectories (solid lines): (a) $n = 20$, (b) $n = 50$, (c) $n = 100$, (d) $n = 200$.*

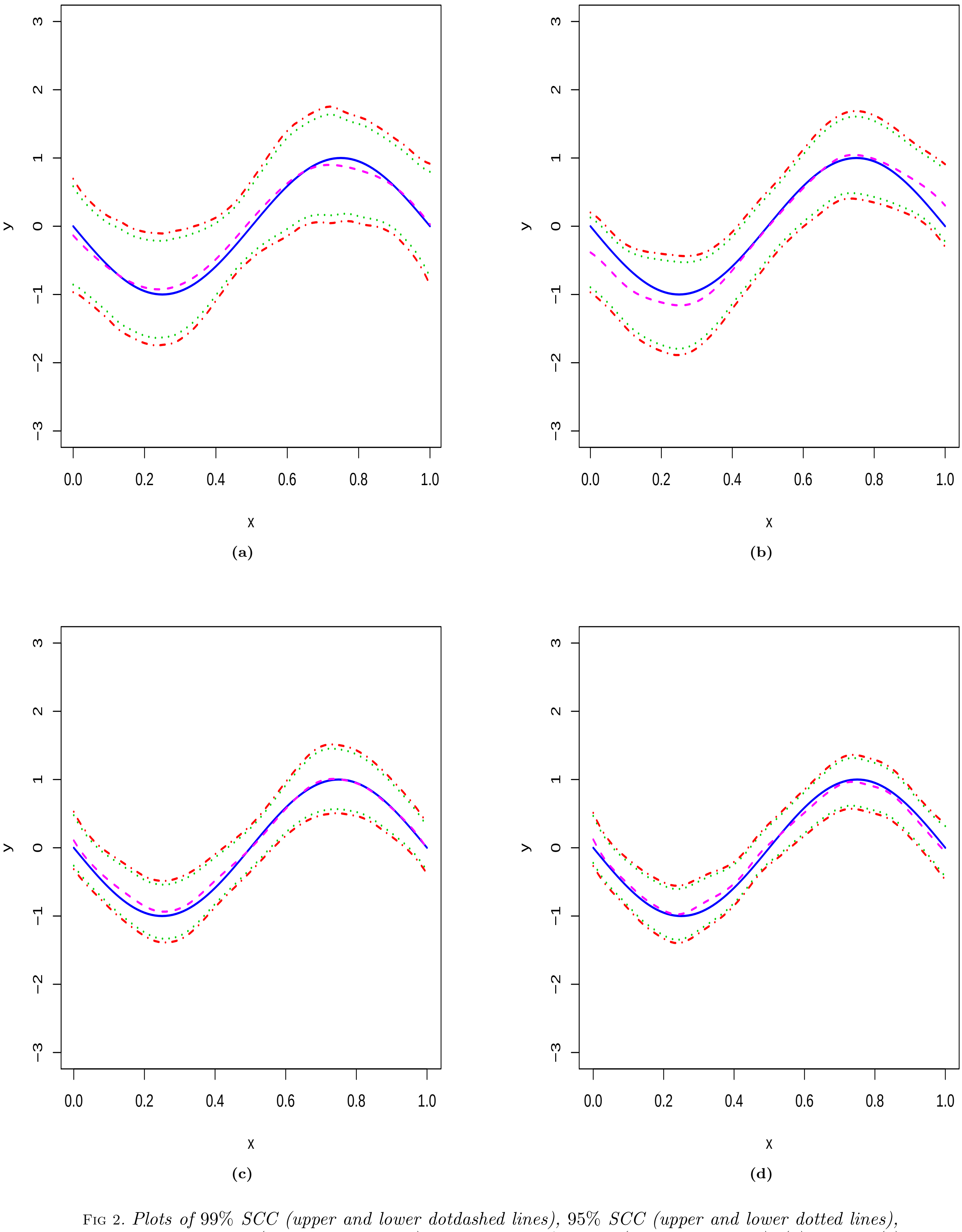

Fig 2. *Plots of 99% SCC (upper and lower dotdashed lines), 95% SCC (upper and lower dotted lines), local linear estimator (median dashed line) and true mean function (median solid line): (a) $n = 20$, (b) $n = 50$, (c) $n = 100$, (d) $n = 200$.*

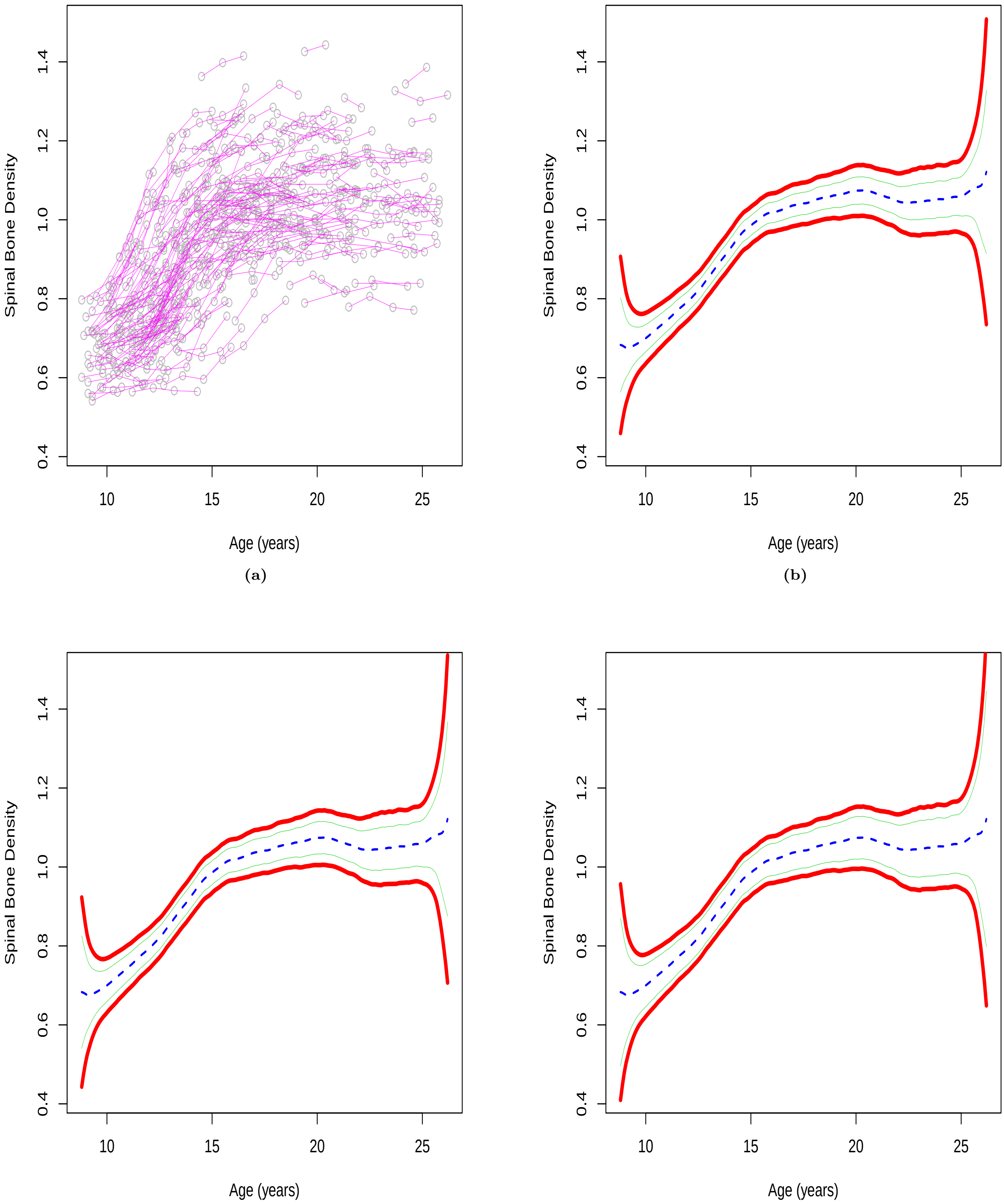

Fig 3. *Plots of the growth curve data, local linear estimator (median dashed line), SCC (upper and lower thick lines) and SCI (upper and lower solid lines): (a) the data, (b) confidence level* $= 90\%$, *(c) confidence level* $= 95\%$, *(d) confidence level* $= 99\%$

SFB 649 Discussion Paper Series 2011

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Localising temperature risk" by Wolfgang Karl Härdle, Brenda López Cabrera, Ostap Okhrin and Weining Wang, January 2011.

002 "A Confidence Corridor for Sparse Longitudinal Data Curves" by Shuzhuan Zheng, Lijian Yang and Wolfgang Karl Härdle, January 2011.

SFB 649, Ziegelstraße 13a, D-10117 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".