Confidence surfaces for the mean of locally stationary functional time series

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany

Weichi Wu
Tsinghua University
Center for Statistical Science
Department of Industrial Engineering
100084 Beijing China

Abstract

The problem of constructing a simultaneous confidence band for the mean function of a locally stationary functional time series $\{X_{i,n}(t)\}_{i=1,\dots n}$ is challenging as these bands can not be built on classical limit theory. On the one hand, for a fixed argument t of the functions $X_{i,n}$, the maximum absolute deviation between an estimate and the time dependent regression function exhibits (after appropriate standardization) an extreme value behaviour with a Gumbel distribution in the limit. On the other hand, for stationary functional data, simultaneous confidence bands can be built on classical central theorems for Banach space valued random variables and the limit distribution of the maximum absolute deviation is given by the sup-norm of a Gaussian process. As both limit theorems have different rates of convergence, they are not compatible, and a weak convergence result, which could be used for the construction of a confidence surface in the locally stationary case, does not exist.

In this paper we propose new bootstrap methodology to construct a simultaneous confidence band for the mean function of a locally stationary functional time series, which is motivated by a Gaussian approximation for the maximum absolute deviation. We prove the validity of our approach by asymptotic theory, demonstrate good finite sample properties by means of a simulation study and illustrate its applicability analyzing a data example.

AMS subject classification: 62R10; 62M10; 62M20
Keywords and phrases: locally stationary time series, functional data, confidence bands, Gaussian approximation

1 Introduction

In the big data era data gathering technologies provide enormous amounts of data with complex structure. In many applications the observed high-dimensional data exhibits certain degrees of dependence and smoothness and thus may naturally be regarded as discretized functions. A major tool for the statistical analysis of such data is functional data analysis (FDA) which has found considerable attention in the statistical literature (see, for example, the monographs of Bosq, 2000; Ramsay and Silverman, 2005; Ferraty and Vieu, 2010; Horváth and Kokoszka, 2012; Hsing and Eubank, 2015, among others). In FDA the considered parameters, such as the mean or the (auto-)covariance (operator) are functions themselves, which makes the development of statistical methodology challenging. Most of the literature considers Hilbert space-based methodology for which there exists by now a well developed theory. In particular, this approach allows the application of dimension reduction techniques such as (functional) principal components. On the other hand, in many applications it is reasonable to assume that functions are at least continuous (see also Ramsay and Silverman, 2005, for a discussion of the integral role of smoothness) and fully functional methods can prove advantageous. More recently, Aue et al. (2018), Bucchia and Wendler (2017) and Horváth et al. (2014) discuss fully functional methodology in a Hilbert space framework and Dette et al. (2020); Dette and Kokot (2020) develop inference methods for functional data in a Banach space framework.

In this paper we are interested in statistical inference regarding the mean functions of a not necessarily stationary functional time series $(X_{i,n})_{i=1,\dots,n}$ in the space $L^2[0,1]$ of square integrable functions on the interval $[0,1]$. More precisely, we consider the model

$$X_{i,n}(t) = m(\tfrac{i}{n}, t) + \varepsilon_{i,n}(t) \ , \quad i = 1, \dots, n \ , \tag{1.1}$$

where $(\varepsilon_{i,n})_{i=1,\dots,n}$ is a centred locally stationary process (see Section 2 for a precise definition) and $m : [0,1] \times [0,1] \to \mathbb{R}$ is a smooth mean function. This means that at each time point "i" we observe a function $t \to X_{i,n}(t)$ with mean function $t \to m(\frac{i}{n}, t)$, and our goal is a simultaneous confidence surface for the (time dependent) mean function $(u,t) \to m(u,t)$ of the locally stationary process $\{X_{i,n}(t)\}_{i=1,\dots,n}$.

Locally stationary functional time series have found considerable interest in the recent literature (see, for example, van Delft and Eichler, 2018; Aue and van Delft, 2020; Bücher et al., 2020; Kurisu, 2021a,b; van Delft and Dette, 2021). In the context of functional data analysis the "regression" model (1.1) has been mostly investigated in the stationary case, where $m(u,t) = m(t)$ (see, for example, Berkes et al., 2009; Horváth et al., 2013; Dette et al., 2020, among many others). In this case the model reduces to $X_i(t) = m(t) + \varepsilon_i$ with a stationary error process $(\varepsilon_i)_{i=1,\dots,n}$, and a common approach is based on a functional central limit theorem of the form

$$\sqrt{n}(\hat{m} - m) \rightsquigarrow \mathbb{G} \ , \tag{1.2}$$

where $\hat{m}$ is an appropriate estimate of the mean function and $\mathbb{G}$ is a centered Gaussian process in the space of continuous functions $C[0,1]$ (here the symbol $\rightsquigarrow$ denotes weak convergence in $C[0,1]$).

Several authors have used results of this type to construct (asymptotic) simultaneous confidence bands under different model assumptions. For example, Degras (2011, 2017) and Cao et al. (2012) assume that a random sample of functions is observed on a fine grid and define the band by

$$\Big\{ f : [0,1] \to \mathbb{R} \,\Big|\, \sup_{t\in[0,1]} \mid \hat{m}(t) - f(t) \mid \leq \tfrac{q_{1-\alpha}}{\sqrt{n}} \Big\}$$

where $q_{1-\alpha}$ is an appropriate quantile (obtained from a weak convergence result of the form (1.2) or by resampling). On the other hand Dette et al. (2020) assume that the full trajectory can be observed and use the estimate $\hat{m} = \bar{X}_n$ and multiplier bootstrap for this purpose. More recently, alternative simultaneous confidence (asymptotic) bands have been constructed by Liebl and Reimherr (2019); Telschow and Schwartzman (2022) using (1.2) and the Gaussian Kinematic formula.

Alternatively, if the data in model (1.1) are no functions and do not depend on the variable t we obtain the classical nonparametric regression model $X_{i,n} = m(\frac{i}{n}) + \varepsilon_{i,n}$, where the problem of constructing confidence bands for the regression function $u \to m(u)$ has a long history. Most authors consider the case of independent identically distributed errors (see Konakov and Piterbarg, 1984; Xia, 1998; Proksch, 2014, among others) and a kernel estimate for the regression function. The simultaneous (asymptotic) confidence band is based on a weak convergence result of the type

$$\sqrt{ab_n \log(bb_n^{-1})} \sup_{t\in[0,1]} \mid \hat{m}(t) - \mathbb{E}[\hat{m}(t)] \mid -2\log(bb_n^{-1}) \xrightarrow{\mathcal{D}} G \tag{1.3}$$

where the symbol $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution, G denotes a Gumbel distribution, b_n is the bandwidth of the nonparametric estimator and a, b are known constants (a depends on the variance of the errors). Wu and Zhao (2007) develop a simultaneous confidence band for the regression function in a model with a stationary series errors and Zhou (2010) derives a simultaneous confidence band in quantile regression model with locally stationary errors. A similar extreme value type result for the absolute maximum of a normalized deviation is used by Ma et al. (2012) and Zheng et al. (2014) to develop a confidence band for sparse functional data.

As the limit theorems (1.2) and (1.3) are not compatible, a weak convergence result for the maximum absolute deviation between the estimate and the regression function in the locally stationary case does not exist. As a consequence, the construction of simultaneous confidence surfaces for the regression function $(u, t) \to m(u, t)$ in model (1.1) is more challenging. In this paper we propose a general solution to this problem, which is not based on weak convergence results. As an alternative to “classical” limit theory (which does not exist in the present situation) we develop Gaussian approximations for the maximum absolute deviation between the estimate and the regression function. These results are then used to construct a non-standard multiplier bootstrap procedure for the construction of simultaneous confidence bands for the mean function of a locally stationary functional time series. We prove the validity of our approach by asymptotic theory, demonstrate good finite sample properties by means of a simulation study and illustrate its applicability analyzing a data example. As a by-product of our approach, we also derive new confidence bands

for the functions $t \to m(u,t)$ (for fixed u) and $u \to m(u,t)$ (for fixed t), which provide efficient alternatives to the commonly used confidence bands for stationary functional data or real valued locally stationary data, respectively.
The remaining part of the paper is organized as follows. The statistical model is introduced in Section 2. In Section 3 and 4 we develop Gaussian approximations, use these results for the construction of simultaneous confidence bands and prove the validity of our approach. The finite sample properties are illustrated in Section 5 by means of a simulation study and a data example. Finally, all technical proofs are deferred to Section B and C in the appendix, while Section A there provides more details on further algorithms.

2 Locally stationary functional time series

Consider the model (1.1) with a smooth regression function $m : [0,1]^2 \to \mathbb{R}$. Note that there are three types of simultaneous confidence bands/surfaces) which can be considered in the present context.

(1) Simultaneous confidence bands *for fixed* t, which have the form

$$\mathcal{C}(t) = \left\{ f : [0,1] \to \mathbb{R} \mid \ \hat{L}_1(u,t) \le f(u) \le \hat{U}_1(u,t) \ \ \forall u[0,1] \right\} , \tag{2.1}$$

where $\hat{L}_1$ and $\hat{U}_1$ are appropriate lower and upper bounds calculated from the data. As $t \in [0,1]$ is fixed these bound can be derived using results of the type (1.3). An alternative approach based on multiplier bootstrap is given in the online supplement to this paper (see Section A).

(2) Simultaneous confidence bands *for fixed* u, which have the form

$$\mathcal{C}(u) = \left\{ f : [0,1] \to \mathbb{R} \mid \ \hat{L}_2(u,t) \le f(t) \le \hat{U}_2(u,t) \ \ \forall t \in [0,1] \right\} , \tag{2.2}$$

where $\hat{L}_2$ and $\hat{U}_2$ are appropriate lower and upper bounds calculated from the data. Note that these bounds can not be directly calculated using results of the type (1.2) as the expectation of $X_{i,n}$ varies with i.

(3) Simultaneous confidence surfaces, which have the form

$$\mathcal{C} = \left\{ f : [0,1]^2 \to \mathbb{R} \mid \ \hat{L}_3(u,t) \le f(u,t) \le \hat{U}_3(u,t) \ \ \forall u,t \in [0,1] \right\} \tag{2.3}$$

where $\hat{L}_3$ and $\hat{U}_3$ are appropriate lower and upper bounds calculated from the data.

In the theoretical part of this paper we develop (asymptotic) simultaneous confidence bands of the form (2.2) and (2.3). Algorithms for confidence bands of the form (2.1) can be derived similarly, and are given in Section A of the online supplement for the sake of completeness. Our approach is based on the maximum deviations

$$\sup_{t\in[0,1]} |\hat{m}(u,t) - m(u,t)| \tag{2.4}$$

for the simultaneous confidence band of the form (2.2) and

$$\sup_{t,u\in[0,1]} |\hat{m}(u,t) - m(u,t)| \tag{2.5}$$

for simultaneous confidence surface of the form (2.3), where

$$\begin{aligned} \hat{m}(u,t) &= \frac{1}{nb_n}\sum_{i=1}^{n} X_{i,n}(t) K\Big(\frac{\frac{i}{n}-u}{b_n}\Big) \\ &= \frac{1}{nb_n}\sum_{i=1}^{n} m(\tfrac{i}{n},t) K\Big(\frac{\frac{i}{n}-u}{b_n}\Big) + \frac{1}{nb_n}\sum_{i=1}^{n} \varepsilon_{i,n}(t) K\Big(\frac{\frac{i}{n}-u}{b_n}\Big) \end{aligned} \tag{2.6}$$

denotes the common Nadaraya-Watson estimate with bandwidth b_n. Other estimates as local linear regression could be considered as well without changing our main theoretical results (note that we consider a uniform design and therefore the local linear and Nadaraya Watson estimator behave very similarly within the interval $[b_n, 1-b_n]$). In order to obtain quantiles for the maximum deviations estimates (2.4) and (2.5) we will develop a bootstrap procedure for the stochastic expansion

$$\sqrt{nb_n}\big(\hat{m}(u,t_s) - m(u,t_s)\big) \approx \frac{1}{\sqrt{nb_n}}\sum_{i=1}^{n} \varepsilon_{i,n}(t_s) K\Big(\frac{\frac{i}{n}-u}{b_n}\Big) \tag{2.7}$$

at discrete time points $t_1, \ldots, t_p \in [0,1]$, where p is increasing with the sample size such that the set $t_1, \ldots, t_p$ is asymptotically dense in the interval $[0,1]$. For this purpose we require several technical assumptions which will be introduced next.

Assumption 2.1 (mean function)**.** For each fixed $t \in [0,1]$ the function $u \to m(u,t)$ is four times continuously differentiable with bounded fourth order derivative, that is

$$\sup_{t,u\in[0,1]} \Big|\frac{\partial^4}{\partial u^4} m(u,t)\Big| \leq M_0$$

for some constant M_0.

Note that for the consistency of the estimate in (2.7) at a given point u it is not necessary to assume smoothness of the function m in the second argument. In fact, in Remark 3.1(ii) below we show that the difference between $\hat{m}(u,t)$ and $m(u,t)$ can be uniformly approximated by a weighted sum of the random variables $\varepsilon_{1,n}(t), \ldots, \varepsilon_{n,n}(t)$. As a consequence, a uniform approximation of the form (2.7) for an increasing number of points $\{t_1, \ldots, t_p\}$ is guaranteed by an appropriate smoothness condition on the error process $\{\varepsilon_{i,n}(t)\}_{i=1,\dots,n}$, which will be introduced next.

Assumption 2.2 (error process)**.** The error process has the form

$$\varepsilon_{i,n}(t) = G(\tfrac{i}{n}, t, \mathcal{F}_i) \ , \quad i = 1, \dots, n$$

where $\mathcal{F}_i = (\dots, \eta_{i-1}, \eta_i)$, $(\eta_i)_{i \in \mathbb{Z}}$ is a sequence of independent identically distributed random variables in some measurable space $\mathcal{S}$ and $G : [0,1] \times [0,1] \times \mathcal{S}^{\mathbb{Z}} \to \mathbb{R}$ denotes a filter with the following properties

(1) There exists a constant $t_0 > 0$ such that

$$\sup_{u,t \in [0,1]} \mathbb{E}(t_0 \exp(G(u,t,\mathcal{F}_0))) < \infty. \tag{2.8}$$

(2) Let $(\eta_i')_{i \in \mathbb{N}}$ denote a sequence of independent identically distributed random variables which is independent of but has the same distribution as $(\eta_i)_{i \in \mathbb{Z}}$. Define $\mathcal{F}_i^* = (\dots, \eta_{-1}, \eta_0', \eta_1, \dots, \eta_i)$ and consider for some $q \geq 2$ the dependence measure

$$\delta_q(G, i) = \sup_{u,t \in [0,1]} \|G(u,t,\mathcal{F}_i) - G(u,t,\mathcal{F}_i^*)\|_q. \tag{2.9}$$

There exists a constant $\chi \in (0,1)$ such that

$$\delta_q(G, i) = O(\chi^i) \ . \tag{2.10}$$

(3) For the same constant q as in (2) there exists a positive constant M such that

$$\sup_{t \in [0,1], u_1, u_2 \in [0,1]} \|G(u_1,t,\mathcal{F}_i) - G(u_2,t,\mathcal{F}_i)\|_q \leq M|u_1 - u_2|.$$

(4) The *long run variance*

$$\sigma^2(u,t) := \sum_{k=-\infty}^{\infty} \mathrm{Cov}(G(u,t,\mathcal{F}_0), G(u,t,\mathcal{F}_k)). \tag{2.11}$$

of the process $(G(u,t,\mathcal{F}_i))_{i \in \mathbb{Z}}$ satisfies

$$\inf_{u,t \in [0,1]} \sigma^2(u,t) > 0.$$

Assumption 2.2(2) requires that the dependence measure is geometrically decaying. Similar results as presented in Section 3 and 4 of this paper can be obtained under summability assumptions with substantially more intensive mathematical arguments and complicated notation, see Remark 3.1(ii) for some details. Assumption 2.2(3) means that the locally stationary functional time series is smooth in u and is crucial for constructing simultaneous confidence surfaces of the form (2.3).

Assumption 2.3. The filter G in Assumption 2.2 is differentiable with respect to t. If $G_2(u,t,\mathcal{F}_i) = \frac{\partial}{\partial t}G(u,t,\mathcal{F}_i)$, $G_2(u,0,\mathcal{F}_i) = G_2(u,0+,\mathcal{F}_i)$, $G_2(u,1,\mathcal{F}_i) = G_2(u,1-,\mathcal{F}_i)$, we assume

(1) There exists a constant $q^* > 2$ such that for some $\chi \in (0,1)$

$$\delta_{q^*}(G_2, i) = O(\chi^i).$$

(2) For the same constant $q^* > 2$ as in (1) there exists a constant M such that

$$\sup_{t\in[0,1],u\in[0,1]} \|G_2(u,t,\mathcal{F}_i)\|_{q^*} \le M.$$

Assumption 2.4 (kernel)**.** The kernel $K(\cdot)$ is a symmetric continuous function which vanishes outside the interval $[-1,1]$ and satisfies $\int_{\mathbb{R}} K(x)dx = 1$, $\int_{\mathbb{R}} K(v)v^2dv = 0$. Additionally, the second order derivative K'' is Lipschitz continuous on the interval $(-1,1)$.

We conclude this section discussing several examples for the error process, which are also used to illustrate Assumption 2.2.

Example 2.1. Let $(B_j)_{j\ge0}$ denote a basis of $L^2\big([0,1]^2\big)$ and let $(\eta_{i,j})_{i\ge0,j\ge0}$ denote an array of independent identically distributed centred random variables with variance σ^2. We define the error process

$$\epsilon_i(u,v) = \sum_{j=0}^{\infty} \eta_{i,j} B_j(u,v),$$

assume that

$$\sup_{u\in[0,1]} \int_0^1 \mathbb{E}(\epsilon_i^2(u,v))dv = \sigma^2 \sup_{u\in[0,1]} \sum_{s=0}^{\infty} \int B_s^2(u,v)dv < \infty.$$

Next, consider the locally stationary MA(∞) functional linear model

$$\varepsilon_{i,n}(t) = \sum_{j=0}^{\infty} \int_0^1 a_j(t,v)\epsilon_{i-j}(\tfrac{i}{n}, v)dv \ , \tag{2.12}$$

where $(a_j)_{j\geq 0}$ is a sequence of square integrable functions $a_j : [0,1]^2 \to \mathbb{R}$ satisfying

$$\sum_{j=0}^{\infty} \sup_{u,v\in[0,1]} |a_j(u,v)| < \infty.$$

Define $\mathcal{F}_i = (\dots, \eta_{i-1}, \eta_i)$, then we obtain from (2.12) the representation of the form $\varepsilon_{i,n}(t) = G(\frac{i}{n}, t, \mathcal{F}_i)$, where

$$G(u,t,\mathcal{F}_i) = \sum_{j=0}^{\infty} \int_0^1 a_j(t,v) \sum_{s=0}^{\infty} \eta_{i-j,s} B_s(u,v)dv.$$

Further, assume that $\|\eta_{1,1}\|_q < \infty$ for some $q > 2$, then by Burkholder's and Cauchy's inequality the physical dependence measure defined in (2.9) satisfies

$$\begin{aligned}\delta_q(G,i) &= \sup_{u,t\in[0,1]} \Big\| \sum_{s=0}^{\infty} \int_0^1 a_i(t,v) B_s(u,v)dv(\eta_{0,s} - \eta'_{0,s}) \Big\|_q \\ &= O\Big(\sup_{u,t\in[0,1]} \Big(\sum_{s=0}^{\infty} \Big(\int_0^1 a_i(t,v)B_s(u,v)dv \Big)^2 \Big)^{1/2} \Big) \\ &= O\Big(\sup_{t\in[0,1]} \Big[\int_0^1 a_i^2(t,v)dv \Big]^{1/2} \Big) .\end{aligned}$$

Similarly, it follows for $q \geq 2$ that

$$\begin{aligned}\|G(u,t,\mathcal{F}_0)\|_q^2 &\leq Mq \sum_{j=0}^{\infty} \sum_{s=0}^{\infty} (\int_0^1 a_j(t,v)B_s(u,v)dv)^2 \|\eta_{1,1}\|_q^2 \\ &\leq Mq \sum_{j=0}^{\infty} \int_0^1 a_j^2(t,v)dv \sum_{s=0}^{\infty} \int_0^1 B_s^2(u,v)dv \|\eta_{1,1}\|_q^2 \qquad (2.13)\end{aligned}$$

for some sufficiently large constant M. Consequently, the filter G has finite moment of order q, if

$$\sum_{j=0}^{\infty} \int_0^1 a_j^2(t,v)dv < \infty \ . \qquad (2.14)$$

Furthermore, if there exists positive constants M_0 and α such that $\|\eta_{1,1}\|_q \leq M_0 q^{1/2-\alpha}$, Assumption 2.2(1) is also satisfied, because for any fixed t_0, the sequence

$$\frac{t_0^q \|G(u,t,\mathcal{F}_0)\|_q^q}{q!} = O\Big(\frac{C^q t_0^q q^{q-\alpha q}}{q!} \Big) = O\Big(\frac{1}{\sqrt{2\pi q}} (\frac{Ct_0 e}{q^{\alpha}})^q \Big)$$

is summable, where

$$C = \sup_{t\in[0,1],u\in[0,1]} M_0 \sqrt{M \sum_{j=0}^{\infty} \int_0^1 a_j^2(t,v)dv \sum_{s=0}^{\infty} \int_0^1 B_s^2(u,v)dv.}$$

Moreover, if $b_s(u,v) := \frac{\partial}{\partial u} B_s(u,v)$ exists for $u \in (0,1), v \in [0,1]$, then it follows observing (2.13) that Assumption 2.2(3) holds under (2.14) and

$$\sup_{u\in[0,1]} \sum_{s=0}^{\infty} \int b_s^2(u,v)dv < \infty.$$

Finally, if $\|\eta_{1,1}\|_{q^*} < \infty$ and

$$\sup_{t\in[0,1]} \Big[\int_0^1 \Big(\frac{\partial}{\partial t} a_i(t,v) \Big)^2 dv \Big]^{1/2} = O(\chi^i) \ ,$$

it can be shown by similar arguments as given above that Assumption 2.3 is satisfied.

Example 2.2. For a given orthonormal basis $(\phi_k(t))_{k\geq 1}$ of $L^2([0,1])$ consider the functional time series $(G(u,t,\mathcal{F}_i))_{i\in\mathbb{Z}}$ defined by

$$G(u,t,\mathcal{F}_i) = \sum_{k=1}^{\infty} H_k(u,\mathcal{F}_i)\phi_k(t) \ , \tag{2.15}$$

where for each $k \in \mathbb{N}$ and $u \in [0,1]$ the random coefficients $(H_k(u,\mathcal{F}_i))_{i\in\mathbb{Z}}$ are stationary time series. A parsimonious choice of (2.15) is to consider $\mathcal{F}_i = \cup_{k=1}^{\infty} \mathcal{F}_{i,k}$ where $\{\mathcal{F}_{i,k}\}_{k=1}^{\infty}$ are independent filtrations. In this case we obtain

$$G(u,t,\mathcal{F}_i) = \sum_{k=1}^{\infty} H_k(u,\mathcal{F}_{i,k})\phi_k(t), \tag{2.16}$$

and the random coefficients $H_k(u,\mathcal{F}_{i,k})$ are stochastically independent. A sufficient condition for Assumption 2.2(2) in model (2.16) is

$$\sup_{t\in[0,1]} \sum_{k=0}^{\infty} |\phi_k(t)|\delta_q(H_k,i) = O(\chi^i) \ ,$$

where $\delta_q(H_k,i) := \sup_{u\in[0,1]} \|H_k(u,\mathcal{F}_{i,k}) - H_k(u,\mathcal{F}_{i,k}^*)\|_q$. The qth moment of the process G in (2.16) exists for $q \geq 2$, if

$$\Delta_q := \sup_{t\in[0,1],u\in[0,1]} \sum_{k=0}^{\infty} \phi_k^2(t)\|H_k(u,\mathcal{F}_{0,k})\|_q^2 < \infty.$$

If further $\Delta_q = O(q^{1/2-\alpha})$ for some $\alpha > 0$, then similar arguments as given in Example 2.1 show that Assumption 2.2(1) is satisfied as well. Finally, if the inequality

$$\sum_{k=0}^{\infty} \phi_k^2(t)\Big\|\frac{\partial}{\partial u}H_k(u,\mathcal{F}_{0,k})\Big\|_q^2 < \infty$$

holds uniformly with respect to $t, u \in (0,1)$, Assumption 2.2(3) is also satsified.
On the other hand, in model (2.15) we have $H_k(u,\mathcal{F}_i) = \int_0^1 G(u,t,\mathcal{F}_i)\phi_k(t)dt$, and consequently the magnitude of $\|H_k\|_q$ and $\delta_q(H_k,i)$ can be determined by Assumption 2.2. For example, if the basis of $L^2([0,1])$ is given by $\phi_k(t) = \cos(k\pi t)$ $(k = 0,1,\dots)$ and the inequality

$$\|G(u,0,\mathcal{F}_1)\|_q + \Big\|\frac{\partial}{\partial t}G(u,0,\mathcal{F}_1)\Big\|_q + \sup_{u\in[0,1]}\Big\|\frac{\partial^2}{\partial t^2}G(u,t,\mathcal{F}_1)\Big\|_q < \infty,$$

holds for $u \in [0,1]$, it follows by similar arguments as given in Zhou and Dette (2020) that

$$\sup_{u\in[0,1]}\big\|H_k(u,\mathcal{F}_k)\big\|_q = O(k^{-2}), \quad \delta_q(H_k,i) = O\big(\min(k^{-2},\delta_G(i,q))\big). \tag{2.17}$$

Similarly, assume that the basis of $L^2([0,1])$ is given by the Legendre polynomials and that

$$\sup_{u\in[0,1]}\max_{s=1,2,3}\Big\|\int_{-1}^{1}\frac{|\frac{\partial^s}{\partial t^s}G(u,t,\mathcal{F}_0)|}{\sqrt{1-x^2}}dx\Big\|_q < \infty.$$

If additionally for every $\varepsilon > 0$, there exists a constant $\delta > 0$ such that

$$\sum_{s=1,2}\sum_{k}\Big\|\frac{\partial^s}{\partial t^s}G(u,x_k,\mathcal{F}_i) - \frac{\partial^s}{\partial t^s}G(u,x_{k-1},\mathcal{F}_i))\Big\|_q < \varepsilon$$

for any finite sequence of pairwise disjoint sub-intervals (x_{k-1},x_k) of the interval $(0,1)$ such that $\sum_k (x_k - x_{k-1}) < \delta$, it follows from Theorem 2.1 of Wang and Xiang (2012) that (2.17) holds as well.
Finally, if

$$\sup_{t\in[0,1]}\sum_{k=0}^{\infty}|\phi_k'(t)|\delta_{q^*}(H_k,i) = O(\chi^i)$$

and

$$\sup_{t\in[0,1],u\in[0,1]} \sum_{k=0}^{\infty} \phi_k'^2(t)\|H_k(u,\mathcal{F}_{0,k})\|_{q^*}^2 < \infty,$$

it can be shown by similar arguments as given above that Assumption 2.3 is also satisfied.

3 Gaussian approximation and bootstrap

In this section we will develop a resampling procedure to approximate the quantiles of the distribution of the the maximum deviations defined in (2.2) and (2.3). For this purpose we first derive non-standard Gaussian approximations, which are the basis for the proposed bootstrap procedure.

3.1 Gaussian approximation

To be precise, define for $1 \le i \le n$ the p-dimensional vector

$$\begin{aligned} Z_i(u) &= (Z_{i,1}(u), \dots, Z_{i,p}(u))^\top \\ &= K\Big(\frac{\frac{i}{n}-u}{b_n}\Big)\Big(G(\tfrac{i}{n},\tfrac{1}{p},\mathcal{F}_i), G(\tfrac{i}{n},\tfrac{2}{p},\mathcal{F}_i), \dots, G(\tfrac{i}{n},\tfrac{p-1}{p},\mathcal{F}_i), G(\tfrac{i}{n},1,\mathcal{F}_i)\Big)^\top, \end{aligned} \tag{3.1}$$

where $K(\cdot)$ and b_n are the kernel and bandwidth used in the estimate (2.6), respectively. Next we define the p-dimensional vector

$$Z_{i,l} = Z_i(\tfrac{l}{n}) = (Z_{i,l,1}, \dots Z_{i,l,p})^\top, \tag{3.2}$$

where

$$Z_{i,l,k} = G(\tfrac{i}{n},\tfrac{k}{p},\mathcal{F}_i)K\Big(\frac{\frac{i}{n}-\frac{l}{n}}{b_n}\Big) \quad (1 \le k \le p)$$

(note that all entries in the vector $Z_{i,l}$ will be zero if $|i-l|/(nb_n) \ge 1$). Define

$$\Delta(u,t) := \hat{m}(u,t) - m(u,t)$$

as the difference between the regression function m and its estimate (2.6), then the following theorem provides a Gaussian approximation for the maximum deviation of $\max_t |\Delta(u,t)|$ for fixed u and is the basic tool for the development of a simultaneous confidence band of the form (2.2) based on bootstrap. Throughout this paper we use the notation

$$\Theta(a,b) = a\sqrt{1 \vee \log((b/a))}$$

for positive constants a, b, and $|v|_\infty = \max_{1\le i\le k}|v_i|$ denotes the maximum norm of a k-dimensional vector $v = (v_1, ..., v_k)^\top$ (the dimension k will always be clear from the context). The notation $a \vee b$

denotesmax(a,b).

Theorem 3.1. *Let Assumptions 2.1 - 2.4 be satisfied and assume that the bandwidth in* (2.6) *satisfies that* $n^{1+a}b_n^9 = o(1)$, $n^{a-1}b_n^{-1} = o(1)$ *for some* $0 < a < 4/5$. *For any fixed* $u \in (0,1)$ *there exists a sequence of centred p-dimensional Gaussian vectors* $\big(Y_i(u)\big)_{i\in\mathbb{N}}$ *with the same covariance structure as the vector* $Z_i(u)$ *in* (3.1)*, such that*

$$\begin{aligned}\mathfrak{P}_n(u) &:= \sup_{x\in\mathbb{R}} \Big|\mathbb{P}\Big(\max_{0\le t\le 1}\sqrt{nb_n}|\Delta(u,t)| \le x\Big) - \mathbb{P}\Big(\Big|\frac{1}{\sqrt{nb_n}}\sum_{i=1}^n Y_i(u)\Big|_\infty \le x\Big)\Big| \\ &= O\Big((nb_n)^{-(1-11\iota)/8} + \Theta\Big(\sqrt{nb_n}\Big(b_n^4 + \frac{1}{nb_n}\Big), p\Big) + \Theta\Big(p^{\frac{1-q^*}{1+q^*}}, p\Big)\Big)\end{aligned}$$

for any sequence $p \to \infty$ *with* $p = O(\exp(n^\iota))$ *for some* $0 \le \iota < 1/11$. *In particular,*

$$\mathfrak{P}_n(u) = o(1)$$

if $p = \sqrt{n}$ *and the constant* q^* *in Assumption 2.3 is sufficiently large.*

A Gaussian approximation for the maximum deviation $\max_{u,t}\sqrt{nb_n}|\Delta(u,t)|$ is more intricate. We recall the definition of the vector $Z_{i,l}$ in (3.2) and consider the $(n-2\lceil nb_n\rceil+1)p$-dimensional vectors $\tilde{Z}_1,\ldots\tilde{Z}_{2\lceil nb\rceil-1}$ defined by

$$\tilde{Z}_j = \big(Z^\top_{j,\lceil nb_n\rceil}, Z^\top_{j+1,\lceil nb_n\rceil+1},\ldots,Z^\top_{n-2\lceil nb_n\rceil+j,n-\lceil nb_n\rceil}\big)^\top . \tag{3.3}$$

Note that $\tilde{Z}_{2\lceil nb_n\rceil} = 0$ and that

$$\max_{\substack{1\le v\le p\\ \lceil nb_n\rceil\le l\le n-\lceil nb_n\rceil}} \Big|\frac{1}{\sqrt{nb_n}}\sum_{i=1}^n \varepsilon_{i,n}(\tfrac{v}{p})K\Big(\frac{\frac{i}{n}-\frac{l}{n}}{b_n}\Big)\Big| = \Big|\frac{1}{\sqrt{nb_n}}\sum_{i=1}^{2\lceil nb_n\rceil-1}\tilde{Z}_i\Big|_\infty . \tag{3.4}$$

Heuristically, we have by (2.6)

$$\max_{\substack{b_n\le u\le 1-b_n,\\ 0\le t\le 1}} \sqrt{nb_n}|\Delta(u,t)| \approx \max_{\substack{1\le v\le p\\ \lceil nb_n\rceil\le l\le n-\lceil nb_n\rceil}} \Big|\frac{1}{\sqrt{nb_n}}\sum_{i=1}^n \varepsilon_{i,n}(\tfrac{v}{p})K\Big(\frac{\frac{i}{n}-\frac{l}{n}}{b_n}\Big)\Big|,$$

and therefore the right hand side of (3.4) is an approximation of the maximum absolute deviation $\max_{u,t}\sqrt{nb_n}|\Delta(u,t)|$. The following results makes this intuition rigorous.

Theorem 3.2. *Let Assumptions 2.1 - 2.4 be satisfied and assume that* $n^{1+a}b_n^9 = o(1)$, $n^{a-1}b_n^{-1} = o(1)$ *for some* $0 < a < 4/5$, *and* $\frac{n^{2/q^*}}{nb_n} = o(1)$. *There exists a sequence of centred* $(n-2\lceil nb_n\rceil+1)p$*-dimensional centred Gaussian vectors* $\tilde{Y}_1,\ldots,\tilde{Y}_{2\lceil nb_n\rceil-1}$ *with the same auto-covariance structure as* *the vector $\tilde{Z}_i$ in* (3.3) *such that*

$$\mathfrak{P}_n := \sup_{x\in\mathbb{R}} \Big|\mathbb{P}\Big(\max_{b_n\le u\le 1-b_n, 0\le t\le 1} \sqrt{nb_n}|\Delta(u,t)| \le x\Big) - \mathbb{P}\Big(\Big|\frac{1}{\sqrt{nb_n}}\sum_{i=1}^{2\lceil nb_n\rceil -1} \tilde{Y}_i\Big|_\infty \le x\Big)\Big|$$
$$= O\Big((nb_n)^{-(1-11\iota)/8} + \Theta\Big(\sqrt{nb_n}(b_n^4 + \frac{1}{nb_n}), np\Big) + \Theta\big(((np)^{1/q^*}((nb_n)^{-1} + 1/p))^{\frac{q^*}{q^*+1}}, np\big)\Big)$$

for any sequence $p \to \infty$ with $np = O(\exp(n^\iota))$ for some $0 \le \iota < 1/11$. In particular, for the choice $p = \sqrt{n}$ we have

$$\mathfrak{P}_n = o(1)$$

if the constant q^* in Assumption 2.3 is sufficiently large.

Remark 3.1.

(i) A careful inspection of the proofs in Section B of the online supplement shows that it is possible to prove similar results under alternative moment assumptions. More precisely, Theorem 3.1 remains valid if condition (2.8) is replaced by

$$\mathbb{E}\Big[\sup_{0\le t\le 1} (G(u,t,\mathcal{F}_0))^4\Big] < \infty \quad . \tag{3.5}$$

Similarly, Theorem 3.2 holds under the assumption

$$\mathbb{E}\Big[\sup_{0\le u,t\le 1} (G(u,t,\mathcal{F}_0))^4\Big] < \infty \quad . \tag{3.6}$$

The details are omitted for the sake of brevity. Note that the sup in (3.5) and (3.6) appears inside the expectation, while it appears outside the expectation in (2.8). Thus neither (2.8) implies (3.5) and (3.6) nor vice versa.

(ii) Assumption 2.2(2) requires geometric decay of the dependence measure $\delta_q(G,i)$ and a careful inspection of the proofs in Section B of the online supplement shows that similar (but weaker) results can be obtained under less restrictive assumptions. To be precise, define $\Delta_{k,q} = \sum_{i=k}^\infty \delta_q(G,i)$, $\Xi_M = \sum_{i=M}^\infty i\delta_2(G,i)$ and consider the following assumptions.

(a) $\sum_{i=0}^\infty i\delta_3(G,i) < \infty$.

(b1) There exist constants $M = M(n) > 0$, $\gamma = \gamma(n) \in (0,1)$ and $C_1 > 0$ such that

$$(2\lceil nb_n\rceil)^{3/8} M^{-1/2} l_n^{-5/8} \ge C_1 \max\{l_n, l_n^{1/2}\}$$

where $l_n = \max(\log(2\lceil nb_n\rceil p/\gamma), 1)$.

(b2) There exist constants $M = M(n) > 0$, $\gamma = \gamma(n) \in (0,1)$ and $C_2 > 0$ such that

$$(2\lceil nb_n\rceil)^{3/8} M^{-1/2} {l'_n}^{-5/8} \geq C_2 \max\{l'_n, {l'_n}^{1/2}\}$$

where $l'_n = \max(\log(2\lceil nb_n\rceil(n - 2\lceil nb_n\rceil + 1)p/\gamma), 1)$.

Then, if the assumptions of Theorem 3.1 hold, where Assumption 2.2 (ii) is replaced by (a) and (b1), we have

$$\mathfrak{P}_n(u) = O\Big(\eta_n + \Theta\Big(\sqrt{nb_n}(b_n^4 + \frac{1}{nb_n}), p\Big) + \Theta\Big(p^{\frac{1-q^*}{1+q^*}}, p\Big)\Big)$$

with

$$\begin{aligned}\eta_n = {} & (nb_n)^{-1/8} M^{1/2} l_n^{7/8} + \gamma + \big((nb_n)^{1/8} M^{-1/2} l_n^{-3/8}\big)^{q/(1+q)} \big(p\Delta_{Mq}^q\big)^{1/(1+q)} \\ & + \Xi_M^{1/3} \left(1 \vee \log\left(p/\Xi_M\right)\right)^{2/3}.\end{aligned}$$

Similarly under the conditions of Theorem 3.2 with Assumption 2.2 (ii) replaced by (a) and (b2), we have

$$\mathfrak{P}_n = O\Big(\eta'_n + \Theta\Big(\sqrt{nb_n}(b_n^4 + \frac{1}{nb_n}), np\Big) + \Theta\Big(\big((np)^{1/q^*}((nb_n)^{-1} + 1/p)\big)^{\frac{q^*}{q^*+1}}, np\Big)\Big)$$

with

$$\begin{aligned}\eta'_n = {} & (nb_n)^{-1/8} M^{1/2} {l'_n}^{7/8} + \gamma + \Big((nb_n)^{1/8} M^{-1/2} {l'_n}^{-3/8}\Big)^{q/(1+q)} \big(np\Delta_{M,q}^q\big)^{1/(1+q)} \\ & + \Xi_M^{1/3} \left(1 \vee \log\left(np/\Xi_M\right)\right)^{2/3}.\end{aligned}$$

3.2 Bootstrap

We will use Theorem 3.1 and 3.2 to construct simultaneous confidence bands and surfaces for the regression function m. Therefore it is important to generate the Gaussian random vectors $\tilde{Y}_i$ with the same auto-covariance structure as the vector $\tilde{Z}_i(u)$ in (3.1) or the vector $\tilde{Z}_i$ in (3.3). For this purpose we consider (for fixed t) the local linear estimator of m with bandwidth $d_n > 0$, that is

$$\big(\hat{m}_{d_n}(u,t), \widehat{\tfrac{\partial}{\partial u} m_{d_n}}(u,t)\big)^\top = \underset{\beta_0,\beta_1}{\operatorname{argmin}} \sum_{i=1}^n \big(X_{i,n}(t) - \beta_0 - \beta_1(\tfrac{i}{n} - u)\big)^2 H\Big(\frac{\frac{i}{n} - u}{d_n}\Big) \tag{3.7}$$

where

$$H(x) = 0.75(1 - x^2)\mathbf{1}(|x| \leq 1) \tag{3.8}$$

is the Epanechnikov Kernel. We define the residuals

$$\hat{\varepsilon}_{i,n}(t) = X_{i,n}(t) - \hat{m}_{d_n}(\tfrac{i}{n}, t) \ , \tag{3.9}$$

and the p-dimensional vector

$$\begin{aligned} \hat{Z}_i(u) &= (\hat{Z}_{i,1}(u), \ldots, \hat{Z}_{i,p}(u))^\top \\ &= K\Big(\frac{\frac{i}{n} - u}{b_n}\Big)\big(\hat{\varepsilon}_{i,n}(\tfrac{1}{p}), \hat{\varepsilon}_{i,n}(\tfrac{2}{p}), \ldots, \hat{\varepsilon}_{i,n}(\tfrac{p-1}{p}), \hat{\varepsilon}_{i,n}(1)\big)^\top . \end{aligned} \tag{3.10}$$

as an analog of (3.1). Similarly we define the analog of (3.3) by

$$\hat{\tilde{Z}}_j = \big(\hat{Z}^\top_{j,\lceil nb_n \rceil}, \hat{Z}^\top_{j+1,\lceil nb_n \rceil+1}, \ldots, \hat{Z}^\top_{n-2\lceil nb_n \rceil+j, n-\lceil nb_n \rceil}\big)^\top, \tag{3.11}$$

where

$$\hat{Z}_{i,l} = \hat{Z}_i(\tfrac{l}{n}) = (\hat{Z}_{i,l,1}, \ldots \hat{Z}_{i,l,p})^\top$$

(note that we replace $Z_{i,l}$ in (3.3) by $\hat{Z}_{i,l}$). We propose Algorithm 1 and 2 to calculate a simultaneous confidence band and a simultaneous confidence surface for the regression function m.

Algorithm 1:

Result: simultaneous confidence band for fixed $u \in [b_n, 1 - b_n]$ as defined in (2.2)

(a) Calculate the p-dimensional vectors $\hat{Z}_i(u)$ in (3.10)

(b) For window size m_n, let $m_n' = 2\lfloor m_n/2 \rfloor$, define the vectors $\hat{S}^*_{jm_n}(u) = \sum_{r=j}^{j+m_n-1} \hat{Z}_r(u)$,

$$\hat{S}_{jm_n'}(u) = \hat{S}^*_{j,\lfloor m_n/2 \rfloor}(u) - \hat{S}^*_{j+\lfloor m_n/2 \rfloor,\lfloor m_n/2 \rfloor}(u)$$

(c) **for** r=1, ..., B **do**

- Generate independent standard normal distributed random variables $\{R_i^{(r)}\}_{i=\lceil nu-nb_n \rceil}^{\lfloor nu+nb_n \rfloor}$
- Calculate the bootstrap statistic

$$T^{(r)}(u) = \Big| \sum_{j=\lceil nu-nb_n \rceil}^{\lfloor nu+nb_n \rfloor - m_n' + 1} \hat{S}_{jm_n'}(u) R_j^{(r)} \Big|_\infty$$

end

(d) Define $T_{\lfloor (1-\alpha)B \rfloor}(u)$ as the empirical $(1-\alpha)$-quantile of the sample $T^{(1)}(u), \ldots, T^{(B)}(u)$ and

$$\hat{L}_1(u,t) = \hat{m}(u,t) - \hat{r}_1(u) \quad , \quad \hat{U}_1(u,t) = \hat{m}(u,t) + \hat{r}_1(u)$$

where

$$\hat{r}_1(u) = \frac{\sqrt{2} T_{\lfloor (1-\alpha)B \rfloor}(u)}{\sqrt{nb_n}\sqrt{m_n'(\lfloor nu + nb_n \rfloor - \lceil nu - nb_n \rceil - m_n' + 2)}}$$

Output:

$$\mathcal{C}_n(u) = \big\{ f : [0,1]^2 \to \mathbb{R} \mid \ \hat{L}_1(u,t) \le f(u,t) \le \hat{U}_1(u,t) \ \ \forall t \in [0,1] \big\}. \tag{3.12}$$

Algorithm 2:

Result: simultaneous confidence surface of the form (2.3)

(a) Calculate the $(n-2\lceil nb_n\rceil+1)p$-dimensional vectors $\hat{\hat{Z}}_i$ in (3.11)

(b) For window size m_n, let $m_n' = 2\lfloor m_n/2\rfloor$, define the vectors $\hat{S}^*_{jm_n} = \sum_{r=j}^{j+m_n-1} \hat{\hat{Z}}_r$, and

$$\hat{S}_{jm_n'} = \hat{S}^*_{j,\lfloor m_n/2\rfloor} - \hat{S}^*_{j+\lfloor m_n/2\rfloor,\lfloor m_n/2\rfloor} \tag{3.13}$$

For $a \le b$ let $\hat{S}_{jm_n',[a:b]}$ be the $(b-a+1)$-dimensional sub-vector of the vector $\hat{S}_{jm_n'}$ in (3.13) containing its ath - bth components.

(c) **for** *r=1, ..., B* **do**

- Generate independent standard normal distributed random variables $\{R_i^{(r)}\}_{i\in[1,n-m_n']}$, and define the $(2\lceil nb_n\rceil - m_n')$-dimensional random vectors

$$V_k^{(r)} = (V_{k,1}^{(r)},\ldots,V_{k,2\lceil nb_n\rceil-m_n'}^{(r)})^\top := (R_k^{(r)},\ldots,R_{k+2\lceil nb_n\rceil-m_n'-1}^{(r)})^\top\ . \tag{3.14}$$

- Calculate

$$T_k^{(r)} = \sum_{j=1}^{2\lceil nb_n\rceil-m_n'} \hat{S}_{jm_n',[(k-1)p+1:kp]} V_{k,j}^{(r)}\ ,\quad k=1,\ldots,n-2\lceil nb_n\rceil+1, \tag{3.15}$$

$$T^{(r)} = \max_{1\le k\le n-2\lceil nb_n\rceil+1} |T_k^{(r)}|_\infty .$$

end

(d) Define $T_{\lfloor(1-\alpha)B\rfloor}$ as the empirical $(1-\alpha)$-quantile of the bootstrap sample $T^{(1)},\ldots,T^{(B)}$ and

$$\hat{L}_2(u,t) = \hat{m}(u,t) - \hat{r}_2\quad ,\qquad \hat{U}_2(u,t) = \hat{m}(u,t) + \hat{r}_2$$

where

$$\hat{r}_2 = \frac{\sqrt{2}T_{\lfloor(1-\alpha)B\rfloor}}{\sqrt{nb_n}\sqrt{m_n'(2\lceil nb_n\rceil - m_n').}}$$

Output:

$$\mathcal{C}_n = \big\{f:[0,1]^2\to\mathbb{R}\ |\ \ \hat{L}_2(u,t)\le f(u,t)\le \hat{U}_2(u,t)\ \ \forall u\in[b_n,1-b_n]\ \forall t\in[0,1]\big\}. \tag{3.16}$$

Remark 3.2. We would like to point out that there is a substantial difference between Algorithm 2 and existing bootstrap methods for inference by simultaneous confidence bands in other contexts (see Zhou and Wu, 2010; Wang et al., 2020; Karmakar et al., 2021, among others). Note that for a given k the random variables $V_{k,j}^{(r)}$ in Algortihm 2 are independent standard normally distributed. However, the Gaussian random vector $V_k^{(r)}$ and $V_{k'}^{(r)}$ defined by (3.14) are not independent if the indices k and k' are close. For example, most of the elements of the vectors $V_1 = (R_1, R_2, ..., R_{i+p-m'_n})^\top$ and $V_2 = (R_2, R_3, ..., R_{i+p-m'_n}, R_{i+p-m'_n+1})^\top$ are the same. This is necessary to mimic the structure of the vectors $\tilde{Z}_i$ in (3.3), which appear in the Gaussian approximation provided by Theorem 3.2. For example, we have $\tilde{Z}_1 = \left(Z_{1,\lceil nb_n\rceil}^\top, Z_{2,\lceil nb_n\rceil+1}^\top, \ldots, Z_{n-2\lceil nb_n\rceil+1,n-\lceil nb_n\rceil}^\top\right)^\top$, $\tilde{Z}_2 = \left(Z_{2,\lceil nb_n\rceil}^\top, Z_{3,\lceil nb_n\rceil+1}^\top, \ldots, Z_{n-2\lceil nb_n\rceil+2,n-\lceil nb_n\rceil}\right)^\top$, and therefore most of the elements of the vectors $\tilde{Z}_1$ and $\tilde{Z}_2$ coincide as well.

We have the following theorem regarding the validity of the bootstrap simultaneous confidence band $\mathcal{C}_n(u)$ and the surface $\mathcal{C}_n$ defined in Algorithms 1 and 2, respectively.

Theorem 3.3. *Assume that the conditions of Theorem 3.2 hold, and that $nd_n^3 \to \infty$, $nd_n^6 = o(1)$. Define*

$$\vartheta'_n = \frac{\log^2 n}{m_n} + \frac{m_n}{nb_n} + \sqrt{\frac{m_n}{nb_n}} p^{4/q} \ , \quad \textit{and} \ \ \vartheta_n = \frac{\log^2 n}{m_n} + \frac{m_n}{nb_n} + \sqrt{\frac{m_n}{nb_n}} (np)^{4/q}.$$

(i) If $p \to \infty$ such that $p = O(\exp(n^\iota))$ for some $0 \le \iota < 1/11$ and

$$(\vartheta')_n^{1/3} \Big\{1 \vee \log\Big(\frac{p}{\vartheta'_n}\Big)\Big\}^{2/3} + \Theta\Big(\Big(\sqrt{m_n \log p}\Big(\frac{1}{\sqrt{nd_n}} + d_n^2\Big) p^{\frac{1}{q}}\Big)^{q/(q+1)}, p\Big) = o(1),$$

then the simultaneous confidence band (3.12) *in Algorithm 1 satisfies*

$$\lim_{n\to\infty} \lim_{B\to\infty} \mathbb{P}(m(u,\cdot) \in \mathcal{C}_n(u) \mid \mathcal{F}_n) = 1 - \alpha$$

in probability.

(ii) If $p \to \infty$ such that $np = O(\exp(n^\iota))$ for some $0 \le \iota < 1/11$ and

$$\vartheta_n^{1/3} \Big\{1 \vee \log\Big(\frac{np}{\vartheta_n}\Big)\Big\}^{2/3} + \Theta\Big(\Big(\sqrt{m_n \log np}\Big(\frac{1}{\sqrt{nd_n}} + d_n^2\Big) (np)^{\frac{1}{q}}\Big)^{q/(q+1)}, np\Big) = o(1),$$

then the simultaneous confidence band (3.16) *in Algorithm 2 satisfies*

$$\lim_{n\to\infty} \lim_{B\to\infty} \mathbb{P}(m \in \mathcal{C}_n \mid \mathcal{F}_n) = 1 - \alpha$$

in probability.

Remark 3.3. Several authors consider (stationary) functional data models with noisy observation (see Cao et al., 2012; Chen and Song, 2015, among others) and we expect that the results presented in this section can be extended to this scenario. More precisely, consider the model

$$Y_{ij} = X_{i,n}(\tfrac{j}{N}) + \sigma(\tfrac{j}{N})z_{ij},\ 1 \le i \le n, 1 \le j \le N\ ,$$

where $X_{i,n}$ is the functional time series defined in (1.1), $\{z_{ij}\}_{i=1,\dots,n,j=1,\dots,N}$ is an array of centred independent identically distributed observations and $\sigma(\cdot)$ is a positive function on the interval $[0,1]$. This means that one can not observe the full trajectory of $\{X_{i,n}(t) \mid t \in [0,1]\}$, but only the function $X_{i,n}$ evaluated at the discrete time points $1/N, 2/N, \dots, (N-1)/N, 1$ subject to some random error. If $N \to \infty$ as $n \to \infty$, and the regression function m in (1.1) is sufficiently smooth, we expect that we can construct simultaneous confidence bands and surfaces by applying the procedure described in this section to smoothed trajectories. A similar comment applies to the confidence regions developed in the following section.
For example, we can consider the smooth estimate

$$\tilde{m}(u,\cdot) = \operatorname*{argmin}_{g \in \mathcal{S}_p} \sum_{i=\lfloor nu-\sqrt{n}\rfloor}^{\lceil nu+\sqrt{n}\rceil} \sum_{j=1}^{N} \big(Y_{i,j} - g(\tfrac{j}{N})\big)^2\ , \tag{3.17}$$

where $\mathcal{S}_p$ denotes the set of splines of order p, which depends on the smoothness of the function $t \to m(u,t)$. We can now construct confidence bands applying the methodology to the data $\tilde{X}_{i,n}(\cdot) = \tilde{m}(\frac{i}{\sqrt{n}},\cdot)$, $i = 1,\dots,\sqrt{n}$ due to the asymptotic efficiency of the spline estimate (see Proposition 3.2-3.4 in Cao et al., 2012).
Alternatively, we can also obtain smooth estimates $t \to \check{X}_{i,n}(t)$ of the trajectory using local polynomials, and we expect that the proposed methodology applied to the data $\check{X}_{1,n},\dots,\check{X}_{n,n}$ will yield valid simultaneous confidence bands and surfaces, where the range for the variable t is restricted to the interval $[c_n, 1-c_n]$ and c_n denotes the bandwidth of the local polynomial estimator used in smooth estimator of the trajectory.

4 Confidence bands and surfaces with varying width

The confidence bands and surfaces in Section 3 have a constant width and do not reflect the variability of the estimate $\hat{m}$ at the point (u,t). In this section we will construct simultaneous confidence bands and surfaces adjusted by the long-run variance defined in (2.11) (more precisely by an appropriate estimator). Among others, this approach has been proposed by Degras (2011) and Zheng et al. (2014) for repeated measurement data from independent subjects where a variance estimator is used for standardization. It has also been considered by Zhou and Wu (2010) who derived a simultaneous confidence tube for the parameter of a time varying coefficients linear model with a (real-valued) locally stationary error process. In the situation considered in the present paper this task is challenging as a uniformly consistent estimator of the long-run variance σ^2 in (2.11) is required. For the sake of brevity we will restrict ourselves to the construction of simultaneous confidence surfaces of the form (2.3), but similar results can be obtained for the bands (2.1) and (2.2) as well. For these types of confidence bands the corresponding algorithms are given in Section A of the appendix.

If the long-run variance in (2.11) would be known, Algorithm 2 can easily be modified to obtain a confidence band of the form

$$\left\{f : [0,1]^2 \to \mathbb{R} \mid \; \hat{m}(u,t) - \sigma(u,t)\hat{c}_{n,1-\alpha} \leq f(u,t) \leq \hat{m}(u,t) + \sigma(u,t)\hat{c}_{n,1-\alpha} \;\; \forall u,t \in [0,1]\right\}, \tag{4.1}$$

where $\hat{c}_{n,1-\alpha}$ is the empirical quantile calculated by bootstrap. We will investigate a Gaussian approximation in this case first and discuss the problem of variance estimation afterwards. For this purpose we consider the "normalized" maximum deviation of

$$\Delta^\sigma(u,t) = \frac{\hat{m}(u,t) - m(u,t)}{\sigma(u,t)}.$$

We define for $1 \leq i \leq n$ the p dimensional vector

$$\begin{aligned} Z_i^\sigma(u) &= (Z_{i,1}^\sigma(u), \ldots, Z_{i,p}^\sigma(u))^\top \\ &= K\Big(\frac{\frac{i}{n} - u}{b_n}\Big)\Big(G^\sigma(\tfrac{i}{n}, \tfrac{1}{p}, \mathcal{F}_i), G^\sigma(\tfrac{i}{n}, \tfrac{2}{p}, \mathcal{F}_i), \ldots, G^\sigma(\tfrac{i}{n}, \tfrac{p-1}{p}, \mathcal{F}_i), G^\sigma(\tfrac{i}{n}, 1, \mathcal{F}_i)\Big)^\top, \end{aligned}$$

where $G^\sigma(\frac{i}{n}, t, \mathcal{F}_i) = G(\frac{i}{n}, t, \mathcal{F}_i)/\sigma(\frac{i}{n}, t)$. Similarly as in Section 3 we consider the p-dimensional vector

$$Z_{i,l}^\sigma = Z_i^\sigma(\tfrac{l}{n}) = (Z_{i,l,1}^\sigma, \ldots Z_{i,l,p}^\sigma)^\top ,$$

where

$$Z_{i,l,k}^\sigma = G^\sigma(\tfrac{i}{n}, \tfrac{k}{p}, \mathcal{F}_i) K\Big(\frac{\frac{i}{n} - \frac{l}{n}}{b_n}\Big) \quad (1 \leq k \leq p).$$

Finally, we define the $(n - 2\lceil nb_n \rceil + 1)p$-dimensional vectors $\tilde{Z}_1^\sigma, \ldots \tilde{Z}_{2\lceil nb \rceil - 1}^\sigma$by

$$\tilde{Z}_j^\sigma = \big(Z_{j,\lceil nb_n \rceil}^{\sigma,\top}, Z_{j+1,\lceil nb_n \rceil+1}^{\sigma,\top}, \ldots, Z_{n-2\lceil nb_n \rceil+j,n-\lceil nb_n \rceil}^{\sigma,\top}\big)^\top \tag{4.2}$$

and obtain the following result.

Theorem 4.1. *Let the Assumptions of Theorem 3.2 be satisfied and assume that the partial derivative* $\frac{\partial^2 \sigma(u,t)}{\partial u \partial t}$ *exists and is bounded on* $(0,1)^2$. *Then there exist* $(n-2\lceil nb_n \rceil+1)p$*-dimensional centred*

Gaussian vectors $\tilde{Y}_1^\sigma, \ldots, \tilde{Y}_{2\lceil nb_n\rceil-1}^\sigma$ *with the same auto-covariance structure as the vector* $\tilde{Z}_i^\sigma$ *in* (4.2) *such that*

$$\begin{aligned}\mathfrak{P}_n^\sigma &:= \sup_{x\in\mathbb{R}} \Big|\mathbb{P}\Big(\max_{b_n\le u\le 1-b_n, 0\le t\le 1} \sqrt{nb_n}|\Delta^\sigma(u,t)| \le x\Big) - \mathbb{P}\Big(\Big|\frac{1}{\sqrt{nb_n}}\sum_{i=1}^{2\lceil nb_n\rceil-1} \tilde{Y}_i^\sigma\Big|_\infty \le x\Big)\Big| \\ &= O\Big((nb_n)^{-(1-11\iota)/8} + \Theta\big(\sqrt{nb_n}(b_n^4 + \frac{1}{nb_n}), np\big) \\ &\quad + \Theta\big(((np)^{1/q^*}((nb_n)^{-1} + 1/p))^{\frac{q^*}{q^*+1}}, np\big) + \Theta\big(b_n^{\frac{q-2}{q+1}}, np\big)\Big)\end{aligned}$$

for any sequence $p \to \infty$ *with* $np = O(\exp(n^\iota))$ *for some* $0 \le \iota < 1/11$*. In particular, for the choice* $p = \sqrt{n}$ *we have*

$$\mathfrak{P}_n^\sigma = o(1)$$

if the constant q^* *in Assumption 2.3 is sufficiently large.*

In practice, the long-run variance function in (4.1) is unknown and has to be replaced by a uniformly consistent estimator. For this propose recall the definition of the Epanechnikov kernel H in (3.8), and define for some bandwidth $\tau_n \in (0,1)$ the weights

$$\bar{\omega}(t,i) = H\Big(\frac{\frac{i}{n}-t}{\tau_n}\Big)\Big/\sum_{i=1}^n H\Big(\frac{\frac{i}{n}-t}{\tau_n}\Big).$$

Let $S_{k,r}(\cdot) = \sum_{i=k}^r X_{i,n}(\cdot)$ denote the partial sum of the data $X_{k,n}(\cdot), \ldots, X_{r,n}(\cdot)$ (note that these are functions) and define for $w \ge 2$

$$\Delta_j(t) = \frac{S_{j-w+1,j}(t) - S_{j+1,j+w}(t)}{w}.$$

An estimator of the long-run variance in (2.11) is then defined by

$$\hat{\sigma}^2(u,t) = \sum_{j=1}^n \frac{w\Delta_j^2(t)}{2}\bar{\omega}(u,j), \tag{4.3}$$

if $u \in [w/n, 1-w/n]$. For $u \in [0, w/n)$ and $u \in (1-w/n, 1]$ we define to be constant, that is $\hat{\sigma}^2(u,t) = \hat{\sigma}^2(w/n,t)$ and $\hat{\sigma}^2(u,t) = \hat{\sigma}^2(1-w/n,t)$, respectively. Our next result establishes the uniform consistency of this estimator.

Proposition 4.1. *Let the assumptions of Theorem 3.2 be satisfied and assume that the partial derivative* $\frac{\partial^2\sigma(u,t)}{\partial^2 u}$ *exists on the square* $(0,1)^2$*, is bounded and is continuous in* $u \in (0,1)$*. If*

$w \to \infty$, $w = o(n^{2/5})$, $w = o(n\tau_n)$, $\tau_n \to 0$ and $n\tau_n \to \infty$ we have that

$$\Big\| \sup_{\substack{u\in[\gamma_n,1-\gamma_n]\\ t\in(0,1)}} |\hat{\sigma}^2(u,t) - \sigma^2(u,t)| \Big\|_{q'} = O\Big(\frac{w^{5/2}}{n}\tau_n^{-1/q'} + w^{1/2}n^{-1/2}\tau_n^{-1/2-4/q'} + w^{-1} + \tau_n^2\Big),$$
$$\Big\| \sup_{\substack{u\in[0,\gamma_n)\cup(1-\gamma_n,1]\\ t\in(0,1)}} |\hat{\sigma}^2(u,t) - \sigma^2(u,t)| \Big\|_{q'} = O(g_n) \ ,$$

where

$$g_n = \frac{w^{5/2}}{n}\tau_n^{-1/q'} + w^{1/2}n^{-1/2}\tau_n^{-1/2-4/q'} + w^{-1} + \tau_n \ , \tag{4.4}$$

$\gamma_n = \tau_n + w/n$, $q' = \min(q, q^*)$ and q, q^* are defined in Assumptions 2.2 and 2.3, respectively.

To state the bootstrap algorithm for a simultaneous confidence surface of the form (2.3) with varying width, we introduce the following notation

$$\begin{aligned}\hat{Z}_i^{\hat{\sigma}}(u) &= (\hat{Z}_{i,1}^{\hat{\sigma}}(u), \ldots, \hat{Z}_{i,p}^{\hat{\sigma}}(u))^\top \\ &= K\Big(\frac{\frac{i}{n}-u}{b_n}\Big)\Big(\frac{\hat{\varepsilon}_{i,n}(\frac{1}{p})}{\hat{\sigma}(\frac{i}{n},\frac{1}{p})}, \frac{\hat{\varepsilon}_{i,n}(\frac{2}{p})}{\hat{\sigma}(\frac{i}{n},\frac{2}{p})}, \ldots, \frac{\hat{\varepsilon}_{i,n}(\frac{p-1}{p})}{\hat{\sigma}(\frac{i}{n},\frac{p-1}{p})}, \frac{\hat{\varepsilon}_{i,n}(1)}{\hat{\sigma}(\frac{i}{n},1)}\Big)^\top\end{aligned}$$

and consider the empirical analog

$$\hat{\tilde{Z}}_j^{\hat{\sigma}} = \big(\hat{Z}_{j,\lceil nb_n\rceil}^{\hat{\sigma},\top}, \hat{Z}_{j+1,\lceil nb_n\rceil+1}^{\hat{\sigma},\top} \ldots, \hat{Z}_{n-2\lceil nb_n\rceil+j,n-\lceil nb_n\rceil}^{\hat{\sigma},\top}\big)^\top \tag{4.6}$$

of the vector $\tilde{Z}_j^{\sigma}$ in (4.2), where $\hat{Z}_{i,l}^{\hat{\sigma}} = \hat{Z}_i^{\hat{\sigma}}(\frac{l}{n}) = (\hat{Z}_{i,l,1}^{\hat{\sigma}}, \ldots \hat{Z}_{i,l,p}^{\hat{\sigma}})^\top$.

Theorem 4.2. *Assume that the conditions of Theorem 3.3(ii), Proposition 4.1 and Theorem 4.1 hold and that there exists a sequence $\eta_n \to \infty$ such that*

$$\Theta\big(\sqrt{m_n \log np}(g_n\eta_n)(np)^{\frac{1}{q}}\big)^{q/(q+1)}, np) + \eta_n^{-q'} = o(1) \ ,$$

where γ_n and g_n are defined by (4.4) and q' and l are the constants in Proposition 4.1 and Theorem 3.3, respectively. Then the simultaneous confidence surface (4.5) in Algorithm 3 satisfies

$$\lim_{n\to\infty}\lim_{B\to\infty} \mathbb{P}(m \in \mathcal{C}_n^{\hat{\sigma}} \mid \mathcal{F}_n) = 1 - \alpha$$

in probability.

Remark 4.1. The methodology presented so far can be extended to construct a simultaneous confidence band for the vector of mean functions of a multivariate locally stationary functional

Algorithm 3:

Result: simultaneous confidence surface of the form (2.3)

(a) Calculate the the estimate of the long-run variance $\hat{\sigma}^2$ in (4.3)

(b) Calculate the $(n-2\lceil nb_n\rceil+1)p$-dimensional vectors $\hat{\tilde{Z}}_i^{\hat{\sigma}}$ in (4.6)

(c) For window size m_n, let $m_n' = 2\lfloor m_n/2\rfloor$,define the vectors $\hat{S}_{jm_n}^{\hat{\sigma},*} = \sum_{r=j}^{j+m_n-1}\hat{\tilde{Z}}_r^{\hat{\sigma}}$, and

$$\hat{S}_{jm_n'}^{\hat{\sigma}} = \hat{S}_{j,\lfloor m_n/2\rfloor}^{\hat{\sigma},*} - \hat{S}_{j+\lfloor m_n/2\rfloor,\lfloor m_n/2\rfloor}^{\hat{\sigma},*}$$

For $a \le b$ let $\hat{S}_{jm_n',[a:b]}^{\hat{\sigma}}$ be the $(b-a+1)$-dimensional sub-vector of the vector $\hat{S}_{jm_n'}^{\hat{\sigma}}$ containing its ath - bth components.

(d) **for** *r=1, ..., B* **do**

- Generate independent standard normal distributed random variables $\{R_i^{(r)}\}_{i\in[1,n-m_n']}$ and define the random vectors $V_k^{(r)} = (V_{k,1}^{(r)},\ldots,V_{k,2\lceil nb_n\rceil-m_n'+1}^{(r)})^\top$ by (3.14)
- Calculate

$$T_k^{\hat{\sigma},(r)} = \sum_{j=1}^{2\lceil nb_n\rceil-m_n'} \hat{S}_{jm_n',[(k-1)p+1:kp]}^{\hat{\sigma}} V_{k,j}^{(r)}\ , \quad k=1,\ldots,n-2\lceil nb_n\rceil+1,$$

$$T^{\hat{\sigma},(r)} = \max_{1\le k\le n-2\lceil nb_n\rceil+1} |T_k^{\hat{\sigma},(r)}|_\infty .$$

end

(e) Define $T_{\lfloor(1-\alpha)B\rfloor}^{\hat{\sigma}}$ as the empirical $(1-\alpha)$-quantile of the sample $T^{\hat{\sigma},(1)},\ldots,T^{\hat{\sigma},(B)}$ and

$$\hat{L}_3^{\hat{\sigma}}(u,t) = \hat{m}(u,t) - \hat{r}_3(u,t), \qquad \hat{U}_3^{\hat{\sigma}}(u,t) = \hat{m}(u,t) + \hat{r}_3(u,t),$$

where

$$\hat{r}_3(u,t) = \frac{\hat{\sigma}(u,t)\sqrt{2}T_{\lfloor(1-\alpha)B\rfloor}^{\hat{\sigma}}}{\sqrt{nb_n}\sqrt{m_n'(2\lceil nb_n\rceil - m_n')}}$$

Output:

$$\mathcal{C}_n^{\hat{\sigma}} = \big\{ f : [0,1]^2 \to \mathbb{R} \mid \ \hat{L}_3^{\hat{\sigma}}(u,t) \le f(u,t) \le \hat{U}_3^{\hat{\sigma}}(u,t) \ \ \forall u \in [b_n, 1-b_n]\ \forall t\in[0,1]\big\}. \quad (4.5)$$

time series. For simplicity we consider a 2-dimensional series of the form

$$\begin{pmatrix} X^1_{i,n}(t) \\ X^2_{i,n}(t) \end{pmatrix} = \begin{pmatrix} m_1(\frac{i}{n},t) \\ m_2(\frac{i}{n},t) \end{pmatrix} + \begin{pmatrix} \varepsilon^1_{i,n}(t) \\ \varepsilon^2_{i,n}(t) \end{pmatrix} , \tag{4.7}$$

and define for $a=1,2$

$$\begin{aligned} \hat{Z}_i^{a,\hat{\sigma}}(u) &= (\hat{Z}_{i,1}^{a,\hat{\sigma}}(u),\ldots,\hat{Z}_{i,p}^{a,\hat{\sigma}}(u))^\top \\ &= K\Big(\frac{\frac{i}{n}-u}{b_n}\Big)\Big(\frac{\hat{\varepsilon}^a_{i,n}(\frac{1}{p})}{\hat{\sigma}_a(\frac{i}{n},\frac{1}{p})},\frac{\hat{\varepsilon}^a_{i,n}(\frac{2}{p})}{\hat{\sigma}_a(\frac{i}{n},\frac{2}{p})},\ldots,\frac{\hat{\varepsilon}^a_{i,n}(\frac{p-1}{p})}{\hat{\sigma}_a(\frac{i}{n},\frac{p-1}{p})},\frac{\hat{\varepsilon}^a_{i,n}(1)}{\hat{\sigma}_a(\frac{i}{n},1)}\Big)^\top, \end{aligned}$$

where $\hat{\varepsilon}^a_{i,n}(t) = X^a_{i,n}(t) - \hat{m}_{a,d_n}(\frac{i}{n},t)$ and $\hat{m}_{a,d_n}(\frac{i}{n},t)$ is the local linear estimator of the function $m_a(\frac{i}{n},t)$ in (4.7) with bandwidth d_n (see equation (3.7) for its definition) and $\hat{\sigma}^2_a(\frac{i}{n},t)$ is the estimator of long-variance of $\varepsilon^a_{i,n}(t)$ defined in (4.3). Next we consider the $2(n-2\lceil nb_n\rceil+1)p$ vector

$$\hat{\tilde{Z}}_j^{\hat{\sigma}} = \big(\hat{Z}^{\hat{\sigma},\top}_{j,\lceil nb_n\rceil},\hat{Z}^{\hat{\sigma},\top}_{j+1,\lceil nb_n\rceil+1}\ldots,\hat{Z}^{\hat{\sigma},\top}_{n-2\lceil nb_n\rceil+j,n-\lceil nb_n\rceil}\big)^\top$$

where $\hat{Z}^{\hat{\sigma}}_{i,l} = \hat{Z}^{\hat{\sigma}}_i(\frac{l}{n}) = (\hat{Z}^{1,\hat{\sigma}}_{i,l,1},\hat{Z}^{2,\hat{\sigma}}_{i,l,1}\ldots\hat{Z}^{1,\hat{\sigma}}_{i,l,p},\hat{Z}^{2,\hat{\sigma}}_{i,l,p})^\top$ contains information from both components. Define for $a=1,2$

$$\hat{L}^{\hat{\sigma}}_{3,a}(u,t) = \hat{m}_a(u,t) - \hat{r}_{3,a}(u,t), \quad \hat{U}^{\hat{\sigma}}_{3,a}(u,t) = \hat{m}_a(u,t) + \hat{r}_{3,a}(u,t)$$

where

$$\hat{r}_{3,a}(u,t) = \frac{\hat{\sigma}_a(u,t)\sqrt{2}T^{\hat{\sigma}}_{\lfloor(1-\alpha)B\rfloor}}{\sqrt{nb_n}\sqrt{m'_n(2\lceil nb_n\rceil - m'_n)}}$$

and $T^{\hat{\sigma}}_{\lfloor(1-\alpha)B\rfloor}$ is generated in the same way as in step (e) of Algorithm 3 with p replaced by $2p$, $\hat{m}_a(u,t)$ is the kernel estimator of $m_a(u,t)$ defined in (2.6). Further, define for $a=1,2$ the set of functions

$$\mathcal{C}^{\hat{\sigma}}_{a,n} = \big\{f:[0,1]^2\to\mathbb{R} \mid \ \hat{L}_{3,a}(u,t)\le f(u,t)\le \hat{U}_{3,a}(u,t) \ \ \forall u\in[b_n,1-b_n]\ \forall t\in[0,1]\big\}.$$

Suppose that the mean functions and error processes of $X^1_{i,n}(t)$ and $X^2_{i,n}(t)$ satisfy the conditions of Theorem 4.2, then it can be proved that

$$\lim_{n\to\infty}\lim_{B\to\infty}\mathbb{P}(m_1\in\mathcal{C}^{\hat{\sigma}}_{1,n}, m_2\in\mathcal{C}^{\hat{\sigma}}_{2,n}\mid \mathcal{F}_n) = 1-\alpha$$

in probability.

5 Finite sample properties

In this section we study the finite sample performance of the simultaneous confidence bands and surfaces proposed in the previous sections. We start giving some more details regarding the general implementation of the algorithms. A simulation study and a data example are presented in Section 5.2 and 5.3, respectively.

5.1 Implementation

For the estimator of the regression function in (2.6) we use the kernel (of order 4)

$$K(x) = (45/32 - 150x^2/32 + 105x^4/32)\mathbf{1}(|x| \leq 1) \ ,$$

and for the bandwidth we choose $b_n = 1.2d_n$. Here, for the confidence band in (2.3), the parameter d_n is chosen as the minimizer of

$$MGCV(b) = \max_{1 \leq s \leq p} \frac{\sum_{i=1}^{n} (\hat{m}_b(\frac{i}{n}, \frac{s}{p}) - X_{i,n}(\frac{s}{p}))^2}{(1 - \mathrm{tr}(Q_s(b))/n)^2} \ , \tag{5.1}$$

$p = \lceil \sqrt{n} \rceil$ and $Q_s(b)$ is an $n \times n$ is the matrix defining the local linear estimator in (3.7), that is

$$\big(\hat{m}_b(\tfrac{1}{n}, \tfrac{s}{p}), \hat{m}_b(\tfrac{2}{n}, \tfrac{s}{p}), \ldots, \hat{m}_b(1, \tfrac{s}{p})\big)^\top = Q_s(b)\big(X_{1,n}(\tfrac{s}{p}), ..., X_{n,n}(\tfrac{s}{p})\big)^\top.$$

For the simultaneous confidence band (for a fixed $u \in (0, 1)$) in (2.2) d_n is defined similarly, where the loss function in (5.1) is replaced by

$$MGCV(b) = \max_{1 \leq s \leq p} \frac{\sum_{i=\lceil nu - nb_n \rceil}^{\lfloor nu + nb_n \rfloor} (\hat{m}_b(\frac{i}{n}, \frac{s}{p}) - X_{i,n}(\frac{s}{p}))^2}{(1 - \mathrm{tr}(Q_s(b,u))/(2nb))^2}. \tag{5.2}$$

and

$$\big(\hat{m}_b(\tfrac{\lceil nu - nb_n \rceil}{n}, \tfrac{s}{p}), ..., \hat{m}_b(\tfrac{\lfloor nu + nb_n \rfloor}{n}, \tfrac{s}{p})\big)^\top = Q_s(b,u)\big(X_{\lceil nu - nb_n \rceil, n}(\tfrac{s}{p}), ..., X_{\lfloor nu + nb_n \rfloor, n}(\tfrac{s}{p})\big)^\top,$$

The criteria (5.1) and (5.2) are motivated by the generalized cross validation criterion introduced by Craven and Wahba (1978) and will be called Maximal Generalized Cross Validation (MGCV) method throughout this paper.

For the estimator of the long-run variance in (4.3) we use $w = \lfloor n^{2/7} \rfloor$ and $\tau_n = n^{-1/7}$ as recommended in Dette and Wu (2019). The window size in the multiplier bootstrap is then selected by the minimal volatility method advocated by Politis et al. (1999). For the sake of brevity, we discuss this method only for Algorithm 3 in detail (the choice for Algorithm 1 and 2, and for the algorithms in the online supplement is similar). We consider a grid of window sizes $\tilde{m}_1 < \ldots < \tilde{m}_M$ (for some integer M). We first calculate $\hat{S}^{\hat{\sigma}}_{j\tilde{m}_s}$ defined in step (c) of Algorithm 3 for each $\tilde{m}_s$. Let

$\hat{S}^{\hat{\sigma},\diamond}_{\tilde{m}_s}$ denote the $(n-2\lceil nb_n\rceil+1)p$ dimensional vector with r_{th} entry defined by

$$\hat{S}^{\hat{\sigma},\diamond}_{\tilde{m}_s,r} = \frac{1}{\tilde{m}_s(2\lceil nb_n\rceil - \tilde{m}_s)} \sum_{j=1}^{2\lceil nb_n\rceil-\tilde{m}_s} (\hat{S}^{\hat{\sigma}}_{j\tilde{m}_s,r})^2 ,$$

and consider the standard error of $\{\hat{S}^{\hat{\sigma},\diamond}_{\tilde{m}_s,r}\}_{s=k-2}^{k+2}$, that is

$$\mathrm{se}\big(\{\hat{S}^{\hat{\sigma},\diamond}_{\tilde{m}_s,r}\}_{s=k-2}^{k+2}\big) = \Big(\frac{1}{4}\sum_{s=k-2}^{k+2}\Big(\hat{S}^{\hat{\sigma},\diamond}_{\tilde{m}_s,r} - \frac{1}{5}\sum_{s=k-2}^{k+2}\hat{S}^{\hat{\sigma},\diamond}_{\tilde{m}_s,r}\Big)^2\Big)^{1/2}$$

Then we choose $m_n' = \tilde{m}_j$ where j is defined as the minimizer of the function

$$MV(k) = \frac{1}{(n-2\lceil nb_n\rceil+1)p} \sum_{r=1}^{(n-2\lceil nb_n\rceil+1)p} se\big(\{\hat{S}^{\hat{\sigma},\diamond}_{\tilde{m}_s,r}\}_{s=k-2}^{k+2}\big) .$$

5.2 Simulated data

We consider two regression functions

$$m_1(u,t) = (1+u)(6(t-0.5)^2+1),$$
$$m_2(u,t) = (1+u^2)(6(t-0.5)^2(1+\mathbf{1}(t>0.3))+1)$$

(note that m_2 is discontinuous at the point $t=0.3$). For the definition of the error processes let $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ be a sequence of independent standard normally distributed random variables and $\{\eta_i\}_{i\in\mathbb{Z}}$ be a sequence of independent t-distributed random variables with 8 degrees of freedom. Define the functions

$$a(t) = 0.5\cos(\pi t/3), \qquad b(t) = 0.4t, \qquad c(t) = 0.3t^2,$$
$$d_1(t) = 1+0.5\sin(\pi t), \qquad d_{2,1}(t) = 2t-1, \qquad d_{2,2}(t) = 6t^2-6t+1,$$

and $\mathcal{F}_i^1 = (\ldots,\varepsilon_{i-1},\varepsilon_i)$, $\mathcal{F}_i^2 = (\ldots,\eta_{i-1},\eta_i)$. We consider the following two locally stationary time series models defined by

$$G_1(t,\mathcal{F}_i^1) = a(t)G_1(t,\mathcal{F}_{i-1}^1) + \varepsilon_i,$$
$$G_2(t,\mathcal{F}_i^2) = b(t)G_2(t,\mathcal{F}_{i-1}^2) + \eta_i - c(t)\eta_{i-1}.$$

Note that G_1 is a locally stationary AR(1) process (or equivalently a locally stationary MA(∞) process), and that G_2 is a locally stationary ARMA$(1,1)$ model. With these processes we define the following functional time series model (for $1\le i\le n$, $0\le t\le 1$)

(a) $X_{i,n}(t) = m_1(\frac{i}{n},t) + G_1(\frac{i}{n},\mathcal{F}_i^1)d_1(t)/3.$

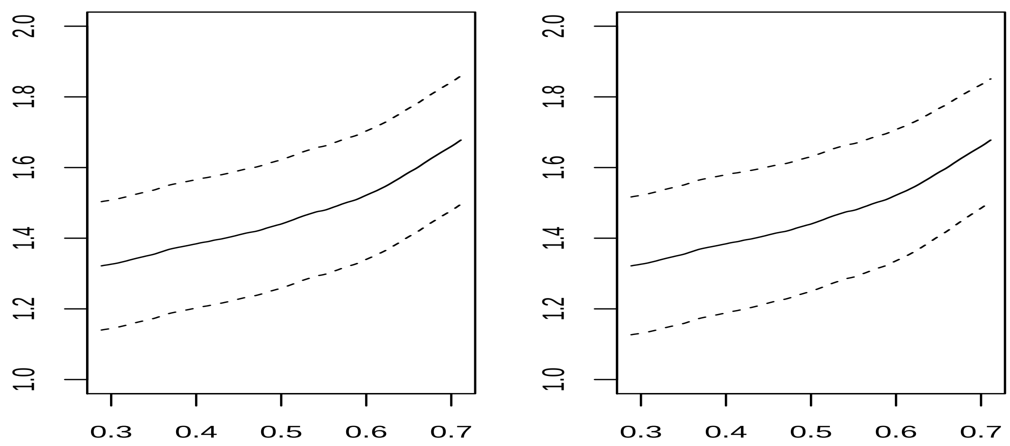

Figure 1: 95% *simultaneous confidence bands of the form* (2.1) *(fixed* $t = 0.5$*) for the regression function in model (a) from* $n = 800$ *observations. Left panel: constant width (Algorithm A.4); Right panel: varying width (Algorithm A.5).*

(b) $X_{i,n}(t) = m_1(\frac{i}{n}, t) + G_1(\frac{i}{n}, \mathcal{F}_i^1)d_{2,1}(t)/2 + G_2(\frac{i}{n}, \mathcal{F}_i^2)d_{2,2}(t)/2$

(c) $X_{i,n}(t) = m_2(\frac{i}{n}, t) + G_1(\frac{i}{n}, \mathcal{F}_i^1)d_1(t)/3$.

(d) $X_{i,n}(t) = m_2(\frac{i}{n}, t) + G_1(\frac{i}{n}, \mathcal{F}_i^1)d_{2,1}(t)/2 + G_2(\frac{i}{n}, \mathcal{F}_i^2)d_{2,2}(t)/2.$

We begin displaying typical 95% simultaneous confidence bands obtained from one simulation run for model (a) with sample size $n = 800$. Figure 1 shows the confidence bands of the form (2.1) (for fixed t) with constant and variable width (calculated by Algorithm A.4 and A.5, see Section A of the online supplement). In Figure 2 we display the simultaneous confidence bands of the form (2.2) (for fixed u) with constant width (Algorithm 1) and variable width (Algorithm A.6, see Section A of the online supplement for details), while Figure 3 shows the simultaneous confidence surface of the form (2.3) calculated by Algorithm 2 and 3 described in Section 3 and 4, respectively. We observe that in all cases there exist differences between the bands (surfaces) with constant and variable width, but they are not substantial.
We next investigate the coverage probabilities of the different confidence bands constructed in this paper for sample sizes $n = 500$ and $n = 800$. All results presented in the following discussion are based on 1000 simulation runs and $B = 1000$ bootstrap replications. In all tables the left part shows the coverage probabilities of the bands with constant width while the results in the right part correspond to the bands with varying width. In Table 1 we give some results for the confidence bands of the form (2.1) (for fixed $t = 0.5$) with constant and variable width (Algorithm A.4 and Algorithm A.5 in Section A of the online supplement), while we present in Table 2 the simulated coverage probabilities of the simultaneous confidence bands of the form (2.2), where $u = 0.5$ is

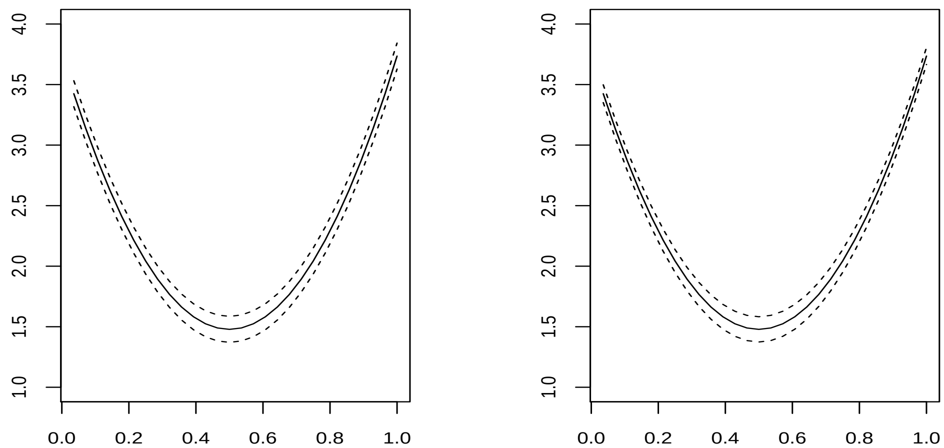

Figure 2: 95% *simultaneous confidence band of the form* (2.1) *(fixed* $u = 0.5$*) for the regression function in model (a) from* $n = 800$ *observations. Left panel: constant width (Algorithm 1); Right panel: varying width (Algorithm A.6).*

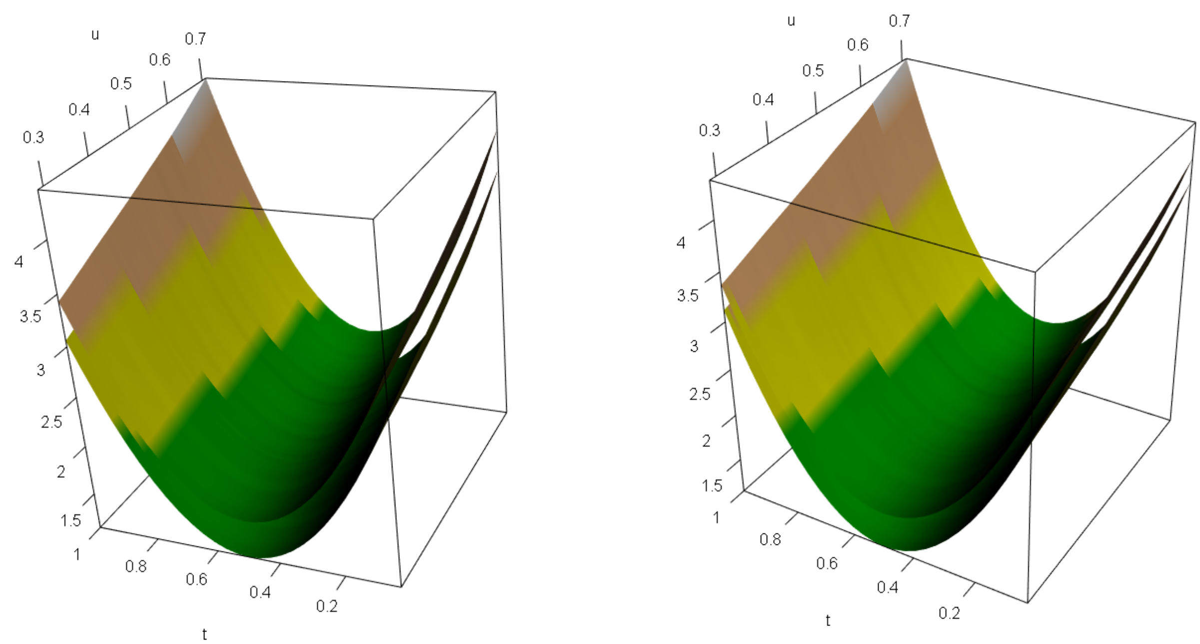

Figure 3: 95% *simultaneous confidence surface of the form* (2.3) *for the regression function in model (a) from* $n = 800$ *observations. Left panel: constant width (Algorithm 2); Right panel: varying width (Algorithm 3)*

fixed (Algorithm 1 and Algorithm A.6 in the online supplement). Corresponding results for the simultaneous confidence surfaces of the form (2.3) can be found in Table 3 (Algorithm 2 (constant width) and 3 (varying width)). We observe that the simulated coverage probabilities are close to their nominal levels in all cases under consideration, which illustrates the validity of our methods for finite sample sizes.

	constant width				varying width			
	model (a)		model (b)		model (a)		model (b)	
level	90%	95%	90%	95%	90%	95%	90%	95%
$n = 500$	90.4 %	94.9%	89.8 %	95.9%	90.1.5 %	96.3%	89.9 %	94.6%
$n = 800$	89.3 %	94.5 %	90.0 %	95.5 %	89.4 %	95.0 %	90.3 %	95.6 %
	model (c)		model (d)		model (c)		model (d)	
level	90%	95%	90%	95%	90%	95%	90%	95%
$n = 500$	90.1 %	95.9%	90.4 %	96.0%	90.4 %	95.3%	91.0 %	96.4%
$n = 800$	90.6 %	95.6 %	90.0 %	95.4 %	88.9 %	94.9 %	89.3 %	95.1 %

Table 1: *Simulated coverage probabilities of the simultaneous confidence band of the form* (2.1) *for fixed* $t = 0.5$ *calculated by Algorithm A.4 (constant width) and A.5 (varying width).*

	constant width				varying width			
	model (a)		model (b)		model (a)		model (b)	
level	90%	95%	90%	95%	90%	95%	90%	95%
$n = 500$	87.8 %	92.3%	88.5 %	93.6%	88.5 %	93.3%	88.1 %	92.5%
$n = 800$	88.7 %	93.9 %	89.0 %	94.4 %	90.8 %	94.3 %	88.2 %	93.9 %
	model (c)		model (d)		model (c)		model (d)	
level	90%	95%	90%	95%	90%	95%	90%	95%
$n = 500$	87.4 %	92.4%	88.0 %	93.1%	88.1 %	93.3%	89.6 %	94.6%
$n = 800$	88.9 %	93.6 %	88.7 %	94.5 %	89.6 %	93.8 %	89.9 %	95.2 %

Table 2: *Simulated coverage probabilities of the simultaneous confidence band of the form* (2.2) *for fixed* $u = 0.5$ *calculated by Algorithms 1 (constant width) and A.6 (varying width).*

5.3 Real data

In this section we illustrate the proposed methodology analyzing the implied volatility (IV) of the European call option of SP500. These options are contracts such that their holders have the right to buy the SP500 at a specified price (strike price) on a specified date (expiration date). The implied volatility is derived from the observed SP500 option prices, directly observed parameters, such as risk-free rate and expiration date, and option pricing methods and is widely used in the studies of quantitative finance. For more details, we refer to Hull (2003).

	constant width				varying width			
	model (c)		model (d)		model (c)		model (d)	
level	90%	95%	90%	95%	90%	95%	90%	95%
$n = 500$	90.1 %	95.1%	89.1 %	95.1%	90.3 %	95.6%	89.5 %	95.3%
$n = 800$	89.6 %	94.8 %	90.5 %	95.0 %	90.4 %	95.0 %	89.4 %	94.8 %
	model (c)		model (d)		model (c)		model (d)	
level	90%	95%	90%	95%	90%	95%	90%	95%
$n = 500$	88.9 %	94.8%	89.7 %	95.4%	91.0 %	95.7%	88.3 %	94.3%
$n = 800$	90.1 %	95.1 %	90.8 %	95.5 %	90.1 %	95.8 %	89.7 %	95.5 %

Table 3: *Simulated coverage probabilities of the simultaneous confidence band of the form* (2.3) *calculated by Algorithm 2 (constant width) and Algorithm 3 (varying width).*

We collect the implied volatility and the strike price from the 'optionmetrics' database and the SP500 index from the CRSP database. Both databases can be accessed from Wharton Research Data Service (WRDS). We calulate the simultaneous confidence band for the implied volatility surface, which is a two variate function of time (more precisely time to maturity) and moneyness, where the moneyness is calculated using strike price divided by SP500 indices. The options are collected from December 21, 2016 to July 19, 2019, and the expiration date is December 20, 2019. Therefore the length of time series is 647. Within each day we observe the volatility curve, which is the implied volatility as a function of moneyness.

Recently, Liu et al. (2016) models IV via functional time series. Following their perspective, we shall study the IV data via model (1.1), where $X_{i,n}(t)$ represents the observed volatility curve at a day i, with total sample size $n = 647$. We consider the options with moneyness in the range of $[0.8, 1.4]$, corresponding to options that have been actively traded in this period (note that, our methodology was developed for functions on the interval $[0, 1]$, but it is obvious how to extend this to an arbitrary compact interval $[a, b]$). The number of observations for each day varies from 34 to 56, and we smooth the implied volatility using linear interpolation and constant extrapolation.

To study the well documented volatility smile, we calculate confidence bands of the form (2.2) for fixed $u = 0.5$, by Algorithm 1 (constant width) and Algorithm A.6 (varying width). The parameter selection procedure proposed in Section 5.3 yields $b_n = 0.216$ and $m_n = 18$, and the resulting simultaneous confidence bands of the form (2.2) are presented in Figure 4. We observe that both 95% simultaneous confidence bands indicate that the implied volatility is a quadratic function of moneyness, which supports the well documented phenomenon of 'volatility smile'. In Figure 5 We also display 95% simultaneous confidence bands of the form (2.1) for fixed $t = 0.5$ (which corresponds to Moneyness=1.1). We observe that the volatility changes with time when moneyness (or equivalently, the strike price and underlying asset price) is specified. We observe that the differences between the bands with constant and variable width are rather small.

In practice it is important to determine whether the volatility curve changes with time i.e., to test $H_0 : m(u, t) \equiv m(t)$. As pointed out by Daglish et al. (2007), the volatility surface of an asset would be flat and unchanging if the assumptions of Black–Scholes (Black and Scholes, 1973) hold. In particular, Daglish et al. (2007) demonstrate that for most assets the volatility surfaces are

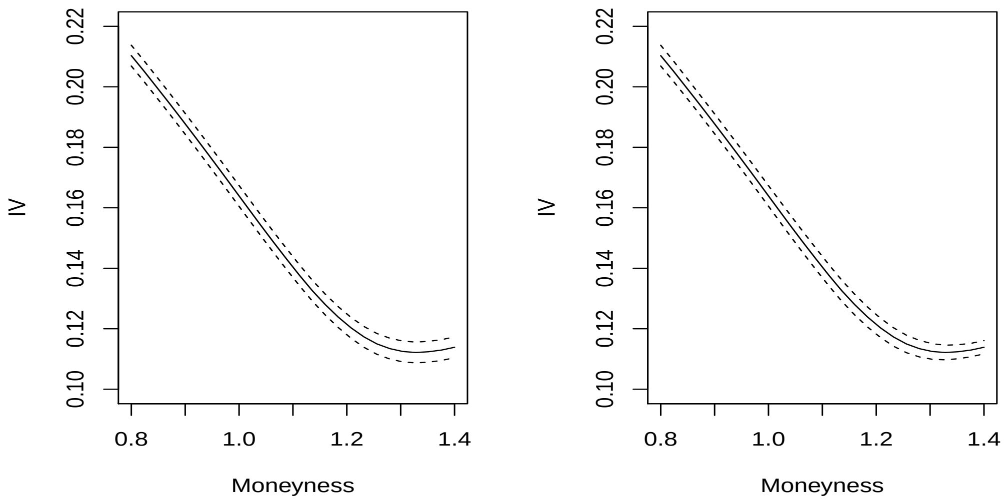

Figure 4: *95% simultaneous confidence bands of the form* (2.2) *(fixed* $u = 0.5$*) for the IV surface. Left panel: constant width (Algorithm 1); Right panel: variable width (Algorithm A.6).*

not flat and are stochastically changing in practice. We can provide an inference tool for such a conclusion using the simultaneous confidence bands and surfaces developed in Section 3 and 4. For example, note, that by the duality between confidence regions and hypotheses tests, an asymptotic level α test for the hypothesis $H_0 : m(u, t) \equiv m(t)$ is obtained by rejecting the null hypothesis, whenever the surface of the form $m(u, t) = m(t)$ is not contained in an $(1 - \alpha)$ simultaneous confidence surface of the form (2.3).
Therefore we construct the 95% simultaneous confidence surface for the regression function m with constant and varying width using Algorithms 2 and Algorithm 3, respectively. The parameter chosen by procedures in Section 5.3 are $b_n = 0.12$ and $m_n = 18$. The results are depicted in Figure 6 (for a better illustration the z-axis shows 100× implied volatility). We observe from both figures that simultaneous confidence bands do not contain a surface of the form $m(u, t) = m(t)$ and therefore reject the null hypothesis (at significance level 0.05%).

Online Supplemental Material: the online supplemental material contains further algorithms of confidence bands of type (2.1), and of type (2.2) with varying width. It also provides all the detailed technical proofs.

Acknowledgements This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt A1,C1) of the German Research Foundation (DFG). Weichi Wu gratefully acknowledges NSFC Young Program (no. 11901337) and BJNSF (no. Z190001) of China. The authors would like to thank Martina Stein, who typed parts of this paper with considerable expertise.

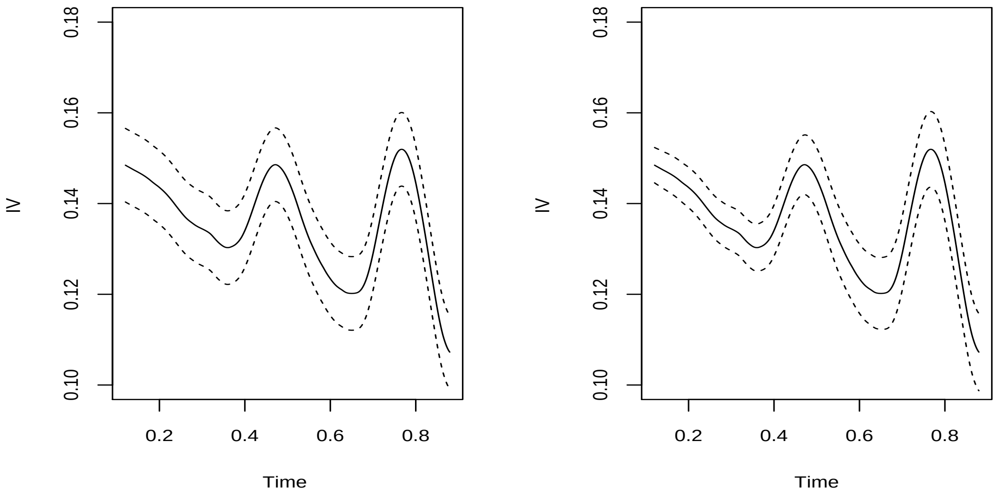

Figure 5: *95% simultaneous confidence bands of the form* (2.2) *(fixed* $t = 0.5$*). Left panel: constant width (Algorithm A.4); Right panel: variable width (Algorithm A.5).*

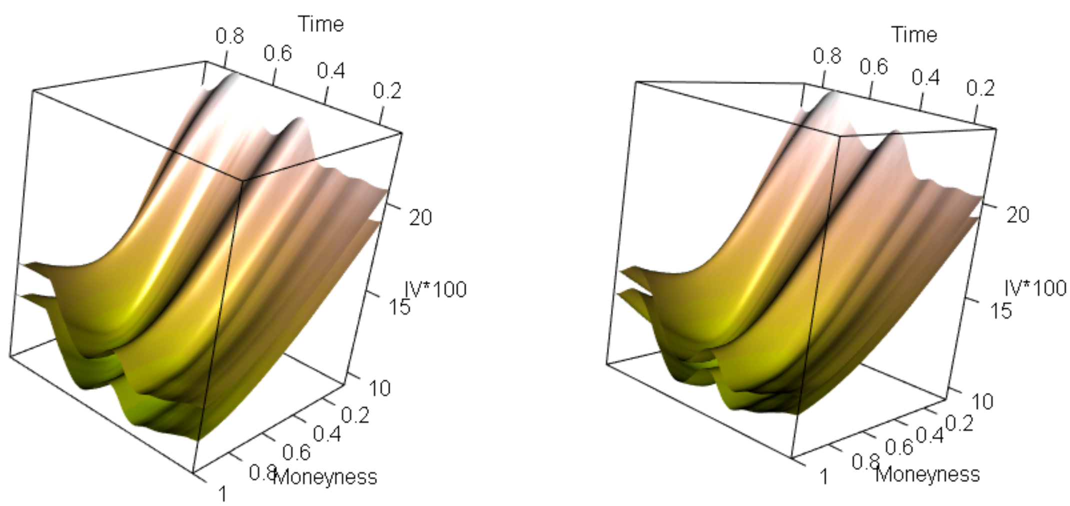

Figure 6: *95% simultaneous confidence band of the form* (2.3) *for the IV surface. Left panel: constant width (Algorithm 2); Right panel: variable width (Algorithm 3).*

Supplemental for 'Confidence surfaces for the mean of locally stationary functional time series'

Section A of the supplemental material contains algorithms for the simultaneous confidence bands of the form (2.1) and for the confidence surface of the form (2.3) with varying width. Section B.1 provides proofs of our theoretical results in Sections 3 and 4 of the main article, while Section C provides further auxiliary results.

A Further algorithms

In this section we provide the remaining algorithms for the calculation of the simultaneous confidence bands of the form (2.1) (fixed t) and for a simultaneous confidence surface of the form (2.3) with varying width. Recall the definition of the residuals $\hat{\varepsilon}_{i,n}(t)$ in (3.9) and of the long run variance estimator $\hat{\sigma}$ in (4.3) in the main article.
We begin with the simultaneous confidence bands for a fixed $t \in [0,1]$ defined in (2.1) and define

$$\hat{Z}_i(u,t) = K\Big(\frac{\frac{i}{n}-u}{b_n}\Big)\hat{\varepsilon}_{i,n}(t), \hat{Z}_{i,l}(t) = \hat{Z}_i(\tfrac{l}{n},t),$$
$$\hat{Z}_i^{\hat{\sigma}}(u,t) = K\Big(\frac{\frac{i}{n}-u}{b_n}\Big)\frac{\hat{\varepsilon}_{i,n}(t)}{\hat{\sigma}(\frac{i}{n},t)}, \quad \hat{Z}_{i,l}^{\hat{\sigma}}(t) = \hat{Z}_i^{\hat{\sigma}}(\tfrac{l}{n},t).$$

Next we consider the $(n-2\lceil nb_n\rceil+1)$-dimensional vectors

$$\hat{\hat{Z}}_j(t) = \big(\hat{Z}_{j,\lceil nb_n\rceil}(t), \hat{Z}_{j+1,\lceil nb_n\rceil+1}(t), \ldots, \hat{Z}_{n-2\lceil nb_n\rceil+j,n-\lceil nb_n\rceil}(t)\big)^\top, \tag{A.3}$$
$$\hat{\hat{Z}}_j^{\hat{\sigma}}(t) = \big(\hat{Z}^{\hat{\sigma}}_{j,\lceil nb_n\rceil}(t), \hat{Z}^{\hat{\sigma}}_{j+1,\lceil nb_n\rceil+1}(t), \ldots, \hat{Z}^{\hat{\sigma}}_{n-2\lceil nb_n\rceil+j,n-\lceil nb_n\rceil}(t)\big)^\top \tag{A.4}$$

$(1 \le j \le 2\lceil nb_n\rceil - 1)$, then a simultaneous confidence band for fixed $t \in [0,1]$ can be generated by the following algorithms A.4 (constant width) algorithm A.5 (varying width). The proof is omitted for the sake of brevity.

Theorem A.1. *Assume that the conditions of Theorem 3.2 hold. Define*

$$\vartheta_n^\dagger = \frac{\log^2 n}{m_n} + \frac{m_n}{nb_n} + \sqrt{\frac{m_n}{nb_n}} n^{4/q}.$$

(i) *If* $\vartheta_n^{\dagger,1/3}\{1 \vee \log(\frac{n}{\vartheta_n^\dagger})\}^{2/3} + \Theta\big(\big(\sqrt{m_n \log n}(\frac{1}{\sqrt{nd_n}} + d_n^2)(n)^{\frac{1}{q}}\big)^{q/(q+1)}, n\big) = o(1)$ *we have that for any* $\alpha \in (0,1)$

$$\lim_{n\to\infty} \lim_{B\to\infty} \mathbb{P}(m \in \mathcal{C}_n(t) \mid \mathcal{F}_n) = 1-\alpha$$

in probability.

(ii) *If further the conditions of Theorem 4.1 and Proposition 4.1 hold, then*

$$\lim_{n\to\infty} \lim_{B\to\infty} \mathbb{P}(m \in \mathcal{C}_n^{\hat\sigma}(t) \mid \mathcal{F}_n) = 1-\alpha$$

in probability.

Next we present a detailed algorithm to estimate simultaneous confidence band for a fixed u (of the form (2.2)) with varying width. For this purpose we define the p-dimensional vector

$$\begin{aligned}\hat{Z}_i^{\hat\sigma_u}(u) &= (\hat{Z}_{i,1}^{\hat\sigma_u}(u), \ldots, \hat{Z}_{i,p}^{\hat\sigma_u}(u))^\top \\ &= K\Big(\frac{\frac{i}{n}-u}{b_n}\Big)\Big(\frac{\hat\varepsilon_{i,n}(\frac{1}{p})}{\hat\sigma(u,\frac{1}{p})}, \frac{\hat\varepsilon_{i,n}(\frac{2}{p})}{\hat\sigma(u,\frac{2}{p})}, \ldots, \frac{\hat\varepsilon_{i,n},(\frac{p-1}{p})}{\hat\sigma(u,\frac{p-1}{p})}, \frac{\hat\varepsilon_{i,n}(1)}{\hat\sigma(u,1)}\Big)^\top.\end{aligned} \tag{A.6}$$

We present the following Algorithm A.6 which gives out an asymptotic correct simultaneous confidence band with varying width. The validity of the algorithm can be proved in a similar way to the proof of Theorem 4.2. The details are omitted for the sake of brevity.

Algorithm A.4:

Result: simultaneous confidence band of the form (2.1) with fixed width

(a) Calculate the the $(n - 2\lceil nb_n \rceil + 1)$-dimensional vector $\hat{\tilde{Z}}_j(t)$ in (A.3);

(b) For window size m_n, let $m'_n = 2\lfloor m_n/2 \rfloor$, define vectors $\hat{S}^*_{jm_n}(t) = \sum_{r=j}^{j+m_n-1} \hat{\tilde{Z}}_r(t)$, and

$$\hat{S}_{jm'_n}(t) = \hat{S}^*_{j,\lfloor m_n/2 \rfloor}(t) - \hat{S}^*_{j+\lfloor m_n/2 \rfloor, \lfloor m_n/2 \rfloor}(t) \tag{A.5}$$

Let $\hat{S}_{jm'_n,k}(t)$ be the k_{th} component of $\hat{S}_{jm'_n}(t)$.

(c) **for** *r=1, …, B* **do**

- Generate independent standard normal distributed random variables $\{R_i^{(r)}\}_{i \in [1, n-m'_n]}$ and $(2\lceil nb_n \rceil - m'_n)$-dimensional random vectors

$$V_k^{(r)} = (V_{k,1}^{(r)}, \ldots, V_{k,2\lceil nb_n \rceil - m'_n}^{(r)})^\top := (R_k^{(r)}, \ldots, R_{k+2\lceil nb_n \rceil - m'_n - 1}^{(r)})^\top .$$

- Calculate

$$T_k^{(r)}(t) = \sum_{j=1}^{2\lceil nb_n \rceil - m'_n} \hat{S}_{jm'_n,k}(t) V_{k,j}^{(r)} \ , \quad k = 1, \ldots, n - 2\lceil nb_n \rceil + 1,$$

$$T^{(r)}(t) = \max_{1 \le k \le n - 2\lceil nb_n \rceil + 1} |T_k^{(r)}(t)|.$$

end

(d) Define $T_{\lfloor (1-\alpha)B \rfloor}(t)$ as the empirical $(1-\alpha)$-quantile of the sample $T^{(1)}(t), \ldots, T^{(B)}(t)$ and

$$\hat{L}_4(u,t) = \hat{m}(u,t) - \hat{r}_4(t) \quad , \qquad \hat{U}_4(u,t) = \hat{m}(u,t) + \hat{r}_4(t)$$

where

$$\hat{r}_4(t) = \frac{\sqrt{2} T_{\lfloor (1-\alpha)B \rfloor}(t)}{\sqrt{nb_n} \sqrt{m'_n (2\lceil nb_n \rceil - m'_n)}}$$

Output: $\mathcal{C}_n(t) = \left\{ f : [0,1]^2 \to \mathbb{R} \mid \ \hat{L}_4(u,t) \le f(u,t) \le \hat{U}_4(u,t) \ \ \forall u \in [0,1] \right\}$

Algorithm A.5:

Result: simultaneous confidence band of the form (2.1) with varying width

(a) Calculate the estimate of the long-run variance $\hat{\sigma}^2$ in (4.3)

(b) Calculate the $(n - 2\lceil nb_n \rceil + 1)$-dimensional vectors $\hat{\tilde{Z}}_j^{\hat{\sigma}}(t)$ in (A.4)

(c) For window size m_n, let $m_n' = 2\lfloor m_n/2 \rfloor$, define the vectors $\hat{S}_{jm_n}^{\hat{\sigma},*}(t) = \sum_{r=j}^{j+m_n-1} \hat{\tilde{Z}}_r^{\hat{\sigma}}(t)$ and

$$\hat{S}_{jm_n'}^{\hat{\sigma}}(t) = \hat{S}_{j,\lfloor m_n/2 \rfloor}^{\hat{\sigma},*}(t) - \hat{S}_{j+\lfloor m_n/2 \rfloor,\lfloor m_n/2 \rfloor}^{\hat{\sigma},*}(t)$$

Let $\hat{S}_{jm_n',k}^{\hat{\sigma}}(t)$ be the k_{th} component of $\hat{S}_{jm_n'}^{\hat{\sigma}}(t)$.

(d) **for** *r=1, ..., B* **do**

- Generate independent standard normal distributed random variables $\{R_i^{(r)}\}_{i \in [1, n-m_n']}$ and define $V_k^{(r)} = (V_{k,1}^{(r)}, \ldots, V_{k,2\lceil nb_n \rceil - m_n' + 1}^{(r)})^\top$ by (3.14)
- Calculate

$$T_k^{\hat{\sigma},(r)}(t) = \sum_{j=1}^{2\lceil nb_n \rceil - m_n'} \hat{S}_{jm_n',k}^{\hat{\sigma}}(t) V_{k,j}^{(r)} \ , \quad k = 1, \ldots, n - 2\lceil nb_n \rceil + 1,$$

$$T^{\hat{\sigma},(r)}(t) = \max_{1 \le k \le n - 2\lceil nb_n \rceil + 1} |T_k^{\hat{\sigma},(r)}(t)|.$$

end

(e) Define $T_{\lfloor (1-\alpha)B \rfloor}^{\hat{\sigma}}(t)$ as the empirical $(1-\alpha)$-quantile of the sample $T^{\hat{\sigma},(1)}(t), \ldots, T^{\hat{\sigma},(B)}(t)$ and

$$\hat{L}_5^{\hat{\sigma}}(u,t) = \hat{m}(u,t) - \hat{r}_5(u,t), \qquad \hat{U}_5^{\hat{\sigma}}(u,t) = \hat{m}(u,t) + \hat{r}_5(u,t)$$

where

$$\hat{r}_5(u,t) = \frac{\hat{\sigma}(u,t)\sqrt{2} T_{\lfloor (1-\alpha)B \rfloor}^{\hat{\sigma}}(t)}{\sqrt{nb_n}\sqrt{m_n'(2\lceil nb_n \rceil - m_n')}}$$

Output:

$$\mathcal{C}_n^{\hat{\sigma}}(t) = \left\{ f : [0,1]^2 \to \mathbb{R} \mid \ \hat{L}_5^{\hat{\sigma}}(u,t) \le f(u,t) \le \hat{U}_5^{\hat{\sigma}}(u,t) \ \ \forall u \in [0,1] \right\}.$$

Algorithm A.6:

Result: simultaneous confidence band of the form (2.2) with varying width.

(a) For given $u \in [b_n, 1-b_n]$, calculate the the estimate of the long-run variance $\hat{\sigma}^2(u,\cdot)$ in (4.3)

(b) Calculate the vector $\hat{Z}_i^{\hat{\sigma}_u}(u)$ in (A.6);

(c) For window size m_n, let $m_n' = 2\lfloor m_n/2 \rfloor$ and define the p-dimensional random vectors $\hat{S}_{jm_n}^{\hat{\sigma}_u,*}(u) = \sum_{r=j}^{j+m_n-1} \hat{Z}_r^{\hat{\sigma}_u}(u)$,

$$\hat{S}_{jm_n'}^{\hat{\sigma}_u}(u) = \hat{S}_{j,\lfloor m_n/2 \rfloor}^{\hat{\sigma}_u,*}(u) - \hat{S}_{j+\lfloor m_n/2 \rfloor,\lfloor m_n/2 \rfloor}^{\hat{\sigma}_u,*}(u)$$

(d) **for** *r=1, ..., B* **do**

- Generate independent standard normal distributed random variables $\{R_i^{(r)}\}_{i=\lceil nu-nb_n \rceil}^{\lfloor nu+nb_n \rfloor}$
- Calculate the bootstrap statistic

$$T^{\hat{\sigma}_u,(r)}(u) = \Big| \sum_{j=\lceil nu-nb_n \rceil - m_n'+1}^{\lfloor nu+nb_n \rfloor} \hat{S}_{jm_n'}^{\hat{\sigma}_u}(u) R_j^{(r)} \Big|_\infty$$

end

(e) Define $T_{\lfloor (1-\alpha)B \rfloor}^{\hat{\sigma}_u}(u)$ as the empirical $(1-\alpha)$-quantile of the sample $T^{\hat{\sigma}_u,(1)}(u), \ldots, T^{\hat{\sigma}_u,(B)}(u)$ and

$$\hat{L}_6^{\hat{\sigma}_u}(u,t) = \hat{m}(u,t) - \hat{r}_6^{\hat{\sigma}_u}(u,t) \quad , \quad \hat{U}_6^{\hat{\sigma}_u}(u,t) = \hat{m}(u,t) + \hat{r}_6^{\hat{\sigma}_u}(u,t),$$

where

$$\hat{r}_6^{\hat{\sigma}_u}(u,t) = \frac{\hat{\sigma}(u,t)\sqrt{2}T_{\lfloor (1-\alpha)B \rfloor}^{\hat{\sigma}_u}(u)}{\sqrt{nb_n}\sqrt{m_n'(\lfloor nu+nb_n \rfloor - \lceil nu-nb_n \rceil - m_n' + 2)}}$$

Output: $\mathcal{C}_n^{\hat{\sigma}_u}(u) = \big\{ f : [0,1]^2 \to \mathbb{R} \mid \hat{L}_6^{\hat{\sigma}_u}(u,t) \le f(u,t) \le \hat{U}_6^{\hat{\sigma}_u}(u,t) \ \ \forall u \in [b_n, 1-b_n] \big\}$.

B Proofs

In the following proofs, for two real sequence a_n and b_n we write $a_n \lesssim b_n$, if there exists a universal positive constant M such that $a_n \le M b_n$. Let $\mathbf{1}(\cdot)$ be the usual indicator function.

B.1 Proof of Theorem 3.1 and 3.2

B.1.1 Gaussian approximations on a finite grid

The proofs of Theorem 3.1 and 3.2 are based on an auxiliary result providing a Gaussian approximation for the maximum deviation of the quantity $\sqrt{nb_n}|\Delta(u,t_v)|$ over the grid of $\{1/n,...,n/n\} \times \{t_1,...,t_p\}$ where $t_v = \frac{v}{p}$ $(v = 1,\dots,p)$.

Proposition B.1. *Assume that $n^{1+a}b_n^9 = o(1)$, $n^{a-1}b_n^{-1} = o(1)$ for some $0 < a < 4/5$, and let Assumptions 2.1, 2.2 and 2.4 be satisfied.*

(i) For a fixed $u \in (0,1)$, let $Y_1(u),\dots,Y_n(u)$ denote a sequence of centred p-dimensional Gaussian vectors such that $Y_i(u)$ has the same auto-covariance structure of the vector $Z_i(u)$ defined in (3.1). If $p = O(\exp(n^\iota))$ for some $0 \le \iota < 1/11$, then

$$\mathfrak{P}_{p,n}(u) := \sup_{x\in\mathbb{R}} \Big|\mathbb{P}\Big(\max_{1\le v\le p} \sqrt{nb_n}|\Delta(u,t_v)| \le x\Big) - \mathbb{P}\Big(\Big|\frac{1}{\sqrt{nb_n}}\sum_{i=1}^n Y_i(u)\Big|_\infty \le x\Big)\Big|$$
$$= O\Big((nb_n)^{-(1-11\iota)/8} + \Theta\Big(\sqrt{nb_n}\Big(b_n^4 + \frac{1}{nb_n}\Big), p\Big)\Big)$$

(ii) Let $\tilde{Y}_1,\dots,\tilde{Y}_{2\lceil nb_n\rceil-1}$ denote independent $(n-2\lceil nb_n\rceil+1)p$-dimensional centred Gaussian vectors with the same auto-covariance structure as the vector $\tilde{Z}_i$ in (3.3). If $np = O(\exp(n^\iota))$ for some $0 \le \iota < 1/11$, then

$$\mathfrak{P}_{p,n} := \sup_{x\in\mathbb{R}} \Big|\mathbb{P}\Big(\max_{\lceil nb_n\rceil \le l \le n-\lceil nb_n\rceil, 1\le v\le p} \sqrt{nb_n}|\Delta(\tfrac{l}{n},t_v)| \le x\Big) - \mathbb{P}\Big(\Big|\frac{1}{\sqrt{nb_n}}\sum_{i=1}^{2\lceil nb_n\rceil-1} \tilde{Y}_i\Big|_\infty \le x\Big)\Big|$$
$$= O\Big((nb_n)^{-(1-11\iota)/8} + \Theta\Big(\sqrt{nb_n}\Big(b_n^4 + \frac{1}{nb_n}\Big), np\Big)\Big)$$

Proof of Proposition B.1. Using Assumptions 2.1, 2.4 and a Taylor expansion we obtain

$$\sup_{u\in[b_n,1-b_n],t\in[0,1]} \Big|\mathbb{E}(\hat{m}(u,t)) - m(u,t) - b_n^2\int K(v)v^2 dv \frac{\partial^2}{\partial u^2}m(u,t)/2\Big| \le M\Big(\frac{1}{nb_n} + b_n^4\Big) \quad \text{(B.1)}$$

for some constant M. Notice that for $u \in [b_n, 1-b_n]$,

$$\hat{m}(u,t) - \mathbb{E}(\hat{m}(u,t)) = \frac{1}{nb_n}\sum_{i=1}^{n} G(\tfrac{i}{n}, t, \mathcal{F}_i) K\Big(\frac{\frac{i}{n}-u}{b_n}\Big) \tag{B.2}$$
$$= \frac{1}{nb_n}\sum_{i=\lceil n(u-b_n)\rceil}^{\lfloor n(u+b_n)\rfloor} G(\tfrac{i}{n}, t, \mathcal{F}_i) K\Big(\frac{\frac{i}{n}-u}{b_n}\Big).$$

Therefore, observing the definition of $Z_i(u)$ in (3.1) we have (notice that $Z_i(u)$ is a vector of zero if $|\frac{i}{n}-u| \geq b_n$)

$$\max_{1\leq v\leq p} \sqrt{nb_n}|\hat{m}(u,t_v) - \mathbb{E}(\hat{m}(u,t_v))| = \Big|\frac{1}{\sqrt{nb_n}}\sum_{i=\lceil n(u-b_n)\rceil}^{\lfloor n(u+b_n)\rfloor} Z_i(u)\Big|_\infty.$$

We will now apply Corollary 2.2 of Zhang and Cheng (2018) and check its assumptions first. By Assumption 2.2(2) and the fact that the kernel is bounded it follows that

$$\max_{1\leq l\leq p}\sup_i \|Z_{i,l}(u) - Z_{i,l}^{(i-j)}(u)\|_2 = O(\chi^j),$$

where for any (measurable function) $g = g(\mathcal{F}_i)$, we define for $j \leq i$ the function $g^{(j)}$ by $g^{(j)} = g(\mathcal{F}_i^{(j)})$, where $\mathcal{F}_i^{(j)} = (\ldots, \eta_{j-1}, \eta_j', \eta_{j+1}, \ldots, \eta_i)$ and $\{\eta_i'\}_{i\in\mathbb{Z}}$ is an independent copy of $\{\eta_i\}_{i\in\mathbb{Z}}$ (recall that $\mathcal{F}_i = (\eta_{-\infty}, ..., \eta_i)$). Lemma C.3 in Section C shows that condition (9) in the paper of Zhang and Cheng (2018) is satisfied. Moreover Assumption 2.2(1) implies condition (13) in this reference. Observing that for random vector $v = (v_1, ..., v_p)^\top$ and all $x \in \mathbb{R}$

$$\{|v|_\infty \leq x\} = \Big\{\max_{1\leq i\leq p}(v_1, ..., v_p, -v_1, ..., -v_p) \leq x\Big\}$$

we can use Corollary 2 of Zhang and Cheng (2018) and obtain

$$\sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big(\Big|\frac{1}{\sqrt{nb_n}}\sum_{i=1}^{n} Y_i(u)\Big|_\infty \leq x\Big) - \mathbb{P}\Big(\frac{1}{\sqrt{nb_n}}\Big|\sum_{i=1}^{n} Z_i(u)\Big|_\infty \leq x\Big)\Big| = O((nb_n)^{-(1-11\iota)/8}). \tag{B.3}$$

Therefore part (i) of the assertion follows from (B.1), (B.3) and Lemma C.1 in Appendix C.

For part (ii), note that by the definition of the vector $\tilde{Z}_i$ in (3.3) we have that

$$\max_{1\leq v\leq p}\max_{\lceil nb_n\rceil\leq l\leq n-\lceil nb_n\rceil}|W_n(\tfrac{l}{n}, t_v)| = \max_{\lceil nb_n\rceil\leq l\leq n-\lceil nb_n\rceil}\Big|\frac{1}{\sqrt{nb_n}}\sum_{i=1}^{n} Z_i(\frac{l}{n})\Big|_\infty = \Big|\frac{1}{\sqrt{nb_n}}\sum_{i=1}^{2\lceil nb_n\rceil-1}\tilde{Z}_i\Big|_\infty,$$

where we use the notation

$$W_n(u,t) = \sqrt{nb_n}\big(\hat{m}(u,t) - \mathbb{E}(\hat{m}(u,t))\big) = \frac{1}{\sqrt{nb_n}} \sum_{i=1}^{n} G(\tfrac{i}{n}, t, \mathcal{F}_i) K\Big(\frac{\frac{i}{n} - u}{b_n}\Big) \tag{B.4}$$

for the sake of brevity. Let $\tilde{Z}_{i,s}$ denote the sth entry of the vector $\tilde{Z}_i$ defined in (3.3) ($1 \leq s \leq (n - 2\lceil nb_n\rceil + 1)p$). By Assumption 2.2(2) it follows that

$$\max_{1\leq s\leq (n-2\lceil nb_n\rceil+1)p} \sup_i \|\tilde{Z}_{i,s} - \tilde{Z}_{i,s}^{(i-j)}\|_2 = O(\chi^j).$$

By Lemma C.3 in Section C we obtain the inequality

$$c_1 \leq \min_{1\leq j\leq (n-2\lceil nb_n\rceil+1)p} \tilde{\sigma}_{j,j} \leq \max_{1\leq j\leq (n-2\lceil nb_n\rceil+1)p} \tilde{\sigma}_{j,j} \leq c_2$$

for the quantities

$$\tilde{\sigma}_{j,j} := \frac{1}{2\lceil nb_n\rceil - 1} \sum_{i,l=1}^{2\lceil nb_n\rceil -1} \mathrm{Cov}(\tilde{Z}_{i,j}, \tilde{Z}_{l,j}).$$

Therefore condition (9) in the paper of Zhang and Cheng (2018) holds, and condition (13) in this reference follows from Assumption 2.2(1). As a consequence, Corollary 2.2 in Zhang and Cheng (2018) yields

$$\sup_{x\in\mathbb{R}} \Big|\mathbb{P}\Big(\max_{\substack{\lceil nb_n\rceil\leq l_1\leq n-\lceil nb_n\rceil \\ 1\leq l_2\leq p}} |W_n(\tfrac{l_1}{n}, \tfrac{l_2}{p})| \leq x\Big) - \mathbb{P}\Big(\Big|\frac{1}{\sqrt{nb_n}} \sum_{i=1}^{2\lceil nb_n\rceil -1} \tilde{Y}_i\Big|_\infty \leq x\Big)\Big| = O((nb_n)^{-(1-11\iota)/8}). \tag{B.5}$$

Consequently part (ii) follows by the same arguments given in the proof of part (i) via an application of Lemma C.1 in Section C. $\diamondsuit$

B.1.2 Proof of Theorem 3.1

The proof of the first assertion is similar (but simpler) than the proof of Theorem 3.2. Therefore details are omitted. The second assertion follows from the bandwidth condition such that $\sqrt{nb_n}(b_n^4 + \frac{1}{nb_n}) = o(n^{-a/2})$.

B.1.3 Proof of Theorem 3.2

For $p \in \mathbb{N}$ define by $t_v = \frac{v}{p}$, ($v = 0, \ldots, p$) an equidistant partition of the interval $[0, 1]$ and let M be a sufficiently large generic constant which may vary from line to line. Recalling the notation of

$W_n(u,t)$ in (B.4) we have by triangle inequality

$$\Big|\sup_{b_n\le u\le 1-b_n, 0\le t\le 1}|W_n(u,t)|-\max_{\substack{\lceil nb_n\rceil\le l_1\le n-\lceil nb_n\rceil\\ 1\le s\le p}}|W_n(\tfrac{l_1}{n},\tfrac{s}{p})|\Big|\le \tilde{W}_n\ ,$$

where

$$\tilde{W}_n\ =\max_{\substack{\lceil nb_n\rceil\le l_1\le n-\lceil nb_n\rceil, 1\le s\le p,\\ |u-\frac{l_1}{n}|\le 1/n, |t-\frac{s}{p}|\le 1/p, u,t\in[0,1]}}|W_n(u,t)-W_n(\tfrac{l_1}{n},\tfrac{s}{p})|.$$

By Assumption 2.3, Burkholder's inequality and similar arguments as given in the proof of Proposition 1.1 of Dette and Wu (2020) we obtain

$$\begin{aligned}&\sup_{u,t\in[0,1]}\Big\|\frac{\partial}{\partial u}W_n(u,t)\Big\|_{q^*}\le\frac{M}{b_n},\ \sup_{u,t\in[0,1]}\Big\|\frac{\partial}{\partial t}W_n(u,t)\Big\|_{q^*}\le M,\\ &\sup_{u,t\in[0,1]}\Big\|\frac{\partial^2}{\partial u\partial t}W_n(u,t)\Big\|_{q^*}\le\frac{M}{b_n}.\end{aligned}\tag{B.6}$$

Note that we have for $\tau_s>0$, $s=1,2$ and $x,y\in[0,1)$,

$$\begin{aligned}&\Big\|\sup_{\substack{0\le t_1\le\tau_1\\0\le t_2\le\tau_2}}|W_n(t_1+x,t_2+y)-W_n(x,y)|\Big\|_{q^*}\\ &\le\int_0^{\tau_1}\Big\|\frac{\partial}{\partial u}W_n(x+u,y)\Big\|_{q^*}du+\int_0^{\tau_2}\Big\|\frac{\partial}{\partial t}W_n(x,y+v)\Big\|_{q^*}dv+\int_0^{\tau_1}\int_0^{\tau_2}\Big\|\frac{\partial^2}{\partial x\partial t}W_n(x+u,y+v)\Big\|_{q^*}dudv.\end{aligned}$$

Therefore, (B.6) and similar arguments as in the proof of Proposition B.2 of Dette et al. (2019) show

$$\|\tilde{W}_n\|_{q^*}=O((np)^{1/q^*}((nb_n)^{-1}+1/p)).\tag{B.7}$$

Observing (B.5) in the proof of Proposition B.1, Lemma C.1 in Section C and (B.7) it therefore follows that

$$\begin{aligned}\mathfrak{P}_n&\lesssim(nb_n)^{-(1-11\iota)/8}+\Theta\Big(\sqrt{nb_n}\Big(b_n^4+\frac{1}{nb_n}\Big),np\Big)+\Theta(\delta,np)+\mathbb{P}(\tilde{W}_n>\delta)\\ &\lesssim(nb_n)^{-(1-11\iota)/8}+\Theta\Big(\sqrt{nb_n}\Big(b_n^4+\frac{1}{nb_n}\Big),np\Big)+\Theta(\delta,np)\\ &\quad+\Big((np)^{1/q^*}((nb_n)^{-1}+1/p)/\delta\Big)^{q^*}.\end{aligned}$$

Solving $\delta = \big((np)^{1/q^*}((nb_n)^{-1}+1/p)/\delta\big)^{q^*}$ we get $\delta = \big((np)^{1/q^*}((nb_n)^{-1}+1/p)\big)^{\frac{q^*}{q^*+1}}$ and the assertion of the theorem follows.

B.2 Proof of Theorem 3.3

We only show part (ii), part (i) follows by similar arguments. Let T_k denote the statistic generated by (3.15) in one bootstrap iteration of Algorithm 2 and define for integers a, b the quantities

$$
\begin{aligned}
T^\diamond_{ap+b} &= \sum_{j=1}^{2\lceil nb_n\rceil - m_n'} \hat{S}_{jm_n',(a-1)p+b} V_{k,j}, a = 1, ...n-2\lceil nb_n\rceil+1, 1\leq b\leq p \\
T^\diamond &:= ((T_1^\diamond)^\top, \dots, (T^\diamond_{(n-2\lceil nb_n\rceil+1)p})^\top)^\top = \big(T_1{}^\top, \dots, T_{n-2\lceil nb_n\rceil+1}{}^\top\big)^\top \\
T &= |T^\diamond|_\infty = \max_{1\leq k\leq n-2\lceil nb_n\rceil+1} |T_k|_\infty
\end{aligned}
$$

It suffices to show that the following inequality holds

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}} \Big|\mathbb{P}(|T^\diamond/\sqrt{m_n'(2\lceil nb_n\rceil - m_n')}|_\infty \leq x|\mathcal{F}_n) - \mathbb{P}\Big(\frac{1}{\sqrt{2nb_n}}\Big|\sum_{i=1}^{2\lceil nb_n\rceil-1} \tilde{Y}_i\Big|_\infty \leq x\Big)\Big| \\
&= O_p\Big(\vartheta_n^{1/3}\{1\vee \log(\frac{np}{\vartheta_n})\}^{2/3} + \Theta\Big(\Big(\sqrt{m_n\log np}\Big(\frac{1}{\sqrt{nd_n}} + d_n^2\Big)(np)^{\frac{1}{q}}\Big)^{q/(q+1)}, np\Big)\Big). \qquad \text{(B.8)}
\end{aligned}
$$

If this estimate has been established, assertion (ii) of Theorem 3.3 follows from Theorem 3.2, which shows that the probabilities $\mathbb{P}\big(\max_{b_n\leq u\leq 1-b_n, 0\leq t\leq 1} \sqrt{nb_n}|\Delta(u,t)| \leq x\big)$ can be approximated by the probabilities $\mathbb{P}\big(\frac{1}{\sqrt{2nb_n}}|\sum_{i=1}^{2\lceil nb_n\rceil-1}\tilde{Y}_i|_\infty \leq x\big)$ (uniformly with respect to $x\in\mathbb{R}$).

For a proof of (B.8) we assume without loss of generality that m_n is even so that $m_n' = m_n$. For convenience, let $\sum_{i=a}^b Z_i = 0$ if the indices a and b satisfy $a > b$. Given the data, it follows for the conditional covariance

$$
\begin{aligned}
&m_n((2\lceil nb_n\rceil - 1) - m_n + 1)\sigma^{T^\diamond}_{(k_1-1)p+j_1,(k_2-1)p+j_2} := \mathbb{E}(T^\diamond_{(k_1-1)p+j_1} T^\diamond_{(k_2-1)p+j_2}|\mathcal{F}_n) \qquad \text{(B.9)} \\
&= \mathbb{E}\Big(\sum_{r=1}^{2\lceil nb_n\rceil - m_n} \hat{S}_{rm_n,(k_1-1)p+j_1} V_{k_1,r} \sum_{r=1}^{2\lceil nb_n\rceil - m_n} \hat{S}_{rm_n,(k_2-1)p+j_2} V_{k_2,r}\Big|\mathcal{F}_n\Big) \\
&= \sum_{r=1}^{2\lceil nb_n\rceil - m_n-(k_2-k_1)} \hat{S}_{(r+k_2-k_1)m_n,(k_1-1)p+j_1}\hat{S}_{rm_n,(k_2-1)p+j_2}.
\end{aligned}
$$

where $1\leq k_1\leq k_2\leq (n-2\lceil nb_n\rceil+1)$, $1\leq j_1, j_2\leq p$, and $\hat{S}_{rm_n,j}$ is the j_{th} entry of the vector $\hat{S}_{rm_n}$. Here, without generality, we assume $k_1\leq k_2$. Define $\tilde{T}^\diamond$, $\tilde{S}^*_{j,m_n}$ and $\tilde{S}_{jm_n}$ in the same way as $T^\diamond$, $\hat{S}^*_{jm_n}$ and $\hat{S}_{jm_n}$ in (3.15) and (3.13), respectively, where the residuals $\hat{\tilde{Z}}_i$ defined in (3.11) and used in step (a) of Algorithm 2 have been replaced by quantities $\tilde{Z}_i$ defined in (3.3).Then we obtain by similar arguments

$$m_n((2\lceil nb_n\rceil-1)-m_n+1)\sigma^{\tilde{T}^\diamond}_{(k_1-1)p+j_1,(k_2-1)p+j_2} := \mathbb{E}(\tilde{T}^\diamond_{(k_1-1)p+j_1}\tilde{T}^\diamond_{(k_2-1)p+j_2}|\mathcal{F}_n)$$
$$= \sum_{r=1}^{\lceil 2nb_n\rceil-m_n-(k_2-k_1)} \tilde{S}_{(r+k_2-k_1)m_n,(k_1-1)p+j_1}\tilde{S}_{rm_n,(k_2-1)p+j_2}. \tag{B.10}$$

Recall the definition of the random variable $\tilde{Y}_j$ in Proposition B.1 and denote by $\tilde{Z}_{j,i}$, $\tilde{Y}_{j,i}$ the ith component of the vectors $\tilde{Z}_j$ and $\tilde{Y}_j$, respectively $(1\le i\le (n-2\lceil nb_n\rceil+1)p$, $1\le j\le 2\lceil nb_n\rceil-1)$. Then we obtain

$$\sigma^{\tilde{Y}}_{(k_1-1)p+j_1,(k_2-1)p+j_2} := \mathbb{E}\Big(\frac{1}{2\lceil nb_n\rceil-1}\sum_{i_1=1}^{2\lceil nb_n\rceil-1}\tilde{Y}_{i_1,(k_1-1)p+j_1}\sum_{i_2=1}^{2\lceil nb_n\rceil-1}\tilde{Y}_{i_2,(k_2-1)p+j_2}\Big)$$
$$= \frac{\mathbb{E}(\sum_{i_1=1}^{2\lceil nb_n\rceil-1}\tilde{Z}_{i_1,(k_1-1)p+j_1}\sum_{i_2=1}^{2\lceil nb_n\rceil-1}\tilde{Z}_{i_2,(k_2-1)p+j_2})}{2\lceil nb_n\rceil-1}$$
$$= \frac{\mathbb{E}(\sum_{i_1=1}^{2\lceil nb_n\rceil-1}Z_{i_1+(k_1-1),\lceil nb_n\rceil+(k_1-1),j_1}\sum_{i_2=1}^{2\lceil nb_n\rceil-1}Z_{i_2+(k_2-1),\lceil nb_n\rceil+(k_2-1),j_2})}{2\lceil nb_n\rceil-1}, \tag{B.11}$$

where $Z_{i_1+(k_1-1),\lceil nb_n\rceil,j_1}$ is the j_1th entry of the $p-$dimensional random vector $Z_{i_1+(k_1-1),\lceil nb_n\rceil,j_1}$ and $Z_{i_2+(k_2-1),\lceil nb_n\rceil,j_2}$ is defined similarly. We will show at the end of this section that

$$\Big\|\max_{k_1,k_2,j_1,j_2}|\sigma^{\tilde{Y}}_{(k_1-1)p+j_1,(k_2-1)p+j_2}-\sigma^{\tilde{T}^\diamond}_{(k_1-1)p+j_1,(k_2-1)p+j_2}\Big\|_{q/2} = O(\vartheta_n). \tag{B.12}$$

If (B.12) holds, it follows from Lemma C.3 in the appendix that there exists a constant $\eta_0>0$ such that

$$\mathbb{P}\Big(\min_{\substack{1\le k\le(n-2\lceil nb_n\rceil+1),\\ 1\le j\le p}}\sigma^{\tilde{T}^\diamond}_{(k-1)p+j,(k-1)p+j}\ge\eta_0\Big)\ge 1-O(\vartheta_n^{q/2}).$$

Then, by Theorem 2 of Chernozhukov et al. (2015), we have

$$\sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big(\frac{|\tilde{T}^\diamond|_\infty}{\sqrt{m_n'(2\lceil nb_n\rceil-m_n)}}\le x\ \Big|\mathcal{F}_n\Big)-\mathbb{P}\Big(\frac{1}{\sqrt{2nb_n}}\Big|\sum_{i=1}^{2\lceil nb_n\rceil-1}\tilde{Y}_i\Big|_\infty\le x\Big)\Big|$$
$$= O_p(\vartheta_n^{1/3}\{1\vee\log(\frac{np}{\vartheta_n})\}^{2/3}). \tag{B.13}$$

Since conditional on $\mathcal{F}_n$, $\big(\tilde{T}^\diamond-T^\diamond\big)$ is a $(n-2\lceil nb_n\rceil+1)p$ dimensional Gaussian random vector we obtain by the (conditional) Jensen inequality and conditional inequality for the concentration of the maximum of a Gaussian process (see Chapter 5 in Appendix A of Chatterjee, 2014, where a similar result has been derived in Lemma A.1) that

$$\mathbb{E}(|\tilde{T}^{\diamond} - T^{\diamond}|_{\infty}^{q}|\mathcal{F}_n) \le M|\sqrt{\log np}\max_{r=1}^{(n-2\lceil nb_n\rceil+1)p}\Big(\sum_{j=1}^{2\lceil nb_n\rceil-m_n'}(\hat{S}_{jm_n',r}-S_{jm_n',r})^2\Big)^{1/2}|^q \tag{B.14}$$

for some large constant M almost surely. Observing that

$$\max_{1\le i\le n}|Z_i|^l \le \sum_{1\le i\le n}|Z_i|^l \qquad \text{for any } l>0, n\in\mathbb{N} \tag{B.15}$$

and using a similar argument as given in the proof of Proposition 1.1 in Dette and Wu (2020) yields

$$\frac{1}{\sqrt{m_n(2\lceil nb_n\rceil - m_n)}}\Big\|\max_{r=1}^{(n-2\lceil nb_n\rceil+1)p}\Big(\sum_{j=1}^{\lceil 2nb_n\rceil-m_n'}(\hat{S}_{jm_n',r}-S_{jm_n',r})^2\Big)^{1/2}\Big\|_q = O\Big(\sqrt{m_n}\Big(\frac{1}{\sqrt{nd_n}}+d_n^2\Big)(np)^{\frac{1}{q}}\Big),$$

(recall that d_n is the bandwidth of the local linear estimator (3.7)) and combining this result with the (conditional version) of Lemma C.1 in Appendix C and (B.14) yields

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big(\frac{|T^{\diamond}|_\infty}{\sqrt{m_n(2\lceil nb_n\rceil-m_n)}}>x\;\Big|\mathcal{F}_n\Big)-\mathbb{P}\Big(\frac{1}{\sqrt{2nb_n}}\Big|\sum_{i=1}^{2\lceil nb_n\rceil-1}\tilde{Y}_i\Big|_\infty>x\Big)\Big|\\
\le&\sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big(\frac{|\tilde{T}^{\diamond}|_\infty}{\sqrt{m_n(2\lceil nb_n\rceil-m_n)}}>x\;\Big|\mathcal{F}_n\Big)-\mathbb{P}\Big(\frac{1}{\sqrt{2nb_n}}\Big|\sum_{i=1}^{2\lceil nb_n\rceil-1}\tilde{Y}_i\Big|_\infty>x\Big)\Big|\\
&\qquad+\mathbb{P}\Big(\frac{|\tilde{T}^{\diamond}-T^{\diamond}|_\infty}{\sqrt{m_n(2\lceil nb_n\rceil-m_n)}}>\delta\Big|\mathcal{F}_n\Big)+O\big(\Theta(\delta,np)\big)\\
\le&\sup_{x\in\mathbb{R}}|\mathbb{P}\Big(\frac{|\tilde{T}^{\diamond}|_\infty}{\sqrt{m_n(2\lceil nb_n\rceil-m_n)}}>x\;\Big|\mathcal{F}_n\Big)-\mathbb{P}\Big(\frac{1}{\sqrt{2nb_n}}\Big|\sum_{i=1}^{2\lceil nb_n\rceil-1}\tilde{Y}_i\Big|_\infty>x\Big)\Big|\\
&\qquad+O_p\big(\delta^{-q}\big(\sqrt{m_n\log np}\big(\frac{1}{\sqrt{nd_n}}+d_n^2\big)(np)^{\frac{1}{q}}\big)^q\big)+O\big(\Theta(\delta,np)\big),
\end{aligned} \tag{B.16}$$

where we have used the Markov's inequality. Taking $\delta=\big(\sqrt{m_n\log np}(\frac{1}{\sqrt{nd_n}}+d_n^2)(np)^{\frac{1}{q}}\big)^{q/(q+1)}$ in (B.16), and combining this estimate with (B.13) yields (B.8) completes the proof.

Proof of (B.12)**.** To simplify the notation, write

$$G_{j,i,k}=G(\tfrac{i+k-1}{n},j/p,\mathcal{F}_{i+k-1}),\quad G_{j,i,k,u}=G(\tfrac{i+k-1+u}{n},j/p,\mathcal{F}_u)$$

Without loss of generality, we consider the case $k_1 \le k_2$. We calculate $\sigma^{\tilde{Y}}_{(k_1-1)p+j_1,(k_2-1)p+j_2}$ observing the representation

$$Z_{i_1+(k_1-1),\lceil nb_n\rceil+(k_1-1),j_1} = G_{j_1,i_1,k_1} K\Big(\tfrac{i_1-\lceil nb_n\rceil}{nb_n}\Big).$$

By Lemma C.2 it follows that

$$\mathbb{E}\big[Z_{i_1+(k_1-1),\lceil nb_n\rceil+(k_1-1),j_1} Z_{i_2+(k_2-1),\lceil nb_n\rceil+(k_2-1),j_2}\big] = O(\chi^{|i_1-i_2+k_1-k_2|}). \tag{B.17}$$

uniformly for $1 \le i_1, i_2 \le 2\lceil nb_n\rceil - 1$, $1 \le j_1, j_2 \le p$, $1 \le k_1, k_2 \le n - 2\lceil nb_n\rceil + 1$. We first show that (B.12) holds whenever $k_2 - k_1 > 2\lceil nb_n\rceil - m_n$. On the one hand, observing and (B.9) and (B.10) that if $2\lceil nb_n\rceil - m_n - (k_2 - k_1) < 0$ then

$$\sigma^{\tilde{T}^\diamond}_{(k_1-1)p+j_1,(k_2-1)p+j_2} = 0 \quad a.s. \tag{B.18}$$

Moreover, by (B.11) and (B.17), straightforward calculations show that

$$\sigma^{\tilde{Y}}_{(k_1-1)p+j_1,(k_2-1)p+j_2} = \frac{1}{2\lceil nb_n\rceil - 1} O\Big(\sum_{i_1=1}^{2\lceil nb_n\rceil-1}\sum_{i_2=1}^{2\lceil nb_n\rceil-1} \chi^{|i_1-i_2+k_1-k_2|}\Big) = O\Big(\frac{m_n}{nb_n}\Big). \tag{B.19}$$

Combining (B.18), (B.19) and by applying similar argument to $k_1 \ge k_2$, we obtain

$$\Big\| \max_{\substack{k_1,k_2,j_1,j_2\\ |k_2-k_1|>2\lceil nb_n\rceil-m_n}} |\sigma^{\tilde{Y}}_{(k_1-1)p+j_1,(k_2-1)p+j_2} - \sigma^{\tilde{T}^\diamond}_{(k_1-1)p+j_1,(k_2-1)p+j_2}\Big\|_{q/2} = O\Big(\frac{m_n}{nb_n}\Big). \tag{B.20}$$

Now consider the case that $k_2 - k_1 \le 2\lceil nb_n\rceil - m_n$. Without generality we consider $k_1 \le k_2$. Again by (B.11)

$$\begin{aligned}
&\mathbb{E}\Big(\sum_{i_1=1}^{k_2-k_1} Z_{i_1+(k_1-1),\lceil nb_n\rceil+(k_1-1),j_1} \sum_{i_2=1}^{2\lceil nb_n\rceil-1} Z_{i_2+(k_2-1),\lceil nb_n\rceil+(k_2-1),j_2}\Big)\\
&= O\Big(\sum_{i_1=1}^{k_2-k_1}\sum_{i_2=1}^{2\lceil nb_n\rceil-1} \chi^{|i_2-i_1+k_2-k_1|}\Big) = O\Big(\sum_{i_1=1}^{k_2-k_1}\sum_{i_2=1}^{2\lceil nb_n\rceil-1} \chi^{i_2-i_1+k_2-k_1}\Big) = O(1),\\
&\mathbb{E}\Big(\sum_{i_1=1}^{2\lceil nb_n\rceil-1} Z_{i_1+(k_1-1),\lceil nb_n\rceil+(k_1-1),j_1} \sum_{i_2=2\lceil nb_n\rceil-(k_2-k_1)}^{2\lceil nb_n\rceil-1} Z_{i_2+(k_2-1),\lceil nb_n\rceil+(k_2-1),j_2}\Big)\\
&= O\Big(\sum_{i_1=1}^{2\lceil nb_n\rceil-1}\sum_{i_2=2\lceil nb_n\rceil-(k_2-k_1)}^{2\lceil nb_n\rceil-1} \chi^{|i_2-i_1+k_2-k_1|}\Big) = O\Big(\sum_{i_1=1}^{2\lceil nb_n\rceil-1}\sum_{i_2=2\lceil nb_n\rceil-(k_2-k_1)}^{2\lceil nb_n\rceil-1} \chi^{i_2-i_1+k_2-k_1}\Big) = O(1).
\end{aligned}$$

Let $a = \lfloor M\log n\rfloor$ for a sufficiently large constant M. Using (B.11), it follows (considering the lags up to a) that

$$\begin{aligned}
&\sigma^{\tilde{Y}}_{(k_1-1)p+j_1,(k_2-1)p+j_2}\\
&= \tfrac{1}{2\lceil nb_n\rceil-1}\mathbb{E}\Big(\sum_{i_1=k_2-k_1+1}^{2\lceil nb_n\rceil-1} Z_{i_1+(k_1-1),\lceil nb_n\rceil+(k_1-1),j_1}\sum_{i_2=1}^{2\lceil nb_n\rceil-(k_2-k_1)-1} Z_{i_2+(k_2-1),\lceil nb_n\rceil+(k_2-1),j_2}\Big)\\
&+O((nb_n)^{-1})\\
&= \tfrac{1}{2\lceil nb_n\rceil-1}\mathbb{E}\Big(\sum_{i_1,i_2=1}^{2\lceil nb_n\rceil-(k_2-k_1)-1} G_{j_1,i_1,k_2}K(\tfrac{i_1+k_2-k_1-\lceil nb_n\rceil}{nb_n})G_{j_2,i_2,k_2}K(\tfrac{i_2-\lceil nb_n\rceil}{nb_n})\Big)+O((nb_n)^{-1})\\
&= A+B+O(nb_n\chi^a+(nb_n)^{-1}), \qquad \text{(B.21)}
\end{aligned}$$

where the terms A and B are defined by

$$\begin{aligned}
A &:= \tfrac{1}{(2\lceil nb_n\rceil-1)}\sum_{i=1}^{2\lceil nb_n\rceil-(k_2-k_1)-1} A_i, \qquad \text{(B.22)}\\
A_i &= \mathbb{E}(G_{j_1,i,k_2,0}G_{j_2,i,k_2,0})K(\tfrac{i+k_2-k_1-\lceil nb_n\rceil}{nb_n})K(\tfrac{i-\lceil nb_n\rceil}{nb_n})\\
B &= \tfrac{1}{(2\lceil nb_n\rceil-1)}\sum_{u=1}^{a}(B_{1,u}+B_{2,u}),\\
B_{1,u} &= \sum_{i=1}^{2\lceil nb_n\rceil-(k_2-k_1)-1-u} B_{1,u,i}, \qquad \text{(B.23)}\\
B_{2,u} &=: \sum_{i=1}^{2\lceil nb_n\rceil-(k_2-k_1)-1-u} B_{2,u,i}. \qquad \text{(B.24)}
\end{aligned}$$

and

$$\begin{aligned}
B_{1,u,i} &= \mathbb{E}(G_{j_1,i,k_2,u}G_{j_1,i,k_2,0})K(\tfrac{i+u+k_2-k_1-\lceil nb_n\rceil}{nb_n})K(\tfrac{i-\lceil nb_n\rceil}{nb_n}))\\
B_{2,u,i} &= \mathbb{E}(G_{j_1,i,k_2,0}G_{j_2,i,k_2,u})K(\tfrac{i+k_2-k_1-\lceil nb_n\rceil}{nb_n})K(\tfrac{i+u-\lceil nb_n\rceil}{nb_n})
\end{aligned}$$

Therefore, by (B.21), we have that

$$\sigma^{\tilde{Y}}_{(k_1-1)p+j_1,(k_2-1)p+j_2} = \frac{1}{2\lceil nb_n\rceil-1}\Big(\sum_{i=1}^{2\lceil nb_n\rceil-1-(k_2-k_1)} A_i+\sum_{u=1}^{a}\sum_{i=1}^{2\lceil nb_n\rceil-1-(k_2-k_1)-u}(B_{1,u,i}+B_{2,u,i})\Big) +O(nb_n\chi^a+(nb_n)^{-1}). \qquad \text{(B.25)}$$

Now for the term in (B.10) we have

$$\tilde{S}_{(r+k_2-k_1)m_n,(k_1-1)p+j_1}\tilde{S}_{rm_n,(k_2-1)p+j_2} = \Big(\sum_{i=r+k_2-k_1}^{r+k_2-k_1+m_n/2-1} - \sum_{i=r+k_2-k_1+m_n/2}^{r+k_2-k_1+m_n}\Big) Z_{i+k_1-1,\lceil nb_n\rceil+k_1-1,j_1}$$
$$\times\Big(\sum_{i=r}^{r+m_n/2-1} - \sum_{i=r+m_n/2}^{r+m_n}\Big) Z_{i+k_2-1,\lceil nb_n\rceil+k_2-1,j_2}$$
$$=\Big(\sum_{i=r}^{r+m_n/2-1} - \sum_{i=r+m_n/2}^{r+m_n}\Big) G_{j_1,i,k_2} K(\tfrac{i+k_2-k_1-\lceil nb_n\rceil}{nb_n})$$
$$\times\Big(\sum_{i=r}^{r+m_n/2-1} - \sum_{i=r+m_n/2}^{r+m_n}\Big) G_{j_2,i,k_2} K(\tfrac{i-\lceil nb_n\rceil}{nb_n}).$$

By Lemma C.2, it follows that uniformly for $|k_2-k_1| \le 2\lceil nb_n\rceil - m_n$ and $1 \le r \le \lceil 2nb_n\rceil - m_n - (k_2-k_1)$,

$$\begin{aligned}
&\mathbb{E}\tilde{S}_{(r+k_2-k_1)m_n,(k_1-1)p+j_1}\tilde{S}_{rm_n,(k_2-1)p+j_2}\\
&=\sum_{i=r}^{r+m_n}\mathbb{E}(G_{j_1,i,k_2}G_{j_2,i,k_2})K(\tfrac{i+k_2-k_1-\lceil nb_n\rceil}{nb_n})K(\tfrac{i-\lceil nb_n\rceil}{nb_n})\\
&+\sum_{u=1}^{a}\Big(\sum_{i=r}^{r+m_n-u}\big(\mathbb{E}(G_{j_1,i,(k_2+u)}G_{j_2,i,k_2})K(\tfrac{i+u+k_2-k_1-\lceil nb_n\rceil}{nb_n})K(\tfrac{i-\lceil nb_n\rceil}{nb_n})\\
&+\mathbb{E}(G_{j_2,i,(k_2+u)}G_{j_1,i,k_2})K(\tfrac{i+k_2-k_1-\lceil nb_n\rceil}{nb_n})K(\tfrac{i+u-\lceil nb_n\rceil}{nb_n})\big)\Big) + O(m_n\chi^a + a^2),
\end{aligned} \tag{B.26}$$

where the the term $m_n\chi^a$ corresponds to the error of omitting terms in the sum with a large index a, and the term a^2 summarizes the error due to ignoring different signs in the product $\tilde{S}_{(r+k_2-k_1)m_n,(k_1-1)p+j_1}\tilde{S}_{rm_n,(k_2-1)p+j_2}$ (for each index u, we omit $2u$). Furthermore, by Assumption 2.4 and 2.2(3) it follows that uniformaly for $|u| \le a$

$$\frac{1}{m_n}\sum_{i=r}^{r+m_n}\mathbb{E}(G_{j_1,i,k_2}G_{j_2,i,k_2})K(\tfrac{i+k_2-k_1-\lceil nb_n\rceil}{nb_n})K(\tfrac{i-\lceil nb_n\rceil}{nb_n}) = A_r + O(\frac{m_n}{nb_n}), \tag{B.27}$$
$$\frac{1}{m_n}\sum_{i=r}^{r+m_n-u}\mathbb{E}(G_{j_1,i,(k_2+u)}G_{j_2,i,k_2})K(\tfrac{i+u+k_2-k_1-\lceil nb_n\rceil}{nb_n})K(\tfrac{i-\lceil nb_n\rceil}{nb_n}) = B_{1,u,r} + O(\frac{m_n}{nb_n}+\frac{a}{m_n}), \tag{B.28}$$
$$\frac{1}{m_n}\sum_{i=r}^{r+m_n-u}\mathbb{E}(G_{j_2,i,(k_2+u)}G_{j_1,i,k_2})K(\tfrac{i+k_2-k_1-\lceil nb_n\rceil}{nb_n})K(\tfrac{i+u-\lceil nb_n\rceil}{nb_n}) = B_{2,u,r} + O(\frac{m_n}{nb_n}+\frac{a}{m_n})\big), \tag{B.29}$$

where terms $A_r, B_{1,u,r}$ and $B_{2,u,r}$ are defined in equations (B.22), (B.23) and (B.24), respectively.

Notice that (B.10) and expressions (B.26), (B.27), (B.28) and (B.29) yield that

$$\mathbb{E}\sigma^{\tilde{T}^\diamond}_{(k_1-1)p+j_1,(k_2-1)p+j_2} = \frac{1}{2\lceil nb_n\rceil - m_n}\Big\{\sum_{r=1}^{2\lceil nb_n\rceil - m_n-(k_2-k_1)}(A_r + O(\frac{m_n}{nb_n}))$$
$$+\sum_{u=1}^{a}\sum_{r=1}^{2\lceil nb_n\rceil - m_n-(k_2-k_1)}(B_{1,u,r}+B_{2,u,r}+O(\frac{m_n}{nb_n}+\frac{a}{m_n}))\Big\} + O\big(\chi^a+\frac{a^2}{m_n}\big). \tag{B.30}$$

Lemma C.2 implies

$$\max_{\substack{1\le r\le 2\lceil nb_n\rceil-(k_2-k_1)-1,\\ 1\le k_1\le k_2\le (n-2\lceil nb_n\rceil+1),s=1,2}} B_{s,u,r} = O(\chi^u),$$

which yields in combination with equations (B.25), (B.30) and a similar argument applied to the case that $k_1 \ge k_2$,

$$\max_{\substack{1\le k_1,k_2\le (n-2\lceil nb_n\rceil+1)\\ |k_2-k_1|\le 2\lceil nb_n\rceil - m_n, 1\le j_1,j_2\le p}} \Big|\mathbb{E}\sigma^{\tilde{T}^\diamond}_{(k_1-1)p+j_1,(k_2-1)p+j_2} - \sigma^{\tilde{Y}}_{(k_1-1)p+j_1,(k_2-1)p+j_2}\Big| = O\Big(\frac{\log^2 n}{m_n}+\frac{m_n}{nb_n}\Big). \tag{B.31}$$

Furthermore, using (B.15), the Cauchy-Schwartz inequality, a similar argument as given in the proof of Lemma 1 of Zhou (2013) and Assumption 2.2(2) yield that

$$\Big\|\max_{\substack{1\le k_1\le k_2\le (n-2\lceil nb_n\rceil+1),\\ 1\le j_1,j_2\le p}} |\mathbb{E}\sigma^{\tilde{T}^\diamond}_{(k_1-1)p+j_1,(k_2-1)p+j_2} - \sigma^{\tilde{T}^\diamond}_{(k_1-1)p+j_1,(k_2-1)p+j_2}|\Big\|_{q/2} = O\Big(\sqrt{\frac{m_n}{nb_n}}(np)^{4/q}\Big). \tag{B.32}$$

Combining (B.31) and (B.32), we obtain

$$\Big\|\max_{\substack{k_1,k_2,j_1,j_2\\ |k_2-k_1|\le 2\lceil nb_n\rceil - m_n}} |\sigma^{\tilde{Y}}_{(k_1-1)p+j_1,(k_2-1)p+j_2} - \sigma^{\tilde{T}^\diamond}_{(k_1-1)p+j_1,(k_2-1)p+j_2}|\Big\|_{q/2} = O\Big(\frac{\log^2 n}{m_n}+\frac{m_n}{nb_n}+\sqrt{\frac{m_n}{nb_n}}(np)^{4/q}\Big). \tag{B.33}$$

Therefore the estimate (B.12) follows combining (B.20) and (B.33).

B.3 Proof of the results in Section 4

B.3.1 Proof of Theorem 4.1

Similarly to (B.1) and (B.2) in the proof of Theorem B.1 we obtain

$$\sup_{\substack{u\in[b_n,1-b_n]\\ t\in[0,1]}} \frac{1}{\sigma(u,t)}\Big|\mathbb{E}(\hat{m}(u,t)) - m(u,t)\Big| \leq M\Big(\frac{1}{nb_n} + b_n^4\Big) \tag{B.34}$$

for some constant M, where we have used the fact that, by Assumption 2.4, $\int K(v)v^2dv = 0$. Moreover, by a similar but simpler argument as given in the proof of equation (B.7) in Lemma B.3 of Dette et al. (2019) we have for the quantity

$$\frac{\hat{m}(u,t) - \mathbb{E}(\hat{m}(u,t))}{\sigma(u,t)} = \Psi^\sigma(u,t) := \frac{1}{nb_n}\sum_{i=1}^n \frac{G(\frac{i}{n},t,\mathcal{F}_i)}{\sigma(u,t)} K\Big(\frac{\frac{i}{n}-u}{b_n}\Big)$$

the estimate

$$\Big\|\sup_{u\in[b_n,1-b_n],t\in(0,1)} \sqrt{nb_n}|\Phi^\sigma(u,t) - \Psi^\sigma(u,t)|\Big\|_q = O(b_n^{1-2/q}), \tag{B.35}$$

where

$$\Phi^\sigma(u,t) = \frac{1}{nb_n}\sum_{i=1}^n \frac{G(\frac{i}{n},t,\mathcal{F}_i)}{\sigma(\frac{i}{n},t)} K(\frac{\frac{i}{n}-u}{b_n}).$$

Following the proof of Theorem 3.2 we find that

$$\sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big(\max_{b_n\leq u\leq 1-b_n,0\leq t\leq 1}\sqrt{nb_n}\big|\Phi^\sigma(u,t)\big| \leq x\Big) - \mathbb{P}\Big(\Big|\frac{1}{\sqrt{nb_n}}\sum_{i=1}^{2\lceil nb_n\rceil-1}\tilde{Y}_i^\sigma\Big|_\infty \leq x\Big)\Big|$$
$$= O\Big((nb_n)^{-(1-11\iota)/8} + \Theta\Big(\big((np)^{1/q^*}((nb_n)^{-1}+1/p)\big)^{\frac{q^*}{q^*+1}}, np\Big)\Big).$$

Combining this result with Lemma C.1 (with $X = \max_{b_n \le u \le 1-b_n, 0\le t\le 1} \sqrt{nb_n}\big|\Phi^\sigma(u,t)\big|$, $Y = \frac{1}{\sqrt{nb_n}}\sum_{i=1}^{2\lceil nb_n\rceil -1}\tilde{Y}_i^\sigma$, $X' = \max_{b_n \le u \le 1-b_n, 0\le t\le 1} \sqrt{nb_n}\big|\Psi^\sigma(u,t)\big|$ with a sufficiently large constant M) and (B.35) gives

$$\begin{aligned}
\sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big(\max_{b_n\le u\le 1-b_n, 0\le t\le 1}\sqrt{nb_n}\big|\Psi^\sigma(u,t)\big| \le x\Big) - \mathbb{P}\Big(\Big|\frac{1}{\sqrt{nb_n}}\sum_{i=1}^{2\lceil nb_n\rceil-1}\tilde{Y}_i^\sigma\Big|_\infty \le x\Big)\Big| \\
= O\Big((nb_n)^{-(1-11\iota)/8} + \Theta\big(\big((np)^{1/q^*}((nb_n)^{-1}+1/p)\big)^{\frac{q^*}{q^*+1}}, np\big) \\
+\mathbb{P}\Big(\sup_{u\in[b_n,1-b_n],t\in(0,1)}\sqrt{nb_n}|\Phi^\sigma(u,t)-\Psi^\sigma(u,t)|>\delta\Big)+\Theta(\delta,np)\Big) \\
= O\Big((nb_n)^{-(1-11\iota)/8} + \Theta\big(\big((np)^{1/q^*}((nb_n)^{-1}+1/p)\big)^{\frac{q^*}{q^*+1}}, np\big)+\Theta(\delta,np)+\frac{b_n^{q-2}}{\delta^q}\Big). \qquad \text{(B.36)}
\end{aligned}$$

Taking $\delta = b_n^{\frac{q-2}{q+1}}$ we obtain for the last two terms in in (B.36)

$$\Theta(\delta,np)+\frac{b_n^{q-2}}{\delta^q} = O\Big(\Theta(b_n^{\frac{q-2}{q+1}},np)\Big).$$

On the other hand, (B.34), (B.36) and Lemma C.1 (with $X = \max_{b_n \le u \le 1-b_n, 0\le t\le 1} \sqrt{nb_n}\big|\Psi^\sigma(u,t)\big|$, $Y = \frac{1}{\sqrt{nb_n}}\sum_{i=1}^{2\lceil nb_n\rceil -1}\tilde{Y}_i^\sigma$, $X' = \max_{b_n \le u \le 1-b_n, 0\le t\le 1} \sqrt{nb_n}\big|\Delta^\sigma(u,t)\big|$ and $\delta = M\sqrt{nb_n}(\frac{1}{nb_n}+b_n^4)$ with a sufficiently large constant M) yield

$$\begin{aligned}
\sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big(\max_{b_n\le u\le 1-b_n, 0\le t\le 1}\sqrt{nb_n}\big|\Delta^\sigma(u,t)\big| \le x\Big) - \mathbb{P}\Big(\Big|\frac{1}{\sqrt{nb_n}}\sum_{i=1}^{2\lceil nb_n\rceil-1}\tilde{Y}_i^\sigma\Big|_\infty \le x)\Big| \\
= O\Big((nb_n)^{-(1-11\iota)/8} + \Theta\big(\big((np)^{1/q^*}((nb_n)^{-1}+1/p)\big)^{\frac{q^*}{q^*+1}}, np\big)\Big) \\
+\Theta\Big(\sqrt{nb_n}(b_n^4+\frac{1}{nb_n}),np\Big)+\Theta(b_n^{\frac{q-2}{q+1}},np)\Big).
\end{aligned}$$

B.3.2 Proof of Proposition 4.1

Define $\tilde{S}_{k,r}(t) = \sum_{i=k}^r G(i,t,\mathcal{F}_i)$, and define for $u\in[w/n, 1-w/n]$

$$\tilde{\Delta}_j(t) = \frac{\tilde{S}_{j-w+1,j}(t)-\tilde{S}_{j+1,j+w}(t)}{w}\ , \quad \tilde{\sigma}^2(u,t) = \sum_{j=1}^n \frac{w\tilde{\Delta}_j^2(t)}{2}\bar{\omega}(u,j)$$

as the analogs of $\Delta_j(t)$ defined in the main article and the quantities in (4.3), respectively. We also use the convention $\tilde{\sigma}^2(u,t) = \tilde{\sigma}^2(w/n,t)$ and $\tilde{\sigma}^2(u,t) = \tilde{\sigma}^2(1-w/n,t)$ if $u\in[0,w/n)$ and

$u \in (1 - w/n, 1]$, respectively. Assumption 2.1 and the mean value theorem yield

$$\max_{w \le j \le n-w} \sup_{0<t<1} |\tilde{\Delta}_j(t) - \Delta_j(t)| = \max_{w \le j \le n-w} \sup_{0<t<1} \Big| \sum_{r=j-w+1}^{j} m(r/n,t) - \sum_{r=j+1}^{j+w} m(r/n,t) \Big| = O(w/n). \tag{B.37}$$

On the other hand, Assumption 2.2 and Assumption 2.3 and similar arguments as given in the proof of Lemma 3 of Zhou and Wu (2010) give

$$\max_j \|\tilde{\Delta}_j(t)\|_{q'} = O(\sqrt{w}), \;\; \max_j \Big\| \frac{\partial}{\partial t} \tilde{\Delta}_j(t) \Big\|_{q'} = O(\sqrt{w}). \tag{B.38}$$

Moreover, Proposition B.1. of Dette et al. (2019) yields

$$\max_j \Big\| \sup_t |\tilde{\Delta}_j(t)| \Big\|_{q'} = O(\sqrt{w}). \tag{B.39}$$

Now we introduce the notation $C_j(t) = \tilde{\Delta}_j(t) - \Delta_j(t)$ (note that this quantity is not random) and obtain by (B.37) the representation

$$\begin{aligned} \tilde{\sigma}^2(u,t) - \hat{\sigma}^2(u,t) &= \sum_{j=1}^{n} \frac{w(2\tilde{\Delta}_j(t) - C_j(t))C_j(t)}{2} \bar{w}(u,j) \\ &= \sum_{j=1}^{n} w\tilde{\Delta}_j(t) C_j(t) \bar{\omega}(u,j) + O(w^3/n^2) \end{aligned} \tag{B.40}$$

uniformly with respect to $u, t \in (0,1)^2$. Furthermore, by (B.37) we have

$$\sup_{t \in (0,1)} \Big| \sum_{j=1}^{n} w\tilde{\Delta}_j(t) C_j(t) \bar{\omega}(u,j) \Big| \le W^{\diamond}(u) := M(w/n) \sum_{j=1}^{n} w \sup_{t \in (0,1)} |\Delta_j(t)| \bar{\omega}(u,j) \;, \tag{B.41}$$

where M is a sufficiently large constant. Notice that $W^{\diamond}(u)$ is differentiable with respect to the variable u. Therefore it follows from the triangle inequality, (B.39) and Proposition B.1 of Dette et al. (2019), that

$$\Big\| \sup_{u \in [\gamma_n, 1-\gamma_n]} |W^{\diamond}(u)| \Big\|_{q'} = O\Big(\frac{w^{5/2}}{n} \tau_n^{-1/q'} \Big). \tag{B.42}$$

Combining (B.40) and (B.42), we obtain

$$\Big\| \sup_{\substack{u\in[\gamma_n,1-\gamma_n]\\ t\in(0,1)}} |\tilde\sigma^2(u,t)-\hat\sigma^2(u,t)| \Big\|_{q'} = O\Big(\frac{w^{5/2}}{n}\tau_n^{-1/q'}+w^3/n^2\Big). \tag{B.43}$$

By Burkholder inequality (see for example Wu, 2005) in $\mathcal{L}^{q'/2}$ norm, (B.38) and similar arguments as given in the proof of Lemma 3 in Zhou and Wu (2010) we have

$$\begin{aligned}
&\sup_{\substack{u\in[\gamma_n,1-\gamma_n]\\ t\in(0,1)}} \big\|\tilde\sigma^2(u,t)-\mathbb{E}(\tilde\sigma^2(u,t))\big\|_{q'/2} = O\big(w^{1/2}n^{-1/2}\tau_n^{-1/2}\big),\\
&\sup_{\substack{u\in[\gamma_n,1-\gamma_n]\\ t\in(0,1)}} \Big\|\frac{\partial}{\partial t}(\tilde\sigma^2(u,t)-\mathbb{E}(\tilde\sigma^2(u,t)))\Big\|_{q'/2} = O\big(w^{1/2}n^{-1/2}\tau_n^{-1/2}\big),\\
&\sup_{\substack{u\in[\gamma_n,1-\gamma_n]\\ t\in(0,1)}} \Big\|\big(\frac{\partial}{\partial u}+\frac{\partial^2}{\partial u\partial t}\big(\tilde\sigma^2(u,t)-\mathbb{E}(\tilde\sigma^2(u,t))\big)\Big\|_{q'/2} = O\big(w^{1/2}n^{-1/2}\tau_n^{-1/2-1}\big).
\end{aligned} \tag{B.44}$$

It can be shown by similar but simpler argument as given in the proof of Proposition B.2 of Dette et al. (2019) that these estimates imply

$$\sup_{\substack{u\in[\gamma_n,1-\gamma_n]\\ t\in(0,1)}} \big\|\tilde\sigma^2(u,t)-\mathbb{E}(\tilde\sigma^2(u,t))\big\|_{q'/2} = O\big(w^{1/2}n^{-1/2}\tau_n^{-1/2-4/q'}\big). \tag{B.45}$$

Moreover, it follows from the proof of Theorem 4.4 of Dette and Wu (2019) that

$$\begin{aligned}
&\sup_{\substack{u\in[\gamma_n,1-\gamma_n]\\ t\in(0,1)}} \Big|\mathbb{E}\tilde\sigma^2(u,t)-\sigma^2(u,t)\Big| = O\Big(\sqrt{w/n}+w^{-1}+\tau_n^2\Big),\\
&\sup_{\substack{u\in[0,\gamma_n)\cup(1-\gamma_n,1]\\ t\in(0,1)}} \Big|\mathbb{E}\tilde\sigma^2(u,t)-\sigma^2(u,t)\Big| = O\Big(\sqrt{w/n}+w^{-1}+\tau_n\Big)
\end{aligned} \tag{B.46}$$

and the assertion is a consequence of (B.43), (B.45) and (B.46).

B.3.3 Proof of Theorem 4.2

Recall that $g_n = \frac{w^{5/2}}{n}\tau_n^{-1/q'} + w^{1/2}n^{-1/2}\tau_n^{-1/2-4/q'} + w^{-1} + \tau_n$ and let η_n be a sequence of positive numbers such that $\eta_n \to \infty$ and $g_n\eta_n \to 0$ (note that g_n is the convergence rate of the estimator $\hat\sigma^2$ in Proposition 4.1). Let $T_k^{\hat\sigma}$ denote the statistic $T_k^{\hat\sigma,(r)}$ in step (d) of Algorithm 3 generated by one bootstrap iteration and define for integers a, b the quantities

$$T^{\hat{\sigma},\diamond}_{ap+b} = \sum_{j=1}^{2\lceil nb_n \rceil - m_n'} \hat{S}^{\hat{\sigma}}_{jm_n',(a-1)p+b} V_{k,j}, a = 1, ...n - 2\lceil nb_n \rceil + 1, 1 \le b \le p$$
$$T^{\hat{\sigma},\diamond} := ((T^{\hat{\sigma},\diamond}_1)^\top, \dots, (T^{\hat{\sigma},\diamond}_{(n-2\lceil nb_n \rceil+1)p})^\top)^\top = \big(T^{\hat{\sigma}\top}_1, \dots, T^{\hat{\sigma}}_{n-2\lceil nb_n \rceil+1}{}^\top\big)^\top$$

and therefore

$$T^{\hat{\sigma}} = |T^{\hat{\sigma},\diamond}|_\infty = \max_{1 \le k \le n-2\lceil nb_n \rceil+1} |T^{\hat{\sigma}}_k|_\infty$$

We recall the notation (4.2), introduce the $(n - 2\lceil nb_n \rceil + 1)p$-dimensional random vectors $\hat{S}^{\sigma,*}_{jm_n} = \sum_{r=j}^{j+m_n-1} \tilde{Z}^\sigma_r$, and

$$\hat{S}^\sigma_{jm_n'} = \hat{S}^{\sigma,*}_{j,\lfloor m_n/2 \rfloor} - \hat{S}^{\sigma,*}_{j+\lfloor m_n/2 \rfloor+1,\lfloor m_n/2 \rfloor} \ ,$$

and consider

$$T^\sigma_k = \sum_{j=1}^{2\lceil nb_n \rceil - m_n'} \hat{S}^\sigma_{jm_n',[(k-1)p+1:kp]} V_{k,j} \ , \quad k = 1, \dots, n - 2\lceil nb_n \rceil + 1,$$
$$T^{\sigma,\diamond} := ((T^{\sigma,\diamond}_1)^\top, \dots, (T^{\sigma,\diamond}_{(n-2\lceil nb_n \rceil+1)p})^\top)^\top = \big(T^{\sigma\top}_1, \dots, T^\sigma_{n-2\lceil nb_n \rceil+1}{}^\top\big)^\top \ ,$$

where $T^{\sigma,\diamond}$ is obtained from $T^{\hat{\sigma},\diamond}$ by replacing $\hat{\sigma}$ by σ. Similar arguments as given in the proof of Theorem 3.3 show, that it is sufficient to show the estimate

$$\begin{aligned} \sup_{x\in\mathbb{R}} \Big| \mathbb{P}(|T^{\hat{\sigma},\diamond}/\sqrt{m_n'(2\lceil nb_n \rceil - m_n')}|_\infty \le x | \mathcal{F}_n) - \mathbb{P}\Big(\frac{1}{\sqrt{2nb_n}} \Big| \sum_{i=1}^{2\lceil nb_n \rceil - 1} \tilde{Y}^\sigma_i \Big|_\infty \le x\Big) \Big| \\ = O_p\Big(\vartheta_n^{1/3}\{1 \vee \log(\frac{np}{\vartheta_n})\}^{2/3} + \Theta\big(\sqrt{m_n \log np}(\frac{1}{\sqrt{nd_n}} + d_n^2)(np)^{\frac{1}{q}}\big)^{q/(q+1)}, np\big) \\ + \Theta\big(\big(\sqrt{m_n \log np}(g_n \eta_n)(np)^{\frac{1}{q}}\big)^{q/(q+1)}, np\big) + \eta_n^{-q'}\Big), \end{aligned} \tag{B.47}$$

where ϑ_n and d_n are defined in Theorem 3.3. The assertion of Theorem 4.2 then follows from Theorem 4.1.
Now we prove (B.47). By the first step in the proof of Theorem 3.3 it follows that

$$\begin{aligned} \sup_{x\in\mathbb{R}} \Big| \mathbb{P}(|T^{\sigma,\diamond}/\sqrt{m_n'(2\lceil nb_n \rceil - m_n')}|_\infty \le x | \mathcal{F}_n) - \mathbb{P}\Big(\frac{1}{\sqrt{2nb_n}} \Big| \sum_{i=1}^{2\lceil nb_n \rceil - 1} \tilde{Y}^\sigma_i \Big|_\infty \le x\Big) \Big| \\ = O_p\Big(\vartheta_n^{1/3}\{1 \vee \log(\frac{np}{\vartheta_n})\}^{2/3} + \Theta\Big(\big(\sqrt{m_n \log np}(\frac{1}{\sqrt{nd_n}} + d_n^2)(np)^{\frac{1}{q}}\big)^{q/(q+1)}, np\Big)\Big). \end{aligned} \tag{B.48}$$

Define the $\mathcal{F}_n$ measurable event

$$A_n = \Big\{ \sup_{u\in[0,1],t\in(0,1)} |\hat{\sigma}^2(u,t) - \sigma^2(u,t)| > g_n\eta_n \Big\} ,$$

then Proposition 4.1 and Markov's inequality yield

$$\mathbb{P}(A_n) = O\big(\eta_n^{-q'}\big). \tag{B.49}$$

By similar arguments as given in the proof of Theorem 3.3 we have

$$\mathbb{E}\big(|T^{\sigma,\diamond} - T^{\hat{\sigma},\diamond}|_\infty^q \mathbf{1}(A_n)\big|\mathcal{F}_n\big) \le M\Big|\sqrt{\log np} \max_{r=1}^{(n-2\lceil nb_n\rceil+1)p} \Big(\sum_{j=1}^{\lceil 2nb_n\rceil - m'_n} (\hat{S}^{\sigma}_{jm'_n,r} - \hat{S}^{\hat{\sigma}}_{jm'_n,r})^2 \mathbf{1}(A_n)\Big)^{1/2}\Big|^q \tag{B.50}$$

for some large constant M almost surely, and the triangle inequality, a similar argument as given in the proof of Proposition 1.1 in Dette and Wu (2020) and (B.15) yield

$$\frac{1}{\sqrt{m_n(2\lceil nb_n\rceil - m_n)}}\Big\| \max_{r=1}^{(n-2\lceil nb_n\rceil+1)p} \Big(\sum_{j=1}^{\lceil 2nb_n\rceil - m'_n} (\hat{S}^{\sigma}_{jm'_n,r} - \hat{S}^{\hat{\sigma}}_{jm'_n,r})^2 \mathbf{1}(A)\Big)^{1/2}\Big\|_q = O\big(\sqrt{m_n} g_n\eta_n (np)^{\frac{1}{q}}\big).$$

This together with the (conditional version) of Lemma C.1 and (B.50) shows that

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big(\frac{|T^{\hat{\sigma},\diamond}|_\infty}{\sqrt{m_n(2\lceil nb_n\rceil - m_n)}} > x\Big|\mathcal{F}_n\Big) - \mathbb{P}\Big(\frac{1}{\sqrt{2nb_n}}\Big|\sum_{i=1}^{2\lceil nb_n\rceil -1}\tilde{Y}_i^{\sigma}\Big|_\infty > x\Big)\Big| \\
&\le \sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big(\frac{|T^{\sigma,\diamond}|_\infty}{\sqrt{m_n(2\lceil nb_n\rceil - m_n)}} > x\Big|\mathcal{F}_n\Big) - \mathbb{P}\Big(\frac{1}{\sqrt{2nb_n}}\Big|\sum_{i=1}^{2\lceil nb_n\rceil -1}\tilde{Y}_i^{\sigma}\Big|_\infty > x)\Big| \\
&\qquad + \mathbb{P}\Big(\frac{|T^{\diamond,\sigma} - T^{\diamond,\hat{\sigma}}|_\infty}{\sqrt{m_n(2\lceil nb_n\rceil - m_n)}} > \delta\Big|\mathcal{F}_n\Big) + O\big(\Theta(\delta, np)\big) \\
&\le \sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big(\frac{|T^{\sigma,\diamond}|_\infty}{\sqrt{m_n(2\lceil nb_n\rceil - m_n)}} > x\Big|\mathcal{F}_n\Big) - \mathbb{P}\Big(\frac{1}{\sqrt{2nb_n}}\Big|\sum_{i=1}^{2\lceil nb_n\rceil -1}\tilde{Y}_i^{\sigma}\Big|_\infty > x\Big)\Big| \\
&\qquad + O_p\big(\delta^{-q}\big(\sqrt{m_n\log np}(g_n\eta_n)(np)^{\frac{1}{q}}\big)^q\big) + O\big(\Theta(\delta, np) + \eta_n^{-q'}\big),
\end{aligned}$$

where we used Markov's inequality and (B.49). Taking

$$\delta = \big(\sqrt{m_n\log np}(g_n\eta_n)(np)^{\frac{1}{q}}\big)^{q/(q+1)}$$

and observing (B.48) yields (B.47) and proves the assertion.

C Some auxiliary results

This section contains several technical lemmas, which will be used in the proofs of the main results in Section B.

Lemma C.1. *For any random vectors X, X', Y, and $\delta \in \mathbb{R}$, we have that*

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(|X'| > x) - \mathbb{P}(|Y| > x)| \leq \sup_{x\in\mathbb{R}} |\mathbb{P}(|X| > x) - \mathbb{P}(|Y| > x)| + \mathbb{P}(|X - X'| > \delta) + 2\sup_{x\in\mathbb{R}} \mathbb{P}(|Y - x| \leq \delta). \tag{C.1}$$

Furthermore, if $Y = (Y_1, ..., Y_p)^\top$ is a p-dimensional Gaussian vector and there exist positive constants $c_1 \leq c_2$ such that for all $1 \leq j \leq p$, $c_1 \leq \mathbb{E}(Y_j^2) \leq c_2$, then

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(|X'| > x) - \mathbb{P}(|Y|_\infty > x)| \leq \sup_{x\in\mathbb{R}} |\mathbb{P}(|X| > x) - \mathbb{P}(|Y|_\infty > x)| + \mathbb{P}(|X - X'| > \delta) + C\Theta(\delta, p), \tag{C.2}$$

where C is a constant only dependent on c_1 and c_2.

Proof of Lemma C.1. By triangle inequality, we shall see that

$$\mathbb{P}(|X'| > x) - \mathbb{P}(|Y| > x) \leq \mathbb{P}(|X' - X| > \delta) + \mathbb{P}(|X| > x - \delta) - \mathbb{P}(|Y| > x), \tag{C.3}$$

$$\mathbb{P}(|X'| > x) - \mathbb{P}(|Y| > x) \geq -\mathbb{P}(|X' - X| > \delta) + \mathbb{P}(|X| > x + \delta) - \mathbb{P}(|Y| > x). \tag{C.4}$$

Notice that right-hand side of (C.3) is

$$\mathbb{P}(|X' - X| > \delta) + \mathbb{P}(|X| > x - \delta) - \mathbb{P}(|Y| > x - \delta) + \mathbb{P}(|Y| > x - \delta) - \mathbb{P}(|Y| > x).$$

The absolute value of the above expression is then uniformly bounded by

$$\mathbb{P}(|X' - X| > \delta) + \sup_{x\in\mathbb{R}} |\mathbb{P}(|X| > x) - \mathbb{P}(|Y| > x)| + 2\sup_{x\in\mathbb{R}} \mathbb{P}(|Y - x| \leq \delta). \tag{C.5}$$

Similarly, the absolute value of right-hand side of (C.4) is also uniformly bounded by (C.5), which proves (C.1). Finally, (C.2) follows from (C.1) and an application of Corollary 1 in Chernozhukov et al. (2015). Note that in this result the constant C is determined by $\max_{1\leq j\leq p} \mathbb{E}(Y_j^2) \leq c_2$ and $\min_{1\leq j\leq p} \mathbb{E}(Y_j^2) \geq c_1$. ♢

The following result is a consequence of of Lemma 5 of Zhou and Wu (2010).

Lemma C.2. *Under the assumption 2.2(2) , we have that*

$$\sup_{u_1,u_2,t_1,t_2\in[0,1]} |\mathbb{E}(G(u_1,t,\mathcal{F}_i)G(u_2,t_2,\mathcal{F}_j))| = O(\chi^{|i-j|}).$$

Lemma C.3. *Define*

$$\sigma_{j,j}(u) = \frac{1}{nb_n}\sum_{i,l=1}^{n} \mathrm{Cov}(Z_{i,j}(u), Z_{l,j}(u))$$

where $Z_{i,j}$ are the components of the vector $Z_i(u)$ defined in (3.1)*. If $b_n = o(1)$, $\frac{\log n}{nb_n} = o(1)$ and Assumption 2.4 and Assumption 2.2 are satisfied, there exist positive constants c_1 and c_2 such that for sufficiently large n*

$$0 < c_1 \le \min_{1\le j\le p} \sigma_{j,j}(u) \le \max_{1\le j\le p} \sigma_{j,j}(u) \le c_2 < \infty.$$

for all $u \in [b_n, 1-b_n]$. Moreover, we have for

$$\tilde{\sigma}_{j,j} := \frac{1}{2\lceil nb_n\rceil - 1}\sum_{i,l=1}^{2\lceil nb_n\rceil -1} \mathrm{Cov}(\tilde{Z}_{i,j}, \tilde{Z}_{l,j}), \tag{C.6}$$

the estimates

$$c_1 \le \min_{1\le j\le (n-2\lceil nb_n\rceil+1)p} \tilde{\sigma}_{j,j} \le \max_{1\le j\le (n-2\lceil nb_n\rceil+1)p} \tilde{\sigma}_{j,j} \le c_2.$$

Proof of Lemma C.3. By definition,

$$\sigma_{j,j}(u) = \frac{1}{nb_n}\sum_{i,l=1}^{n} \mathbb{E}\Big(G(\tfrac{i}{n}, t_j, \mathcal{F}_i)K\Big(\frac{\frac{i}{n}-u}{b_n}\Big)G(\tfrac{l}{n}, t_j, \mathcal{F}_l)K\Big(\frac{\frac{l}{n}-u}{b_n}\Big)\Big).$$

Observing Assumption 2.2 and Lemma C.2, we have

$$\mathbb{E}(G(\tfrac{i}{n}, t_j, \mathcal{F}_i)G(\frac{l}{n}, t_j, \mathcal{F}_l) - G(u, t_j, \mathcal{F}_i)G(u, t_j, \mathcal{F}_l)) = O\big(\min(\chi^{|l-i|}, b_n)\big)$$

uniformly with respect to $u \in [b_n, 1-b_n]$, $|\frac{i}{n} - u| \le b_n$ and $|\frac{l}{n} - u| \le b_n$. Consequently, observing Assumption 2.4 it follows that

$$\sigma_{j,j}(u) = \frac{1}{nb_n}\sum_{i,l=1}^{n} \mathbb{E}\Big(G(u, t_j, \mathcal{F}_i)K\Big(\frac{\frac{i}{n}-u}{b_n}\Big)G(u, t_j, \mathcal{F}_l)K\Big(\frac{\frac{l}{n}-u}{b_n}\Big)\Big) + O(-b_n \log b_n) \tag{C.7}$$

On the other hand, if r_n is a sequence such that $r_n = o(1)$ and $nb_nr_n \to \infty$, $A(u, r_n) := \{l : |\frac{\frac{l}{n}-u}{b_n}| \leq 1 - r_n, u \in [b_n, 1 - b_n]\}$ we obtain by (C.7) and Lemma C.2 that

$$
\begin{aligned}
\sigma_{j,j}(u) &= \frac{1}{nb_n} \sum_{l=1}^{n} \sum_{i=1}^{n} \mathbf{1}(|i-l| \leq nb_nr_n) \mathbb{E}\Big(G(u, t_j, \mathcal{F}_i) K\Big(\frac{\frac{i}{n}-u}{b_n}\Big) G(u, t_j, \mathcal{F}_l) K\Big(\frac{\frac{l}{n}-u}{b_n}\Big)\Big) \\
&\quad + O(-b_n \log b_n + \chi^{nb_nr_n}) \\
&= \frac{1}{nb_n} \sum_{l=1}^{n} K^2\Big(\frac{\frac{l}{n}-u}{b_n}\Big) \sum_{\substack{1 \leq i \leq n, \\ |i-l| \leq nb_nr_n}} \mathbb{E}\Big(G(u, t_j, \mathcal{F}_i) G(u, t_j, \mathcal{F}_l) \mathbf{1}\Big(\Big|\frac{\frac{i}{n}-u}{b_n}\Big| \leq 1\Big)\Big) \\
&\quad + O(-b_n \log b_n + \chi^{nb_nr_n} + r_n) \\
&= \frac{1}{nb_n} \sum_{\substack{1 \leq l \leq n, \\ l \in A(u, r_n)}} K^2\Big(\frac{\frac{l}{n}-u}{b_n}\Big) \sum_{\substack{1 \leq i \leq n, \\ |i-l| \leq nb_nr_n}} \mathbb{E}\Big(G(u, t_j, \mathcal{F}_i) G(u, t_j, \mathcal{F}_l) \mathbf{1}\Big(\Big|\frac{\frac{i}{n}-u}{b_n}\Big| \leq 1\Big)\Big) \\
&\quad + O(-b_n \log b_n + \chi^{nb_nr_n} + r_n) \qquad \text{(C.8)}
\end{aligned}
$$

uniformly for $j \in \{1, \ldots, p\}$. We obtain by the definition of the long-run variance $\sigma^2(u, t)$ in Assumption 2.2(4) and Lemma C.2 that

$$
\Big| \sum_{i=1}^{n} \mathbb{E}\Big(G(u, t_j, \mathcal{F}_i) G(u, t_j, \mathcal{F}_l) \mathbf{1}\Big(\Big|\frac{\frac{i}{n}-u}{b_n}\Big| \leq 1, |i-l| \leq nb_nr_n\Big)\Big) - \sigma^2(u, t_j)\Big| = O(\chi^{nb_nr_n}) \qquad \text{(C.9)}
$$

uniformly with respect to $l \in A(u, r_n) := \{l : |\frac{\frac{l}{n}-u}{b_n}| \leq 1 - r_n, u \in [b_n, 1 - b_n]\}$ and $j \in \{1, \ldots, p\}$. Combining (C.8) and (C.9) and using Lemma C.2 yields

$$
\begin{aligned}
\sigma_{j,j}(u) &= \frac{1}{nb_n} \sum_{l=1}^{n} K^2\Big(\frac{\frac{l}{n}-u}{b_n}\Big) \sigma^2(u, t_j) + O(-b_n \log b_n + \chi^{nb_nr_n} + r_n) \\
&= \sigma^2(u, t_j) \int_{-1}^{1} K^2(t) dt + O\Big(-b_n \log b_n + \chi^{nb_nr_n} + r_n + \frac{1}{nb_n}\Big). \qquad \text{(C.10)}
\end{aligned}
$$

Let $r_n = \frac{a \log n}{nb_n}$ for some sufficiently large positive constant a, then the assertion of the lemma follows in view of Assumption 2.2(4)).

For the second assertion, consider the case that $j = k_1 p + k_2$ for some $0 \leq k_1 \leq n - 2\lceil nb_n \rceil - 1$ and $1 \leq k_2 \leq p$. Therefore by definition (3.3) in the main article,

$$
\tilde{Z}_{i,k_1p+k_2} = G(\tfrac{i+k_1}{n}, \tfrac{k_2}{p}, \mathcal{F}_{i+k_1}) K(\tfrac{i-\lceil nb_n \rceil}{nb_n}),
$$

which gives for the quantity in (C.6)

$$\tilde{\sigma}_{k_1p+k_2,k_1p+k_2} = \frac{1}{2\lceil nb_n\rceil - 1}\sum_{i,l=1}^{2\lceil nb_n\rceil-1} \mathbb{E}\Big(G(\tfrac{i+k_1}{n},\tfrac{k_2}{p},\mathcal{F}_{i+k_1})K(\tfrac{i-\lceil nb_n\rceil}{nb_n})G(\tfrac{l+k_1}{n},\tfrac{k_2}{p},\mathcal{F}_{l+k_1})K(\tfrac{l-\lceil nb_n\rceil}{nb_n})\Big)$$

Consequently, putting $i + k_1 = s_1$ and $i + k_2 = s_2$ and using a change of variable, we obtain that

$$\tilde{\sigma}_{k_1p+k_2,k_1p+k_2} = \sigma_{k_2,k_2}\Big(\tfrac{k_1+\lceil nb_n\rceil}{nb_n}\Big), \tag{C.11}$$

which finishes the proof. ♢

References

Aue, A., G. Rice, and O. Sönmez (2018). Detecting and dating structural breaks in functional data without dimension reduction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80*(3), 509–529.

Aue, A. and A. van Delft (2020). Testing for stationarity of functional time series in the frequency domain. *The Annals of Statistics 48*(5), 2505 – 2547.

Berkes, I., R. Gabrys, L. Horváth, and P. Kokoszka (2009). Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society: Series B 71*, 927–946.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy 81*(3), 637–654.

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. Lecture Notes in Statistics. Springer New York.

Bucchia, B. and M. Wendler (2017). Change-point detection and bootstrap for Hilbert space valued random fields. *Journal of Multivariate Analysis 155*, 344–368.

Bücher, A., H. Dette, and F. Heinrichs (2020). Detecting deviations from second-order stationarity in locally stationary functinal time series. *Annals of the Institute of Statistical Mathematics 72*, 1055–1094.

Cao, G., L. Yang, and D. Todem (2012). Simultaneous inference for the mean function based on dense functional data. *Journal of Nonparametric Statistics 24*(2), 359–377.

Chatterjee, S. (2014). *Superconcentration and related topics*, Volume 15. Springer.

Chen, M. and Q. Song (2015). Simultaneous inference of the mean of functional time series. *Electronic Journal of Statistics 9*(2), 1779–1798.

Chernozhukov, V., D. Chetverikov, and K. Kato (2015). Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probability Theory and Related Fields 162*(1), 47–70.

Craven, P. and G. Wahba (1978). Smoothing noisy data with spline functions. *Numerische mathematik 31*(4), 377–403.

Daglish, T., J. Hull, and W. Suo (2007). Volatility surfaces: theory, rules of thumb, and empirical evidence. *Quantitative Finance 7*(5), 507–524.

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

Degras, D. (2017). Simultaneous confidence bands for the mean of functional data. *WIREs Computational Statistics 9*(3), e1397.

Dette, H. and K. Kokot (2020). Bio-equivalence tests in functional data by maximum deviation. *Biometrika, to appear*.

Dette, H., K. Kokot, and A. Aue (2020). Functional data analysis in the Banach space of continuous functions. *Annals of Statistics 48*(2), 1168–1192.

Dette, H., K. Kokot, and S. Volgushev (2020). Testing relevant hypotheses in functional time series via self-normalization. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 82*(3), 629–660.

Dette, H. and W. Wu (2019). Detecting relevant changes in the mean of nonstationary processes—a mass excess approach. *The Annals of Statistics 47*(6), 3578–3608.

Dette, H. and W. Wu (2020). Prediction in locally stationary time series. *Journal of Business & Economic Statistics*, 1–12.

Dette, H., W. Wu, and Z. Zhou (2019). Change point analysis of correlation in non-stationary time series. *Statistica Sinica 29*(2), 611–643.

Ferraty, F. and P. Vieu (2010). *Nonparametric Functional Data Analysis*. Springer-Verlag, New York.

Horváth, L. and P. Kokoszka (2012). *Inference for Functional Data with Applications*. Springer-Verlag, New York.

Horváth, L., P. Kokoszka, and R. Reeder (2013). Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society: Series B 75*, 103–122.

Horváth, L., P. Kokoszka, and G. Rice (2014). Testing stationarity of functional time series. *Journal of Econometrics 179*, 66–82.

Hsing, T. and R. Eubank (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to linear Operators*. New York: Wiley.

Hull, J. C. (2003). *Options futures and other derivatives*. Pearson Education India.

Karmakar, S., S. Richter, and W. B. Wu (2021). Simultaneous inference for time-varying models. *Journal of Econometrics*.

Konakov, V. D. and V. I. Piterbarg (1984). On the convergence rate of maximal deviation distribution for kernel regression estimate. *Journal of Multivariate Analysis 15*, 279–294.

Kurisu, D. (2021a). Nonparametric regression for locally stationary functional time series.

Kurisu, D. (2021b). On the estimation of locally stationary functional time series.

Liebl, D. and M. Reimherr (2019). Fast and fair simultaneous confidence bands for functional parameters. *arXiv:1910.00131*.

Liu, X., H. Xiao, and R. Chen (2016). Convolutional autoregressive models for functional time series. *Journal of Econometrics 194*(2), 263–282.

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

Politis, D. N., J. P. Romano, and M. Wolf (1999). *Subsampling*. Springer Science & Business Media.

Proksch, K. (2014). On confidence bands for multivariate nonparametric regression. *Annals of the Institute of*

Statistical Mathematics 68, 209–236.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis* (Second ed.). Springer, New York.

Telschow, F. and A. Schwartzman (2022). Simultaneous confidence bands for functional data using the Gaussian Kinematic formula. *Journal of Statistical Planning and Inference 216*, 79–94.

van Delft, A. and H. Dette (2021). A similarity measure for second order properties of non-stationary functional time series with applications to clustering and testing. *Bernoulli 27*(1), 469 – 501.

van Delft, A. and M. Eichler (2018). Locally stationary functional time series. *Electronic Journal of Statistics 12*(1), 107 – 170.

Wang, H. and S. Xiang (2012). On the convergence rates of legendre approximation. *Mathematics of Computation 81*(278), 861–877.

Wang, J., G. Cao, L. Wang, and L. Yang (2020). Simultaneous confidence band for stationary covariance function of dense functional data. *Journal of Multivariate Analysis 176*, 104584.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences 102*(40), 14150–14154.

Wu, W. B. and Z. Zhao (2007). Inference of trends in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69*(3), 391–410.

Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *Journal of the Royal Statistical Society, Ser. B 60*, 797–811.

Zhang, X. and G. Cheng (2018). Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli 24*(4A), 2640–2675.

Zheng, S., L. Yan, and W. K. Härdle (2014). A smooth simultaneous confidence mean of sparse functional data. *Journal of the American Statistical Association 109*(506), 1779–1798.

Zhou, Z. (2010). Nonparametric inference of quantile curves for nonstationary time series. *The Annals of Statistics 38*(4), 2187–2217.

Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association 108*(502), 726–740.

Zhou, Z. and H. Dette (2020). Statistical inference for high dimensional panel functional time series. *arXiv preprint arXiv:2003.05968*.

Zhou, Z. and W. B. Wu (2010). Simultaneous inference of linear models with time varying coefficients. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72*(4), 513–531.