Statistica Sinica Preprint No: SS-2023-0150	
Title	Confidence Surfaces for the Mean of Locally Stationary Functional Time Series
Manuscript ID	SS-2023-0150
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0150
Complete List of Authors	Holger Dette and Weichi Wu
Corresponding Authors	Weichi Wu
E-mails	weichi.wu1@gmail.com
Notice: Accepted version subject to English editing.	

Confidence surfaces for the mean of locally stationary functional time series

Holger Dette, Weichi Wu

Ruhr-Universität Bochum and Tsinghua University

Abstract: The problem of constructing a simultaneous confidence surface for the 2-dimensional mean function of a non-stationary functional time series is challenging as these bands can not be built on classical limit theory for the maximum absolute deviation between an estimate and the time-dependent regression function. In this paper, we propose a new bootstrap methodology to construct such a region. Our approach is based on a Gaussian approximation for the maximum norm of sparse high-dimensional vectors approximating the maximum absolute deviation which is suitable for nonparametric inference of high-dimensional time series. The elimination of the zero entries produces (besides the time dependence) additional dependencies such that the "classical" multiplier bootstrap is not applicable. To solve this issue we develop a novel multiplier bootstrap, where blocks of the coordinates of the vectors are multiplied with random variables, which mimic the specific structure between the vectors appearing in the Gaussian approximation. 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 by analyzing a data example.

Key words and phrases: locally stationary times series, functional data, confidence surface, Gaussian approximation, multiplier bootstrap

1. Introduction

In the big data era data gathering technologies provide enormous amounts of data with complex structure. In many applications the observed 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 (see, for example, Shang, 2014). On the other hand, in many applications data is observed on a very fine gird and 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). In such cases fully functional methods can prove advantageous and have been recently, developed by Horváth et al. (2014), Bucchia and Wendler (2017), Aue et al. (2018), Dette et al. (2020) and Dette and Kokot (2020) among others.

In this paper we are interested in statistical inference regarding the smooth mean functions of a not necessarily stationary functional time series $(X_{i,n})$, $1 \leq i \leq n$ in the space $L^2[0,1]$ of square integrable functions on the interval $[0,1]$. As we do not assume stationarity, the mean function $t \to \mathbb{E}[X_{i,n}(t)]$ is changing with i and we assume that it is given by $\mathbb{E}[X_{i,n}(t)] = m(\frac{i}{n}, t)$, where m is a smooth function on the unit square. Our goal is the construction of simultaneous confidence surfaces (SCSs) for the (time dependent) mean function $(u,t) \to m(u,t)$ of the locally stationary functional time series $\{X_{i,n}\}_{i=1,\ldots,n}$. As an illustration we display in Figure 1 the implied volatility of an SP500 index as a function of moneyness (t) at different times to maturity (u, which is scaled to the interval $[0,1]$). These functions are quadratic and known as "volatility smiles" in the literature on option pricing. They seem to slightly vary in time. In practice, it is important to assess whether these "smiles" are time-invariant. We refer the interested reader to Section 4 for a more detailed discussion (in particular

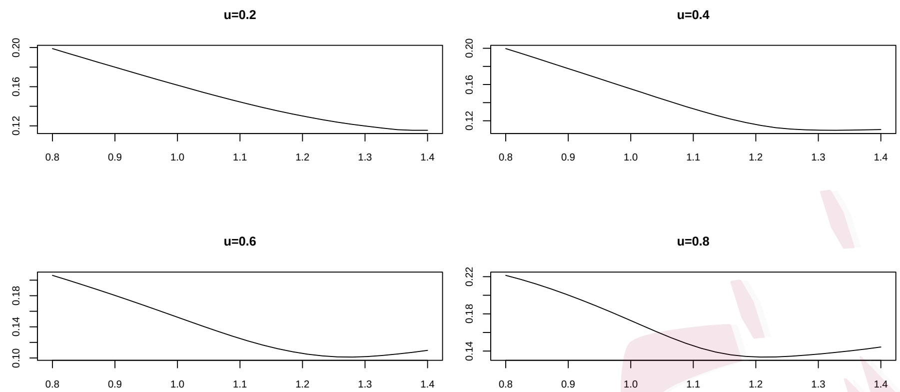

Figure 1: *Volatility Smile at different (one minus) times to maturity (*$u =$ 0.2*,* 0.4*,* 0.6 *and* 0.8*). The x-axis corresponds to 'moneyness'.*

we construct there a confidence surface for the function $(u, t) \to m(u, t))$. To our best knowledge confidence bands have only been considered in the stationary case, where $m(u, t) = m(t)$. Under the assumption of stationarity they can be constructed using the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_{i,n}$ and the weak convergence of $\sqrt{n}(\bar{X}_n - m)$ to a centered Gaussian process (see, for example Degras, 2011; Cao et al., 2012; Degras, 2017; Dette et al., 2020, who either assume that data is observed on a dense grid or that the full trajectory can be observed). More recently, alternative simultaneous confidence (asymptotic) bands have been constructed by Liebl and Reimherr (2019); Telschow and Schwartzman (2022) using the Gaussian Kinematic formula.

On the other hand, although non-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), the problem of constructing a confidence surface for the mean function has not been considered in the literature so far. A potential explanation that a solution is still not available, consists in the fact that due to non-stationarity smoothing is required to estimate the function $u \to m(u,t)$ (for a fixed t). This results in an estimator converging with a $1/\sqrt{b_n n}$ rate (here b_n denotes a bandwidth). On the other hand, in the stationary case (where m does not depend on u), the sample mean $\bar{X}_n$ can be used, resulting in a $1/\sqrt{n}$ rate.

As a consequence, a weak convergence result for the sample mean in the non-stationary case is not available and the construction of SCSs for the regression function $(u,t) \to m(u,t)$ is 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 (for which it is not clear if it exists 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 SCSs for the mean function of a locally stationary functional time series. Our approach is based on approximating the maximal absolute deviation $\hat{\Delta} = \max_{u,t\in[0,1]} |\hat{m}_{b_n}(u,t) - m(u,t)|$ by a maximum taken over a discrete grid, which becomes dense with increasing sample size. We thus relate $\hat{\Delta}$ to the maximum norm of a **sparse** high-dimensional vector. We then further develop Gaussian approximations for the maximum norm of **sparse** high-dimensional random vectors based on the methodology proposed by Chernozhukov et al. (2013) Zhang and Cheng (2018). Finally, the covariance structure of this vector (which is actually a high-dimensional long-run variance) is mimicked by a multiplier bootstrap. Our approach is non-standard in the following sense: due to the sparsity, the Gaussian approximations in the cited literature cannot be directly used. In order to make these applicable we reduce the dimension by deleting vanishing entries. However, this procedure produces additional spatial dependencies (besides the dependencies induced by time series), such that the common multiplier bootstrap is not applicable. Therefore we propose a novel multiplier bootstrap, where instead of the full vector, individual blocks of the vector are multiplied with independent random variables, such that for different vectors a certain amount (depending on the lag) of these multipliers coincide. Our proposed Gaussian approximation and the bootstrap scheme are suitable for nonparametric inference of means of high-dimensional time series. In fact, we first discretize our functional time series to nonstationary high-dimensional time series and then utilize the above-mentioned Gaussian approximation and bootstrap scheme for simultaneous inference.

The remaining part of the paper is organized as follows. The statistical model is introduced in Section 2, where we describe our approach in an informal way and propose two confidence surfaces for the mean function m. Section 3 is devoted to rigorous statements under which conditions our method provides valid (asymptotic) confidence surfaces. As a by-product of our approach, we also derive in the online supplement of this paper 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 (see Section S4 for details). Although our main focus is on SCSs excluding the boundary (as most work in the literature does), we also provide - as a complement - simultaneous inference at the boundary, which is of independent interest, see Remark 2. In Section 4 we demonstrate the usefulness of our approach by means of analyzing a data example. Finally, all technical results are deferred to the online supplement. There, we also give remarks regarding noisy and multivariate locally stationary functional time series and some concrete examples of locally stationary functional time series. Moreover, the online supplement contains implementation details and a simulation study illustrating the finite sample properties of the asymptotic results.

2. SCSs for non-stationary time series

Throughout this paper we consider the model

$$X_{i,n}(t) = m(\tfrac{i}{n}, t) + \varepsilon_{i,n}(t) \ , \quad i = 1, \ldots, n \ , \tag{2.1}$$

where $(\varepsilon_{i,n})_{i=1,\ldots,n}$ is a centered locally stationary process in $L^2[0,1]$ of square integrable functions on the interval $[0,1]$ (see Section 3 for a precise mathematical 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 $t \to X_{i,n}(t)$.

Let $\mathcal{C}^{a,b}$ denote the set of functions $f : [0,1]^2 \to \mathbb{R}$, which are a-times and b-times partially differentiable with respect to the first and second coordinate, respectively, such that for fixed t and for fixed u the functions $u \to \frac{\partial^a}{\partial u^a} f(u,v)$ and $t \to \frac{\partial^b}{\partial t^b} f(u,t)$ are Lipschitz continuous with a uniformly bounded Lipschitz constant on the interval $[0,1]$. In this paper, we are interested in a SCS

$$\mathcal{C}_n = \left\{ f \in \mathcal{C}^{3,0} \mid \ \hat{L}_1(u,t) \le f(u,t) \le \hat{U}_1(u,t) \ \ \forall u,t \ \right\} \tag{2.2}$$

for the mean function $(u,t) \to m(u,t)$, where $\hat{L}_1$ and $\hat{U}_1$ are appropriate lower and upper bounds calculated from the data. The methodology developed in this paper allows the construction of SCSs for the functions $t \to m(u,t)$ (for fixed u) and $u \to m(u,t)$ (for fixed t), which are developed in Section S4 of the online supplement for the sake of completeness. Moreover, in the main part of this paper we investigate SCSs for $u \in [b_n, 1-b_n]$. SCSs on the boundary can be constructed in a similar spirit, even though their form is distinct from their interior counterparts, see Remark 2.

Our approach is based on the maximum deviation

$$\hat{\Delta}_n = \sup_{t,u} |\hat{m}_{b_n}(u,t) - m(u,t)| \ ,$$

where for $u \in [b_n, 1-b_n]$

$$\hat{m}_{b_n}(u,t) = \sum_{i=1}^{n} X_{i,n}(t) K\Big(\frac{\frac{i}{n}-u}{b_n}\Big) \Big/ \sum_{i=1}^{n} K\Big(\frac{i/n-u}{b_n}\Big) \tag{2.3}$$

denotes the common Nadaraya-Watson estimate with kernel K. For $u \in [0, b_n)$ and for $u \in (1-b_n, 1]$ we use boundary kernels, say $K_l(\cdot)$ and $K_r(\cdot)$, respectively, in the definition of $\hat{m}_{b_n}(u,t)$. Throughout this paper, we make the following assumptions on these kernels.

Assumption 2.1. The kernel $K(\cdot)$ is symmetric continuous, supported on the interval $[-1, 1]$ and satisfies $\int_{\mathbb{R}} K(x)dx = 1$, $\int_{\mathbb{R}} K(v)v^2 dv = 0$ and $\int_{\mathbb{R}} K(v)v^4 dv > 0$. K_l (used for estimation on $[0, b_n)$) is supported on the interval $[0, 1]$ and satisfies $K_l(0) = K_l(1) = 0$, $\int K_l(x)x^j dx = 0$ for $j = 1, 2$, $\int K_l(x)dx = 1$, $\int K_l(x)x^3 dx > 0$. Additionally, both K and K_l are twice differentiable on their support, respectively and K'' and K_l'' are Lipschitz continuous. The kernel K_r (used for estimation on $(1 - b_n, 1]$) is given by $K_r(x) = K_l(-x)$.

As a consequence of Assumption 2.1, the bias of the estimator (2.3) is of order $O(b_n^4)$ in the interval $[b_n, 1 - b_n]$, and of order $O(b_n^3)$ at the region $[0, b_n)$ and $(1 - b_n, 1]$ if the function $u \to \frac{\partial^3}{\partial u^3} m(u, v)$ is Lipschitz continuous with bounded Lipschitz constant. As alternative one could consider the local polynomial regression estimate of order 3. As this will make the theoretical analysis even more technical, we leave the study of the statistical properties of local polynomial estimator of higher order for a 2-dimensional mean function of a non-stationary functional time series for future work.

2.1 Confidence surfaces with fixed width

Note that under smoothness assumption the deterministic term in (2.3) approximates $m(u, t)$. For an increasing sample size n, we can approximate the maximum deviation on a discrete grid, i.e.

$$\hat{\Delta}_{b_n} := \max_{\substack{b_n \le u \le 1-b_n, \\ 0 \le t \le 1}} \sqrt{nb_n} |\hat{\Delta}(u,t)| \approx \max_{\substack{\lceil nb_n \rceil \le l \le n - \lceil nb_n \rceil \\ 1 \le k \le p}} \sqrt{nb_n} |\hat{\Delta}(\tfrac{l}{n}, \tfrac{k}{p})|$$

$$\approx \max_{\substack{\lceil nb_n \rceil \le l \le n - \lceil nb_n \rceil \\ 1 \le k \le p}} \Big| \frac{1}{\sqrt{nb_n}} \sum_{i=1}^{n} \varepsilon_{i,n}(\tfrac{k}{p}) K\Big(\frac{\frac{i}{n} - \frac{l}{n}}{b_n}\Big)\Big|, \quad (2.4)$$

where p is increasing with n as well. Therefore, the bootstrap procedure will be based on a Gaussian approximation of the right hand side of (2.4), which is the maximum norm of high-dimensional sparse vector. In this section our approach will be stated in a rather informal way, rigorous statements can be found in Section 3.

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

$$Z_i(u) = (Z_{i,1}(u), \dots, Z_{i,p}(u))^\top \quad (2.5)$$
$$= K\Big(\frac{\frac{i}{n} - u}{b_n}\Big)\Big(\varepsilon_{i,n}(\tfrac{1}{p}), \varepsilon_{i,n}(\tfrac{2}{p}), \dots, \varepsilon_{i,n}(\tfrac{p-1}{p}), \varepsilon_{i,n}(1)\Big)^\top,$$

where $K(\cdot)$ and b_n are the interior kernel and bandwidth used in the estimate (2.3), respectively. Next we define the p-dimensional vector

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

where

$$Z_{i,l,k} = \varepsilon_{i,n}(\tfrac{k}{p}) K\Big(\frac{\frac{i}{n} - \frac{l}{n}}{b_n}\Big) \quad (1 \le k \le p) .$$

Note that, by (2.4),

$$\hat{\Delta}_{b_n} \approx \max_{\substack{\lceil nb_n \rceil \leq l \leq n - \lceil nb_n \rceil \\ 1 \leq k \leq p}} \Big| \frac{1}{\sqrt{nb_n}} \sum_{i=1}^{n} Z_{i,l,k} \Big| \approx \Big| \frac{1}{\sqrt{nb_n}} \sum_{i=1}^{n} (Z_{i,\lceil nb_n \rceil}^\top, \ldots, Z_{i,n-\lceil nb_n \rceil}^\top)^\top \Big|_\infty , \tag{2.6}$$

where $|a|_\infty$ denotes the maximum norm of a finite dimensional vector a (the dimension will always be clear from context). The entries in the vector $Z_{i,l}$ are zero whenever $|i-l|/(nb_n) \geq 1$. Therefore, the high-dimensional vector $(Z_{i,\lceil nb_n \rceil}^\top, \ldots, Z_{i,n-\lceil nb_n \rceil}^\top)^\top$ is sparse and common Gaussian approximations for its maximum norm (see, for example, Chernozhukov et al., 2013), Zhang and Cheng (2018) are not applicable.

To address this issue we reconstruct high-dimensional vectors, say $\tilde{Z}_j$, by eliminating vanishing entries in the vectors $Z_{i,l}$ and rearranging the nonzero ones. While this approach is very natural it produces additional dependencies, which require a substantial modification of the common multiplier bootstrap as considered, for example, in Zhou and Wu (2010), Zhou (2013), Karmakar et al. (2021) or Mies (2021) for (low dimensional) locally stationary time series. More precisely, we define the $(n - 2\lceil nb_n \rceil + 1)p$-dimensional vectors $\tilde{Z}_1, \ldots, \tilde{Z}_{2\lceil nb \rceil - 1}$ by

$$\tilde{Z}_i = \left(Z_{i,\lceil nb_n \rceil}^\top, Z_{i+1,\lceil nb_n \rceil+1}^\top, \ldots, Z_{n-2\lceil nb_n \rceil+i, n-\lceil nb_n \rceil}^\top \right)^\top . \tag{2.7}$$

We also put $\tilde{Z}_{2\lceil nb_n\rceil} = 0$ and note that

$$\Big|\frac{1}{\sqrt{nb_n}}\sum_{i=1}^{n}\bar{Z}_i\Big|_\infty = \Big|\frac{1}{\sqrt{nb_n}}\sum_{i=1}^{2\lceil nb_n\rceil-1}\tilde{Z}_i\Big|_\infty, \tag{2.8}$$

where $\bar{Z}_i := (Z_{i,\lceil nb_n\rceil}^\top, Z_{i,\lceil nb_n\rceil+1}^\top, \ldots, Z_{i,n-\lceil nb_n\rceil}^\top)^\top$. Note that the right hand side of (2.8) is a sum of the $(n-2\lceil nb_n\rceil-1)p$ dimensional vectors

$$\begin{aligned}
\tilde{Z}_1 &= K\big(\tfrac{1-\lceil nb_n\rceil}{nb_n}\big)(\vec{\varepsilon}_1, \vec{\varepsilon}_2, \ldots, \vec{\varepsilon}_{n-2\lceil nb_n\rceil+1})^\top,\\
\tilde{Z}_2 &= K\big(\tfrac{2-\lceil nb_n\rceil}{nb_n}\big)(\vec{\varepsilon}_2, \vec{\varepsilon}_3, \ldots, \vec{\varepsilon}_{n-2\lceil nb_n\rceil+2})^\top,\\
&\vdots\\
\tilde{Z}_{2\lceil nb_n\rceil-1} &= K\big(\tfrac{\lceil nb_n\rceil-1}{nb_n}\big)\big(\vec{\varepsilon}_{2\lceil nb_n\rceil-1}, \vec{\varepsilon}_{2\lceil nb_n\rceil}, \ldots, \vec{\varepsilon}_{n-1}\big)^\top,
\end{aligned} \tag{2.9}$$

where $\vec{\varepsilon}_i = (\varepsilon_{i,n}(\frac{1}{p}), \ldots, \varepsilon_{i,n}(\frac{p}{p}))$. On the other hand the left hand side of (2.8) is a sum of the **sparse** vectors

$$\begin{aligned}
\bar{Z}_1 &= \Big(K\big(\tfrac{1-\lceil nb_n\rceil}{nb_n}\big)\vec{\varepsilon}_1\,,\quad 0\quad, 0\,, \ldots, 0\,,\quad 0\quad\Big)^\top,\\
\bar{Z}_2 &= \Big(K\big(\tfrac{2-\lceil nb_n\rceil}{nb_n}\big)\vec{\varepsilon}_2\,,\ K\big(\tfrac{1-\lceil nb_n\rceil}{nb_n}\big)\vec{\varepsilon}_2\,,\ 0\,,\ \ldots,\ 0\,,\quad 0\quad\Big)^\top,\\
&\vdots\\
\bar{Z}_{n-1} &= \Big(\quad 0\quad,\quad 0\quad, 0\,,\ \ldots,\ 0\,,\ K\big(\tfrac{\lceil nb_n\rceil-1}{nb_n}\big)\vec{\varepsilon}_{n-1}\Big)^\top.
\end{aligned}$$

Although, the vectors on both sides of (2.8) are very different, and the number of terms in the sum is different, the non-vanishing elements over which the maximum is taken on both sides coincide. We note that this transformation yields some computational advantages and, even more important, it allows the development of a Gaussian approximation and a corresponding multiplier bootstrap, which is explained next.

To be precise, observing (2.6), we see that the right hand side of (2.8) is an approximation of the maximum absolute deviation $\max_{u,t} \sqrt{nb_n}|\hat{\Delta}(u,t)|$. In Theorem 1 in Section 3.2 we will show that the vectors $\tilde{Z}_1, \dots, \tilde{Z}_{2\lceil nb \rceil -1}$ in (2.8) can be replaced by Gaussian vectors. More precisely we prove the existence of $(n - 2\lceil nb_n \rceil + 1)p$-dimensional centered Gaussian vectors $\tilde{Y}_1, \dots, \tilde{Y}_{2\lceil nb_n \rceil -1}$ with the same auto-covariance structure as the vector $\tilde{Z}_i$ in (2.7) such that

$$\sup_{x\in\mathbb{R}} \Big|\mathbb{P}\Big(\max_{\substack{b_n \le u \le 1-b_n \\ 0\le t\le 1}} \sqrt{nb_n}|\hat{\Delta}(u,t)| \le x\Big) - \mathbb{P}\Big(\Big|\frac{1}{\sqrt{nb_n}} \sum_{i=1}^{2\lceil nb_n \rceil -1} \tilde{Y}_i\Big|_\infty \le x\Big)\Big| = o(1) \quad (2.10)$$

if p is an appropriate sequence converging to infinity with the sample size (for example, $p = \sqrt{n}$).

The estimate (2.10) is the basic tool for the construction of a SCS for the regression function m. For its application it is necessary to generate Gaussian random vectors $\tilde{Y}_i$ with the same auto-covariance structure as the vector $\tilde{Z}_i$ in (2.7), which is not trivial. To see this, note that the common multiplier bootstrap approach for approximating the distribution of $\frac{1}{\sqrt{nb_n}} \sum_{i=1}^{2\lceil nb_n \rceil -1} \tilde{Z}_i$ replaces the $\tilde{Z}_i$ by block sums multiplied with independent random variables, such as $R_i \sum_{s=i}^{i+m} \tilde{Z}_s/\sqrt{m}$ (see Zhang and Cheng, 2014) or $R_i \sum_{s=i}^{i+m}(Z_s - \frac{1}{2\lceil nb_n\rceil - 1}\sum_{s=1}^{2\lceil nb_n\rceil - 1}\tilde{Z}_i)/\sqrt{m}$ (see Zhou, 2013), where $R_1, \ldots, R_{2\lceil nb_n\rceil}$ are independent standard normally distributed random variables. However, this would not yield to valid approximation due to the additional dependencies between $\tilde{Z}_1, \ldots, \tilde{Z}_{2\lceil nb_n\rceil - 1}$. As an alternative we therefore propose a multiplier bootstrap, which also mimics this dependence structure by multiplying p-dimensional blocks of block sums of $\tilde{Z}_i$ by standard normally distributed random variables, which reflects the specific dependencies of these vectors. In other words the vectors $\tilde{Z}_1, \tilde{Z}_2, \tilde{Z}_3, \ldots$ in (2.9) are replaced by

$$K\big(\tfrac{1-\lceil nb_n\rceil}{nb_n}\big)(\vec{\varepsilon}_{1:1+m_n}R_1, \vec{\varepsilon}_{2:2+m_n}R_2, \ldots, \vec{\varepsilon}_{n-2\lceil nb_n\rceil+1:n-2\lceil nb_n\rceil+1+m_n}R_{n-2\lceil nb_n\rceil+1})^\top,$$
$$K\big(\tfrac{2-\lceil nb_n\rceil}{nb_n}\big)(\vec{\varepsilon}_{2:2+m_n}R_2, \vec{\varepsilon}_{3:3+m_n}R_3, \ldots, \vec{\varepsilon}_{n-2\lceil nb_n\rceil+2:n-2\lceil nb_n\rceil+2+m_n}R_{n-2\lceil nb_n\rceil+2})^\top,$$
$$K\big(\tfrac{3-\lceil nb_n\rceil}{nb_n}\big)(\vec{\varepsilon}_{3:3+m_n}R_3, \vec{\varepsilon}_{4:4+m_n}R_4, \ldots, \vec{\varepsilon}_{n-2\lceil nb_n\rceil+3:n-2\lceil nb_n\rceil+3+m_n}R_{n-2\lceil nb_n\rceil+3})^\top,$$
$$\vdots \tag{2.11}$$

respectively, where

$$\vec{\varepsilon}_{j:j+m_n} = \frac{1}{\sqrt{m}_n}\sum_{r=j}^{j+\lfloor m_n/2\rfloor - 1}\vec{\varepsilon}_r - \frac{1}{\sqrt{m}_n}\sum_{r=j+\lfloor m_n/2\rfloor}^{j+2\lfloor m_n/2\rfloor - 1}\vec{\varepsilon}_r.$$

Here we consider local block sums (of increasing length) to mimic the dependence structure of the error process. A difference of local block sums is used to mitigate the effect of bias if all elements of the unknown errors $\vec{\varepsilon}_i$ are replaced by corresponding nonparametric residuals $\hat{\varepsilon}_{i,n}(s/p)$, $1 \le s \le p$ where $\hat{\varepsilon}_{i,n}(t) = X_{i,n}(t) - \hat{m}_{b_n}(\frac{i}{n}, t)$, where $\hat{m}_{b_n}$ is defined (2.3). We emphasize that the use of boundary kernels in this estimate (see Assumption 2.1) allows us to construct SCSs for the boundary region as well. Details are given in Remark 2 below. With the residuals we define the p-dimensional vector

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

as an analog of (2.5). Similarly, we define the analog of (2.7) by

$$\hat{\hat{Z}}_j = \big(\hat{Z}^\top_{j,\lceil nb_n\rceil}, \hat{Z}^\top_{j+1,\lceil nb_n\rceil+1}, \ldots, \hat{Z}^\top_{n-2\lceil nb_n\rceil+j, n-\lceil nb_n\rceil}\big)^\top, \qquad (2.12)$$

where $\hat{Z}_{i,l} = \hat{Z}_i(\frac{l}{n}) = (\hat{Z}_{i,l,1}, \ldots \hat{Z}_{i,l,p})^\top$. Note that we have replaced $Z_{i,l}$ in (2.7) by $\hat{Z}_{i,l}$, which can be calculated from the data. These vectors will be used in Algorithm 1 to define empirical versions of the vectors in (2.11), which then mimic the dependence structure of the vectors $\tilde{Y}_1, \ldots, \tilde{Y}_{2\lceil nb_n\rceil-1}$ in the Gaussian approximation (2.10) (see equations (2.14) and (2.15) in Algorithm 1). The SCS for the mean function m is finally defined by

$$\mathcal{C}_n = \big\{ f \in \mathcal{C}^{3,0} : [0,1]^2 \to \mathbb{R} |\ \hat{L}_1(u,t) \le f(u,t) \le \hat{U}_1(u,t)\ \forall u \in [b_n, 1-b_n]\ \forall t \in [0,1] \big\}\ , \qquad (2.13)$$

where the definition of functions $\hat{L}_1, \hat{U}_1 : [0,1]^2 \to \mathbb{R}$ is given in Algorithm 1. Finally, Theorem 2 in Section 3 shows that $\mathcal{C}_n$ defines a valid asymptotic $(1-\alpha)$ confidence surface for the regression function m in model (2.1).

Remark 1. In this paper we assume that at each time point the full trajectory is observed. Therefore, smoothing with respect to the variable t is not necessary. Smoothing with respect to both variables becomes necessary if the trajectories are observed with measurement error. Our method is also applicable to dense and discrete observations from the trajectory. In these cases smoothing with respect to the variable t yields a further bias. Another scenario when smoothing is important is the situation where the trajectory is observed at sparse discrete points. This case is beyond the scope of our paper because it requires a different theory.

2.2 Confidence surfaces with varying width

The confidence surface in Algorithm 1 has a constant width and does not reflect the variability of the estimate $\hat{m}$ at the point (u,t). In this section we will construct a SCS adjusted by an estimator of the long-run variance (see equation (4.1) in Section 3.1 for the exact definition). 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

Algorithm 1

(a) Calculate the $(n-2\lceil nb_n\rceil+1)p$-dimensional vectors $\hat{\hat{Z}}_i$ in (2.12)

(b) For window size m_n, let $m_n' = 2\lfloor m_n/2\rfloor$, define the vectors

$$\hat{S}_{jm_n'} = \frac{1}{\sqrt{m_n'}} \sum_{r=j}^{j+\lfloor m_n/2\rfloor-1} \hat{\hat{Z}}_r - \frac{1}{\sqrt{m_n'}} \sum_{r=j+\lfloor m_n/2\rfloor}^{j+m_n'-1} \hat{\hat{Z}}_r \tag{2.14}$$

and denote by $\hat{\vec{\varepsilon}}_{j:j+m_n',k}$ the p-dimensional sub-vector of the vector $\hat{S}_{jm_n'}$ in (2.14) containing its $(k-1)p+1$st $-$ kpth components.

(c) **For** r=1, …, B, **do** Generate i.i.d. $N(0,1)$ random variables $\{R_i^{(r)}\}_{i=1,\dots,n-m_n'}$. Calculate

$$T_k^{(r)} = \sum_{j=1}^{2\lceil nb_n\rceil-m_n'} \hat{\vec{\varepsilon}}_{j:j+m_n',k} R_{k+j-1}^{(r)}, \quad 1\le k\le n-2\lceil nb_n\rceil+1 \tag{2.15}$$

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

end

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

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

where

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

Output: SCS (2.13) for the mean function m.

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 of non-stationary functional data as considered here this task is challenging as an estimator of the long-run variance is required, which is uniformly consistent on the square $[0,1]^2$.

In order to define such an estimator let H denote the Epanechnikov kernel 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 $S_{k,r}^X = \frac{1}{\sqrt{r}} \sum_{i=k}^{k+r-1} X_{i,n}$ denote the normalized partial sum of the data $X_{k,n}, \ldots, X_{k+r-1,n}$ (note that these are functions) and define for $w \geq 2$

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

An estimator of the long-run variance (where the exact definition is in (4.1)) is then defined by

$$\hat{\sigma}^2(u,t) = \sum_{j=1}^{n} \frac{w\Delta_j^2(t)}{2} \bar{\omega}(u,j), \tag{2.16}$$

if $u \in [w/n, 1-w/n]$. For $u \in [0, w/n)$ and $u \in (1-w/n, 1]$ we define it as $\hat{\sigma}^2(u,t) = \hat{\sigma}^2(w/n,t)$ and $\hat{\sigma}^2(u,t) = \hat{\sigma}^2(1-w/n,t)$, respectively. We will show in Proposition 1 in Section 3 that this estimator is uniformly consistent.

To state the bootstrap algorithm for a SCS of the form (2.2) 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 normalized analog

$$\hat{\hat{Z}}_j^{\hat{\sigma}} = \big(\hat{Z}_{j,\lceil nb_n\rceil}^{\hat{\sigma},\top}, \hat{Z}_{j+1,\lceil nb_n\rceil+1}^{\hat{\sigma},\top} \ldots, \hat{Z}_{n-2\lceil nb_n\rceil+j,n-\lceil nb_n\rceil}^{\hat{\sigma},\top}\big)^\top \tag{2.17}$$

of the vector $\hat{\hat{Z}}_j$ in (2.12), where $\hat{Z}_{i,l}^{\hat{\sigma}} = \hat{Z}_i^{\hat{\sigma}}(\frac{l}{n}) = (\hat{Z}_{i,l,1}^{\hat{\sigma}}, \ldots \hat{Z}_{i,l,p}^{\hat{\sigma}})^\top$. The SCS with varying width for the mean function m is then defined by

$$\mathcal{C}_n^{\hat{\sigma}} = \big\{f \in \mathcal{C}^{3,0} : [0,1]^2 \to \mathbb{R} \mid \hat{L}_2^{\hat{\sigma}}(u,t) \le f(u,t) \le \hat{U}_2^{\hat{\sigma}}(u,t) \;\; \forall u \in [b_n, 1-b_n] \, \forall t \in [0,1]\big\}, \tag{2.18}$$

where the functions $\hat{L}_2$ and $\hat{U}_2$ are constructed in Algorithm 2. Theorem 4 in Section 3.2 shows that this defines a valid asymptotic $(1-\alpha)$ confidence surface for the function m in model (2.1).

We also emphasize that we can construct similar SCSs for noisy and multivariate locally stationary functional time series; see Remarks S1 and S2 in the supplementary material for details.

Algorithm 2

(a) Calculate the the estimate of the long-run variance $\hat{\sigma}^2$ in (2.16).

(b) Calculate the $(n-2\lceil nb_n\rceil+1)p$-dimensional vectors $\hat{\tilde{Z}}_i^{\hat{\sigma}}$ in (2.17).

(c) For window size m_n, let $m_n' = 2\lfloor m_n/2\rfloor$, define

$$\hat{S}_{jm_n'}^{\hat{\sigma}} = \frac{1}{\sqrt{m_n'}} \sum_{r=j}^{j+\lfloor m_n/2\rfloor-1} \hat{\tilde{Z}}_r^{\hat{\sigma}} - \frac{1}{\sqrt{m_n'}} \sum_{r=j+\lfloor m_n/2\rfloor}^{j+m_n'-1} \hat{\tilde{Z}}_r^{\hat{\sigma}}$$

and denote by $\hat{\vec{\varepsilon}}_{j:j+m_n',k}^{\,\hat{\sigma}}$ be the p-dimensional sub-vector of the vector $\hat{S}_{jm_n'}^{\hat{\sigma}}$ containing its $(k-1)p+1$st $-$ kpth components.

(d) **For** $r=1,\dots,B$ **do** Generate i.i.d. $N(0,1)$ random variables $\{R_i^{(r)}\}_{i=1,\dots,n-m_n'}$. Calculate

$$T_k^{\hat{\sigma},(r)} = \sum_{j=1}^{2\lceil nb_n\rceil-m_n'} \hat{\vec{\varepsilon}}_{j:j+m_n',k}^{\,\hat{\sigma}} R_{k+j-1}^{(r)}, \quad k=1,\dots,n-2\lceil nb_n\rceil+1,$$

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

end

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

$$\hat{L}_2^{\hat{\sigma}}(u,t) = \hat{m}_{b_n}(u,t) - \hat{r}_2(u,t), \qquad \hat{U}_2^{\hat{\sigma}}(u,t) = \hat{m}_{b_n}(u,t) + \hat{r}_2(u,t),$$

where

$$\hat{r}_2(u,t) = \frac{\hat{\sigma}(u,t)\sqrt{2}T_{\lfloor(1-\alpha)B\rfloor}^{\hat{\sigma}}}{\sqrt{nb_n}\sqrt{2\lceil nb_n\rceil - m_n'}}$$

Output: SCS (2.18) with varying width for the mean function m.

3. Theoretical justification

In this section, we first present the locally stationary functional time series model for which the theoretical results of this paper are derived (Section 3.1). We also describe under which conditions Algorithm 1 and 2 provide valid asymptotic $(1-\alpha)$ confidence surfaces for the regression function m in model (2.1) (Section 3.2 and 3.3). Throughout this paper we use the notation

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

for positive constants a, b, and the notation $a \vee b$ denotes the maximum of the real numbers a and b.

3.1 Locally stationary processes and physical dependence

We begin with an assumption for the mean function m in model (2.1).

Assumption 3.1. $m \in \mathcal{C}^{3,0}$.

In fact, in the proof of Theorem 3, we show that the difference between $\hat{m}_{b_n}(u,t)$ 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, an approximation of the form (2.4) 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 3.2. 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

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

(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 $q > 2$ the dependence measure

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

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

$$\delta_q(G, i) = O(\chi^i) \ .$$

(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*

$$\sigma^2(u,t) := \sum_{k=-\infty}^{\infty} \mathrm{Cov}(G(u,t,\mathcal{F}_0), G(u,t,\mathcal{F}_k)) \tag{4.1}$$

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 3.2(2) requires that the dependence measure is geometrically decaying. Similar results as presented in this section can be obtained under summability assumptions with substantially more intensive mathematical arguments and complicated notation, see Remark S3(ii) in the supplemental material for some details. Assumption 3.2(3) means that the locally stationary functional time series is smooth in u, while the smoothness in t is provided in the next assumption. They are crucial for constructing SCSs of the form (2.2).

Assumption 3.3. The filter G in Assumption 3.2 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 that there exists a constant $q^* > 2$ such that for some $\chi \in (0, 1)$ and $i \geq 0$,

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

In the online supplement we present several examples of locally stationary functional time series satisfying these assumptions (see Section S5).

3.2 Theoretical analysis of the methodology in Section 2.1

The bootstrap methodology introduced in Section 2 is based on the Gaussian approximation (2.10), which will be stated rigorously in Theorem 1. Theorem 2 shows under which conditions the confidence surface (2.13) has asymptotic level $(1 - \alpha)$.

Theorem 1 (Justification of Gaussian approximation (2.10))**.** *Let Assumptions 2.1, 3.1 - 3.3 be satisfied and assume that* $n^{1+a} b_n^9 = o(1)$, $n^{a-1} b_n^{-1} = o(1)$ *for some* $0 < a < 4/5$. *Then there exists a sequence of centered* $(n - 2\lceil nb_n \rceil + 1)p$*-dimensional centered 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* (2.7) *such that the distance* $\mathfrak{P}_n$ *defined in* (2.10) *satisfies*

$$\mathfrak{P}_n = O\Big((nb_n)^{-(1-11\iota)/8} + \Theta\Big(\sqrt{nb_n}(b_n^4 + \frac{1}{n}), np\Big) + \Theta\big(((np)^{1/q^*}((nb_n)^{-1} + 1/p))^{\frac{q^*}{q^*+1}}, np\big)\Big)$$

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

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

if the constant q^* *in Assumption 3.3 is sufficiently large.*

In Section S2.4 of the supplementary material we investigate the finite sample properties of the approximation in Theorem 1 by means of a simulation study. Moreover, Theorem 1 is the main ingredient to prove the validity of the bootstrap SCS $\mathcal{C}_n$ defined in (2.13) by Algorithm 1. More precisely, we have the following result.

Theorem 2. *Assume that the conditions of Theorem 1 hold. Recall that* m_n *is the block size defined in* (2.11) *Define*

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

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

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

then the SCS (2.13) *constructed by Algorithm 1 satisfies*

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

in probability.

Theorem 2 can also be built on alternative assumptions, such as the polynomial decaying instead of geometric decaying dependence measure as in Assumption 3.2(2). Due to page limit, we relegate further discussions on conditions to Section S3 of the supplemental material.

3.3 Theoretical analysis of the methodology in Section 2.2

In this section we will prove that the surface (2.18) defines an asymptotic $(1-\alpha)$ confidence surface with varying width for the mean function m. If the long-run variance in (4.1) would be known, a confidence surface could be based on the "normalized" maximum deviation of

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

Therefore we will derive a Gaussian approximation for the vector $\big(\hat{\Delta}^{\sigma}(\frac{l}{n},\frac{k}{p})\big)$, $l=1,\ldots,n; k=1,\ldots,p$ first and define for $1 \le i \le n$ the p dimensional vector

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

where $\varepsilon_{i,n}^{\sigma}(t) = \varepsilon_{i,n}(t)/\sigma(\frac{i}{n},t)$. Similarly as in Section 2.2 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

$$Z^{\sigma}_{i,l,k} = \varepsilon^{\sigma}_{i,n}(\tfrac{k}{p})K\Big(\frac{\frac{i}{n}-\frac{l}{n}}{b_n}\Big) \quad (1 \le k \le p).$$

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

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

and obtain the following result.

Theorem 3. *Let the Assumptions of Theorem 1 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 centered Gaussian vectors* $\tilde{Y}^{\sigma}_1, \ldots, \tilde{Y}^{\sigma}_{2\lceil nb_n\rceil-1}$ *with the same auto-covariance structure as the vector* $\tilde{Z}^{\sigma}_i$ *in* (4.2) *such that*

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

for any sequence $p \to \infty$ *with* $np = O(\exp(n^{\iota}))$ *for some* $0 \le \iota < 1/11$. *In particular, for the choice* $p = n^c$ *for any* $c > 0$ *we have* $\mathfrak{P}^{\sigma}_n = o(1)$ *if the constant* q^* *in Assumption 3.3 is sufficiently large, such that*

$$\Theta\big(\big[(np)^{1/q^*}((nb_n)^{-1} + 1/p)\big]^{\frac{q^*}{q^*+1}}, np\big) = o(1).$$

The next result shows that the estimator $\hat{\sigma}$ defined by (2.16) is uniformly consistent. Thus the unknown long-run variance σ^2 in Theorem 3 can be replaced by $\hat{\sigma}^2$ and the result can be used to prove the validity of the confidence surface (2.18) defined by Algorithm 2.

Proposition 1. *Let the assumptions of Theorem 1 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 Lipschitz 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*

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

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

where

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

$\gamma_n = \tau_n + w/n$*,* $q' = \min(q, q^*)$ *and* q, q^* *are defined in Assumptions 3.2 and 3.3, respectively.*

We investigate the finite sample performance of the long-run variance estimator $\hat{\sigma}^2$ in Section S2.5 of the supplementary material by means of a simulation study. Proposition 1 and Theorem 3 yield that $\mathfrak{P}_n^{\hat{\sigma}} = o_p(1)$ provided that $\mathfrak{P}_n^{\sigma} = o(1)$.

Theorem 4. *Assume that the conditions of Theorem 2, Proposition 1 hold, that* $p = n^c$ *for some* $c > 0$*, and that* q^* *in Theorem 3 satisfies*

$\Theta\big(\big[(np)^{1/q^*}((nb_n)^{-1}+1/p)\big]^{\frac{q^*}{q^*+1}}, np\big) = o(1)$. *Further assume there exists a sequence* $\eta_n \to \infty$ *such that*

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

where γ_n, g_n *and* q' *are defined in Proposition 1. Then the SCS* (2.18) *defined by Algorithm 2 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 2. Note that the SCS derived so far exclude the boundary region $[0, b_n) \cup (1-b_n, b_n]$ for the varable u. This is common practice in the context of simultaneous inference for kernel based estimates, as inference at the boundary is very difficult due to the inaccurate estimation and sophisticated statistical properties of most nonparametric estimators at the boundary. The problem of simultaneous confidence bands including the boundary region has even not been thoroughly investigated for one-dimensional responses (see for example Zhou and Wu, 2010; Wu and Zhou, 2017). In the following discussion we provide a first solution to this problem. Since the bias of the usual local linear estimates at the boundary is of order $O(b_n^2)$ and therefore too large for simultaneous inference, we use higher order one sided kernel for the boundary region. Simple calculations show that the bias of the NW estimator with this kernel can be of order $O(\frac{1}{n}+b_n^3)$. We discuss the constant width SCS for the boundary, while the varying SCS could be constructed in a similar way. Let $\tilde{K}_l(v)=\sum_{i=1}^n K_l(\frac{i/n-v/n}{b_n})$ and $\tilde{K}_r(v)=\sum_{i=1}^n K_r(\frac{i/n-v/n}{b_n})$. Then, similar to (2.12), we define

$$\hat{Z}_i^l(u)=K_l\Big(\frac{\frac{i}{n}-u}{b_n}\Big)\big(\hat{\varepsilon}_{i,n}(\tfrac{1}{p}),\varepsilon_{i,n}(\tfrac{2}{p}),\ldots,\hat{\varepsilon}_{i,n}(\tfrac{p-1}{p}),\hat{\varepsilon}_{i,n}(1)\big)^\top$$
$$\hat{Z}_i^r(u)=K_r\Big(\frac{\frac{i}{n}-u}{b_n}\Big)\big(\hat{\varepsilon}_{i,n}(\tfrac{1}{p}),\hat{\varepsilon}_{i,n}(\tfrac{2}{p}),\ldots,\hat{\varepsilon}_{i,n}(\tfrac{p-1}{p}),\hat{\varepsilon}_{i,n}(1)\big)^\top$$

and consider for $1\le s\le\lceil nb_n\rceil$ the $\lceil nb_n\rceil$ and $\lceil nb_n\rceil+1$ dimensional vectors $\hat{\tilde{Z}}_s^l$and $\hat{\tilde{Z}}_s^r$ as

$$\hat{\tilde{Z}}_s^l=(\hat{Z}_s^{l\top}(\tfrac{1}{n})/\tilde{K}_l(1),\hat{Z}_{s+1}^{l\top}(\tfrac{2}{n})/\tilde{K}_l(2),\ldots,Z_{s+\lceil nb_n\rceil-2}^{l\top}(\tfrac{\lceil nb_n\rceil-1}{n})/\tilde{K}_l(\lceil nb_n\rceil-1))^\top$$
$$\hat{\tilde{Z}}_s^r=(\hat{Z}_{n'+s}^{r\top}(\tfrac{n'+\lceil nb_n\rceil}{n})/\tilde{K}_r(n'+\lceil nb_n\rceil),\hat{Z}_{n'+s+1}^{r\top}(\tfrac{n'+\lceil nb_n\rceil+1}{n})/\tilde{K}_r(n'+\lceil nb_n\rceil+1),\ldots,$$
$$Z_{n'+s+\lceil nb_n\rceil-1}^{r\top}(1)/\tilde{K}_r(n))^\top$$

where $n'=n-2\lceil nb_n\rceil+1$. For window size m_n, let $m_n'=2\lfloor m_n/2\rfloor$, define the vectors

$$\hat{S}_{jm_n'}^l=\frac{1}{\sqrt{m_n'}}\sum_{r=j}^{j+\lfloor m_n/2\rfloor-1}\hat{\tilde{Z}}_r^l-\frac{1}{\sqrt{m_n'}}\sum_{r=j+\lfloor m_n/2\rfloor}^{j+m_n'-1}\hat{\tilde{Z}}_r^l.$$

Let $\hat{\vec{\varepsilon}}^{\,l}_{j:j+m_n',k}$ be the p-dimensional sub-vector of the vector $\hat{S}_{jm_n'}^l$ containing its $(k-1)p+1$st $-$ kpth components. Then for $v=1,\ldots,B$, we generate i.i.d. $N(0,1)$ random variables $\{R_i^{(v)}\}_{i=1,\ldots,2\lceil nb_n\rceil-m_n'}$. and calculate for $k = 1,\ldots,\lceil nb_n\rceil$

$$T_k^{l,(v)} = \sum_{j=1}^{\lceil nb_n\rceil-m_n'} \hat{\vec{\tilde{\varepsilon}}}_{j:j+m_n',k}^{\,l} R_{k+j-1}^{(v)}\ , \quad T^{l,(r)} = \max_{1\le k\le\lceil nb_n\rceil} |T_k^{(r)}|_\infty.$$

Similarly, using $\hat{\tilde{Z}}_s^r$ and another sequence of $i.i.d.$ $N(0,1)$ random variables $\{V_i^{(v)}\}_{i=1,\ldots,2\lceil nb_n\rceil-m_n'}$ we could generate $T^{r,(v)}$ for $v = 1,\ldots B$, where $\{V_i^{(v)}\}$ and $\{R_i^{(v)}\}$ are independent. Define $T_{\lfloor(1-\alpha)B\rfloor}$ as the empirical $(1-\alpha)$-quantile of the bootstrap sample $T^{l,(1)},\ldots,T^{l,(B)}, T^{r,(1)},\ldots,T^{r,(B)}$, then the lower and upper bound of the $(1-\alpha)$-SCS for $u\in[0,b_n)\cup(1-b_n,1]$ are given by $\hat{L}_{\text{boundary}}(u,t) = \hat{m}_{b_n}(u,t) - \hat{r}_{\text{boundary}}$ and $\hat{U}_{\text{boundary}}(u,t) = \hat{m}_{b_n}(u,t) + \hat{r}_{\text{boundary}}$, respectively where $\hat{r}_{\text{boundary}} = \frac{\sqrt{nb_n}T_{\lfloor(1-\alpha)B\rfloor}}{\sqrt{\lceil nb_n\rceil-m_n'}.}$ and $u\in[0,b_n)\cup(1-b_n,1]$, $t\in[0,1]$. We examine the empirical coverage probabilities of this SCS in Section S2.3 of the online supplement.

4. 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).

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 calculate the SCSs 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 IV 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 (2.1), 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 straightforward to extend this to an arbitrary com-

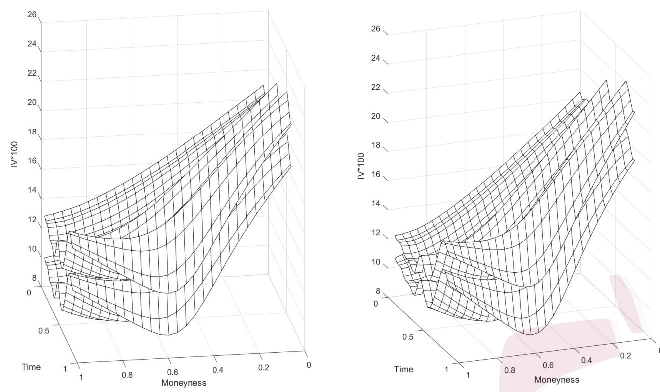

Figure 2: *95% SCS of the form* (2.2) *for the IV surface. Left panel: constant width (Algorithm 1); Right panel: variable width (Algorithm 2).*

pact 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.

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 not flat and are stochastically changing in practice. We can provide an inference tool for such a conclusion using the SCSs developed in Section 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 $(1-\alpha)$ SCS of the form (2.2).

Therefore we construct the 95% SCS for the regression function m with constant and varying width using Algorithm 1 and Algorithm 2, respectively. Following Section S2.1 of supplemental material we choose $b_n = 0.1$ and $m_n = 36$. The results are depicted in Figure 4 (for a better illustration the z-axis shows 100× implied volatility). We observe from both figures that the SCSs do not contain a surface of the form $m(u,t) = m(t)$ and therefore reject the null hypothesis (at significance level 5%). In the supplement, we construct the simultaneous confidence bands of IV w.r.t. fixed t and fixed u.

Supplementary Materials

contains further results for noisy and multivariate locally stationary functional time series, the confidence bands for the functions $t \to m(u,t)$ (fixed u) and $u \to m(u,t)$ (fixed t), implementation details, simulation and additional data analysis results, examples of locally stationary functional time series. It also includes all the detailed proofs.

Acknowledgements

Holger Dette gratefully acknowledges Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Project A1, C1) of the German Research Foundation (DFG). Weichi Wu gratefully acknowledges NSFC no.12271287 and no. 11901337. Weichi Wu is the corresponding author.

References

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

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

Bücher, A., Dette, H., and Heinrichs, F. (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., Yang, L., and Todem, D. (2012). Simultaneous inference for the mean function based on dense functional data. *Journal of Nonparametric Statistics*, 24(2):359–377.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6):2786–2819.

Daglish, T., Hull, J., and Suo, W. (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 Kokot, K. (2020). Bio-equivalence tests in functional data by maximum deviation. *Biometrika, to appear*.

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

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

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*.

Springer-Verlag, New York.

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

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

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

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

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

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

Mies, F. (2021). Functional estimation and change detection for nonstationary time series. *Journal of the American Statistical Association*, pages 1–12.

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

Shang, H. L. (2014). A survey of functional principal component analysis. *AStA Advances in Statistical Analysis*, 98(2):121–142.

Telschow, F. and Schwartzman, A. (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 Dette, H. (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 Eichler, M. (2018). Locally stationary functional time series. *Electronic Journal of Statistics*, 12(1):107 – 170.

Wu, W. and Zhou, Z. (2017). Nonparametric inference for time-varying coefficient quantile regression. *Journal of Business & Economic Statistics*, 35(1):98–109.

Zhang, X. and Cheng, G. (2014). Bootstrapping high dimensional time series. *arXiv preprint arXiv:1406.1037*.

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

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

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

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