Statistical Inference for High Dimensional Panel Functional Time Series

Zhou Zhou and Holger Dette
University of Toronto and Ruhr-Universität Bochum

February 27, 2020

Abstract

In this paper we develop statistical inference tools for high dimensional functional time series. We introduce a new concept of physical dependent processes in the space of square integrable functions, which adopts the idea of basis decomposition of functional data in these spaces, and derive Gaussian and multiplier bootstrap approximations for sums of high dimensional functional time series. These results have numerous important statistical consequences. Exemplarily, we consider the development of joint simultaneous confidence bands for the mean functions and the construction of tests for the hypotheses that the mean functions in the spatial dimension are parallel. The results are illustrated by means of a small simulation study and in the analysis of Canadian temperature data.

AMS Subject Classification: G2M10, G2G10, G2G09
Key words: *high dimensional functional time series, physical dependence, Gaussian approximation, simultaneous confidence bands, hypotheses tests, spatio-temporal data*

1 Introduction

In many fields of statistics data, say $X_{i,j}$, are recorded in time (denoted by the index i) at different locations (denoted by the index j). This type of data is called panel or spatio-temporal data and appears in numerous applications. For example, in economics data are often collected for different firms or regions over several years, or in geo-statistics and climate research data are recorded over a period of time at different locations. Statistical methods for this type of data are meanwhile well developed, in particular in the field of econometrics and spatial statistics, and we refer to the monographs of Hsiao (2003), Baltagi (2005), Wooldridge (2010), Cressie and Wikle (2015) and Haining and Li (2020) among many others. In recent years there has also been substantial interest in statistical inference tools for high dimensional time series and we mention exemplarily the work of Cressie and Johannesson (2008), Galvao and Montes-Rojas (2010), Belloni et al. (2016), Banerjee (2017) and Kock and Tang (2019) who developed methodology for various high dimensional spatio-temporal models.

A common feature of most of the literature on statistical methodology in this context consists in the fact that the data $X_{i,j}$ are real valued or multivariate. However in many modern applications more and more data are recorded continuously during a time interval or intermittently at a very dense grid of time points. In such circumstances it is often reasonable to use a high-dimensional functional time series model, where the data $X_{i,j}$ observed at time i in panel j is a function. Throughout this paper we call this type of data functional panel data or high-dimensional functional time series. For example, $X_{i,j}$ could represent the (smoothed) curve of the price of a stock for a day, where i denotes the index for the different days and j is the index for the different stocks. For another example, $X_{i,j}$ could represent the (smoothed) temperature or precipitation curve for a year i at a location j.

Although functional data analysis is nowadays a rather well developed and very broad field [see for example the monographs of Bosq (2000), Ramsay and Silverman (2005), Ferraty and Vieu (2010), Horváth and Kokoszka (2012) and Hsing and Eubank (2015)], there does not exist much literature for panels of functional data, in particular for high dimensional panels. Several authors have developed statistical methodology for spatio-temporal data [see Delicado et al. (2010), Gromenko et al. (2017), Kokoszka and Reimherr (2019) among others], but this literature usually does not consider the high-dimensional case, where the spatial dimension is increasing with the sample size. High dimensional functional time series have been investigated in the context of forecasting and factor analysis [see Gao et al. (2019) or Nisol et al. (2019)]. Other authors considered statistical inference tools for longitudinal functional data with a general hierarchical structure, where independent subjects or units are observed at repeated times and at each time, a functional observation (curve) is recorded [see for example, Greven et al. (2010), Park and Staicu (2015) or Chen et al. (2017) among others].

The goal of this paper is to develop statistical methodology for high-dimensional functional panel data, which exhibits dependencies in the time **and** spatial directions. In Section 2 we introduce a new concept of physical dependent processes in the space of square integrable functions on the interval $[0, 1]$. Our approach is similar in spirit to the model of m-approximable functional time series introduced by Hörmann and Kokoszka (2010), but our formulation further adopts the idea of basis decomposition of functional data and consequently separates the functional index and time index in the mathematical representation of the functional processes. We also refer to the work of Bosq (2002), Bradley (2007) or Panaretos and Tavakoli (2013) for other concepts modelling functional dependent data such as mixing conditions or autocorrelations and cumulants. It is found in our theoretical investigations that it is simpler and more efficient to establish Gaussian approximation, chaining and bootstrap results for high dimensional functional time series using our physical representation. In Section 3 we provide some probabilistic results which are useful for the statistical analysis of high dimensional functional pane data. In particular we derive a Gaussian approximation for sums (with respect to time) of high dimensional panels, which is used for the construction of a multiplier bootstrap procedure to approximate the distribution of sums uniformly with respect to the spatial dimension. These results have numerous applications in the statistical inference of high-dimensional functional time series and we illustrate the potential of our approach in Section 4. Here we derive joint simultaneous confidence bands for the mean functions of a high-dimensional functional time series and construct a test to check the hypothesis of parallelism of the mean functions. Section 5 is devoted to a small simulation study to illustrate the finite sample properties of the new methodology. We also present a data example analyzing yearly temperature curves from from different Canadian cities. Finally, all proofs are deferred to the online supplemental material of the paper.

2 Physical representation of functional time series

In this section we introduce the basic functional time series model considered in this paper. To be precise, let $\mathcal{L}^2[0,1]$ denote the set of all real-valued square integrable functions defined on the interval $[0,1]$ and let $\{X_i\}_{i\in\mathbb{Z}}$ be a stationary functional time series in $\mathcal{L}^2[0,1]$. We begin with a motivating example for the general model defined below.

Example 2.1. Let $\mathcal{B}=\{B_j\}_{j=0}^{\infty}$ be a basis of $\mathcal{L}^2[0,1]$, then the i-th element X_i of a time series $\{X_i\}_{i\in\mathbb{Z}}$ in $\mathcal{L}^2[0,1]$ can be represented as

$$X_i(u)=\sum_{j=0}^{\infty} L_{i,j}B_j(u) \quad (i\in\mathbb{Z}) \ , \tag{2.1}$$

where $L_{i,j}=\int_a^b X_i(u)B_j(u)du$ is the jth Fourier coefficient of X_i with respect to the basis $\mathcal{B}$. If $\mathbf{L}_i=(L_{i,0},L_{i,1},\cdots)^\top$ denotes the infinite-dimensional vector of Fourier coefficients of X_i, one can use any dependence concept for the infinite dimensional stationary time series $\{\mathbf{L}_i\}_{i\in\mathbb{Z}}$ to describe the dependence structure of the functional time series $\{X_i\}_{i\in\mathbb{Z}}$ in $\mathcal{L}^2[0,1]$. Following Wu (2005), a general model for the infinite dimensional stationary time series $\{\mathbf{L}_i\}_{i\in\mathbb{Z}}$ is given by

$$\mathbf{L}_i=\mathbf{G}(\mathcal{F}_i), \quad i\in\mathbb{Z},$$

where $\mathbf{G} : \mathcal{S}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ is a given filter, $\mathcal{F}_i = (\cdots, \eta_{i-1}, \eta_i)$ and $\{\eta_i\}_{i\in\mathbb{Z}}$ is a sequence of independent and identically distributed random variables taking values in some space $\mathcal{S}$. In this case each element X_i of the time series can be written in the form $X_i(u) = H(u, \mathcal{F}_i)$ for an appropriate function H mapping a sequence of Fourier coefficients to its corresponding function in $\mathcal{L}^2[0,1]$.

Using Example 2.1 as motivation, we now introduce the following formulation of a stationary functional time series and the associated dependence concept, which will be considered in this paper.

Definition 2.1. *A stationary functional time series $\{X_i(u)\}_{i\in\mathbb{Z}}$ in $\mathcal{L}^2[0,1]$ has a physical representation, if there exists a measurable function $H : [0,1] \times \mathcal{S}^{\mathbb{N}} \to \mathbb{R}$ such that*

$$X_i(u) = H(u, \mathcal{F}_i), \quad \forall u \in [0,1], \tag{2.2}$$

where $\mathcal{F}_i = (\cdots, \eta_{i-1}, \eta_i)$ and $\{\eta_i\}_{i\in\mathbb{Z}}$ is a sequence of independent identically distributed random elements in $\mathcal{S}$.

Note that a functional time series $\{X_i\}_{i\in\mathbb{Z}} \subset \mathcal{L}^2[0,1]$ of the form (2.2) is strictly stationary by this definition. In order to specify its dependence, we define for a real valued random variable Z and a constant $q \geq 1$ (in the case of existence) its norm $||Z||_q = \big(\mathbb{E}[|Z|^q|]\big)^{1/q}$ and introduce the following measure of physical dependence.

Definition 2.2. *Let $\{X_i\}_{i\in\mathbb{Z}}$ denote a functional time series in $\mathcal{L}^2[0,1]$ of the form (2.2), such that $\|X_i(u)\|_q < \infty$. The dependence measures of $\{X_i\}_{i\in\mathbb{Z}}$ are defined by*

$$\delta_H(k,q) = \sup_{u\in[0,1]} \|H(u, \mathcal{F}_k) - H(u, \mathcal{F}_k^*)\|_q, \quad k \geq 0, \tag{2.3}$$

where $\mathcal{F}_k^* = (\cdots, \eta_{-2}, \eta_{-1}, \eta_0^*, \eta_1, \cdots, \eta_{k-1}, \eta_k)$ and η_0^* is identically distributed as η_0 and is independent of the sequence $\{\eta_i\}_{i\in\mathbb{Z}}$.

In (2.2) the representation H can be viewed as a filter or data generating function of a stochastic system and $\{\eta_i\}_{i\in\mathbb{Z}}$ can be viewed as random shocks or innovations of the system. Adopting the latter point of view, the physical dependence measures $\delta_H(k,q)$ quantify the influence of the innovations k steps ahead on the current output of the system. Weak dependence is characterized by the fast decay of $\delta_H(k,q)$ as k grows. Note that our definition of a functional time series and their dependence measures are different from the popular concept of $\mathcal{L}^p$-m-approximability (see, for example, Hörmann and Kokoszka (2010) and Horváth and Kokoszka (2012)), where $X_i(u)$ is expressed as $G(\epsilon_i(u), \epsilon_{i-1}(u), \cdots)$ with

$\{\epsilon_i(u)\}_{i\in\mathbb{Z}}$ i.i.d. random functions and the dependence is measured by the changes in $X_i(u)$ when certain $\epsilon_j(u)$ in the past are replaced with i.i.d. copies. Both formulations are based on Bernoulli shifts and coupling to quantify and control the dependence strength of functional time series. However, the formulation considered in this paper further adopts the idea of basis decomposition of functional data and hence separates the functional index u and time index i in (2.2). It is found in our theoretical investigations that it is simpler and more efficient to establish Gaussian approximation and bootstrap results for high dimensional functional time series using the representation (2.2), conditions on the decay of the dependence measure (2.3) and the associated basis decomposition.

We continue giving two examples on how to check the rate of decay of $\delta_H(k,q)$ for linear and nonlinear functional time series models.

Example 2.2. Consider the following MA(∞) functional linear model

$$X_i(u) = \sum_{j=0}^{\infty} \int_0^1 a_j(u,v)\epsilon_{i-j}(v)\,dv,$$

where $\{a_j\}_{j\geq 0}$ is a sequence of square integrable functions $a_j : [0,1]^2 \to \mathbb{R}$ satisfying $\sum_{j=0}^{\infty} \sup_{u,v\in[0,1]} |a_j(u,v)| < \infty$, and $\{\epsilon_i\}_{i\in\mathbb{Z}}$ is a sequence of i.i.d. random functions in $\mathcal{L}^2[0,1]$. Writing $\epsilon_i(u) = \sum_{j=0}^{\infty} \eta_{i,j} B_j(u)$ with $\eta_i = (\eta_{i,j})_{j\geq 0}^{\top}$, we see that X_i can be represented in the form of (2.2). Assume that, for some $q > 1$, $\sup_{u\in[0,1]} \|\epsilon_i(u)\|_q < \infty$, then simple calculations using Definition 2.3 show that the physical dependence measures in (2.3) satisfy

$$\delta_H(k,q) = O\Big(\sup_{u\in[0,1]} \int_0^1 |a_k(u,v)|\,dv\Big).$$

Example 2.3. In this example we continue our investigation of the nonlinear model introduced in Example 2.1. Recall the representation (2.1) and let $b_j = \sup_{u\in[0,1]} |B_j(u)|$ and $l_j = \mathbb{E}|L_{i,j}|$. The rate of decay of l_j is determined by the smoothness of the functions X_i as well as the basis functions B_j used in the decomposition (2.1). See Proposition 2.1 below for the calculation of l_j when $X_i(u)$ is twice continuously differentiable and B_j is the cosine basis. Now write

$$X_i(u) = \sum_{j=0}^{\infty} l_j \tilde{L}_{i,j} B_j(u),$$

with $\tilde{L}_{i,j} = L_{ij}/l_j = \tilde{G}_j(\mathcal{F}_i)$. Denote the physical dependence measure of $\{\tilde{L}_{i,j}\}_{i\in\mathbb{Z}}$ by (cf. Wu (2005)) $\theta_j(k,q) := \|\tilde{G}_j(\mathcal{F}_k) - \tilde{G}_j(\mathcal{F}_k^*)\|_q$. Then using (2.3), we obtain that

$$\delta_H(k,q) = O\Big(\sum_{j=0}^{\infty} l_j b_j \theta_j(k,q)\Big). \tag{2.4}$$

For a wide class of nonlinear time series models, Wu (2005) contains detailed calculations of the rate of decay of their physical dependence measures. As l_j and b_j can be calculated as discussed above, the estimate (2.4) can be used to bound the dependence measures of a wide class of functional time series. A further simplification can be carried out if $\sum_{j=0}^{\infty} l_j b_j < \infty$. In this case we have $\delta_H(k,q) = O(\theta(k,q))$, where $\theta(k,q) = \sup_{j\geq 0} \theta_j(k,q)$.

Throughout this paper we will work with the basis $\{\cos(k\pi u)\}_{k=0,1,\ldots}$ of $\mathcal{L}^2[0,1]$. If $\{X_i\}_{i\in\mathbb{Z}}$ is a functional time series in $\mathcal{L}^2[0,1]$ such that (2.2) holds for some filter H, it can be represented as a Fourier series

$$X_i(u) = \sum_{k=0}^{\infty} a_{i,k} \cos(k\pi u),$$

where

$$a_{i,k} = 2\int_0^1 \cos(k\pi u) X_i(u) du, \tag{2.5}$$

$k \geq 1$ and $a_{i,0} = \int_0^1 X_i(u)du$ are random variables which can be written in the form

$$a_{i,k} = G_k(\mathcal{F}_i) = 2\int_0^1 \cos(k\pi u) H(u, \mathcal{F}_i) du$$

for some filter G_k. The following result specifies the rate of decay of the Fourier coefficients under smoothness conditions on the function X_i.

Proposition 2.1. *Suppose that $\{X_i\}_{i\in\mathbb{Z}}$ is a time series in $\mathcal{L}^2[0,1]$ such that (2.2) is satisfied and $\|X_i(u)\|_q < \infty$ for some $q \geq 2$. Assume X_i is twice continuously differentiable on the interval [0,1] a.s. and that*

$$\|X_1(0)\|_q + \|X_1'(0)\|_q + \sup_{u\in[0,1]} \|X_1''(u)\|_q < \infty. \tag{2.6}$$

Then the Fourier coefficients in (2.5) satisfy

$$\|a_{i,k}\|_q \leq \frac{C_q}{k^2}$$

for some constant C_q which is independent of i, k. Furthermore,

$$\|a_{i,k} - a_{i,k}^*\|_q = O(\min(\delta_H(i,q), 1/k^2)),$$

where $a_{i,k}^* = G_k(\mathcal{F}_i^*) = \int_0^1 \cos(k\pi u) H(u, \mathcal{F}_i^*) du$.

Proposition 2.1 establishes the rate of decay of $a_{i,k}$ as a function of k as well as the dependence measure of $a_{i,k}$ when viewed as a time series indexed by i. In the following we will consider the standardized sum

$$S_n(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i(u)$$

and its approximation

$$S_n(k,u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i(k,u),$$

where $X_i(k,u) = \sum_{j=1}^{k} a_{i,j} \cos(j\pi u)$ is the kth Fourier sum of $X_i(u)$.

Proposition 2.2. *If the assumptions of Proposition 2.1 are satisfied, $\mathbb{E}(X_i(u)) = 0$ and $\delta_H(i,q) = O(i^{-\beta})$ for some $\beta > 2$, we have*

$$\Big\| \sup_{u\in[0,1]} |S_n(u) - S_n(k,u)| \Big\|_q \leq C k^{(2-\beta)/\beta},$$

where C is a finite constant which does not depend on k or n.

Proposition 2.2 establishes that $S_n(u)$ can be well approximated by $S_n(k,u)$ if k is sufficiently large. The latter is an important result due to the fact that the theoretical investigation of $S_n(k,u)$ is much easier as it can be written as a linear combination of *finitely many* random variables.

We conclude this section with a result on chaining of functional time series. To be precise define a grid $\mathcal{U}_n = \{t_{i,n}\}_{i=0}^{l_n}$ where $t_{i,n} = i/l_n$ and l_n is a positive integer that diverges to infinity. The following proposition establishes that $\sup_{u\in[0,1]} |S_n(k,u)|$ can be well approximated by $\max_{u\in\mathcal{U}_n} |S_n(k,u)|$ for a sufficiently dense grid (that is $l_n \to \infty$).

Proposition 2.3. *If the assumptions of Proposition 2.2 are satisfied, we have*

$$\Big\| \max_{0\leq i\leq l_n-1} \sup_{u\in[t_{i,n},t_{i+1,n}]} |S_n(k,u) - S_n(k,t_{i,n})| \Big\|_q \leq C \frac{k^{2/\beta}}{l_n^{1-1/q}},$$

where C is a finite constant which does not depend on k or n.

3 Main results

In this section we derive a Gaussian approximation for partial sums of high dimensional functional time series which can be used to define a multiplier bootstrap procedure.

Throughout this paper $\{\mathbf{X}_i\}_{i=1}^n$ denotes an r-dimensional stationary functional times with $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,r})^\top$, where each component $X_{i,j}$ of the vector $\mathbf{X}_i$ is an element of $\mathcal{L}^2[0,1]$ satisfying (2.2) for some filter H_j, that is $X_{i,j}(u) = H_j(u, \mathcal{F}_i)$ $(j = 1, 2, \cdots, r)$. Consequently, the r-dimensional vector can be represented as

$$\mathbf{X}_i(u) = \mathbf{H}(u, \mathcal{F}_i) \tag{3.1}$$

$(i = 1, \ldots, n,\ u \in [0,1])$, where the r-dimensional filter is defined by $\mathbf{H} = (H_1, \ldots, H_r)^\top$. We assume that the dimension

$$r \asymp n^{\theta_1} \tag{3.2}$$

increases at a polynomial rate with the sample size n, where $\theta_1 > 0$ is a given constant. We are interested in the probabilistic properties of the sums of high-dimensional vectors

$$\mathbf{S}_n(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{X}_i(u) \tag{3.3}$$

as $n, r \to \infty$, and denote by

$$S_{n,j}(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_{i,j}(u) \qquad (j = 1, \ldots, r) \tag{3.4}$$

the j-th component of the vector $\mathbf{S}_n = (S_{n,1}, \ldots, S_{n,r})^T$.

3.1 Gaussian approximation

Consider the component-wise Fourier expansion of the r-dimensional function $\mathbf{X}_i$

$$\mathbf{X}_i(u) = \sum_{j=0}^\infty \mathbf{a}_{i,j} \cos(j\pi u)$$

and define the vector

$$\mathbf{A}_{n,j} = (A_{n,j,1}, \ldots A_{n,j,r})^\top = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{a}_{i,j}. \tag{3.5}$$

For $n \in \mathbb{N}$ let $\{\mathbf{N}_{n,k}\}_{k=0}^\infty$ be a sequence of independent centered r-dimensional Gaussian random variables that preserves the covariance structure of $\{\mathbf{A}_{n,k}\}_{k=0}^\infty$, i.e. Cov$(\mathbf{A}_{n,k})$ =Cov$(\mathbf{N}_{n,k})$, define

$$\mathbf{S}_n^N(u) = \left(S_{n,1}^N(u), \ldots, S_{n,r}^N(u)\right)^\top = \sum_{k=0}^\infty \mathbf{N}_{n,k} \cos(k\pi u) \tag{3.6}$$

and denote by $S^N_{n,j}(u)$ the j-th entry of the vector $\mathbf{S}^N_n(u) = (S^N_{n,1}(u), \ldots, S^N_{n,r}(u))^\top$. By the proof of Proposition 2.2, we have for l-th coordinates $A_{n,j,l}$ of the vectors $\mathbf{A}_{n,j}$ in (3.5)

$$\sum_{j=0}^{\infty} \|A_{n,j,l}\|_q < \infty \ ,$$

and therefore the random variable $\mathbf{S}^N_n(u)$ in (3.6) is well defined (almost surely).

Theorem 3.1. *Let $\{\mathbf{X}_i\}_{i\in\mathbb{Z}}$ denote a stationary r-dimensional time series satisfying (3.1). For each j, suppose that $X_{i,j}$ is twice continuously differentiable on the interval $[0,1]$ a.s.; satisfies $\|X_{ij}\|_q < \infty$ and (2.6) for some $q \geq 4$. Assume $\mathbb{E}(\mathbf{X}_i(u)) = 0$ and*

$$\max_{1\leq j\leq r} \delta_{H_j}(i,q) = O(i^{-\beta}) \tag{3.7}$$

for some $\beta > 3$. If there exists a positive constant δ such that

$$\min_{1\leq j\leq r} \inf_{u\in[0,1]} \mathbb{E}[S^2_{n,j}(u)] \geq \delta \tag{3.8}$$

for sufficiently large n and the exponent θ_1 in (3.2) satisfies

$$\theta_1 < f(\beta,q) := \frac{q-2}{2[1+\beta/[(q-1)(\beta-2)]]}, \tag{3.9}$$

we have as $n, r \to \infty$

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\Big[\max_{1\leq j\leq r} \sup_{0\leq u\leq 1} |S_{n,j}(u)| \leq x\Big] - \mathbb{P}\Big[\max_{1\leq j\leq r} \sup_{0\leq u\leq 1} |S^N_{n,j}(u)| \leq x\Big]\right| \to 0. \tag{3.10}$$

It is easy to show that, under assumptions of Theorem 3.1, the weak convergence

$$S_{n,j}(u) \Rightarrow \mathcal{N}(0, \sigma^2(j,u))$$

holds for any $j = 1, \ldots, r$ and $u \in [0,1]$, where the symbol $\mathcal{N}(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2. Consequently, assumption (3.8) is rather mild and means that, uniformly in u and j, $S_{n,j}(u)$ will not converge to a degenerate limit.

3.2 Bootstrapping panel functional time series

Theorem 3.1 provides an approximation (in distribution) of $\max_{1\leq j\leq r} \sup_{u\in[0,1]} |S_{n,j}(u)|$ by a corresponding expression using normally distributed random variables. The latter distribution can be easily simulated if the dependence structure of the process $\{\mathbf{N}_{n,k}\}_{k=0}^{\infty}$

would be known. In order to mimic the dependencies, we propose a multiplier bootstrap method. To be precise, define for a block size m the local (r-dimensional) mean

$$\mathbf{T}_{i,m}(u) = \frac{1}{m}\sum_{j=i}^{i+m} \mathbf{X}_j(u) \tag{3.11}$$

and consider the vector

$$\mathbf{\Phi}_m(u) = \sqrt{\frac{m}{n-m}} \sum_{i=1}^{n-m} [\mathbf{T}_{i,m}(u) - \mathbf{S}_n(u)/\sqrt{n}]N_i, \tag{3.12}$$

where N_i are i.i.d. one-dimensional standard normal random variables. Let $\Phi_{m,j}(u)$ be the j-th entry of $\mathbf{\Phi}_m(u)$, then we have the following result.

Theorem 3.2. *Let $\{\mathbf{X}_i\}_{i\in\mathbb{Z}}$ denote a stationary r-dimensional time series satisfying (3.1). For each j, suppose that $X_{i,j}$ is twice continuously differentiable on the interval $[0,1]$ a.s. satisfying $\|X_{ij}\|_q < \infty$ and (2.6) for some $q > 2$ and assume that*

$$\|X''_{i,j}(u) - X''_{i,j}(v)\|_q \leq C|u - v| \tag{3.13}$$

holds for all $u, v \in [0,1]$ for sufficiently large n. Further assume that and (3.8) holds and that

$$\max_{1\leq j\leq r} [\delta_{H_j}(i,q) + \delta_{H'_j}(i,q)] = O(i^{-\beta}) \tag{3.14}$$

for some $\beta > 2$ and $q > 4$, where $H'_j(u, \mathcal{F}_i)$ is the filter corresponding to the derivative $X'_{i,j}(u)$ of the process of $X_{i,j}(u) = H_j(u, \mathcal{F}_i)$. If the block size m satisfies $m \asymp n^{\phi}$ with $0 < \phi < 1$ and the exponent θ_1 in (3.2) satisfies $\theta_1 < \phi' q^2/[2(q+1)]$ where $\phi' = -\max\{(\phi - 1)/2, -\phi\}$, then on a sequence of events E_n such that $\mathbb{P}(E_n) \to 1$ if $r, n \to \infty$, we have

$$\sup_{x\in\mathbb{R}} \Big| \mathbb{P}\Big[\max_{1\leq j\leq r} \sup_{0\leq u\leq 1} |\Phi_{m,j}(u)| \leq x \;\Big| \{\mathbf{X}_i\}_{i=1}^n\Big] - \mathbb{P}\Big[\max_{1\leq j\leq r} \sup_{0\leq u\leq 1} |S^N_{n,j}(u)| \leq x\Big]\Big| \to 0.$$

4 Some statistical applications

The results of Section 3 can be used for statistical inference of high-dimensional time series, if the statistical analysis is based on a vector of sums of the form (3.4). Typical applications are tests for parametric assumptions on the mean function, two (or more) sample comparisons, the construction of confidence regions for parameters of interest, change point analysis, to name just a few. Exemplarily, we illustrate in this section two applications, namely the construction of joint simultaneous confidence bands for the mean functions and the development of a test that the mean functions at different spatial locations are parallel.

4.1 Joint simultaneous confidence bands (JSCB)

let $\mathcal{C}^2[0,1]$ denote the space of twice continuously differentiable functions on the interval $[0,1]$, let $\{\mathbf{X}_i\}_{i=1}^n$ be an r-dimensional stationary functional time series in $(\mathcal{C}^2[0,1])^r$ satisfying (3.1) and denote by $\mathbf{g}(u) := \mathbb{E}(\mathbf{X}_i(u)) = (g_1(u),\ldots,g_r(u))^\top$ the expectation of $\mathbf{X}_i$. This subsection is devoted to the construction of joint (asymptotic) simultaneous confidence bands (JSCB) for the components of the vector $\mathbf{g}$ if the sample size n and dimension r converge to infinity. More precisely, we aim to find (random) functions $c_1, d_1, \ldots, c_r, d_r$ defined on the interval $[0,1]$, such that $c_j(u) \le d_j(u)$ for all $u \in [0,1]$ and such that the subset

$$\mathcal{C}_{n,\alpha} = \Big\{\mathbf{h} \in (\mathcal{C}^2[0,1])^r \mid c_j(u) \le h_j(u) \le d_j(u)\ \forall u \in [0,1]\ \forall j = 1,\ldots,r\Big\}. \tag{4.1}$$

of the set of r-dimensional functions $\mathbf{h} = (h_1,\ldots,h_r)^\top$ on the interval $[0,1]$ satisfies

$$\lim_{n,r\to\infty} \mathbb{P}(\mathbf{g} \in \mathcal{C}_{n,\alpha}) = 1-\alpha$$

for some pre-specified constant $\alpha \in (0,1)$. To this end, we denote by

$$v_j(u) = \sqrt{\mathrm{Var}(X_{i,j}(u))}, \quad j = 1,2,\cdots,r,$$

the variance of $X_{i,j}(u)$ and make the following assumption.

(V) There exists a positive constant $\delta > 0$ such that

$$\inf_{u\in[0,1]} \min_{1\le j\le r} v_j(u) \ge \delta$$

Proposition 4.1. *Let $\{\mathbf{X}_i\}_{i\in\mathbb{Z}}$ denote a stationary r-dimensional time series satisfying (3.1). For each j, suppose that $X_{i,j}$ is twice continuously differentiable on the interval $[0,1]$ a.s. satisfies $\|X_{ij}\|_q < \infty$ and (2.6) for some $q > 2$. Assume that (3.7), (3.8), (3.9) and condition (V) hold. Let $\check{X}_{i,j}(u) = (X_{i,j}(u) - g_j(u))$ and define*

$$\check{S}_{n,j}(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \check{X}_{i,j}(u) \qquad j = 1,2,\cdots,r. \tag{4.2}$$

Then, as $r,n\to\infty$, we have

$$\sup_{x\in\mathbb{R}} \Big|\mathbb{P}\Big[\max_{1\le j\le r}\sup_{0\le u\le 1} |\check{S}_{n,j}(u)/v_j(u)| \le x\Big] - \mathbb{P}\Big[\max_{1\le j\le r}\sup_{0\le u\le 1} |S^N_{n,j}(u)/v_j(u)| \le x\Big]\Big| \to 0,$$

where $S^N_{n,j}(u)$ is the jth component of the vector $\mathbf{S}^N_n = (S^N_{n,1}(u),\ldots,S^N_{n,r}(u))^\top$ defined in (3.6).

Proposition 4.1 implies that if one can find a critical value $c^n_{1-\alpha}$ such that

$$\lim_{r,n\to\infty} \mathbb{P}\Big[\max_{1\le j\le r}\sup_{0\le u\le 1}|S^N_{n,j}(u)/v_j(u)| \le c^n_{1-\alpha}\Big] = 1-\alpha,$$

then a $100(1-\alpha)\%$ JSCB for the vector $\mathbf{g} = (g_1,\dots,g_r)^\top$ is obtained by (4.1) using the functions

$$c_j(u) \;=\; \frac{S_{n,j}(u) - c^n_{1-\alpha}v_j(u)}{\sqrt{n}}, \qquad j = 1,\dots,r, \tag{4.3}$$

$$d_j(u) \;=\; \frac{S_{n,j}(u) + c^n_{1-\alpha}v_j(u)}{\sqrt{n}}, \qquad j = 1,\dots,r, \tag{4.4}$$

where $S_{n,j}(u)$ is defined in (3.4). As the standard deviation v_j and the quantile $c^n_{1-\alpha}$ are not known, we have to estimate these from the data. In particular v_j^2 is estimated by the sample variance of $\{X_{i,j}(u)\}_{i=1}^n$ defined as

$$\hat{v}_j^2(u) = \frac{1}{n}\sum_{i=1}^n \big(X_{i,j}(u) - S_{n,j}(u)/\sqrt{n}\big)^2.$$

The following lemma establishes that $\hat{v}_j^2$ is a uniformly consistent estimator.

Lemma 4.1. *For each j, suppose $X_{i,j}$ is twice continuously differentiable on the interval [0,1] a.s. and satisfies (2.6) for some $q > 2$. Assume that (3.7) holds and that $\theta_1 < q/4$. Then we have $r, n \to \infty$*

$$\Big\|\max_{1\le j\le r}\sup_{0\le u\le 1}|\hat{v}_j^2(u) - v_j^2(u)|\Big\|_{q/2} = O(r^{2/q}/\sqrt{n}) = o(1).$$

In order to obtain the critical value $c^n_{1-\alpha}$, we now utilize the multiplier bootstrap procedure established in Theorem 3.2. Detailed steps for the implementation are listed as follows.

(a) Select block size m such that $m \to \infty$, $m = o(n)$.

(b) For a large integer B generate independent standard normal distributed random variables $\{N_i^j\}_{i=1}^{n-m}$, $j = 1, 2, \cdots, B$. For each j, calculate

$$\mathbf{\Phi}^j_m(u) = \sqrt{\frac{m}{n-m}}\sum_{i=1}^{n-m}[\mathbf{T}_{i,m}(u) - \mathbf{S}_n(u)/\sqrt{n}]N_i^j$$

and

$$\tilde{\Phi}^j_m = \max_{1\le k\le r}\sup_{0\le u\le 1}[|\Phi^j_{m,k}(u)|/\hat{v}_j(u)],$$

where $\Phi^j_{m,k}(u)$ is the k-th component of the vector $\mathbf{\Phi}^j_m(u) = (\Phi^j_{m,1}(u),\dots,\Phi^j_{m,r}(u))^\top$.

(c) Let $\tilde{\Phi}_m^{(1)} \leq \cdots \leq \tilde{\Phi}_m^{(B)}$ be the ordered statistics of $\{\tilde{\Phi}_m^j\}_{j=1}^B$. Then the quantile $c_{1-\alpha}^n$ is estimated by $\tilde{\Phi}_m^{(\lfloor(1-\alpha)B\rfloor)}$.

(d) The JSCB is then defined by (4.1) using the functions in (4.3) and (4.4) with these estimates, that is

$$\begin{aligned} c_j(u) &= \frac{S_{n,j}(u) - \tilde{\Phi}_m^{(\lfloor(1-\alpha)B\rfloor)}\hat{v}_j(u)}{\sqrt{n}}, \\ d_j(u) &= \frac{S_{n,j}(u) + \tilde{\Phi}_m^{(\lfloor(1-\alpha)B\rfloor)}\hat{v}_j(u)}{\sqrt{n}}. \end{aligned} \qquad j = 1, \ldots, r, \tag{4.5}$$

A data driven rule for the choice of the block length in step (a) will be given in Section 4.3. The following proposition establishes that these definitions yield an asymptotically correct JSCB.

Proposition 4.2. *Under the conditions of Theorem 3.2 and Proposition 4.1, we have for the JSCB defined in (4.1) with the functions c_j and d_j defined in (4.5)*

$$\lim_{n,r\to\infty} \lim_{B\to\infty} \mathbb{P}(\mathbf{g} \in \mathcal{C}_{n,\alpha}) = 1 - \alpha.$$

Proposition 4.2 follows directly from Theorem 3.2, Proposition 4.1 and Lemma 4.1. The details are omitted for the sake of brevity. The finite coverage probabilities of these joint simultaneous confidence bands are investigated in Section 5.1 by means of a simulation study.

4.2 Test of parallelism

Suppose we observe a panel of functional time series $\{\mathbf{X}_i\}_{i=1}^n$ of dimension r satisfying (3.1). Adopting the same notation as in Section 4.1, for $j = 1, 2, \cdots, r$, let $g_j(u) = \mathbb{E}[X_{i,j}(u)]$ be the mean function of panel j. Denote by $g_j^o = \int_0^1 g_j(u)\, du$ the average of $g_j(u)$ with respect to u, $j = 1, 2, \cdots, r$. In this subsection, we are interested in testing the null hypothesis that the mean functions in the different panels are parallel, that is

$$H_0 : g_1(\cdot) \parallel g_2(\cdot) \parallel \cdots \parallel g_r(\cdot), \tag{4.6}$$

versus the alternative that at least two mean functions are not parallel. Here the symbol $\parallel$ denotes parallelism of two functions; that is, $g_i \parallel g_j$ if and only if the difference $g_i(u) - g_j(u)$ is a constant function.

Now let

$$\hat{g}_j(u) = \frac{1}{n}\sum_{i=1}^{n} X_{i,j}(u) \text{ and } \hat{g}_j^o = \int_0^1 \hat{g}_j(u)\,du$$

be the corresponding sample estimates of $g_j(u)$ and g_j^o, respectively, and define

$$\nu_j(u) = g_j(u) - g_j^o \ ; \quad \hat{\nu}_j(u) = \hat{g}_j(u) - \hat{g}_j^o.$$

To test the hypothesis (4.6) we consider the following statistic

$$T_n = \sqrt{n} \max_{1\le j<k\le r} \sup_{0\le u\le 1} |\hat{\nu}_j(u) - \hat{\nu}_k(u)|/v_{j,k}(u), \tag{4.7}$$

where $v_{j,k}(u) > 0$ is a measure of scale for $\hat{\nu}_j(u) - \hat{\nu}_k(u)$ which will be defined below. Observe that the null hypothesis in (4.6) is satisfied if and only if $\nu_j(u) = \nu_k(u)$ for all $u \in [0,1]$ and all pairs k and j, $1 \le j < k \le r$. As a result, the test statistic T_n should be large if there exists at least one pair of non-parallel functions. Furthermore, T_n can be viewed as a family-wise-error-controlled pair-wise test of parallelism among the curves $g_1, \ldots, g_r$ and one can use the test T_n to decide which curves are parallel to each other and which ones are not with a given family-wise error (asymptotically). Consequently, the statistic T_n can be used to cluster curves with similar shapes together.
In order to obtain the critical values for a test, which rejects the null hypothesis (4.6) for large values of T_n, we define the quantities

$$\begin{aligned} W_{i,j}(u) &= X_{i,j}(u) - \int_0^1 X_{i,j}(u)\,du \\ W_{i,j,k}(u) &= W_{i,j}(u) - W_{i,k}(u) \end{aligned}$$

and

$$\mathring{S}_{n,j,k}(u) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_{i,j,k}(u), \quad 1 \le j < k \le r. \tag{4.8}$$

With these notations it is easy to see that the statistic T_n in (4.7) can be represented as

$$T_n = \max_{1\le j<k\le r} \sup_{0\le u\le 1} |\mathring{S}_{n,j,k}(u)|/v_{j,k}(u). \tag{4.9}$$

From this representation we observe immediately that Theorems 3.1 and 3.2 can be used to determine its critical values. Furthermore, it is easy to see from the above representation that a natural choice for the scaling factors is

$$v_{j,k}(u) = \sqrt{\mathrm{Var}[W_{i,j,k}(u)]},$$

where in practice, one replaces $v_{j,k}(u)$ with its sample version, that is

$$\hat{v}_{j,k}(u) = \Big\{ \frac{1}{n} \sum_{i=1}^{n} \Big[W_{i,j,k}(u) - \frac{1}{n} \sum_{s=1}^{n} W_{s,j,k}(u) \Big]^2 \Big\}^{1/2}.$$

Let $\mathring{\mathbf{W}}_i(u)$ be the vector of dimension $r(r-1)/2$ with entries $W_{i,j,k}(u)$, $1 \le j < k \le r$, consider its (coordinate-wise) Fourier expansion

$$\mathring{\mathbf{W}}_i(u) = \sum_{j=0}^{\infty} \mathring{\mathbf{a}}_{i,j} \cos(j\pi u)$$

and define the $r(r-1)/2$-dimensional vector

$$\mathring{\mathbf{A}}_{n,j} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathring{\mathbf{a}}_{i,j} \ .$$

For $n \in \mathbb{N}$ let $\{\mathbf{N}_{n,k}\}_{k=0}^{\infty}$ be a sequence of independent centered $r(r-1)/2$-dimensional Gaussian vectors that preserves the covariance structure of $\{\mathring{\mathbf{A}}_{n,k}\}_{k=0}^{\infty}$, i.e.

$$\mathrm{Cov}(\{\mathring{\mathbf{A}}_{n,k}\}_{k=0}^{\infty}) = \mathrm{Cov}(\{\mathbf{N}_{n,k}\}_{k=0}^{\infty}),$$

and define the vector

$$\mathring{\mathbf{S}}_n^N(u) = \sum_{l=0}^{\infty} \mathring{\mathbf{N}}_{n,l} \cos(l\pi u) = \big(\mathring{S}_{n,j,k}^N(u)\big)_{1 \le j < k \le r},$$

where the random variables $\mathring{S}_{n,j,k}^N(u)$ denote the entries of $\mathring{\mathbf{S}}_n^N(u)$. By the proof of Proposition 2.2, we have $\sum_{k=0}^{\infty} \|\mathring{A}_{n,k,a,b}\|_q < \infty$ where the random variables $\mathring{A}_{n,k,a,b}$ denote the entries of the vector $\mathring{\mathbf{A}}_{n,k}$ $(1 \le a < b \le r)$. Hence the random variable $\mathring{\mathbf{S}}_n^N(u)$ is well defined almost surely.

Proposition 4.3. *For each j, suppose that $X_{i,j}$ is twice continuously differentiable on the interval [0,1] a.s. and satisfies (2.6) for some $q > 2$. Assume that (3.7) holds; that (3.8) holds with $S_{n,j}^2(u)$ therein replaced by $\mathring{S}_{n,j,k}^2$; that (3.9) holds with θ_1 therein replaced by $2\theta_1$; and that condition (V) hold with v_j therein replaced by $v_{j,k}$, $1 \le j < k \le r$. If $n, r \to \infty$, we have, under the null hypothesis (4.6) of parallelism*

$$\sup_{x \in \mathbb{R}} \Big| \mathbb{P}\Big[\max_{1 \le j < k \le r} \sup_{0 \le u \le 1} \Big| \frac{\mathring{S}_{n,j,k}(u)}{v_{j,k}(u)} \Big| \le x \Big] - \mathbb{P}\Big[\max_{1 \le j < k \le r} \sup_{0 \le u \le 1} \Big| \frac{\mathring{S}_{n,j,k}^N(u)}{v_{j,k}(u)} \Big| \le x \Big] \Big| \to 0.$$

Proposition 4.3 reveals that under the null hypothesis of parallel mean functions $g_1, \ldots, g_r$ the distribution of the statistic T_n in (4.7) can be well approximated by the law of the L^∞ norm of a high-dimensional vector of Gaussian random functions. Theorem 3.2 implies that the multiplier bootstrap can be used to approximate this L^∞ norm. A combination of these results motivates the following bootstrap test for the hypothesis (4.6).

(i) Select a block size m, such that $m \to \infty$, $m = o(n)$.

(ii) For a large integer B generate independent standard normal distributed random variables $\{N_i^h\}_{i=1}^{n-m}$, $h = 1, 2, \cdots, B$. For each h and pair (j, k), $1 \le j < k \le r$, calculate

$$\mathring{\Phi}_{m,j,k}^h(u) = \sqrt{\frac{m}{n-m}} \sum_{i=1}^{n-m} [\mathring{T}_{i,m,j,k}(u) - \mathring{S}_{n,j,k}(u)/\sqrt{n}] N_i^h / \hat{v}_{j,k}(u),$$

where $\mathring{T}_{i,m,j,k}(u) = \sum_{a=i}^{i+m} W_{a,j,k}(u)/m$, and define

$$\breve{\Phi}_m^h = \max_{1 \le j < k \le r} \sup_{0 \le u \le 1} |\mathring{\Phi}_{m,j,k}^h(u)|.$$

(iii) Let $\breve{\Phi}_m^{(1)} \le \cdots \le \breve{\Phi}_m^{(B)}$ be the ordered statistics of $\{\breve{\Phi}_m^j\}_{j=1}^B$ and estimate the quantile $\mathring{c}_{1-\alpha,n}$ of the distribution of T_n in (4.9) by $\breve{\Phi}_m^{(\lfloor (1-\alpha)B \rfloor)}$.

(iv) Reject the null hypothesis (4.6) of parallel mean functions, whenever

$$T_n > \breve{\Phi}_m^{(\lfloor (1-\alpha)B \rfloor)}. \tag{4.10}$$

A data driven rule for the choice of the block size in step (a) will be given in Section 4.3. The asymptotic validity of the above multiplier bootstrap procedure is established in the following two propositions.

Proposition 4.4. *Suppose that (3.13) and (3.14) hold and that $\theta_1 < \phi' q^2/[4(q+1)]$. Further assume that the conditions of Proposition 4.3 are satisfied. Then, under the null hypothesis (4.6) of parallel mean functions we have*

$$\lim_{n,r \to \infty} \lim_{B \to \infty} \mathbb{P}\big(T_n > \breve{\Phi}_m^{(\lfloor (1-\alpha)B \rfloor)}\big) = \alpha.$$

Proposition 4.4 follows directly from Theorem 3.2, Proposition 4.3 and Lemma 4.1. The details are omitted for the sake of brevity. Next, we study the power of the test considering the local alternative

$\mathbf{H}_a$: There exists a pair (j,k), $1 \le j < k \le r$, such that

$$\inf_{c\in\mathbb{R}} \sup_{u\in[0,1]} |g_j(u) - g_k(u) - c|\sqrt{n/\log n} \to \infty. \tag{4.11}$$

The following proposition shows that the test (4.10) achieves asymptotic power 1 under alternatives of the form (4.11). Hence this test is able to detect alternatives that deviate from the null hypothesis of parallel mean functions at the rate $\sqrt{\log n/n}$.

Proposition 4.5. *If the conditions of Proposition 4.4 are satisfied, we have under alternatives of the form (4.11) that*

$$\lim_{n,r\to\infty} \lim_{B\to\infty} \mathbb{P}\big(T_n > \breve{\Phi}_m^{(\lfloor(1-\alpha)B\rfloor)}\big) = 1.$$

4.3 Tuning parameter selection

To implement the multiplier bootstrap, one needs to choose the block size m. Observe that the quality of the bootstrap depends on how well the conditional covariance operator of $\mathbf{\Phi}_m$ in (3.12) approximates the covariance operator of $\mathbf{S}_n^N$ in (3.6). The tuning parameter m controls the accuracy for the latter approximation. When m is too large, the variance of the conditional covariance operator of $\mathbf{\Phi}_m$ will be too large; while if m is too small, then the latter conditional covariance operator will have too much bias. In this paper, we adopt the minimum volatility (MV) method proposed in Politis et al. (1999) to select the block size m. The idea behind the MV method is that the conditional covariance operator of $\mathbf{\Phi}_m$ should behave stably as a function of m when m is in an appropriate range. Therefore one could select m that minimizes the variability of the conditional covariance operator of $\mathbf{\Phi}_m$ as a function of m. We refer the readers to Politis et al. (1999) and Zhou (2013) for more detailed discussions of the MV method and its applications in time series analysis. Specifically, one first determines a sequence of equally-spaced candidate block sizes $m_1 < m_2 < \cdots < m_k$ and defines $m_0 = 2m_1 - m_2$ and $m_{k+1} = 2m_k - m_{k-1}$. For each $i \in \{0, \ldots, k+1\}$, one calculates

$$\Xi_j^{(i)}(u) := \sqrt{\frac{m_i}{n-m_i} \sum_{k=1}^{n-m_i} [T_{i,j,m} - S_{n,j}(u)/\sqrt{n}]^2}, \; j = 1, 2, \cdots, r,$$

where $T_{i,j,m}$ denotes the jth component of $\mathbf{T}_{i,m}$. Note that $\Xi_j^{(i)}(u)$ is an estimator of the marginal variance operator of $S_{n,j}^N(u)$. Next, one calculates for $i = 1, 2, \cdots, k$

$$\Xi^{(i)} := \sum_{j=1}^{r} \int_0^1 \mathrm{sd}(\{\Xi_j^{(l)}(u)\}_{l=i-1}^{i+1}) \, du,$$

where sd denotes standard deviation. Observe that $\Xi^{(i)}$ measures the overall variability of $\Xi_j^{(i)}(u)$ (with respect to i) at candidate block size m_i. We finally recommend selecting the block size m_i such that $i = \text{argmin}_{1\leq i\leq k}\,\Xi^{(i)}$. The MV method performs well in our simulation studies, which are presented in the following section.

5 Numerical experiments

In this section we illustrate the application of the methodology by means of a small simulation study and by the analysis of a real data example.

5.1 Simulation study

In the Monte Carlo simulations, we consider the following two types of simple panel time series models:

The **PAR(a)-model** is defined as the panel autoregressive functional model

$$\mathbf{X}_i(u) = \mathbf{g}(u) + A_r\mathbf{e}_i(u),$$

where A_r is an $r\times r$ tridiagonal matrix with 1's on the diagonal and 1/2's on the off-diagonal, $\mathbf{e}_i(u) = (e_{i,1}(u), \cdots, e_{i,r}(u))^\top$ is an r-dimensional vector with independent entries satisfying the AR-equation

$$e_{i,j}(u) = ae_{i-1,j}(u) + \epsilon_{i,j}(u)\ ,\ j = 1,\dots,r\ . \tag{5.1}$$

and $a \in [0, 1)$ denotes the auto-regressive (AR) coefficient that controls the strength of the temporal dependence. The innovations in (5.1) are given by i.i.d. random functions

$$\epsilon_{i,j}(u) = \sum_{k=1}^{\infty} k^{-3}[\cos(2\pi ku)\epsilon_{i,j,k} + \sin(2\pi ku)\epsilon_{i,j,k}],\ i = 1, 2, \dots,\ j = 1,\dots,r\ ,$$

where the infinite-dimensional vector $\boldsymbol{\epsilon}_{i,j} := (\epsilon_{i,j,1}, \epsilon_{i,j,2}, \cdots)^\top$ satisfies $\boldsymbol{\epsilon}_{i,j} = A_\infty\boldsymbol{\epsilon}'_{i,j}$, A_∞ is the infinite-dimensional tridiagonal matrix with 1's on the diagonal and 1/2's on the off-diagonal and

$$\boldsymbol{\epsilon}'_{i,j} := (\epsilon'_{i,j,1}, \epsilon'_{i,j,2}, \cdots)^\top \tag{5.2}$$

are i.i.d. infinite-dimensional random vectors with i.i.d. entries at each component.

The **PMA(a)-model** is defined as the panel moving average functional model. The setup is the same as that of PAR-model except that equation (5.1) is replaced by a moving average (MA) representation

$$e_{i,j}(u) = \epsilon_{i,j}(u) + a\epsilon_{i-1,j}(u),$$

where a is the MA-coefficient that controls the temporal dependence.

Observe that the functions $\mathbf{X}_i$ in the PAR(a)- and PMA(a)-model are twice but not three times differentiable and that the components $X_{i,j_1}(u)$ and $X_{i,j_2}(u)$ are independent if $|j_1 - j_2| > 1$. By construction, the Fourier coefficients $\epsilon_{i,j,k}$ as a sequence of k are correlated. In the simulation studies, the entries $\epsilon'_{i,j,k}$ in the vector (5.2) are either i.i.d. standard normal distributed random variables (denoted by $\mathcal{N}(0,1)$) or i.i.d. standardized t-distributed random variables with 6 degrees of freedom (denoted by $\sqrt{2/3}t_6$) to represent light tailed and heavy tailed cases, respectively (note that the standard deviation of a $\sqrt{2/3}t_6$-distributed random variable is 1). Throughout the simulations, the block size is selected by the MV method described in Section 4.3.

5.1.1 Coverage accuracy of JSCB

In order to investigate the approximation of the confidence level of the JSCB for finite sample sizes, we consider the function $\mathbf{g}(u) = 0$ in the PAR(a)- and PMA(a)-model. The sample size n is chosen as 200 and 400 and the dimension r varies from 5 to 80 in order to investigate the impact of the dimensionality on the coverage probability. The number of bootstrap replications is chosen as $B = 1000$ and for each scenario 1000 simulation runs are performed. The simulated coverage probabilities of the JSCB under both light and heavy tailed errors and various levels of time series dependence are reported in Tables 1 and 2.

We observe that the performances of the JSCB under light and heavy tailed errors are similar. When $n = 200$, the JSCB is reasonably accurate for all dimensions when the time series dependence is moderate. Under stronger time series dependence, such as for the PAR(0.5)-model, the coverage probabilities of the JSCB are slightly smaller than the nominal confidence level. However, the approximation improves significantly if the sample size increases to $n = 400$. In this case, the coverage of the JSCB is reasonably accurate in all cases under consideration. The impact of the dimension on the accuracy of the approximation of the confidence level is hardly visible, but we observe a slightly better performance of the JSCB for smaller dimensions $r = 5, 10$ compared to the cases $r = 20, 40, 80$.

$n = 200$										
	$1-\alpha = 0.95$					$1-\alpha = 0.9$				
r	PAR(0)	PAR(.2)	PAR(.5)	PMA(.5)	PMA(1)	PAR(0)	PAR(.2)	PAR(.5)	PMA(.5)	PMA(1)
5	.941	.944	.907	.939	.942	.893	.878	.830	.878	.875
10	.940	.934	.916	.934	.941	.891	.870	.824	.881	.878
20	.938	.943	.919	.945	.936	.884	.890	.817	.887	.869
40	.942	.942	.914	.943	.941	.885	.871	.818	.873	.873
80	.943	.941	.912	.944	.934	.902	.889	.808	.871	.867
$n = 400$										
	$1-\alpha = 0.95$					$1-\alpha = 0.9$				
r	PAR(0)	PAR(.2)	PAR(.5)	PMA(.5)	PMA(1)	PAR(0)	PAR(.2)	PAR(.5)	PMA(.5)	PMA(1)
5	.948	.947	.941	.945	.944	.892	.880	.887	.899	.893
10	.948	.957	.934	.939	.939	.898	.890	.862	.877	.872
20	.954	.953	.946	.943	.943	.897	.891	.878	.885	.891
40	.955	.944	.947	.949	.940	.905	.885	.873	.885	.888
80	.956	.946	.942	.955	.943	.903	.908	.857	.896	.880

Table 1: *Simulated coverage probabilities of the JSCB in the PAR- and PMA-model. The errors in (5.2) are* $\mathcal{N}(0,1)$*-distributed and* $1-\alpha$ *represents the nominal coverage probability.*

5.1.2 Accuracy and power of parallelism test

In order to investigate the finite sample accuracy of the test for parallelism, we consider the functions

$$g_i(u) = u^2 - u + c_i, \quad i = 1, 2, \cdots, r,$$

in the PAR(a) and PMA(a)-model, where c_i are i.i.d. standard normal distributed random variables. Observe that the mean functions $g_i(u)$, $1 \le i \le r$ are parallel. Various values for the dimension r from 5 to 30 and various values of the parameter a are selected to investigate the impact of the dimension and temporal dependence on the accuracy of the test, respectively. Since the parallelism test compares mean curves for every pair of the r component functional time series, the effective dimensionality of the test is $\binom{r}{2}$. Hence when $r = 30$, we are effectively testing a 435 dimensional function. Throughout this subsection, the number of bootstrap replications is chosen as $B = 1000$ and for each scenario, 500 simulations are used for the calculation of the rejection probabilities.

The simulated Type I error rates of the test for parallelism under light and heavy tailed

$n = 200$										
	$1-\alpha = 0.95$					$1-\alpha = 0.9$				
r	PAR(0)	PAR(.2)	PAR(.5)	PMA(.5)	PMA(1)	PAR(0)	PAR(.2)	PAR(.5)	PMA(.5)	PMA(1)
5	.941	.944	.919	.939	.937	.884	.887	.83	.873	.874
10	.942	.943	.918	.938	.943	.895	.880	.827	.873	.878
20	.942	.944	.911	.936	.945	.892	.880	.803	.876	.883
40	.934	.943	.916	.937	.944	.877	.888	.812	.877	.885
80	.944	.954	.925	.935	.940	.877	.871	.814	.872	.868
$n = 400$										
	$1-\alpha = 0.95$					$1-\alpha = 0.9$				
r	PAR(0)	PAR(.2)	PAR(.5)	PMA(.5)	PMA(1)	PAR(0)	PAR(.2)	PAR(.5)	PMA(.5)	PMA(1)
5	.952	.939	.939	.948	.940	.906	.889	.883	.885	.888
10	.947	.945	.936	.944	.943	.898	.899	.875	.883	.874
20	.955	.952	.923	.946	.950	.904	.886	.857	.881	.890
40	.946	.944	.941	.950	.942	.896	.887	.861	.887	.885
80	.947	.956	.929	.954	.945	.893	.901	.856	.902	.886

Table 2: *Simulated coverage probabilities of the JSCB in the PAR- and PMA-model. The errors in (5.2) are $\sqrt{2/3}t_6$-distributed and $1-\alpha$ represents the nominal coverage probability.*

errors are reported in Tables 3 and 4, respectively. We observe similar results as in Section 5.1.1. There are no significant differences in the approximation of the nominal level for light and heavy tails, and the test (4.10) is reasonably accurate when the time series dependence is not too strong. Moreover, the test becomes also reasonably accurate under stronger temporal dependence for larger sample size (here $n = 400$). The approximation of the nominal level by the test (4.10) is only slightly better for lower dimensions ($r = 5, 10$) compared with higher dimensions ($r = 20, 30$).
Next, we investigate the finite sample performance of the test (4.10) under the alternative. To this end, we let

$$g_1(u) = u^2 + (b-1)u + c_1 \quad \text{and} \quad g_i(u) = u^2 - u + c_i, \quad i = 2, 3, \cdots, r,$$

where c_i, $i = 1, 2, \cdots, r$ are i.i.d. standard normal random variables, $b \geq 0$ is a constant that controls the magnitude of deviation from the null hypothesis. In particular the choice $b = 0$ corresponds to parallel mean functions and larger values for b indicate a more substantial deviation from the null hypothesis. Further observe that there is only one

$n = 200$										
	$\alpha = 0.05$					$\alpha = 0.1$				
r	PAR(0)	PAR(.2)	PAR(.5)	PMA(.5)	PMA(1)	PAR(0)	PAR(.2)	PAR(.5)	PMA(.5)	PMA(1)
5	.042	.062	.076	.052	.066	.114	.130	.142	.102	.120
10	.058	.054	.078	.062	.060	.124	.118	.164	.110	.110
20	.060	.050	.080	.066	.058	.126	.122	.202	.132	.152
30	.066	.064	.094	.056	.066	.128	.140	.190	.134	.140
$n = 400$										
	$\alpha = 0.05$					$\alpha = 0.1$				
r	PAR(0)	PAR(.2)	PAR(.5)	PMA(.5)	PMA(1)	PAR(0)	PAR(.2)	PAR(.5)	PMA(.5)	PMA(1)
5	.052	.048	.056	.052	.052	.108	.100	.124	.106	.110
10	.048	.052	.066	.052	.056	.100	.122	.134	.116	.102
20	.058	.058	.068	.054	.048	.114	.110	.138	.120	.114
30	.050	.048	.072	.038	.060	.102	.110	.124	.106	.136

Table 3: *Simulated type I errors of the test* (4.10) *for parallelism of the mean functions in the PAR- and PMA-model. The errors in (5.2) are* $\mathcal{N}(0,1)$*-distributed and* α *represents the nominal level of the test.*

curve that is not parallel to the rest. For the simulations of the rejection probabilities under the alternative, we investigate a PAR(0.2) model with $n = 200$ observations. The simulated rejection rates as a function of b are plotted in Figure 1 for dimensions $r = 20$ and 30, respectively. We observe from Figure 1 that the simulated power of the test (4.10) increases very fast as b increases. In particular, the simulated power of the test becomes reasonably high when b is larger than 0.15. This observation is consistent with our theoretical finding that the parallelism test is able to detect alternatives converging to the null hypothesis with a rate $\sqrt{\log r/n}$. Additionally, we find that the simulated power for the dimension $r = 20$ is higher than for dimension $r = 30$. This observation is also consistent with the theoretical $\sqrt{\log r/n}$ rate for the detection of local alternatives.

5.2 Analysis of Canadian temperature data

In this section we analyze daily mean temperature records from 1902 to 2018 in 15 representative Canadian cities as an empirical illustration of our methodology. The cities chosen are Calgary, Charlottetown, Edmonton, Halifax, Hamilton, Kitchener, London,

$n=200$										
	$\alpha=0.05$					$\alpha=0.1$				
r	PAR(0)	PAR(.2)	PAR(.5)	PMA(.5)	PMA(1)	PAR(0)	PAR(.2)	PAR(.5)	PMA(.5)	PMA(1)
5	.050	.054	.070	.050	.058	.098	.118	.166	.116	.122
10	.050	.056	.086	.058	.054	.104	.124	.172	.128	.134
20	.048	.058	.088	.054	.066	.108	.108	.196	.116	.140
30	.064	.054	.102	.066	.042	.112	.118	.222	.134	.124
$n=400$										
	$\alpha=0.05$					$\alpha=0.1$				
r	PAR(0)	PAR(.2)	PAR(.5)	PMA(.5)	PMA(1)	PAR(0)	PAR(.2)	PAR(.5)	PMA(.5)	PMA(1)
5	.058	.058	.062	.052	.062	.112	.104	.124	.110	.112
10	.050	.046	.060	.054	.050	.100	.112	.130	.124	.124
20	.056	.058	.056	.064	.056	.106	.108	.126	.112	.128
30	.056	.048	.068	.064	.046	.108	.104	.133	.126	.116

Table 4: *Simulated type I errors of the test* (4.10) *for parallelism of the mean functions in the PAR- and PMA-model. The errors in (5.2) are* $\sqrt{2/3}t_6$*-distributed and* α *represents the nominal level of the test.*

Montreal, Ottawa, Quebec, Saskatoon, Toronto, Vancouver, Victoria, and Windsor. The data are publicly available from the government of Canada website: `https://climate.weather.gc.ca/historical_data/search_historic_data_e.html`. The daily temperature records of each year are smoothed using local linear kernel smoothing to represent the overall yearly functional pattern of the temperature records. As a result, we obtain a panel functional time series with temporal dimension $n = 117$ and spatial dimension $r = 15$. Figure 2 below shows 3D plots of the functional time series of mean temperature in Toronto from two different angles. It can be seen that the yearly mean temperature curves are of similar shape which depicts clear seasonal variation. Additionally, substantial temperature variability can be seen from the curves among different years.

Note that we expect spatial correlation among the panels as some selected cities are close in distance. We are interested in constructing JSCB of the overall yearly pattern of temperature for the 15 cities across Canada. Additionally, we are interested in identifying which cities are similar in yearly temperature patterns in the sense that their temperature curves are parallel to each other. As the cities chosen in the study are scattered across four climate zones: Atlantic Canada, Great Lakes/St. Lawrence Lowlands, Prairies, and

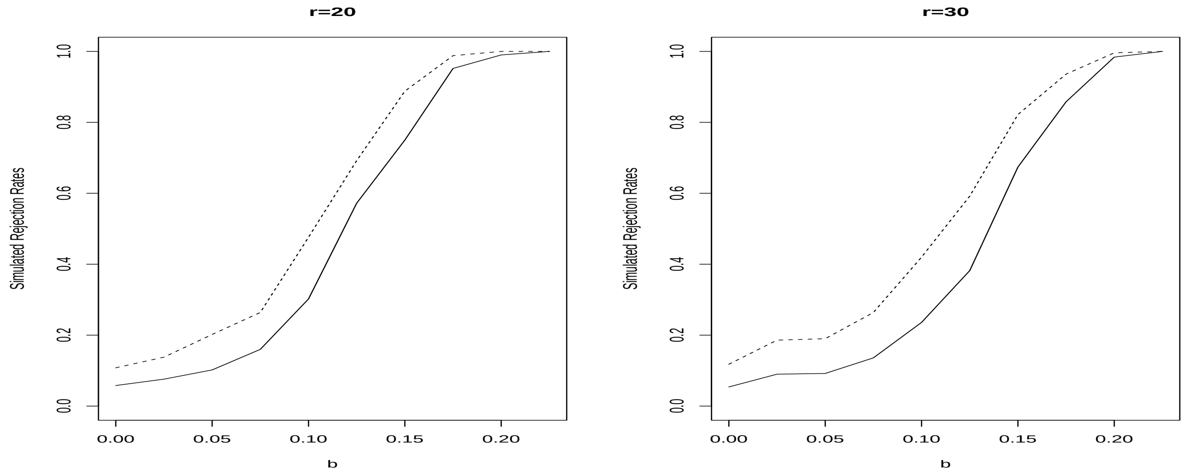

Figure 1: *Simulated rejection rates of the test* (4.10) *for the hypothesis of parallel mean functions with* $r = 20$ *and* 30. *Solid and dashed lines represent the rejection rates under nominal levels* 0.05 *and* 0.1, *respectively.*

Pacific Coast, we are also interested in checking whether the temperature patterns of the cities can be grouped according to their climate zones.

The 95% JSCB of the yearly temperature pattern of the 15 cities are plotted in Figure 3. This JSCB is based on 10,000 bootstrap replications with block size $m = 10$ which was selected by the MV method. We can see from the plots that all yearly temperature patterns reflect clear seasonal variation of winter lows and summer highs. A closer look at the plots shows that Calgary, Edmonton and Saskatoon have significantly wider bands in winter than the other cities. Therefore we conclude that the winter mean temperature in cities of the Prairies zone has more year-to-year variation than the other zones.

As all temperature curves are similar in shape to the sinusoidal functions, we would like to test the hypothesis

$$H_0^k : g_i(u) = c_i + \sum_{j=1}^{k}(a_{i,j}\cos(2\pi ju) + b_{i,j}\sin(2\pi ju)) \ , \ i = 1, 2, \cdots, 15.$$

Observe that one will encounter multiple testing problems if the cities are tested separately. Here we perform the tests by fitting the harmonic functions to the mean temperature curves and then checking whether all fitted functions can be fully embedded into the 95% JSCB. When $k = 1$, most fitted harmonic curves are not covered by the JSCB. When $k = 2$, the fitted harmonic curves are plotted in Figure 3. It can be seen that all 15 harmonic curves are fully covered by the 95% JSCB. These results suggest that one can use the

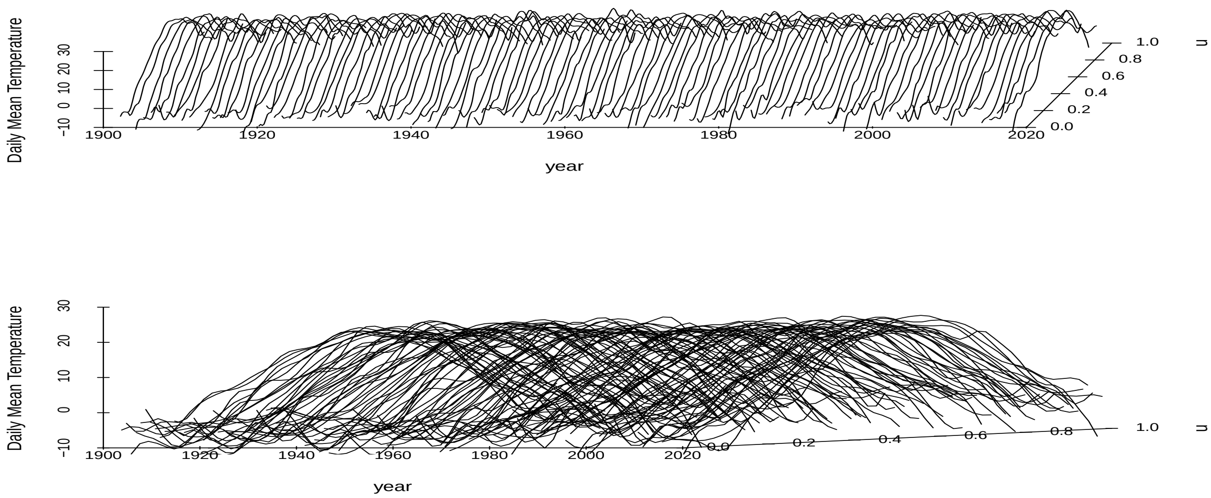

Figure 2: *3D functional time series plots from two different angles for the mean temperature curves of Toronto (1902-2018).*

parametric mean model $g_i(u) = c_i + \sum_{j=1}^{2}(a_{i,j}\cos(2\pi j u) + b_{i,j}\sin(2\pi j u))$, $i = 1, 2, \cdots, 15$ for the Canadian temperature data.

For the test of parallelism, we follow the implementation steps in Section 4.2 with block sizes $m = 10$ and $B = 10,000$ bootstrap replications. Observe that there are 105 pairs to be compared and that the sample size is $n = 117$. Therefore the dimensionality here is almost as large as the sample size. The resulted pairwise test statistics range from 3.0 to 93 and the critical values of the test at 0.1%, 0.5%, 1%, 5% and 10% are 7.9, 7.04, 6.61, 5.50 and 5.04, respectively. In Table 5 we display the pairwise p-values of the parallelism test obtained by the critical values listed above. Note that those pair-wise p-values do not suffer from multiple testing problems as (4.10) is a joint test of parallelism. Clearly, the overall test of parallelism is rejected with a very strong evidence. A closer look at Table 5 shows that Hamilton, London, Kitchener, Toronto and Windsor can be grouped together in term of temperature parallelism. And another group of cities with parallel temperature patterns is Quebec, Ottawa and Montreal. The first group of cities are all in the area of southeast Ontario surrounded by the great lakes. The second group of cities are located near the St. Lawrence River in the St. Lawrence Lowlands. On the other hand, we also observe at some close cities significantly different temperature patterns. Two such

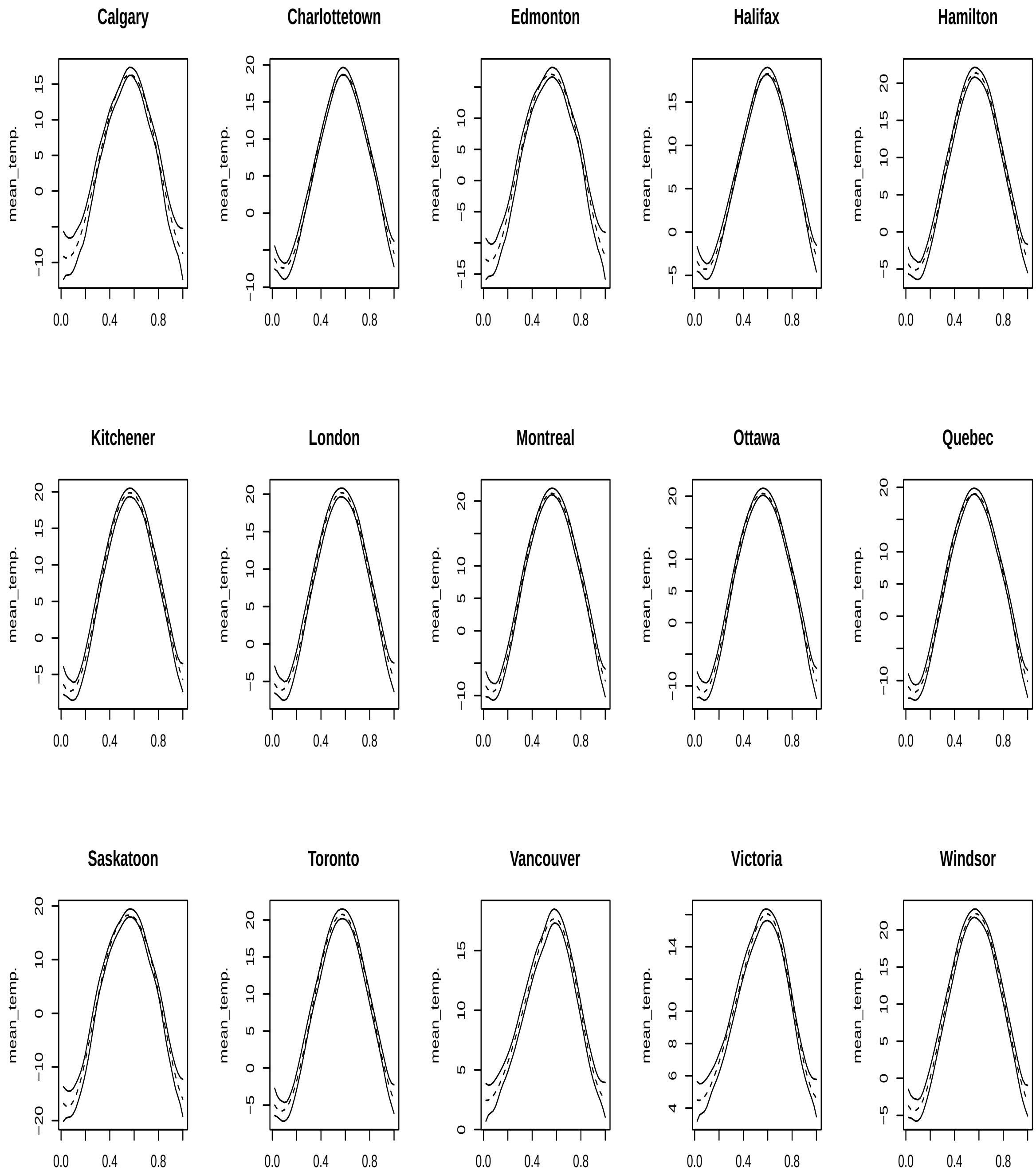

Figure 3: *95% JSCB for the Canadian temperature data. The solid lines represent the upper and lower bounds of the JSCB in (4.1) with the functions defined in (4.5). The dotted lines in the middle represent the fitted harmonic curves* $g_i(u) = c_i + \sum_{j=1}^{2}(a_{i,j}\cos(2\pi ju) + b_{i,j}\sin(2\pi ju))$, $i = 1, 2, \cdots, 15$.

	Ham	Lon	Kit	Tor	Win	Que	Ott	Mon	Cal	Edm	Cha	Hal	Vic	Van	Sas
Ham															
Lon	0.2														
Kit	< 0.1	0.3													
Tor	> 10	3.6	1.8												
Win	> 10	> 10	> 10	> 10											
Que	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1										
Ott	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	5									
Mon	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	> 10	< 0.1								
Cal	0.6	< 0.1	0.3	0.1	0.4	< 0.1	< 0.1	< 0.1							
Edm	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	0.5	< 0.1	< 0.1						
Cha	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1					
Hal	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1				
Vic	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1			
Van	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1		
Sas	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	

Table 5: *Table of pairwise parallelism test p-values (in percentage) for the Canadian temperature data.*

interesting pairs are Vancouver and Victoria and Calgary and Edmonton.

In order to carry out a further investigation, the 95% JSCB for the mean temperature differences of the two pairs of cities are plotted in Figure 4. Vancouver and Victoria are both on the southern Pacific coast of Canada and are within 100 kilometers in distance. However, Figure 4 shows that Victoria is 1-2 degrees Celsius warmer than Vancouver in the winter and 1.5 to 2.5 degrees Celsius cooler than Vancouver in the summer. Therefore the two cities are significantly different in yearly temperature pattern. This may be due to the fact that Victoria is located in the Vancouver Island and the island adjusts the temperature significantly. Calgary and Edmonton are major cities of Alberta. Both cities are in the Prairies with Edmonton about 300 kilometers north of Calgary. Figure 4 shows that Edmonton is significantly colder than Calgary in the winter but significantly warmer in the summer. Interestingly, we observe from Table 5 that the temperature pattern of Edmonton has some similarities (highest p-value around 0.5%) to that of the St. Lawrence Lowland group while the temperature pattern of Calgary has some similarities (highest p-value around 0.5%) to that of the great lakes group. It is unknown why this is the case and further investigations with geography and climatology knowledge are needed.

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) and NSERC of Canada (fund number : 489079). We gratefully acknowledge Professors Alois Kneip, Dominik Liebl and Hans Georg Müller for some helpful discussions on related literature.

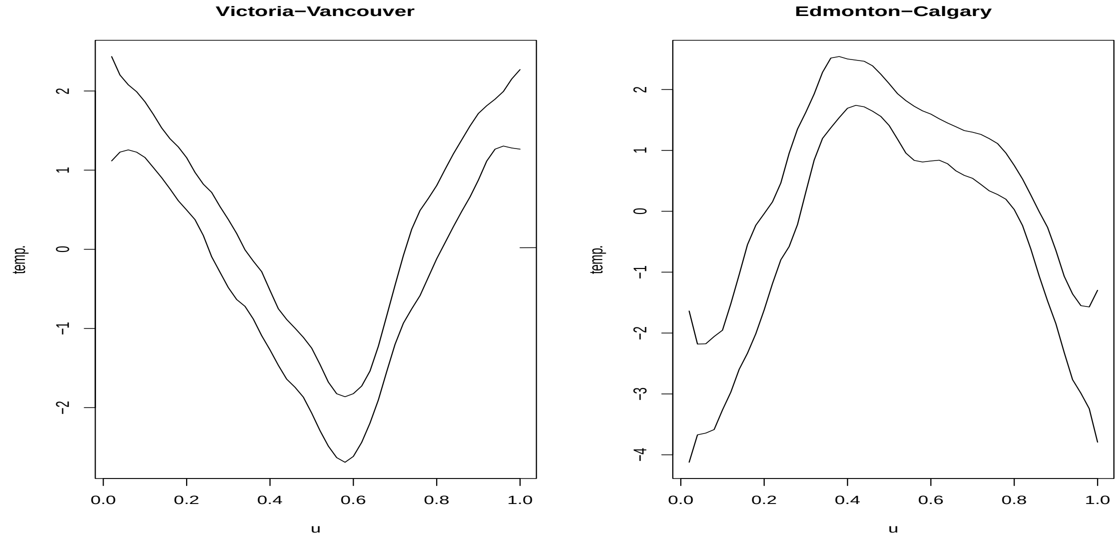

Figure 4: *95% JSCB for the mean yearly temperature differences between Victoria and Vancouver (left panel) and Edmonton and Calgary (right panel).*

References

Baltagi, B. (2005). *Econometric Analysis of Panel Data (3rd ed.).* Wiley, New York.

Banerjee, S. (2017). High-dimensional bayesian geostatistics. *Bayesian Anal.*, 12(2):583–614.

Belloni, A., Chernozhukov, V., Hansen, C., and Kozbur, D. (2016). Inference in high-dimensional panel models with an application to gun control. *Journal of Business & Economic Statistics*, 34(4):590–605.

Bosq, D. (2000). *Linear Processes in Function Spaces.* Springer-Verlag, New York.

Bosq, D. (2002). Estimation of mean and covariance operator of autoregressive processes in Banach spaces. *Statistical Inference for Stochastic Processes.*, 5(3):287–306.

Bradley, R. C. (2007). *Introduction to Strong Mixing Conditions. Vols. 1, 2, 3.* Kendrick Press, Heber City, UT.

Chen, K., Delicado, P., and Müller, H.-G. (2017). Modelling function-valued stochastic processes, with applications to fertility dynamics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(1):177–196.

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):209–226.

Cressie, N. A. C. and Wikle, C. K. (2015). *Statistics for Spatio-Temporal Data.* John Wiley & Sons.

Delicado, P., Giraldo, R., Comas, C., and Mateu, J. (2010). Statistics for spatial functional data: some recent contributions. *Environmetrics*, 21:224–239.

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

Galvao, A. F. and Montes-Rojas, G. V. (2010). Penalized quantile regression for dynamic panel data. *Journal of Statistical Planning and Inference*, 140(11):3476 – 3497.

Gao, Y., Shang, H. L., and Yang, Y. (2019). High-dimensional functional time series forecasting: An application to age-specific mortality rates. *Journal of Multivariate Analysis*, 170:232 – 243. Special Issue on Functional Data Analysis and Related Topics.

Greven, S., Crainiceanu, C., Caffo, B., and Reich, D. (2010). Longitudinal functional principal component analysis. *Electron. J. Statist.*, 4:1022–1054.

Gromenko, O., Kokoszka, P., and Reimherr, M. (2017). Detection of change in the spatiotemporal mean function. *Journal of the Royal Statistical Society, Ser. B*, 79:29–50.

Haining, R. P. and Li, G. (2020). *Modelling Spatial and Spatial-Temporal Data: A Bayesian Approach.* Chapman & Hall/CRC, Boca Raton.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *Annals of Statistics.*, 38(3):1845–1884.

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

Hsiao, C. (2003). *Analysis of Panel Data (Second ed.).* Cambridge University Press, New York.

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

Kock, A. B. and Tang, H. (2019). Uniform inference in high-dimensional dynamic panel data models with approximately sparse fixed effects. *Econometric Theory*, 35(2):295–359.

Kokoszka, P. and Reimherr, M. (2019). Some recent developments in inference for geostatistical functional data. *Revista Colombiana de Estadística*, 42:101–122.

Nisol, G., Tavakoli, S., and Hallin, M. (2019). High-dimensional functional factor models. *arXiv:1905.10325.*

Panaretos, V. M. and Tavakoli, S. (2013). Fourier analysis of stationary time series in function space. *Ann. Statist.*, 41(2):568–603.

Park, S. Y. and Staicu, A.-M. (2015). Longitudinal functional data analysis. *Stat*, 4(1):212–226.

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

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

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data (2nd ed.)*. MIT Press.

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

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

Supplemental Material for "Statistical Inference for High Dimensional Panel Functional Time Series"

Zhou Zhou and Holger Dette
University of Toronto and Ruhr-Universität Bochum

February 27, 2020

Abstract

This supplemental material contains proofs of the theoretical results of the paper.

In this supplemental material, the symbol C denotes a positive finite constant which may vary from place to place.

Proof of the results in Section 2

Proof of Proposition 2.1. Using integration by parts, we have for the Fourier coefficients in (2.5) and $k \geq 1$,

$$a_{i,k} = \frac{2}{(k\pi)^2}\Big(- \int_0^1 X_i''(u)\cos(k\pi u)\,du + X_i'(1)\cos(k\pi) - X_i'(0)\Big).$$

Note that $\|X_i'(1)\|_q \leq \|X_i'(0)\|_q + \int_0^1 \|X_i''(u)\|_q\,du < \infty$. As a result $\|a_{i,k}\|_q \leq C/k^2$. Hence $\|a^*_{i,k}\|_q \leq C/k^2$ as $a_{i,k}$ and $a^*_{i,k}$ are identically distributed. On the other hand,

$$\begin{aligned} \|a_{i,k} - a^*_{i,k}\|_q &\leq \int_0^1 \|H(u,\mathcal{F}_i) - H(u,\mathcal{F}_i^*)\|_q |\cos(k\pi u)|\,du \\ &\leq \sup_{u\in[0,1]} \|H(u,\mathcal{F}_i) - H(u,\mathcal{F}_i^*)\|_q = \delta_X(i,q), \end{aligned}$$

and the proposition follows. ◇

Proof of Proposition 2.2. Let $A_{n,k} = \sum_{i=1}^n a_{i,k}/\sqrt{n}$. Observe that $\mathbb{E}(a_{j,k}) = 0$. For any $j \in \mathbb{Z}$ and a random variable Z, let

$$\mathcal{P}_j(Z) = \mathbb{E}(Z|\mathcal{F}_j) - \mathbb{E}(Z|\mathcal{F}_{j-1}).$$

We have

$$\|A_{n,j}\|_q \leq C\sum_{l=0}^{\infty} \|\mathcal{P}_0(a_{l,j})\|_q$$

$$\leq \; C\sum_{l=0}^{\infty}\|a_{l,j}-a^*_{l,j}\|_q \leq C\Big[\sum_{l=0}^{\lfloor j^{2/\beta}\rfloor} j^{-2} + \sum_{l=\lfloor j^{2/\beta}\rfloor+1}^{\infty} l^{-\beta}\Big] \leq Cj^{2(1-\beta)/\beta}$$

by Proposition 2.1, Theorem 1 (i), (ii) of Wu (2005) and Theorem 1(iii) of Wu (2007), where we utilized the facts that $q \geq 2$ and $\sum_{l=0}^{\infty}\|a_{l,j}-a^*_{l,j}\|_q < \infty$. Observing the representation

$$S_n(u) - S_n(k,u) = \sum_{j=k+1}^{\infty} A_{n,j}\cos(j\pi u)$$

we obtain

$$\Big\|\sup_{u\in[0,1]}|S_n(u)-S_n(k,u)|\Big\|_q \leq \sum_{j=k+1}^{\infty}\|A_{n,j}\|_q \leq C\sum_{j=k+1}^{\infty} j^{2(1-\beta)/\beta} \leq Ck^{(2-\beta)/\beta},$$

and the assertion of the proposition follows. ♢

Proof of Proposition 2.3. Observe that $S_n(k,u) - S_n(k,t_{i,n}) = \int_{t_{i,n}}^{u} S'_n(k,s)\,ds$, where $S'_n(k,s)$ is the derivative of $S_n(k,s)$ with respect to s. Therefore

$$\sup_{u\in[t_{i,n},t_{i+1,n}]}|S_n(k,u)-S_n(k,t_{i,n})| \leq \int_{t_{i,n}}^{t_{i+1,n}}|S'_n(k,s)|\,ds,$$

which implies

$$\Big\|\sup_{u\in[t_{i,n},t_{i+1,n}]}|S_n(k,u)-S_n(k,t_{i,n})|\Big\|_q \leq \int_{t_{i,n}}^{t_{i+1,n}}\|S'_n(k,s)\|_q\,ds. \quad (1)$$

Now for any $s \in [0,1]$, $-S'_n(k,s) = \pi\sum_{j=1}^{k} A_{n,j}j\sin(j\pi s)$. By the same calculations as those in the proof of Proposition 2.2 we have

$$\|S'_n(k,s)\|_q \leq \sum_{j=1}^{k} j\|A_{n,j}\|_q \leq C\sum_{j=1}^{k} j\cdot j^{2(1-\beta)/\beta} \leq Ck^{2/\beta}. \quad (2)$$

Combining (1) and (2), we obtain that

$$\Big\|\sup_{u\in[t_{i,n},t_{i+1,n}]}|S_n(k,u)-S_n(k,t_{i,n})|\Big\|_q \leq Ck^{2/\beta}/l_n$$

for any i. Hence

$$\Big\|\max_{0\leq i\leq l_n-1}\sup_{u\in[t_{i,n},t_{i+1,n}]}|S_n(k,u)-S_n(k,t_{i,n})|\Big\|_q \leq Ck^{2/\beta}/l_n^{1-1/q}$$

by an L^q maximum inequality. ♢

Proof of the results in Section 3

Recall the definition (3.3) and define

$$\mathbf{S}_n(k,u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{X}_i(k,u),$$

where

$$\mathbf{X}_i(k,u) = \sum_{j=1}^{k} \mathbf{a}_{i,j} \cos(j\pi u)$$

and $\mathbf{a}_{i,j} = 2\int_0^1 \mathbf{X}_i(u)\cos(j\pi u)\,du$ for $j \geq 1$; $\mathbf{a}_{i,0} = \int_0^1 \mathbf{X}_i(u)\,du$. Let

$$\mathbf{S}_n^N(k,u) = \sum_{j=0}^{k} \mathbf{N}_{n,j} \cos(j\pi u) \tag{3}$$

and denote the j-th entry of $\mathbf{S}_n(k,u)$ (resp. $\mathbf{S}_i^N(k,u)$) by $S_{n,j}(k,u)$ (resp. $S_{n,j}^N(k,u)$).

Lemma 0.1. *Denote the dependence measure of* $\{X_j(k,\cdot)\}_{j\in\mathbb{Z}}$ *by* $\delta_{H,k}(i,q)$. *If* $\delta_H(i,q) = O(i^{-\beta})$, *then*

$$\delta_{H,k}(i,q) = O(i^{-\beta/2+\epsilon}) \quad \textit{for any } \epsilon > 0.$$

Proof. By definition, we have

$$\delta_{H,k}(i,q) = \Big\| \sum_{j=0}^{k} (a_{i,j} - a_{i,j}^*) \cos(j\pi u) \Big\|_q,$$

and Proposition 2.1 implies that $\|a_{i,j} - a_{i,j}^*\|_q \leq C\min(j^{-2}, i^{-\beta}) \leq Cj^{-2(1/2+\alpha)} i^{-\beta(1/2-\alpha)}$ for an arbitrarily small $\alpha > 0$. We obtain

$$\delta_{H,k}(i,q) \leq \sum_{j=0}^{k} \|a_{i,j} - a_{i,j}^*\|_q \leq \sum_{j=0}^{k} Cj^{-2(1/2+\alpha)} i^{-\beta(1/2-\alpha)} \leq Ci^{-\beta/2+\alpha\beta}.$$

As α can be made arbitrarily small, the lemma follows. ♢

Lemma 0.2. *Under the assumptions of Theorem 3.1, we have*

$$\sup_{x\in\mathbb{R}} \Big| \mathbb{P}\Big[\max_{1\leq j\leq r} \max_{1\leq i\leq l_n} |S_{n,j}(k,t_{i,n})| \leq x \Big] - \mathbb{P}\Big[\max_{1\leq j\leq r} \max_{1\leq i\leq l_n} |S_{n,j}^N(k,t_{i,n})| \leq x \Big] \Big| \to 0.$$

Proof. Let

$$\tilde{\mathbf{X}}_{i,k} = (\mathbf{X}_i^\top(k, t_{1,n}), \cdots, \mathbf{X}_i^\top(k, t_{l_n,n}))^\top \tag{4}$$

denote the vector of length rl_n in which the l-th block contains the coordinates of the vector $\mathbf{X}_i(k, t_{l,n})$ and define

$$\tilde{\mathbf{S}}_{k,n} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \tilde{\mathbf{X}}_{i,k}$$

as the corresponding standardized sum.
Finally, recall the notation of $\mathbf{S}_n^N(k, \cdot)$ in (3), define

$$\tilde{\mathbf{S}}_{k,n}^N = \left([\mathbf{S}_n^N(k, t_{1,n})]^\top, \cdots, [\mathbf{S}_n^N(k, t_{l_n,n})]^\top\right)^\top$$

and note that $\tilde{\mathbf{S}}_{k,n}$ and $\tilde{\mathbf{S}}_{k,n}^N$ share the same covariance structure. By Proposition 2.2, we have $\|S_{n,j}(u) - S_{n,j}(k,u)\|_2 \to 0$ uniformly in j and u if $k \to \infty$. Observing assumption (3.8) we conclude that, uniformly in j and u, $\mathbb{E}(S_{n,j}^2(k,u)) \geq \delta/2$ for sufficiently large n. Now pick $l_n \asymp n^{\theta_2}$ and $k \asymp n^{\theta_3}$ with positive constants θ_2 and θ_3. By assumption (3.9), we can choose θ_2 and θ_3 such that

$$\begin{cases} \theta_1 + \theta_2 < q/2 - 1, \\ \theta_1/q + [(2-\beta)/\beta]\theta_3 < 0, \\ \theta_1/q + (1/q - 1)\theta_2 + (2/\beta)\theta_3 < 0. \end{cases}$$

Let $\Delta = \sum_{j=-\infty}^{\infty} \mathbb{E}(\tilde{\mathbf{X}}_{i,k}\tilde{\mathbf{X}}_{i-j,k}^\top)$ denote the long-run-covariance of the time series $\{\tilde{\mathbf{X}}_{i,k}\}_{i\in\mathbb{Z}}$ defined by (4). Then as $\theta_1 + \theta_2 < q/2 - 1$ and $\max_{1\leq k\leq r} \delta_{H_j,k}(i,q) = O(i^{-\beta/2+\epsilon})$ for an arbitrarily small positive ϵ (see Lemma 0.1), we can easily check that the assumptions of part (i) of Theorem 3.2 of Zhang and Wu (2017) hold (define α therein as $\beta/2 - 1 - \epsilon$ and note that $\alpha > 1/2 - 1/q$ since $\beta > 3$). To provide more details, we denote by $X_{\cdot j}$ the j-th component process of $\tilde{\mathbf{X}}_{i,k}$, $j = 1, 2, \cdots, rl_n$, and check the magnitudes of the quantities $\delta_{i,q,j}$ $\Delta_{m,q,j}$, $\|X_{\cdot j}\|_{q,\alpha}$, $\Psi_{q,\alpha}$, $\Upsilon_{q,\alpha}$, $\omega_{i,q}$, $\Omega_{m,q}$, $\||X_\cdot|_\infty\|_{q,\alpha}$, $\Theta_{q,\alpha}$, L_1, L_2, N_1, N_2, W_1 and W_2 defined in Zhang and Wu (2017). By Lemma 0.1, $\delta_{i,q,j} = O(i^{-\beta/2+\epsilon})$ for any j which implies $\Delta_{m,q,j} = O(m^{-\beta/2+1+\epsilon})$. Elementary but tedious calculations using the assumptions of Theorem 3.1 yield that

$$\|X_{\cdot j}\|_{q,\alpha} = O(1)\ , \quad \Psi_{q,\alpha} = O(1)\ , \quad \Upsilon_{q,\alpha} = O(p^{1/q})$$

(note that here $p = O(n^{\theta_1+\theta_2})$),

$$\begin{aligned} \omega_{i,q} &= O(p^{1/q} i^{-\beta/2+\epsilon})\ ,\ \Omega_{m,q} = O(p^{1/q} m^{-\beta/2+1+\epsilon})\ ,\ \Theta_{q,\alpha} = O(p^{1/q})\ , \\ L_1 &= o(1)\ ,\ L_2 = O(\log^{2/\alpha} p)\ ,\ \||X_\cdot|_\infty\|_{q,\alpha} = O(p^{1/q})\ , \\ W_1 &= O(\log^7(pn))\ ,\ W_2 = O(\log^4(pn)), \end{aligned}$$

and $N_1 \geq c(n/\log p)^{q/2}/p$, $N_2 \geq cn/\log^2 p$, where c is a positive constant. Therefore the assumptions of part (i) of Theorem 3.2 of Zhang and Wu (2017) hold provided that $\theta_1 + \theta_2 < q/2 - 1$ and ϵ is sufficiently small. Thus we obtain

$$\sup_{x\in\mathbb{R}} \Big| \mathbb{P}[|\tilde{\mathbf{S}}_{k,n}|_\infty \leq x] - \mathbb{P}[|\Delta^{1/2}\mathbf{G}|_\infty \leq x] \Big| \to 0, \tag{5}$$

where $|\cdot|_\infty$ denotes the maximum norm on $\mathbb{R}^{rl_n}$ and $\mathbf{G}$ is a standard rl_n dimensional Gaussian vector. In fact, Theorem 3.2 in Zhang and Wu (2017) performs a normalization and it requires that the long-run variance of each $S_{n,j}(k, t_{i,n})$ equals 1, $j = 1, 2, \cdots, r$, $i = 1, 2, \cdots, l_n$. But their arguments easily apply to the case of arbitrary marginal long-run variances with a positive lower bound. Now observe that

$$\Delta_n := \mathbb{E}([\tilde{\mathbf{S}}_{k,n}\tilde{\mathbf{S}}_{k,n}^\top]) = \sum_{j=-n}^{n} (1 - |j|/n)\mathbb{E}(\tilde{\mathbf{X}}_{i,k}\tilde{\mathbf{X}}_{i-j,k}^\top),$$

and that Lemma 6 of Zhou (2014) and Lemma 0.1 yield

$$|\mathbb{E}[X_{i,p,k}X_{i-j,s,k}]| \leq Cj^{-\beta/2+\epsilon}, \quad 1 \leq p \leq rl_n, 1 \leq s \leq rl_n$$

for any $\epsilon > 0$, where $X_{i,p,k}$ is the p-th entry of $\tilde{\mathbf{X}}_{i,k}$. Hence simple calculations yield for the entries of the matrices Δ_n and Δ that

$$\max_{1\leq i,j\leq lr_n} |\Delta_n(i,j) - \Delta(i,j)| = O(n^{-\beta/2+\epsilon+1}).$$

By the Gaussian comparison inequality in Lemma 3.1 of Chernozhukov et al. (2013), we have

$$\sup_{x\in\mathbb{R}} \Big| \mathbb{P}[|\tilde{\mathbf{S}}_{k,n}^N|_\infty \leq x] - \mathbb{P}[|\Delta_n^{1/2}\mathbf{G}|_\infty \leq x] \Big| \leq C(n^{(-\beta/2+\epsilon+1)/3}[\log(rl_n/n^{-\beta/2+\epsilon+1}]^{2/3} \to 0,$$

which implies (observing (5))

$$\sup_{x\in\mathbb{R}} |\mathbb{P}[|\tilde{\mathbf{S}}_{k,n}|_\infty \leq x] - \mathbb{P}[\tilde{\mathbf{S}}_{k,n}^N|_\infty \leq x]| \to 0.$$

And this is exactly the claim of the lemma. $\diamond$

Proof of Theorem 3.1. Proposition 2.2 and an application of an L^q maximal inequality yield

$$\Big\| \max_{1\leq j\leq r} \sup_{0\leq u\leq 1} |S_{n,j}(u) - S_{n,j}(k,u)| \Big\|_q \leq Cr^{1/q}k^{(2-\beta)/\beta}.$$

Since $\theta_1/q + [(2-\beta)/\beta]\theta_3 < 0$, the right hand side of the above inequality converges to 0 faster than $n^{-\alpha_1}$ for some positive α_1. By the triangular and Markov inequalities, for $\delta_{1,n} := \sqrt{r^{1/q}k^{(2-\beta)/\beta}}$, we have

$$\begin{aligned}
\mathbb{P}\Big(\max_{1\le j\le r}\sup_{0\le u\le 1}|S_{n,j}(u)| \le x\Big) &\le \mathbb{P}\Big(\max_{1\le j\le r}\sup_{0\le u\le 1}|S_{n,j}(k,u)| \le x+\delta_{1,n}\Big)\\
&+ \mathbb{P}\Big(\max_{1\le j\le r}\sup_{0\le u\le 1}|S_{n,j}(u)-S_{n,j}(k,u)| \ge \delta_{1,n}\Big)\\
&\le \mathbb{P}\Big(\max_{1\le j\le r}\sup_{0\le u\le 1}|S_{n,j}(k,u)| \le x+\delta_{1,n}\Big) + \epsilon_{1,n},
\end{aligned}$$

where $\epsilon_{1,n} = Crk^{q(2-\beta)/\beta}/\delta_{1,n}^q \to 0$ with some polynomial rate. Similarly, since $\theta_1/q + (1/q-1)\theta_2 + (2/\beta)\theta_3 < 0$, it follows that $r^{1/q}k^{2/\beta}/l_n^{1-1/q}$ converges to 0 faster than $n^{-\alpha_2}$ for some positive α_2. By Proposition 2.3, we have, for $\delta_{2,n} := \sqrt{r^{1/q}k^{2/\beta}/l_n^{1-1/q}}$,

$$\mathbb{P}\Big(\max_{1\le j\le r}\sup_{0\le u\le 1}|S_{n,j}(k,u)| \le x+\delta_{1,n}\Big) \le \mathbb{P}\Big(\max_{1\le j\le r}\sup_{1\le i\le l_n}|S_{n,j}(k,t_{i,n})| \le x+\delta_{1,n}+\delta_{2,n}\Big) + \epsilon_{2,n},$$

where $\epsilon_{2,n} = C\{\delta_{2,n}/[r^{1/q}k^{2/\beta}/l_n^{1-1/q}]\}^{-q} \to 0$ with some polynomial rate. Now by Lemma 0.2, we obtain

$$\mathbb{P}\Big(\max_{1\le j\le r}\sup_{1\le i\le l_n}|S_{n,j}(k,t_{i,n})| \le x+\delta_{1,n}+\delta_{2,n}\Big) - \mathbb{P}\Big(\max_{1\le j\le r}\sup_{1\le i\le l_n}|S^N_{n,j}(k,t_{i,n})| \le x+\delta_{1,n}+\delta_{2,n}) \to 0$$

uniformly in x. Denote the latter uniform convergence rate by $\epsilon_{3,n}$. Furthermore, by Nazarov's anti-concentration inequality (Nazarov (2003)),

$$\begin{aligned}
\mathbb{P}\Big(\max_{1\le j\le r}\sup_{1\le i\le l_n}|S^N_{n,j}(k,t_{i,n})| \le x+\delta_{1,n}+\delta_{2,n}\Big) &- \mathbb{P}\Big(\max_{1\le j\le r}\sup_{1\le i\le l_n}|S^N_{n,j}(k,t_{1,n})| \le x\Big)\\
&\le C(\delta_{1,n}+\delta_{2,n})\sqrt{\log(rl_n)}.
\end{aligned}$$

Following the same arguments as above, we have

$$\begin{aligned}
\mathbb{P}\Big(\max_{1\le j\le r}\sup_{1\le i\le l_n}|S^N_{n,j}(k,t_{i,n})| \le x\Big) &- \mathbb{P}\Big(\max_{1\le j\le r}\sup_{0\le u\le 1}|S^N_{n,j}(u)| \le x\Big)\\
&\le \epsilon_{1,n}+\epsilon_{2,n}+C(\delta_{1,n}+\delta_{2,n})\sqrt{\log(rl_n)}.
\end{aligned}$$

Therefore we conclude that

$$\begin{aligned}
\mathbb{P}\Big[\max_{1\le j\le r}\sup_{0\le u\le 1}|S_{n,j}(u)| \le x\Big] &- \mathbb{P}\Big[\max_{1\le j\le r}\sup_{0\le u\le 1}|S^N_{n,j}(u)| \le x\Big]\\
&\le C[\epsilon_{1,n}+\epsilon_{2,n}+\epsilon_{3,n}+(\delta_{1,n}+\delta_{2,n})\sqrt{\log(rl_n)}] \to 0
\end{aligned}$$

uniformly in x. Similarly, we also obtain the lower bound

$$\begin{aligned}\mathbb{P}\Big[\max_{1\le j\le r}\sup_{0\le u\le 1}|S_{n,j}(u)|\le x\Big] \ &- \ \mathbb{P}\Big[\max_{1\le j\le r}\sup_{0\le u\le 1}|S^N_{n,j}(u)|\le x\Big]\\ &\ge \ -C[\epsilon_{1,n}+\epsilon_{2,n}+\epsilon_{3,n}+(\delta_{1,n}+\delta_{2,n})\sqrt{\log(rl_n)}]\to 0\end{aligned}$$

uniformly in x. Hence we conclude that Theorem 3.1 holds. $\diamondsuit$

Recall the definition of the vector $\mathbf{T}_{i,m}(u)$ in (3.11) and denote its j-th entry by $T_{i,j,m}(u)$. Similarly define $\Phi_{m,j}(u)$ as the j-th component of the vector $\mathbf{\Phi}_m$ in (3.12).

Lemma 0.3. *Under the assumptions of Theorem 3.2, there exists a sequence of events $E_{1,n}$ such that $\mathbb{P}(E_{1,n})\to 1$. On $E_{1,n}$, we have*

$$\sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big(\max_{1\le j\le r}\sup_{1\le i\le l_n}|\Phi_{m,j}(t_{i,n})|\le x\Big|\{\mathbf{X}_i(u)\}_{i=1}^n\Big)-\mathbb{P}\Big(\max_{1\le j\le r}\sup_{1\le i\le l_n}|S^N_{n,j}(t_{i,n})|\le x\Big)\Big|\to 0.$$

Proof. By Theorem 4 of Zhou (2013), we have, for each $t_{l,n}$, $1\le l\le l_n$ and j, $1\le j\le r$,

$$\Big\|\sum_{i=1}^{n-m}\frac{m}{n-m}\big[T_{i,j,m}(t_{l,n})-S_{n,j}(t_{l,n})/\sqrt{n}\big]^2-\tilde{\Sigma}_{(l-1)r+j,(l-1)r+j}\Big\|_{q'}\le C(\sqrt{m/n}+1/m),$$

where $q'=q/2$, $\tilde{\Sigma}_{i,j}$ denotes the (i,j)-th entry of the matrix $\tilde{\Sigma}$

$$\tilde{\Sigma}=\sum_{j=-\infty}^{\infty}\mathbb{E}[\bar{\mathbf{X}}_i-\mathbb{E}\bar{\mathbf{X}}_i][\bar{\mathbf{X}}_{i-j}^\top-\mathbb{E}\bar{\mathbf{X}}_{i-j}^\top],$$

and $\bar{\mathbf{X}}_i=(\mathbf{X}_i^\top(t_{1,n}),\cdots,\mathbf{X}_i^\top(t_{l_n,n}))^\top$.

Let $\tilde{\Sigma}_n=\mathbb{E}(\mathbf{S}_n^N[\mathbf{S}_n^N]^\top)$, where $\mathbf{S}_n^N=([\mathbf{S}_n^N(t_{1,n})]^\top,\cdots,[\mathbf{S}_n^N(t_{l_n,n})]^\top)^\top$. Observe that for each fixed l and j, the k-th dependence measure of the sequence $\{X_{i,j}(t_{i,n})\}_{i=1}^n$ decays at the rate $O(k^{-\beta})$. Hence by similar arguments as those given in the proof of Lemma 0.2, we obtain

$$\max_{1\le i,j\le rl_n}|\tilde{\Sigma}_{n,i,j}-\tilde{\Sigma}_{i,j}|\le Cn^{-\beta+1},$$

and we conclude that

$$\Big\|\sum_{i=1}^{n-m}\frac{m}{n-m}[T_{i,j,m}(t_{l,n})-S_{n,j}(t_{l,n})/\sqrt{n}]^2-\tilde{\Sigma}_{n,(l-1)r+j,(l-1)r+j}\Big\|_{q'}\le C(\sqrt{m/n}+1/m). \quad (6)$$

From these considerations and by arguments similar to those in the proof of Theorem 4 of Zhou (2013), we also obtain that

$$\Big\|\sum_{i=1}^{n-m}\frac{m}{n-m}[T_{i,j,m}(t_{l,n})-\frac{S_{n,j}(t_{l,n})}{\sqrt{n}}][T_{i,h,m}(t_{w,n})-\frac{S_{n,h}(t_{w,n})}{\sqrt{n}}] \ - \ \tilde{\Sigma}_{n,(l-1)r+j,(w-1)r+h}\Big\|_{q'}$$

$$\leq \quad C(\sqrt{m/n}+1/m),$$

for any $1 \leq j,h \leq r$ and $1 \leq l,w \leq l_n$. Therefore, an $L^{q'}$ maximum inequality yields

$$\Big\| \max_{1\leq a,b\leq rl_n} |[\mathrm{Cov}(\bar{\mathbf{\Phi}}_m|\{\mathbf{X}_i(u)\}_{i=1}^n)]_{a,b} - \tilde{\Sigma}_{n,a,b}|\Big\|_{q'} \leq [rl_n]^{2/q}(\sqrt{m/n}+1/m) := \gamma_{1,n},$$

where $\bar{\mathbf{\Phi}}_m = (\mathbf{\Phi}_m^\top(t_{1,n}),\cdots,\mathbf{\Phi}_m^\top(t_{l_n,n}))^\top$. By the assumption $\theta_1 < \phi' q^2/[2(q+1)]$, we can choose $l_n = n^{\theta_2}$ such that

$$\begin{cases} \theta_1/q - \theta_2 < 0 \\ 2(\theta_1+\theta_2)/q < \phi' \end{cases}$$

Since $2(\theta_1+\theta_2)/q < \phi'$, we conclude that $\gamma_{1,n}$ converges to zero at a polynomial rate. Hence the event

$$E_{1,n} := \max_{1\leq a,b\leq rl_n} \left|[\mathrm{Cov}(\bar{\mathbf{\Phi}}_m|\{\mathbf{X}_i(u)\}_{i=1}^n)]_{a,b} - \tilde{\Sigma}_{n,a,b}\right| \leq \sqrt{\gamma_{1,n}}$$

has probability at least $1-\gamma_{1,n}^{q/2}$. On event $E_{1,n}$, we have by Lemma 3.1 of Chernozhukov et al. (2013) that

$$\begin{aligned}\sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big(\max_{1\leq j\leq r}\sup_{1\leq i\leq l_n}|\Phi_{m,j}(t_{i,n})|\leq x\Big|\{\mathbf{X}_i(u)\}_{i=1}^n\Big) &- \mathbb{P}\Big(\max_{1\leq j\leq r}\sup_{1\leq i\leq l_n}|S_{n,j}^N(t_{i,n})|\leq x\Big)\Big| \\ &\leq C\gamma_{1,n}^{1/6}\log(rl_n\gamma_{1,n}^{-1/2})^{2/3}\to 0,\end{aligned}$$

which completes the proof of Lemma 0.3. ♢

Lemma 0.4. *Assume $m/n \to 0$. Then under the assumptions of Theorem 3.2 , we have*

$$\Big\|\sup_u \sum_{i=1}^{n-m}\frac{m}{n-m}[T'_{i,j,m}(u) - S'_{n,j}(u)/\sqrt{n}]^2\Big\|_{q/2} \leq C.$$

Proof. Without loss of generality, we assume $\mathbb{E}(X'_{i,j}(u)) = 0$. First we write $X'_{i,j}(u) := \sum_{j=0}^\infty b_{i,j,k}\cos(k\pi u)$. Then, similar to the proof of Proposition 2.1 utilizing $\|X''_{1,j}(u) - X''_{1,j}(v)\|_q \leq C|u-v|$, it follows

$$\|b_{i,j,k} - b^*_{i,j,k}\|_q \leq C\min(1/k^2, i^{-\beta}).$$

Let $B_{j,k} = \sum_{i=1}^n b_{i,j,k}/\sqrt{n}$. Then by the proof of Proposition 2.2, we have $\|B_{j,k}\|_q \leq Ck^{2(1-\beta)/\beta}$ which implies

$$\|\sup_{0\leq u\leq 1}|S'_{n,j}(u)|^2\|_{q/2} \leq \Big(\sum_{j=0}^\infty \|B_{j,k}\|_q\Big)^2 \leq C.$$

Similarly, we have for any i,

$$m\| \sup_{0\le u\le 1} |T'_{i,j,m}(u)|^2\|_{q/2} \le C.$$

As a result the lemma follows. ◊

Lemma 0.5. *Assume $m/n \to 0$. Then under the assumptions of Theorem 3.2 and the choice of θ_2 in Lemma 0.3, there exists a sequence of events $E_{2,n}$ such that $\mathbb{P}(E_{2,n}) \to 1$, and on $E_{2,n}$ we have the conditional Orlicz norm*

$$\Big\| \max_j \max_{1\le i\le l_n-1} \sup_{u\in[t_{i,n},t_{i+1,n}]} |\Phi_{m,j}(u) - \Phi_{m,j}(t_{i,n})| \Big\|_{\Psi_2} \to 0.$$

Proof. Note that, for any $u \in [t_{i,n}, t_{i+1,n}]$

$$|\Phi_{m,j}(u) - \Phi_{m,j}(t_{i,n})| = |\int_{t_{i,n}}^{u} \Phi'_{m,j}(s)\, ds| \le \int_{t_{i,n}}^{t_{i+1,n}} |\Phi'_{m,j}(s)|\, ds. \tag{7}$$

For each j and s, we have for the conditional variance

$$\mathrm{Var}(\Phi'_{m,j}(s)|\{\mathbf{X}_h(u)\}_{h=1}^n) = \sum_{i=1}^{n-m} \frac{m}{n-m}[T'_{i,j,m}(s) - S'_{n,j}(s)/\sqrt{n}]^2,$$

and Lemma 0.4 implies (using a maximal inequality) that

$$\Big\| \max_j \sup_u \sum_{i=1}^{n-m} \frac{m}{n-m}[T'_{i,j,m}(u) - S'_{n,j}(u)/\sqrt{n}]^2 \Big\|_{q/2} \le Cr^{2/q}.$$

Let $\gamma_{3,n} = r^{2/q+\eta}$ for an arbitrarily small positive constant η, and define the event

$$E_{2,n} = \max_j \sup_u \sum_{i=1}^{n-m} \frac{m}{n-m}[T'_{i,j,m}(u) - S'_{n,j}(u)/\sqrt{n}]^2 \le \gamma_{3,n}.$$

By Markov's inequality $\mathbb{P}(E_{2,n}) \ge 1 - r^{-\eta q/2}$. On the event $E_{2,n}$, we have that, conditional on $\{\mathbf{X}_h(u)\}_{h=1}^n$, $\mathrm{Std}(\Phi'_{m,j}(s)) \le \sqrt{\gamma_{3,n}}$ uniformly in j and s, where Std stands for "standard deviation". Note that the Orcliz norm is proportional to the standard deviation for centered Gaussian random variables. Therefore, on the event $E_{2,n}$ and conditional on $\{\mathbf{X}_h(u)\}_{h=1}^n$, we have $\|\Phi'_{m,j}(s)\|_{\Psi_2} \le \sqrt{\gamma_{3,n}}$ uniformly in j and s. Hence it follows by a maximal inequality that, on the event $E_{2,n}$ and conditional on $\{\mathbf{X}_h(u)\}_{h=1}^n$, the Orlicz norm

$$\Big\| \max_j \max_{1\le i\le l_n-1} \sup_{u\in[t_{i,n},t_{i+1,n}]} |\Phi_{m,j}(u) - \Phi_{m,j}(t_{i,n})| \Big\|_{\Psi_2}$$

$$
\begin{aligned}
&\leq \sqrt{\log(rl_n)} \max_{i,j} \Big\| \sup_{u\in[t_{i,n},t_{i+1,n}]} |\Phi_{m,j}(u) - \Phi_{m,j}(t_{i,n})| \Big\|_{\Psi_2} \\
&\leq \sqrt{\log(rl_n)} \max_{j} \sup_{s} \|\Phi'_{m,j}(s)\|_{\Psi_2}/l_n \leq \sqrt{\gamma_{3,n}}\sqrt{\log(rl_n)}/l_n.
\end{aligned}
$$

The right-hand side converges to zero polynomially fast for sufficiently small η since $\theta_1/q - \theta_2 < 0$.

♢

Proof of Theorem 3.2. Note that, for each fixed j,

$$
\Big\| \sup_{0\leq s\leq 1} |[S^N_{n,j}]'(k,s)| \Big\|_{\Psi_2} \leq \sum_{l=0}^{k} l\|N_{n,l,j}\|_{\Psi_2} \leq C\sum_{l=0}^{k} l\|N_{n,l,j}\|_2 \leq C\sum_{l=0}^{k} l \times l^{2(1-\beta)/\beta} \leq Ck^{2/\beta},
$$

where we have used the fact that the Orcliz norm is proportional to the standard deviation for centered Gaussian random variables. Hence we obtain that

$$
\Big\| \max_j \sup_{0\leq s\leq 1} |[S^N_{n,j}]'(k,s)| \Big\|_{\Psi_2} \leq Ck^{2/\beta}\sqrt{\log r}. \tag{8}
$$

On the other hand, we have

$$
\Big\| \sup_u |S^N_{n,j}(u) - S^N_{n,j}(k,u)| \Big\|_{\Psi_2} \leq C\sum_{l=k+1}^{\infty} \|N_{n,l,j}\|_{\Psi_2} \leq Ck^{2/\beta-1},
$$

which implies

$$
\Big\| \max_j \sup_u |S^N_{n,j}(u) - S^N_{n,j}(k,u)| \Big\|_{\Psi_2} \leq Ck^{2/\beta-1}\sqrt{\log r}. \tag{9}
$$

Combining (8) and (9) and by the triangular inequality, we obtain that

$$
\Big\| \max_j \max_{1\leq i\leq l_n} \sup_{t_i\leq u\leq t_{i+1}} |S^N_{n,j}(u) - S^N_{n,j}(t_{i,n})| \Big\|_{\Psi_2} \leq C[k^{2/\beta}/l_n + k^{2/\beta-1}]\sqrt{\log r}. \tag{10}
$$

Choose k diverging to infinity with a slowly enough polynomial rate such that the right hand side of (10) converges to 0 with a polynomial rate $n^{-\eta_1}$ for some $\eta_1 > 0$. We have that, for any $x \in \mathbb{R}$,

$$
\begin{aligned}
\mathbb{P}\Big[\max_{1\leq j\leq r} \sup_{0\leq u\leq 1} |S^N_{n,j}(u)| \leq x\Big] &- \mathbb{P}\Big[\max_{1\leq j\leq r} \max_{1\leq i\leq l_n} |S^N_{n,j}(t_{i,n})| \leq x - \tau_{1,n}\Big] \\
&\geq -\mathbb{P}[\max_j \max_{1\leq i\leq l_n} \sup_{t_{i,n}\leq u\leq t_{i+1,n}} |S^N_{n,j}(u) - S^N_{n,j}(t_{i,n})| \geq \tau_{1,n}] \\
&\geq -n^{-C_0}
\end{aligned}
$$

for some constant $C_0 > 0$, where $\tau_{1,n} = n^{-\eta_1}\sqrt{\log n}$ and we have used Markov's inequality in the second inequality above. Similarly, by Lemma 0.5, we have that, on the event $E_{2,n}$,

$$\begin{aligned}\mathbb{P}\Big[\max_{1\le j\le r}\sup_{0\le u\le 1}|\Phi_{m,j}(u)| &\le x \;\Big|\{\mathbf{X}_i\}_{i=1}^n\Big] - \mathbb{P}\Big[\max_{1\le j\le r}\max_{1\le i\le l_n}|\Phi_{m,j}(t_{i,n})| \le x+\tau_{2,n}\;\Big|\{\mathbf{X}_i\}_{i=1}^n\Big] \\ &\le \mathbb{P}[\max_j \max_{1\le i\le l_n}\sup_{t_{i,n}\le u\le t_{i+1,n}}|\Phi_{m,j}(u)-\Phi_{m,j}(t_{i,n})| \ge \tau_{2,n}|\{\mathbf{X}_i\}_{i=1}^n] \\ &\le n^{-C_1}\end{aligned}$$

for some constant $C_1 > 0$, where $\tau_{2,n} = n^{-\eta_2}\sqrt{\log n}$ for some constant $\eta_2 > 0$. Therefore, by Lemma 0.3 and on the event $E_{1,n} \cap E_{2,n}$, we have

$$\begin{aligned}P_n &:= \mathbb{P}\Big[\max_{1\le j\le r}\sup_{0\le u\le 1}|\Phi_{m,j}(u)| \le x\;\Big|\{\mathbf{X}_i\}_{i=1}^n\Big] - \mathbb{P}\Big[\max_{1\le j\le r}\sup_{0\le u\le 1}|S_{n,j}^N(u)| \le x\Big] \\ &\le \mathbb{P}\Big[\max_{1\le j\le r}\max_{1\le i\le l_n}|\Phi_{m,j}(t_{i,n})| \le x+\tau_{2,n}\;\Big|\{\mathbf{X}_i\}_{i=1}^n\Big] - \mathbb{P}\Big[\max_{1\le j\le r}\max_{1\le i\le l_n}|S_{n,j}^N(t_{i,n})| \le x-\tau_{1,n}\Big] + O(n^{-\tilde{C}}), \\ &\le \eta_{3,n} + \mathbb{P}\Big[x-\tau_{1,n} < \max_{1\le j\le r}\max_{1\le i\le l_n}|S_{n,j}^N(t_{i,n})| \le x+\tau_{2,n}\Big] + O(n^{-\tilde{C}}),\end{aligned}$$

where $\tilde{C} = \min(C_0, C_1)$ and $\eta_{3,n}$ is the approximation error in Lemma 0.3. By Nazarov's Inequality (Nazarov (2003)),

$$\mathbb{P}\Big[x-\tau_{1,n} < \max_{1\le j\le r}\max_{1\le i\le l_n}|S_{n,j}^N(t_{i,n})| \le x+\tau_{2,n}\Big] \le C(\tau_{2,n}+\tau_{1,n})\sqrt{\log(rl_n)}$$

which converges to 0 at a polynomial rate. Hence, it follows

$$P_n \le \eta_{3,n} + O((\tau_{2,n}+\tau_{1,n})\sqrt{\log(rl_n)}) + O(n^{-\tilde{C}})$$

uniformly with respect to $x \in \mathbb{R}$. Note that the right hand side of the above inequality converges to 0 at a polynomial rate. By a similar argument, we obtain on the event $E_{1,n} \cap E_{2,n}$ that

$$P_n \ge -\big[\eta_{3,n} + O((\tau_{2,n}+\tau_{1,n})\sqrt{\log(rl_n)}) + O(n^{-\tilde{C}})\big]$$

uniformly with respect to $x \in \mathbb{R}$, which implies

$$\sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big[\max_{1\le j\le r}\sup_{0\le u\le 1}|\Phi_{m,j}(u)| \le x\;\Big|\{\mathbf{X}_i\}_{i=1}^n\Big] - \mathbb{P}\Big[\max_{1\le j\le r}\sup_{0\le u\le 1}|S_{n,j}^N(u)| \le x\Big]\Big| \to 0.$$

◇

Proofs of the results in Section 4

Proof of Proposition 4.1. Observe that $\check{X}_{i,j}(u)$ is a centered process. Hence this proposition follows from Theorem 3.1 if we can show that $v_j(u)$ is twice continuously differentiable on $[0,1]$ for all j. Observe that $g_j(u) = \mathbb{E}(X_{i,j}(u))$. By the assumption that $X_{i,j}(u)$ is twice continuously differentiable, (2.6) and a version of the Dominated Convergence Theorem (c.f. Theorem 1.6.8 of Durrett (2010)), we have that $g_j(u)$ is twice continuously differentiable. As a result $X_{i,j}(u) - g_j(u)$ is twice continuously differentiable a.s.. Since $v_j^2(u) = \mathbb{E}[X_{i,j}(u) - g_j(u)]^2$, (2.6) and the above mentioned version of the Dominated Convergence Theorem imply that $v_j(u)$ is twice continuously differentiable for all j. $\diamondsuit$

Proof of Lemma 4.1. Recall the notation (4.2) and observe that

$$\hat{v}_j^2(u) - v_j^2(u) = \frac{1}{n}\sum_{i=1}^{n}[\check{X}_{i,j}^2(u) - \mathbb{E}\check{X}_{i,j}^2(u)] - \frac{1}{n}\check{S}_{n,j}^2(u). \tag{11}$$

Let $Z_{i,j}(u) = \check{X}_{i,j}^2(u) - \mathbb{E}\check{X}_{i,j}^2(u)$. Then $\{Z_{i,j}(u)\}_{i=1}^n$ is a centered functional time series with dependence measures

$$\begin{aligned}\delta_{Z_{i,j}}(k, q/2) &= \sup_{u\in[0,1]} \|Z_{k,j}(u) - Z_{k,j}^*(u)\|_{q/2} \le \sup_{u\in[0,1]} \|\check{X}_{k,j}(u) + \check{X}_{k,j}^*(u)\|_q \|X_{k,j}(u) - X_{k,j}^*(u)\|_q \\ &\le C\delta_{G_j}(k,q) \le Ck^{-\beta}.\end{aligned}$$

Therefore, by the arguments in the proof of Lemma 0.4, we obtain that

$$\Big\| \sup_{u\in[0,1]} \frac{1}{n}\Big|\sum_{i=1}^{n}[\check{X}_{i,j}^2(u) - \mathbb{E}\check{X}_{i,j}^2(u)]\Big|\Big\|_{q/2} \le C/\sqrt{n}$$

and an $L^{q/2}$ maximal inequality yields

$$\Big\| \max_{1\le j\le r} \sup_{u\in[0,1]} \frac{1}{n}\Big|\sum_{i=1}^{n}[\check{X}_{i,j}^2(u) - \mathbb{E}\check{X}_{i,j}^2(u)]\Big|\Big\|_{q/2} \le Cr^{2/q}/\sqrt{n}.$$

Similarly, we also obtain that

$$\Big\| \max_{1\le j\le r} \sup_{u\in[0,1]} \frac{1}{n}\check{S}_{n,j}^2(u)\Big\|_{q/2} \le Cr^{2/q}/n.$$

The Lemma follows by these two inequalities and (11). $\diamondsuit$

Proof of Proposition 4.3. Recalling the Fourier expansion $\mathbf{X}_i(u) = \sum_{k=0}^{\infty} \mathbf{a}_{i,k} \cos(k\pi u)$. It follows that $W_{i,j}(u)$ admits the following cosine representation

$$W_{i,j}(u) = \sum_{k=1}^{\infty} a_{i,j,k} \cos(k\pi u), \tag{12}$$

where $a_{i,j,k}$ is the j-th component of $\mathbf{a}_{i,k}$, $j = 1, 2, \cdots, r$. As a result, $W_{i,j,k}(u)$ can be represented in the form

$$W_{i,j,k}(u) = \sum_{c=1}^{\infty} f_{i,j,k,c} \cos(c\pi u), \tag{13}$$

where $f_{i,j,k,c} = a_{i,j,c} - a_{i,k,c}$. From the proof of Proposition 2.1, it follows

$$\max_{j,k} \|f_{i,j,k,c} - f^*_{i,j,k,c}\|_q \le C \min(1/c^2, i^{-\beta}), \tag{14}$$

for all $c \ge 1$, where $c > 0$ is a constant. Furthermore, the dependence measures of $\{W_{i,j,k}(u)\}_{i=1}^n$ satisfy

$$\delta_{W_{j,k}}(h,q) \le \delta_{W_j}(h,q) + \delta_{W_k}(h,q) \le 4\delta_H(h,q) \le Ch^{-\beta} \tag{15}$$

uniformly in j, k, where the second inequality is derived from the fact that

$$\begin{aligned}
\delta_{W_j}(h,q) &= \sup_u \left\| X_h(u) - \int_0^1 X_h(u)\, du - [X_h^*(u) - \int_0^1 X_h^*(u)\, du] \right\|_q \\
&\le \sup_u \|X_h(u) - X_h^*(u)\|_q + \int_0^1 \|X_h(u) - X_h^*(u)\|_q \, du \\
&\le 2 \sup_u \|X_h(u) - X_h^*(u)\|_q = 2\delta_H(h,q).
\end{aligned}$$

Recall that $\mathring{\mathbf{W}}_i(u)$ is the $r(r-1)/2$-dimensional vector with entries $W_{i,j,k}(u)$. The proof of Proposition 4.3 now follows from applying the same arguments as those given in the proof of Theorem 3.1 and Proposition 4.1 to $\mathring{\mathbf{W}}_i(u)$. $\diamondsuit$

Proof of Proposition 4.5. Recall the definition of $\mathring{S}_{n,j,k}$ in (4.8) and define

$$T^* = \max_{1 \le j < k \le r} \sup_{0 \le u \le 1} |\mathring{S}_{n,j,k}(u) - \mathbb{E}\mathring{S}_{n,j,k}(u)|/v_{j,k}(u).$$

Then following the proof of Propositions 4.3 and 4.4 it follows

$$\mathbb{P}(T^* > \mathring{c}_{1-\alpha}) \to 1 - \alpha. \tag{16}$$

Meanwhile, from the proofs of Propositions 4.3 and 4.4 we observe that T^* is asymptotically equivalent to the L^∞ norm of a $r(r-1)/2$ dimensional Gaussian random function and the Orcliz norm of each Gaussian random function is $O(1)$. Hence it is straightforward to conclude by a maximal inequality that the Orcliz norm of the latter L^∞ norm is of order $O(\sqrt{\log n})$ in view of $r \asymp n^{\theta_1}$. Therefore $\mathring{c}_{1-\alpha} = O_{\mathbb{P}}(\sqrt{\log n})$. On the other hand, it is straightforward to conclude that, under $\mathbf{H}_a$,

$$\max_{1\le j<k\le r} \sup_{u\in[0,1]} |\mathbb{E}\mathring{S}_{n,j,k}(u)|/v_{j,k}(u)|/\sqrt{\log n} \to \infty.$$

Together with equation (16), we conclude that $\mathbb{P}(T/\sqrt{\log n} > d_n) \to 1$ for any divergent sequence d_n. Hence the result follows in view of the fact that $\mathring{c}_{1-\alpha} = O_{\mathbb{P}}(\sqrt{\log n})$. $\diamondsuit$

References

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

Durrett, R. (2010). *Probability, Theory and Examples (Fourth Edition).* Cambridge University Press.

Nazarov, F. (2003). On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In *Geometric Aspects of Functional Analysis, Lecture Notes in Mathematics, Volume 1807*, pages 169–187. Springer.

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

Wu, W. B. (2007). Strong invariance principles for dependent random variables. *Annals of Probability.*, 35(6):2294–2320.

Zhang, D. and Wu, W. B. (2017). Gaussian approximation for high dimensional time series. *Annals of Statistics*, 45(5):1895–1919.

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

Zhou, Z. (2014). Inference of weighted V-statistics for non-stationary time series and its applications. *Annals of Statistics*, 42(1):87–114.