On Functional Data Analysis with Dependent Errors

Dissertation

submitted for the degree of Doctor of Natural Sciences (Dr. rer. nat.)

Presented by

Haiyan Liu

at the

Universität Konstanz

Faculty of Sciences
Department of Mathematics and Statistics

Date of the oral examination: 22 June, 2016

First referee: Professor Dr. Jan Beran
Second referee: Professor Dr. Michael Kupper

Acknowledgement

Mathematics is fascinating and amazing which provides me with the non-stop steps of my life.

It is my pleasure to express my sincerest gratitude to my supervisor Prof. Jan Beran for his greatest trust, patient guidance and continuous support of my PhD study. No matter how stupid things I made, he forgave me and showed me the correct directions. So that I can have the opportunity and confidence to do my PhD research and to pursue my future research. To be his student I feel extremely honoured and lucky. I would also like to show my gratitude to Prof. Michael Kupper and Prof. Robert Denk for their valuable suggestions and discussions as members of my dissertation committee. I would also like to thank all my other teachers and advisors in my life. Without their help, I would be the most unenlightened one.

Moreover, greatest thanks go to my colleagues. Thanks to Ms. Veronika Weißer for her help and care. Thanks to Dr. Volker Bürkel, Dr. Hao Liu, Dr. Dieter Schell and Dr. Zhaoxia Yang for their help and advice especially at the beginning of my study. Thanks to Klaus for the happy and valuable collaboration. Thanks to Britta, Ludovic, Nadiya, Peng, Reinhard, Valeria, Yusufu for their help and kindness. Many thanks go to all my friends for the accompany and joy.

Last but not least, my deepest gratitude goes to my parents Wen Liu and Xiuyun Liu for giving me a health body and mind. Owing to my parents and my elder brother Jianpeng Liu's love and support over the years, I have the hardware and software to pursue everything. I also want to thank myself for working hard and not so frequently being lazy.

This work is financially supported by the German Research Foundation (DFG) through research unit FOR 1882 "Psychoeconomics".

Abstract

Classical functional data analysis (FDA) is based on directly observed random curves. However, in a more realistic setting such as for certain types of EEG data, the observations are perturbed by noise even strongly dependent noise. In this dissertation the influence of long memory noise on trend and covariance estimation, functional principal component analysis and two sample inference is investigated.

Firstly, the kernel estimation of trend function and covariance function in repeated time series with long memory errors is considered. Functional central limit theorem of estimated trend and estimated covariance is established. Since the main quantity of interest in FDA is the covariance, the trend plays the role of a nuisance parameter. Therefore, the orthogonal contrast transformation is proposed to eliminate the trend before estimating the covariance. In order to relax the constrain between the number of random curves and the number of sampling points on each curve, higher order kernels are used.

Secondly, we consider the estimation of eigenvalues, eigenfunctions (functional principal components) and functional principal component scores in FDA models with short or long memory errors. It turns out that there is no difference between short and long memory errors when considering the asymptotic distribution of estimated eigenvalues and estimated eigenfunctions. However, the asymptotic distribution of estimated scores and the rate of convergence differ significantly between weakly and strongly dependent errors. Moreover, long memory property not only lead to a slower rate of convergence, but the dependence of score estimators.

Thirdly, two sample inference for eigenspaces in FDA models with dependent errors is discussed. A test for testing the equality of subspaces spanned by a finite number of eigenfunctions is constructed and its asymptotic distribution under the null hypothesis is derived. This provides the basis for defining suitable test procedures. In order to obtain asymptotically exact rejection regions, the joint asymptotic distribution of the residual process is required. However, since the dimension of the subspace is in most cases very small, we propose to use a simple Bonferroni adjusted test. A more practical solution is a bootstrap test which is also applicable even for small samples.

Zusammenfassung

Klassische “functional data analysis” (FDA) geht von direkt beobachteten zufälligen Funktionen aus. Realistischer ist vielen Situation (wie zum Beispiel bei EEG-Daten), von der Annahme auszugehen, dass die Beobachtungen mit zufälligen Messfehlern behaftet sind, der möglicherweise langfristig abhängig sein können. In dieser Dissertation wird der Effekt langfristiger Abhängigkeit auf die Schätzung von Trend- und Kovarianzfunktion, sowie auf FDA und Zweistichprobentests untersucht.

Zuerst betrachten wir Kernschätzung des Trends und der Kovarianzfuntion für wiederholte Zeitreihen mit langfristiger Abhängigkeit. Funktional Grenzwertsätze werden hergeleitet. Da das Hauptinteresse bei FDA der Kovarianz gilt, ist der Trend ein Störparameter. Wir führen deshalb eine “orthogonal contrast transformation” ein, mit der Trend eliminiert wird bevor man zur Schätzung der Kovarianzfuntion übergeht. Um die Annahmen für die Anzahl replizierter Zeitreihen zu verbessern, verwenden wir zudem Kerne höherer Ordnung.

Als nächstes betrachten wir die Schätzung von Eigenwerten, Eigenfunktionen (functional principal components) und der sogenannten “function principal component scores” unter kurzfristiger und langfristiger Abhängigkeit. Es stellt sich heraus, dass es in Bezug auf Eigenwerte und Eigenfunktionen keinen Unterschied gibt zwischen kurz- und langfristiger Abhängigkeit. Jedoch ist die asymptotische Verteilung und Konvergenzrate der geschätzten Scores völlig unterschiedlich. Langfristige Abhängigkeit führt nicht nur zu einer langsameren Konvergenzrate sondern auch zu einem Verlust der Unabhängigkeit verschiedener Score-Schätzungen.

Drittens diskutieren wir Zweistichproben-Inferenz für Eigenräume von FDA-Modellen. Ein Test der Nullhypothese, dass zwei endliche Eigenräume identisch sind, wird eingeführt und die asymptotische Verteilung unter der Nullhypothese hergeleitet. Um asymptotisch korrekte Verwerfungsbereiche zu definieren, wird die gemeinsame Verteilung von Residualprozessen benötigt. Da die Dimension der zu testenden Eigenräume meist klein ist, führen wir eine einfache Bonferroni-Korrektur ein. Eine praktikablere Lösung ist ein Bootstrap Test, der auch für kleine Stichproben anwendbar ist.

Contents

Chapter 1

Introduction

It is assumed that, in classical functional data analysis (FDA), one observes n independent random curves

$$\{X_1(t), ..., X_n(t)\}$$

on the interval $[0, 1]$ directly. These n independent random curves $\{X_i(t)\}$ ($i = 1, ..., n$) are assumed to come from an underlying unknown stochastic process $X(t) \in L^2[0, 1]$ with expectation $\mu(t) = E[X(t)]$ and covariance function $C(s, t) = cov(X(s), X(t))$. By Karhunen-Loève (K.L.) expansion, the random curves $X_i(t)$ admit the form

$$X_i(t) = \mu(t) + \sum_{l=1}^{\infty} \xi_{il}\phi_l(t) \quad (i = 1, ..., n) \tag{1.1}$$

where the coefficients $\{\xi_{il}\}$ ($i = 1, ..., n$, $l \in \mathbb{N}$) (functional principal component scores) are uncorrelated random variables with $E[\xi_{il}] = 0$, $E[\xi_{il}^2] = \lambda_l \geq 0$ and $\sum \lambda_l < \infty$, the functions $\{\phi_l(t)\}$ ($l \in \mathbb{N}$) are continuous real-valued functions on $[0, 1]$ that are pairwise orthogonal in $L^2[0, 1]$. By Mercer's theorem, $C(s, t)$ can be written as

$$C(s, t) = \sum_{l=1}^{\infty} \lambda_l\phi_l(s)\phi_l(t) \quad (s, t \in [0, 1]) \tag{1.2}$$

where $\{\lambda_l\}$, $\{\phi_l(t)\}$ ($l \in \mathbb{N}$) denote the eigenvalues and the corresponding eigenfunctions (functional principal components) of the covariance operator $\mathbf{C}(y) = E\left[\langle (X - \mu), y\rangle (X - \mu)\right]$ ($y \in L^2[0, 1]$), $\{\phi_l(t)\}$ ($l \in \mathbb{N}$) build an orthonormal $L^2[0, 1]$-basis.

There are three basic questions about this classical FDA model (1.1), which are very important in analyzing the essential behavior of the underlying unknown random curve $X(t)$:

- How to estimate the trend function $\mu(t)$ and covariance function $C(s,t)$?
- How to estimate the eigenvalues λ_l, eigenfunctions $\phi_l(t)$ and the functional principal component scores of the observed sample paths $\xi_{il} = \xi_{il}(\omega)$?
- How to make the two sample inference for two independent functional samples $\left\{X_i^{(1)}(t)\right\}$ and $\left\{X_i^{(2)}(t)\right\}$?

There exists a huge number of literature on this classical FDA model where one can find satisfactory and successful methods to deal with the above three questions. For instance, monographs of Bosq (2000), Ferraty (2011), Ferraty and Romain (2011), Ferraty and Vieu (2006), Horváth and Kokoszka (2012), Ramsay and Silverman (2002, 2005) and Shi and Choi (2011) give the fundamental concepts and methods.

Nonparametric estimation of $\mu(t)$ is fully considered by many statisticians. The introduction to spline smoothing method can be found in de Boor (2001), Eubank (1999), Green and Silverman (1993), Wahba (1990), Wang (2011) and the references therein. Wavelet smoothing method can be found in Antoniadis and Oppenheim(2012), Chui (1992), Daubechies (1992), Donoho et al. (1995), Johnstone and Silverman (1997), Ogden (2012), Percival and Walden (2006) and the references therein. For kernel smoothing and local polynomial smoothing methods see for instance Fan and Gijbels (1996), Härdle and Vieu (1992), Hart and Wehrly (1986), John (1984), Wand and Jones (1994), Lin and Carroll (2000) and Loader (2012). References about nonparametric estimation of $C(s,t)$ in FDA can be found in works such as Bevilacqua et al. (2012), Bigot et al. (2010), Bigot et al. (2011), Fan et al. (2013), Hall et al. (1994), Lirio et al. (2014), Sancetta (2014), Shaby and Ruppert (2012) and Zhang and Chen (2007).

Many works on the estimation of λ_l, $\phi_l(t)$ and ξ_{il} have also been reported. Bosq (2000) and Dauxois et al. (1982) consider the estimation of λ_l and $\phi_l(t)$ and the corresponding central limit theorem. Johnstone and Lu (2009) give a counter example if the regular conditions are not hold and propose different approaches to deal with non-smooth data. Under some additional conditions, Hall and Hosseini-Nasab (2006) and Hall and Hosseini-Nasab (2009) derive stochastic expansions of λ_l and $\phi_l(t)$. Gervini (2008) propose the method by using functional median as a more robust measure of central tendency and also construct a robust estimation of $\phi_l(t)$. A definition of the mode of the distribution of a random function is considered by Delaigle and Hall (2010) who also give the functional principal component expansions. Boente and Fraiman (2000) discuss the kernel based functional principal components analysis. Other contributions include Cardot (2000), Ma (2013), Ocana et al. (1999), Reiss and Ogden (2007) and Yang et al. (2011).

Since there are possibly different functional principal components' structures, it may be different to work with two functional samples. Hall and Keilegom (2007) consider a bootstrap test of two samples with iid random errors. Fremdt et al. (2013) construct a robust test for the equality of the covariance structures in two functional samples. Benko et al. (2009) and Boente et al. (2011) construct a bootstrap test for testing these equalities in the situation without errors. Horváth et al. (2009) propose a method of comparing two functional linear models where the explanatory variables are functions and the response variables can be either scalars or functions. Panaretos et al. (2010) test whether the two functional samples of continuous independent identically Gaussian processes with zero mean have the same covariance or not. Horváth et al. (2013) discuss the two sample problem of dependent time series. Gromenko and Kokoszka (2012) construct a test for testing the equality of the mean fucntions of the curves from two disjoint spatial regions. Other contributions are given in Cuevas et al. (2004), Ferraty et al. (2007), Gabrys et al. (2010).

Most of these literature deal with the situation where the random curves $X_i(t)$ are observed directly. However, as pointed out in Yao et al. (2003), Yao et al. (2005) and Yao (2007), in a more realistic setting, the observations may be perturbed by random noise. Yao (2007) therefore consider the question of nonparametric estimation in the FDA context where $X_i(t)$ are randomly perturbed by noises ϵ and series $Y_{ij} = X_i(t_j) + \epsilon_i(j)$ with independent identically distributed (iid) random errors $\epsilon_i(j)$ are observed (also see e.g. Cai and Yuan 2010, Hall et al. 2006, Ramsay and Ramsay 2002, and Staniswalis and Lee 1998). Typically it is assumed that, for fixed i, $\{\epsilon_i(j)\}$ $(j \in \mathbb{N})$ are iid random variables. However, this assumption is too restrictive in some applications. In particular, for example for certain types of EEG signals, long memory properties in $\{\epsilon_i(j)\}$ $(j \in \mathbb{N})$ can exist (see e.g. Bornas et al. 2013, Linkenkaer-Hansen et al. 2001, Parish et al. 2004, Nikulin and Brismar 2005, Watters 2000). We focus here on this situation, although the corresponding results can be derived in a more generally situation which includs the possibility of weakly dependence. The related inferences are our results in Beran and Liu (2014), Beran and Liu (2016) and Beran, Liu and Telkmann (2016).

In summary, the following situation will be discussed.

- The observations include n independent time series $Y_{i\cdot} = (Y_{i1}, ..., Y_{iN})$ $(i = 1, ..., n)$ and each observation Y_{ij} is defined by:

$$Y_{ij} = X_i(t_j) + \epsilon_i(j) \quad (t_j = j/N; = 1, ..., N), \tag{1.3}$$

  where n is the number of curves, N is the number of sampling points on each curve, $t_j = j/N$ denotes rescaled time.

- The random curves $X_1(t), ..., X_n(t)$ are assumed to come from an underlying random process $X(t) \in L^2[0,1]$ independently and by K.L. expansion, $X_i(t)$ can be written as :

$$X_i(t) = \mu(t) + \sum_{l=1}^{\infty} \xi_{il}\phi_l(t) \quad (t \in [0,1]) \tag{1.4}$$

  with covariance $cov(X(t), X(s)) = C(s,t) = \sum_l \lambda_l \phi_l(s)\phi_l(t)$ and expectation $E[X(t)] = \mu(t)$, where $\{\xi_{il}\}$ $(i = 1, ..., n,\ l \in \mathbb{N})$ are uncorrelated random variables with mean 0 and variance λ_l and $\sum \lambda_l < \infty$, $\{\lambda_l\}$ and $\{\phi_l(t)\}$ $(l \in \mathbb{N})$ denote the eigenvalues and the corresponding eigenfunctions of the covariance operator $\mathbf{C}$, $\{\phi_l(t)\}$ $(l \in \mathbb{N})$ build an orthonormal $L^2[0,1]$-basis.

- Error processes $\{\epsilon_i(j)\}$ $(j \in \mathbb{N})$ are stationary Gaussian with autocovariance

$$\gamma_\epsilon(k) = cov\left(\epsilon_i(j), \epsilon_i(j+k)\right) \underset{k\to\infty}{\sim} c_\gamma |k|^{2d-1} \tag{1.5}$$

and spectral density

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_\epsilon(k) e^{ik\lambda} \underset{|\lambda|\to 0}{\sim} c_f |\lambda|^{-2d} \tag{1.6}$$

for some constants $0 < c_f, c_\gamma < \infty$ and $d \in [0, \frac{1}{2})$, where "$\sim$" has the meaning that the ratio of the left and right hand side goes to one. For the case $d = 0$, $\{\epsilon_i(j)\}$ $(j \in \mathbb{N})$ is a short-range dependent time series. For the case $d \in (0, \frac{1}{2})$, $\{\epsilon_i(j)\}$ $(j \in \mathbb{N})$ is a long-range dependent time series. The error processes are assumed to be independent with the scores ξ_{il}.

Our main purpose in this thesis is to investigate and discuss the influence of short- and long-memeory properties in errors on estimation of $\mu(t)$, $C(s,t)$, λ_l, $\phi_l(t)$, ξ_{il} and on the two functional sample inference. In this thesis, we only consider the equidistant sampling points case and Gaussian $\epsilon_i(j)$ and Gaussian ξ_{il}. The reason is that we mainly focus on the essential effect of short or long range dependence in the errors $\{\epsilon_i(j)\}$ $(j \in \mathbb{N})$. A generalization to nonequidistant FDA models is also possible and the corresponding methods will be considered later. In fact, as pointed out by Menéndez et al. (2010) and Menéndez et al. (2013), for the case of single observed nonequidistant time series, some cautions with respect to the distribution of observational time points is needed while doing estimation.

For the nonparametric regression for single long-range dependent time series, one can refer to for example Beran and Feng (2002a,b,c), Beran and Shumeyko (2012), Csörgö and Mielniczuk(1995), Hall and Hart (1990), Ray and Tsay (1997) and Robinson (1997). Ghosh (2001) gives the first results on nonparametric trend function estimation in replicated long-range dependent time series. Generally, literature on FDA with dependent errors seems to be sparse. The first results on estimation of $\mu(t)$ and $C(s,t)$ in repeated time series under general dependence assumptions on $\epsilon_i(j)$ appear in Beran and Liu (2014). The estimation of λ_l, $\phi_l(t)$ and ξ_{il} and their asymptotic behavior under long memory error processes is firstly considered by Beran and Liu (2016). Beran, Liu and Telkmann (2016) discuss the two sample inference problem in FDA models with weakly and strongly dependent errors. For a general and detailed overview on statistical inference for long-range dependent processes see e.g. Beran (1994), Beran (2010), Beran et al. (2013), Doukhan et al. (2003), Giraitis et al (2012), Robinson (2003) and references therein.

The thesis is organized as follows:

Chapter 2 is a short introduction to long memory processes including FARIMA models and functional data analysis (FDA). The references to this chapter include Beran (1994), Beran et al. (2013), Doukhan et al. (2003), Giraitis et al (2012), Robinson (2003) for long memory process, and Bosq (2000), Ferraty and Vieu (2006) and Horváth and Kokoszka (2012), Ramsay and Silverman (2005) for functional data analysis.

The aim of Chapter 3 is to obtain the mean $\mu(t)$ and covariance $C(s,t)$ estimation in FDA models with long memory error processes and the asymptotic behavior of these estimators. In the first step, one-dimensional boundary kernel estimation of $\mu(t)$ and two-dimensional boundary kernel estimation of $C(s,t)$ are defined. Then, the asymptotic mean, variance of estimated mean $\hat{\mu}(t)$ are obtained. Conditions needed to obtain a functional limit theorem for $\hat{\mu}(t)$ lead to the idea of using contrast transformations before estimating of $C(s,t)$. Therefore the asymptotic mean, asymptotic variance and the functional limit theorems for kernel estimators of $C(s,t)$ based on this contrast transformation model are discussed. A small simulation study illustrates the asymptotic results for $\hat{\mu}(t)$ comes at last. The presentation of this chapter is similar to that of Beran and Liu (2014).

In Chapter 4, we estimate the eigenvalues λ_l, eigenfunctions (functional principal components) $\phi_l(t)$ and functional principal component scores ξ_{il} based on the covariance estimator $\hat{C}(s,t)$ defined in Chapter 3. Asymptotic properties of estimated eigenvalues $\hat{\lambda}_l$ and eigenfunctions $\hat{\phi}_l(t)$ are derived for both short- and long-range dependent errors. Moreover, the asymptotic joint distribution of $\{\hat{\phi}_l(t)\}$ $(l = 1, ..., p)$ which will be used by considering two sample inference is also discussed. The asymptotic joint distribution of $\{\hat{\xi}_{il}\}$ $(l = 1, ..., p)$ for each curve is obtained for the short- and long-range dependent errors cases respectively. Simulations illustrate the asymptotic results of $\hat{\lambda}_l$ and $\hat{\phi}_l(t)$. This chapter is based on our previous work in Beran and Liu (2016).

Chapter 5 considers the two functional sample inference for eigenspaces in FDA with dependent errors. After recalling the basic results in Chapter 3 and Chapter 4, we construct a test for testing the equality of subspaces spanned by a finite number of eigenfunctions. The test is based on the residual process. Then, we derive the asymptotic null distribution. This provides the basis for defining suitable test procedures. However, in order to obtain reasonable rejection regions and to avoid the calculation of the joint asymptotic distribution of the residual process, we propose to use a simple Bonferroni adjusted test (since, in most cases, the dimension of the subspace is very small). A more practical solution - a bootstrap test - is also constructed. Simulations illustrate the results are discussed. This chapter is based on our previous results in Beran, Liu and Telkmann (2016).

We conclude this thesis with some concluding remarks in Chapter 6.

Chapter 2

Basics

This chapter serves to collect preliminary probabilistic properties and statistical methods in long memory processes and functional data analysis (FDA) that will be needed later in this thesis. For the sake of completeness, we include these materials.

In section 2.1, we cite some well-known basic definitions and results in long memory processes. In fact, these results can be found in various textbooks and research monographs, e.g. Beran (1994), Beran et al. (2013), Doukhan et al. (2003), Giraitis et al. (2012) and Robinson (2003).

Section 2.2 briefly reviews some fundamental concepts and facts about FDA. They are also well-known, see e.g. Bosq (2000), Ferraty and Vieu (2006), Horváth and Kokoszka (2012) and Ramsay and Silverman (2005).

2.1 Long memory processes

We will introduce the definition of (linear) dependence structures in this section (see Beran et al. 2013 pages 20 and 30).

Definition 2.1. *Let* $\{X_t\} \in \mathbb{R}$ $(t \in \mathbb{Z})$ *be a stochastic process.* $\{X_t\}$ *is said to be strictly stationary if for all* $k, l \in \mathbb{N}$ *and* $t_1, ..., t_l \in \mathbb{Z}$ *the joint distributions of* $(X_{t_1}, ..., X_{t_l})$ *and* $(X_{t_1+k}, ..., X_{t_l+k})$ *are same.*

Definition 2.2. *Let* $\{X_t\} \in \mathbb{R}$ $(t \in \mathbb{Z})$ *be a stochastic process with* $E\left[X_t^2\right] < \infty$

$(t \in \mathbb{Z})$. Define $\mu_t = E[X_t]$, then the function

$$\gamma_X : \ \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$$
$$(s,t) \mapsto \gamma_X(s,t) = cov(X_s, X_t) = E[(X_s - \mu_s)(X_t - \mu_t)]$$

is called autocovariance function of $\{X_t\}$.

Definition 2.3. *Let $\{X_t\} \in \mathbb{R}$ $(t \in \mathbb{Z})$ be a stochastic process with $E[X_t^2] < \infty$ $(t \in \mathbb{Z})$. Then $\{X_t\}$ is said to be weakly stationary if*

$$(i) \quad \exists \mu \in \mathbb{R} : E[X_t] = \mu \text{ for all } t \in \mathbb{Z},$$
$$(ii) \quad \gamma_X(s,t) = \gamma_X(t-s, 0) =: \gamma_X(t-s) \text{ for all } s,t \in \mathbb{Z}.$$

Definition 2.4. *Let $\{X_t\} \in \mathbb{R}$ $(t \in \mathbb{Z})$ be a weakly stationary stochastic process with autocovariance function $\gamma_X(k)$. Then $\{X_t\}$ is said to exhibit (linear)*

$$(i) \quad \text{short-range dependence if } 0 < \sum_{k=-\infty}^{\infty} \gamma_X(k) < \infty,$$
$$(ii) \quad \text{long-range dependence if } \sum_{k=-\infty}^{\infty} \gamma_X(k) = \infty,$$
$$(iii) \quad \text{antipersistence if } \sum_{k=-\infty}^{\infty} \gamma_X(k) = 0.$$

Definition 2.5. *Let $\{X_t\} \in \mathbb{R}$ $(t \in \mathbb{Z})$ be a weakly stationary stochastic process with autocovariance function $\gamma_X(k)$, then the function*

$$f_X : \ [-\pi, \pi] \to \mathbb{R}$$
$$\lambda \mapsto f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma_X(k) e^{-ik\lambda}$$

is called spectral density function of $\{X_t\}$.

In the following several chapters, for simplicity of presentation, long-range (short-range) dependence is characterized by

$$\gamma(k) = cov(X_j, X_{j+k}) \underset{k \to \infty}{\sim} c_\gamma |k|^{2d-1}$$

for some constants $d \in (0, \frac{1}{2})$ $(d = 0)$ and $0 < c_\gamma < \infty$. For the spectral density $f(\lambda)$, this corresponds to

$$f(\lambda) \underset{|\lambda| \to 0}{\sim} c_f |\lambda|^{-2d}$$

for some constants $d \in (0, \frac{1}{2})$ $(d = 0)$ and $0 < c_f < \infty$.

Example 2.1. *Let* $\{X_t\} \in \mathbb{R}$ $(t \in \mathbb{Z})$ *be a weakly stationary stochastic process with*

$$X_t = \sum_{j=0}^{\infty} a_j \epsilon_{t-j},$$

where ϵ_t *are uncorrelated zero mean random variables,* $\sigma_\epsilon^2 = var(\epsilon_t) < \infty$*, and*

$$a_j = (-1)^j C_{-d}^j = (-1)^j \frac{\Gamma(1-d)}{\Gamma(j+1)\Gamma(1-d-j)}$$

with $-0.5 < d < 0.5$*. Then* a_j *are the coefficients in the power series representation*

$$A(z) = (1-z)^{-d} = \sum_{j=0}^{\infty} a_j z^j.$$

Therefore, the spectral density of X_t *is given by*

$$f_X(\lambda) = \frac{\sigma_\epsilon^2}{2\pi}|A(e^{-i\lambda})|^2 = \frac{\sigma_\epsilon^2}{2\pi}|2(1-cos\lambda)|^{-d} \sim \frac{\sigma_\epsilon^2}{2\pi}|\lambda|^{-2d}.$$

Thus, we obtain short-range dependence for $d = 0$ *(and in fact uncorrelated observations), long-range dependence for* $0 < d < 0.5$*. If* ϵ_t *are independent, then* $\{X_t\}$ *is called a fractional ARIMA*$(0, d, 0)$ *process.*

2.2 Functional data analysis

First, we consider some concepts on a Hilbert space (see Horváth and Kokoszka 2012 pages 21 and 22).

Let $\mathbb{H}$ be a separable Hilbert space (i.e. a Hilbert space with a countable basis $\{e_i, i \in \mathbb{Z}\}$) with inner product $\langle \cdot, \cdot \rangle$ which generates the norm $\| \cdot \|$.

Definition 2.6. *Denote by* $\mathcal{L}$ *the space of bounded linear operators on* $\mathbb{H}$ *with the norm*

$$\|\mathbf{A}\|_{\mathcal{L}} = \sup_{x \in \mathbb{H}, \|x\| \le 1} \{\|\mathbf{A}(x)\|\}.$$

An operator $\mathbf{A} \in \mathcal{L}$ *is said to be compact if there exists two orthonormal bases* $\{u_l\}$ *and* $\{v_l\}$*, and a real (positive) sequence* $\{\lambda_l\}$ *converging to zero, such that*

$$\mathbf{A}(x) = \sum_{l=1}^{\infty} \lambda_l \langle x, u_l \rangle v_l \quad (x \in \mathbb{H}).$$

Remark 2.1. *(1) The* λ_j *are assumed positive because one can replace* v_j *by* $-v_j$ *if needed.*

(2) The representation in Definition 2.6 is called the singular value decomposition, where $\{\lambda_l\}$ *are called the singular values of the operator* $\mathbf{A}$.

Definition 2.7. *A compact operator* $\mathbf{A} \in \mathcal{L}$ *admitting above representation in Definition 2.6 is said to be a Hilbert-Schmidt operator if*

$$\sum_{l=1}^{\infty} \lambda_l^2 < \infty.$$

Remark 2.2. *The space* $\mathcal{S}$ *of Hilbert-Schmidt operators is a separable Hilbert space with the scalar product*

$$\langle \mathbf{A}_1, \mathbf{A}_2 \rangle_{\mathcal{S}} = \sum_{i=1}^{\infty} \langle \mathbf{A}_1(e_i), \mathbf{A}_2(e_i) \rangle \quad (\mathbf{A}_1, \mathbf{A}_2 \in \mathcal{S}),$$

where $\{e_i\}$ *is an arbitrary orthonormal basis.*

Definition 2.8. *An operator* $\mathbf{A} \in \mathcal{L}$ *is said to be symmetric if*

$$\langle \mathbf{A}(x), y \rangle = \langle x, \mathbf{A}(y) \rangle \quad (x, y \in \mathbb{H}),$$

and positive-definite if

$$\langle \mathbf{A}(x), x \rangle \geq 0 \quad (x \in \mathbb{H}).$$

Remark 2.3. *A symmetric positive-definite Hilbert-Schmidt operator* $\mathbf{A}$ *admits the decomposition, by Hilber-schmidt theorem,*

$$\mathbf{A}(x) = \sum_{l=1}^{\infty} \lambda_l \langle x, \phi_l \rangle \phi_l, \quad x \in \mathbb{H}$$

with orthonormal ϕ_l *which are the eifenfunctions of* $\mathbb{A}$, *i.e.* $\mathbb{A}(\phi_l) = \lambda_l \phi_l$, *and the corresponding eigenvalues* λ_l *be positive and which converge to zero.*

Definition 2.9. *Denote by* $L^2 = L^2([0,1])$ *the separable Hilbert space of measurable real-valued square integrable functions on* $[0,1]$ *with the inner product, for* $x, y \in L^2$,

$$\langle x, y \rangle = \int_0^1 x(t) y(t) dt.$$

Remark 2.4. *In this thesis, we consider the random function X defined on some probability space $(\Omega, \mathcal{F}, P)$, which can be viewed as a random element of L^2 equipped with the Borel σ-algebra $\mathcal{B}_{L^2}$, i.e. at time $t \in [0,1]$*

$$X : (\Omega, \mathcal{F}, P) \to (L^2, \mathcal{B}_{L^2})$$
$$w \mapsto X(t,w)$$

Now, the definition of mean and covariance of random elements in L^2 can be given as follows (see Horváth and Kokoszka 2012 pages 23 and 24 or Bosq 2000 page 18).

Definition 2.10. *Let $X = \{X(t), t \in [0,1]\}$ be a random function in L^2. If X is integrable i.e.*

$$E\|X\| = E\left[\int_0^1 X^2(t)dt\right]^{1/2} < \infty$$

for all $t \in [0,1]$, then the mean function of $X(t)$ is defined by the unique function $\mu(t) \in L^2$ such that

$$E\langle y, X\rangle = \langle y, \mu\rangle$$

for any $y \in L^2$. It follows that $\mu(t) = E[X(t)]$ for almost all $t \in [0,1]$.

Definition 2.11. *Let $X = \{X(t), t \in [0,1]\}$ be a random function in L^2. If X is square integrable, i.e.*

$$E\|X\|^2 = E\left[\int_0^1 X^2(t)dt\right] < \infty,$$

and $E[X(t)] = \mu(t)$ the covariance operator of X is defined by

$$\mathbf{C}(y) = E[\langle (X-\mu), y\rangle (X-\mu)] \quad y \in L^2$$

and the covariance function of X is defined by

$$C(s,t) = E[(X(s) - \mu(s))(X(t) - \mu(t))].$$

Remark 2.5. *(1) It is easy to see that $\mathbf{C}(\cdot)$ is an integral operator and the covariance function $C(s,t)$ is its kernel, i.e.*

$$\mathbf{C}(y)(t) = \int_0^1 C(s,t)y(s)ds.$$

(2) The covariance operator $\mathbf{C}(\cdot)$ *can be viewed as a bounded symmetric positive-definite Hilbert-Schmidt operator on* L^2,

$$\mathbf{C} : L^2 \to L^2$$
$$X(t,w) =: x_1(t) \mapsto \mathbf{C}(x_1)(t) := E[\langle (X(s)-\mu(s)), x_1(s)\rangle (X(t)-\mu(t))]$$

The following Mercer's theorem is frequently used in this thesis, which is a representation of a symmetric positive-definite function on a square as a sum of a convergent sequence of product functions. See for instance Horváth and Kokoszka (2012) page 23 or Bosq (2000) pages 24 and 25.

Theorem 2.1. *Let* $C(s,t)$ *be a covariance function continuous on* $[0,1]^2$ *and* $\mathbf{C}(\cdot)$ *be the corresponding integral operator. Then there exists a sequence* $\{\phi_l(t)\}$ *of continuous functions on* $[0,1]$ *and a decreasing sequence* $\{\lambda_l\}$ *of nonnegative numbers such that*

$$\mathbf{C}(\phi_l)(t) = \int_0^1 C(s,t)\phi_l(s)ds = \lambda_l \phi_l(t) \quad (t \in [0,1], l \in \mathbb{N}),$$

and

$$\int_0^1 \phi_{l_1}(t)\phi_{l_2}(s)ds = \delta_{l_1,l_2} = \begin{cases} 1, & l_1 = l_2 \\ 0, & l_1 \neq l_2 \end{cases} \qquad (l_1, l_2 \in \mathbb{N}).$$

Moreover,

$$C(s,t) = \sum_{l=1}^{\infty} \lambda_l \phi_l(s)\phi_l(t) \quad (s,t \in [0,1]), \tag{2.1}$$

where the series converges uniformly on $[0,1]^2$*; hence*

$$\sum_{l=1}^{\infty} \lambda_l = \int_0^1 C(t,t)dt < \infty.$$

Remark 2.6. *(1)* $\{\phi_l(t)\}$ *and* $\{\lambda_l\}$ *are the corresponding eigenfunctions and eigenvalues of the covariance operator* $\mathbf{C}$.

(2) An operator $\mathbf{C} \in \mathcal{L}(L^2)$ *is a covariance operator if and only if it is compact, symmetric, positive-definite and the sum of its eigenvalues is finite, i.e.* $\sum_{l=1}^{\infty} \lambda_l < \infty$.

The following Karhunen-Loève (K.L.) expansion provides an explicit form of the random curve (see Horváth and Kokoszka 2012 page 25 or Bosq 2000 page 25). It is interesting in itself and will be used in the sequel of this thesis.

Theorem 2.2. *Let* $X = \{X(t), t \in [0,1]\}$ *be zero mean measurable random processes in* L^2 *with continuous covariance function* $C(s,t)$*. Then*

$$X(t) = \sum_{l=1}^{\infty} \xi_l \phi_l(t) \quad (t \in [0,1]), \tag{2.2}$$

where $\{\xi_l\}$ *is a sequence of real valued zero mean random variables such that*

$$E[\xi_{l_1}\xi_{l_2}] = \lambda_{l_1}\delta_{l_1,l_2} \quad (l_1, l_2 \in \mathbb{N}),$$

and where the sequences $\{\phi_l(t)\}$ *and* $\{\lambda_l\}$ *are the eigenfunctions and eigenvalues of* $C(s,t)$ *as defined in the Mercer's Theorem 2.1. The series in (2.2) converges uniformly with respect to the* $L^2(\Omega, \mathcal{F}, P)$*-norm.*

Remark 2.7. *In this thesis, in order to focus on the main quantity, we will assume that* $\{\xi_l\}$ *(*$l \in \mathbb{N}$*) are Gaussian random variables.*

Chapter 3

Estimation of trend $\mu(t)$ and covariance $C(s,t)$

In this chapter, we consider the estimation of trend function $\mu(t)$ and covariance function $C(s,t)$ in repeated time series models typically encountered in functional data analysis (FDA) with the modification that the random curves are perturbed by random noise. The random noise processes may exhibit short- or long-range dependence. The one-dimensional boundary kernel estimation of $\mu(t)$ and two-dimensional boundary kernel estimation of $C(s,t)$ are defined. Functional central limit theorem of the estimated trend and estimated covariance is established. Since $\mu(t)$ plays the role of a nuisance parameter that is of no interest when the focus is on $C(s,t)$ only, it is wise to eliminate it before estimating $C(s,t)$. The elimination of $\mu(t)$ can be done without any asymptotic loss of efficiency by using orthonormal contrast transformations. In the sense that, under the Gaussian assumption, the contrast transformed model is equivalent in distribution to the original model. In order to relax the restriction between the number of random curves and the number of sampling points on each curve, higher order kernels are used and additional differentiability assumptions on covariance are imposed while estimating $C(s,t)$. A simple simulation example is given to illustrate the derived asymptotic properties of estimated trend function. This chapter is based on our previous results in Beran and Liu (2014).

3.1 Models and estimators

In this section, we give a detailed explanation of the FDA model with dependent errors. Moreover, the one-dimensional boundary kernel estimation of $\mu(t)$ and two-dimensional boundary kernel estimation of $C(s,t)$ are defined.

3.1.1 Models

We consider the FDA model with long memory errors which is briefly explained in Chapter 1.

The observations which consist of n independent time series $Y_{i\cdot} = (Y_{i1}, ..., Y_{iN})$ $(i = 1, ..., n)$ are defined by

$$Y_{ij} = X_i(t_j) + \epsilon_i(j) \quad (i = 1, ..., n; j = 1, ..., N) \tag{3.1}$$

where $t_j = j/N$ denotes rescaled time, n is number of random curves, N is the number of sampling points on each curve (as mentioned in Chapter 1, in order to focus on the essential effect of dependent structure in error processes, we only consider the equidistant case), $\{\epsilon_i(j)\}_{j\in\mathbb{N}}$ is a random error process.

The random curves $X_i(t)$ are assumed to come from an underlying random process $X(t) \in L^2[0,1]$ with continuous expectation $\mu(t) = E[X(t)]$ and continuous covariance $C(s,t) = cov(X(s), X(t))$.

By Mercer's theorem (Theorem 2.1), $C(s,t)$ can be written as

$$C(s,t) = \sum_{l=1}^{\infty} \lambda_l \phi_l(s)\phi_l(t) \quad (s, t \in [0,1]) \tag{3.2}$$

where $\{\lambda_l\}$, $\{\phi_l(t)\}$ $(l \in \mathbb{N})$ denote the eigenvalues and the corresponding eigenfunctions (functional principal components) of the covariance operator $\mathbf{C}(y) = E\left[\langle (X-\mu), y\rangle (X-\mu)\right]$ $(y \in L^2[0,1])$, $\{\lambda_l\}$ be a decreasing sequence of nonnegative numbers with $\sum \lambda_l < \infty$, $\{\phi_l(t)\}$ are the continuous real-valued functions on $[0,1]$ which build an orthonormal $L^2[0,1]$-basis.

By Karhunen-Loève (K.L.) expansion (Theorem 2.2), the random curve $X_i(t)$ has the form

$$X_i(t) = \mu(t) + \sum_{l=1}^{\infty} \xi_{il}\phi_l(t) \quad (t \in [0,1]) \tag{3.3}$$

where the coefficients $\{\xi_{il}\}$ (functional principal scores) are random variables such that

$$E[\xi_{il}] = 0$$

and

$$E[\xi_{il_1}\xi_{il_2}] = \lambda_{l_1}\delta_{l_1,l_2},$$

$\{\epsilon_i(j)\}$ $(j \in \mathbb{N})$ are independent of $\{\epsilon_{i'}(j)\}$ $(j \in \mathbb{N})$ for $i \neq i'$, $\{\xi_{il}\}$ $(l \in \mathbb{N})$ are independent of $\{\epsilon_{i'}(j)\}$ $(j \in \mathbb{N})$ for all $i, i' \in \{1, ..., n\}$.

In contrast to classical FDA, the error processes $\{\epsilon_i(j)\}$ $(j \in \mathbb{N})$ will be assumed to be existing and be dependent. Instead, asymptotic results for estimators of $\mu(t)$, $C(s,t)$ (Chapter 3), λ_l, $\phi_l(t)$, ξ_{il} (Chapter 4) and two functional sample inference (Chapter 5) will be derived under the condition that $\{\epsilon_i(j)\}$ $(j \in \mathbb{N})$ are stationary Gaussian and short- or long-range dependent. This means that the spectral density of $\{\epsilon_i(j)\}$ $(j \in \mathbb{N})$ has the form

$$f(\lambda) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_\epsilon(k)e^{ik\lambda} \underset{|\lambda|\to 0}{\sim} c_f|\lambda|^{-2d} \tag{3.4}$$

for some constants $0 < c_f < \infty$ and $d \in [0, \frac{1}{2})$, where "$\sim$" means that the ratio of the left and right hand side tends to one. For $d = 0$, (3.4) implies short memory characterized by $0 < \sum_{k\in\mathbb{Z}}\gamma_\epsilon(k) < \infty$, whereas for $0 < d < \frac{1}{2}$ we have long memory with non-summable autocovariances of the form

$$\gamma_\epsilon(k) = cov\left(\epsilon_i(j), \epsilon_i(j+k)\right) \underset{k\to\infty}{\sim} c_\gamma|k|^{2d-1} \tag{3.5}$$

for some constant $0 < c_\gamma < \infty$ (as discussed in Chapter 1).

More specifically, the following assumptions and conditions on the model will be used:

- (A1) The eigenfunctions of the covariance $\{\phi_l(t)\}$ $(l \in \mathbb{N})$ are uniformly continuous and consist an orthonormal basis of $L^2[0,1]$.
- (A2) The functional principal component scores $\{\xi_{il}\}$ $(i, l \in \mathbb{N})$ are assumed to be independent Gaussian variables with $E[\xi_{il}] = 0$, $E[\xi_{il}^2] = \lambda_l \geq 0$ and $\sum\lambda_l < \infty$.

- (A3) For each i, $\{\epsilon_i(j)\}$ $(j \in \mathbb{N})$ is assumed to be a stationary Gaussian process with zero mean and spectral density $f_\epsilon(\lambda)$ satisfying (3.4) or autocovariance $\gamma_\epsilon(k)$ satisfying (3.5), which means that the errors are short memory processes for $d = 0$ and long memory processes for $d \in \left(0, \frac{1}{2}\right)$.

- (A4) The error processes $\{\epsilon_i(j)\}$ $(j \in \mathbb{N})$ and $\{\epsilon_{i'}(j)\}$ $(j \in \mathbb{N})$ are assumed to be independent for $i \neq i'$, and the functional principal scores $\{\xi_{il}\}$ $(l \in \mathbb{N})$ are assumed to be independent of the error processes $\{\epsilon_{i'}(j)\}$ $(j \in \mathbb{N})$ for all $i, i' \in \{1, ..., n\}$.

- (M1) The mean of the unknown underlying random function $\mu(t)$ is assumed to be twice continuously differentiable on $[0,1]$, i. e. $\mu \in C^2[0,1]$.

- (M2) The covariance of the unknown underlying random function $C(s,t)$ is assumed to be twice continuously differentiable on $[0,1]^2$, i. e. $C \in C^2[0,1]^2$.

3.1.2 Definition of the estimators

Let the observations Y_{ij} $(i = 1, ..., n,\ j = 1, ..., N)$ be given by (3.1), (3.3) and (3.4) or (3.5). Now we give the definition of the one-dimensional kernel estimator of $\mu(t)$ and the two-dimensional kernel estimator of $C(s,t)$.

Let $K_1(u)$, $K_2(u,v)$ be two kernel functions with support $[-1,1]$ and $[-1,1]^2$ respectively. By using the notation $\bar{y}_{\cdot j} = n^{-1}\sum_{i=1}^{n} Y_{ij}$, the one-dimensional kernel estimator of $\mu(t)$ based on kernel $K_1(u)$ is defined by

$$\hat{\mu}(t) = \frac{1}{Nb}\sum_{j=1}^{N} K_1\left(\frac{t - t_j}{b}\right)\bar{y}_{\cdot j}. \tag{3.6}$$

Since the bandwidth b will be a function of N, we may write $b = b_N$.

For the covariance function of each time series $Y_{i\cdot} = (Y_{i1}, ..., Y_{iN})^T$ $(i = 1, ..., n)$, note that, under the independent assumption of ξ_{il} and $\epsilon_i(j)$, we have

$$cov(Y_{ij}, Y_{ik}) = cov\left(X(t_j), X(t_k)\right) + cov(\epsilon_i(j), \epsilon_i(k)),$$

and

$$E\left[(Y_{ij} - \hat{\mu}(t_j))(Y_{ik} - \hat{\mu}(t_k))\right] \approx cov(X(t_j), X(t_k)) + cov(\epsilon_i(j), \epsilon_i(k)).$$

Using the notation

$$C_{ijk} = (Y_{ij} - \hat{\mu}(t_j))(Y_{ik} - \hat{\mu}(t_k))$$

the two-dimensional kernel estimator of $C(s,t)$ based on kernel $K_2(u,v)$ can be defined as

$$\begin{aligned} \hat{C}(s,t) &= \frac{1}{(Nb)^2} \sum_{j,k=1}^{N} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) n^{-1} \sum_{i=1}^{n} C_{ijk} \qquad (3.7) \\ &= \frac{1}{(Nb)^2} \sum_{j,k=1}^{N} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) \bar{C}_{\cdot jk}, \end{aligned}$$

where $\bar{C}_{\cdot jk} = n^{-1} \sum_{i=1}^{n} C_{ijk}$.

In order to deal with the boundary effect which is typically encountered in kernel regression, the kernel estimator $\hat{\mu}(t)$ given by (3.6) will be modified by using boundary kernels. Specifically, we use left boundary kernel $K_{left}^{(c)}(t)$ for boundary points $t = cb \in [0, b)$ (with $0 \leq c < 1$) and right boundary kernels $K_{right}^{(c)}(t)$ for boundary points $t = 1 - cb \in (1 - b, 1]$ (with $0 \leq c < 1$). The definition of the boundary kernel estimation will be given below in equation (3.10). Boundary kernels are used to make sure that the bias $B_{N,b}(t) = E[\hat{\mu}(t)] - \mu(t)$ at boundary points $t \in [0, b) \cup (1 - b, 1]$ and that at interior points $t \in [b, 1 - b]$ have the same order $O(b^2)$ (without boundary correction the order is $O(b)$). For details on boundary kernels one can refer for instance Gasser and Müller (1979), Gasser et al. (1985), Müller (1991), Müller and Wang (1994) and Beran and Feng (2002c).

Specifically the following assumptions on the kernels will be used:

- (K1) Let $K_1(t)$ be a symmetric probability density function with support $[-1, 1]$ such that

$$0 < \beta_1 = \int_{-1}^{1} K_1(t) t^2 dt < \infty. \qquad (3.8)$$

- (K2) Let

$$K_{left}^{(1)}(t) = K_1(t).$$

  Let $K_{left}^{(c)}(t)$ ($c \in [0, 1)$) denote functions with support $[-1, c]$ such that

$$\int_{-1}^{c} K_{left}^{(c)}(t) dt = 1, \; \int_{-1}^{c} K_{left}^{(c)}(t) t dt = 0, \; \beta_c := \int_{-1}^{c} K_{left}^{(c)}(t) t^2 dt \neq 0. \qquad (3.9)$$

Moreover, $K^{(c)}_{right}(t)$ ($c \in [0,1)$) are functions with support $[-c,1]$ defined by

$$K^{(c)}_{right}(t) = K^{(c)}_{left}(-t).$$

- (K3) The one-dimensional boundary kernels $K_{1,b}$ ($b \in \left(0, \frac{1}{2}\right)$) are functions with support $[-1,1]$ defined by

$$\begin{aligned}
K_{1,b}(t) &= K_1(t) \quad (t \in [b, 1-b]), \\
K_{1,b}(t) &= K^{(c)}_{left}(t) \quad (t = cb,\ 0 \le c < 1), \\
K_{1,b}(t) &= K^{(c)}_{right}(t) \quad (t = 1 - cb,\ 0 \le c < 1), \\
K_{1,b}(t) &= 0 \quad (|t| > 1).
\end{aligned}$$

- (K4) For any $c \in [0,1)$, $K^{(c)}_{left}(t)$ and $K^{(c)}_{right}(t)$ are almost everywhere continuously differentiable.

- (K5) There exists a positive and finite constant C_K such that for all $c \in [0,1)$,

$$\left\| K^{(c)}_{left}(t) \right\|^2 = \int_{-1}^{1} \left[K^{(c)}_{left}(t) \right]^2 dt < C_K.$$

- (K6) The two-dimensional boundary kernel $K_{2,b}(s,t)$ is defined to be the product of $K_{1,b}(t)$

$$K_{2,b}(s,t) = K_{1,b}(s) K_{1,b}(t).$$

The one-dimensional boundary kernel estimator of $\mu(t)$ based on $K_{1,b}$ is defined by:

$$\hat{\mu}(t) = \frac{1}{Nb} \sum_{j=1}^{N} K_{1,b}\left(\frac{t - t_j}{b}\right) \bar{y}_{\cdot j}. \tag{3.10}$$

The two-dimensional boundary kernel estimator of $C(s,t)$ based on $K_{2,b}$ is defined by:

$$\hat{C}(s,t) = \frac{1}{(Nb)^2} \sum_{j,k=1}^{N} K_{2,b}\left(\frac{s - t_j}{b}, \frac{t - t_k}{b}\right) \bar{C}_{\cdot jk}. \tag{3.11}$$

Define the following notation

$$\beta(t) = \beta_1 \cdot \mathbf{1}\{0 < t < 1\} + \beta_0 \cdot \mathbf{1}\{t \in \{0,1\}\} \tag{3.12}$$

where β_1 and β_0 are defined in (3.8) and (3.9) respectively. This notation will be used in calculating the asymptotic approximation for the bias of $\hat{\mu}(t)$ and $\hat{C}(s,t)$.

Due to $b_N \to 0$ as $N \to \infty$, t is an interior point asymptotically except for $t = 0, 1$. This is the reason that we distinguish $t \in (0,1)$ from $t \in \{0,1\}$.

For $0 < d < \frac{1}{2}$, we will use the following several notations

$$V_{d,interior} = \int_{-1}^{1}\int_{-1}^{1} |x-y|^{2d-1} K_1(x)K_1(y)dxdy,$$
$$V_{d,1} = \int_{-1}^{1}\int_{-1}^{1} |x-y|^{2d-1} K_{left}^{(0)}(x)K_1(y)dxdy,$$
$$V_{d,2} = \int_{-1}^{1}\int_{-1}^{1} |x-y|^{2d-1} K_{left}^{(0)}(x)K_{left}^{(0)}(y)dxdy,$$
$$V_{d,3} = \int_{-1}^{1}\int_{-1}^{1} |x-y|^{2d-1} K_{left}^{(0)}(x)K_{right}^{(0)}(y)dxdy,$$

and

$$\begin{aligned} V_d(s,t) =& V_{d,interior}\mathbf{1}\{0 < s,t < 1\} + V_{d,1}\mathbf{1}\{(s,t) \text{ or } (t,s) \text{ in } \{0,1\}\times(0,1)\} \\ &+ V_{d,2}\mathbf{1}\{(s,t) = (0,0) \text{ or } (1,1)\} + V_{d,3}\mathbf{1}\{(s,t) = (0,1) \text{ or } (1,0)\}. \end{aligned}$$

This notation $V_d(s,t)$ $(s,t \in [0,1])$ will be used in calculating the asymptotic approximation for the variance of $\hat{\mu}(t)$ and $\hat{C}(s,t)$. Note that, if s and/or t are equal to zero or one, then boundary kernels with $c = 0$ will be used. This is the reason for distinguishing the four cases. However, this occurs only at the very edge of the region $[0,1]^2$, i.e. for $s, t \notin (0,1)$. Since $b_N \to 0$ as $N \to \infty$, all other points (i.e. $s, t \in (0,1)$) are asymptotically interior points.

3.2 Asymptotic distribution of $\hat{\mu}(t)$

This section considers the asymptotic expectation, asymptotic variance and weak convergence of $\hat{\mu}(t)$.

3.2.1 Expected value and variance of $\hat{\mu}(t)$

Note that the bias of $\hat{\mu}(t)$ is a linear function of the observations $Y_{ij} = \mu(t_j) + \sum_{l=1}^{\infty} \xi_{il}\phi_l(t_j) + \epsilon_i(j)$ given a one-dimensional boundary kernel function $K_{1,b}(t)$ and a bandwidth b. Note that $E[Y_{ij}] = \mu(t_j)$, the bias of $\hat{\mu}(t)$ is equal to

$$B_{N,b}(t) = E[\hat{\mu}(t)] - \mu(t) = \frac{1}{Nb}\sum_{j=1}^{N} K_{1,b}\left(\frac{t-t_j}{b}\right)\mu(t_j) - \mu(t).$$

Thus, given N and a corresponding bandwidth b, $B_{N,b}(t)$ does not depend on the random functions $\sum_{l=1}^{\infty} \xi_{il}\phi_l(t)$ and the dependence structure of $\epsilon_i(j)$. Therefore, the asymptotic expression for $B_{N,b}(t)$ follows by standard arguments and has the order $O(b^2)$. For the variance of $\hat{\mu}(t)$ one can see that, under the assumption that $n \to \infty$, the dominating term is in order $O\left(n^{-1}\right)$ and is not influenced asymptotically by the dependence structure of $\epsilon_i(j)$. More specifically, these asymptotic properties can be summarized as the following theorem.

Theorem 3.1. *Let Y_{ij} be defined by (3.1), (3.3) and (3.4) or (3.5). Assume that (A1), (A2), (A3), (A4), (K1), (K2), (K3), (K5)and (M1) hold. The estimation of trend function is defined in (3.10). Moreover, let*

$$n \to \infty,\ N \to \infty,\ b \to 0,\ Nb^3 \to \infty. \tag{3.13}$$

Then, for any $t \in [0,1]$,

$$\begin{aligned} E[\hat{\mu}(t)] - \mu(t) &= C_{bias}(t)b^2 + o(b^2) + O\left((Nb)^{-1}\right) \\ &= C_{bias}(t)b^2 + o(b^2), \\ C_{bias}(t) &= \frac{1}{2}\mu''(t)\beta(t), \end{aligned}$$

with $\beta(t)$ defined in (3.12) and all $O(\cdot)$ and $o(\cdot)$ terms uniform in $t \in [0,1]$. Furthermore,

- *if $0 < d < \frac{1}{2}$, then*

$$\begin{aligned} var[\hat{\mu}(t)] &= n^{-1}C_{var}(t)\left[1 + O\left((Nb)^{2d-1}\right) + O\left(b^2\right) + O\left((Nb)^{-1}\right)\right] \\ &= n^{-1}C_{var}(t)\left[1 + O\left((Nb)^{2d-1}\right) + O\left(b^2\right)\right], \end{aligned}$$

- *if $d = 0$, then*

$$var[\hat{\mu}(t)] = n^{-1}C_{var}(t)\left[1 + O\left((Nb)^{-1}\right) + O\left(b^2\right)\right],$$

with

$$C_{var}(t) = \sum_{l=1}^{\infty} \lambda_l \phi_l^2(t),$$

where all $O(\cdot)$ and $o(\cdot)$ terms are uniform in $t \in [0,1]$.

Remark 3.1. *The arguments in Theorem 3.1 can easily be extended to higher order kernel estimator. For estimating of derivatives of $\mu(t)$, one can also easily obtain the corresponding results. Moreover, the corresponding results can be extended to local polynomial estimators due to their representation based on asymptotically equivalent kernels (see e.g. Gasser et al. 1985, Lejeune 1985, Müller 1987, 1988, Lejeune and Sarda 1992, Fan 1992, Ruppert and Wand 1994, Feng 1999, 2004).*

Remark 3.2. *Theorem 3.1 can also be extended to the case of a nonequidistant design, provided that $\hat{\mu}(t)$ is replaced either by the Nadaraya-Watson estimator*

$$\hat{\mu}(t) = \left[\sum_{j'=1}^{N} K_{1,b}\left(\frac{t - t_{j'}}{b}\right)\right]^{-1} \sum_{j=1}^{N} K_{1,b}\left(\frac{t - t_j}{b}\right) \bar{y}_{\cdot j}.$$

or a local polynomial estimator (see e.g. Fan and Gijbels 1996).

3.2.2 Weak convergence of $\hat{\mu}(t)$ in $C[0,1]$

We first define two sequences of processes:

$$Z^0_{n,N}(t) = \sqrt{n}\left(\hat{\mu}(t) - E[\hat{\mu}(t)]\right) \quad (t \in [0,1],\, n, N \in \mathbb{N})$$

and

$$Z_{n,N}(t) = \sqrt{n}\left(\hat{\mu}(t) - \mu(t)\right) \quad (t \in [0,1],\, n, N \in \mathbb{N}).$$

Under additional assumptions on the sequence of bandwidth b, weak convergence of $\hat{\mu}(t)$ in $C[0,1]$ in the supremum norm sense can be obtained as follows.

Theorem 3.2. *Suppose that (A1), (A2), (A3), (A4), (K1), (K2), (K3), (K4), (K5) and (M1) hold. The estimation of trend function is defined in (3.10). Furthermore let*

$$n \to \infty,\, N \to \infty,\, b = b_N \to 0 \tag{3.14}$$

such that

$$\liminf N b^{1+2/(1-2d)} > q \tag{3.15}$$

for a suitable constant $q > 0$. Then,

$$Z^0_{n,N}(t) \Rightarrow Z(t) = \sum_{l=1}^{\infty} \sqrt{\lambda_l}\phi_l(t)\zeta_l$$

where "$\Rightarrow$" denotes weak convergence in $C[0,1]$ equipped with the supremum norm and ζ_l are iid $N(0,1)$ random variables. If in addition

$$n = n_N = o\left(N^{4\frac{1-2d}{3-2d}}\right) \tag{3.16}$$

and b_N (which satisfies (3.15)) is such that

$$n_N b_N^4 \to 0, \tag{3.17}$$

then

$$Z_{n,N}(t) \Rightarrow Z(t) = \sum_{l=1}^{\infty} \sqrt{\lambda_l}\phi_l(t)\zeta_l.$$

Condition (3.15) is required in proving the weak convergence of the error term. Note that, for $0 \leq d < \frac{1}{2}$, we have $1 + 2/(1-2d) \geq 3$. Therefore, condition (3.15) implies $Nb^3 \to \infty$ as stated in condition (3.13) in Theorem 3.1. Condition (3.17) is only required to make sure that the bias of $\hat{\mu}(t)$ is in order $o\left(n^{-\frac{1}{2}}\right)$. Combining conditions (3.15) and (3.17), we have

$$CN^{-\frac{1-2d}{3-2d}} \leq b << n^{-\frac{1}{4}} \tag{3.18}$$

for some constant $C > 0$. In order that (3.18) can be fulfilled by a sequence of bandwidths b_N we need (3.16).

Condition (3.16) means that the number of repeated time series n cannot grow too fast compared to the number of sampling points on each curve N. Clearly, since the upper bound of n is very small when $d \to \frac{1}{2}$, condition (3.16) is an unpleasant condition. To avoid (3.16) one may prefer to find suitable bias correction methods to relax condition (3.17). Alternatively, one may find ways to eliminate $\mu(t)$ before calculating other statistics if it plays the role of a nuisance parameter. This is the case when our focus is on $C(s,t)$ only. Therefore before discussing asymptotic properties of $\hat{C}(s,t)$ we will discuss a simple method for eliminating $\mu(t)$ in the next section.

Remark 3.3. *We mention the asymptotic results on kernel estimation for single time series $Y_j = X(t_j) + \epsilon(j)$ $(j = 1, 2, ..., N)$. As reported by Csörgö and Mielniczuck (1995), for any fixed rescaled time points $0 < t_1 < ... < t_m < 1$, the standardized variables $Z_N^0(t_k) := \dfrac{\hat{\mu}(t_k) - E[\hat{\mu}(t_k)]}{\sqrt{var(\hat{\mu}(t_k))}}$ $(k = 1, 2, ..., m)$ converge in* *distribution to independent standard normal random variables. Therefore, it is impossible to establish weak convergence of the corresponding sequence of continuous time processes $Z_N^0(t)$ $(t \in (0,1))$ in supremum norm sense.*

Remark 3.4. *Suppose that the observations consist of only one time series Y_j $(j = 1, 2, ..., N)$. This corresponds to setting $n = 1$ in our case, i.e. $Y_j = Y_{1j} = X_1(t_j) + \epsilon_1(j)$. Since in this case we only observe one (random) curve $X_1(t) = \mu(t) + \sum_{l=1}^{\infty} \xi_{1l}\phi_l(t)$, thus $\sum_{l=1}^{\infty} \xi_{1l}\phi_l(t)$ and $X_1(t)$ can be considered as deterministic trend functions. So $\hat{\mu}(t)$ is an estimator of the trend $X_1(t) = \mu(t) + \sum_{l=1}^{\infty} \xi_{1l}\phi_l(t)$ instead of $\mu(t)$. Since only one time series is observed (i.e. $n = 1$ is fixed), asymptotic properties can be derived for $N \to \infty$ only. The limiting behavior is thus determined by the dependence structure of $\epsilon(j)$.*

Remark 3.5. *However, the results differ completely when n $(n \to \infty)$ independent replicates are observed. Note that the replicates are independent, $\epsilon_i(j) + \sum_{l=1}^{\infty} \xi_{il}\phi_l(t_j)$ consists the (zero mean) random error of Y_{ij} $(i = 1, ..., n,\ j = 1, ..., N)$. Therefore, the asymptotic distribution of $Z_{n,N}^0(t)$ is dominated by the contribution of $\sum_{l=1}^{\infty} \xi_{il}\phi_l(t_j)$ if (3.14) and (3.15) hold. This is the reason why, in contrast to the case with $n = 1$, we can obtain weak convergence of $Z_{n,N}^0(t)$ and $Z_{n,N}(t)$ $(t \in [0,1])$.*

Remark 3.6. *It should be emphasized that there is no restriction on n for the process $Z_{n,N}^0(t)$. In fact, the influence of the dependence structure of $\epsilon(j)$ on the limit theorem for $Z_{n,N}^0(t)$ is that b_N must satisfy (3.15). However, condition (3.15) has nothing to do with n. Therefore, the conditions (3.14) and (3.15) can always be achieved without any restriction on n and there is no restriction on n for the process $Z_{n,N}^0(t)$.*

Remark 3.7. *It may seem that the mean squared error $E\left[(\hat{\mu}(t) - \mu(t))^2\right]$ can be reduced arbitrarily by increasing n only. Obviously, one can reduce the variance part by increasing n since it is in order $O\left(n^{-1}\right)$. However, the bias part which is in order $O\left(b_N^2\right)$ depends on b_N and not on n. Since (3.14) and (3.15) imply that b_N cannot be chosen arbitrarily small compared to N, the bias cannot be reduced* *just by increasing n only. In fact, from*

$$Z_{n,N}(t) = Z^0_{n,N}(t) + \sqrt{n}\left[C_{bias}(t)b_N^2 + o(b_N^2)\right],$$

one can see that the sequence $\sqrt{n}b_N^2$ has to converge to zero in order to make sure that the bias is asymptotically negligible. This leads to the restriction (3.16).

3.3 Contrast transformation and the asymptotic distribution of $\hat{C}(s,t)$

The orthonormal contrast transformation and the asymptotic expected value, asymptotic variance and the weak convergence of $\hat{C}(s,t)$ are dixcussed in this section.

3.3.1 Orthonormal contrast transformation

The most challenging thing in FDA is dimension reduction. This can be done by considering functional principal component analysis (FPCA) which is based on the covariance function $C(s,t)$. Therefore, the main quantity of interest in FDA is $C(s,t)$. The two-dimensional boundary kernel estimator of $C(s,t)$ defined in (3.11) is based on

$$C_{ijk} = (Y_{ij} - \hat{\mu}(t_j))(Y_{ik} - \hat{\mu}(t_k)).$$

As pointed out in previous section, kernel estimation of $\mu(t)$ leads to an unpleasant restriction (3.16) on n. Since we are interested in $C(s,t)$ only, $\mu(t)$ plays the role of an infinite dimensional nuisance parameter that is of no interest. Therefore, to avoid (3.16) it is wise to eliminate $\mu(t)$ before estimating $C(s,t)$.

The elimination of $\mu(t)$ can be done without any asymptotic loss of efficiency (as $n \to \infty$) by using orthonormal contrast transformations as follows. Let

$$\mathbf{1} = (1, ..., 1)^T \in \mathbb{R}^n$$

and

$$c_{\cdot m} = (c_{1m}, ..., c_{nm})^T \in \mathbb{R}^n, \quad m = 1, ..., n-1$$

be such that

$$\langle \mathbf{1}, c_{\cdot m} \rangle = \sum_{s=1}^{n} c_{sm} = 0$$

and

$$\langle c_{\cdot m}, c_{\cdot m'} \rangle = \sum_{s=1}^{n} c_{sm} c_{sm'} = \delta_{mm'}$$

with $\delta_{mm'}$ denoting the Kronecker delta. This means that $c_{\cdot 1}, ..., c_{\cdot n-1}$ build an orthonormal basis in the subspace which is orthogonal to $\mathbf{1}^T$.

We then define $n-1$ contrast series $Y^c_{m\cdot} = (Y^c_{m1}, ..., Y^c_{mN})$ by

$$Y^c_{mj} = \langle c_{\cdot m}, Y_{\cdot j} \rangle = \sum_{s=1}^{n} c_{si} Y_{sj} \quad (m = 1, ..., n-1; j = 1, ..., N)$$

where $Y_{\cdot j} = (Y_{1j}, ..., Y_{nj})^T$ are the original observations. Note that

$$Y_{ij} = X_i(t_j) + \epsilon_i(j) \quad (i = 1, ..., n; j = 1, ..., N)$$

where $\epsilon_i(j)$ is a random deviation from $X_i(t_j)$ and $X_i(t)$ are random functions defined by

$$X_i(t) = \mu(t) + \sum_{l=1}^{\infty} \xi_{il} \phi_l(t)$$

with random coefficients ξ_{il}. Using the notation $\xi_{\cdot l} = (\xi_{1l}, ..., \xi_{nl})^T$ and $\epsilon_{\cdot}(j) = (\epsilon_1(j), ..., \epsilon_n(j))^T$, then the contrast transformed random coefficients ξ^c_{ml} and contrast transformed errors $\epsilon^c_m(j)$ are

$$\xi^c_{ml} = \langle c_{\cdot m}, \xi_{\cdot l} \rangle = \sum_{s=1}^{n} c_{sm} \xi_{sl}$$

and

$$\epsilon^c_m(j) = \langle c_{\cdot m}, \epsilon_{\cdot}(j) \rangle = \sum_{s=1}^{n} c_{sm} \epsilon_s(j).$$

Therefore we obtain the contrast series

$$Y^c_{mj} = X^c_m(t_j) + \epsilon^c_m(j) \quad (m = 1, \cdots, n-1; j = 1, \cdots, N) \tag{3.19}$$

where $X^c_m(\cdot)$ $(m = 1, ..., n-1)$ are the contrast modified functions

$$X^c_m(t) = \left(\sum_{s=1}^{n} c_{sm} \right) \mu(t) + \sum_{l=1}^{\infty} \xi^c_{ml} \phi_l(t) = \sum_{l=1}^{\infty} \xi^c_{ml} \phi_l(t).$$

Thus the trend $\mu(t)$ is removed.

Moreover, the covariance functions of $X_i(t)$ and $X_m^c(t)$ keep the same. In fact, this can be seen as follows:

$$E[X_m^c(t)X_{m'}^c(t)] = \sum_{l_1,l_2=1}^{\infty} E\left[\xi_{ml_1}^c \xi_{m'l_2}^c\right] \phi_{l_1}(t)\phi_{l_2}(t)$$

with

$$E\left[\xi_{ml_1}^c \xi_{m'l_2}^c\right] = \sum_{s_1,s_2=1}^{n} c_{s_1m}c_{s_2m'}E[\xi_{s_1l_1}\xi_{s_2l_2}] = 0 \quad (l_1 \neq l_2)$$

and

$$E\left[\xi_{ml}^c \xi_{m'l}^c\right] = \sum_{s_1,s_2=1}^{n} c_{s_1m}c_{s_2m'}E[\xi_{s_1l}\xi_{s_2l}] = \sum_{s=1}^{n} c_{sm}c_{sm'}\lambda_l = \langle c_{\cdot m}, c_{\cdot m'}\rangle \lambda_l = \delta_{mm'}\lambda_l.$$

Thus,

$$E\left[X_m^c(t)X_{m'}^c(t)\right] = \delta_{mm'} \sum_{l=1}^{\infty} \lambda_l \phi_l^2(t)$$

which is equal to $E[X_i(t)X_{i'}(t)]$ (see (3.2)). Similarly we can obtain $cov(\epsilon_m^c(j), \epsilon_{m'}^c(j')) = cov(\epsilon_i(j), \epsilon_{i'}(j'))$.

Since $\epsilon_i(j)$ and ξ_{il} are assumed to be Gaussian, the contrast transformed model Y_{mj}^c defined in (3.19) is equivalent in distribution to (i.e. has the same distribution as) the original model defined in Y_{ij} (3.1) with $\mu(t) \equiv 0$, except that n reduces to $n-1$. This equivalence in distribution implies that asymptotic results (as $n \to \infty$) for statistics based on Y_{mj}^c and Y_{ij} with $\mu(t) \equiv 0$ are the same. Therefore, in the following we may assume that the observed series are given as

$$Y_{mj}^c = \sum_{l=1}^{\infty} \xi_{ml}^c \phi_l(t_j) + \epsilon_i^c(j) \underset{d}{=} Y_{ij} := \sum_{l=1}^{\infty} \xi_{il}\phi_l(t_j) + \epsilon_i(j), \tag{3.20}$$

and the covariance estimator (3.11) can be replaced by

$$\hat{C}(s,t) = \frac{1}{(Nb)^2} \sum_{j,k=1}^{N} K_{2,b}\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) n^{-1} \sum_{i=1}^{n} Y_{ij}Y_{ik}. \tag{3.21}$$

3.3.2 Expected value and variance of $\hat{C}(s,t)$

The asymptotic expected value and variance of the two-dimensional boundary kernel estimation of $C(s,t)$ defined by (3.21) are summarized as follows.

Theorem 3.3. *Let $Y_{ij} = Y_{ij}^c$ be defined by (3.20) (and (3.1), (3.3), (3.4) or (3.5)). Assume that (A1), (A2), (A3), (A4), (K1), (K2), (K3), (K5), (K6) and (M2) hold. The estimation of covariance function is defined in (3.21). Moreover, let*

$$n \to \infty,\ N \to \infty,\ b \to 0,\ Nb^3 \to \infty. \tag{3.22}$$

Then the following holds:

- *if $0 < d < \frac{1}{2}$, then for any $s, t \in (0,1)$,*

$$E\left[\hat{C}(s,t)\right] = \sum_{l=1}^{\infty} \lambda_l \phi_l(s)\phi_l(t) + (Nb)^{2d-1} c_\gamma V_d(s,t) + O(b^2) + o\left(\max\left\{(Nb)^{2d-1}, b^2\right\}\right)$$

  and

$$\begin{aligned} var\left[\hat{C}(s,t)\right] &= n^{-1}\left[2\sum_{l=1}^{\infty} \lambda_l^2 \phi_l^2(s)\phi_l^2(t) + \sum_{l_1<l_2} \lambda_{l_1}\lambda_{l_2}\left[\phi_{l_1}(s)\phi_{l_2}(t) + \phi_{l_2}(s)\phi_{l_1}(t)\right]^2\right] \\ &\quad + n^{-1}\left[4(Nb)^{2d-1} c_\gamma V_d(s,t) \sum_{l=1}^{\infty} \lambda_l \phi_l(s)\phi_l(t) + O(b^2) + o\left(\max\{(Nb)^{2d-1}, b^2\}\right)\right], \end{aligned}$$

- *if $d = 0$, then these equations are replaced by*

$$E\left[\hat{C}(s,t)\right] = \sum_{l=1}^{\infty} \lambda_l \phi_l(s)\phi_l(t) + O(b^2)$$

  and

$$\begin{aligned} var\left[\hat{C}(s,t)\right] &= n^{-1}\left[2\sum_{l=1}^{\infty} \lambda_l^2 \phi_l^2(s)\phi_l^2(t) + \sum_{l_1<l_2} \lambda_{l_1}\lambda_{l_2}\left[\phi_{l_1}(s)\phi_{l_2}(t) + \phi_{l_2}(s)\phi_{l_1}(t)\right]^2\right] \\ &\quad + O\left(n^{-1}b^2\right), \end{aligned}$$

where all $O(\cdot)$ and $o(\cdot)$ terms are uniform in $s, t \in [0,1]$.

3.3.3 Weak convergence of $\hat{C}(s,t)$ in $[0,1]^2$

At first we define the following two sequences of processes:

$$Z_{n,N}^0(s,t) = \sqrt{n}\left(\hat{C}(s,t) - E\left[\hat{C}(s,t)\right]\right) \quad (t, s \in [0,1],\ n, N \in \mathbb{N})$$

and

$$Z_{n,N}(s,t) = \sqrt{n}\left(\hat{C}(s,t) - C(s,t)\right) \quad (t,s \in [0,1],\, n, N \in \mathbb{N}).$$

Similar to the estimated trend function $\hat{\mu}(t)$, under additional assumptions, uniform convergence results can be obtained for the estimated covariance function $\hat{C}(s,t)$.

Theorem 3.4. *Let $Y_{ij} = Y_{ij}^c$ be defined by (3.20) (and (3.1), (3.3), (3.4) or (3.5)). Assume that (A1), (A2), (A3), (A4), (K1), (K2), (K3), (K5), (K6) and (M2) hold. The estimation of covariance function is defined in (3.21). Furthermore let*

$$n \to \infty,\ N \to \infty,\ b = b_N \to 0 \tag{3.23}$$

such that

$$\liminf N b^{1+2/(1-2d)} > q \tag{3.24}$$

for a suitable $q > 0$. Then,

$$Z^0_{n,N}(s,t) \Rightarrow Z(s,t) = Z_1(s,t) + Z_2(s,t)$$

where "$\Rightarrow$" denotes weak convergence in $C[0,1]^2$ equipped with the supremum norm and Z_1, Z_2 are zero mean Gaussian processes which are independent from each other, and have covariance functions

$$cov\left(Z_1(t,s), Z_1(t',s')\right) = 2\sum_l \lambda_l^2 \phi_l(s)\phi_l(t)\phi_l(s')\phi_l(t')$$

and

$$cov\left(Z_2(t,s), Z_2(t',s')\right)$$
$$= \sum_{l_1<l_2} \lambda_{l_1}\lambda_{l_2}\left[\phi_{l_1}(s)\phi_{l_2}(t) + \phi_{l_2}(s)\phi_{l_1}(t)\right]\left[\phi_{l_1}(s')\phi_{l_2}(t') + \phi_{l_2}(s')\phi_{l_1}(t')\right].$$

Note that there is no restriction on n for sequence $Z^0_{n,N}(s,t)$, which means that removing $\mu(t)$ by the contrast transformation achieved its purpose. However, the asymptotic bias as given in Theorem 3.3, especially the term of order $O(b^2)$, needs to be considered while dealing with sequence $Z_{n,N}(s,t)$.

Since the bias term of order $O(b^2)$ comes from the properties of $K_{2,b}(u,v)$ and $C(s,t)$ only, the problem of restriction on the growth of n and N can be resolved by using higher order boundary kernels and imposing additional differentiability assumptions on $C(s,t)$. Therefore, we introduce the following conditions:

- (K7) For some $l \in \mathbb{N}$, and all $c \in [0,1]$,

$$\int K_{left}^{(c)}(u)u^j du = 0 \quad (j = 1, 2, ..., 2l-1),$$
$$0 < \int K_{left}^{(c)}(u)u^{2l} du < C_K < \infty$$

  where C_K is a suitable constant.

- (M3) $C(s,t) \in C^{2l+2}[0,1]^2$

It is well known that, under these additional conditions, by using Taylor expansion we have

$$\frac{1}{(Nb)^2}\sum_{j,k} K_{2,b}\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) C(t_j,t_k) = C(s,t) + O\left(b^{2l}\right) + O\left((Nb)^{-1}\right).$$

If we assume

$$Nb^{2l+1} \to \infty,$$

then the error term $O\left((Nb)^{-1}\right)$ is asymptotically negligible compared to $O(b^{2l})$. Together with $nb^{4l} \to 0$, we have

$$N^{-\frac{1}{2l+1}} << b << n^{-\frac{1}{4l}}. \tag{3.25}$$

Condition (3.25) together with condition (3.24) leads to

$$n = o\left(\min\left\{N^{\frac{4l}{2l+1}}, N^{4l\frac{1-2d}{3-2d}}\right\}\right). \tag{3.26}$$

In particular, if $C(s,t)$ is infinitely differentiable, then we may choose l arbitrarily large. For instance, for fixed d, we may choose $l > \frac{1}{1-2d}$, thus we have,

$$n = o\left(N^{\frac{4l}{2l+1}}\right).$$

Obviously, this is a much better bound than $n = o\left(N^{4(1-2d)/(3-2d)}\right)$. Moreover, for fixed d, (3.26) becomes $n = o(N^2)$ for $l \to \infty$. These results can be summarized as the following theorem:

Theorem 3.5. *Let $Y_{ij} = Y_{ij}^c$ be defined by (3.20) (and (3.1), (3.3), (3.4) or (3.5)), and the conditions of Theorem 3.4 hold. In addition, assume (M3) and (K7) with $l \geq 2$. Let*

$$n \to \infty,\ N \to \infty,\ b = b_N \to 0 \tag{3.27}$$

such that

$$Nb^{2l+1} \to \infty,\ nb^{4l} \to 0. \tag{3.28}$$

Then,

$$Z_{n,N}(s,t) \Rightarrow Z(s,t) = Z_1(s,t) + Z_2(s,t)$$

where "$\Rightarrow$" denotes weak convergence in $C[0,1]^2$ equipped with the supremum norm and Z_1, Z_2 are as in Theorem 3.4.

Remark 3.8. *For fixed d, the conditions in Theorem 3.5 are much simpler than in Theorem 3.4. This emphasizes the importance of using higher order kernels whenever (M3) is applicable and the contrast transformation.*

Remark 3.9. *In applications it may be worthwhile applying a simple bias correction of the bias term in the order $O\left((Nb)^{2d-1}\right)$. One can estimation $\hat{d}_i$ and $\hat{c}_{\gamma,i}$ from each independent time series $Y^c_{i\cdot}$ by using one of the well known methods in the literature (see e.g. Giraitis et al. 2012, Chapter 8, and Beran et al. 2013, Chapter 5, for a review on estimation of the long-memory parameter). In fact, consistent estimation is also can be obtained even in the presence of the unknown mean functions $\sum \xi_{il}\phi(t_j)$ (see e.g. Beran and Feng 2002a,b, Beran et al. 2013, Chapter 7.4). The estimation of d and c_γ can be obtained by averaging, i.e. $\hat{d} = n^{-1}\sum \hat{d}_i$, $\hat{c}_\gamma = n^{-1}\sum \hat{c}_{\gamma,i}$.*

Thus, the bias modified estimate can be given as

$$\tilde{C}(s,t) = \hat{C}(s,t) - (Nb)^{2\hat{d}-1}\hat{c}_\gamma V_{\hat{d}}.$$

A similar adjustment has been proposed in by Staniswalis and Lee (1998) and Yao et al. (2003) in the iid context and the authors propose to remove the diagonal terms C_{ijj}. Note that removing the diagonal is not sufficient if the errors exhibt long range dependence. Instead, the complete sum over all covariances - which is asymptotically of the form $(Nb)^{2d-1}c_\gamma V_d(s,t)$ - should be removed.

3.4 Simulations

In this section we illustrate the finite behavior of the proposed asymptotic results in Theorem 3.1 by a simple simulation example. The simulation is designed as follows. We consider the very simple case

$$Y_{ij} = \mu(t_j) + \xi_i \phi(t_j) + \epsilon_i(j) \tag{3.29}$$

with the trend

$$\mu(t) = 10 + 2t^2 \quad (t \in [0,1])$$

and only one functional principal component

$$\phi(t) = \sin 4\pi t \quad (t \in [0,1]).$$

The scores ξ_i are iid standard normal random variables. The error process $\epsilon_i(j)$ with variance one are generated as follows: (a) iid $N(0,1)$; (b) $AR(1)$ with lag-one correlation $\rho = 0.5$; (c) $FARIMA(0, 0.3, 0)$ process which is a long memory process with long memory parameter $d = 0.3$ as discussed in Chapter 2 (see e.g. Granger and Joyeux 1980, Hosking 1981, Beran 1994, Beran et al. 2013).

The number of sampling points on each curve considered here are chosen as $N =$ 100, 200, 400, 600, 800, 1000, 2000 and 4000. According to the restriction (3.16), the number of time series (or sample sizes) n_N should be in the order $o\left(N^{\frac{4}{3}}\right)$ for the short-range dependent cases (a) and (b) or $o\left(N^{\frac{2}{3}}\right)$ for the long-range dependent case (c). Thus, n_N can be set equal to $10N^{0.6}$ rounded to the next integer (i.e. $n_N =$ 158, 240, 364, 464, 552, 631, 956 and 1450 respectively).

For the kernel estimator $\hat{\mu}(t)$, the rectangular kernel $K_1(u) = \frac{1}{2}\mathbf{1}\{-1 \le u \le 1\}$ is used. The bandwidth $b = b_N$ is set equal to $0.05N^{-0.16}$. This is because condition (3.18) should be satisfied. In fact, note that $n_N^{-\frac{1}{4}} = 10^{-1/4}N^{-0.15}$ and $d = 0$ (in case (a) and case (b)) implies $N^{-\frac{1-2d}{3-2d}} = N^{-\frac{1}{3}}$ and $d = 0.3$ (in case (c)) implies $N^{-\frac{1-2d}{3-2d}} = N^{-\frac{1}{6}}$. Therefore $0.15 < 0.16 < \frac{1}{6}$ implies that (3.18) holds. For each pair (N, n_N), 400 hundred simulations are carried out.

Table 3.1 gives simulated values of standardized integrated variance $\sigma_N^2 = n_N \int_0^1 var\left[\hat{\mu}(t)\right] dt$ and standardized integrated mean squared error $IMSE_N = n_N \int_0^1 E\left[(\hat{\mu}(t) - \mu(t))^2\right] dt$ of $\hat{\mu}(t)$. Note that the integrated squared bias of $\hat{\mu}(t)$ is in the order $O(b_N^4) = O(N^{-0.64})$. The integrated variance of $\hat{\mu}(t)$ is in the order $O\left(n_N^{-1}\right) = O(N^{-0.6})$. So the integrated squared bias of $\hat{\mu}(t)$ converges to zero at a slightly faster rate than the integrated variance of $\hat{\mu}(t)$. Therefore, the asymptotic values of σ_N^2 and $IMSE_N$ obtained from Theorem 3.1 are the same, that is $\sigma^2_\infty = IMSE_\infty = \int_0^1 \sin^2(4\pi t)dt = \frac{1}{2}$. The results in Table 3.1 show that this asymptotic value is reached fairly quickly, even if ϵ exhibits strongly dependence.

3.5 Proofs and tables

3.5.1 Proofs

Proof. (of Theorem 3.1)

Because of the conditions on boundary kernels, it is sufficient to consider interior points $t \in (0,1)$ only. The proof is the same for the two boundary points $t = 0$ and $t = 1$ except that $\beta(t)$ equals to β_0 instead of β_1.

For simplicity of presentation we consider the very simple case with only one basis function $\phi(t)$. Since the principal scores ξ_{il} ($l \in \mathbb{N}$) are independent and the functions $\phi_l(t)$ ($l \in \mathbb{N}$) are orthonormal, the extension to the general case is straightforward. Therefore, let

$$Y_{ij} = \mu(t_j) + \xi_i\phi(t_j) + \epsilon_i(j) \quad (i = 1, ..., n,\ j = 1, ..., N).$$

Since the bias neither depends on n nor the temporal dependence structure, the asymptotic expression for it is standard. In fact,

$$\begin{aligned}
&E\left[\hat{\mu}(t)\right] - \mu(t) \\
&= E\left[\frac{1}{Nb}\sum_{j=1}^{N} K_1\left(\frac{t-t_j}{b}\right) n^{-1}\left(\sum_{i=1}^{n}(\mu(t_j) + \xi_i\phi(t_j) + \epsilon_i(j))\right)\right] - \mu(t) \\
&= \frac{1}{Nb}\sum_{j=1}^{N} K_1\left(\frac{t-t_j}{b}\right)\mu(t_j) - \mu(t) \\
&= \frac{1}{2}\mu''(t)\int_{-1}^{1} K_1(u)u^2 du \times b^2 + o(b^2) + O\left((Nb)^{-1}\right) \\
&= \frac{1}{2}\mu''(t)\beta(t)b^2 + o(b^2) + O\left((Nb)^{-1}\right).
\end{aligned}$$

The proof of asymptotic variance consists of decomposing $\hat{\mu}(t) - E\left[\hat{\mu}(t)\right]$ into dominated part and remainder part which is asymptotically negligible. Since under the assumption of long-range dependence the variance of each of the terms converges to zero at a slower rate than under short-range dependence, it is sufficient to consider the case of long-range dependence. Thus, in the following, we assume that (3.5) holds with $0 < d < \frac{1}{2}$. Since $t \in (0,1)$, we have $t \in [b_N, 1-b_N]$ for N large enough. We then have

$$\hat{\mu}(t) - E\left[\hat{\mu}(t)\right] = \frac{1}{Nb}\sum_{j=1}^{N} K_1\left(\frac{t-t_j}{b}\right) n^{-1}\left[\sum_{i=1}^{n} \xi_i \phi(t_j) + \epsilon_i(j)\right].$$

Since ϵ and ξ are independent, then

$$var[\hat{\mu}(t)] = n^{-2}(Nb)^{-2}\left(A_{n,N} + B_{n,N}\right)$$

where

$$A_{n,N} = var\left(\sum_{i=1}^{n}\sum_{j=1}^{N} K_1\left(\frac{t-t_j}{b}\right)\xi_i\phi(t_j)\right),$$
$$B_{n,N} = var\left(\sum_{i=1}^{n}\sum_{j=1}^{N} K_1\left(\frac{t-t_j}{b}\right)\epsilon_i(j)\right).$$

For the part $A_{n,N}$ we have

$$\begin{aligned}
A_{n,N} &= \left(\sum_j \phi(t_j) K_1\left(\frac{t-t_j}{b}\right)\right)^2 var\left(\sum_i \xi_i\right)\\
&= n(Nb)^2\lambda\phi^2(t)\left(1 + O(b^2) + O\left((Nb)^{-1}\right)\right).
\end{aligned}$$

For the part $B_{n,N}$ we obtain

$$\begin{aligned}
B_{n,N} &= \sum_i var\left(\sum_j K_1\left(\frac{t-t_j}{b}\right)\epsilon_i(j)\right)\\
&= n\left(\sigma^2\sum_j K_1^2\left(\frac{t-t_j}{b}\right) + \sum_{j\neq k}\gamma_\epsilon(j-k)K_1\left(\frac{t-t_j}{b}\right)K_1\left(\frac{t-t_k}{b}\right)\right)\\
&= n\left(\sigma^2\sum_j K_1^2\left(\frac{t-t_j}{b}\right) + \sum_{j\neq k}c_\gamma|j-k|^{2d-1}K_1\left(\frac{t-t_j}{b}\right)K_1\left(\frac{t-t_k}{b}\right)\right)\\
&= O\left(n(Nb)\right) + n(Nb)^{2d+1}c_\gamma V_d(s,t)(1+o(1))\\
&= n(Nb)^{2d+1}c_\gamma V_d(s,t)(1+o(1)).
\end{aligned}$$

Thus, as $Nb_N \to \infty$ and $b_N \to 0$,

$$\begin{aligned}
var[\hat{\mu}(t)] &= n^{-2}(Nb)^{-2}\left(A_{n,N} + B_{n,N}\right)\\
&= n^{-1}\lambda\phi^2(t)\left[1 + O\left((Nb)^{2d-1}\right) + O(b^2) + O\left((Nb)^{-1}\right)\right].
\end{aligned}$$

□

Proof. (of Theorem 3.2)

As discussed in the proof of Theorem 3.1, it is to sufficient to consider interior points only. Moreover, it is sufficient to consider the case of long memory, i.e. (3.5) with $0 < d < \frac{1}{2}$.

Condition (3.17) is only required to make the bias in the order $o\left(n^{-\frac{1}{2}}\right)$. Since this condition (3.17), i.e.

$$\sqrt{n}b^2 \to 0$$

together with Theorem 3.1 and (3.14) implies

$$\lim_{n\to\infty} \sqrt{n} \sup_{t\in[0,1]} |E[\hat{\mu}(t)] - \mu(t)| = 0, \tag{3.30}$$

thus, it is sufficient to consider the weak convergence of

$$Z^0_{n,N}(t) = \sqrt{n}\left(\hat{\mu}(t) - E[\hat{\mu}(t)]\right).$$

As discussed in the proof of Theorem 3.1, we may consider the very simple case with only one basis function $\phi(t)$, i.e.

$$X_i(t_j) = \mu(t_j) + \xi_i\phi(t_j).$$

Let

$$c_N(t) = \frac{1}{Nb}\sum_{j=1}^{N} K_1\left(\frac{t-t_j}{b}\right)\phi(t_j),$$

$$u_n = n^{-\frac{1}{2}}\lambda^{-\frac{1}{2}}\sum_{i=1}^{n}\xi_i$$

and

$$e_n(j) = n^{-\frac{1}{2}}\sum_{i=1}^{n}\epsilon_i(j),$$

the sequence $Z^0_{n,N}(t)$ can be separated into two independent components

$$\begin{aligned} Z^0_{n,N}(t) &= \frac{1}{Nb}\sum_{j=1}^{N} K_1\left(\frac{t-t_j}{b}\right) n^{-1}\left[\sum_{i=1}^{n}\xi_i\phi(t_j) + \epsilon_i(j)\right] \\ &= S_{n,N,1}(t) + S_{n,N,2}(t) \end{aligned}$$

where

$$S_{n,N,1}(t) = c_N(t)\sqrt{\lambda}u_n$$

and

$$S_{n,N,2}(t) = \frac{1}{Nb} \sum_{j=1}^{N} K_1 \left(\frac{t - t_j}{b} \right) e_n(j).$$

Firstly, we consider $S_{n,N,1}(t)$. Since ξ_i are Gaussian independent random variables with mean 0 and variance λ, u_n is a standard normal random variable for all n. For $c_N(t)$ we have

$$|c_N(s) - c_N(t)| \leq |\phi(t) - \phi(s)| + C_1 b^2 + C_2 (Nb)^{-1}$$

where C_1, C_2 are suitable constants not depending on s, t and N. Therefore, for any $\Delta > 0$, we have

$$\begin{aligned} w_{n,N}(\Delta) &= \sup_{|s-t| \leq \Delta} |S_{n,N,1}(s) - S_{n,N,1}(t)| \\ &\leq \sqrt{\lambda} \left(w_\phi(\Delta) + C_1 b^2 + C_2 (Nb)^{-1} \right) |u_n| \end{aligned}$$

where

$$w_\phi(\Delta) = \sup_{|s-t| \leq \Delta} |\phi(t) - \phi(s)|$$

is the modulus of continuity of $\phi(t)$. Let $\tau > 0$. Then, we have

$$P \left(w_{n,N}(\Delta) > \tau \right) \leq P \left(\sqrt{\lambda} \left(w_\phi(\Delta) + C_1 b^2 + C_2 (Nb)^{-1} \right) |u_n| > \tau \right).$$

Taking the lim sup over n and N of the two hand sides such that conditions (3.14) and (3.17) hold, we have

$$\limsup_{n,N} P \left(w_{n,N}(\Delta) > \tau \right) \leq 2 \left[1 - \Phi \left(\tau w_\phi^{-1}(\Delta) \lambda^{-\frac{1}{2}} \right) \right].$$

where Φ denotes the cumulative standard normal distribution. Since $\phi(t)$ is uniformly continuous on $[0,1]$, we obtain

$$\lim_{\Delta \to 0} \limsup_{n,N} P \left(w_{n,N}(\Delta) > \tau \right) = 0.$$

Theorem 7.5 in Billingsley (1999, p. 84) then implies

$$S_{n,N,1}(t) \Rightarrow \sqrt{\lambda} \phi(t) \zeta \tag{3.31}$$

where ζ is a standard normal random variable.

Secondly, for $S_{n,N,2}(t)$, note that the error term $e_n(j)$ ($j \in \mathbb{N}$) is a zero mean stationary Gaussian process with autocovariance function

$$\gamma_e(k) = \gamma_\epsilon(k) \underset{k\to\infty}{\sim} c_\gamma k^{2d-1}.$$

Now we show that $S_{n,N,2}(t)$ converges weakly to the zero process $\zeta^0(t) \equiv 0$ in $C[0,1]$ equipped with the supremum norm. Since $S_{n,N,2}(t) \in C[0,1]$ and $\zeta^0(t) \in C[0,1]$, convergence in the supremum norm is equivalent to convergence in the Skorohod norm. Thus, it is sufficient to derive the weak convergence in the Skorohod metric.

First we show convergence of finite dimensional distributions. Note that

$$\begin{aligned} var\left(S_{n,N,2}(t)\right) &= (Nb)^{-2} \sum_{j,j'=1}^{N} K_1\left(\frac{t-t_j}{b}\right) K_1\left(\frac{t-t_{j'}}{b}\right) \gamma_\epsilon(j-j') \\ &= (Nb)^{2d-1} c_\gamma V_d(t,t) + r_N(t) \end{aligned}$$

with

$$\lim_N (Nb)^{1-2d} \sup_{t\in[0,1]} r_N(t) = 0$$

where the limit $N \to \infty$, $b = b_N \to 0$ is taken such that (3.14) and (3.17) hold. Therefore, we have

$$\sup_{t\in[0,1]} var\left(S_{n,N,2}(t)\right) \le C_{var}(Nb)^{2d-1} \quad (n \ge n_0,\ N \ge N_0),$$

for $n_0 \in \mathbb{N}$ and $N_0 \in \mathbb{N}$ large enough and a suitable constant $0 < C_{var} < \infty$. Since ϵ is long-range dependent i.e. $0 < d < \frac{1}{2}$, this implies that, for all $p \in \mathbb{N}$, $t_1, ..., t_p \in [0,1]$, $(S_{n,N,2}(t_1), ..., S_{n,N,2}(t_p))^T$ converges in probability and in distribution to the p-dimensional zero vector $(0, ..., 0)^T$.

Now we consider tightness of $S_{n,N,2}(t)$ ($t \in [0,1]$). To show tightness note that

$$\begin{aligned} &E\left[\left(S_{n,N,2}(s) - S_{n,N,2}(t)\right)^2\right] \\ &= (Nb)^{-2} E\left[\left(\sum_j^N K_1\left(\frac{s-t_j}{b}\right) n^{-\frac{1}{2}} \sum_{i=1}^n \epsilon_i(j) - \sum_j^N K_1\left(\frac{t-t_j}{b}\right) n^{-\frac{1}{2}} \sum_{i=1}^n \epsilon_i(j)\right)^2\right] \\ &\le \kappa^2 (t-s)^2 b^{-2} (Nb)^{-2} \sum_{j,k=1}^{2Nb} |\gamma_\epsilon(j-k)| \\ &\le C(t-s)^2 b^{-2} (Nb)^{2d-1} \\ &= C\,(t-s)^2 \left(N b^{1+2/(1-2d)}\right)^{2d-1} \end{aligned}$$

where

$$\kappa = \sup_{u\in[-1,1]} |K_1'(u)|$$

and C is a suitable constant not depending on s, t, n and N. Therefore, assumption (3.15) implies that there is a finite constant C^* such that

$$E\left[(S_{n,N,2}(s) - S_{n,N,2}(t))^2\right] \leq C^*(t-s)^2 \quad (t, s \in [0,1]^2).$$

Thus, tightness of $S_{n,N,2}(t)$ ($t \in [0,1]$) and weak convergence of $S_{n,N,2}(t)$ to $\zeta^0(t) \equiv 0$ in the Skorohod topology follows from Theorem 13.5 in Billingsley (1999, p. 142). This together with (3.31) implies

$$Z^0_{n,N}(t) \Rightarrow \sqrt{\lambda}\phi(t)\zeta.$$

As discussed above, the additional condition (3.17) and (3.30) then implies

$$Z_{n,N}(t) \Rightarrow \sqrt{\lambda}\phi(t)\zeta.$$

For the general case with an arbitrary number of basis functions $\phi_l(t)$, the corresponding results can be obtained similarly. □

Proof. (of Theorem 3.3)

As discussed in the proof of Theorem 3.1, for simplicity of presentation only interior points will be discussed explicitly, since this issue only matters for the definition of $V_d(s,t)$. Moreover, it is sufficient to consider the long memory case, i.e. (3.5) with $0 < d < \frac{1}{2}$. Also, for simplicity of presentation, we first consider the case with only one basis function $\phi(t_j)$.

For the expected value of $\hat{C}(s,t)$, we have

$$
\begin{aligned}
&E[\hat{C}(s,t)] \\
&= \frac{1}{(Nb)^2}\sum_{j,k=1}^{N} K_2\left(\frac{s-t_j}{b},\frac{t-t_k}{b}\right) n^{-1}\sum_{i=1}^{n} E\left[(\xi_i\phi(t_j)+\epsilon_i(j))(\xi_i\phi(t_k)+\epsilon_i(k))\right] \\
&= \frac{1}{(Nb)^2}\sum_{j,k=1}^{N} K_1\left(\frac{s-t_j}{b}\right) K_1\left(\frac{t-t_k}{b}\right)\left[\lambda\phi(t_j)\phi(t_k)+\gamma_\epsilon(j-k)\right] \\
&= \lambda\left\{\frac{1}{Nb}\sum_{j=1}^{N} K_1\left(\frac{s-t_j}{b}\right)\phi(t_j)\right\}\left\{\frac{1}{Nb}\sum_{j=k}^{N} K_1\left(\frac{t-t_k}{b}\right)\phi(t_k)\right\} \\
&\quad + \frac{1}{(Nb)^2}\sum_{j,k=1}^{N} K_1\left(\frac{s-t_j}{b}\right) K_1\left(\frac{t-t_k}{b}\right)\gamma_\epsilon(j-k) \\
&= \lambda\phi(s)\phi(t)\left[1+O(b^2)+O((Nb)^{-1})\right]+(Nb)^{2d-1}c_\gamma V_d(s,t)\left[1+O((Nb)^{-1})\right] \\
&= \lambda\phi(s)\phi(t)+O(b^2)+O\left((Nb)^{2d-1}\right)
\end{aligned}
$$

where the error terms $O(b^2)$ and $O\left((Nb)^{2d-1}\right)$ are uniform in $s,t\in(0,1)$.

The variance of $\hat{C}(s,t)$ is of the form

$$
\begin{aligned}
var[\hat{C}(s,t)] &= \frac{1}{(Nb)^4} var\left\{\sum_{j,k=1}^{N} K_2\left(\frac{s-t_j}{b},\frac{t-t_k}{b}\right) n^{-1}\sum_{i=1}^{n} Y_{ij}Y_{ik}\right\} \\
&=: (Nb)^{-4}n^{-1}A_N(s,t).
\end{aligned}
$$

Since, for $i \neq i'$, the process $Y_i(t)$ is independent of the process $Y_{i'}(t)$, we have

$$
\begin{aligned}
& A_N(s,t) \\
&= var\left\{\sum_{j,k} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) Y_{ij}Y_{ik}\right\} \\
&= \sum_{j,k,j',k'} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) K_2\left(\frac{s-t_{j'}}{b}, \frac{t-t_{k'}}{b}\right) E[Y_{ij}Y_{ik}Y_{ij'}Y_{ik'}] \\
&\quad - \left(\sum_{j,k} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) E[Y_{ij}Y_{ik}]\right)^2 \\
&= E[\sum_{j,k,j',k'} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) K_2\left(\frac{s-t_{j'}}{b}, \frac{t-t_{k'}}{b}\right) \\
&\qquad \cdot (\xi\phi(t_j)+\epsilon(j))(\xi\phi(t_k)+\epsilon(k))(\xi\phi(t_{j'})+\epsilon(j'))(\xi\phi(t_{k'})+\epsilon(k'))] \\
&\quad - \left\{E\left[\sum_{j,k} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right)(\xi\phi(t_j)+\epsilon(j))(\xi\phi(t_k)+\epsilon(k))\right]\right\}^2 \\
&= A_{N,1}+A_{N,2}+A_{N,3}+A_{N,4}+A_{N,5}+A_{N,6}+A_{N,7}+A_{N,8}-(A_{N,9}+A_{N,10})^2
\end{aligned}
$$

where

$$A_{N,1} = E[\xi^4] \sum_{j,k,j',k'} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) K_2\left(\frac{s-t_{j'}}{b}, \frac{t-t_{k'}}{b}\right) \phi(t_j)\phi(t_k)\phi(t_{j'})\phi(t_{k'}),$$

$$A_{N,2} = \lambda \sum_{j,k,j',k'} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) K_2\left(\frac{s-t_{j'}}{b}, \frac{t-t_{k'}}{b}\right) \phi(t_j)\phi(t_k)E[\epsilon(j')\epsilon(k')],$$

$$A_{N,3} = \lambda \sum_{j,k,j',k'} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) K_2\left(\frac{s-t_{j'}}{b}, \frac{t-t_{k'}}{b}\right) \phi(t_j)\phi(t_{j'})E[\epsilon(k)\epsilon(k')],$$

$$A_{N,4} = \lambda \sum_{j,k,j',k'} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) K_2\left(\frac{s-t_{j'}}{b}, \frac{t-t_{k'}}{b}\right) \phi(t_j)\phi(t_{k'})E[\epsilon(k)\epsilon(j')],$$

$$A_{N,5} = \lambda \sum_{j,k,j',k'} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) K_2\left(\frac{s-t_{j'}}{b}, \frac{t-t_{k'}}{b}\right) \phi(t_k)\phi(t_{j'})E[\epsilon(j)\epsilon(k')],$$

$$A_{N,6} = \lambda \sum_{j,k,j',k'} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) K_2\left(\frac{s-t_{j'}}{b}, \frac{t-t_{k'}}{b}\right) \phi(t_k)\phi(t_{k'})E[\epsilon(j)\epsilon(j')],$$

$$A_{N,7} = \lambda \sum_{j,k,j',k'} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) K_2\left(\frac{s-t_{j'}}{b}, \frac{t-t_{k'}}{b}\right) \phi(t_{j'})\phi(t_{k'})E[\epsilon(j)\epsilon(k)],$$

$$A_{N,8} = \sum_{j,k,j',k'} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) K_2\left(\frac{s-t_{j'}}{b}, \frac{t-t_{k'}}{b}\right) E[\epsilon(j)\epsilon(k)\epsilon(j')\epsilon(k')],$$

$$A_{N,9} = \lambda \left\{ \sum_j K_1 \left(\frac{s-t_j}{b} \right) \phi(t_j) \right\} \left\{ \sum_k K_1 \left(\frac{t-t_k}{b} \right) \phi(t_k) \right\},$$

and

$$A_{N,10} = \sum_{j,k} K_2 \left(\frac{s-t_j}{b}, \frac{t-t_k}{b} \right) \gamma_\epsilon(j-k).$$

Since $E[\xi^4] = 3\lambda^2$, and for $t \neq s$ and N large enough, we have

$$\sum_{j,k,j',k'} K_2 \left(\frac{s-t_j}{b}, \frac{t-t_k}{b} \right) K_2 \left(\frac{s-t_{j'}}{b}, \frac{t-t_{k'}}{b} \right) \phi(t_j)\phi(t_k)\phi(t_{j'})\phi(t_{k'})$$
$$= (Nb)^4 \phi^2(t)\phi^2(s) \left[1 + O(b^2) + O\left((Nb)^{-1} \right) \right],$$

we thus obtain

$$\begin{aligned} A_{N,1} &= (Nb)^4 3\lambda^2 \phi^2(t)\phi^2(s) \left[1 + O(b^2) + O\left((Nb)^{-1} \right) \right] \\ &= (Nb)^4 3\lambda^2 \phi^2(t)\phi^2(s) \left[1 + o(1) \right]. \end{aligned}$$

Similarly, for $t = s$, we have

$$\sum_{j,k} K_1 \left(\frac{t-t_j}{b} \right) K_1 \left(\frac{t-t_k}{b} \right) \phi(t_j)$$
$$= (Nb)^2 \phi^2(t) \|K_1\|^2 \left[1 + O(b^2) + O\left((Nb)^{-1} \right) \right]$$

so that again we obtain

$$A_{N,1} = (Nb)^4 3\lambda^2 \phi^2(t)\phi^2(s) \left[1 + O(b^2) + O\left((Nb)^{-1} \right) \right].$$

Analogous arguments imply that the terms $A_{N,j}$ $(j = 2, ..., 8)$ are of the order $o\left((Nb)^4\right)$ and hence asymptotically negligible, with $A_{N,j}$ $(j = 2, ..., 7)$ being of the order $O\left((Nb)^{4+2d-1}\right)$ and $A_{N,8}$ of the order $o\left((Nb)^{4+2d-1}\right)$. For example, for $A_{N,2}$ and $t \neq s$, we have

$$\begin{aligned} &\sum_{j,k,j',k'} K_2 \left(\frac{s-t_j}{b}, \frac{t-t_k}{b} \right) K_2 \left(\frac{s-t_{j'}}{b}, \frac{t-t_{k'}}{b} \right) \phi(t_j)\phi(t_k) E[\epsilon(j')\epsilon(k')] \\ &= \left\{ \sum_{j,k} K_2 \left(\frac{s-t_j}{b}, \frac{t-t_k}{b} \right) \phi(t_j)\phi(t_k) \right\} \left\{ \sum_{j',k'} \gamma_\epsilon(j'-k') K_2 \left(\frac{s-t_{j'}}{b}, \frac{t-t_{k'}}{b} \right) \right\} \\ &= (Nb)^2 \phi(t)\phi(s) \left[1 + O(b^2) + O\left((Nb)^{-1} \right) \right] (Nb)^{2d+1} c_\gamma V_d(s,t) [1 + o(1)] \\ &= (Nb)^{3+2d} \phi(t)\phi(s) \left[1 + O(b^2) + O\left((Nb)^{-1} \right) \right] c_\gamma V_d(s,t) \left[1 + o(1) \right] \\ &= O\left((Nb)^{3+2d} \right) \\ &= o\left((Nb)^4 \right), \end{aligned}$$

the last equality following from $0 < d < \frac{1}{2}$. For $A_{N,10}$ we obtain

$$A_{N,10} = \sum_{j,k} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right)\gamma_\epsilon(j-k) = (Nb)^{2d+1}c_\gamma V_d(s,t)\,[1+o(1)]$$

thus

$$A_{N,10}^2 = O\left((Nb)^{2+4d}\right) = o\left((Nb)^4\,(Nb)^{2d-1}\right).$$

Finally, for $A_{N,9}$, similarly we have,

$$\begin{aligned} A_{N,9} &= \lambda\sum_{j,k} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right)\phi(t_j)\phi(t_k) \\ &= (Nb)^2\lambda\phi(t)\phi(s)\left[1+O(b^2)+O\left((Nb)^{-1}\right)\right] \end{aligned}$$

so that

$$A_{N,9}^2 = (Nb)^4\lambda^2\phi^2(t)\phi^2(s)\,[1+o(1)]$$

and

$$\begin{aligned} \left(A_{N,9}+A_{N,10}\right)^2 &= (Nb)^4\lambda^2\phi^2(t)\phi^2(s)[1+o(1)] \\ &\quad + (Nb)^{4+2d-1}2\lambda\phi(t)\phi(s)c_\gamma V_d(s,t)[1+o(1)] \end{aligned}$$

In conclude, we have

$$\begin{aligned} A_N(s,t) &= (Nb)^4 2\lambda^2\phi^2(t)\phi^2(s) \\ &\quad + (Nb)^4\left[(Nb)^{2d-1}4c_\gamma V_d(s,t)\lambda\phi(s)\phi(t) + O(b^2) + o\left(\max\left\{(Nb)^{2d-1}, b^2\right\}\right)\right] \end{aligned}$$

and hence

$$\begin{aligned} var[\hat{C}(s,t)] &= n^{-1}2\lambda^2\phi^2(t)\phi^2(s) \\ &\quad + n^{-1}\left[(Nb)^{2d-1}4c_\gamma V_d(s,t)\lambda\phi(s)\phi(t) + O(b^2) + o\left(\max\left\{(Nb)^{2d-1}, b^2\right\}\right)\right]. \end{aligned}$$

In the general case with an arbitrary number of basis functions $\phi_l(t)$, the analogous arguments are applicable except that there is only one place should be changed. That is the variance of $\hat{C}(s,t)$ with $(Nb)^{-4}\left(A_{N,1} - A_{N,9}^2\right)$ containing the additional asymptotically non-negligible (mixed) term

$$\sum_{l_1<l_2}\lambda_{l_1}\lambda_{l_2}\left[\phi_{l_1}(s)\phi_{l_2}(t) + \phi_{l_2}(s)\phi_{l_1}(t)\right]^2.$$

□

Proof. (of Theorem 3.4)

As discussed in the proof of Theorem 3.1, for simplicity of presentation, only interior points will be discussed explicitly. For boundary points nothing will be changed except the expression for $V_d(s,t)$. Moreover, it is sufficient to consider the long memory case, i.e. (3.5) with $0 < d < \frac{1}{2}$. Also, for simplicity of presentation, we again mainly present the proof for the model with one basis function $\phi(t)$ only.

Defining the following notations

$$\zeta_n = n^{-\frac{1}{2}} \sum_{i=1}^{n} \left(\xi_i^2 - \lambda\right),$$

$$e_{n,1}(j) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \xi_i \epsilon_i(j)$$

and

$$e_{n,2}(j,k) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \epsilon_i(j)\epsilon_i(k) - \gamma_\epsilon(j-k).$$

For the sequence of processes $Z^0_{n,N}(s,t)$, we have the following decomposition

$$\sqrt{n}\left(\hat{C}(s,t) - E[\hat{C}(s,t)]\right) = S_{n,1}(s,t) + S_{n,2}(s,t) + S_{n,3}(s,t) + S_{n,4}(s,t)$$

where

$$S_{n,1}(s,t) = \left\{ \frac{1}{(Nb)^2} \sum_{j,k=1}^{N} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) \phi(t_j)\phi(t_k) \right\} \cdot \zeta_n,$$

$$S_{n,2}(s,t) = \frac{1}{(Nb)^2} \sum_{j,k=1}^{N} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) \phi(t_j) \cdot e_{n,1}(k),$$

$$S_{n,3}(s,t) = \frac{1}{(Nb)^2} \sum_{j,k=1}^{N} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) \phi(t_k) \cdot e_{n,1}(j)$$

and

$$S_{n,4}(s,t) = \frac{1}{(Nb)^2} \sum_{j,k=1}^{N} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) \cdot e_{n,2}(j,k).$$

As discussed in the proof of Theorem 3.2, since $S_{n,1}(s,t) \in C[0,1]^2$, $S_{n,2}(s,t) \in C[0,1]^2$, $S_{n,3}(s,t) \in C[0,1]^2$ and $S_{n,4}(s,t) \in C[0,1]^2$ $(s,t \in [0,1])$ it is sufficient to show weak convergence in the Skorohod metric for each of them.

Convergence of finite dimensional distributions can be derived straight. For example, for $S_{n,1}(s,t)$, we have

$$var(\zeta_n) = 2\lambda^2$$

and

$$\frac{1}{(Nb)^2}\sum_{j,k=1}^{N} K_2\left(\frac{s-t_j}{b},\frac{t-t_k}{b}\right)\phi(t_j)\phi(t_k)$$
$$= \phi(s)\phi(t) + O\left(b^2\right) + O\left((Nb)^{-1}\right)$$

with uniformly bounded errors. Thus

$$var\left(S_{n,1}(s,t)\right) = 2\lambda^2\phi^2(s)\phi^2(t)\left[1 + O(b^2) + O\left((Nb)^{-1}\right)\right]$$

converges to $2\lambda^2\phi^2(s)\phi^2(t)$ uniformly in $s \in [0,1]$ and $t \in [0,1]$.

For $S_{n,2}(s,t)$, we have

$$var\left(S_{n,2}(s,t)\right)$$
$$= (Nb)^{-4}\sum_{j,k,j',k'=1}^{N} K_2\left(\frac{s-t_j}{b},\frac{t-t_k}{b}\right)K_2\left(\frac{s-t_{j'}}{b},\frac{t-t_{k'}}{b}\right)$$
$$\cdot\,\phi(t_j)\phi(t_{j'})cov\left(e_{n,1}(k),e_{n,1}(k')\right).$$

Now, since

$$cov\left(e_{n,1}(k),e_{n,1}(k')\right) = \lambda\gamma_\epsilon\left(k-k'\right) = \lambda c_\gamma|k-k'|^{2d-1}$$

and $K_2(u,v) = K_1(u)K_1(v)$ so

$$var\left(S_{n,2}(s,t)\right) = \lambda\left\{\frac{1}{Nb}\sum_{j=1}^{N}K_1\left(\frac{s-t_j}{b}\right)\phi(t_j)\right\}\left\{\frac{1}{Nb}\sum_{j=1}^{N}K_1\left(\frac{t-t_j}{b}\right)\phi(t_j)\right\}$$
$$\cdot\,(Nb)^{-2}\sum_{k,k'=1}^{N}K_1\left(\frac{t-t_k}{b}\right)K_1\left(\frac{t-t_{k'}}{b}\right)\gamma_\epsilon\left(k-k'\right)$$
$$= (Nb)^{2d-1}\lambda\phi(s)\phi\left(t\right)c_\gamma V_d(s,t)\left[1+O\left(b^2\right)+O\left((Nb)^{-1}\right)\right]$$

which converges to zero uniformly in $s \in [0,1]$ and $t \in [0,1]$. The same arguments apply to $S_{n,3}(s,t)$.

Finally, for $S_{n,4}(s,t)$, we note that

$$cov\left(e_{n,2}(j,k),e_{n,2}(j',k')\right)$$
$$= cov\left(\epsilon(j)\epsilon(k),\epsilon(j')\epsilon(k')\right)$$
$$= \gamma_\epsilon(j-j')\gamma_\epsilon(k-k') + \gamma_\epsilon(j-k')\gamma_\epsilon(k-j').$$

By similar arguments as before this leads to

$$var\left(S_{n,4}(s,t)\right) = (Nb)^{4d-2} 2c_\gamma^2 V_d^2(s,t)\left[1 + O(b^2) + O\left((Nb)^{-1}\right)\right].$$

We thus derive for $S_{n,2}(s,t)$, $S_{n,3}(s,t)$ and $S_{n,4}(s,t)$ convergence of finite dimensional distributions to zero. However, for $S_{n,1}(s,t)$ we have for any $p \in \mathbb{N}$, $u^{(i)} = (u_1^{(i)}, u_2^{(i)}) \in [0,1]^2$ $(i = 1, ..., p)$,

$$\sqrt{n}\left(S_{n,1}(u_1^{(1)}, u_2^{(1)}) - C(u_1^{(1)}, u_2^{(1)}), ..., S_{n,1}(u_1^{(p)}, u_2^{(p)}) - C(u_1^{(p)}, u_2^{(p)})\right)$$
$$\underset{d}{\rightarrow} \sqrt{2}\lambda\left[\phi\left(u_1^{(1)}\right)\phi\left(u_2^{(1)}\right), ..., \phi\left(u_1^{(p)}\right)\phi\left(u_2^{(p)}\right)\right]\zeta$$

where ζ is a standard noraml random variable.

As discussed in Theorem 3.2, with respect to tightness in the supremum norm, it is sufficient to show tightness in the Skorohod metric. For example, we consider $S_{n,4}(t,s)$ which is the term with the slowest rate of convergence. For any $\Delta > 0$, let $u = (u_1, u_2)$, $v = (v_1, v_2) \in [0,1]^2$ with Euclidian distance $\|u - v\| < \Delta$. Then, we have

$$\begin{aligned}
S_{n,4}(v_1, v_2) &= \frac{1}{(Nb)^2}\sum_{j,k=1}^{N} K_2\left(\frac{v_1 - t_j}{b}, \frac{v_2 - t_k}{b}\right)\cdot e_{n,2}(j,k) \\
&= \frac{1}{(Nb)^2}\sum_{j,k=1}^{N} K_2\left(\frac{u_1 - t_j}{b}, \frac{u_2 - t_k}{b}\right)\cdot e_{n,2}(j,k) \\
&\quad + \frac{1}{(Nb)^2}b^{-1}\sum_{j,k=1}^{N} \frac{\partial}{\partial x_1}K_2\left(\frac{u_1^* - t_j}{b}, \frac{u_2^* - t_k}{b}\right)\cdot e_{n,2}(j,k) \\
&\quad + \frac{1}{(Nb)^2}b^{-1}\sum_{j,k=1}^{N} \frac{\partial}{\partial x_2}K_2\left(\frac{u_1^{**} - t_j}{b}, \frac{u_2^{**} - t_k}{b}\right)\cdot e_{n,2}(j,k)
\end{aligned}$$

where $\max\{\|u^* - v\|, \|u^{**} - v\|\} < \Delta$ with $u^* = (u_1^*, u_2^*)$ and $u^* = (u_1^{**}, u_2^{**})$. Using the notation

$$\kappa = \sup_{x_1, x_2 \in [-1,1]} \max\left\{\left|\frac{\partial}{\partial x_1}K_2(x_1, x_2)\right|, \left|\frac{\partial}{\partial x_2}K_2(x_1, x_2)\right|\right\},$$

we obtain the upper bound

$$
\begin{aligned}
& E\left[\|S_{n,4}(u_1,u_2)-S_{n,4}(v_1,v_2)\|^2\right] \\
& \le \|u-v\|^2 (Nb)^{-4} b^{-2} 4\kappa^2 \sum_{j,k,j',k'=1}^{2Nb} |\gamma_\epsilon(j-j')\gamma_\epsilon(k-k')+\gamma_\epsilon(j-k')\gamma_\epsilon(k-j')| \\
& \le C(Nb)^{4d-2} b^{-2} 4\kappa^2 \Delta^2 \\
& = \left[C\left(Nb^{1+2/(1-2d)}\right)^{4d-2} 4\kappa^2\right]\Delta^2
\end{aligned}
$$

where C is a suitable finite constant not depending on s, t, n and N. Thus assumption (3.24) implies that there is a finite constant C^* such that

$$E\left[\|S_{n,4}(u)-S_{n,4}(v)\|^2\right] \le C^* \|u-v\|^2.$$

Therefore tightness of $S_{n,4}(s,t)$ $(s,t \in [0,1]^2)$ is derived and the weak convergence of $S_{n,4}(s,t)$ to 0 in the Skorohod metric follows from Theorem 13.5 in Billingsley (1999, p. 142). Similarly it follows that $S_{n,2}(s,t)$ and $S_{n,3}(s,t)$ are tight and converge to 0 in the Skorohod metric.

Finally, for $S_{n,1}(s,t)$, convergence in the supremum norm in $C[0,1]^2$ can be shown directly by noting that ζ_n converges in distribution to a standard normal random variable and

$$S_{n,1}(s,t) = c_N(s,t)\zeta_n$$

where

$$c_N(s,t) = \frac{1}{(Nb)^2}\sum_{j,k=1}^{N} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right)\phi(t_j)\phi(t_k)$$

and, with $u=(u_1,u_2)$, $v=(v_1,v_2)$,

$$|c_N(u)-c_N(v)| \le |\phi(u_1)\phi(u_2)-\phi(v_1)\phi(v_2)| + C_1 b^2 + C_2 (Nb)^{-1}$$

with suitable constants C_1, C_2 not depending on s, t, n and N. Therefore, for $\|u-v\| \le \Delta$, we have

$$
\begin{aligned}
& w_{n,N}(\Delta) \\
& = \sup_{|u-v|\le\Delta} |S_{n,N,1}(u_1,u_2)-S_{n,N,1}(v_1,v_2)| \\
& \le \left(w_{\phi,\phi}(\Delta)+C_1 b^2 + C_2(Nb)^{-1}\right)|\zeta_n|
\end{aligned}
$$

where

$$w_{\phi,\phi}(\Delta) = \sup_{\|u-v\|\leq\Delta} |\phi(u_1)\phi(u_2) - \phi(v_1)\phi(v_2)|$$

is the modulus of continuity of $\phi(\cdot)\phi(\cdot)$. For $\tau > 0$ we have

$$P\left(w_{n,N}(\Delta) > \tau\right) \leq P\left(|\zeta_n| > \frac{\tau}{w_{\phi,\phi}(\Delta) + C_1 b^2 + C_2(Nb)^{-1}}\right).$$

Taking the lim sup over n and N of the two hand sides such that conditions (3.23) and (3.24) hold, we then have

$$\limsup_{n,N} P\left(w_{n,N}(\Delta) > \tau\right) \leq 2\left(1 - \Phi\left(\frac{\tau}{w_{\phi,\phi}(\Delta)}\right)\right).$$

Since $\phi(t)$ is uniformly continuous, let $\Delta \to 0$, we have

$$\lim_{\Delta\to 0}\limsup_{n,N} P\left(w_{n,N}(\Delta) > \tau\right) = 0.$$

Then theorem 7.5 in Billingsley (1999, p. 84) implies

$$S_{n,N,1}(s,t) \Rightarrow \sqrt{2}\lambda\phi(s)\phi(t)\zeta$$

where ζ is a standard normal random variable.

For the general case with an arbitrary number of components $\phi_l(t)$ ($l \in \mathbb{N}$), we can write

$$\begin{aligned} S_{n,1}(s,t) &= \frac{1}{(Nb)^2}\sum_{j,k=1}^{N} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right)\left(\zeta_{n,1}(t_j,t_k) + \zeta_{n,2}(t_j,t_k)\right) \\ &=: S_{n,1;1}(s,t) + S_{n,1;2}(s,t) \end{aligned}$$

with

$$\zeta_{n,1} = n^{-\frac{1}{2}}\sum_{i=1}^{n}\sum_{l}(\xi_{il}^2 - \lambda_l)\phi_l(t_j)\phi_l(t_k),$$

$$\zeta_{n,2} = n^{-\frac{1}{2}}\sum_{i=1}^{n}\sum_{l_1\neq l_2}\xi_{il_1}\xi_{il_2}\phi_{l_1}(t_j)\phi_{l_2}(t_k).$$

It is easy to see that, for all s, t, s', t', we have

$$cov\left(S_{n,1;1}(s,t), S_{n,1;2}(s',t')\right) = 0$$

$$\lim_{n\to\infty} cov\left(S_{n,1;1}(s,t), S_{n,1;1}(s',t')\right) = 2\sum_{l}\lambda_l^2\phi_l(s)\phi_l(t)\phi_l(s')\phi_l(t')$$

and

$$\lim_{n\to\infty} cov\left(S_{n,1;2}(s,t), S_{n,1;2}(s',t')\right)$$
$$= \sum_{l_1<l_2} \lambda_{l_1}\lambda_{l_2}[\phi_{l_1}(s)\phi_{l_2}(t)+\phi_{l_2}(s)\phi_{l_1}(t)][\phi_{l_1}(s')\phi_{l_2}(t')+\phi_{l_2}(s')\phi_{l_1}(t')].$$

Tightness of $S_{n,1;1}(t,s)$ and $S_{n,1;2}(t,s)$ follows by analogous arguments as for the simple case with one basis function we discussed before. □

Proof. (of Theorem 3.5) The proof is analogous to the derivation of Theorem 3.4, and is therefore omitted. □

3.5.2 Tables

	σ_N^2			$IMSE_N$		
	(a)	(b)	(c)	(a)	(b)	(c)
$N = 100$, $n = 158$	0.683	0.933	0.722	0.687	0.937	0.726
$N = 200$, $n = 240$	0.602	0.762	0.645	0.606	0.766	0.648
$N = 400$, $n = 364$	0.551	0.663	0.587	0.554	0.667	0.590
$N = 600$, $n = 464$	0.528	0.618	0.570	0.532	0.622	0.574
$N = 800$, $n = 552$	0.521	0.591	0.554	0.525	0.596	0.558
$N = 1000$, $n = 631$	0.512	0.567	0.540	0.517	0.571	0.544
$N = 2000$, $n = 956$	0.504	0.535	0.523	0.508	0.539	0.528
$N = 4000$, $n = 1450$	0.494	0.514	0.509	0.499	0.519	0.513

Table 3.1: Simulated values of the standardized integrated variance $\sigma_N^2 = n_N \int_0^1 var\left[\hat{\mu}(t)\right] dt$ and the standardized integrated mean squared error $IMSE_N = n_N \int_0^1 E\left[(\hat{\mu}(t)-\mu(t))^2\right] dt$. The results are based on 400 simulations of $Y_{ij} = \mu(t_j) + \xi_i\phi(t_j) + \epsilon_i(j)$ with $\mu(t) = 10 + 2t^2$, $\phi(t) = \sin 4\pi t$, ξ_i iid $N(0,1)$ variables and $\epsilon_i(j)$ with variance one generated by one of the following processes: (a) iid $N(0,1)$; (b) $AR(1)$ with lag-one correlation $\rho = 0.5$; (c) $FARIMA(0, 0.3, 0)$ process.

Chapter 4

Estimation of eigenvalues λ, eigenfunctions $\phi(t)$, and scores ξ

As discussed in Chapter 3, observations in functional data analysis (FDA) are often perturbed by random noise. The random noise may exhibit weakly dependence or strongly dependence. In this chapter we consider estimation of eigenvalues λ_l, eigenfunctions (functional principal components) $\phi_l(t)$ and functional principal component scores ξ_{il} in a FDA model which is perturbed by short- or long-range dependent error process. The corresponding estimators are based on the two-dimensional boundary kernel estimator of covariance $C(s,t)$ of the underlying random curve $X(t) \in L^2[0,1]$ which generates the observations (after orthonormal contrast transformation of the observations) (see Chapter 3). As it turns out, the asymptotic distribution of estimated eigenvalues $\hat{\lambda}_l$ and estimated eigenfunctions $\hat{\phi}_l(t)$ does not depend on the dependence structure of the error process. Although $\left\{\hat{\phi}_l(t)\right\}$ $(l \in \mathbb{N})$ do not have independent property, the joint asymptotic distribution of them for both short and long memory error process are also the same. However, this is not the case for the estimated functional principal component scores $\hat{\xi}_{il}$. In fact, the rate of convergence and the asymptotic distribution of $\left\{\hat{\xi}_{il}\right\}$ $(l \in \mathbb{N})$ differ distinctly between the cases of short- and long-range dependence. Somewhat surprisingly, under long-range dependence, $\hat{\xi}_{i1}, \hat{\xi}_{i2}, ...$ are no longer independent. A simulation example illustrates the asymptotic properties of $\hat{\lambda}_l$ and $\hat{\phi}_l(t)$. This chapter is based on our previous work in Beran and Liu (2016).

4.1 Estimators

Throughout this chapter, observations are still assumed to be of the form (3.1), (3.3), (3.4) or (3.5) as that in Chapter 3.

Recall that the one-dimensional kernel estimation of trend function $\mu(t)$ of the underlying random curve $X(t) \in L^2[0,1]$ is defined as in (3.6), i.e.

$$\hat{\mu}(t) = \frac{1}{Nb}\sum_{j=1}^{N} K_1\left(\frac{t-t_j}{b}\right) n^{-1}\sum_{i=1}^{n} Y_{ij}$$

where Y_{ij} $(i = 1, ..., n,\ j = 1, ..., N)$ are the observations, n is the number of random curves, N is the number of sampling points on each curve, $t_j = j/N$ denotes the rescaled times, $K_1(t)$ is the one-dimensional kernel function with support $[-1,1]$ and $b = b_N > 0$ is a bandwidth. The two-dimensional kernel estimation of covariance function $C(s,t)$ of $X(t)$ is given in (3.7), i.e.

$$\hat{C}(s,t) = \frac{1}{(Nb)^2}\sum_{j,k=1}^{N} K_2\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) n^{-1}\sum_{i=1}^{n}(Y_{ij} - \hat{\mu}(t_j))(Y_{ik} - \hat{\mu}(t_k))$$

where $K_2(s,t)$ is a two-dimensional kernel function with support $[-1,1]^2$ and $b = b_N > 0$ is a bandwidth.

As pointed out in Chapter 3, since $\mu(t)$ plays the role of a nuisance parameter that is of no interest when the focus is on $C(s,t)$ only, it is wise to remover $\mu(t)$ before estimating $C(s,t)$. Eliminating of $\mu(t)$ can be done without any asymptotic loss of efficiency (as $n \to \infty$) by using orthonormal contrasts. In the sense that, under the Gaussian assumption, the transformed model is equivalent in distribution to the original model with $\mu(t) \equiv 0$, except that n reduces to $n-1$. Therefore, we assume that $\mu(t)$ to be known and identically equal to zero, i.e.

$$Y_{ij} = \sum_{l=1}^{\infty} \xi_{il}\phi_l(t) + \epsilon_i(j).$$

To deal with the boundary effect in kernel regression, the boundary kernel is used as discussed in Chapter 3. In order to relax the restriction between n and N, higher order kernels are used and additional differentiability assumptions on $C(s,t)$ are imposed while estimating it (see Chapter 3 or Beran and Liu 2014).

Therefore, after the orthonormal contrast transformation of the original observations, the covariance estimator is given as in (3.21):

$$\hat{C}(s,t) = \frac{1}{(Nb)^2}\sum_{j,k=1}^{N} K_{2,b}\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) n^{-1}\sum_{i=1}^{n} Y_{ij}Y_{ik},$$

where $K_{2,b}(s,t)$ is a two-dimensional boundary higher order kernel function.

The eigenvalues λ_l and eigenfunctions (functional principal components) $\phi_l(t)$ of the covariance operator $\mathbf{C}$ with kernel $C(s,t)$ are estimated from

$$\int_0^1 \hat{C}(s,t)\hat{\phi}_l(s)ds = \hat{\lambda}_l\hat{\phi}_l(t), \tag{4.1}$$

where $\int_0^1 \hat{\phi}_l^2(t)dt = 1$ and $\int_0^1 \hat{\phi}_l(t)\hat{\phi}_m(t)dt = 0$ for $m < l$. Note that (4.1) follows from the Mercer's Theorem (see Chapter 2) on the estimated covairance

$$\hat{C}(s,t) = \sum_l \hat{\lambda}_l\hat{\phi}_l(s)\hat{\phi}_l(t) \quad (s,t \in [0,1]). \tag{4.2}$$

The functional principal component scores ξ_{il} are estimated by

$$\hat{\xi}_{il} = N^{-1}\langle Y_{i\cdot}, \hat{\phi}_{l\cdot}\rangle = N^{-1}\sum_{j=1}^{N} Y_{ij}\hat{\phi}_l(t_j), \tag{4.3}$$

with $t_j = j/N$, $Y_{i\cdot} = (Y_{i1}, ..., Y_{iN})^T$, and $\hat{\phi}_{l\cdot} = \left(\hat{\phi}_l(t_1), ..., \hat{\phi}_l(t_j)\right)^T$.

Note that, since the error processes $\epsilon_i(j)$ are existing and not independent,

$$\begin{aligned} cov(Y_{ij}, Y_{ik}) &= cov\left(X(t_j), X(t_k)\right) + cov(\epsilon_i(j), \epsilon_i(k)) \\ &= C(t_j, t_k) + \gamma_\epsilon(j-k) \end{aligned}$$

means that $\hat{\phi}_l(t)$, $\hat{\lambda}_l$ and $\hat{\xi}_{il}$ based on Y_{ij} differ from the corresponding estimates obtained from $X_i(t_j)$. One of the questions answered in the following will be in how far the effect of $\epsilon_i(j)$ is asymptotically negligible.

4.2 Asymptotic properties of $\hat{\lambda}$ and $\hat{\phi}(t)$

In this section we study asymptotic distribution of estimated eigenvalues $\hat{\lambda}_l$ and asymptotic distribution and joint asymptotic distribution of estimated eigenfunctions $\hat{\phi}_l(t)$ defined by (4.1) and (4.2).

4.2.1 Assumptions and notations

At first, we introduce some assumptions and notations.

- (A1) Assume that the eigenvalues are non-negative and identifiable. That is, for the first p largest eigenvalues, we have $\lambda_1 > \lambda_2 > ... > \lambda_p > \lambda_{p+1} > 0$.

- (A2) Since eigenfunctions are determined only up to a sign, we assume without loss of generality that the signs of $\hat{\phi}_l(t)$ are adjusted in the correct direction. That is $sign\left(\left\langle \hat{\phi}_l, \phi_l \right\rangle\right) = 1$ for $l = 1, ..., p$.

- (N1) Let $\mathbb{H}$ be a separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ which generates the norm $\| \cdot \|$.

- (N2) Denote by $\mathcal{L}$ the space of bounded linear operators on $\mathbb{H}$, equipped with the norm $\|\mathbf{T}\|_{\mathcal{L}} = \sup_{\|f\| \leq 1} \|\mathbf{T}(f)\|$.

- (N3) Define the operator $(g \otimes h)$ by $(g \otimes h)(f) = \langle g, f \rangle h$ for $f, g, h \in \mathbb{H}$.

- (N4) Denote by $\mathcal{S}$ the separable Hilbert space of Hilbert-Schmidt operators on $\mathbb{H}$, equipped with the Hilbert-Schmidt inner product $\langle \mathbf{T}_1, \mathbf{T}_2 \rangle_{\mathcal{S}} = trace(\mathbf{T}_1^* \mathbf{T}_2) = \sum_j \langle \mathbf{T}_1 u_j, \mathbf{T}_2 u_j \rangle$ $(\mathbf{T}_1, \mathbf{T}_2 \in \mathcal{S})$ and the norm $\|\mathbf{T}\|_{\mathcal{S}} = \sqrt{\langle \mathbf{T}, \mathbf{T} \rangle_{\mathcal{S}}}$ $(\mathbf{T} \in \mathcal{S})$, where $\{u_j : j \geq 1\}$ is any complete orthonormal system in $\mathbb{H}$.

- (N5) The covariance operators $\mathbf{C}$ and $\hat{\mathbf{C}}$ generated by the true covariance function $C(s,t)$ and the estimated covariance eigenfunction $\hat{C}(s,t)$ are defined as $\mathbf{C}(f)(t) = \int_0^1 C(s,t) f(s) ds$ and $\hat{\mathbf{C}}(f)(t) = \int_0^1 \hat{C}(s,t) f(s) ds$ respectively. It is obvious that $\mathbf{C}$ and $\hat{\mathbf{C}}$ are symmetric and positive-definite Hilbert-Schmidt operators so they admit the decompositions $\mathbf{C}(f) = \sum_{l=1}^{\infty} \lambda_l \langle f, \phi_l \rangle \phi_l$ and $\hat{\mathbf{C}}(f) = \sum_{l=1}^{\infty} \hat{\lambda}_l \langle f, \hat{\phi}_l \rangle \hat{\phi}_l$ respectively with orthonormal ϕ_l $(l \in \mathbb{N})$ and $\hat{\phi}_l$ $(l \in \mathbb{N})$ which are the eigenfunctions of $\mathbf{C}$ and $\hat{\mathbf{C}}$, i.e. $\mathbf{C}(\phi_l) = \lambda_l \phi_l$ and $\hat{\mathbf{C}}(\hat{\phi}_l) = \hat{\lambda}_l \hat{\phi}_l$.

- (N6) Define the orthooonal projection operators from $L^2[0,1]$ onto the subspaces spanned by the true eigenfunction $\phi_l(t)$ and the estimated eigenfunction $\hat{\phi}_l(t)$ respectively,

$$\mathbf{P}_l = \phi_l \otimes \phi_l, \hat{\mathbf{P}}_l = \hat{\phi}_l \otimes \hat{\phi}_l, \quad l \in \mathbb{N}.$$

4.2.2 Lemmas

To derive asymptotic properties of estimated eigenvalues $\hat{\lambda}_l$ and estimated eigenfunctions $\hat{\phi}_l(t)$, the following several lemmas will be used. First we consider weak convergence of sequence of operators

$$\mathbf{Z}_n = n^{1/2}(\hat{\mathbf{C}} - \mathbf{C})$$

which can be easily obtained from Theorem 3.5:

Lemma 4.1. *Suppose the assumptions in Theorem 3.5 hold. Then there exist iid standard normal random variables $\zeta_{l_1 l_2}$ $(l_1, l_2 \in \mathbb{N})$ such that*

$$\mathbf{Z}_n \Rightarrow \mathbf{Z} \quad (n \to \infty)$$

where $\mathbf{Z}$ is the operator

$$\mathbf{Z} = \sum_{l_1 < l_2} \sqrt{\lambda_{l_1}\lambda_{l_2}}(\phi_{l_1} \otimes \phi_{l_2} + \phi_{l_2} \otimes \phi_{l_1})\zeta_{l_1 l_2} + \sqrt{2}\sum_{l} \lambda_l(\phi_l \otimes \phi_l)\zeta_{ll} \tag{4.4}$$

and "$\Rightarrow$" denotes weak convergence in $\mathcal{S}$ equipped with the norm $\|\cdot\|_{\mathcal{S}}$.

Lemma 4.1 together with the results in Dauxois et al. (1982), Bosq (2000), Dunford and Schwartz (1988), and Kato (1976) implies consistency of $\hat{\mathbf{P}}_l$:

Lemma 4.2. *Let $\hat{\lambda}_l$, $\hat{\phi}_l(t)$ $(l \in \mathbb{N})$ be defined by (4.1) and (4.2), and suppose that the assumptions of Lemma 4.1 hold. Moreover, (A1) and (A2) hold. Then, for each $l \in \{1, ..., p\}$, we have*

$$\left\|\hat{\mathbf{P}}_l - \mathbf{P}_l\right\|_{\mathcal{S}} = O_p\left(n^{-1/2}\right).$$

Remark 4.1. *The assumptions in Lemma 4.2 can be generalized a little bit. Suppose that $\lambda_l \geq \lambda_{l+1}$ $(l \in \mathbb{N})$. For a given threshold $c_\lambda > 0$, let $I \subseteq \mathbb{N}$ denote the (finite) set of indices such that $\lambda_l > c_\lambda$ and $\lambda_l > \lambda_{l+1}$ $(l \in I)$. Then, for each $l \in I$, we have the same results about projection operators as that in Lemma 4.2. The assumptions in the following lemmas and theorems in this chapter can also be generalized similarly and the corresponding results hold.*

We now can formulate a consistency of $\hat{\lambda}_l$ which are defined in (4.1) and (4.2), the consistency of $\hat{\phi}_l(t)$ in (4.1) and (4.2) in the L^2 norm $\|\cdot\|$, and the uniform consistency of $\hat{\phi}_l(t)$ on $[0, 1]$:

Lemma 4.3. *Under the assumptions of Lemma 4.2. Then, for $l \in \{1, ..., p\}$, we have*

$$|\hat{\lambda}_l - \lambda_l| = O_p(n^{-1/2}), \tag{4.5}$$

$$\|\hat{\phi}_l - \phi_l\| = O_p(n^{-1/2}), \tag{4.6}$$

and

$$\sup_{t\in[0,1]} |\hat{\phi}_l(t) - \phi_l(t)| = O_p(n^{-1/2}). \tag{4.7}$$

Remark 4.2. *The rates in (4.5), (4.6) and (4.7) are direct consequences of the rate in Lemma 4.1.*

To derive asymptotic properties of $\hat{\lambda}_l$ and $\hat{\phi}_l(t)$, the following Lemma 4.4 about the convergence of $\hat{\mathbf{P}}_l$ will also be used. It can be proved by extending the approaches in Dauxois et al. (1982), Dunford and Schwartz (1988), and Bosq (2000) to FDA processes with random perturbations.

Lemma 4.4. *Under the assumptions of Lemma 4.2, and $l \in \{1, ..., p\}$, $n^{1/2}(\hat{\mathbf{P}}_l - \mathbf{P}_l)$ converges weakly in $\mathcal{S}$ to a Gaussian random element with mean zero.*

4.2.3 CLT for $\hat{\lambda}$ and $\hat{\phi}(t)$

Asymptotic normality of $n^{1/2}(\hat{\mathbf{P}}_l - \mathbf{P}_l)$ implies a central limit theorem for the estimated eigenvalues and eigenfunctions:

Theorem 4.1. *Assume that the assumptions of Lemma 4.2 hold, and $l \in \{1, ..., p\}$. Then*

$$n^{1/2}\left(\hat{\lambda}_l - \lambda_l\right) \to_d \sqrt{2}\lambda_l\zeta_{ll}, \tag{4.8}$$

and

$$\begin{aligned} & n^{1/2}\left(\hat{\phi}_l(t) - \phi_l(t)\right) \\ & \Rightarrow Z_l(t) := \sum_{k:k>l} \sqrt{\lambda_l\lambda_k}\left(\lambda_l - \lambda_k\right)^{-1} \phi_k\zeta_{lk} + \sum_{k:k<l} \sqrt{\lambda_l\lambda_k}\left(\lambda_l - \lambda_k\right)^{-1} \phi_k\zeta_{kl} \end{aligned} \tag{4.9}$$

where "$\Rightarrow$" denotes weak convergence in $C[0,1]$ equipped with the supremum norm, and ζ_{ij} are iid standard normal variables.

Remark 4.3. *These results coincide with those in Dauxois et al. (1982) (see Proposition 8 and Proposition 10) where the estimation based on the directly observed random curves $X_i(t)$ is considered. However, since the functions $X_i(t)$ are not observed directly, the situation considered here is more complicated. This leads to additional conditions as stated in (3.16) that the number of replicates n should not increase too fast compared with the number of sampling points on each curve N. As discussed in Chapter 3 or in Beran and Liu (2014), in order to relax this unpleasant restriction, we need to use higher order kernels while estimating the covariance and impose differentiable properties on the covariance. On the other hand, from the improved condition (3.26), we can see that, for fixed estimated long memory parameter d, even with the "best" kernels the number of replicates cannot grow faster than $o(N^2)$.*

Remark 4.4. *The spacings of eigenvalues influence the asymptotic distribution of $\hat{\phi}_l(t)$. However, the asymptotic distribution of $\hat{\lambda}_l$ does not depend on the spacing of the eigenvalues.*

The following Theorem 4.2 gives the asymptotic joint distribution of the estimated eigenfunctions $\hat{\phi}_l(t)$.

Theorem 4.2. *Suppose the assumptions from Lemma 4.2 hold, then for any p which satisfies (A1),*

$$\sqrt{n}\left(\hat{\phi}_1(t)-\phi_1(t),...,\hat{\phi}_p(t)-\phi_p(t)\right)^T \Rightarrow (Z_1(t),...,Z_p(t))^T \tag{4.10}$$

with

$$Z_l(t)=\sum_{k:k>l}\sqrt{\lambda_l\lambda_k}(\lambda_l-\lambda_k)^{-1}\phi_k\zeta_{lk}+\sum_{k:k<l}\sqrt{\lambda_l\lambda_k}(\lambda_l-\lambda_k)^{-1}\phi_k\zeta_{kl}.$$

Moreover,

$$cov(Z_l(s),Z_l(t))=\sum_{k:k\neq l}\lambda_l\lambda_k(\lambda_l-\lambda_k)^{-2}\phi_l(s)\phi_l(t)$$

and

$$cov(Z_{l_1}(t),Z_{l_2}(t))=-\lambda_{l_1}\lambda_{l_2}(\lambda_{l_1}-\lambda_{l_2})^{-2}\phi_{l_1}(s)\phi_{l_2}(t)\quad(l_1\neq l_2)$$

where ζ_{lk} are iid standard normal random variables for $k:k>l$, ζ_{kl} are iid standard normal random variables for $k:k<l$, ζ_{lk} are independent of ζ_{kl}, and "$\Rightarrow$" denotes weak convergence in $C\left[0,1\right]^2$ equipped with the supremum norm.

Remark 4.5. *There is no asymptotic independence property of estimated eigenfunctions.*

4.3 Asymptotic properties of $\hat{\xi}$

Our aim here is to discuss asymptotic properties of the estimated principal component scores $\hat{\xi}_{il}$ defined in (4.3). In contrast to $\hat{\lambda}_l$, $\hat{\phi}_l(t)$, the rate of convergence and asymptotic distribution of $\hat{\xi}_{il}$ differ distinctly, depending on whether we have short- or long-range dependent errors. Moreover the different score estimators $\hat{\xi}_{i1}, \hat{\xi}_{i2}, ...$ are no longer independent for long-range dependent case. First we consider short-range dependence.

4.3.1 Short memory case

Theorem 4.3. *Assume that the assumptions of Lemma 4.2 hold, and*

$$\sum_{k=-\infty}^{\infty} |\gamma_\epsilon(k)| < \infty.$$

Define

$$\zeta_{i,n} = (\zeta_{i,n;1}, ..., \zeta_{i,n;p})^T := \sqrt{N}\left(\hat{\xi}_{i,1} - \xi_{i,1}, ..., \hat{\xi}_{i,p} - \xi_{i,p}\right)^T \quad i = 1, ..., n.$$

Then, for each $i = 1, ..., n$,

$$\zeta_{i,n} \underset{d}{\to} \zeta \sim N(0, V)$$

where $V_{l_1 l_2} = 0$ $(l_1 \neq l_2)$ and $V_{ll} = 2\pi f_\epsilon(0)$, with $V = [V_{l_1 l_2}]_{l_1, l_2 = 1, ..., p}$.

4.3.2 Long memory case

The following Theorem 4.4 shows that a completely different result is obtained for strongly dependent error processes.

Theorem 4.4. *Assume that the assumptions of Lemma 4.2 hold, and*

$$\gamma_\epsilon(k) = cov\left(\epsilon_i(j), \epsilon_i(j+k)\right) \underset{k\to\infty}{\sim} c_\gamma |k|^{2d-1}$$

for some $0 < c_\gamma < \infty$, $0 < d < \frac{1}{2}$. Furthermore let

$$V_{l_1 l_2} = c_\gamma \int_{-1}^{1}\int_{-1}^{1} |u - v|^{2d-1} \phi_{l_1}(u)\phi_{l_2}(v) du dv \quad (l_1, l_2 = 1, ..., p)$$

and

$$\zeta_{i,n} := N^{\frac{1}{2}-d}\left(\hat{\xi}_{i,1} - \xi_{i,1}, ..., \hat{\xi}_{i,p} - \xi_{i,p}\right)^T.$$

Then, for each i,

$$\zeta_{i,n} \underset{d}{\to} \zeta_i \sim N(0, V)$$

with $V = [V_{l_1 l_2}]_{l_1,l_2=1,...,p}$.

Remark 4.6. *It is worth noting that, compared to the case of short-range dependence, long-range dependence leads to a slower rate of convergence of $\hat{\xi}_{il}$. Moreover, for the long-range dependent errors, $\hat{\xi}_{il_1}$ and $\hat{\xi}_{il_2}$ ($l_1 \neq l_2$) are no longer asymptotically independent. The strength of the dependence depends on the long memory parameter d and the eigenfunctions $\phi_l(t)$.*

4.4 Simulations

To illustrate the asymptotic results of estimated eigenvalues $\hat{\lambda}_l$ and estimated eigenfunctions $\hat{\phi}_l(t)$, we set up the following simulation. We consider the simple model with only two basis functions defined by

$$Y_{ij} = X_i(t_j) + \epsilon_i(j) = \xi_{i1}\phi_1(t_j) + \xi_{i2}\phi_2(t_j) + \epsilon_i(j) \quad (i = 1, ..., n;\ j = 1, ..., N;\ t_j = j/N) \tag{4.11}$$

with

$$\xi_{i1} \sim N(0, \lambda_1) = N(0, 4),$$

$$\xi_{i2} \sim N(0, \lambda_2) = N(0, 2)$$

and

$$\phi_1(t) = \sqrt{2}\cos(\pi t),$$

$$\phi_2(t) = \sqrt{2}\cos(2\pi t).$$

Similar to the simulation design in Chapter 3, the error process $\epsilon_i(j)$ is assumed to have variance one and is generated by one of the following processes: (a) iid $N(0, 1)$; (b) AR(1) process with lag-one correlation $\rho = 0.5$; (c) FARIMA$(0, d, 0)$ process with long memory parameter $d = 0.3$. Note that in the two short memory cases (a) and (b), $d = 0$. The number of sampling points on each cure N are chosen as 100, 200, 400, 600, 800, 1000, 1500 and 2000. Then according to the conditions in Theorem 4.1 (or Theorem 3.5), the number of sampling points on each curve n_N can be set equal to $n = 10N^{0.6}$ rounded to the next integer (i.e. $n = 158$, 240, 364, 464, 552, 631, 805 and 956 respectively).

For each pair (N, n_N), 200 simulations were carried out. For each simulated series, $\hat{\lambda}_l$ and $\hat{\phi}_l(t)$ are calculated using kernel estimator $\hat{C}(s,t)$ defined by (3.21) based on the bandwidth $b = b_N = 0.05N^{-0.16}$ and product kernel $K_2(u,v) = K_1(u)K_1(v)$ with $K_1(u) = \frac{1}{2}\mathbf{1}\{-1 \le u \le 1\}$. Note that $n_N^{-\frac{1}{4}} = 10^{-1/4}N^{-0.15}$, so the conditions in Theorem 4.1 hold. In order to save calculation time, we set the gird of $\hat{C}(s,t)$ equal to 500, i.e. the discrete matrix $\hat{C}(s_j, t_k)$ $(j, k = 1, ..., 500)$ with $s_j = j/500$ and $t_k = k/500$ is obtained for each simulation. The corresponding eigenvalues and eigenfunctions which are in discrete forms (eigenvectors) with dimension 500 are calculated from the discrete matrix $\hat{C}(s_j, t_k)$ $(j, k = 1, ..., 500)$.

According to Theorem 4.1, we have

$$\sqrt{n}\left(\hat{\lambda}_1 - \lambda_1\right) \underset{d}{\rightarrow} N(0, 32),$$

$$\sqrt{n}\left(\hat{\lambda}_2 - \lambda_2\right) \underset{d}{\rightarrow} N(0, 8)$$

and

$$n^{1/2}\left(\hat{\phi}_1(t) - \phi_1(t)\right) \underset{d}{\rightarrow} N\left(0, \lambda_1\lambda_2(\lambda_1 - \lambda_2)^{-2}\phi_2^2(t)\right) = N\left(0, 4\cos^2(2\pi t)\right),$$

$$n^{1/2}\left(\hat{\phi}_2(t) - \phi_2(t)\right) \underset{d}{\rightarrow} N\left(0, \lambda_2\lambda_1(\lambda_2 - \lambda_1)^{-2}\phi_1^2(t)\right) = N\left(0, 4\cos^2(\pi t)\right).$$

Therefore, the theoretical asymptotic values of the standardized squared bias $B^2_{\lambda_1,N} = n_N[E(\hat{\lambda}_1) - \lambda_1]^2$, variance $\sigma^2_{\lambda_1,N} = n_N var(\hat{\lambda}_1)$ and mean squared error $MSE_{\lambda_1,N} = n_N E[(\hat{\lambda}_1 - \lambda_1)^2] = B^2_{\lambda_1,N} + \sigma^2_{\lambda_1,N}$ of $\hat{\lambda}_1$ should be 0, 32 and 32 respectively. The simulated values of $B^2_{\lambda_1,N}$, $\sigma^2_{\lambda_1,N}$ and $MSE_{\lambda_1,N}$ given in Table 4.1 essentially agree with the theoretical values. The simulation results also show that, under long memory error process (c) with $d = 0.3 \in (0, \frac{1}{2})$, the bias $B_{\lambda_1,N}$ appears to converge to zero at a slower rate than for the two short memory error processes (a) and (b) with $d = 0$. In fact, it can be seen that, in the example considered here, the bias term $B_{\lambda_1,N}$ is dominated by the variance term $\sigma^2_{\lambda_1,N}$ so the MSE is not really affected by $B^2_{\lambda_1,N}$.

Table 4.2 shows the same quantities for $\hat{\lambda}_2$. Here the theoretical asymptotic values of $B^2_{\lambda_2,N}$, $\sigma^2_{\lambda_2,N}$ and $MSE_{\lambda_2,N}$ should be 0, 8 and 8 respectively. Analogous comments are applicable for $\hat{\lambda}_1$.

Table 4.3 shows simulated values of the standardized integrated squared bias $B^2_{\phi_1,N}$, variance $\sigma^2_{\phi_1,N}$ and mean squared error $IMSE_{\phi_1,N}$ of $\hat{\phi}_1(t)$. Specifically,

$$B^2_{\phi_1,N} = n_N \int_0^1 \left\{ E\left[\hat{\phi}_1(t)\right] - \phi_1(t) \right\}^2 dt,$$

$$\sigma^2_{\phi_1,N} = n_N \int_0^1 var\left(\hat{\phi}_l(t)\right) dt = n_N \int_0^1 E\left[\left\{\hat{\phi}_1(t) - E\left[\hat{\phi}_1(t)\right]\right\}^2\right] dt,$$

and $IMSE_{\phi_1,N} = B^2_{\phi_1,N} + \sigma^2_{\phi_1,N}$. The corresponding theoretical asymptotic values of $B^2_{\phi_1,N}$, $\sigma^2_{\phi_1,N}$ and $IMSE_{\phi_1,N}$ are 0, 2 and 2 respectively. This is essentially confirmed by the simulated values in Table 4.3. The same is true for the simulated values of $\hat{\phi}_2(t)$ given in Table 4.4.

4.5 Proofs and tables

4.5.1 Proofs

Proof. (of Lemma 4.1)

As derived in Theorem 3.5, we have, under the given assumptions

$$\begin{aligned} Z_{n,N}(s,t) :&= \sqrt{n}\left(\hat{C}(s,t) - C(s,t)\right) \\ &\Rightarrow Z(s,t) \\ &= Z_1(s,t) + Z_2(s,t) \end{aligned}$$

where "$\Rightarrow$" denotes weak convergence in $C[0,1]^2$ equipped with the supremum norm, $Z_1(s,t)$, $Z_2(s,t)$ are Gaussian processes with zero mean, $Z_1(s,t)$ and $Z_2(s,t)$ are independent from each other, and have covariance functions

$$cov\left(Z_1(t,s), Z_1(t',s')\right) = 2\sum_l \lambda_l^2 \phi_l(s)\phi_l(t)\phi_l(s')\phi_l(t')$$

and

$$cov\left(Z_2(t,s), Z_2(t',s')\right)$$

$$= \sum_{l_1<l_2} \lambda_{l_1}\lambda_{l_2}\left[\phi_{l_1}(s)\phi_{l_2}(t) + \phi_{l_2}(s)\phi_{l_1}(t)\right]\left[\phi_{l_1}(s')\phi_{l_2}(t') + \phi_{l_2}(s')\phi_{l_1}(t')\right].$$

Now, since the sequence of integral operators $\mathbf{Z}_n = n^{1/2}(\hat{\mathbf{C}} - \mathbf{C}) \in \mathcal{S}$ and the integral operator $\mathbf{Z} \in \mathcal{S}$, we have

$$\|\mathbf{Z}_n - \mathbf{Z}\|_{\mathcal{S}}^2 = \int_0^1 \int_0^1 \left\{ \sqrt{n} \left[\hat{C}(s,t) - C(s,t) \right] - Z(s,t) \right\}^2 dsdt.$$

Thus, for any $\Delta > 0$,

$$\lim_{n\to\infty} P\left(\sup_{s,t\in[0,1]} \left| \sqrt{n} \left[\hat{C}(s,t) - C(s,t) \right] - Z(s,t) \right| \geq \Delta \right) = 0$$

and hence

$$\begin{aligned}
&\lim_{n\to\infty} P(\|\mathbf{Z}_n - \mathbf{Z}\|_{\mathcal{S}} \geq \Delta) \\
&\leq \lim_{n\to\infty} P\left(\sup_{s,t\in[0,1]} \left| \sqrt{n} \left[\hat{C}(s,t) - C(s,t) \right] - Z(s,t) \right| \geq \Delta \right) \\
&= 0.
\end{aligned}$$

Therefore, the weak convergence of $\mathbf{Z}_n$ in $\mathcal{S}$ equipped with the norm $\|\cdot\|_{\mathcal{S}}$ follows from Billingsley (1999). □

Proof. (of Lemma 4.2)

Let $l \in \{1, ..., p\}$ and a constant $\rho > 0$ such that

$$0 < \rho < \frac{1}{2} \min_{k,l\in\{1,...,p\},\ k\neq l} |\lambda_k - \lambda_l|.$$

Define a circle with center λ_l and radius ρ on the complex plane

$$\Lambda_{\rho,l} = \{z \in \mathbb{C} : |z - \lambda_l| = \rho\}$$

and consider the resolvents of covariance operators $\mathbf{C}$ and $\hat{\mathbf{C}}$

$$\mathbf{R}(z) = (\mathbf{C} - z\mathbf{I})^{-1},$$

and

$$\hat{\mathbf{R}}(z) = (\hat{\mathbf{C}} - z\mathbf{I})^{-1}.$$

Then the projection operators (Riesz projection, Kato (1976 p. 178)) $\mathbf{P}_l$ and $\hat{\mathbf{P}}_l$ can be written as

$$\mathbf{P}_l = -\frac{1}{2\pi i} \int_{\Lambda_{\rho,l}} \mathbf{R}(z) dz,$$

$$\hat{\mathbf{P}}_l = -\frac{1}{2\pi i}\int_{\Lambda_{\rho,l}} \hat{\mathbf{R}}(z)dz.$$

Since

$$\begin{aligned}
\hat{\mathbf{R}}(z) &= (\hat{\mathbf{C}} - z\mathbf{I})^{-1} \\
&= [(\mathbf{C} - z\mathbf{I}) + (\mathbf{C} - z\mathbf{I})(\hat{\mathbf{C}} - \mathbf{C})(\mathbf{C} - z\mathbf{I})^{-1}]^{-1} \\
&= \mathbf{R}(z)[\mathbf{I} + (\hat{\mathbf{C}} - \mathbf{C})\mathbf{R}(z)]^{-1} \\
&= \mathbf{R}(z)\sum_{k=0}^{\infty}[(\mathbf{C} - \hat{\mathbf{C}})\mathbf{R}(z)]^k, \qquad (4.12)
\end{aligned}$$

we have

$$\begin{aligned}
\|\hat{\mathbf{R}}(z) - \mathbf{R}(z)\|_{\mathcal{S}} &= \left\|\mathbf{R}(z)\frac{(\hat{\mathbf{C}} - \mathbf{C})\mathbf{R}(z)}{\mathbf{I} - (\mathbf{C} - \hat{\mathbf{C}})\mathbf{R}(z)}\right\|_{\mathcal{S}} \\
&\leq \frac{\|\hat{\mathbf{C}} - \mathbf{C}\|_{\mathcal{S}}\|\mathbf{R}(z)\|_{\mathcal{S}}^2}{1 - \|\hat{\mathbf{C}} - \mathbf{C}\|_{\mathcal{S}}\|\mathbf{R}(z)\|_{\mathcal{S}}}.
\end{aligned}$$

This implies

$$\begin{aligned}
\|\hat{\mathbf{P}}_l - \mathbf{P}_l\|_{\mathcal{S}} &\leq \frac{1}{2\pi}\int_{\Lambda_{\rho,l}} \|\hat{\mathbf{R}}(z) - \mathbf{R}(z)\|_{\mathcal{S}}dz \qquad (4.13) \\
&\leq \rho\frac{\|\hat{\mathbf{C}} - \mathbf{C}\|_{\mathcal{S}}M_{\rho,l}^2}{1 - \|\hat{\mathbf{C}} - \mathbf{C}\|_{\mathcal{S}}M_{\rho,l}}
\end{aligned}$$

where

$$M_{\rho,l} = \sup_{z\in\Lambda_{\rho,l}} \|\mathbf{R}(z)\|_{\mathcal{S}} < \infty.$$

Now, note that Lemma 4.1 implies, for an arbitrary δ with $0 < \delta < \frac{1}{2}M_{\rho,l}^{-1}$ there exists $n_0 \in \mathbb{N}$ such that, $\|\hat{\mathbf{C}} - \mathbf{C}\|_{\mathcal{S}} < \delta$ $(n > n_0)$. Therefore (4.13) implies

$$\|\hat{\mathbf{P}}_l - \mathbf{P}_l\|_{\mathcal{S}} \leq 2\rho M_{\rho,l}^2\|\hat{\mathbf{C}} - \mathbf{C}\|_{\mathcal{S}}.$$

Hence, by noticing Lemma 4.1, we have

$$\left\|\hat{\mathbf{P}}_l - \mathbf{P}_l\right\|_{\mathcal{S}} = O_p\left(n^{-1/2}\right).$$

□

Proof. (of Lemma 4.3)

First, recall that

$$\left|\hat{\lambda}_l - \lambda_l\right| \leq \left\|\hat{\mathbf{C}} - \mathbf{C}\right\|_{\mathcal{L}} \leq \left\|\hat{\mathbf{C}} - \mathbf{C}\right\|_{\mathcal{S}}$$

(see e.g. Horváth and Kokoszka 2012, Lemma 2.2; Bosq 2000, Lemma 4.2), (4.5) follows from Lemma 4.1 .

Then we consider $\hat{\phi}_l(t)$. Without loss of generality, we may assume (A2) hold, i.e. $\langle\hat{\phi}_l, \phi_l\rangle \geq 0$. Given a complete orthonormal system $\{u_j : j \geq 1\}$ in $\mathbb{H}$, we have

$$\begin{aligned}
&\left\|\hat{\mathbf{P}}_l - \mathbf{P}_l\right\|_{\mathcal{S}}^2 \\
&= \left\|\hat{\phi}_l \otimes \hat{\phi}_l - \phi_l \otimes \phi_l\right\|_{\mathcal{S}}^2 \\
&= \sum_j \left\langle (\hat{\phi}_l \otimes \hat{\phi}_l - \phi_l \otimes \phi_l)u_j, (\hat{\phi}_l \otimes \hat{\phi}_l - \phi_l \otimes \phi_l)u_j \right\rangle \\
&= \sum_j \left\{ \langle\hat{\phi}_l, u_j\rangle^2 - 2\langle\hat{\phi}_l, u_j\rangle\langle\phi_l, u_j\rangle\langle\hat{\phi}_l, \phi_l\rangle + \langle\phi_l, u_j\rangle^2 \right\} \\
&= 2\left(1 - \langle\hat{\phi}_l \otimes \hat{\phi}_l, \phi_l \otimes \phi_l\rangle\right) \\
&= 2\left(1 - \langle\hat{\phi}_l, \phi_l\rangle^2\right) \\
&= \left\|\hat{\phi}_l - \phi_l\right\|^2 \left(1 + \langle\hat{\phi}_l, \phi_l\rangle\right).
\end{aligned}$$

Therefore,

$$\left\|\hat{\phi}_l - \phi_l\right\|^2 \leq \left\|\hat{\mathbf{P}}_l - \mathbf{P}_l\right\|_{\mathcal{S}}^2 = O_p\left(n^{-1}\right).$$

and (4.6) follows.

To obtain (4.7), note that, for fixed $l \in \{1, ..., p\}$, we have

$$\begin{aligned}
&|\hat{\lambda}_l\hat{\phi}_l(t) - \lambda_l\phi_l(t)| \\
&= \left|\int \hat{C}(s,t)\hat{\phi}_l(s)ds - \int C(s,t)\phi_l(s)ds\right| \\
&\leq \int \left|\hat{C}(s,t) - C(s,t)\right| \cdot \left|\hat{\phi}_l(s)\right| ds + \int |C(s,t)| \cdot \left|\hat{\phi}_l(s) - \phi_l(s)\right| ds \\
&\leq \sqrt{\int \left(\hat{C}(s,t) - C(s,t)\right)^2 ds} + \sqrt{\int C^2(s,t)ds}\left\|\hat{\phi}_l - \phi_l\right\|.
\end{aligned}$$

Since Lemma 4.1 and (4.6) hold and $\lambda_l > 0$, we obtain

$$\sup_{t\in[0,1]} \left|\hat{\lambda}_l\hat{\phi}_l(t)/\lambda_l - \phi_l(t)\right| = O_p(n^{-1/2}),$$

uniformly in $t \in [0,1]$. Then due to (4.5), this implies (4.7). $\square$

Proof. (of Lemma 4.4)

Recall that projection operators $\mathbf{P}_l$ and $\hat{\mathbf{P}}_l$ have the form

$$\mathbf{P}_l = -\frac{1}{2\pi i}\int_{\Lambda_{\rho,l}} \mathbf{R}(z)dz,$$

$$\hat{\mathbf{P}}_l = -\frac{1}{2\pi i}\int_{\Lambda_{\rho,l}} \hat{\mathbf{R}}(z)dz,$$

and the relationship of resolvents $\mathbf{R}$ and $\hat{\mathbf{R}}$ is

$$\begin{aligned}\hat{\mathbf{R}}(z) &= \mathbf{R}(z)[\mathbf{I} + (\hat{\mathbf{C}} - \mathbf{C})\mathbf{R}(z)]^{-1} \\ &= \mathbf{R}(z)\sum_{k=0}^{\infty}[(\mathbf{C} - \hat{\mathbf{C}})\mathbf{R}(z)]^k.\end{aligned}$$

Using the notation $\mathbf{Z}_n = n^{1/2}(\hat{\mathbf{C}} - \mathbf{C})$ and $\mathbf{H}_n(z) = \sum_{k=0}^{\infty} -n^{-1/2}[\mathbf{Z}_n\mathbf{R}(z)]^k$, we have

$$\begin{aligned}n^{1/2}\left(\hat{\mathbf{P}}_l - \mathbf{P}_l\right) &= n^{1/2}\int_{\Lambda_{\rho,l}} \left[\mathbf{R}(z) - \hat{\mathbf{R}}(z)\right] dz/2i\pi \\ &= n^{1/2}\frac{1}{2\pi i}\int_{\Lambda_{\rho,l}} \left\{\mathbf{R}(z) - \mathbf{R}(z)\sum_{k=0}^{\infty}\left[(\mathbf{C} - \hat{\mathbf{C}})\mathbf{R}(z)\right]^k\right\} dz \\ &= -n^{1/2}\frac{1}{2\pi i}\int_{\Lambda_{\rho,l}} \mathbf{R}(z)\sum_{k=1}^{\infty}\left[(\mathbf{C} - \hat{\mathbf{C}})\mathbf{R}(z)\right]^k dz \\ &= -n^{1/2}\frac{1}{2\pi i}\int_{\Lambda_{\rho,l}} \mathbf{R}(z)(\mathbf{C} - \hat{\mathbf{C}})\mathbf{R}(z)\sum_{k=0}^{\infty}\left[(\mathbf{C} - \hat{\mathbf{C}})\mathbf{R}(z)\right]^k dz \\ &= \frac{1}{2\pi i}\int_{\Lambda_{\rho,l}} \mathbf{R}(z)\mathbf{Z}_n\mathbf{R}(z)\mathbf{H}_n(z)dz \\ &=: \eta_l^n(\mathbf{Z}_n).\end{aligned}$$

It is easy to verify that the mapping η_l^n satisfies the Rubin-Billingsley conditions of Theorem 5.5 in Billingsley (1968, p. 34). Let

$$\begin{aligned}\mathbf{A}_n(z) &= n^{1/2}\left[\hat{\mathbf{R}}(z) - \mathbf{R}(z)\right] + \mathbf{R}(z)\mathbf{Z}\mathbf{R}(z) \\ &= -\mathbf{R}(z)\left(\mathbf{Z}_n - \mathbf{Z}\right)\mathbf{R}(z)\mathbf{H}_n(z) - \mathbf{R}(z)\mathbf{Z}\mathbf{R}(z)\left(\mathbf{H}_n(z) - \mathbf{I}\right).\end{aligned}$$

Then

$$\begin{aligned}
&\left\|\eta_l^n(\mathbf{Z}_n) - \frac{1}{2\pi i}\int_{\Lambda_{\rho,l}} \mathbf{R}(z)\mathbf{Z}\mathbf{R}(z)dz\right\|_{\mathcal{S}} \\
&= \left\|-\frac{1}{2\pi i}\int_{\Lambda_{\rho,l}} \mathbf{A}_n(z)dz\right\|_{\mathcal{S}} \\
&\leq \frac{1}{2\pi}\int_{\Lambda_{\rho,l}} \|\mathbf{A}_n(z)\|_{\mathcal{S}}dz.
\end{aligned}$$

Note that $\|\hat{\mathbf{C}} - \mathbf{C}\| \leq (2M_{\rho,l})^{-1}$ (see proof in Lemma 4.2) implies

$$\|\mathbf{H}_n(z)\|_{\mathcal{S}} \leq \sum_{k=0}^{\infty}[\|\hat{\mathbf{C}} - \mathbf{C}\|_{\mathcal{S}}M_{\rho,l}]^k = [1 - \|\hat{\mathbf{C}} - \mathbf{C}\|_{\mathcal{S}}M_{\rho,l}]^{-1} \leq 2,$$

we then have

$$\|\mathbf{H}_n(z) - \mathbf{I}\|_{\mathcal{S}} \leq 2M_{\rho,l}\left\|\hat{\mathbf{C}} - \mathbf{C}\right\|_{\mathcal{S}}.$$

Therefore,

$$\begin{aligned}
\|\mathbf{A}_n(z)\|_{\mathcal{S}} &\leq \|\mathbf{R}(z)(\mathbf{Z}_n - \mathbf{Z})\mathbf{R}(z)\mathbf{H}_n(z)\|_{\mathcal{S}} + \|\mathbf{R}(z)\mathbf{Z}\mathbf{R}(z)(\mathbf{H}_n(z) - \mathbf{I})\|_{\mathcal{S}} \\
&\leq M_{\rho,l}^2\|\mathbf{Z}_n - \mathbf{Z}\|_{\mathcal{S}}\|\mathbf{H}_n(z)\|_{\mathcal{S}} + M_{\rho,l}^2\|\mathbf{Z}\|_{\mathcal{S}}\|\mathbf{H}_n(z) - \mathbf{I}\|_{\mathcal{S}} \\
&\to 0.
\end{aligned}$$

Then separability of $\mathcal{S}$, (4.4) and the Rubin-Billingsley theorem imply

$$\eta_l^n(\mathbf{Z}_n) \Rightarrow \frac{1}{2\pi i}\int_{\Lambda_{\rho,l}} \mathbf{R}(z)\mathbf{Z}\mathbf{R}(z)dz =: \eta_l(\mathbf{Z}).$$

Due to the residue theorem, $\mathbf{R}(z) = \sum_l(\lambda_l - z)^{-1}\phi_l \otimes \phi_l$, and

$$\mathbf{R}(z)\mathbf{Z}\mathbf{R}(z) = \sum_{l_1,l_2}[(\lambda_{l_1} - z)(\lambda_{l_2} - z)]^{-1}(\phi_{l_1} \otimes \phi_{l_1})\,\mathbf{Z}\,(\phi_{l_2} \otimes \phi_{l_2}).$$

Therefore, using the notation $\mathbf{S}_l = \sum_{k:k\neq l}(\lambda_l - \lambda_k)^{-1}\phi_k \otimes \phi_k$, $\eta_l(\mathbf{Z})$ can be expressed explicitly as

$$\begin{aligned}
\eta_l(\mathbf{Z}) &= Res_{z=\lambda_l}\left(\sum_{l_1,l_2}[(\lambda_{l_1}-z)(\lambda_{l_2}-z)]^{-1}(\phi_{l_1}\otimes\phi_{l_1})\mathbf{Z}(\phi_{l_2}\otimes\phi_{l_2})\right) \\
&= -\sum_{l_2:l_2\neq l}(\lambda_{l_2}-\lambda_l)^{-1}(\phi_l\otimes\phi_l)\mathbf{Z}(\phi_{l_2}\otimes\phi_{l_2}) \\
&\quad -\sum_{l_1:l_1\neq l}(\lambda_{l_1}-\lambda_l)^{-1}(\phi_{l_1}\otimes\phi_{l_1})\mathbf{Z}(\phi_l\otimes\phi_l) \\
&= -\sum_{k:k\neq l}(\lambda_k-\lambda_l)^{-1}\mathbf{P}_l\mathbf{Z}(\phi_k\otimes\phi_k)-\sum_{k:k\neq l}(\lambda_k-\lambda_l)^{-1}(\phi_k\otimes\phi_k)\mathbf{Z}\mathbf{P}_l \\
&= \mathbf{P}_l\mathbf{Z}\mathbf{S}_l+\mathbf{S}_l\mathbf{Z}\mathbf{P}_l
\end{aligned}$$

which is Gaussian and with mean zero due to (4.4). □

Proof. (of Theorem 4.1)

Firstly, we show (4.8). Let

$$\mathbf{M}_l^n(\mathbf{Z_n}) = n^{1/2}(\hat{\mathbf{P}}_l\hat{\mathbf{C}}\hat{\mathbf{P}}_l-\lambda_l\hat{\mathbf{P}}_l).$$

Then, for fixed l, we have

$$\begin{aligned}
\mathbf{M}_l^n(\mathbf{Z_n})\hat{\phi}_l &= n^{1/2}(\hat{\mathbf{P}}_l\hat{\mathbf{C}}\hat{\mathbf{P}}_l-\lambda_l\hat{\mathbf{P}}_l)\hat{\phi}_l \\
&= n^{1/2}(\hat{\phi}_l\otimes\hat{\phi}_l\hat{\mathbf{C}}\hat{\phi}_l\otimes\hat{\phi}_l-\lambda_l\hat{\phi}_l\otimes\hat{\phi}_l)\hat{\phi}_l \\
&= n^{1/2}(\hat{\phi}_l\otimes\hat{\phi}_l\hat{\mathbf{C}}(\hat{\phi}_l)-\lambda_l\hat{\phi}_l) \\
&= n^{1/2}(\hat{\phi}_l\otimes\hat{\phi}_l\hat{\lambda}_l\hat{\phi}_l-\lambda_l\hat{\phi}_l) \\
&= n^{1/2}(\hat{\lambda}_l-\lambda_l)\hat{\phi}_l.
\end{aligned}$$

This means that $n^{1/2}(\hat{\lambda}_l-\lambda_l)$ is an eigenvalue of the operator $\mathbf{M}_l^n(\mathbf{Z_n})$. Now, the Rubin-Billingsley theorem implies

$$\mathbf{M}_l^n(\mathbf{Z_n}) \Rightarrow \mathbf{P}_l\mathbf{Z}\mathbf{P}_l =: \mathbf{M}_l(\mathbf{Z}).$$

Therefore, the asymptotic distribution of $n^{1/2}(\hat{\lambda}_l-\lambda_l)$ is given by the distribution of the eigenvalues of $\mathbf{P}_l\mathbf{Z}\mathbf{P}_l$. Note that $\mathbf{P}_l\mathbf{Z}\mathbf{P}_l$ is a Gaussian random element in $\mathcal{S}$

with mean zero. Plugging in the right hand side of (4.4), we have

$$\begin{aligned}\mathbf{P}_l\mathbf{Z}\mathbf{P}_l(\phi_l) &= (\phi_l \otimes \phi_l)\sum_{l_1<l_2}\sqrt{\lambda_{l_1}\lambda_{l_2}}(\phi_{l_1}\otimes\phi_{l_2}+\phi_{l_2}\otimes\phi_{l_1})(\phi_l\otimes\phi_l)(\phi_l)\zeta_{l_1l_2}\\ &+ (\phi_l\otimes\phi_l)\sqrt{2}\sum_l \lambda_l(\phi_l\otimes\phi_l)(\phi_l\otimes\phi_l)(\phi_l)\zeta_{ll}\\ &= 0+\sqrt{2}\lambda_l\phi_l\zeta_{ll}\end{aligned}$$

with ζ_{ll} be standard normal random variable. Therefore,

$$n^{1/2}(\hat{\lambda}_l - \lambda_l) \to_d \sqrt{2}\lambda_l\zeta_{ll} = N(0, 2\lambda_l^2).$$

Next we consider (4.9). Note that, $n^{1/2}(\hat{\phi}_l - \phi_l)$ can be written as

$$\begin{aligned}n^{1/2}(\hat{\phi}_l - \phi_l) &= \mathbf{P}_l(n^{1/2}(\hat{\phi}_l - \phi_l)) + (\mathbf{I}-\mathbf{P}_l)(n^{1/2}(\hat{\phi}_l - \phi_l))\\ &=: I_l + II_l.\end{aligned}$$

First, it is easy to show that I_l converges weakly to 0. In fact, from the definition, we have

$$\begin{aligned}I_l &= \langle n^{1/2}(\hat{\phi}_l - \phi_l), \phi_l\rangle\phi_l\\ &= n^{1/2}(\langle\hat{\phi}_l,\phi_l\rangle - 1)\phi_l\\ &= \langle n^{1/2}(\hat{\mathbf{P}}_l - \mathbf{P}_l), \mathbf{P}_l\rangle_{\mathcal{S}}\cdot(\langle\hat{\phi}_l,\phi_l\rangle+1)^{-1}\cdot\phi_l,\end{aligned}$$

and since $\langle\hat{\phi}_l,\phi_l\rangle$ tends to 1, applying Lemma 4.4 we have

$$\langle n^{1/2}(\hat{\mathbf{P}}_l - \mathbf{P}_l), \mathbf{P}_l\rangle_{\mathcal{S}} \Rightarrow \langle\eta_l(\mathbf{Z}),\mathbf{P}_l\rangle_{\mathcal{S}} = tr(\eta_l(\mathbf{Z})\mathbf{P}_l) = tr(\mathbf{P}_l\mathbf{S}_l\mathbf{Z}) = 0.$$

Therefore, $I_l \Rightarrow 0$. Then we consider the second term II_l and show that II_l converges weakly to $\mathbf{S}_l\mathbf{Z}(\phi_l)$. Since

$$\begin{aligned}(\mathbf{I}-\mathbf{P}_l)(n^{1/2}(\hat{\phi}_l - \phi_l)) &= n^{1/2}(\mathbf{I}-\mathbf{P}_l)\hat{\phi}_l\\ &= n^{1/2}(\mathbf{I}-\mathbf{P}_l)\hat{\mathbf{P}}_l\phi_l\cdot(\langle\hat{\phi}_l,\phi_l\rangle)^{-1}\\ &= (\mathbf{I}-\mathbf{P}_l)[n^{1/2}(\hat{\mathbf{P}}_l-\mathbf{P}_l)](\phi_l)\cdot(\langle\hat{\phi}_l,\phi_l\rangle)^{-1}\\ &\Rightarrow (\mathbf{I}-\mathbf{P}_l)\eta_l(\mathbf{Z})(\phi_l)\\ &= \mathbf{S}_l\mathbf{Z}(\phi_l),\end{aligned}$$

we have

$$
\begin{aligned}
n^{1/2}(\hat{\phi}_l - \phi_l) &\Rightarrow \mathbf{S}_l\mathbf{Z}(\phi_l) \\
&= \sum_{k:k\neq l}(\lambda_l - \lambda_k)^{-1}\phi_k \otimes \phi_k \sum_{l_1<l_2}\sqrt{\lambda_{l_1}\lambda_{l_2}}(\phi_{l_1}\otimes\phi_{l_2} + \phi_{l_2}\otimes\phi_{l_1})(\phi_l)\zeta_{l_1 l_2} \\
&\quad + \sum_{k:k\neq l}(\lambda_l - \lambda_k)^{-1}\phi_k\otimes\phi_k\sqrt{2}\sum_{l}\lambda_l(\phi_l\otimes\phi_l)(\phi_l)\zeta_{ll} \\
&= \sum_{k:k>l}\sqrt{\lambda_l\lambda_k}(\lambda_l-\lambda_k)^{-1}\phi_k\zeta_{lk} + \sum_{k:k<l}\sqrt{\lambda_l\lambda_k}(\lambda_l-\lambda_k)^{-1}\phi_k\zeta_{kl} + 0
\end{aligned}
$$

with ζ_{ij} and ζ_{ii} $(i, j \in \{1, ..., p\})$ are independent standard normal random variables as given as in (4.9) and (4.8) respectively. □

Proof. (of Theorem 4.2)

Following the proof before, for $l_1 \neq l_2$, the covariance operator of $\mathbf{S}_l\mathbf{Z}(\phi_l)$ is

$$
\begin{aligned}
&E[\mathbf{S}_{l_1}\mathbf{Z}(\phi_{l_1})\otimes\mathbf{S}_{l_2}\mathbf{Z}(\phi_{l_2})] \\
&= E\Bigg[\left(\sum_{k_1:k_1>l_1}\sqrt{\lambda_{l_1}\lambda_{k_1}}(\lambda_{l_1}-\lambda_{k_1})^{-1}\phi_{k_1}\zeta_{l_1k_1} + \sum_{k_1:k_1<l_1}\sqrt{\lambda_{l_1}\lambda_{k_1}}(\lambda_{l_1}-\lambda_{k_1})^{-1}\phi_{k_1}\zeta_{k_1l_1}\right) \\
&\quad \otimes\left(\sum_{k_2:k_2>l_2}\sqrt{\lambda_{l_2}\lambda_{k_2}}(\lambda_{l_2}-\lambda_{k_2})^{-1}\phi_{k_2}\zeta_{l_2k_2} + \sum_{k_2:k_2<l_2}\sqrt{\lambda_{l_2}\lambda_{k_2}}(\lambda_{l_2}-\lambda_{k_2})^{-1}\phi_{k_2}\zeta_{k_2l_2}\right)\Bigg] \\
&= \sum_{k_1:k_1>l_1}\sum_{k_2:k_2>l_2}\sqrt{\lambda_{l_1}\lambda_{k_1}}\sqrt{\lambda_{l_2}\lambda_{k_2}}(\lambda_{l_1}-\lambda_{k_1})^{-1}(\lambda_{l_2}-\lambda_{k_2})^{-1}\phi_{k_1}\otimes\phi_{k_2}E[\zeta_{l_1k_1}\zeta_{l_2k_2}] \\
&\quad + \sum_{k_1:k_1>l_1}\sum_{k_2:k_2<l_2}\sqrt{\lambda_{l_1}\lambda_{k_1}}\sqrt{\lambda_{l_2}\lambda_{k_2}}(\lambda_{l_1}-\lambda_{k_1})^{-1}(\lambda_{l_2}-\lambda_{k_2})^{-1}\phi_{k_1}\otimes\phi_{k_2}E[\zeta_{l_1k_1}\zeta_{k_2l_2}] \\
&\quad + \sum_{k_1:k_1<l_1}\sum_{k_2:k_2>l_2}\sqrt{\lambda_{l_1}\lambda_{k_1}}\sqrt{\lambda_{l_2}\lambda_{k_2}}(\lambda_{l_1}-\lambda_{k_1})^{-1}(\lambda_{l_2}-\lambda_{k_2})^{-1}\phi_{k_1}\otimes\phi_{k_2}E[\zeta_{k_1l_1}\zeta_{l_2k_2}] \\
&\quad + \sum_{k_1:k_1<l_1}\sum_{k_2:k_2<l_2}\sqrt{\lambda_{l_1}\lambda_{k_1}}\sqrt{\lambda_{l_2}\lambda_{k_2}}(\lambda_{l_1}-\lambda_{k_1})^{-1}(\lambda_{l_2}-\lambda_{k_2})^{-1}\phi_{k_1}\otimes\phi_{k_2}E[\zeta_{k_1l_1}\zeta_{k_2l_2}] \\
&= -\lambda_{l_1}\lambda_{l_2}(\lambda_{l_1}-\lambda_{l_2})^{-2}\phi_{l_1}\otimes\phi_{l_2}.
\end{aligned}
$$

Therefore, for $l_1 \neq l_2 \in \{1, ..., p\}$,

$$cov(\phi_{l_1}(s), \phi_{l_2}(t)) = -\lambda_{l_1}\lambda_{l_2}(\lambda_{l_1}-\lambda_{l_2})^{-2}\phi_{l_1}(s)\phi_{l_2}(t).$$

For $l \in \{1, ..., p\}$,

$$cov(\phi_l(s), \phi_l(t)) = \sum_{k:k\neq l}\lambda_l\lambda_k(\lambda_l-\lambda_k)^{-2}\phi_k(s)\phi_k(t)$$

follows directly from 4.1. □

Proof. (of Theorem 4.3)

From the definition of $\hat{\xi}_{il}$ in (4.3), we have

$$\begin{aligned}
\hat{\xi}_{il} &= N^{-1}\langle Y_{i\cdot}, \hat{\phi}_l\cdot\rangle \\
&= N^{-1}\langle Y_{i\cdot}, \phi_l \cdot +(\hat{\phi}_l \cdot - \phi_l\cdot)\rangle \\
&= N^{-1}\langle Y_{i\cdot}, \phi_l\cdot\rangle + N^{-1}\langle Y_{i\cdot}, (\hat{\phi}_l \cdot - \phi_l\cdot)\rangle \\
&= N^{-1}\langle Y_{i\cdot}, \phi_l\cdot\rangle + O\left(n^{-1/2}N^{-1}\right) \\
&=: A_{i,l,N} + B_{i,l,N} + O\left(n^{-1/2}N^{-1}\right)
\end{aligned}$$

with two independent terms

$$\begin{aligned}
A_{i,l,N} &= N^{-1}\sum_{j=1}^{N}\left(\sum_{k=1}^{\infty}\xi_{ik}\phi_k(t_j)\phi_l(t_j)\right) = \sum_{k=1}^{\infty}\xi_{ik}N^{-1}\sum_{j=1}^{N}\phi_k(t_j)\phi_l(t_j), \\
B_{i,l,N} &= N^{-1}\sum_{j=1}^{N}\epsilon_i(j)\phi_l(t_j).
\end{aligned}$$

Note that, given the functional principal component scores ξ_{ik} $(k \in \mathbb{N})$, the first term $A_{i,l,N}$ is deterministic thus $var\,(A_{i,l,N}|\xi_{ik}, k \in \mathbb{N}) = 0$. Moreover,

$$A_{i,l,N} = \sum_{k=1}^{\infty}\xi_{ik}\left\langle \phi_k, \phi_l\right\rangle + O\left(N^{-1}\right) = \xi_{il} + O\left(N^{-1}\right).$$

For the second term $B_{i,l,N}$. Since it does not depend on ξ_{ik} $(k \in \mathbb{N})$, the conditional expectation $E[B_{i,l,N}|\xi_{ik}, k \in \mathbb{N}] = E[B_{i,l,N}]$ equals to 0 and the conditional covariance is

$$\begin{aligned}
&cov\left(B_{i,l_1,N}, B_{i,l_2,N}|\xi_{ik}, k \in \mathbb{N}\right) \\
&= cov\left(B_{i,l_1,N}, B_{i,l_2,N}\right) \\
&= N^{-2}\sum_{j_1,j_2=1}^{N}\gamma_\epsilon\left(j_1 - j_2\right)\phi_{l_1}(t_{j_1})\phi_{l_2}(t_{j_2}).
\end{aligned}$$

Let $\delta > 0$ be an arbitrarily small constant. Since the error processes ϵ_i are short memory, i.e. $\sum|\gamma_\epsilon(k)| < \infty$, there exists a positive integer $k_0 = k_0(\delta)$ such that

$$\sum_{|k|>k_0}|\gamma_\epsilon\left(k\right)| < \delta.$$

Moreover, since $p < \infty$ and the eigenfunctions $\phi_l(t)$ ($l \in \mathbb{N}$) are uniformly continuous, we have

$$\max_{1 \le l \le p} \sup_{t \in [-\pi, \pi]} |\phi_l(t)| \le C_\phi < \infty$$

for a suitable constant C_ϕ. We may separate the conditional covariance into two terms

$$cov\left(B_{i,l_1,N}, B_{i,l_2,N} | \xi_{ik}, k \in \mathbb{N}\right) = S_{1,N}(\delta) + S_{2,N}(\delta)$$

with

$$S_{1,N}(\delta) = N^{-2} \sum_{\substack{j_1, j_2 = 1 \\ |j_1 - j_2| \le k_0}}^{N} \gamma_\epsilon\left(j_1 - j_2\right) \phi_{l_1}(t_{j_1}) \phi_{l_2}(t_{j_2}),$$

$$S_{2,N}(\delta) = N^{-2} \sum_{\substack{j_1, j_2 = 1 \\ |j_1 - j_2| > k_0}}^{N} \gamma_\epsilon\left(j_1 - j_2\right) \phi_{l_1}(t_{j_1}) \phi_{l_2}(t_{j_2}).$$

For the term $S_{1,N}(\delta)$. Since $|j_1 - j_2| \le k_0$, $S_{1,N}(\delta)$ can be written as

$$S_{1,N}(\delta) = N^{-2} \sum_{j_1=1}^{N} \sum_{j_2 = \max(1, j_1 - k_0)}^{\min(N, j_1 + k_0)} \gamma_\epsilon\left(j_1 - j_2\right) \phi_{l_1}(t_{j_1}) \phi_{l_2}(t_{j_2}).$$

Note that

$$\max_{\substack{1 \le j_1, j_2 \le N \\ |j_1 - j_2| \le k_0}} |t_{j_1} - t_{j_2}| = \max_{\substack{1 \le j_1, j_2 \le N \\ |j_1 - j_2| \le k_0}} \left| \frac{j_1 - j_2}{N} \right| \le \frac{k_0}{N} \underset{N \to \infty}{\to} 0,$$

uniform continuity of $\phi_l(t)$ implies

$$\max_{\substack{1 \le j_1, j_2 \le N \\ |j_1 - j_2| \le k_0}} |\phi_{l_2}(t_{j_2}) - \phi_{l_2}(t_{j_1})| < \delta \quad (n \ge n_0)$$

for $n_0 = n_0(\delta)$ large enough. Therefore,

$$\begin{aligned} S_{1,N}(\delta) &= N^{-2} \sum_{j_1=1}^{N} \phi_{l_1}(t_{j_1}) \sum_{j_2 = \max(1, j_1 - k_0)}^{\min(N, j_1 + k_0)} \gamma_\epsilon(j_1 - j_2) \left[\phi_{l_2}(t_{j_1}) + r_{l_2, j_2, N}\right] \\ &=: S_{1,N,\phi}(\delta) + S_{1,N,r}(\delta) \end{aligned}$$

with $|r_{l_2,j_2,N}|$ bounded uniformly by δ. Since, for the first term $S_{1,N,\phi}(\delta)$,

$$\begin{aligned} &\lim_{N \to \infty} N^{-1} \sum_{j_1=1}^{N} \phi_{l_1}(t_{j_1}) \phi_{l_2}(t_{j_1}) \sum_{j_2 = \max(1, j_1 - k_0)}^{\min(N, j_1 + k_0)} \gamma_\epsilon\left(j_1 - j_2\right) \\ &= \langle \phi_{l_1}, \phi_{l_2} \rangle \sum_{k = -k_0}^{k_0} \gamma_\epsilon(k), \end{aligned}$$

therefore, we have

$$S_{1,N,\phi}(\delta) = N^{-2} \sum_{j_1=1}^{N} \phi_{l_1}(t_{j_1}) \sum_{j_2=\max(1,j_1-k_0)}^{\min(N,j_1+k_0)} \gamma_\epsilon (j_1 - j_2) \phi_{l_2}(t_{j_1})$$
$$= N^{-1} \langle \phi_{l_1}, \phi_{l_2} \rangle \sum_{k=-k_0}^{k_0} \gamma_\epsilon(k) + o(N^{-1}).$$

For the second term $S_{1,N,r}(\delta)$, we have the upper bound

$$|S_{1,N,r}(\delta)| = \left| N^{-2} \sum_{j_1=1}^{N} \phi_{l_1}(t_{j_1}) \sum_{j_2=\max(1,j_1-k_0)}^{\min(N,j_1+k_0)} \gamma_\epsilon (j_1 - j_2) r_{l_2,j_2,N} \right|$$
$$\leq \delta N^{-2} \sum_{j_1=1}^{N} |\phi_{l_1}(t_{j_1})| \sum_{k=-\infty}^{\infty} |\gamma_\epsilon(k)|$$
$$\leq \delta N^{-1} C$$

with a constant

$$C = C_\phi \sum_{k=-\infty}^{\infty} |\gamma_\epsilon(k)| < \infty.$$

Therefore, we obtain

$$\lim_{N\to\infty} N \cdot S_{1,N}(\delta) = \langle \phi_{l_1}, \phi_{l_2} \rangle \sum_{k=-k_0}^{k_0} \gamma_\epsilon(k) + o(\delta).$$

Letting $k_0 \to \infty$, we have

$$\lim_{N\to\infty} N \cdot S_{1,N}(\delta) = \langle \phi_{l_1}, \phi_{l_2} \rangle \sum_{k=-\infty}^{\infty} \gamma_\epsilon(k) + o(\delta)$$
$$= 2\pi f_\epsilon(0) \langle \phi_{l_1}, \phi_{l_2} \rangle + o(\delta)$$
$$= 2\pi f_\epsilon(0) \delta_{l_1 l_2} + o(\delta).$$

Similarly, for the term $S_{2,N}(\delta)$, we have the upper bound

$$|S_{2,N}(\delta)| \leq CN^{-1} \sum_{|k|>k_0} |\gamma_\epsilon(k)| \leq C\delta N^{-1},$$

with a constant $C = C_\phi^2$, and hence

$$\lim_{N\to\infty} N \cdot S_{2,N}(\delta) = o(\delta).$$

In conclusion, we have

$$\lim_{N\to\infty} cov\left(B_{i,l_1,N}, B_{i,l_2,N}\right) = 2\pi f_\epsilon(0) \delta_{l_1 l_2} = V_{l_1 l_2}.$$

Moreover, for each i, since $\epsilon_i(j)$ $(j \in \mathbb{Z})$ is a Gaussian process so

$$\zeta_{i,n} := N^{\frac{1}{2}} \left(\hat{\xi}_{i,1} - \xi_{i,1}, ..., \hat{\xi}_{i,p} - \xi_{i,p} \right)^T \underset{d}{\rightarrow} \zeta \sim N(0, V)$$

with $V = [V_{l_1 l_2}]_{l_1, l_2 = 1,...,p}$. □

Proof. (of Theorem 4.4)

As discussed in the proof of Theorem 4.3 we have

$$\hat{\xi}_{il} = N^{-1} \langle Y_{i\cdot}, \hat{\phi}_{l\cdot} \rangle = A_{i,l,N} + B_{i,l,N} + O\left(n^{-1/2} N^{-1}\right)$$

where $A_{i,l,N} = \xi_{il} + O(N^{-1})$ is deterministic, and $B_{i,l,N}$ does not depending on the functional principal component scores ξ_{ik} $(k \in \mathbb{N})$.

For the covariance of $B_{i,l,N}$, since the error processes $\epsilon_i(j)$ $(j \in \mathbb{Z})$ are long-range dependent, we have

$$\begin{aligned}
& cov\left(B_{i,l_1,N}, B_{i,l_2,N}\right) \\
&= N^{-2} \sum_{j_1, j_2 = 1}^{N} \gamma_\epsilon\left(j_1 - j_2\right) \phi_{l_1}(t_{j_1}) \phi_{l_2}(t_{j_2}) \\
&= c_\gamma N^{2d-1} \sum_{\substack{j_1, j_2 = 1 \\ j_1 \neq j_2}}^{N} \left| \frac{j_1}{N} - \frac{j_2}{N} \right|^{2d-1} \phi_{l_1}\left(\frac{j_1}{N}\right) \phi_{l_2}\left(\frac{j_2}{N}\right) \frac{1}{N^2} + o\left(N^{2d-1}\right) \\
&= N^{2d-1} c_\gamma \int_{-1}^{1} \int_{-1}^{1} |u - v|^{2d-1} \phi_{l_1}(u) \phi_{l_2}(v) du dv + o\left(N^{2d-1}\right) \\
&= N^{2d-1} V_{l_1 l_2} + o\left(N^{2d-1}\right).
\end{aligned}$$

Since $\epsilon_i(j)$ $(j \in \mathbb{Z})$ is a Gaussian process and $N^{1/2-d} \cdot N^{-1} = N^{-d-1/2} = o(1)$, we obtain

$$\zeta_{i,n} \underset{d}{\rightarrow} \zeta_i \sim N(0, V)$$

with $V = [V_{l_1 l_2}]_{l_1, l_2 = 1,...,p}$ defined above in Theorem 4.3. □

4.5.2 Tables

N	$B^2_{\lambda_1,N}$			$\sigma^2_{\lambda_1,N}$			$MSE_{\lambda_1,N}$		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
100	0.047	0.279	0.462	32.881	35.056	32.950	32.928	35.335	33.412
200	0.093	0.237	0.633	30.260	30.661	31.442	30.353	30.898	32.076
400	0.076	0.112	0.551	30.285	30.698	31.317	30.361	30.809	31.868
600	0.098	0.174	0.743	27.863	28.181	29.633	27.961	28.354	30.376
800	0.024	0.053	0.337	30.668	31.004	31.039	30.692	31.057	31.376
1000	0.008	0.020	0.311	30.842	30.753	30.032	30.850	30.774	30.343
1500	0.254	0.294	1.005	31.949	31.805	32.511	32.203	32.100	33.516
2000	0.047	0.057	0.433	32.990	33.128	32.623	33.037	33.186	33.056

Table 4.1: Simulated values of the standardized squared bias $B^2_{\lambda_1,N} = n_N[E(\hat{\lambda}_1) - \lambda_1]^2$, variance $\sigma^2_{\lambda_1,N} = n_N var(\hat{\lambda}_1)$ and mean squared error $MSE_{\lambda_1,N} = n_N E[(\hat{\lambda}_1 - \lambda_1)^2]$ of $\hat{\lambda}_1$. For each N, the results are based on 200 simulations of model (4.11) with error process $\epsilon_i(j)$ generated by (a) iid $N(0,1)$ variables; (b) an AR(1) process with lag-one correlation $\rho = 0.5$; (c) a $FARIMA(0,d,0)$ process with $d = 0.3$.

N	$B^2_{\lambda_2,N}$			$\sigma^2_{\lambda_2,N}$			$MSE_{\lambda_2,N}$		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
100	0.177	0.037	0.008	7.467	7.527	7.959	7.643	7.564	7.967
200	0.177	0.068	0.001	8.302	7.981	8.206	8.478	8.049	8.207
400	0.063	0.015	0.035	8.150	8.250	8.441	8.213	8.264	8.475
600	0.006	0.000	0.081	8.842	8.768	8.894	8.848	8.768	8.975
800	0.021	0.006	0.059	8.001	7.999	8.110	8.022	8.004	8.169
1000	0.005	0.001	0.070	7.808	7.786	7.835	7.814	7.787	7.905
1500	0.023	0.015	0.019	8.332	8.265	8.428	8.355	8.280	8.447
2000	0.033	0.028	0.032	8.227	8.253	8.258	8.260	8.281	8.289

Table 4.2: Simulated values of the standardized squared bias $B^2_{\lambda_2,N} = n_N[E(\hat{\lambda}_2) - \lambda_2]^2$, variance $\sigma^2_{\lambda_2,N} = n_N var(\hat{\lambda}_2)$ and mean squared error $MSE_{\lambda_2,N} = n_N E[(\hat{\lambda}_2 - \lambda_2)^2]$ of $\hat{\lambda}_2$. For each N, the results are based on 200 simulations of model (4.11) with error process $\epsilon_i(j)$ generated by (a) iid $N(0,1)$ variables; (b) an AR(1) process with lag-one correlation $\rho = 0.5$; (c) a $FARIMA(0,d,0)$ process with $d = 0.3$.

N	$B^2_{\phi_1,N}$			$\sigma^2_{\phi_1,N}$			$MSE_{\phi_1,N}$		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
100	0.126	0.116	0.126	2.208	2.243	2.194	2.334	2.359	2.320
200	0.077	0.070	0.078	2.219	2.283	2.269	2.296	2.353	2.347
400	0.077	0.074	0.090	2.111	2.128	2.168	2.188	2.202	2.258
600	0.070	0.071	0.067	1.881	1.889	1.889	1.951	1.959	1.957
800	0.065	0.068	0.057	1.891	1.926	1.950	1.956	1.994	2.007
1000	0.080	0.083	0.070	1.852	1.879	1.845	1.932	1.963	1.915
1500	0.066	0.063	0.065	1.689	1.677	1.713	1.755	1.740	1.778
2000	0.029	0.030	0.028	1.797	1.823	1.873	1.826	1.852	1.901

Table 4.3: Simulated values of the standardized integrated squared bias $B^2_{\phi_1,N}$, variance $\sigma^2_{\phi_1,N}$ and mean squared error $IMSE_{\phi_1,N}$ of $\hat{\phi}_1$. For each N, the results are based on 200 simulations of model (4.11) with error process $\epsilon_i(j)$ generated by (a) iid $N(0,1)$ variables; (b) an AR(1) process with lag-one correlation $\rho = 0.5$; (c) a $FARIMA(0,d,0)$ process with $d = 0.3$.

N	$B^2_{\phi_2,N}$			$\sigma^2_{\phi_2,N}$			$MSE_{\phi_2,N}$		
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
100	0.210	0.202	0.210	2.318	2.416	2.378	2.528	2.618	2.588
200	0.126	0.117	0.133	2.278	2.404	2.387	2.404	2.521	2.520
400	0.126	0.122	0.145	2.144	2.207	2.262	2.270	2.329	2.408
600	0.124	0.125	0.123	1.905	1.950	1.970	2.029	2.074	2.093
800	0.123	0.128	0.110	1.909	1.975	2.018	2.032	2.103	2.129
1000	0.150	0.155	0.135	1.867	1.920	1.906	2.017	2.075	2.041
1500	0.140	0.135	0.138	1.700	1.707	1.773	1.840	1.842	1.911
2000	0.089	0.091	0.087	1.805	1.847	1.925	1.895	1.938	2.013

Table 4.4: Simulated values of the standardized integrated squared bias $B^2_{\phi_2,N}$, variance $\sigma^2_{\phi_2,N}$ and mean squared error $IMSE_{\phi_2,N}$ of $\hat{\phi}_2$. For each N, the results are based on 200 simulations of model (4.11) with error process $\epsilon_i(j)$ generated by (a) iid $N(0,1)$ variables; (b) an AR(1) process with lag-one correlation $\rho = 0.5$; (c) a $FARIMA(0,d,0)$ process with $d = 0.3$.

Chapter 5

Two sample inference for eigenspaces

In this chapter, we discuss two sample inference for eigenspaces in functional data analysis (FDA) with dependent errors. Specifically, we study the similarities of two independent functional data samples by proposing a test statistic for testing the equality of the two eigenspaces $\mathcal{U}$ and $\mathcal{V}$ without assuming that individual eigenfunctions or eigenvalues are identical. Subspaces $\mathcal{U}$ and $\mathcal{V}$ are generated respectively by the first m eigenfunctions of the two functional data samples with a noise component that may exhibit short- or long-range dependence. The test statistic is constructed by considering the residue processes $r_l(t) = \phi_l^{(2)}(t) - \sum_{i=1}^{m} a_{ij}\phi_i^{(1)}(t)$ $(l = 1, ..., m)$, where $\phi^{(1)}(t)$ and $\phi^{(2)}(t)$ are the eigenfunctions of the two samples respectively. Then, the asymptotic distribution of the standardized residuals under the null hypothesis $\mathcal{U} = \mathcal{V}$ is derived. This provides the basis for defining the suitable test procedures where we consider the squared integral of the standardized residuals. However, in order to obtain less conservative rejection regions, the joint asymptotic distribution of the standardized residuals will be required. Since, in most cases, the dimension of the subspace m is very small, we prefer to use a simple Bonferroni adjusted test. In order to avoid the shortcomings of the above test, a more practical solution - bootstrap test - is also discussed. Simulations illustrated the performance of the test are carried out. In particular, the accuracy of the level of significance does not seem to be influenced by the dependence structure and, under the alternative hypothesis, the test is very powerful. This chapter is based on our results in Beran, Liu and Telkmann (2016).

5.1 Definitions

Throughout this chapter, observations are still assumed to be of the form (3.1), (3.3), (3.4) or (3.5) as that in Chapter 3. Moreover, the assumptions in Chapter 3 and Chapter 4 for one sample analysis are assumed to be hold.

5.1.1 Auxiliary results

Recall that, after the orthonormal contrast transformation of the original observations, the two-dimensional boundary kernel estimator of covariance of the underlying random curve $X(t)$ ($t \in [0,1]$) can be given as in (3.21):

$$\hat{C}(s,t) = \frac{1}{(Nb)^2} \sum_{j,k=1}^{N} K_{2,b}\left(\frac{s-t_j}{b}, \frac{t-t_k}{b}\right) n^{-1} \sum_{i=1}^{n} Y_{ij} Y_{ik},$$

where $K_{2,b}(s,t)$ is a two-dimensional boundary higher order kernel function with support $[-1,1]^2$ (see Chapter 3 or Beran and Liu 20146).

The eigenvalues λ_l and eigenfunctions (functional principal components) $\phi_l(t)$ of the covariance function $C(s,t) = cov(X(s), X(t))$ are estimated from (4.1):

$$\int_0^1 \hat{C}(s,t)\hat{\phi}_l(s)ds = \hat{\lambda}_l \hat{\phi}_l(t),$$

where $\int_0^1 \hat{\phi}_l^2(t)dt = 1$ and $\int_0^1 \hat{\phi}_l(t)\hat{\phi}_m(t)dt = 0$ for $m < l$. Note that the above formula follows from the Mercer's Theorem on $\hat{C}(s,t)$ as in (4.2):

$$\hat{C}(s,t) = \sum_l \hat{\lambda}_l \hat{\phi}_l(s)\hat{\phi}_l(t) \quad (s,t \in [0,1]).$$

Note that, since the error processes $\epsilon_i(j)$ are existing and not independent,

$$\begin{aligned} cov(Y_{ij}, Y_{ik}) &= cov\left(X(t_j), X(t_k)\right) + cov(\epsilon_i(j), \epsilon_i(k)) \\ &= C(t_j, t_k) + \gamma_\epsilon(j-k) \end{aligned}$$

means that $\hat{\phi}_l(t)$, $\hat{\lambda}_l$ based on Y_{ij} differ from the corresponding estimates obtained from $X_i(t_j)$ (see Chapter 4 or Beran and Liu 2016).

5.1.2 Two independent samples

One of the main objectives in FDA is to obtain a low-dimensional representation of $X(t)$ in terms of eigenfunctions $\phi_l(t)$ $(l = 1, ..., m)$ with the largest m eigenvalues. Therefore, in this chapter, we consider the following two sample problem. Suppose that we observe two independent samples defined by

$$Y_{ij}^{(1)} = X_i^{(1)}(t_j) + \epsilon_i^{(1)}(j) \quad (i = 1, ..., n^{(1)}; j = 1, ..., N^{(1)}) \tag{5.1}$$

and

$$Y_{ij}^{(2)} = X_i^{(2)}(t_j) + \epsilon_i^{(2)}(j) \quad (i = 1, ..., n^{(2)}; j = 1, ..., N^{(2)}) \tag{5.2}$$

where $X_i^{(k)}, \epsilon_i^{(k)}$ $(k = 1, 2)$ are defined as that in (3.3), (3.4) or (3.5) in Chapter 3. For $k = 1, 2$, the error processes $\epsilon_i^{(k)}$ $(i \in \mathbb{N})$ are independent copies of a Gaussian process $\epsilon^{(k)}$ with zero mean and spectral density $f^{(k)}$ satisfying (3.4). This means that for $d = 0$, $\epsilon_i^{(k)}$ $(i \in \mathbb{N})$ exhibit short memory and for $0 < d < \frac{1}{2}$, $\epsilon_i^{(k)}$ $(i \in \mathbb{N})$ exhibit long memory. Note that the spectral densities $f^{(1)}$, $f^{(2)}$ of the processes $\epsilon^{(1)}$ and $\epsilon^{(2)}$ may be different. The random functions $X_i^{(k)}(t)$ are assumed to have the Karhunen-Loève (K.L.) expansions

$$X_i^{(k)}(t) = \mu^{(k)}(t) + \sum_{l=1}^{p^{(k)}} \xi_{il}^{(k)} \phi_l^{(k)}(t) \quad (k = 1, 2) \tag{5.3}$$

with $p^{(1)}, p^{(2)} \leq \infty$.

Given a finite fixed dimension $m \in \mathbb{N}$, we are interested in testing whether the subspaces (of $L^2[0, 1]$) spanned by $\phi_l^{(1)}(t)$ $(l = 1, ..., m)$ and $\phi_l^{(2)}(t)$ $(l = 1, ..., m)$ are the same. In next section we will give the definition of one test based on residual functions and derive its asymptotic distribution.

5.2 Test statistics and asymptotic properties

The comparison of the eigenspaces generated by the first several eigenfunctions (functional principal components) which leads to the two sample problem as above is discussed in this section.

5.2.1 Definition of test statistic $\tilde{r}_l(t)$

Let $m \leq \min\{p^{(1)}, p^{(2)}\}$ be a positive integer (in practice, m is small). Denote by $\mathcal{U}$ and $\mathcal{V}$ be the m-dimensional linear function spaces generated by the eigenfunctions $\phi_1^{(1)}(t)$,...,$\phi_m^{(1)}(t)$ and $\phi_1^{(2)}(t)$,...,$\phi_m^{(2)}(t)$ respectively. Our aim is to test the null hypothesis

$$H_0 : \mathcal{U} = \mathcal{V}$$

against the alternative hypothesis

$$H_A : \mathcal{U} \neq \mathcal{V}.$$

A natural approach is to consider, for each $l \in \{1, ..., m\}$, the difference between $\phi_l^{(2)}(t)$ and its orthogonal projection to $\mathcal{U}$. Under H_0, this difference is identically equal to zero. However, under H_A, there is at least one l where this difference is not identically equal to zero. Therefore, we consider the residual functions $r_l(t)$ $(l = 1, ..., m)$ defined by

$$r_l(t) = \phi_l^{(2)}(t) - \sum_{i=1}^{m} a_{il}\phi_i^{(1)}(t)$$

where, for each $l \in \{1, ..., m\}$, the coefficients a_{il} $(i = 1, ..., m)$ are chosen to minimize $\left\|\phi_l^{(2)} - \sum a_{il}\phi_i^{(1)}\right\|^2 = \int_0^1 \left|\phi_l^{(2)}(t) - \sum a_{il}\phi_i^{(1)}(t)\right|^2 dt$. Note that, under H_0, we have $r_l(t) \equiv 0$, $\sum_{i=1}^m a_{il}^2 = 1$ and $\sum_{i=1}^m a_{il}a_{il'} = 0$ for all $l \neq l'$, i.e. $\langle a_{\cdot l}, a_{\cdot l'}\rangle = \delta_{ll'}$ where $a_{\cdot l} = (a_{1l}, ..., a_{ml})^T$. This means that, there exists an orthogonal transformation (orthogonal matrix) between two basis functions $\phi_1^{(1)}(t)$,...,$\phi_m^{(1)}(t)$ and $\phi_1^{(2)}(t)$,...,$\phi_m^{(2)}(t)$. On the other hand, under H_A, we can find at least one $l \in \{1, ..., m\}$ such that $\|r_l\|^2 > 0$.

Now we suppose that the observations consist of two independent samples $Y_{ij}^{(k)}$ $(k = 1, 2)$ as defined in (5.1) and (5.2). The corresponding covariance functions $C^{(1)}(s, t)$, $C^{(2)}(s, t)$ and their eigenvalues $\lambda_l^{(1)}$, $\lambda_l^{(2)}$ and eigenfunctions $\phi_l^{(1)}(t)$, $\phi_l^{(2)}(t)$ are estimated by $\hat{C}^{(1)}(s, t)$, $\hat{C}^{(2)}(s, t)$, $\hat{\lambda}_l^{(1)}$, $\hat{\lambda}_l^{(2)}$, $\hat{\phi}_l^{(1)}(t)$, $\hat{\phi}_l^{(2)}(t)$ $(l = 1, ..., m)$ respectively (see (3.21) and (4.1) or (4.2) as mentioned in Auxiliary results section). Then the residual functions $r_l(t)$ $(l = 1, ..., m)$ can be estimated by

$$\hat{r}_l(t) = \hat{r}_l(t) = \hat{\phi}_l^{(2)}(t) - \sum_{i=1}^{m} \hat{a}_{il}\hat{\phi}_i^{(1)} \quad (l = 1, ..., m) \tag{5.4}$$

where $\hat{a}_{il} = \left\langle \hat{\phi}_l^{(2)}, \hat{\phi}_i^{(1)} \right\rangle$. Standardized residual functions are defined by

$$\tilde{r}_l(t) = \sqrt{\frac{n^{(1)}n^{(2)}}{n^{(1)} + n^{(2)}}}\hat{r}_l(t) \tag{5.5}$$

where $n^{(1)}$ and $n^{(2)}$ are the sampling points of the two independent samples respectively. The asymptotic distribution of $\tilde{r}_l(t)$ will be discussed in the next section.

5.2.2 Asymptotic distribution of test statistic $\tilde{r}_l(t)$

In this section, we derive the asymptotic distribution of standardized residual functions $\tilde{r}_l(t)$. For simplicity of presentation we assume that $p^{(1)} = p^{(2)} =: p$. This assumption can be dropped without any change in the proof.

Theorem 5.1. *Suppose that H_0 is true, $X_i^{(1)}(t)$ and $X_i^{(2)}(t)$ admit the representation (5.3) for some $p \leq \infty$, $\mathcal{U}$ and $\mathcal{V}$ are defined as above with $m \leq p$ and the assumptions of Theorem 4.2 hold. Assume also that there exists a constant $\eta \in (0,1)$ such that*

$$\frac{n^{(1)}}{n^{(1)} + n^{(2)}} \rightarrow \eta \ (\text{as } n^{(1)}, n^{(2)} \rightarrow \infty).$$

Furthermore, define, for $i \neq j \in \{1, ..., p\}$,

$$\Lambda_{ij}^{(1)} = \sqrt{\lambda_i^{(1)}\lambda_j^{(1)}} \left(\lambda_i^{(1)} - \lambda_j^{(1)}\right)^{-1}$$

and

$$\Lambda_{ij}^{(2)} = \sqrt{\lambda_i^{(2)}\lambda_j^{(2)}} \left(\lambda_i^{(2)} - \lambda_j^{(2)}\right)^{-1}$$

and denote by $\zeta_{ij}^{(1)}$, $\zeta_{ij}^{(2)}$ $(i, j \in \mathbb{N})$ independent $N(0,1)$ random variables. Then

$$\tilde{r}_l(t) \Rightarrow Z_{res;l}(t)$$

where "$\Rightarrow$" denotes weak convergence in $C[0,1]$ equipped with the uniform metric,

$$Z_{res;l}(t) = Z_{res;l,1}(t) - Z_{res;l,2}(t)$$

and

$$Z_{res;l,1}(t) = \sqrt{\eta} \sum_{k=m+1}^{p} \Lambda_{lk}^{(2)} \left(\phi_k^{(2)}(t) - \sum_{i=1}^{m} a_{ik}\phi_i(t)^{(1)} \right) \zeta_{lk}^{(2)},$$

$$Z_{res;l,2}(t) = \sqrt{1-\eta} \sum_{k=m+1}^{p} \sum_{i=1}^{m} \Lambda_{ik}^{(1)} \left(a_{kl}\phi_i^{(1)}(t) + a_{il}\phi_k^{(1)}(t) \right) \zeta_{ik}^{(1)}.$$

Remark 5.1. *If $m = p$, then $\tilde{r}_l(t) \Rightarrow 0$. In order to obtain a nondegenerate limiting distribution, one may need to consider a higher order expansion of $\tilde{r}_l(t)$.*

Remark 5.2. *In general, the processes $Z_{res;l_1}(t)$ ($t \in [0,1]$) and $Z_{res;l_2}(t)$ ($t \in [0,1]$) are not independent.*

Remark 5.3. *Since $Z_{res;l,1}(t)$ and $Z_{res;l,2}(t)$ are Gaussian and independent of each other, $Z_{res;l,1}(t) - Z_{res;l,2}(t)$ may be replaced by $Z_{res;l,1}(t) + Z_{res;l,2}(t)$. The sign of $Z_{res;l,2}(t)$ is relevant only, if for instance the joint asymptotic distribution of $\hat{r}_l(t)$ and $\hat{\phi}_l(t)$ is of interest.*

Theorem 5.1 can be used to define tests for H_0. For instance, we consider the squared integral of the estimated standardized residual functions:

$$\tilde{U}_l := \int_0^1 \tilde{r}_l^2(t)dt \quad (l = 1, 2, ..., m).$$

Then Theorem 5.1 implies that under H_0,

$$\tilde{U}_l \underset{d}{\rightarrow} U_l = \int_0^1 \left(Z_{\text{res};l,1}(t) - Z_{\text{res};l,2}(t) \right)^2 dt \quad (l = 1, 2, ..., m).$$

Note that under H_0, we expect that for all $l = 1, 2, ..., m$, $\tilde{U}_l \underset{d}{\rightarrow} U_l$. Therefore, using only one of these test statistics is not sufficient, since the test would not be consistent in general. One may prefer to use some adjusted correction to deal with this multiple comparitions problem. For instance, since in most applications m is small, one may prefer to use a simple Bonferroni adjusted test.

A simple Bonferroni adjusted test at the level of significance $\alpha \in (0,1)$ can be defined as follows. Denote by F_{U_l} the distribution function of U_l and by $q_{\alpha,m;l} = F_{U_l}^{-1}(1 - \alpha/m)$ its $(1-\alpha/m)-$quantile. A Bonferroni corrected rejection region can be defined by

$$K_\alpha = \left\{ Y_{ij}^{(k)}(k = 1, 2) : \ \tilde{U}_l > q_{\alpha,m;l} \text{ for at least one } l \in \{1, ..., m\} \right\}. \qquad (5.6)$$

By definition, the actual level of significance does not exceed α.

In order to obtain less conservative or even asymptotically exact rejection regions, we need to calculate the joint asymptotic distribution of $\tilde{r}_1(t), ..., \tilde{r}_m(t)$. In fact, this joint distribution can be derived straightforwardly by using analogous arguments as in Theorem 5.1, since $Z_{\mathrm{res};l,1}(t)$ and $Z_{\mathrm{res};l,2}(t)$ are Gaussian and independent of each other. However, the formulas are quite complicated and depend on unknown quantities.

In applications where m is small one may prefer to use a simple Bonferroni adjusted test, such as the one given here. A more practical solution is the bootstrap test discussed in the next section.

Remark 5.4. *In particular, Theorem 5.1 implies the following formulas for the first two moments of U_l:*

$$E[U_l] = \eta \sum_{k=m+1}^{p} \left\{\Lambda_{lk}^{(2)}\right\}^2 \left(1 - \sum_{i=1}^{m} a_{ik}^2\right) + (1-\eta) \sum_{k=m+1}^{p} \sum_{i=1}^{m} \left\{\Lambda_{ik}^{(1)}\right\}^2 \left(a_{kl}^2 + a_{il}^2\right)$$

and

$$var(U_l) = A_1 + A_2 + A_3$$

with

$$A_1 = 2\eta^2 \sum_{k=m+1}^{p} \left\{\Lambda_{lk}^{(2)}\right\}^2 \left(A_{11} - 2A_{12} + A_{13}\right),$$

$$A_{11} = \left\{\Lambda_{lk}^{(2)}\right\}^2,$$

$$A_{12} = \left\{\Lambda_{lk}^{(2)}\right\}^2 \sum_{i=1}^{m} a_{ik}^2,$$

$$A_{13} = \sum_{k'=m+1}^{p} \left\{\Lambda_{lk'}^{(2)}\right\}^2 \left(\sum_{i=1}^{m} a_{ik} a_{ik'}\right)^2,$$

$$A_2 = 4\eta(1-\eta) \sum_{k,k'=m+1}^{p} \sum_{i=1}^{m} \left\{\Lambda_{lk}^{(2)}\right\}^2 \left\{\Lambda_{ik'}^{(2)}\right\}^2 a_{il} a_{k'k},$$

$$A_3 = 2(1-\eta)^2 \left(A_{31} + 2A_{32} + A_{33}\right),$$

$$A_{31} = \sum_{i=1}^{m} \left(\sum_{k=m+1}^{p} a_{kl}^2 \left\{\Lambda_{ik}^{(1)}\right\}^2\right)^2,$$

$$A_{32} = \sum_{k=m+1}^{p} a_{kl}^2 \sum_{i=1}^{m} a_{il}^2 \left\{\Lambda_{ik}^{(1)}\right\}^4$$

and

$$A_{33} = \sum_{k=m+1}^{p} \left(\sum_{i=1}^{m} a_{il}^2 \left\{ \Lambda_{ik}^{(1)} \right\}^2 \right)^2 .$$

5.2.3 A bootstrap test

The test considered above has the following several shortcomings:

- Using the statistics $\tilde{U}_l$ together with a Bonferroni correction is likely to lead to a conservative test, in particular for large values of m.
- There is an artificial asymmetry in that the estimated eigenfunctions of the second sample are projected on those of the first sample.
- The asymptotic distribution depends on unknown parameters, including the dimension p as well as all p eigenfunctions and eigenvalues.

Therefore, in this section, we propose an improved test to solve these problems. The idea is to define a test statistic that is symmetric and at the same time suitable for a simple bootstrap procedure that does not require knowledge of unknown nuisance parameters.

First we address the question of symmetry. Define estimated residual functions

$$\hat{r}_l(t;1) = \hat{\phi}_l^{(2)}(t) - \sum_{i=1}^{m} \hat{a}_{il}(1)\hat{\phi}_i^{(1)}(t) \ (l = 1, ..., m),$$
$$\hat{r}_l(t;2) = \hat{\phi}_l^{(1)}(t) - \sum_{i=1}^{m} \hat{a}_{il}(2)\hat{\phi}_i^{(2)}(t) \ (l = 1, ..., m)$$

where $\hat{a}_{il}(1) = \left\langle \hat{\phi}_l^{(2)}, \hat{\phi}_i^{(1)} \right\rangle$, $\hat{a}_{il}(2) = \left\langle \hat{\phi}_l^{(1)}, \hat{\phi}_i^{(2)} \right\rangle$. Thus, $\hat{r}_l(t;1)$ and $\hat{r}_l(t;2)$ are the residuals obtained from orthogonal projection of $\hat{\phi}_l^{(2)}$ on *span* $\left\{ \hat{\phi}_1^{(1)}, ..., \hat{\phi}_m^{(1)} \right\}$ and the projection of $\hat{\phi}_l^{(1)}$ on *span* $\left\{ \hat{\phi}_1^{(2)}, ..., \hat{\phi}_m^{(2)} \right\}$ respectively. The corresponding standardized statistics $\tilde{U}_l$ are denoted by $\tilde{U}_l(1)$ and $\tilde{U}_l(2)$. We may then define an overall statistic

$$T_{n,N} = \max \left\{ \tilde{U}_1(1), ..., \tilde{U}_m(1), \tilde{U}_1(2), ..., \tilde{U}_m(2) \right\} .$$

By definition, $T_{n,N}$ is symmetric with respect to the two samples. Moreover, no Bonferroni correction is needed, if the correct quantiles of $T_{n,N}$ can be calculated or approximated with sufficient accuracy. In principle, asymptotic rejection regions based on $T_{n,N}$ follow directly from the joint asymptotic distribution of the standardized residuals $\tilde{r}_l(t;k)$ $(k = 1, 2;\ l = 1, ..., m)$. However, the asymptotic formulas are very complicated and depend on unknown quantities (as we can see in Theorem 5.1). Therefore, it is more useful to define a suitable bootstrap procedure that avoids explicit estimation of nuisance parameters. Since $\tilde{r}_l(t;k)$ $(k = 1, 2;\ l = 1, ..., m)$ are asymptotically jointly normal and the observed series are independent, a bootstrap method can be applied. However, some care needs to be taken to guarantee that the bootstrap distribution of $T_{n,N}$ approximates the null distribution even if H_1 is true.

Before describing the algorithm, we start with some notations and basic observations. Let $m \leq p$, denote

$$\mathcal{U}_m = span\left\{\phi_1^{(1)}, ..., \phi_m^{(1)}\right\},$$

$$\mathcal{V}_m = span\left\{\phi_1^{(2)}, ..., \phi_m^{(2)}\right\}$$

and $\mathcal{U}_{m,p}^{\perp}$ and $\mathcal{V}_{m,p}^{\perp}$ be the orthogonal complement of $\mathcal{U}_m$ and $\mathcal{V}_m$ respectively. When estimated eigenfunctions $\hat{\phi}_l^{(k)}$ are used, we will write $\widehat{\mathcal{U}}_m$, $\widehat{\mathcal{V}}_m$, $\widehat{\mathcal{U}}_{m,p}^{\perp}$ and $\widehat{\mathcal{V}}_{m,p}^{\perp}$ instead. Moreover, orthogonal projections to the corresponding subspaces will be denoted by $P_{\mathcal{U}_m}$, $P_{\mathcal{V}_m}$, $P_{\mathcal{U}_{m,p}^{\perp}}$ and $P_{\mathcal{V}_{m,p}^{\perp}}$. We are testing $H_0 : \mathcal{U}_m = \mathcal{V}_m$. Let

$$X_{i,m}^{(k)}(t) = \sum_{l=1}^{m} \xi_{il}^{(k)} \phi_l^{(k)}(t)$$

and

$$R_{i,m,p}^{(k)}(t) = \sum_{l=m+1}^{p} \xi_{il}^{(k)} \phi_l^{(k)}(t).$$

The observed series $Y_{ij}^{(k)}$ $(k = 1, 2;\ i = 1, ..., n,\ j = 1, ...N)$ can be written as

$$\begin{aligned} Y_{ij}^{(k)} &= X_i^{(k)}(t_j) + \epsilon_i^{(k)}(j) \\ &= X_{i,m}^{(k)}(t_j) + \left(R_{i,m,p}^{(k)}(t_j) + \epsilon_i^{(k)}(j)\right) \\ &= X_{i,m}^{(k)}(t_j) + E_i^{(k)}(t_j) \end{aligned}$$

where $t_j = j/N$, and

$$E_i^{(k)}(t_j) = R_{i,m,p}^{(k)}(t_j) + \epsilon_i^{(k)}(j).$$

For the functions $X_{i,m}^{(k)}(t)$, $R_{i,m,p}^{(k)}(t) \in L^2[0,1]$, we have under H_0,

$$X_{i,m}^{(1)}(t) \in \mathcal{U}_m = \mathcal{V}_m = \mathcal{U}_m \cap \mathcal{V}_m,$$

$$X_{i,m}^{(2)}(t) \in \mathcal{V}_m = \mathcal{U}_m = \mathcal{U}_m \cap \mathcal{V}_m,$$

$$R_{i,m,p}^{(1)}(t) \in \mathcal{U}_{m,p}^{\perp} = \mathcal{V}_{m,p}^{\perp} = \mathcal{U}_{m,p}^{\perp} \cap \mathcal{V}_{m,p}^{\perp},$$

and

$$R_{i,m,p}^{(2)}(t) \in \mathcal{V}_{m,p}^{\perp} = \mathcal{U}_{m,p}^{\perp} = \mathcal{U}_{m,p}^{\perp} \cap \mathcal{V}_{m,p}^{\perp}.$$

Thus, in particular, we obtain

$$\begin{aligned} P_{\mathcal{U}_m}\left(X_{i,m}^{(2)}(t)\right) &= P_{\mathcal{V}_m}\left(X_{i,m}^{(2)}(t)\right) \\ &= \sum_{s=1}^{m} \left\langle X_{i,m}^{(2)}(t), \phi_s^{(2)}(t) \right\rangle \phi_s^{(2)}(t) \\ &= \sum_{s=1}^{m} \left\langle \sum_{l=1}^{m} \xi_{il}^{(2)} \phi_l^{(2)}(t), \phi_s^{(2)}(t) \right\rangle \phi_s^{(2)}(t) \\ &= X_{i,m}^{(2)}(t), \end{aligned}$$

$$\begin{aligned} P_{\mathcal{U}_{m,p}^{\perp}}\left(R_{i,m,p}^{(2)}(t)\right) &= P_{\mathcal{V}_{m,p}^{\perp}}\left(\sum_{l=m+1}^{p} \xi_{il}^{(2)} \phi_l^{(2)}(t)\right) \\ &= \sum_{s=m+1}^{p} \left\langle \sum_{l=m+1}^{p} \xi_{il}^{(2)} \phi_l^{(2)}(t), \phi_s^{(2)}(t) \right\rangle \phi_s^{(2)}(t) \\ &= \sum_{l=m+1}^{p} \xi_{il}^{(2)} \phi_l^{(2)}(t) \\ &= R_{i,m,p}^{(2)}(t), \end{aligned}$$

$$\begin{aligned} P_{\mathcal{V}_m}\left(X_{i,m}^{(1)}(t)\right) &= P_{\mathcal{U}_m}\left(X_{i,m}^{(1)}(t)\right) \\ &= \sum_{s=1}^{m} \left\langle X_{i,m}^{(1)}(t), \phi_s^{(1)}(t) \right\rangle \phi_s^{(1)}(t) \\ &= \sum_{s=1}^{m} \left\langle \sum_{l=1}^{m} \xi_{il}^{(1)} \phi_l^{(1)}(t), \phi_s^{(1)}(t) \right\rangle \phi_s^{(1)}(t) \\ &= X_{i,m}^{(1)}(t), \end{aligned}$$

and

$$
\begin{aligned}
P_{\mathcal{V}_{m,p}^{\perp}}\left(R_{i,m,p}^{(1)}(t)\right) &= P_{\mathcal{U}_{m,p}^{\perp}}\left(\sum_{l=m+1}^{p} \xi_{il}^{(1)}\phi_l^{(1)}(t)\right) \\
&= \sum_{s=m+1}^{p}\left\langle \sum_{l=m+1}^{p} \xi_{il}^{(1)}\phi_l^{(1)}(t), \phi_s^{(1)}(t)\right\rangle \phi_s^{(1)}(t) \\
&= \sum_{l=m+1}^{p} \xi_{il}^{(1)}\phi_l^{(1)}(t) \\
&= R_{i,m,p}^{(1)}(t).
\end{aligned}
$$

Due to weak convergence of the eigenfunctions $\hat{\phi}_l^{(k)}(t)$ in $C[0,1]$, the projection operators $P_{\mathcal{U}_m}$, $P_{\mathcal{V}_m}$, $P_{\mathcal{U}_{m,p}^{\perp}}$ and $P_{\mathcal{V}_{m,p}^{\perp}}$ may be replaced by estimates (see Chapter 4 or Beran and Liu 2016).

The bootstrap algorithm is defined as follows:

- Step 1: Calculate $T_{n,N}$ for the observed data.
- Step 2: Obtain estimates $\hat{\phi}_1^{(k)}(t_j), ..., \hat{\phi}_m^{(k)}(t_j)$ and $\hat{\lambda}_1^{(k)}, ..., \hat{\lambda}_m^{(k)}$ $(k = 1, 2,\ t_j = j/M,\ j = 1, ..., M,\ M \in \mathbb{N})$ from sample 1 $(k = 1)$ and sample 2 $(k = 2)$ respectively.
- Step 3: Calculate $\hat{R}_{i,m,p}^{(k)}(t_j)$ $(k = 1, 2;\ i = 1, ..., n,\ t_j = j/M,\ j = 1, ..., M)$ based on estimates from Step 2. That is

$$\hat{R}_{i,m,p}^{(k)}(t_j) = \sum_{l=m+1}^{p} \hat{\xi}_{il}^{(k)}\hat{\phi}_l^{(k)}(t_j),$$

  where $\hat{\xi}_{il}^{(k)} \sim N\left(0, \hat{\lambda}_l^{(k)}\right)$.
- Step 4: For $i = 1, ..., n$ and $l = 1, ..., m$, set

$$\tilde{\phi}_l^{(1|1)}(t_j) = \frac{P_{\widehat{\mathcal{U}}_m}\left(\hat{\phi}_l^{(1)}(t_j)\right)}{\left\|P_{\widehat{\mathcal{U}}_m}\left(\hat{\phi}_l^{(1)}(\cdot)\right)\right\|_{L^2}} = \hat{\phi}_l^{(1)}(t_j),$$

$$\tilde{\phi}_l^{(2|1)}(t_j) = \frac{P_{\widehat{\mathcal{U}}_m}\left(\hat{\phi}_l^{(2)}(t_j)\right)}{\left\|P_{\widehat{\mathcal{U}}_m}\left(\hat{\phi}_l^{(2)}(\cdot)\right)\right\|_{L^2}},$$

$$\tilde{\phi}_l^{(1|2)}(t_j) = \frac{P_{\widehat{\mathcal{V}}_m}\left(\hat{\phi}_l^{(1)}(t_j)\right)}{\left\|P_{\widehat{\mathcal{V}}_m}\left(\hat{\phi}_l^{(1)}(\cdot)\right)\right\|_{L^2}},$$

$$\tilde{\phi}_l^{(2|2)}(t_j) = \frac{P_{\widehat{\mathcal{V}}_m}\left(\hat{\phi}_l^{(2)}(t_j)\right)}{\left\|P_{\widehat{\mathcal{V}}_m}\left(\hat{\phi}_l^{(2)}(\cdot)\right)\right\|_{L^2}} = \hat{\phi}_l^{(2)}(t_j),$$

$$\tilde{E}_i^{(1|1)}(t_j) = \hat{E}_i^{(1)}(t_j),$$

$$\tilde{E}_i^{(2|1)}(t_j) = P_{\widehat{\mathcal{U}}_m}\left(\hat{E}_i^{(2)}(t_j)\right),$$

$$\tilde{E}_i^{(1|2)}(t_j) = P_{\widehat{\mathcal{V}}_m}\left(\hat{E}_i^{(1)}(t_j)\right),$$

and

$$\tilde{E}_i^{(2|2)}(t_j) = \hat{E}_i^{(2)}(t_j).$$

All projections are evaluated on the grid $t_j = j/M$ $(j = 1, ..., M)$.

- Step 5: For each $s = 1, ..., N_{boot}$ simulate the bootstrap statistic $T_{s,1}^*$ as follows: First simulate n independent vectors of scores

$$\tilde{\xi}_{i\cdot} = \left(\tilde{\xi}_{i1}^{(1)}, ..., \tilde{\xi}_{im}^{(1)}, \tilde{\xi}_{i1}^{(2)}, ..., \tilde{\xi}_{im}^{(2)}\right) \ (i = 1, ..., n)$$

where $\tilde{\xi}_{il}^{(k)}$ are independent $N\left(0, \hat{\lambda}_l^{(k)}\right)$-variables. For each $i = 1, ..., n$, calculate the simulated series

$$\tilde{X}_{i,m}^{(k|1)*}(t_j) = \sum_{l=1}^{m} \tilde{\xi}_{il}^{(k)} \tilde{\phi}_l^{(k|1)}(t_j) \ (k = 1, 2, \ t_j = j/M).$$

Sample n series $\tilde{E}_i^{(1|1)*}(t_j)$ $(i = 1, ..., n, \ t_j = j/M)$ independently with replacement from $\left\{\tilde{E}_1^{(1|1)}(t_j), ..., \tilde{E}_n^{(1|1)}(t_j)\right\}$. Similarily, sample n series $\tilde{E}_i^{(2|1)*}(t_j)$ $(i = 1, ..., n)$ independently with replacement from $\left\{\tilde{E}_1^{(2|1)}(t_j), ..., \tilde{E}_n^{(2|1)}(t_j)\right\}$. Define the bootstraped samples 1 and 2 by

$$Y_{ij}^{(k|1)*} = \tilde{X}_{i,m}^{(k|1)*}(t_j) + \tilde{E}_i^{(k|1)*}(t_j) \ (k = 1, 2; \ i = 1, ..., n, \ t_j = j/M).$$

Denote the corresponding values of $\tilde{U}_l(k)$ $(k = 1, 2; \ l = 1, ..., m)$ by $\tilde{U}_l^*(k)$. Calculate the sth bootstrap statistic

$$T_{s,1}^* = max\left\{\tilde{U}_1^*(1), ..., \tilde{U}_m^*(1), \tilde{U}_1^*(2), ..., \tilde{U}_m^*(2)\right\}.$$

- Step 6: Carry out step 5 using $\tilde{\phi}_l^{(k|2)}(t_j)$ and $\tilde{E}_i^{(k|2)}(t_j)$ $(k = 1, 2,\ t_j = j/M)$ instead, to obtain the corresponding statistic $T^*_{s,2}$.

- Step 7: Given $\alpha \in (0, 1)$ and the bootstrap statistics $T^*_{s,1}$, $T^*_{s,2}$ $(s = 1, ..., N_{boot})$, calculate the empirical $(1 - \alpha)$ quantiles $q^*_{1-\alpha}(1)$ and $q^*_{1-\alpha}(2)$ of $T^*_{s,1}$ $(s = 1, ..., N_{boot})$ and $T^*_{s,2}$ $(s = 1, ..., N_{boot})$ respectively. Reject H_0 at the level of significance α, if

$$T_{n,N} > \min\left\{q^*_{1-\alpha}(1), q^*_{1-\alpha}(2)\right\}.$$

Note that for the calculation of $T^*_{s,1}$, the estimated basis functions $\hat{\phi}_l^{(2)}$ are replaced by $P_{\widehat{\mathcal{U}}_m}(\hat{\phi}_l^{(2)})/||P_{\widehat{\mathcal{U}}_m}(\hat{\phi}_l^{(2)})||$. This is necessary in order that the bootstraped distribution approximates the null distribution even if the data were generated under the alternative hypothesis. For the same reason the estimated basis functions $\hat{\phi}_l^{(1)}$ are replaced by $P_{\widehat{\mathcal{V}}_m}(\hat{\phi}_l^{(1)})/||P_{\widehat{\mathcal{V}}_m}(\hat{\phi}_l^{(1)})||$ when calculating $T^*_{s,2}$. Note also that asymptotically, $||P_{\widehat{\mathcal{U}}_m}(\hat{\phi}_l^{(2)})|| = 1$ so that the standardization by the norm is not needed asymptotically. Finally note that using the minimum of the bootstrap quantiles $q^*_{1-\alpha}(1)$ and $q^*_{1-\alpha}(2)$ is justified by the fact that under H_0 the statistics $T^*_{s,1}$ and $T^*_{s,2}$ have the same asymptotic distribution. Defining the rejection region by the minimum of $q^*_{1-\alpha}(1)$ and $q^*_{1-\alpha}(2)$ leads to improved power for finite samples. A similar idea is discussed in Ghosh and Beran (2000) in the context of two-sample tests for univariate distributions.

In contrast to the asymptotic test discussed in the previous section, the bootstrap test does not require knowledge of unknown nuisance parameters. In particular, the value of p does not have to be specified and the remaining eigenfunctions $\phi_l^{(k)}$ and eigenvalues $\lambda_l^{(k)}$ $(k = 1, 2;\ l = m + 1, ..., p)$ do not have to be estimated. Note also that there is no need to model the residual processes $\epsilon_i^{(1)}(j)$ and $\epsilon_i^{(2)}(j)$.

We conclude this section by noting that an alternative test statistic, defined by

$$D_{4,m} = \int\int\left\{\sum_{l=1}^{m}\hat{\phi}_l^{(1)}(t)\,\hat{\phi}_l^{(1)}(s) - \sum_{l=1}^{m}\hat{\phi}_l^{(2)}(t)\,\hat{\phi}_l^{(2)}(s)\right\}^2 dtds, \qquad (5.7)$$

has been proposed previously by Benko et al. (2009). For the noiseless case with $\epsilon_i(j) \equiv 0$, Benko et al. (2009) derive the asymptotic null distribution of $D_{4,m}$ and define a bootstrap procedure. Moreover, arguments are also given why the asymptotic results should not change when iid noise is added to the curves X_i. In view of Lemmas 1, 2 and 3 given above (sections 2.1 and 2.2.), the asymptotic distribution of $D_{4,m}$ may be derived for general error processes as defined in (3.1), (3.3), (3.4) or (3.5) (and under the conditions specified in Theorem 5.1). Though a detailed proof would have to be carried out, it may be conjectured that the asymptotic distribution is the same as in the noiseless case.

5.3 Simulations

5.3.1 Distribution of $\tilde{U}_l$

To illustrate the performance of the test based on (5.6), the following simulation study is carried out. It consists four cases: 1) $m = 1$, $p = 2$; 2) $m = 1$, $p = 10$; 3) $m = 2$, $p = 3$; 4) $m = 2$, $p = 10$. For each case a) the same eigenspaces generated by the same eigenfunctions, b) the same eigenspaces generated by the rotated eigenfunctions, and c) the unequal eigenspaces are discussed.

We consider two independent samples of sizes $n^{(1)} = n^{(2)} = n$ as defined in (5.1), (5.2). The error process $\epsilon_i^{(1)}(j)$ $(j \in \mathbb{N})$ and $\epsilon_i^{(2)}(j)$ $(j \in \mathbb{N})$ are assumed to be the same and generated by one of the following Gaussian processes: (a) iid $N(0, 1)$; (b) $AR(1)$ with lag-one correlation $\rho = 0.5$ and variance one; (c) $FARIMA(0, 0.3, 0)$ process with variance one (see e.g. Granger and Joyeux 1980, Hosking 1981). The number of time series are chosen as $N^{(1)} = N^{(2)} = N = 200$, 400, 600, 800, and 1000. According to the conditions in Theorem 4.1 (or Theorem 3.5) the corresponding number of sampling points on each time series are set as $n_N^{(1)} = n_N^{(2)} = n_N = 10N^{0.6}$ rounded to the next integer (i.e. $n = 240$, 364, 464, 552, and 631 respectively).

For the kernel estimators $\hat{C}^{(1)}(s,t)$ and $\hat{C}^{(2)}(s,t)$, the same estimation procedure is used. We choose the rectangular product kernel $K_2(u, v) = K_1(u)K_1(v)$ with $K_1(u) = \frac{1}{2}\mathbf{1}\{-1 \le u \le 1\}$. The bandwidth is defined as $b = b_N = 0.05N^{-0.16}$. In this case $n_N^{-\frac{1}{4}} = 10^{-1/4}N^{-0.15}$, so the conditions in Theorem 4.1 hold. Note that $d = 0$ for iid and AR(1) errors, whereas $d = 0.3$ for the FARIMA process. In order to save calculation time, we set the girds of $\hat{C}^{(1)}(s,t)$ and $\hat{C}^{(2)}(s,t)$ equal to 500, i.e. the matrices $\hat{C}^{(1)}(s_j, t_k)$ and $\hat{C}^{(2)}(s_j, t_k)$ $(j, k = 1, ..., 500)$ with $s_j = j/500$ and $t_k = k/500$ is calculated for each simulation. The corresponding eigenvalues $\hat{\lambda}_l$ $(l = 1, ..., m)$ and eigenfunctions $\hat{\phi}_l(t)$ $(l = 1, ..., m)$ are calculated from the matrices $\hat{C}^{(1)}(s_j, t_k)$ and $\hat{C}^{(2)}(s_j, t_k)$ $(j, k = 1, ..., 500)$. In particularly, the eigenfunctions are in discrete forms which means that they are eigenvectors $(\hat{\phi}_l(t_1), ..., \hat{\phi}_l(t_{500}))$ $(t_j = j/500)$ with dimension 500.

The results summarized below are based on 400 simulations (for each model and combination of N and n_N).

More specifically, the following situations are considered:

- Model 1a) ($m = 1$, $p = 2$, H_0 with equal eigenfunction):

  For the first sample, we have

  $$Y_{ij}^{(1)} = \xi_{i1}^{(1)}\phi_1^{(1)}(t_j) + \xi_{i2}^{(1)}\phi_2^{(1)}(t_j) + \epsilon_i(j) \quad (i = 1, ..., n; j = 1, ..., N) \tag{5.8}$$

  with $\xi_{il}^{(1)} \sim N\left(0, \lambda_l^{(1)}\right)$ $(l = 1, 2)$, $\lambda_1^{(1)} = 6$, $\lambda_2^{(1)} = 4$,

  $$\phi_1^{(1)}(t) = \sqrt{2}\cos(\pi t),\ \phi_2^{(1)}(t) = \sqrt{2}\cos(2\pi t).$$

  For the second sample, we have

  $$Y_{ij}^{(2)} = \xi_{i1}^{(2)}\phi_1^{(2)}(t_j) + \xi_{i2}^{(2)}\phi_2^{(2)}(t_j) + \epsilon_i(j) \quad (i = 1, ..., n; j = 1, ..., N) \tag{5.9}$$

  with $\xi_{il}^{(2)} \sim N\left(0, \lambda_l^{(2)}\right)$ $(l = 1, 2)$, $\lambda_2^{(2)} = 2$, $\lambda_1^{(2)} = 6$,

  $$\phi_1^{(2)}(t) = \sqrt{2}\cos(\pi t),\ \phi_2^{(2)}(t) = \sqrt{2}\cos(4\pi t).$$

  This means that the spaces $\mathcal{U}$ and $\mathcal{V}$ spanned by $\phi_1^{(1)}(t)$, $\phi_2^{(1)}(t)$ and $\phi_1^{(2)}(t)$, $\phi_2^{(2)}(t)$ respectively are different. However, since $\phi_1^{(1)}(t) = \phi_1^{(2)}(t)$ and we only test for $m = 1$ (i.e. $\mathcal{U}$ and $\mathcal{V}$ are based on $\phi_1^{(1)}(t)$, $\phi_1^{(2)}(t)$ only respectively), we have $\mathcal{U} = \mathcal{V}$ and H_0 is true.

- Model 1b) ($m = 1$, $p = 2$, H_0 with rotated eigenfunction): The two samples $Y_{ij}^{(1)}$ and $Y_{ij}^{(2)}$ are defined as in (5.8) and (5.9) respectively, with the exception that

  $$\phi_1^{(2)}(t) = -\sqrt{2}\cos(\pi t).$$

  This model is different from Model 1a) in that $\phi_1^{(2)}(t)$ is not equal to $\phi_1^{(1)}(t)$ but instead it is a rotated version of $\phi_1^{(1)}(t)$. Since we only test for $m = 1$ and rotation does not change the eigenspace, we have $\mathcal{U} = \mathcal{V}$ and H_0 is true.

- Model 1c) ($m = 1$, $p = 2$, H_1): The two samples $Y_{ij}^{(1)}$ and $Y_{ij}^{(2)}$ are defined as in (5.8) and (5.9) respectively, with the exception that

$$\phi_1^{(2)}(t) = \sqrt{2}\cos(2\pi t).$$

In this case, we have $\mathcal{U} \neq \mathcal{V}$.

- Model 2a) ($m = 1$, $p = 10$, H_0 with equal eigenfunction):

For the first sample, we have

$$Y_{ij}^{(1)} = \xi_{i1}^{(1)}\phi_1^{(1)}(t_j) + \sum_{l=2}^{10} \xi_{il}^{(1)}\phi_l^{(1)}(t_j) + \epsilon_i(j) \tag{5.10}$$

with $\xi_{il}^{(1)} \sim N\left(0, \lambda_l^{(1)}\right)$ ($l = 1, ..., 10$), $\lambda_1^{(1)} = 6$,

$$\left(\lambda_2^{(1)}, ..., \lambda_{10}^{(1)}\right) = (4, 1, 1/2, 1/4, 1/6, 1/8, 1/10, 1/12, 1/14),$$

$$\phi_1^{(1)}(t) = \sqrt{2}\cos(\pi t),$$

$$\begin{aligned}\left(\phi_2^{(1)}(t), ..., \phi_{10}^{(1)}(t)\right) =&\sqrt{2}(\cos 2\pi t, \cos 6\pi t, \cos 10\pi t, \cos 14\pi t,\\ &\cos 18\pi t, \cos 22\pi t, \cos 26\pi t, \cos 30\pi t, \cos 34\pi t).\end{aligned}$$

For the second sample, we have

$$Y_{ij}^{(2)} = \xi_{i1}^{(2)}\phi_1^{(2)}(t_j) + \sum_{l=2}^{10} \xi_{il}^{(2)}\phi_l^{(2)}(t_j) + \epsilon_i(j) \tag{5.11}$$

with $\xi_{il}^{(2)} \sim N\left(0, \lambda_l^{(2)}\right)$ ($l = 1, ..., 10$), $\lambda_1^{(2)} = 6$,

$$\left(\lambda_2^{(2)}, ..., \lambda_{10}^{(2)}\right) = (2, 1, 1/3, 1/5, 1/7, 1/9, 1/11, 1/13, 1/15),$$

$$\phi_1^{(2)}(t) = \sqrt{2}\cos(\pi t)$$

and

$$\begin{aligned}\left(\phi_2^{(2)}(t), ..., \phi_{10}^{(2)}(t)\right) =&\sqrt{2}(\cos 4\pi t, \cos 8\pi t, \cos 12\pi t, \cos 16\pi t,\\ &\cos 20\pi t, \cos 24\pi t, \cos 28\pi t, \cos 32\pi t, \cos 36\pi t).\end{aligned}$$

This means that the spaces $\mathcal{U}$ and $\mathcal{V}$ spanned by $\phi_1^{(1)}(t), \phi_2^{(1)}(t), ..., \phi_{10}^{(1)}(t)$ and $\phi_1^{(2)}(t), \phi_2^{(2)}(t), ..., \phi_{10}^{(2)}(t)$ respectively are different. However, since $\phi_1^{(1)}(t) = \phi_1^{(2)}(t)$ and we only test for $m = 1$ (i.e. $\mathcal{U}$ and $\mathcal{V}$ are based on $\phi_1^{(1)}(t)$, $\phi_1^{(2)}(t)$ only), we have $\mathcal{U} = \mathcal{V}$ and H_0 is true.

- Model 2b) ($m = 1$, $p = 10$, H_0 and rotated eigenfunction): The two samples $Y_{ij}^{(1)}$ and $Y_{ij}^{(2)}$ are defined as in (5.10) and (5.11) respectively, with the exception that
$$\phi_1^{(2)}(t) = -\sqrt{2}\cos(\pi t).$$
This model is different from Model 2a) in that $\phi_1^{(2)}(t)$ is not equal to $\phi_1^{(1)}(t)$ but instead it is a rotated version of $\phi_1^{(1)}(t)$. Since we only test for $m = 1$ and rotation does not change the eigenspace, we have $\mathcal{U} = \mathcal{V}$ and H_0 is true.

- Model 2c) ($m = 1$, $p = 10$, H_1): The two samples $Y_{ij}^{(1)}$ and $Y_{ij}^{(2)}$ as in (5.10) and (5.11) respectively, with the exception that
$$\phi_1^{(2)}(t) = \sqrt{2}\cos(2\pi t).$$
In this case, we have $\mathcal{U} \neq \mathcal{V}$.

- Model 3a) ($m = 2$, $p = 3$, H_0 with equal eigenfunctions):
For the first sample, we have
$$Y_{ij}^{(1)} = \xi_{i1}^{(1)}\phi_1^{(1)}(t_j) + \xi_{i2}^{(1)}\phi_2^{(1)}(t_j) + \xi_{i3}^{(1)}\phi_3^{(1)}(t_j) + \epsilon_i(j) \tag{5.12}$$
with $\xi_{il}^{(1)} \sim N\left(0, \lambda_l^{(1)}\right)$ $(l = 1, 2, 3)$, $\lambda_1^{(1)} = 6$, $\lambda_2^{(1)} = 4$,
$$\lambda_3^{(1)} = 2,$$
$$\left(\phi_1^{(1)}(t), \phi_2^{(1)}(t)\right) = \sqrt{2}\left(\cos(\pi t), \cos(2\pi t)\right),$$
$$\phi_3^{(1)}(t) = \sqrt{2}\cos(6\pi t).$$
For the second sample, we have
$$Y_{ij}^{(2)} = \xi_{i1}^{(2)}\phi_1^{(2)}(t_j) + \xi_{i2}^{(2)}\phi_2^{(2)}(t_j) + \xi_{i3}^{(2)}\phi_3^{(2)}(t_j) + \epsilon_i(j), \tag{5.13}$$
with $\xi_{il}^{(2)} \sim N\left(0, \lambda_l^{(2)}\right)$ $(l = 1, 2, 3)$, $\lambda_1^{(2)} = 6$, $\lambda_2^{(2)} = 3$,
$$\lambda_3^{(2)} = 1,$$
$$\left(\phi_1^{(2)}(t), \phi_2^{(2)}(t)\right) = \left(\sqrt{2}\cos(\pi t), \sqrt{2}\cos(2\pi t)\right),$$
$$\phi_3^{(2)}(t) = \sqrt{2}\cos(8\pi t).$$

In this case, the spaces $\mathcal{U}$ and $\mathcal{V}$ spanned by $\phi_1^{(1)}(t), \phi_2^{(1)}(t), \phi_3^{(1)}(t)$ and $\phi_1^{(2)}(t), \phi_2^{(2)}(t), \phi_3^{(2)}(t)$ respectively are different. However, since $\phi_1^{(1)}(t) = \phi_1^{(2)}(t)$ and $\phi_2^{(1)}(t) = \phi_2^{(2)}(t)$, and we only test for $m = 2$, we have $\mathcal{U} = \mathcal{V}$ and H_0 is true.

- Model 3b) ($m = 2$, $p = 3$, H_0 with rotated eigenfunctions): The two samples $Y_{ij}^{(1)}$ and $Y_{ij}^{(2)}$ are defined as in (5.12) and (5.13) respectively, with the exception that

$$\left(\phi_1^{(2)}(t), \phi_1^{(2)}(t)\right) = \left(\cos(\pi t) + \cos(2\pi t), \cos(\pi t) - \cos(2\pi t)\right).$$

In this case, we have $\phi_1^{(2)}(t)$ and $\phi_2^{(2)}(t)$ are different from $\phi_1^{(1)}(t)$ and $\phi_2^{(1)}(t)$, $\phi_1^{(2)}(t)$ and $\phi_2^{(2)}(t)$ but can be obtained by rotation of $\phi_1^{(1)}(t)$ and $\phi_2^{(1)}(t)$. Since we only test for $m = 2$ ($\mathcal{U}$ and $\mathcal{V}$ are based on $\phi_1^{(1)}(t)$, $\phi_2^{(1)}(t)$ and $\phi_1^{(2)}(t)$, $\phi_2^{(2)}(t)$) and rotation does not change the eigenspace, we have $\mathcal{U} = \mathcal{V}$ and H_0 is true.

- Model 3c) ($m = 2$, $p = 3$, H_1): The two samples $Y_{ij}^{(1)}$ and $Y_{ij}^{(2)}$ as in (5.12) and (5.13) respectively, with the exception that

$$\left(\phi_1^{(2)}(t), \phi_1^{(2)}(t)\right) = \left(\sqrt{2}\cos(\pi t), \sqrt{2}\cos(4\pi t)\right).$$

In this case, we have $\mathcal{U} \neq \mathcal{V}$.

- Model 4a) ($m = 2$, $p = 10$, H_0 with equal eigenfunctions):
For the first sample, we have

$$Y_{ij}^{(1)} = \xi_{i1}^{(1)}\phi_1^{(1)}(t_j) + \xi_{i2}^{(1)}\phi_2^{(1)}(t_j) + \sum_{l=3}^{10} \xi_{il}^{(1)}\phi_l^{(1)}(t_j) + \epsilon_i(j) \tag{5.14}$$

with $\xi_{il}^{(1)} \sim N\left(0, \lambda_l^{(1)}\right)$ $(l = 1, ..., 10)$, $\lambda_1^{(1)} = 6$, $\lambda_2^{(1)} = 4$,

$$\left(\lambda_3^{(1)}, ..., \lambda_{10}^{(1)}\right) = (2, 1/2, 1/4, 1/6, 1/8, 1/10, 1/12, 1/14),$$

$$\left(\phi_1^{(1)}(t), \phi_2^{(1)}(t)\right) = \sqrt{2}\left(\cos(\pi t), \cos(2\pi t)\right),$$

$$\begin{aligned}\left(\phi_3^{(1)}(t), ..., \phi_{10}^{(1)}(t)\right) =& \sqrt{2}(\cos 6\pi t, \cos 10\pi t, \cos 14\pi t, \cos 18\pi t,\\ &\cos 22\pi t, \cos 26\pi t, \cos 30\pi t, \cos 34\pi t).\end{aligned}$$

For the second sample , we have

$$Y_{ij}^{(2)} = \xi_{i1}^{(2)}\phi_1^{(2)}(t_j) + \xi_{i2}^{(2)}\phi_2^{(2)}(t_j) + \sum_{l=3}^{10} \xi_{il}^{(2)}\phi_l^{(2)}(t_j) + \epsilon_i(j) \tag{5.15}$$

with $\xi_{il}^{(2)} \sim N\left(0, \lambda_l^{(2)}\right)$ $(l = 1, ..., 10)$, $\lambda_1^{(2)} = 6$, $\lambda_2^{(2)} = 3$,

$$\left(\lambda_3^{(2)}, ..., \lambda_{10}^{(2)}\right) = (1, 1/3, 1/5, 1/7, 1/9, 1/11, 1/13, 1/15),$$

$$\left(\phi_1^{(2)}(t), \phi_1^{(2)}(t)\right) = \left(\sqrt{2}\cos(\pi t), \sqrt{2}\cos(2\pi t)\right),$$

$$\left(\phi_3^{(2)}(t), ..., \phi_{10}^{(2)}(t)\right) = \sqrt{2}(\cos 8\pi t, \cos 12\pi t, \cos 16\pi t, \cos 20\pi t, \cos 24\pi t, \cos 28\pi t, \cos 32\pi t, \cos 36\pi t).$$

In this case the spaces $\mathcal{U}$ and $\mathcal{V}$ spanned by $\phi_1^{(1)}(t)$,...,$\phi_{10}^{(1)}(t)$ and $\phi_1^{(2)}(t)$,...,$\phi_{10}^{(2)}(t)$ respectively are different. However, since $\phi_1^{(1)}(t) = \phi_1^{(2)}(t)$ and $\phi_2^{(1)}(t) = \phi_2^{(2)}(t)$, and we only test for $m = 2$, we have $\mathcal{U} = \mathcal{V}$ and H_0 is true.

- Model 4b) ($m = 2$, $p = 10$, H_0 with rotated eigenfunctions): The two samples $Y_{ij}^{(1)}$ and $Y_{ij}^{(2)}$ are defined as in (5.14) and (5.15) respectively, with the exception that

$$\left(\phi_1^{(2)}(t), \phi_1^{(2)}(t)\right) = \left(\cos(\pi t) + \cos(2\pi t), \cos(\pi t) - \cos(2\pi t)\right).$$

In this case, $\phi_1^{(2)}(t)$ and $\phi_2^{(2)}(t)$ differ from $\phi_1^{(1)}(t)$ and $\phi_2^{(1)}(t)$, but can be obtained by rotation of $\phi_1^{(1)}(t)$ and $\phi_2^{(1)}(t)$. Since we only test for $m = 2$ and rotation does not change the eigenspace, we have $\mathcal{U} = \mathcal{V}$ and H_0 is true.

- Model 4c) ($m = 2$, $p = 10$, H_1): The two samples $Y_{ij}^{(1)}$ and $Y_{ij}^{(2)}$ are defined as in (5.14) and (5.15) respectively, with the exception that

$$\left(\phi_1^{(2)}(t), \phi_1^{(2)}(t)\right) = \left(\sqrt{2}\cos(\pi t), \sqrt{2}\cos(4\pi t)\right).$$

In this case, we have $\mathcal{U} \neq \mathcal{V}$.

Tables 5.1, 5.2 and 5.3 give the simulation results for Models 1a), 1b) and 1c) respectively. It can be seen from Tables 5.1 and 5.2 that under the null hypothesis H_0 the simulated values converge to the asymptotic values in a reasonably fast speed, even in the presence of a long range dependent error process. In particular, the dependence structure in error process does not influence the accuracy of the level of significance. Moreover, under the alternative hypothesis H_A chosen here, the test is obviously very powerful. This agrees with our simulation design. Similar comments apply to the results for Model 2, given in Tables 5.4, 5.5 and 5.6 respectively.

For Models 3 and 4, for each case, we have two test statistics $\tilde{U}_1$ and $\tilde{U}_2$. Simulated rejection frequencies for the individual test based on $\tilde{U}_1$ and the individual test based on $\tilde{U}_2$ are given in the fourth and third column from the right respectively (in Tables 5.7, 5.8 and 5.9 for Model 3, and Tables 5.10, 5.11 and 5.12 for Model 4). Note that using only one of these test statistics is not sufficient, since the test would not be consistent in general. In fact, as we can see in our design, for the unequal spaces case Models 3c) and 4c), that $\phi_1^{(1)}(t) = \phi_1^{(2)}(t)$ but $\phi_2^{(1)}(t) \neq \phi_2^{(2)}(t)$. Therefore, in the fourth and third columns from the right hand side in Tables 5.9 and 5.12, the level of significance for $\tilde{U}_1$ tends to some constant smaller than 1, but for $\tilde{U}_2$ the level of significance is always 1. Therefore, we give the rejection frequencies for the joint Bonferroni corrected test defined in (5.6) in last two columns . From the last two columns in Tables 5.7, 5.8, 5.10 and 5.11 we can see that the level of significance does not exceed the nominal level α, but is generally lower. In addition, the results in Tables 5.9 and 5.12 illustrate that the test based on $\tilde{U}_1$ alone would not be very powerful. However the combined Bonferroni corrected test always rejects H_0. These results are what we expected. Moreover, the accuracy of the level of significance does not seem to be influenced by the dependence structure.

5.3.2 Bootstrap test based on $T_{n,N}$

Here, the finite sample performance of the bootstrap test based on $T_{n,N}$ (Section 4) is illustrated. For sample sizes $N^{(1)} = N^{(2)} = N = 200$ and 400, $n_N^{(1)} = n_N^{(2)} = n_N = 10N^{0.6}$ rounded to the next integer, $N_{simul} = 100$ simulations are carried out. The errors $\epsilon_i^{(1)}(j)$ and $\epsilon_i^{(2)}(j)$ are generated by one of the Gaussian processes specified in (a), (b) and (c) above. The following situations are considered:

- Model 1 (H_0):

$$m = 2,\ p = 10,$$

$$\left(\lambda_1^{(1)}, \lambda_2^{(1)}, ..., \lambda_{10}^{(1)}\right) = (6, 4, 2, 1/2, 1/4, 1/6, 1/8, 1/10, 1/12, 1/14)\,,$$

$$\left(\lambda_1^{(2)}, \lambda_2^{(2)}, ..., \lambda_{10}^{(2)}\right) = (6, 3, 2, 1/3, 1/5, 1/7, 1/9, 1/11, 1/13, 1/15),$$

$$\left(\phi_1^{(1)}(t), ..., \phi_{10}^{(1)}(t)\right) = \sqrt{2}(\cos \pi t, \cos 2\pi t, \cos 6\pi t, \cos 10\pi t, \cos 14\pi t, \cos 18\pi t, \cos 22\pi t, \cos 26\pi t, \cos 30\pi t, \cos 34\pi t),$$

$$\left(\phi_1^{(2)}(t), ..., \phi_{10}^{(2)}(t)\right) = \sqrt{2}(\frac{\cos \pi t + \cos 2\pi t}{\sqrt{2}}, \frac{\cos \pi t - \cos 2\pi t}{\sqrt{2}}, \cos 6\pi t, \cos 8\pi t, \cos 16\pi t, \cos 20\pi t, \cos 24\pi t, \cos 28\pi t, \cos 32\pi t, \cos 36\pi t).$$

- Model 2 (H_1):

$$m = 2,\ p = 10,$$

$$\left(\lambda_1^{(1)}, \lambda_2^{(1)}, ..., \lambda_{10}^{(1)}\right) = (6, 4, 2, 1/2, 1/4, 1/6, 1/8, 1/10, 1/12, 1/14)\,,$$

$$\left(\lambda_1^{(2)}, \lambda_2^{(2)}, ..., \lambda_{10}^{(2)}\right) = (6, 3, 2, 1/3, 1/5, 1/7, 1/9, 1/11, 1/13, 1/15),$$

$$\left(\phi_1^{(1)}(t), ..., \phi_{10}^{(1)}(t)\right) = \sqrt{2}(\cos \pi t, \cos 2\pi t, \cos 6\pi t, \cos 10\pi t, \cos 14\pi t, \cos 18\pi t, \cos 22\pi t, \cos 26\pi t, \cos 30\pi t, \cos 34\pi t),$$

$$\left(\phi_1^{(2)}(t), ..., \phi_{10}^{(2)}(t)\right) = \sqrt{2}(\frac{\cos \pi t + \cos 2\pi t}{\sqrt{2}}, \cos 6\pi t, \frac{\cos \pi t - \cos 2\pi t}{\sqrt{2}}, \cos 8\pi t, \cos 16\pi t, \cos 20\pi t, \cos 24\pi t, \cos 28\pi t, \cos 32\pi t, \cos 36\pi t).$$

- Model 3 (H_0):

$$m = 3,\ p = 10,$$

$$\left(\lambda_1^{(1)}, \lambda_2^{(1)}, ..., \lambda_{10}^{(1)}\right) = (6, 4, 2, 1/2, 1/4, 1/6, 1/8, 1/10, 1/12, 1/14)\,,$$

$$\left(\lambda_1^{(2)}, \lambda_2^{(2)}, ..., \lambda_{10}^{(2)}\right) = (6, 4, 3.5, 3, 1/5, 1/7, 1/9, 1/11, 1/13, 1/15),$$

$$\left(\phi_1^{(1)}(t), ..., \phi_{10}^{(1)}(t)\right) = \sqrt{2}(\cos \pi t, \cos 2\pi t, \cos 6\pi t, \cos 10\pi t, \cos 14\pi t, \cos 18\pi t, \cos 22\pi t, \cos 26\pi t, \cos 30\pi t, \cos 34\pi t),$$

$$\left(\phi_1^{(2)}(t), ..., \phi_{10}^{(2)}(t)\right) = \sqrt{2}(\cos \pi t, \cos 2\pi t, \cos 6\pi t, \cos 8\pi t, \cos 12\pi t,$$
$$\cos 16\pi t, \cos 20\pi t, \cos 22\pi t, \cos 24\pi t, \cos 26\pi t).$$

- Model 4 (H_1):

$$m = 3,\ p = 10,$$
$$\left(\lambda_1^{(1)}, \lambda_2^{(1)}, ..., \lambda_{10}^{(1)}\right) = (6, 4, 2, 1/2, 1/4, 1/6, 1/8, 1/10, 1/12, 1/14),$$
$$\left(\lambda_1^{(2)}, \lambda_2^{(2)}, ..., \lambda_{10}^{(2)}\right) = (6, 4, 3.5, 3, 1/5, 1/7, 1/9, 1/11, 1/13, 1/15),$$

$$\left(\phi_1^{(1)}(t), ..., \phi_{10}^{(1)}(t)\right) = \sqrt{2}(\cos \pi t, \cos 2\pi t, \cos 6\pi t, \cos 10\pi t, \cos 14\pi t,$$
$$\cos 18\pi t, \cos 22\pi t, \cos 26\pi t, \cos 30\pi t, \cos 34\pi t),$$

$$\left(\phi_1^{(2)}(t), ..., \phi_{10}^{(2)}(t)\right) = \sqrt{2}(\cos \pi t, \cos 2\pi t, \cos 10\pi t, \cos 8\pi t, \cos 12\pi t,$$
$$\cos 16\pi t, \cos 20\pi t, \cos 22\pi t, \cos 24\pi t, \cos 26\pi t).$$

- Model 5 (H_1):

$$m = 3,\ p = 10,$$
$$\left(\lambda_1^{(1)}, \lambda_2^{(1)}, ..., \lambda_{10}^{(1)}\right) = (6, 4, 2, 1/2, 1/4, 1/6, 1/8, 1/10, 1/12, 1/14),$$
$$\left(\lambda_1^{(2)}, \lambda_2^{(2)}, ..., \lambda_{10}^{(2)}\right) = (6, 4, 3, 2, 1/5, 1/7, 1/9, 1/11, 1/13, 1/15),$$

$$\left(\phi_1^{(1)}(t), ..., \phi_{10}^{(1)}(t)\right) = \sqrt{2}(\cos \pi t, \cos 2\pi t, \cos 6\pi t, \cos 10\pi t, \cos 14\pi t,$$
$$\cos 18\pi t, \cos 22\pi t, \cos 26\pi t, \cos 30\pi t, \cos 34\pi t),$$

$$\left(\phi_1^{(2)}(t), ..., \phi_{10}^{(2)}(t)\right) = \sqrt{2}(\cos \pi t, \cos 2\pi t, \frac{\cos 6\pi t - \cos 8\pi t}{\sqrt{2}}, \cos 10\pi t, \cos 12\pi t,$$
$$\cos 16\pi t, \cos 20\pi t, \cos 22\pi t, \cos 24\pi t, \cos 26\pi t).$$

Simulated rejection probabilities at the nominal level of $\alpha = 0.05$ are summarized in Table 5.13. The results for Models 1 and 3 indicate that, under H_0, the nominal level of significance is reached with good precision already for $N = 200$. With respect to power (Models 2, 4 and 5), the most difficult alternative is Model 5. The reason is that the deviation of $\mathcal{V}_3$ from $\mathcal{U}_3$ is relatively small, because

$\langle \phi_3^{(1)}, \phi_3^{(2)} \rangle = \sqrt{2}||\cos 6\pi t||^2 = 1/\sqrt{2} \neq 0$. In contrast, for Model 3, $\mathcal{U}_3$ is orthogonal to $\mathcal{V}_3$. Similarily, for Model 2, $\mathcal{U}_2$ is orthogonal to $\mathcal{V}_2$. The simulations show that even for Model 5 the power of the bootstrap test is reasonably high for moderately large sample sizes. The result is not influenced by the type of dependence in the noise process.

Finally, Table 5.14 shows simulation results for the same models, but using the bootstrap test (based on $D_{4,m}$) proposed in Benko et al. (2009). The results support the conjecture that this procedure is consistent even in the presence of very general noise processes. A detailed study of the asymptotic properties of $D_{4,m}$ under the general assumptions of Theorem 5.1, and a theoretical comparison of power properties of the tests based on $T_{n,N}$ and $D_{4,m}$ respectively will need to be addressed by future research.

5.4 Proofs and tables

5.4.1 Proofs

Proof. (of Theorem 5.1)

From the results in Chapter 4 about the joint distribution of estimated eigenfunctions, $\sqrt{n^{(1)}}\left(\hat{\phi}_1^{(1)} - \phi_1^{(1)}, ..., \hat{\phi}_m^{(1)} - \phi_m^{(1)}\right)$ and $\sqrt{n^{(2)}}\left(\hat{\phi}_1^{(2)} - \phi_1^{(2)}, ..., \hat{\phi}_m^{(2)} - \phi_m^{(2)}\right)$ converge weakly to the processes $Z^{(1)} = \left(Z_1^{(1)}, ..., Z_m^{(1)}\right)$ and $Z^{(2)} = \left(Z_1^{(2)}, ..., Z_m^{(2)}\right)$ respectively where

$$Z_l^{(1)}(t) = \sum_{k=l+1}^{p} \Lambda_{lk}^{(1)} \phi_k^{(1)} \zeta_{lk}^{(1)} + \sum_{k=1}^{l-1} \Lambda_{lk}^{(1)} \phi_k^{(1)} \zeta_{kl}^{(1)},$$

$$Z_l^{(2)}(t) = \sum_{k=l+1}^{p} \Lambda_{lk}^{(2)} \phi_k^{(2)} \zeta_{lk}^{(2)} + \sum_{k=1}^{l-1} \Lambda_{lk}^{(2)} \phi_k^{(2)} \zeta_{kl}^{(2)}.$$

Since the two samples are independent of each other, the processes $Z^{(1)}(t)$, $Z^{(2)}(t)$

are also independent of each other. Now $\tilde{r}_l(t)$ $(l = 1, ..., m)$ can be written as

$$\begin{aligned}
&\hat{r}_l(t) \\
&= \hat{\phi}_l^{(2)}(t) - \sum_{i=1}^{m} \hat{a}_{il}\hat{\phi}_i^{(1)}(t) \\
&= \hat{\phi}_l^{(2)} - \sum_{i=1}^{m} \left\langle \hat{\phi}_l^{(2)}, \hat{\phi}_i^{(1)} \right\rangle \hat{\phi}_i^{(1)} \\
&= \phi_l^{(2)} + \left(\hat{\phi}_l^{(2)} - \phi_l^{(2)}\right) \\
&- \sum_{i=1}^{m} \left\langle \left(\phi_l^{(2)} + \left(\hat{\phi}_l^{(2)} - \phi_l^{(2)}\right)\right), \left(\phi_i^{(1)} + \left(\hat{\phi}_i^{(1)} - \phi_i^{(1)}\right)\right) \right\rangle \left(\phi_i^{(1)} + \left(\hat{\phi}_i^{(1)} - \phi_i^{(1)}\right)\right) \\
&= \phi_l^{(2)} - \sum_{i=1}^{m} \left\langle \phi_l^{(2)}, \phi_i^{(1)} \right\rangle \phi_i^{(1)} + \left(\hat{\phi}_l^{(2)} - \phi_l^{(2)}\right) \\
&- \sum_{i=1}^{m} \left[\left\langle \phi_l^{(2)}, \left(\hat{\phi}_i^{(1)} - \phi_i^{(1)}\right) \right\rangle \phi_i^{(1)} + \left\langle \left(\hat{\phi}_l^{(2)} - \phi_l^{(2)}\right), \phi_i^{(1)} \right\rangle \phi_i^{(1)} \right] \\
&- \sum_{i=1}^{m} \left\langle \left(\hat{\phi}_l^{(2)} - \phi_l^{(2)}\right), \left(\hat{\phi}_i^{(1)} - \phi_i^{(1)}\right) \right\rangle \phi_i^{(1)} \\
&- \sum_{i=1}^{m} \left\langle \phi_l^{(2)}, \phi_i^{(1)} \right\rangle \left(\hat{\phi}_i^{(1)} - \phi_i^{(1)}\right) + \sum_{i=1}^{m} \left\langle \phi_l^{(2)}, \left(\hat{\phi}_i^{(1)} - \phi_i^{(1)}\right) \right\rangle \left(\hat{\phi}_i^{(1)} - \phi_i^{(1)}\right) \\
&- \sum_{i=1}^{m} \left\langle \left(\hat{\phi}_l^{(2)} - \phi_l^{(2)}\right), \phi_i^{(1)} \right\rangle \left(\hat{\phi}_i^{(1)} - \phi_i^{(1)}\right) + \sum_{i=1}^{m} \left\langle \left(\hat{\phi}_l^{(2)} - \phi_l^{(2)}\right), \left(\hat{\phi}_i^{(1)} - \phi_i^{(1)}\right) \right\rangle \left(\hat{\phi}_i^{(1)} - \phi_i^{(1)}\right).
\end{aligned}$$

Note that

$$\phi_l^{(2)} - \sum_{i=1}^{m} \left\langle \phi_l^{(2)}, \phi_i^{(1)} \right\rangle \phi_i^{(1)} = 0.$$

Moreover

$$\frac{n^{(1)}}{n^{(1)} + n^{(2)}} \to \eta \text{ as } n^{(1)}, n^{(2)} \to \infty$$

implies

$$\sqrt{\frac{n^{(1)}n^{(2)}}{n^{(1)} + n^{(2)}}} \sim \sqrt{\eta n^{(2)}} \sim \sqrt{(1-\eta)\, n^{(1)}}.$$

Therefore, the standardized residual functions

$$\tilde{r}_l(t) = \sqrt{\frac{n^{(1)}n^{(2)}}{n^{(1)} + n^{(2)}}}\hat{r}_l(t)$$

converges weakly to

$$\begin{aligned}
&\sqrt{\eta} Z_l^{(2)} - \sqrt{1-\eta} \sum_{i=1}^{m} \left\langle \phi_l^{(2)}, Z_i^{(1)} \right\rangle \phi_i^{(1)} \\
&\quad - \sqrt{\eta} \sum_{i=1}^{m} \left\langle Z_l^{(2)}, \phi_i^{(1)} \right\rangle \phi_i^{(1)} - \sqrt{1-\eta} \sum_{i=1}^{m} a_{il} Z_i^{(1)} \\
&= \sqrt{\eta} \left(\sum_{k=l+1}^{p} \Lambda_{lk}^{(2)} \phi_k^{(2)} \zeta_{lk}^{(2)} + \sum_{k=1}^{l-1} \Lambda_{lk}^{(2)} \phi_k^{(2)} \zeta_{kl}^{(2)} \right) \\
&\quad - \sqrt{1-\eta} \sum_{i=1}^{m} \left\langle \phi_l^{(2)}, \left(\sum_{k=i+1}^{p} \Lambda_{ik}^{(1)} \phi_k^{(1)} \zeta_{ik}^{(1)} + \sum_{k=1}^{i-1} \Lambda_{ik}^{(1)} \phi_k^{(1)} \zeta_{ki}^{(1)} \right) \right\rangle \phi_i^{(1)} \\
&\quad - \sqrt{\eta} \sum_{i=1}^{m} \left\langle \left(\sum_{k=l+1}^{p} \Lambda_{lk}^{(2)} \phi_k^{(2)} \zeta_{lk}^{(2)} + \sum_{k=1}^{l-1} \Lambda_{lk}^{(2)} \phi_k^{(2)} \zeta_{kl}^{(2)} \right), \phi_i^{(1)} \right\rangle \phi_i^{(1)} \\
&\quad - \sqrt{1-\eta} \sum_{i=1}^{m} a_{il} \left(\sum_{k=i+1}^{p} \Lambda_{ik}^{(1)} \phi_k^{(1)} \zeta_{ik}^{(1)} + \sum_{k=1}^{i-1} \Lambda_{ik}^{(1)} \phi_k^{(1)} \zeta_{ki}^{(1)} \right).
\end{aligned}$$

Since $l = 1, ..., m$ and $m \leq p \leq \infty$, we can write

$$\tilde{r}_l(t) \Rightarrow J_1(t) + J_2(t)$$

where $J_1(t)$ is the term with sum of k over $\{1, ..., l-1, l+1, ..., m\}$ and $J_2(t)$ is the term with sum of k over $\{m, ..., p\}$. It is easy to see that $J_1(t) = 0$ and $J_2(t)$ is Gaussian as follows.

Specifically,

$$
\begin{aligned}
J_1(t) = {} & \sqrt{\eta}\left(\sum_{k=l+1}^{m} \Lambda_{lk}^{(2)}\phi_k^{(2)}\zeta_{lk}^{(2)} + \sum_{k=1}^{l-1}\Lambda_{lk}^{(2)}\phi_k^{(2)}\zeta_{kl}^{(2)}\right) \\
& - \sqrt{\eta}\sum_{i=1}^{m}\left(\sum_{k=l+1}^{m}\Lambda_{lk}^{(2)}a_{ik}\zeta_{lk}^{(2)} + \sum_{k=1}^{l-1}\Lambda_{lk}^{(2)}a_{ik}\zeta_{kl}^{(2)}\right)\phi_i^{(1)} \\
& - \sqrt{1-\eta}\sum_{i=1}^{m}\left(\sum_{k=i+1}^{m}\Lambda_{ik}^{(1)}a_{kl}\zeta_{ik}^{(1)} + \sum_{k=1}^{i-1}\Lambda_{ik}^{(1)}a_{kl}\zeta_{ki}^{(1)}\right)\phi_i^{(1)} \\
& - \sqrt{1-\eta}\sum_{i=1}^{m}a_{il}\left(\sum_{k=i+1}^{m}\Lambda_{ik}^{(1)}\phi_k^{(1)}\zeta_{ik}^{(1)} + \sum_{k=1}^{i-1}\Lambda_{ik}^{(1)}\phi_k^{(1)}\zeta_{ki}^{(1)}\right) \\
= {} & \sqrt{\eta}\left(\sum_{k=l+1}^{m} \Lambda_{lk}^{(2)}\phi_k^{(2)}\zeta_{lk}^{(2)} + \sum_{k=1}^{l-1}\Lambda_{lk}^{(2)}\phi_k^{(2)}\zeta_{kl}^{(2)}\right) \\
& - \sqrt{\eta}\left(\sum_{k=l+1}^{m} \Lambda_{lk}^{(2)}\phi_k^{(2)}\zeta_{lk}^{(2)} + \sum_{k=1}^{l-1}\Lambda_{lk}^{(2)}\phi_k^{(2)}\zeta_{kl}^{(2)}\right) \\
& - \sqrt{1-\eta}\sum_{i=1}^{m}\left(\sum_{k=i+1}^{m}\Lambda_{ik}^{(1)}a_{kl}\zeta_{ik}^{(1)} + \sum_{k=1}^{i-1}\Lambda_{ik}^{(1)}a_{kl}\zeta_{ki}^{(1)}\right)\phi_i^{(1)} \\
& - \sqrt{1-\eta}\sum_{i=1}^{m}a_{il}\left(\sum_{k=i+1}^{m}\Lambda_{ik}^{(1)}\phi_k^{(1)}\zeta_{ik}^{(1)} + \sum_{k=1}^{i-1}\Lambda_{ik}^{(1)}\phi_k^{(1)}\zeta_{ki}^{(1)}\right) \\
= {} & 0
\end{aligned}
$$

and

$$
\begin{aligned}
J_2(t) = {} & \sqrt{\eta}\sum_{k=m+1}^{p}\Lambda_{lk}^{(2)}\phi_k^{(2)}\zeta_{lk}^{(2)} \\
& - \sqrt{\eta}\sum_{i=1}^{m}\sum_{k=m+1}^{p}\Lambda_{lk}^{(2)}a_{ik}\zeta_{lk}^{(2)}\phi_i^{(1)} \\
& - \sqrt{1-\eta}\sum_{i=1}^{m}\sum_{k=m+1}^{p}\Lambda_{ik}^{(1)}a_{kl}\zeta_{ik}^{(1)}\phi_i^{(1)} \\
& - \sqrt{1-\eta}\sum_{i=1}^{m}a_{il}\sum_{k=m+1}^{p}\Lambda_{ik}^{(1)}\phi_k^{(1)}\zeta_{ik}^{(1)} \\
= {} & \sqrt{\eta}\sum_{k=m+1}^{p}\Lambda_{lk}^{(2)}\left(\phi_k^{(2)} - \sum_{i=1}^{m}a_{ik}\phi_i^{(1)}\right)\zeta_{lk}^{(2)} \\
& - \sqrt{1-\eta}\sum_{k=m+1}^{p}\sum_{i=1}^{m}\Lambda_{ik}^{(1)}\left(a_{kl}\phi_i^{(1)} + a_{il}\phi_k^{(1)}\right)\zeta_{ik}^{(1)}
\end{aligned}
$$

is Gaussian. Therefore, we obtain

$$\tilde{r}_l(t) \Rightarrow Z_{\mathrm{res};l}(t) = Z_{\mathrm{res};l,1}(t) - Z_{\mathrm{res};l,2}(t)$$

where

$$Z_{\mathrm{res};l,1}\left(t\right) = \sqrt{\eta} \sum_{k=m+1}^{p} \Lambda_{lk}^{(2)} \left(\phi_k^{(2)} - \sum_{i=1}^{m} a_{ik} \phi_i^{(1)} \right) \zeta_{lk}^{(2)},$$

$$Z_{\mathrm{res};l,2}\left(t\right) = \sqrt{1-\eta} \sum_{k=m+1}^{p} \sum_{i=1}^{m} \Lambda_{ik}^{(1)} \left(a_{kl} \phi_i^{(1)} + a_{il} \phi_k^{(1)} \right) \zeta_{ik}^{(1)}.$$

□

5.4.2 Tables

N	E	V	Q_{25}	Q_{50}	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$\alpha(0.05)$	$\alpha(0.01)$
∞	3.35	18.1	0.647	1.783	4.398	8.463	11.993	20.026	0.050	0.010
$\epsilon_i(j)$: iid										
200	4.36	36.9	0.922	2.503	5.530	10.246	13.737	30.879	0.072	0.025
400	3.88	22.6	0.780	2.105	5.188	9.575	12.822	20.679	0.062	0.012
600	4.01	24.1	0.736	2.131	5.684	9.289	14.281	22.209	0.065	0.018
800	4.21	33.1	0.826	2.462	5.113	10.323	13.437	26.176	0.072	0.020
1000	3.85	23.7	0.890	2.222	4.928	9.191	12.206	20.148	0.055	0.012
$\epsilon_i(j)$: AR(1)										
200	4.47	38.3	1.077	2.513	5.539	9.837	13.883	31.214	0.072	0.028
400	3.94	22.9	0.833	2.180	5.322	9.872	13.715	21.342	0.065	0.015
600	4.05	24.1	0.731	2.151	5.560	9.591	14.352	22.970	0.065	0.015
800	4.27	33.3	0.823	2.457	5.107	10.009	13.417	25.539	0.075	0.020
1000	3.88	23.9	0.860	2.220	5.069	9.430	12.405	19.972	0.058	0.008
$\epsilon_i(j)$: FARIMA$(0, 0.3, 0)$										
200	4.37	31.8	1.000	2.591	5.503	10.334	13.103	33.646	0.075	0.028
400	3.92	23.3	0.739	2.185	5.106	10.318	13.449	19.917	0.072	0.010
600	4.10	24.3	0.742	2.277	5.695	9.526	14.626	20.583	0.068	0.018
800	4.25	33.5	0.865	2.545	5.061	10.140	13.856	25.934	0.075	0.022
1000	3.86	25.0	0.845	2.122	4.730	9.357	12.725	21.418	0.055	0.015

Table 5.1: Model 1a) ($m = 1$, $p = 2$, H_0 with equal eigenfunctions): Asymptotic ($N = \infty$) and simulated values of $E[\tilde{U}_1]$ (denoted by E), $var(\tilde{U}_1)$ (denoted by V), and 25%−, 50%−, 75%−, 90%−, 95%− and 99%−quantiles (Q_{25}, Q_{50}, Q_{75}, Q_{90}, Q_{95}, Q_{99}) of $\tilde{U}_1$ are given. The last two columns show the simulated rejection probabilities for the nominal levels $\alpha = 0.05$ and 0.01 respectively.

N	E	V	Q_{25}	Q_{50}	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$\alpha(0.05)$	$\alpha(0.01)$
∞	3.35	18.1	0.647	1.783	4.398	8.463	11.993	20.026	0.050	0.010
$\epsilon_i(j)$: iid										
200	4.35	36.9	0.886	2.423	5.493	9.938	13.893	31.270	0.070	0.025
400	3.88	22.7	0.785	2.159	5.148	9.548	13.029	20.506	0.062	0.012
600	4.00	24.1	0.693	2.137	5.630	9.394	14.513	22.518	0.065	0.018
800	4.21	33.0	0.830	2.490	5.150	10.295	13.394	25.937	0.070	0.020
1000	3.85	23.8	0.859	2.237	4.855	9.224	12.223	19.841	0.052	0.010
$\epsilon_i(j)$: AR(1)										
200	4.47	38.2	1.051	2.566	5.539	10.039	14.369	30.397	0.072	0.025
400	3.93	22.7	0.838	2.144	5.410	9.560	13.832	20.842	0.065	0.015
600	4.05	24.1	0.733	2.182	5.616	9.483	14.185	23.026	0.065	0.015
800	4.27	33.3	0.856	2.442	5.249	10.135	13.769	25.694	0.070	0.020
1000	3.88	23.9	0.861	2.211	5.010	9.447	12.481	19.912	0.058	0.010
FARIMA$(0, 0.3, 0)$										
200	4.37	31.5	1.030	2.588	5.596	10.112	13.218	33.341	0.068	0.025
400	3.92	23.1	0.805	2.209	4.906	10.277	13.315	21.178	0.068	0.012
600	4.10	24.0	0.784	2.347	5.878	9.512	14.509	21.471	0.065	0.015
800	4.26	33.3	0.908	2.529	5.173	9.983	13.853	26.300	0.070	0.022
1000	3.82	24.5	0.858	2.101	4.676	9.016	12.887	20.932	0.055	0.015

Table 5.2: Model 1b) ($m = 1$, $p = 2$, H_0 with rotated eigenfunction): Asymptotic ($N = \infty$) and simulated values of $E[\tilde{U}_1]$ (denoted by E), $var(\tilde{U}_1)$ (denoted by V), and 25%−, 50%−, 75%−, 90%−, 95%− and 99%−quantiles (Q_{25}, Q_{50}, Q_{75}, Q_{90}, Q_{95}, Q_{99}) of $\tilde{U}_1$ are given. The last two columns show the simulated rejection probabilities for the nominal levels $\alpha = 0.05$ and 0.01 respectively.

N	E	V	Q_{25}	Q_{50}	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$\alpha(0.05)$	$\alpha(0.01)$
					$\epsilon_i(j)$: iid					
200	116.1	37.0	115.0	118.1	119.5	119.9	120.0	120.0	1.00	1.00
400	178.5	23.1	177.1	180.3	181.7	181.9	182.0	182.0	1.00	1.00
600	228.4	24.2	226.7	230.3	231.7	232.0	232.0	232.0	1.00	1.00
800	272.2	32.7	271.3	274.1	275.5	275.9	276.0	276.0	1.00	1.00
1000	312.0	23.7	310.9	313.7	315.1	315.4	315.5	315.5	1.00	1.00
					$\epsilon_i(j)$: AR(1)					
200	116.0	38.4	115.0	118.0	119.5	119.9	120.0	120.0	1.00	1.00
400	178.5	23.2	177.0	180.3	181.7	182.0	182.0	182.0	1.00	1.00
600	228.4	24.2	226.7	230.3	231.7	231.9	232.0	232.0	1.00	1.00
800	272.1	33.0	271.1	274.0	275.5	275.9	276.0	276.0	1.00	1.00
1000	312.0	23.8	310.8	313.7	315.0	315.4	315.5	315.5	1.00	1.00
					$\epsilon_i(j)$: FARIMA$(0, 0.3, 0)$					
200	116.1	31.8	114.8	118.0	119.4	119.9	120.0	120.0	1.00	1.00
400	178.5	23.4	177.4	180.5	181.7	182.0	182.0	182.0	1.00	1.00
600	228.3	24.3	226.7	230.2	231.7	232.0	232.0	232.0	1.00	1.00
800	272.2	33.5	271.4	274.1	275.5	275.9	276.0	276.0	1.00	1.00
1000	312.1	24.7	311.1	313.8	315.1	315.4	315.5	315.5	1.00	1.00

Table 5.3: Model 1c) ($m = 1$, $p = 2$, H_1): Simulated values of $E[\tilde{U}_1]$ (denoted by E), $var(\tilde{U}_1)$ (denoted by V), and 25%−, 50%−, 75%−, 90%−, 95%− and 99%−quantiles (Q_{25}, Q_{50}, Q_{75}, Q_{90}, Q_{95}, Q_{99}) of $\tilde{U}_1$ are given. The last two columns show the simulated rejection probabilities for the nominal levels of significance $\alpha = 0.05$ and 0.01 respectively.

N	E	V	Q_{25}	Q_{50}	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$\alpha(0.05)$	$\alpha(0.01)$
∞	3.81	18.1	1.120	2.251	4.886	9.016	12.316	20.266	0.050	0.010
$\epsilon_i(j)$: iid										
200	4.74	34.1	1.331	2.905	5.919	10.780	14.028	30.470	0.072	0.025
400	4.28	22.4	1.222	2.536	5.777	9.741	13.194	20.821	0.062	0.012
600	4.42	24.1	1.191	2.555	6.068	9.727	14.633	22.371	0.065	0.018
800	4.63	33.1	1.255	2.888	5.564	10.815	13.651	26.330	0.075	0.020
1000	4.27	24.0	1.324	2.559	5.444	9.581	12.589	20.641	0.055	0.015
$\epsilon_i(j)$: AR(1)										
200	4.85	35.8	1.485	2.918	5.990	10.459	14.308	31.576	0.078	0.028
400	4.33	22.8	1.220	2.550	5.771	10.191	14.187	21.497	0.065	0.015
600	4.46	24.2	1.185	2.640	6.089	9.864	14.699	23.114	0.065	0.015
800	4.68	33.3	1.299	2.900	5.652	10.521	13.585	25.684	0.080	0.020
1000	4.29	24.2	1.336	2.600	5.681	9.727	12.917	20.270	0.058	0.012
$\epsilon_i(j)$: FARIMA$(0, 0.3, 0)$										
200	4.73	30.2	1.422	2.958	5.909	10.659	13.751	27.199	0.072	0.028
400	4.31	23.0	1.200	2.566	5.483	10.691	13.955	20.052	0.075	0.010
600	4.51	24.3	1.281	2.568	6.184	10.141	15.139	20.920	0.068	0.018
800	4.67	33.6	1.343	2.858	5.531	10.616	14.288	26.562	0.072	0.022
1000	4.27	25.3	1.260	2.579	5.335	9.634	13.123	21.886	0.055	0.015

Table 5.4: Model 2a) ($m = 1$, $p = 10$, H_0 with equal eigenfunctions): Asymptotic ($N = \infty$) and simulated values of $E[\tilde{U}_1]$ (denoted by E), $var(\tilde{U}_1)$ (denoted by V), and 25%−, 50%−, 75%−, 90%−, 95%− and 99%−quantiles (Q_{25}, Q_{50}, Q_{75}, Q_{90}, Q_{95}, Q_{99}) of $\tilde{U}_1$ are given. The last two columns show the simulated rejection probabilities for the nominal levels $\alpha = 0.05$ and 0.01 respectively.

N	E	V	Q_{25}	Q_{50}	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$\alpha(0.05)$	$\alpha(0.01)$
∞	3.81	18.1	1.120	2.251	4.886	9.016	12.316	20.266	0.050	0.010
$\epsilon_i(j)$: iid										
200	4.73	34.2	1.405	2.933	5.799	10.538	14.241	30.794	0.072	0.028
400	4.29	22.4	1.193	2.637	5.696	9.743	13.304	20.664	0.062	0.012
600	4.41	24.1	1.172	2.595	5.998	9.637	14.865	22.680	0.065	0.018
800	4.63	33.0	1.276	2.900	5.615	10.736	13.539	26.075	0.075	0.020
1000	4.27	24.0	1.313	2.553	5.490	9.587	12.771	20.359	0.055	0.015
$\epsilon_i(j)$: AR(1)										
200	4.85	35.7	1.471	2.939	6.051	10.594	14.837	30.839	0.078	0.028
400	4.34	22.6	1.219	2.575	5.758	9.929	14.304	21.119	0.068	0.015
600	4.45	24.2	1.206	2.635	5.997	9.911	14.502	23.162	0.065	0.015
800	4.69	33.4	1.319	2.860	5.690	10.431	13.951	25.855	0.072	0.020
1000	4.30	24.2	1.315	2.568	5.528	9.684	12.946	20.336	0.055	0.012
$\epsilon_i(j)$: FARIMA$(0, 0.3, 0)$										
200	4.73	30.0	1.449	2.957	6.058	10.688	13.567	26.589	0.068	0.025
400	4.31	22.8	1.251	2.573	5.418	10.633	13.398	21.343	0.068	0.012
600	4.50	24.0	1.276	2.684	6.324	9.948	15.038	21.661	0.068	0.015
800	4.67	33.3	1.388	2.871	5.518	10.533	14.199	26.405	0.072	0.022
1000	4.24	24.8	1.266	2.544	5.365	9.296	13.160	21.289	0.058	0.015

Table 5.5: Model 2b) ($m = 1$, $p = 10$, H_0 with rotated eigenfunction): Asymptotic ($N = \infty$) and simulated values of $E[\tilde{U}_1]$ (denoted by E), $var(\tilde{U}_1)$ (denoted by V), and 25%−, 50%−, 75%−, 90%−, 95%− and 99%−quantiles (Q_{25}, Q_{50}, Q_{75}, Q_{90}, Q_{95}, Q_{99}) of $\tilde{U}_1$ are given. The last two columns show the simulated rejection probabilities for the nominal levels $\alpha = 0.05$ and 0.01 respectively.

N	E	V	Q_{25}	Q_{50}	Q_{75}	Q_{90}	Q_{95}	Q_{99}	$\alpha(0.05)$	$\alpha(0.01)$
$\epsilon_i(j)$: iid										
200	116.1	33.9	115.1	118.0	119.5	119.9	120.0	120.0	1.00	1.00
400	178.5	22.7	177.1	180.3	181.6	181.9	182.0	182.0	1.00	1.00
600	228.4	23.9	226.6	230.2	231.7	232.0	232.0	232.0	1.00	1.00
800	272.2	32.4	271.2	274.1	275.5	275.9	276.0	276.0	1.00	1.00
1000	312.0	23.6	310.9	313.7	315.1	315.4	315.5	315.5	1.00	1.00
$\epsilon_i(j)$: AR(1)										
200	116.0	35.6	114.9	117.9	119.5	119.9	120.0	120.0	1.00	1.00
400	178.5	22.9	177.1	180.3	181.7	182.0	182.0	182.0	1.00	1.00
600	228.4	23.9	226.7	230.3	231.7	231.9	232.0	232.0	1.00	1.00
800	272.2	32.7	271.2	274.1	275.5	275.9	276.0	276.0	1.00	1.00
1000	312.0	23.8	310.9	313.7	315.0	315.4	315.5	315.5	1.00	1.00
$\epsilon_i(j)$: FARIMA$(0, 0.3, 0)$										
200	116.1	30.0	114.8	118.0	119.5	119.9	120.0	120.0	1.00	1.00
400	178.5	22.9	177.4	180.5	181.7	182.0	182.0	182.0	1.00	1.00
600	228.3	24.0	226.7	230.2	231.7	232.0	232.0	232.0	1.00	1.00
800	272.2	33.2	271.4	274.1	275.5	275.9	276.0	276.0	1.00	1.00
1000	312.1	24.6	311.1	313.8	315.1	315.4	315.5	315.5	1.00	1.00

Table 5.6: Model 2c) ($m = 1$, $p = 10$, H_1): Simulated values of $E[\tilde{U}_1]$ (denoted by E), $var(\tilde{U}_1)$ (denoted by V), and 25%−, 50%−, 75%−, 90%−, 95%− and 99%−quantiles (Q_{25}, Q_{50}, Q_{75}, Q_{90}, Q_{95}, Q_{99}) of $\tilde{U}_1$ are given. The last two columns show the simulated rejection probabilities for the nominal levels of significance $\alpha = 0.05$ and 0.01 respectively.

N		E	V	Q_{90}	Q_{95}	Q_{99}	$\alpha(0.05)$	$\alpha(0.01)$	$\alpha_{\text{Bonf}}(0.05)$	$\alpha_{\text{Bonf}}(0.01)$
∞	U_1	0.49	0.31	1.166	1.602	2.636	0.050	0.010		
	U_2	1.36	2.25	3.191	4.308	7.177	0.050	0.010		
$\epsilon_i(j)$: iid										
200	U_1	0.55	0.29	1.272	1.614	2.413	0.052	0.008	0.0300	0.0025
	U_2	1.36	2.18	3.257	4.326	7.800	0.052	0.012		
400	U_1	0.56	0.36	1.266	1.648	2.642	0.052	0.012	0.0350	0.0075
	U_2	1.37	2.46	3.033	4.046	6.677	0.045	0.005		
600	U_1	0.56	0.36	1.352	1.826	2.988	0.065	0.015	0.0300	0.0100
	U_2	1.34	2.46	3.147	4.619	6.983	0.060	0.010		
800	U_1	0.55	0.32	1.235	1.707	2.725	0.055	0.015	0.0325	0.0050
	U_2	1.38	2.32	3.368	4.509	7.258	0.055	0.012		
1000	U_1	0.53	0.30	1.148	1.513	2.580	0.042	0.010	0.0225	0.0050
	U_2	1.39	2.32	3.535	4.494	6.596	0.058	0.010		
$\epsilon_i(j)$: AR(1)										
200	U_1	0.59	0.31	1.325	1.754	2.597	0.060	0.008	0.0275	0.0075
	U_2	1.45	2.24	3.485	4.366	7.271	0.052	0.012		
400	U_1	0.59	0.36	1.322	1.690	2.647	0.055	0.012	0.0350	0.0075
	U_2	1.44	2.49	3.351	4.159	6.622	0.038	0.008		
600	U_1	0.58	0.36	1.320	1.786	2.948	0.068	0.012	0.0300	0.0100
	U_2	1.38	2.35	3.221	4.568	7.131	0.058	0.010		
800	U_1	0.57	0.33	1.266	1.640	2.895	0.052	0.012	0.0325	0.0075
	U_2	1.41	2.32	3.394	4.291	7.439	0.050	0.012		
1000	U_1	0.55	0.31	1.230	1.600	2.654	0.050	0.012	0.0200	0.0050
	U_2	1.42	2.30	3.485	4.517	6.874	0.062	0.010		
$\epsilon_i(j)$: FARIMA(0,0.3,0)										
200	U_1	0.57	0.27	1.325	1.565	2.182	0.048	0.002	0.0225	0.0050
	U_2	1.42	2.15	3.175	4.654	7.309	0.062	0.012		
400	U_1	0.60	0.39	1.319	1.735	2.931	0.058	0.015	0.0400	0.0075
	U_2	1.46	2.43	3.208	4.210	6.906	0.045	0.008		
600	U_1	0.60	0.35	1.394	1.865	2.849	0.070	0.018	0.0300	0.0100
	U_2	1.42	2.41	3.271	4.582	7.476	0.060	0.015		
800	U_1	0.58	0.33	1.301	1.799	2.766	0.065	0.012	0.0325	0.0050
	U_2	1.44	2.32	3.290	4.877	7.315	0.055	0.012		
1000	U_1	0.56	0.30	1.207	1.505	2.727	0.042	0.012	0.0200	0.0050
	U_2	1.44	2.27	3.544	4.512	6.642	0.055	0.010		

Table 5.7: Model 3a) ($m = 2$, $p = 3$, H_0 with equal eigenfunctions): Asymptotic ($N = \infty$) and simulated values of $E[\tilde{U}_1]$, $E[\tilde{U}_2]$ (denoted by E), $var(\tilde{U}_1)$, $var(\tilde{U}_2)$ (denoted by V), and 90%−, 95%− and 99%−quantiles (Q_{90}, Q_{95}, Q_{99}) of $\tilde{U}_1$ and $\tilde{U}_2$ are given. Also given are simulated rejection probabilities ($\alpha(0.05)$, $\alpha(0.01)$) based on $\tilde{U}_1$ and $\tilde{U}_2$ respectively and on the combined test with a Bonferroni correction ($\alpha_{\text{Bonf}}(0.05)$, $\alpha_{\text{Bonf}}(0.01)$).

N		E	V	Q_{90}	Q_{95}	Q_{99}	$\alpha(0.05)$	$\alpha(0.01)$	$\alpha_{\mathrm{Bonf}}(0.05)$	$\alpha_{\mathrm{Bonf}}(0.01)$
∞	U_1	0.80	0.94	1.960	2.710	4.620	0.050	0.010		
	U_2	1.06	1.24	2.488	3.303	5.236	0.050	0.010		
$\epsilon_i(j)$: iid										
200	U_1	0.82	1.08	1.952	2.523	4.929	0.050	0.012	0.0225	0.0100
	U_2	1.08	1.09	2.562	3.262	4.882	0.048	0.008		
400	U_1	0.83	1.17	1.993	2.720	5.995	0.052	0.018	0.0275	0.0125
	U_2	1.09	1.17	2.438	3.184	5.127	0.048	0.010		
600	U_1	0.85	1.08	1.923	2.544	4.841	0.050	0.018	0.0325	0.0075
	U_2	1.05	1.21	2.480	3.242	4.921	0.045	0.005		
800	U_1	0.80	0.90	1.955	2.762	4.706	0.055	0.012	0.0250	0.0050
	U_2	1.13	1.35	2.620	3.567	4.722	0.062	0.005		
1000	U_1	0.79	0.83	2.022	2.798	4.174	0.058	0.005	0.0325	0.0075
	U_2	1.13	1.33	2.725	3.395	5.740	0.060	0.012		
$\epsilon_i(j)$: AR(1)										
200	U_1	0.88	1.15	2.050	2.640	4.337	0.048	0.010	0.0225	0.0100
	U_2	1.15	1.08	2.627	3.277	4.698	0.048	0.010		
400	U_1	0.88	1.19	2.046	2.807	5.951	0.055	0.018	0.0275	0.0125
	U_2	1.15	1.17	2.518	3.217	5.196	0.048	0.008		
600	U_1	0.87	1.04	1.904	2.556	5.042	0.048	0.020	0.0300	0.0075
	U_2	1.08	1.16	2.368	3.412	4.881	0.052	0.005		
800	U_1	0.82	0.89	1.952	2.796	4.632	0.052	0.012	0.0250	0.0050
	U_2	1.16	1.36	2.755	3.677	4.879	0.068	0.005		
1000	U_1	0.81	0.85	2.058	2.946	4.147	0.060	0.008	0.0300	0.0075
	U_2	1.16	1.33	2.686	3.441	5.799	0.062	0.012		
$\epsilon_i(j)$: FARIMA$(0, 0.3, 0)$										
200	U_1	0.86	1.07	1.954	2.608	4.384	0.042	0.010	0.0250	0.0100
	U_2	1.13	1.04	2.648	3.273	4.519	0.048	0.008		
400	U_1	0.90	1.24	2.029	2.799	5.630	0.062	0.018	0.0350	0.0125
400	U_2	1.18	1.31	2.498	3.378	5.523	0.058	0.015		
600	U_1	0.91	1.07	2.061	2.679	5.524	0.048	0.012	0.0325	0.0125
	U_2	1.12	1.21	2.501	3.256	4.881	0.048	0.005		
800	U_1	0.84	0.90	2.019	2.642	4.745	0.048	0.012	0.0275	0.0050
	U_2	1.18	1.41	2.829	3.648	4.983	0.072	0.005		
1000	U_1	0.82	0.80	2.094	2.802	4.037	0.058	0.002	0.0400	0.0100
	U_2	1.20	1.36	2.721	3.587	5.479	0.062	0.015		

Table 5.8: Model 3b) ($m = 2$, $p = 3$, H_0 with rotated eigenfunctions): Asymptotic ($N = \infty$)and simulated values of $E[\tilde{U}_1]$, $E[\tilde{U}_2]$ (denoted by E), $var(\tilde{U}_1)$, $var(\tilde{U}_2)$ (denoted by V), and 90%−, 95%− and 99%−quantiles (Q_{90}, Q_{95}, Q_{99}) of $\tilde{U}_1$ and $\tilde{U}_2$ are given. Also given are simulated rejection probabilities ($\alpha(0.05)$, $\alpha(0.01)$) based on $\tilde{U}_1$ and $\tilde{U}_2$ respectively and on the combined test with a Bonferroni correction ($\alpha_{\mathrm{Bonf}}(0.05)$, $\alpha_{\mathrm{Bonf}}(0.01)$).

N		E	V	Q_{90}	Q_{95}	Q_{99}	$\alpha(0.05)$	$\alpha(0.01)$	$\alpha_{\text{Bonf}}(0.05)$	$\alpha_{\text{Bonf}}(0.01)$
$\epsilon_i(j)$: iid										
200	U_1	1.70	3.62	3.765	4.943	9.670	0.342	0.198	1	1
	U_2	118.83	3.42	120.0	120.0	120.0	1.00	1.00		
400	U_1	1.61	2.61	3.479	4.906	7.177	0.358	0.182	1	1
	U_2	180.94	2.28	182.0	182.0	182.0	1.00	1.00		
600	U_1	1.62	2.70	3.625	4.787	7.680	0.375	0.185	1	1
	U_2	230.93	2.36	232.0	232.0	232.0	1.00	1.00		
800	U_1	1.61	2.78	3.521	4.737	7.542	0.368	0.188	1	1
	U_2	274.93	2.43	276.0	276.0	276.0	1.00	1.00		
1000	U_1	1.57	2.64	3.700	4.760	7.208	0.332	0.168	1	1
	U_2	314.45	2.37	315.5	315.5	315.5	1.00	1.00		
$\epsilon_i(j)$: AR(1)										
200	U_1	1.76	3.58	3.908	5.017	8.769	0.378	0.200	1	1
	U_2	118.82	3.37	120.0	120.0	120.0	1.00	1.00		
400	U_1	1.65	2.67	3.566	4.827	7.826	0.348	0.185	1	1
	U_2	180.93	2.37	182.0	182.0	182.0	1.00	1.00		
600	U_1	1.65	2.80	3.679	4.762	8.648	0.378	0.208	1	1
	U_2	230.92	2.47	232.0	232.0	232.0	1.00	1.00		
800	U_1	1.63	2.77	3.528	4.688	7.586	0.370	0.198	1	1
	U_2	274.93	2.41	276.0	276.0	276.0	1.00	1.00		
1000	U_1	1.59	2.65	3.662	4.770	7.304	0.342	0.182	1	1
	U_2	314.45	2.36	315.5	315.5	315.5	1.00	1.00		
$\epsilon_i(j)$: FARIMA$(0, 0.3, 0)$										
200	U_1	1.74	3.56	3.828	5.125	9.179	0.365	0.202	1	1
	U_2	118.80	3.34	120.0	120.0	120.0	1.00	1.00		
400	U_1	1.65	2.68	3.569	5.126	7.502	0.350	0.192	1	1
	U_2	180.93	2.36	182.0	182.0	182.0	1.00	1.00		
600	U_1	1.65	2.65	3.527	4.673	7.993	0.372	0.195	1	1
	U_2	230.94	2.33	232.0	232.0	232.0	1.00	1.00		
800	U_1	1.64	2.72	3.502	4.829	7.446	0.372	0.185	1	1
	U_2	274.93	2.38	276.0	276.0	276.0	1.00	1.00		
1000	U_1	1.60	2.50	3.754	4.848	7.011	0.350	0.182	1	1
	U_2	314.44	2.21	315.5	315.5	315.5	1.00	1.00		

Table 5.9: Model 3c) ($m = 2$, $p = 3$, H_1): Simulated values of $E[\tilde{U}_1]$, $E[\tilde{U}_2]$ (denoted by E), $var(\tilde{U}_1)$, $var(\tilde{U}_2)$ (denoted by V), and 90%−, 95%− and 99%−quantiles (Q_{90}, Q_{95}, Q_{99}) of $\tilde{U}_1$ and $\tilde{U}_2$ are given. Also given are simulated rejection probabilities ($\alpha(0.05)$, $\alpha(0.01)$) based on $\tilde{U}_1$ and $\tilde{U}_2$ respectively and on the combined test with a Bonferroni correction ($\alpha_{\text{Bonf}}(0.05)$, $\alpha_{\text{Bonf}}(0.01)$).

N		E	V	Q_{90}	Q_{95}	Q_{99}	$\alpha(0.05)$	$\alpha(0.01)$	$\alpha_{\text{Bonf}}(0.05)$	$\alpha_{\text{Bonf}}(0.01)$
∞	U_1	0.70	0.32	1.406	1.839	2.866	0.050	0.010		
	U_2	1.40	2.04	3.100	4.216	7.059	0.050	0.010		
$\epsilon_i(j)$: iid										
200	U_1	0.70	0.29	1.406	1.832	2.748	0.045	0.008	0.0450	0.0100
	U_2	1.73	2.66	3.719	5.301	8.071	0.072	0.020		
400	U_1	0.68	0.26	1.275	1.653	2.409	0.035	0.008	0.0475	0.0125
	U_2	1.72	2.85	3.331	5.231	8.762	0.068	0.015		
600	U_1	0.67	0.27	1.306	1.758	2.352	0.038	0.002	0.0375	0.0100
	U_2	1.71	2.63	3.468	4.527	7.801	0.060	0.018		
800	U_1	0.70	0.34	1.329	1.734	2.933	0.045	0.012	0.0350	0.0125
	U_2	1.71	2.66	3.278	4.695	8.815	0.065	0.018		
1000	U_1	0.70	0.34	1.331	1.736	2.989	0.042	0.015	0.0400	0.0125
	U_2	1.73	2.56	3.320	4.739	8.525	0.065	0.022		
$\epsilon_i(j)$: AR(1)										
200	U_1	0.74	0.29	1.448	1.887	2.718	0.058	0.010	0.0525	0.0175
	U_2	1.86	2.98	3.658	5.474	8.539	0.075	0.028		
400	U_1	0.71	0.28	1.341	1.684	2.535	0.038	0.008	0.0450	0.0125
	U_2	1.77	2.79	3.438	4.967	8.339	0.068	0.018		
600	U_1	0.70	0.27	1.357	1.782	2.268	0.048	0.002	0.0375	0.0100
	U_2	1.77	2.77	3.560	4.494	8.185	0.060	0.020		
800	U_1	0.72	0.34	1.382	1.808	3.018	0.048	0.012	0.0350	0.0125
	U_2	1.74	2.65	3.334	4.837	9.127	0.068	0.018		
1000	U_1	0.71	0.33	1.344	1.799	2.884	0.048	0.012	0.0375	0.0125
	U_2	1.77	2.57	3.366	4.707	8.458	0.070	0.022		
$\epsilon_i(j)$: FARIMA$(0, 0.3, 0)$										
200	U_1	0.73	0.28	1.440	1.765	2.737	0.045	0.008	0.0500	0.0100
	U_2	1.84	2.77	3.626	5.333	8.264	0.072	0.025		
400	U_1	0.72	0.26	1.368	1.654	2.392	0.040	0.008	0.0550	0.0125
	U_2	1.81	3.15	3.597	5.498	8.384	0.068	0.025		
600	U_1	0.71	0.26	1.350	1.801	2.201	0.048	0.002	0.0350	0.0100
	U_2	1.77	2.73	3.504	4.250	8.102	0.052	0.022		
800	U_1	0.74	0.34	1.387	1.783	3.019	0.045	0.015	0.0325	0.0125
	U_2	1.75	2.63	3.384	4.714	9.175	0.070	0.015		
1000	U_1	0.73	0.34	1.348	1.823	3.045	0.050	0.015	0.0375	0.0150
	U_2	1.78	2.58	3.534	4.704	8.522	0.065	0.025		

Table 5.10: Model 4a) ($m = 2$, $p = 10$, H_0 with equal eigenfunctions): Asymptotic ($N = \infty$) and simulated values of $E[\tilde{U}_1]$, $E[\tilde{U}_2]$ (denoted by E), $var(\tilde{U}_1)$, $var(\tilde{U}_2)$ (denoted by V), and 90%−, 95%− and 99%−quantiles (Q_{90}, Q_{95}, Q_{99}) of $\tilde{U}_1$ and $\tilde{U}_2$ are given. Also given are simulated rejection probabilities ($\alpha(0.05)$, $\alpha(0.01)$) based on $\tilde{U}_1$ and $\tilde{U}_2$ respectively and on the combined test with a Bonferroni correction ($\alpha_{\text{Bonf}}(0.05)$, $\alpha_{\text{Bonf}}(0.01)$).

N		E	V	Q_{90}	Q_{95}	Q_{99}	$\alpha(0.05)$	$\alpha(0.01)$	$\alpha_{\text{Bonf}}(0.05)$	$\alpha_{\text{Bonf}}(0.01)$
∞	U_1	1.04	0.93	2.238	2.983	4.706	0.050	0.010		
	U_2	1.06	1.01	2.277	3.059	5.010	0.050	0.010		
$\epsilon_i(j)$: iid										
200	U_1	1.09	1.19	2.249	3.097	5.114	0.058	0.018	0.0350	0.0100
	U_2	1.34	1.16	2.664	3.396	5.715	0.062	0.020		
400	U_1	1.02	1.20	2.007	2.756	5.079	0.042	0.020	0.0425	0.0075
	U_2	1.38	1.31	2.904	3.558	5.558	0.072	0.022		
600	U_1	1.03	1.12	2.182	2.864	4.397	0.048	0.005	0.0400	0.0125
	U_2	1.35	1.28	2.782	3.456	6.126	0.082	0.018		
800	U_1	1.06	1.15	2.120	3.070	4.897	0.062	0.018	0.0500	0.0100
	U_2	1.35	1.31	2.651	3.859	5.782	0.080	0.022		
1000	U_1	1.06	1.10	2.345	3.137	5.409	0.058	0.015	0.0475	0.0100
	U_2	1.37	1.24	2.788	3.769	5.186	0.085	0.015		
$\epsilon_i(j)$: AR(1)										
200	U_1	1.16	1.24	2.339	3.201	5.248	0.070	0.022	0.0425	0.0125
	U_2	1.44	1.28	2.769	3.537	6.039	0.075	0.022		
400	U_1	1.05	1.12	2.063	2.894	5.176	0.050	0.018	0.0450	0.0125
	U_2	1.43	1.33	2.921	3.438	5.954	0.082	0.025		
600	U_1	1.07	1.20	2.197	3.010	4.421	0.052	0.008	0.0450	0.0150
	U_2	1.40	1.34	2.837	3.688	6.254	0.080	0.020		
800	U_1	1.08	1.14	2.107	3.158	4.962	0.065	0.020	0.0550	0.0125
	U_2	1.38	1.28	2.680	3.941	5.963	0.080	0.022		
1000	U_1	1.09	1.09	2.426	3.171	5.163	0.058	0.015	0.0475	0.0100
	U_2	1.40	1.26	2.763	3.797	5.579	0.082	0.015		
$\epsilon_i(j)$: FARIMA$(0, 0.3, 0)$										
200	U_1	1.12	1.19	2.256	3.164	5.650	0.065	0.018	0.0400	0.0125
	U_2	1.42	1.20	2.618	3.469	5.715	0.072	0.020		
400	U_1	1.07	1.30	2.095	2.888	5.288	0.048	0.018	0.0475	0.0100
	U_2	1.44	1.35	2.962	3.672	5.622	0.090	0.025		
600	U_1	1.07	1.09	2.164	2.958	4.599	0.048	0.010	0.0400	0.0125
	U_2	1.44	1.38	3.087	3.820	6.025	0.102	0.025		
800	U_1	1.09	1.11	2.262	3.122	4.958	0.058	0.022	0.0475	0.0150
	U_2	1.40	1.32	2.706	3.768	6.084	0.085	0.020		
1000	U_1	1.10	1.09	2.482	3.186	5.196	0.065	0.015	0.0475	0.0100
	U_2	1.44	1.32	3.009	3.854	5.292	0.098	0.018		

Table 5.11: Model 4b) ($m = 2$, $p = 10$, H_0 with rotated eigenfunctions): Asymptotic ($N = \infty$) and simulated values of $E[\tilde{U}_1]$, $E[\tilde{U}_2]$ (denoted by E), $var(\tilde{U}_1)$, $var(\tilde{U}_2)$ (denoted by V), and 90%−, 95%− and 99%−quantiles (Q_{90}, Q_{95}, Q_{99}) of $\tilde{U}_1$ and $\tilde{U}_2$ are given. Also given are simulated rejection probabilities ($\alpha(0.05)$, $\alpha(0.01)$) based on $\tilde{U}_1$ and $\tilde{U}_2$ and on the combined test with a Bonferroni correction ($\alpha_{\text{Bonf}}(0.05)$, $\alpha_{\text{Bonf}}(0.01)$).

N		E	V	Q_{90}	Q_{95}	Q_{99}	$\alpha(0.05)$	$\alpha(0.01)$	$\alpha_{\text{Bonf}}(0.05)$	$\alpha_{\text{Bonf}}(0.01)$
$\epsilon_i(j)$: iid										
200	U_1	1.85	3.63	3.954	5.034	10.062	0.338	0.172	1	1
	U_2	118.84	3.44	120.0	120.0	120.0	1.00	1.00		
400	U_1	1.73	2.54	3.618	4.720	8.878	0.300	0.170	1	1
	U_2	180.94	2.26	182.0	182.0	182.0	1.00	1.00		
600	U_1	1.74	2.63	3.631	4.820	7.972	0.350	0.160	1	1
	U_2	230.93	2.33	232.0	232.0	232.0	1.00	1.00		
800	U_1	1.76	2.71	3.753	4.618	7.789	0.335	0.175	1	1
	U_2	274.94	2.37	276.0	276.0	276.0	1.00	1.00		
1000	U_1	1.74	2.65	3.839	5.035	6.539	0.335	0.162	1	1
	U_2	314.46	2.31	315.5	315.5	315.5	1.00	1.00		
$\epsilon_i(j)$: AR(1)										
200	U_1	1.90	3.59	4.235	5.144	9.447	0.345	0.190	1	1
	U_2	118.82	3.39	120.0	120.0	120.0	1.00	1.00		
400	U_1	1.77	2.66	3.649	4.814	9.386	0.305	0.185	1	1
	U_2	180.93	2.34	182.0	182.0	182.0	1.00	1.00		
600	U_1	1.78	2.75	3.711	4.996	8.733	0.345	0.170	1	1
	U_2	230.92	2.44	232.0	232.0	232.0	1.00	1.00		
800	U_1	1.78	2.70	3.759	4.596	7.771	0.335	0.192	1	1
	U_2	274.94	2.34	276.0	276.0	276.0	1.00	1.00		
1000	U_1	1.76	2.64	3.857	5.152	6.479	0.332	0.172	1	1
	U_2	314.46	2.30	315.5	315.5	315.5	1.00	1.00		
$\epsilon_i(j)$: FARIMA$(0, 0.3, 0)$										
200	U_1	1.90	3.58	4.097	5.303	9.203	0.342	0.182	1	1
	U_2	118.81	3.36	120.0	120.0	120.0	1.00	1.00		
400	U_1	1.76	2.62	3.667	4.753	8.032	0.298	0.172	1	1
	U_2	180.94	2.34	182.0	182.0	182.0	1.00	1.00		
600	U_1	1.75	2.63	3.633	4.908	8.014	0.350	0.158	1	1
	U_2	230.94	2.31	232.0	232.0	232.0	1.00	1.00		
800	U_1	1.79	2.67	3.770	4.808	8.355	0.335	0.182	1	1
	U_2	274.94	2.31	276.0	276.0	276.0	1.00	1.00		
1000	U_1	1.77	2.53	3.906	5.044	6.766	0.338	0.168	1	1
	U_2	314.45	2.16	315.5	315.5	315.5	1.00	1.00		

Table 5.12: Model 4c) ($m = 2$, $p = 10$, H_1): Simulated values of $E[\tilde{U}_1]$, $E[\tilde{U}_2]$ (denoted by E), $var(\tilde{U}_1)$, $var(\tilde{U}_2)$ (denoted by V), and 90%−, 95%− and 99%−quantiles (Q_{90}, Q_{95}, Q_{99}) of $\tilde{U}_1$ and $\tilde{U}_2$ are given. Also given are simulated rejection probabilities ($\alpha(0.05)$, $\alpha(0.01)$) based on $\tilde{U}_1$ and $\tilde{U}_2$ respectively and on the combined test with a Bonferroni correction ($\alpha_{\text{Bonf}}(0.05)$, $\alpha_{\text{Bonf}}(0.01)$).

N	Model 1	Model 2	Model 3	Model 4	Model 5
50	0.07	0.81	0.06	0.99	0.45
	0.06	0.79	0.05	1	0.35
	0.05	0.79	0.04	1	0.75
100	0.08	0.92	0.11	1	0.82
	0.09	0.92	0.06	1	0.83
	0.09	0.92	0.10	1	0.84
200	0.03	0.97	0.06	1	0.95
	0.03	0.98	0.06	1	0.96
	0.04	0.97	0.07	1	0.96

Table 5.13: Simulated rejection probabilities of the bootstrap test based on $T_{n,N}$ at the nominal level of significance $\alpha = 0.05$. For each value of $N = 50$, 100 and 200, $N_{simul} = 100$ simulations were carried out. For each simulation, the emipirical quantiles $q^*_{0.95}(1)$, $q^*_{0.95}(2)$ were obtained from $T^*_{s,1}$ $(s = 1, ..., N_{boot})$ and $T^*_{s,2}$ $(s = 1, ..., N_{boot})$ respectively, with $N_{boot} = 100$. For each value of N and each model, the results are listed in the following sequence from top to bottom: (a) iid error process, (b) AR(1), (c) FARIMA$(0, d, 0)$.

N	Model 1	Model 2	Model 3	Model 4	Model 5
50	0.09	0.75	0.06	1	0.89
	0.08	0.76	0.06	0.99	0.91
	0.05	0.74	0.06	1	0.91
100	0.04	0.87	0.05	1	0.95
	0.04	0.89	0.02	1	0.95
	0.06	0.89	0.05	1	0.94
200	0.03	1	0.08	1	1
	0.03	1	0.08	1	1
	0.03	1	0.05	1	1

Table 5.14: Simulated rejection probabilities of the bootstrap test based on $D_{4,m}$ (Benko et al. 2009) at the nominal level of significance $\alpha = 0.05$. For each value of $N = 50$, 100 and 200, $N_{simul} = 100$ simulations were carried out. For each value of N and each model, the results are listed in the following sequence from top to bottom: (a) iid error process, (b) AR(1), (c) FARIMA$(0, d, 0)$.

Chapter 6

Concluding remarks

In this thesis we investigated statistical inference of functional data analysis (FDA) model with long memory random errors. The estimation methods of trend function $\mu(t)$, covariance function $C(s,t)$, eigenvalues λ_l, eigenfunctions (functional principal components) $\phi_l(t)$ and functional principal component scores ξ_{il} have been discussed. Two sample inference has also been discussed and a test statistic for testing the equality of eigenspaces has been constructed. All methods have been theoretically discussed and illustrated with simulation examples. Actually, the usefulness of these methods goes far beyond a purely mathematical device. Finally, we give some concluding remarks and mention several open problems in FDA which have been discussed in this thesis.

- **Nonequidistant or Random Designs:** In Chapter 3, we considered estimation of trend function $\mu(t)$ and covariance function $C(s,t)$ in repeated time series with long memory noise errors within a FDA model. In order to focus on the essential effect of dependence in error process, only the equidistant case has been discussed. It is well known that, for the case of single time series, there is no functional central limit theorem. However, if the number of repeated time series n does not grow too fast compared to the number of sampling points of each time series N, a functional central limit theorem can be derived for kernel estimation of $\mu(t)$. Obviously, this is an unpleasant condition. If the estimation of $\mu(t)$ is the main aim, one may use higher order kernel to relax the restriction provided that higher order derivatives of

$\mu(t)$ exist. Since the main quantity of interest in FDA is $C(s,t)$, it is wise to eliminate $\mu(t)$ before estimating $C(s,t)$. Therefore, we propose to remove $\mu(t)$ by using an orthonormal contrast transformation. Under the equidistant and Gaussian assumption, the contrast model is equivalent in distribution to the original model with $\mu(t) = 0$, except that n reduces to $n-1$. The functional central limit theorem of higher order kernel estimator of $C(s,t)$ based on the contrast model is available with a reasonable upper bound of n. Contrasts are not directly applicable for nonequidistant or random designed model. One of the questions to be discussed in future is how to extend these results to nonequidistant or random designed case.

- **Principal Component Scores with Strongly Dependent Errors:** In Chapter 4, the estimation of eigenvalues λ_l, eigenfunctions (functional principal components) $\phi_l(t)$ and functional principal component scores ξ_{il} in a FDA model perturbed by error processes (short- or long-range dependent) has been discussed. Based on the kernel estimation of $C(s,t)$ that we defined in Chapter 3, the asymptotic distribution of the estimated eigenvalues $\hat{\lambda}_l$ does not depend on the dependence structure of the error processes. Moreover, the asymptotic distribution of estimated eigenfunctions $\hat{\phi}_l(t)$ and the joint distribution of them also do not depend on the dependence structure of the error processes. But the different eigenfunction estimators $\hat{\phi}_1(t), \hat{\phi}_2(t), ...$ are no longer independent. However, this is not the case for the estimated functional principal component scores $\hat{\xi}_{il}$. The rates of convergence of $\hat{\xi}_{il}$ are different for the cases of short- and long-range dependent error processes. Moreover, there is no independent property of $\hat{\xi}_{i1}, \hat{\xi}_{i2}, ...$ The strength of the dependence is a function of $\phi_l(t)$ and the long memory parameter d. Therefore, new statistical inference methods for ξ_{il} should be established to incorporate the possibility of strongly dependent error processes.

- **Determine the Dimension of Eigenspace:** In Chapter 5, two sample inference for eigenspaces $\mathcal{U}$ and $\mathcal{V}$ in FDA with dependent errors was discussed and a test for testing the equality of $\mathcal{U}$ and $\mathcal{V}$ was developed. Note that the individual eigenfunctions or eigenvalues are not required to be identical. The test is applicable to FDA models with an error process that may exhibit weakly or strongly dependence. The standardized residual process $\tilde{r}_l(t)$ converges weakly to a Gaussian process with a relatively simple structure. Note that the asymptotic covariance structure of $\tilde{r}_l(t)$ depends on the eigenfunctions (and eigenvalues) spanning $\mathcal{U}$ and $\mathcal{V}$. Moreover, it also depends on the remaining $p - m$ eigenfunctions in the representation of the FDA component $X_i^{(1)}(t)$ and $X_i^{(2)}(t)$. In the worst case, $p - m$ may be even infinite. Therefore, from the practical point of view, determination of the dimension of eigenspace m is very important. This leads to the question of how good approximations based on a few estimated eigenfunctions are and how to design data driven algorithms for choosing a suitable approximation.

- **High Dimensional Model:** Image data and spatial data play an important role in many areas. These observations are obtained on high dimensional lattices not on one dimensional points. We are working on this high dimensional model. So how to extend the methods in this thesis to high dimensional model will be considered.

Bibliography

[1] Antoniadis, A. and Oppenheim, G. (2012). *Wavelets and statistics*. Springer.

[2] Benko, M., Härdle, W. and Kneip, A. (2009). Common functional principal components. *The Annals of Statistics*, **37(1)**, 1-34.

[3] Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman and Hall/CRC Press.

[4] Beran, J. (2010). Long-range dependence. *In: Wiley Interdisciplinary Reviews: Computational Statistics, Edward J. Wegman, Yasmin H. Said, and David W. Scott*, **2(1)**, 26-35.

[5] Beran, J. and Feng, Y. (2002a). SEMIFAR models–a semiparametric framework for modelling trends, long-range dependence and nonstationarity. *Computational Statistics and Data Analysis*, **40(2)**, 393-419.

[6] Beran, J. and Feng, Y. (2002b). Iterative plug-in algorithms for SEMIFAR models–definition, convergence, and asymptotic properties. *Journal of Computational and Graphical Statistics*, **11(3)**, 690-713.

[7] Beran, J. and Feng, Y. (2002c). Local polynomial fitting with long memory, short memory and antipersistent errors. *Annals of the Institute of Statistical Mathematics*, **54(2)**, 291-311.

[8] Beran, J., Feng, Y., Ghosh, S. and Kulik, R. (2013). *Long-Memory Processes: Probabilistic Properties and Statistical Methods*. Springer.

[9] Beran, J. and Liu, H. (2014). On estimation of mean and covariance functions in repeated time series with long-memory errors. *Lithuanian Mathematical Journal*, **54(1)**, 8-34.

[10] Beran, J. and Liu, H. (2016). Estimation of eigenvalues, eignevectors and scores in FDA models with strongly dependent errors. *Journal of Multivariate Analysis*, **147**, 218-233.

[11] Beran, J., Liu, H. and Telkmann, K.(2016). On two sample inference for eigenspaces in FDA with dependent errors. *Journal of Statistical Planning and Inference*, **174**, 20-37.

[12] Beran, J. and Shumeyko, Y. (2012). On asymptotically optimal wavelet estimation of trend functions under long-range dependence. *Bernoulli*, **18(1)**, 137-176.

[13] Bevilacqua, M., Gaetan, C., Mateu, J. and Porcu, E. (2012). Estimating space and space-time covariance functions for large data sets: a weighted composite likelihood approach. *Journal of the American Statistical Association*, **107(497)**, 268-280.

[14] Bigot, J., Biscay, R.J., Loubes, J.M. and Muniz-Alvarez, L. (2010). Nonparametric estimation of covariance functions by model selection. *Electronic Journal of Statistics*, **4**, 822-855.

[15] Bigot, J., Biscay, R.J., Loubes, J.M. and Muniz-Alvarez, L. (2011). Group lasso estimation of high-dimensional covariance matrices. *The Journal of Machine Learning Research*, **12**, 3187-3225.

[16] Billingsley, P. (1999). *Convergence of probability measures*. Wiley.

[17] Boente, G. and Fraiman, R. (2000). Kernel-based functional principal components. *Statistics and Probability Letters*, **48(4)**, 335-345.

[18] Boente, G., Rodriguez, D. and Sued, M. (2011). Testing the equality of covariance operators. *Recent advances in functional data analysis and related topics*, Physica-Verlag HD, 49-53.

[19] de Boor, C. (2001). *A Practical Guide to Splines (Revised Edition)* Springer.

[20] Bornas, X., Noguera, M., Balle, M., Morillas-Romero, A. Aguayo-Siquier, B., Tortella-Feliu, M. and Llabrés, J. (2013). Long-range temporal correlations in resting EEG: Its associations with depression-related emotion regulation strategies. *Journal of Psychophysiology*, **27(2)**, 60.

[21] Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications.* Springer.

[22] Brockwell, P.J. and Davis, R.A. (1991). *Time series: theory and methods (2nd Edition).* Springer.

[23] Cai, T.T. and Yuan, M. (2010). Nonparametric covariance function estimation for functional and longitudinal data. *University of Pennsylvania and Georgia inistitute of technology.*

[24] Chui, C.K. (1992). *An Introduction to Wavelets.* Academic press.

[25] Clarkson, D.B., Fraley, C., Gu, C. and Ramsay J.O. (2005). *S+Functional Data Analysis.* Springer.

[26] Csörgö, S. and Mielniczuk, J. (1995). Nonparametric regression under long-range dependent normal errors. *The Annals of Statistics*, **23(3)**, 1000-1014.

[27] Cuevas, A., Febrero, M. and Fraiman, R. (2004). An anova test for functional data. *Computational Statistics and Data Analysis*, **47(1)**, 111-122.

[28] Daubechies, I. (1992). *Ten Lectures on Wavelets.* CBMS-NSF Regional Conference Series in Applied Mathematics. Philadelphia: Society for industrial and applied mathematics.

[29] Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *Journal of Multivariate Analysis*, **12(1)**, 136-154.

[30] Delaigle, A. and Hall, P. (2010). Defining probability density for a distribution of random functions. *The Annals of Statistics*, **38(2)**, 1171-1193.

[31] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society: Series B*, **57(2)**, 301-369.

[32] Doukhan, P., Oppenheim, G. and Taqqu, M.S. (2003). *Theory and Applications of Long-Range Dependence*. Springer.

[33] Dunford, N. and Schwartz, J.T. (1988). *Linear operators. Part I. General theory*. Wiley-interscience, New York

[34] Eubank, R.L.(1999). *Nonparametric Regression and Spline smoothing*. Chapman and Hall/CRC Press.

[35] Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, **87(420)**, 998-1004.

[36] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman and Hall/CRC Press.

[37] Fan, J., Liao, Y. and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B*, **75(4)**, 603-680.

[38] Feng, Y. (1999). *Kernel- and locally weighted regression–with application to time series decomposition*. Verlag für Wissenschaft und Forschung.

[39] Feng, Y. (2004). Simultaneously modelling conditional heteroskedasticity and scale change. *Econometric Theory*, **20(3)**, 563-596.

[40] Ferraty, F. (2011). *Recent Advances in Functional Data Analysis and Related Topics*. Springer.

[41] Ferraty, F. and Romain, Y. (2011). *The Oxford Handbook of Functional Data Analysis*. Oxford University Press.

[42] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer.

[43] Ferraty, F., Vieu, P. and Viguier-Pla, S. (2007). Factor-based comparison of groups of curves. *Computational statistics and data analysis*, **51(10)**, 4903-4910.

[44] Fremdt, S., Steinebach, J.G., Horváth, L. and Kokoszka, P. (2013). Testing the equality of covariance operators in functional samples. *Scandinavian Journal of Statistics*, **40(1)**, 138-152.

[45] Gabrys, R., Horváth, L. and Kokoszka, P. (2010). Tests for error correlation in the functional linear model. *Journal of the American Statistical Association*, **105(491)**, 1113-1125.

[46] Gasser, T. and Müller, H.G. (1979). *Kernel Estimation of Regression Functions*. Springer Berlin Heidelberg.

[47] Gasser, T., Müller, H.G. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society: Series B*, **47(2)**, 238-252.

[48] Gervini, D. (2008). Robust functional estimation using the median and spherical principal components. *Biometrika*, **95(3)**, 587-600.

[49] Ghosh, S. and Beran, J. (2000). Two-sample T3 plot: a graphical comparison of two distributions. *Journal of Computational and Graphical Statistics*, **9(1)**, 167-179.

[50] Ghosh, S. (2001). Nonparametric trend estimation in replicated time series. *Journal of Statistical Planning and Inference*, **97(2)**, 263-274.

[51] Giraitis, L., Koul, H.L. and Surgailis, D. (2012). *Large sample inference for long memory processes*. Imperial College Press.

[52] Granger, C.W.J. and Joyeux, R. (1980). An introduction to long-range time series models and fractional differencing. *Journal of Time Series Analysis*, **1(1)**, 15-30.

[53] Green, P.J. and Silverman, B.W. (1993). *Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach*. Chapman and Hall/CRC Press.

[54] Gromenko, O. and Kokoszka, P. (2012). Testing the equality of mean functions of ionospheric critical frequency curves. *Journal of the Royal Statistical Society: Series C*, **61(5)**, 715-731.

[55] Hall, P., Fisher, N.I. and Hoffmann, B. (1994). On the nonparametric estimation of covariance functions. *The Annals of Statistics*, **22(4)**, 2115-2134.

[56] Hall, P. and Hart, J.D. (1990). Nonparametric regression with long-range dependence. *Stochastic Processes and their Applications*, **36(2)**, 339-351.

[57] Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B*, **68(1)**, 109-126.

[58] Hall, P. and Hosseini-Nasab, M. (2009). Theory for high-order bounds in functional principal components analysis. *Mathematical Proceedings of the Cambridge Philosophical Society*, **146(1)**, 225-256.

[59] Hall, P., Müller, H.G. and Wang, J.L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics*, **34(3)**, 1493-1517.

[60] Hall, P. and van Keilegom, I. (2007). Two-sample tests in functional data analysis starting from discrete data. *Statistica Sinica*, **17(4)**, 1511-1531.

[61] Härdle, W. and Vieu, P. (1992). Kernel regression smoothing of time series. *Journal of Time Series Analysis*, **13(3)**, 209-232.

[62] Hart, J.D. and Wehrly, T.E. (1986). Kernel regression estimation using repeated measurements data. *Journal of the American Statistical Association*, **81(396)**, 1080-1088.

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

[64] Horváth, L., Kokoszka, P. and Reeder, R. (2013). Estimation of the mean of functional time series and a two sample problem. *Journal of the Royal Statistical Society: Series B*, **75(1)**, 103-122.

[65] Horváth, L., Kokoszka, P. and Reimherr, M. (2009). Two sample inference in functional linear models. *Canadian Journal of Statistics*, **37(4)**, 571-591.

[66] Hosking, J.R.M. (1981). Fractional differencing. *Biometrika*, **68(1)**, 165-176.

[67] John, R. (1984). Boundary modification for kernel regression. *Communications in Statistics Theory and Methods*, **13(7)**, 893-900.

[68] Johnstone, I.M. and Silverman, B.W. (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society: Series B*, **59(2)**, 319-351.

[69] Johnstone, I.M. and Lu, A.Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, **104(486)**, 682-693.

[70] Kato, T. (1976). *Perturbation theory for linear operators*. Springer.

[71] Lejeune, M. (1985). Estimation non-paramétrique par noyaux: régression polynômiale mobile. *Revue de Statistiques Appliquées*, **33**, 43-68.

[72] Lejeune, M. and Sarda, P. (1992). Smooth estimators of distribution and density functions. *Computational Statistics and Data Analysis*, **14(4)**, 457-471.

[73] Lin, X. and Carroll, R.J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, **95(450)**, 520-534.

[74] Linkenkaer-Hansen, K., Nikouline, V.V., Palva, J.M. and Ilmoniemi, R.J. (2001). Long-range temporal correlations and scaling behavior in human brain oscillations. *The Journal of Neuroscience*, **21(4)**, 1370-1377.

[75] Lirio, R.B., Camejo, D.G., Loubes, J.M. and Alvarez, L.M. (2014). Estimation of covariance functions by a fully data-driven model selection procedure and its application to Kriging spatial interpolation of real rainfall data. *Statistical Methods and Applications*, **23(2)**, 149-174.

[76] Loader, C. (2012). Smoothing: local regression techniques. *In Handbook of Computational Statistics*, 571-596.

[77] Ma, Z. (2013). Sparse principal component analysis and iterative thresholding. *The Annals of Statistics*, **41(2)**, 772-801.

[78] Menéndez, P., Ghosh, S. and Beran, J. (2010). On rapid change points under long memory. *Journal of Statistical Planning and Inference*, **140(11)**, 3343-3354.

[79] Menéndez, P., Ghosh, S., Künsch, H. and Tinner, W. (2013). On trend estimation under monotone Gaussian subordination with long-memory: application to fossil pollen series. *Journal of Nonparametric Statistics*, **25(4)**, 765-785.

[80] Müller, H.G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *Journal of the American Statistical Association*, **82(397)**, 231-238.

[81] Müller, H.G. (1988). *Nonparametric analysis of longitudinal data*. Springer.

[82] Müller, H.G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika*, **78(3)**, 521-530.

[83] Müller, H.G. and Wang, J.L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics*, **50(1)**, 61-76.

[84] Nikulin, V.V. and Brismar, T. (2005). Long-range temporal correlations in electroencephalographic oscillations: Relation to topography, frequency band, age and gender. *Neuroscience*, **130(2)**, 549-558.

[85] Ocana, F.A., Aguilera, A.M. and Valderrama, M.J. (1999). Functional principal components analysis by choice of norm. *Journal of Multivariate Analysis*, **71(2)**, 262-276.

[86] Ogden, T. (2012). *Essential wavelets for statistical applications and data analysis*. Springer Science and Business Media.

[87] Panaretos, V.M., Kraus, D. and Maddocks, J.H. (2010). Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *Journal of the American Statistical Association*, **105(490)**, 670-682.

[88] Parish, L.M., Worrell, G.A., Cranstoun, S.D., Stead, S.M., Pennell,P. and Litt, B. (2004). Long-range temporal correlations in epileptogenic and non-epileptogenic human hippocampus. *Neuroscience*, **125(4)**, 1069-1076.

[89] Percival, D.B. and Walden, A.T. (2006). *Wavelet methods for time series analysis*. Cambridge University Press.

[90] Ramsay, J.O. and Ramsey, J.B. (2002). Functional data analysis of the dynamics of the monthly index of nondurable goods production. *Journal of Econometrics*, **107(1)**, 327-344.

[91] Ramsay, J.O. and Silverman, B.W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer.

[92] Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis (Second Edition)*. Springer.

[93] Ray, B.K. and Tsay, R.S. (1997). Bandwidth selection for kernel regression with long-range dependent errors. *Biometrika*, **84(4)**, 791-802.

[94] Reiss, P.T. and Ogden, R.T. (2007). Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association*, **102(479)**, 984-996.

[95] Robinson, P.M. (1997). Large sample inference for nonparametric regression with dependent errors. *The Annals of Statistics*, **25(5)**, 2054-2083.

[96] Robinson, P.M. (2003). *Time Series with Long Memory*. Oxford University Press.

[97] Ruppert, D. and Wand, M.P. (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics*, **22(3)**, 1346-1370.

[98] Sancetta, A. (2014). A Nonparametric estimator for the covariance function of functional data. *Econometric Theory*, 1-23.

[99] Shaby, B. and Ruppert, D. (2012). Tapered covariance: Bayesian estimation and asymptotics. *Journal of Computational and Graphical Statistics*, **21(2)**, 433-452.

[100] Shi, J.Q. and Choi, T. (2011). *Gaussian Process Regression Analysis for Functional Data*. CRC Press.

[101] Staniswalis, J.G. and Lee, J.J. (1998). Nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association*, **93(444)**, 1403-1418.

[102] Wahba, G. (1990). *Spline Models for Observational Data*. Siam.

[103] Wand, M. P. and Jones, M. C. (1994). *Kernel Smoothing*. Chapman and Hall/CRC Press.

[104] Wang, Y. (2011). *Smoothing Splines: Methods and Applications*. Chapman and Hall/CRC Press.

[105] Watters, P.A. (2000). Time-invariant long-range correlations in electroencephalogram dynamics. *International Journal of Systems Science*, **31(7)**, 819-825.

[106] Yang, W., Müller, H.G. and Stadtmüller, U. (2011). Functional singular component analysis. *Journal of the Royal Statistical Society: Series B*, **73(3)**, 303-324.

[107] Yao, F. (2007). Asymptotic distributions of nonparametric regression estimators for longitudinal or functional data. *Journal of Multivariate Analysis*, **98(1)**, 40-56.

[108] Yao, F., Müller, H.G., Clifford, A.J., Dueker, S.R., Follett, J., Lin, L., Buchholz, B.A. and Vogel, J.S. (2003). Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. *Biometrics*, **59(3)**, 676-685.

[109] Yao, F., Müller, H.G. and Wang, J.L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, **100(470)**, 577-590.

[110] Zhang, J.T. and Chen, J. (2007). Statistical inferences for functional data. *The Annals of Statistics*, **35(3)**, 1052-1079.