DISSERTATION

TEST OF CHANGE POINT VERSUS LONG-RANGE DEPENDENCE IN FUNCTIONAL TIME SERIES

Submitted by

Xiangdong Meng

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2024

Doctoral Committee:

Advisor: Piotr S. Kokoszka

Dan Cooley
Haonan Wang
Hong Miao

ABSTRACT

TEST OF CHANGE POINT VERSUS LONG-RANGE DEPENDENCE IN FUNCTIONAL TIME SERIES

In scalar time series analysis, a long-range dependent (LRD) series cannot be easily distinguished from certain non-stationary models, such as the change in mean model with short-range dependent (SRD) errors. To be specific, realizations of LRD series usually have a characteristic of changing local mean if the time span taken into account is long enough, which resembles the behavior of change in mean models. Test procedure for distinguishing between these two types of model has been investigated a lot in scalar case, see e.g. Berkes *et al.* (2006) and Baek and Pipiras (2012) and references therein. However, no analogous test for functional observations has been developed yet, partly because of omitted methods and theory for analyzing functional time series with long-range dependence. My dissertation establishes a procedure for testing change in mean models with SRD errors against LRD processes in functional case, which is an extension of the method of Baek and Pipiras (2012). The test builds on the local Whittle (LW) (or Gaussian semiparametric) estimation of the self-similarity parameter, which is based on the estimated level 1 scores of a suitable functional residual process. Remarkably, unlike other parametric methods such as Whittle estimation, whose asymptotic properties heavily depend on validity of the underlying spectral density on the full frequency range $(-\pi, \pi]$, LW estimation imposes mild restrictions on the spectral density only near the origin and is thus more robust to model misspecification. We shall prove that the test statistic based on LW estimation is asymptotically normally distributed under the null hypothesis and it diverges to infinity under the LRD alternative.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my esteemed advisor, Professor Piotr Kokoszka, for his illuminating guidance throughout my arduous journey toward obtaining my Ph.D. degree and securing employment. I thank him for teaching me mathematical statistics and probability theory and patiently answering my questions along the way, which laid a solid foundation for my forthcoming research. He inspired me to break various conundrums into doable parts, without which I could not make any breakthrough at my early stages. He also motivated me to focus on big picture without indulging in minutiae, thereby enabling me to stay on track. His enthusiasm and passion for research encouraged me to move forward and finally became a confident researcher.

I would like to extend my sincere thanks to my committee members: Professor Daniel Cooley, Professor Haonan Wang and Professor Hong Miao. Their insightful feedback and invaluable suggestions were conducive for enhancing and refining this paper. I express special thanks to Professor Haonan Wang, whose generous support during summer terms was instrumental in advancing my academic pursuits. His guidance, encompassing coursework, research, and life, has been immeasurably valuable.

I very much appreciate Professor Changryong Baek for his expertise in local Whittle estimation and his prominent contributions to the collaborative project. I appreciate his significant assistance in acquainting me with relevant research papers and optimizing my `R` code.

I thank John Wiley and Sons for permission to include Chapter 3 of my dissertation, which was originally published in *Journal of Time Series Analysis*. Financial support was provided by the United States National Science Foundation grants DMS-1914882 and DMS-2123761.

I would like to thank faculty in the Department of Statistics at Colorado State University, who taught me and shared his/her statistical knowledge with me. My role as a teaching assistant alongside these dedicated educators brought me great satisfaction. I gratefully acknowledge the assistance of my cohorts. In particular, I want to thank my dearest friend at CSU, Zifeng Zhang, who generously helped me in my coursework and cheered me up when times got rough.

Lastly, I must thank my family for their tremendous understanding and encouragement throughout the doctoral program. My achievements would not have been possible without their endless support.

DEDICATION

I would like to dedicate this dissertation to all the people who supported me along the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Functional data analysis (FDA) has grown as a crucial and dynamic subfield of statistics over the past decades, partly because of its rigorous mathematical foundation and broad applicability. It provides powerful tools for data analysis, e.g., functional principal component analysis (FPCA), as well as a novel perspective to look at the data: each observation is thought of as a smooth curve or function. As a result, research questions arising from a wide range of areas, such as finance, physics, psychology and climatology, can be addressed in an efficient and elegant way. A comprehensive introduction to methodological and mathematical foundations of FDA can be found in many monographs, such as Ramsay and Silverman (2005), Horváth and Kokoszka (2012) and Kokoszka and Reimherr (2017).

A salient feature of FDA is that it can precisely capture the temporal dependence structure between curves, which has become one of the most important topics in functional time series analysis. A large number of inferential procedures in functional time series rely on the ubiquitous stationarity assumption, which, in particular, implies that observations share a mean function. However, in reality, such an assumption can be violated by an abrupt change in the mean function, which generally leads to distortions on any further statistical analysis. To resolve this problem, numerous change point tests and estimation methods have been developed and refined in recent years. Most of these methodologies benefit from the prominent dimension reduction technique, the FPCA, which facilitates derivations of the asymptotic null distribution and makes practical computations more efficient. Change point tests based on iid functional observations were established and boosted by Berkes *et al.* (2009), Aue *et al.* (2009), Gromenko *et al.* (2017), etc. Weak dependence among observations were taken into account by Aston and Kirch (2012a), Aston and Kirch (2012b), Horváth *et al.* (2014), Aue *et al.* (2018), among many others. In particular, Aue *et al.* (2018) proposed a fully functional approach to detect and estimate the change point in the mean function, without involving any dimension reduction.

Long memory processes, as an independent topic in scalar time series analysis, have recently been of interest to FDA researchers. There are abundant accounts of scalar long memory processes, such as Giraitis *et al.* (2012), Beran *et al.* (2013) among many others. In contrast, there is even no universal or unified definition suitable for long memory processes in functional case. One way to define functional long-range dependence is to impose a specific model structure, namely, the functional moving average model, which is somehow inspired by the FPCA, see Li *et al.* (2020) and Li *et al.* (2021). An important characteristic of long memory processes is that their realizations exhibit long term fluctuations in the level of the series, the so-called Joseph effect. In other words, it can be rather difficult to distinguish between realizations from a long memory process and those from a proper change point model. Such discrimination problem in scalar time series was resolved by Berkes *et al.* (2006) and Baek and Pipiras (2012), but a similar issue needs to be addressed in functional setting.

This dissertation aims to develop a statistical test to distinguish a change in mean model with short-range dependent (SRD) errors from a functional linear long-range dependent (LRD) process. The paper is organized as follows. In the remainder of Chapter 1, we will introduce some basic concepts and notations in FDA that will be used throughout the paper. In particular, we will focus on the FPCA, the long-run covariance function and a brief literature review on the change point tests for the mean function. We will also discuss characterizations of scalar long memory processes, some useful limit theorems in frequency-domain analysis and estimation methods for the memory parameter. Chapter 2 deals with a simple situation, where the error functions are iid under the null hypothesis. The test procedure is presented for ease of practical implementation. We derive the asymptotic null distribution and prove the consistency of the test. A small simulation study is carried out to validate the theoretical results. Chapter 3 uses a parallel structure as in Chapter 2 and accomplishes the final goal of this dissertation, as stated by the first sentence of this paragraph. In addition, we apply the test to intraday stock price data to see whether there is a sudden change in the mean function of the log return rates. In particular, Chapter 3 is based on the following paper:

C. Baek, P. Kokoszka and X. Meng, Test of change point versus long-range dependence in functional time series. Journal of Time Series Analysis, published online.

1.1 Functional data

Before touching any technical results, we first introduce some basic notations and fundamental concepts in FDA that will be used throughout the dissertation, which mainly come from Horváth and Kokoszka (2012) and Kokoszka and Reimherr (2017). In the FDA setting, samples are usually treated as random functions, which are (measurable) mappings from a probability space to a suitable function space. In FDA, the underlying function space is typically the space of $L^2 = L^2([0,1])$, which is the collection of Lebesgue measurable complex-valued functions x defined on the interval $[0,1]$ such that $\int_0^1 |x(t)|^2 dt < \infty$. We can, without loss of generality, focus on the unit interval because any compact interval can be converted to a subset of $[0,1]$ via a proper normalization. For this reason, oftentimes we omit the limits of integration without causing any confusion. The space L^2 forms a separable Hilbert space with the inner product

$$\langle x, y\rangle = \int x(t)\bar{y}(t)dt,$$

where $\bar{a}$ denotes the complex conjugate of a.

Recall that a statistic is a measurable function, or roughly speaking, a transformation of random samples. In the FDA framework, such a transformation that we often encounter is called a linear operator. More specifically, we may just work with the class of bounded (continuous) linear operators defined on a separable Hilbert space $\mathcal{H}$. Such a class, denoted by $\mathcal{L}$, forms a Banach space, i.e. a complete normed space, with the norm

$$\|\Psi\|_{\mathcal{L}} = \sup\{\|\Psi(x)\| : \|x\| = 1, x \in \mathcal{H}\}.$$

However, to establish satisfying statistical results, it is more convenient to restrict our focus on the Hilbert-Schmidt operator. A bounded linear operator $\Psi : \mathcal{H} \to \mathcal{H}$ is said to be Hilbert-Schmidt if $\sum_{j=1}^{\infty} \|\Psi(e_j)\|^2 < \infty$, where $\{e_j, j \geq 1\}$ is an arbitrary orthonormal basis of $\mathcal{H}$. The class of Hilbert-Schmidt operators, denoted by $\mathcal{S}$, forms a separable Hilbert space with the inner product

$$\langle \Psi, \Phi \rangle_{\mathcal{S}} = \sum_{j=1}^{\infty} \langle \Psi(e_j), \Phi(e_j) \rangle.$$

The Hilbert-Schmidt norm is thus defined as

$$\|\Psi\|_{\mathcal{S}} = \left(\sum_{j=1}^{\infty} \|\Psi(e_j)\|^2 \right)^{1/2}.$$

A concrete example of such operator that often arises in FDA is the integral operator, which is defined by

$$\Psi(x)(t) = \int \psi(t,s)x(s)ds, \quad x \in L^2,$$

where the bivariate function ψ is call the kernel of Ψ. One can show that Ψ is Hilbert-Schmidt if and only if $\iint \psi^2(t,s)dtds < \infty$, in which case $\|\Psi\|_{\mathcal{S}} = \left(\iint \psi^2(t,s)dtds\right)^{1/2}$ and $\|\Psi\|_{\mathcal{L}} \leq \|\Psi\|_{\mathcal{S}}$.

Now we turn to the problem of statistical inference. A random function $X = \{X(t), t \in [0,1]\}$ in L^2 is said to be square integrable if $E\|X\|^2 = E\int X^2(t)dt < \infty$. For such X we can define the mean and covariance functions by

$$\mu(t) = E[X(t)],$$

$$c(t,s) = E[(X(t) - \mu(t))(X(s) - \mu(s))]. \quad (1.1.1)$$

Notably, (1.1.1) defines an integral operator C, which is called the covariance operator, via

$$C(y)(t) = \int c(t,s)y(s)ds = E[\langle X - \mu, y \rangle (X - \mu)], \quad y \in L^2. \quad (1.1.2)$$

Suppose that the functions $X_1, X_2, \ldots, X_N$ are iid in L^2 which are square integrable. Then we can estimate the above three quantities by their sample counterparts:

$$\begin{aligned}
\hat{\mu}(t) &= \frac{1}{N}\sum_{i=1}^{N} X_i(t), \\
\hat{c}(t,s) &= \frac{1}{N}\sum_{i=1}^{N}(X_i(t)-\hat{\mu}(t))(X_i(s)-\hat{\mu}(s)), \\
\widehat{C}(y) &= \frac{1}{N}\sum_{i=1}^{N}\langle X_i-\hat{\mu}, y\rangle(X_i-\hat{\mu}). \qquad (1.1.3)
\end{aligned}$$

Remarkably, one can show that C and $\widehat{C}$ are Hilbert-Schmidt, which implies that $C, \widehat{C} \in L^2([0,1]\times[0,1])$.

1.1.1 Functional principal components and their estimation

Functional principal component analysis (FPCA), as a natural extension of principal component analysis in multivariate statistics, has become a powerful dimension reduction technique in FDA. The idea behind it is surprisingly simple and intuitive. It aims to represent an infinite dimensional curve by a few fundamental and parsimonious functions, i.e. the basis functions, while retaining as much information as possible. Technically speaking, for a square integrable random function X, the goal of FPCA is to find an orthonormal system $u_1, u_2, \ldots u_p$ such that the mean squared error

$$S(u_1, u_2 \ldots, u_p) = E\left\|X - \sum_{k=1}^{p}\langle X, u_k\rangle u_k\right\|^2 \qquad (1.1.4)$$

is minimized over all orthonormal systems consisting of p functions. One can show that with additional condition the above optimization problem can be converted to the eigen-analysis of the covariance operator (1.1.2), that is, finding the eigenvalues λ_j and eigenfunctions v_j such that

$$C(v_j) = \lambda_j v_j, \quad j \geq 1.$$

This result can be summarized in the following theorem (see, e.g. Kokoszka and Reimherr (2017)).

THEOREM 1.1.1 *Suppose the random function* X *is in* L^2 *and square integrable with covariance operator* C *defined by* (1.1.2). *In addition, the eigenvalues of* C *satisfy the separation condition*

$$\lambda_1 > \lambda_2 > \cdots \lambda_p > \lambda_{p+1}. \tag{1.1.5}$$

Then for any fixed $p \geq 1$*, the MSE* (1.1.4) *is minimized if* $u_j = v_j$*, where* v_j *are the eigenfunctions of* C *with unit norm.*

The above theorem leads to the famous Karhunen-Loéve expansion of a square integrable random function X, that is,

$$X - \mu = \sum_{j=1}^{\infty} \xi_j v_j, \quad \xi_j = \langle X - \mu, v_j \rangle, \tag{1.1.6}$$

where $\mu = EX$ and v_j are called the functional principal components (FPC's). The information contained in X is thus summarized in the random variables ξ_j, which are called scores of X. The infinite symbol in (1.1.6) implies that X may lie in an infinite dimensional space, which makes any practical calculation infeasible. Thus, in practice, we must reduce the dimension to a finite number, which can be done by truncating the series (1.1.6) at a finite depth, e.g. p, and then replacing ξ_j and v_j by their estimated counterparts. With this notion in mind, suppose $X_1, \ldots, X_N$ are iid in L^2 which are square integrable functional observations. After centering the data (subtracting each observation by their sample mean), each curve can be approximated by the following representation in a p-dimensional space:

$$X_i(t) = \sum_{j=1}^{p} \widehat{\xi}_{ji} \hat{v}_j(t).$$

Naturally, $\hat{v}_j$ should derive from the eigen-analysis of the sample covariance operator (1.1.3), i.e. $\widehat{C}(\hat{v}_j) = \widehat{\lambda}_j \hat{v}_j,\ j \geq 1$. The $\hat{v}_j$ are called the empirical functional principal components (EFPC's) of the data $X_1, \ldots, X_N$. The estimated scores $\widehat{\xi}_{ji}$ can be computed by $\widehat{\xi}_{ji} = \langle X_i, \hat{v}_j \rangle = \int X_i(t)\hat{v}_j(t)dt$. In practical implementation, the value of p (typically a single digit number) can be determined by the cumulative percentage of total variance (CPV) method: we choose p for which

$CPV(p) = \frac{\sum_{k=1}^{p}\widehat{\lambda}_k}{\sum_{k=1}^{N}\widehat{\lambda}_k}$ exceeds a desirable level, e.g. 85%. The computations of $\widehat{\lambda}_j$ and $\hat{v}_j$ can be easily carried out by the `R` function `pca.fd` in the `fda` package.

We conclude this section by two asymptotic results, which can also be found in Kokoszka and Reimherr (2017). One states the central limit theorem and the other indicates that the EFPC's consistently estimate the FPC's.

THEOREM 1.1.2 *Suppose that $\{X_i\}$ is a sequence of iid square integrable random functions with expected value μ and covariance operator C. Then,*

$$N^{-1/2}\sum_{i=1}^{N}(X_i-\mu) \xrightarrow{d} Z,$$

where Z is normal with $EX = 0$ and covariance operator C.

THEOREM 1.1.3 *Suppose that the random functions $X_1, \ldots, X_N$ are iid in L^2 and have the same distribution as X, which satisfies $E\|X\|^4 < \infty$. In addition, condition* (1.1.5) *holds. Then, for each $1 \le j \le p$,*

$$\limsup_{N\to\infty} N \cdot E\left[\|\hat{c}_j\hat{v}_j - v_j\|^2\right] < \infty, \quad \limsup_{N\to\infty} N \cdot E\left[\left|\widehat{\lambda}_j - \lambda_j\right|^2\right] < \infty.$$

where $\hat{c}_j = sign(\langle \hat{v}_j, v_j\rangle)$ ensures that $\hat{v}_j$ points to the same direction as v_j.

Note that $\hat{c}_j$ depend on the unknown v_j, so it cannot be computed from the data and we must make sure that the statistics of interest do not depend on it.

1.1.2 The long-run covariance kernel

In Section 1.1.1 we have assumed that the functional observations are iid and then introduced the FPCA in accordance with such assumption. However, oftentimes it is unrealistic to assume that observations are iid, which excludes serial correlations within samples. As a result, it may be improper to use the covariance function (1.1.1) or the covariance operator (1.1.2) to capture their dependence structure. Thus, it is necessary to introduce a general definition of "weak dependence", which takes temporal dependence into account. Such weak dependence can be characterized by the so-called $L^p - m-$approximability (see, Hörmann and Kokoszka (2010)). The $L^p - m-$approximability is one of many ways of quantifying weak dependence. Various mixing conditions have been used for decades. We use $L^p - m-$approximability because it allows us to use existing asymptotic theory relevant to the problem we want to solve.

For $p \geq 1$, let $L^p_{\mathcal{H}} = L^p_{\mathcal{H}}(\Omega, \mathcal{A}, \mathcal{P})$ be the space of $\mathcal{H} = L^2$ valued random functions X such that

$$v_p(X) = (E\|X\|^p)^{1/p} = \left(E\left\{\int X^2(t)dt\right\}^{p/2}\right)^{1/p} < \infty.$$

DEFINITION 1.1.1 A sequence $\{X_n\} \in L^p_{\mathcal{H}}$ is called $L^p - m-$approximable if each X_n admits the representation

$$X_n = f(\varepsilon_n, \varepsilon_{n-1}, \ldots),$$

where the ε_i are iid elements taking values in a measurable space $\mathcal{S}$, and f is a measurable function $f : \mathcal{S}^\infty \to \mathcal{H}$. Moreover let

$$X_n^{(m)} = f(\varepsilon_n, \varepsilon_{n-1}, \ldots, \varepsilon_{n-m+1}, \varepsilon^{(m)}_{n,n-m}, \varepsilon^{(m)}_{n,n-m-1}, \ldots),$$

where $\{\varepsilon^{(m)}_{n,\ell},\ m \geq 1, -\infty < n, \ell < \infty\}$ are iid copies of ε_0 and independent of $\{\varepsilon_k\}$. Then we have

$$\sum_{m=1}^{\infty} v_p\left(X_n - X_n^{(m)}\right) < \infty.$$

The idea of Definition 1.1.1 is to approximate $\{X_n, n \in \mathbb{Z}\}$ by strictly stationary and $m-$ dependent processes $\{X_n^{(m)},\ n \in \mathbb{Z}, m \geq 1\}$ in the sense that for each n, $X_n^{(m)}$ converges to X_n as $m \to \infty$. Note that Definition 1.1.1 indicates that $\{X_n\}$ is strictly stationary and that $X_n^{(m)}$ is equal in distribution to X_n for each $m \geq 1$. Amazingly, a broad class of linear and nonlinear processes fit into Definition 1.1.1 with additional moderately mild conditions. Examples can be seen in Horváth and Kokoszka (2012). It can be shown that $L^p - m-$approximability can be preserved under various types of transformations (see, e.g. Lemma 16.1 in Horváth and Kokoszka (2012)).

As in the iid case, many asymptotic results can be established in the setting of $L^p - m-$ approximability (see, e.g. Horváth and Kokoszka (2012)). We start with the central limit theorem of the sample mean, which is an analogue of Theorem 1.1.2, under the weak dependence framework.

THEOREM 1.1.4 *If* $\{X_i\}$ *is a zero mean* $L^2 - m-$*approximable sequence, then*

$$N^{-1/2}\sum_{i=1}^{N} X_i \xrightarrow{d} G \text{ in } L^2,$$

where G *is a Gaussian process with*

$$EG(t) = 0 \quad \text{and} \quad E[G(t)G(s)] = c(t,s);$$

$$c(t,s) = E[X_0(t)X_0(s)] + \sum_{i\geq 1} E[X_0(t)X_i(s)] + \sum_{i\geq 1} E[X_0(s)X_i(t)]. \tag{1.1.7}$$

Fortunately, asymptotic results related to the FPCA remain roughly the same under $L^p - m-$approximability, compared to their counterparts in the iid case.

THEOREM 1.1.5 *Suppose that* $\{X_n\} \in L^4_{\mathcal{H}}$ *is an* $L^4 - m-$*approximable sequence and the separation condition* (1.1.5) *holds. Then, for each* $1 \leq j \leq p$,

$$\limsup_{N\to\infty} N \cdot E\left[\|\hat{c}_j\hat{v}_j - v_j\|^2\right] < \infty, \quad \limsup_{N\to\infty} N \cdot E\left[\left|\widehat{\lambda}_j - \lambda_j\right|^2\right] < \infty.$$

where $\hat{c}_j = sign(\langle \hat{v}_j, v_j\rangle)$ *and* $\widehat{\lambda}_j$, $\hat{v}_j$ *are the eigenvalues and the eigenfunctions of the sample covariance operator* (1.1.3)*, respectively.*

Under weak dependence, it seems tempting to stick to the covariance function (1.1.1) when it comes to the FPCA, because of Theorem 1.1.5. However, Theorem 1.1.4 suggests that under some situations it may be more appropriate to consider the long-run covariance function (1.1.7) in conducting the FPCA. Before we can touch that, we need to find a proper way to estimate (1.1.7).

A popular approach for the long-run covariance function estimation is to use the kernel estimator (see, e.g. Horváth *et al.* (2013)):

$$\hat{c}_N(t,s) = \widehat{\gamma}_0(t,s) + \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right)\{\widehat{\gamma}_i(t,s) + \widehat{\gamma}_i(s,t)\}, \tag{1.1.8}$$

where

$$\widehat{\gamma}_i(t,s) = \frac{1}{N}\sum_{j=i+1}^{N}\{X_j(t) - \hat{\mu}(t)\}\{X_{j-i}(s) - \hat{\mu}(s)\}, \quad 0 \le i \le N-1,$$

K is the kernel function and the parameter h is the bandwidth.

Under some mild conditions, we can establish the consistency of the long-run covariance function estimator (1.1.8) as well as the corresponding eigenfunctions, summarized in the following theorem and corollary.

THEOREM 1.1.6 *Suppose that $\{X_n\}_{n=1}^N \in L^2_{\mathcal{H}}$ is an $L^2 - m-$approximable sequence such that $EX_n = \mu$ for each n. The kernel function K is symmetric and continuous, $K(0) = 1$, and $K(u) = 0$ if $|u| > c$ for some $c > 0$. The bandwidth $h = h(N)$ satisfies $h(N) \to \infty$ and $h(N)/N \to 0$ as $N \to \infty$. Moreover, assume that*

$$v_2\left(X_n - X_n^{(m)}\right) = \left(E\|X_n - X_n^{(m)}\|^2\right)^{1/2} = o\left(m^{-1}\right).$$

Then,

$$\iint \{\hat{c}_N(t,s) - c(t,s)\}^2 dtds = o_P(1),$$

where c and $\hat{c}_N$ are defined in (1.1.7) *and* (1.1.8)*, respectively.*

The proof of Theorem 1.1.6 can be found in Horváth *et al.* (2013). The following corollary follows immediately from Lemma 2.3 in Horváth and Kokoszka (2012).

COROLLARY 1.1.1 *If the Assumptions of Theorem 1.1.6 hold and, in addition, the eigenvalues of* (1.1.7) *satisfy the following separation condition:*

$$\lambda_1 > \lambda_2 > \cdots \lambda_p > \lambda_{p+1}.$$

Then, for any fixed $p \geq 1$*,*

$$\|\hat{v}_j - v_j\| = o_P(1), \quad \text{as } N \to \infty,$$

where v_j *and* $\hat{v}_j$ *are the eigenfunctions of* (1.1.7) *and* (1.1.8)*, respectively.*

1.1.3 Change point detection and estimation

So far we have discussed two types of dependence structures of a functional time series, i.e. iid and $L^p - m-$approximability. Note that both of them implicitly assume that observations have the same mean function (if exists), which may not be the case in reality. For example, we cannot expect that the daily temperatures in a certain region always stay the same over a large time span. Instead, we may expect that the average daily temperatures gradually increase due to the global warming. Changes in the mean function of functional observations can distort statistical inferences and lead to ridiculous conclusions. Hence, it is necessary to conduct a test on whether observations have a common mean at the beginning of the statistical analysis. We shall present several formal statistical tests for doing this in both iid and $L^p - m-$approximability settings. We will also discuss how to locate the change point once it is detected by the test. Here we just focus on a single change point detection and estimation, because it closely relates to the problem that this dissertation is trying to solve.

Under the null hypothesis, we assume that functional observations come from the model

$$X_t = \mu + Y_t, \quad 1 \leq t \leq N, \tag{1.1.9}$$

whereas under the alternative hypothesis,

$$X_t = \begin{cases} \mu_1 + Y_t, & 1 \leq t \leq k^*, \\ \mu_2 + Y_t, & k^* < t \leq N, \end{cases} \tag{1.1.10}$$

where the random functions Y_t and the mean functions μ, μ_1, μ_2 are in L^2 and $k^* = \lfloor N\theta \rfloor$ for some $\theta \in (0, 1)$.

To distinguish between (1.1.9) and (1.1.10), Berkes *et al.* (2009) constructed the cumulative sum (CUSUM) statistic based on the largest d scores of the covariance function, that is,

$$S_{N,d} = \frac{1}{N^2} \sum_{l=1}^{d} \widehat{\lambda}_l^{-1} \sum_{k=1}^{N} \left(\sum_{1 \le i \le k} \widehat{\eta}_{i,l} - \frac{k}{N} \sum_{1 \le i \le N} \widehat{\eta}_{i,l} \right)^2, \tag{1.1.11}$$

where the $\widehat{\lambda}_l$ is the l-th largest eigenvalue of the covariance operator (1.1.3) and $\{\widehat{\eta}_{i,l}\}_{i=1}^N$ are the corresponding scores. Berkes *et al.* (2009) verified the limiting distribution of the test statistic $S_{N,d}$ under the null hypothesis.

THEOREM 1.1.7 *Suppose that model* (1.1.9) *holds with the Y_t iid in L^2 such that $EY_0 = 0$ and $E\|Y_0\|^4 < \infty$. In addition, the eigenvalues of* (1.1.2) *satisfy $\lambda_1 > \lambda_2 > \cdots \lambda_d > \lambda_{d+1}$ for some $d > 0$. Then,*

$$S_{N,d} \xrightarrow{d} \int_0^1 \sum_{1 \le l \le d} B_l^2(x) dx,$$

where the B_l are iid standard Brownian bridges.

Moreover, they proposed a change point estimator defined as

$$\widehat{\theta}_N = \inf \left\{ x : T_N(x) = \sup_{0 \le y \le 1} T_N(y) \right\}, \tag{1.1.12}$$

where $T_N(x) = \frac{1}{N} \sum_{l=1}^{d} \widehat{\lambda}_l^{-1} \left(\sum_{1 \le i \le Nx} \widehat{\eta}_{i,l} - x \sum_{1 \le i \le N} \widehat{\eta}_{i,l} \right)^2$. They proved the consistency of the test as well as the change point estimator (1.1.12) if the alternative hypothesis is true.

THEOREM 1.1.8 *Suppose that model* (1.1.10) *holds with the Y_t iid in L^2 such that $EY_0 = 0$ and $E\|Y_0\|^2 < \infty$. If the jump function $\Delta(t) = \mu_1(t) - \mu_2(t)$ is not orthogonal to the subspace that is spanned by the first d eigenfunctions of the covariance operator* (1.1.2)*. Then,*

$$S_{N,d} \xrightarrow{P} \infty, \quad \widehat{\theta}_N \xrightarrow{P} \theta, \quad \text{as } N \to \infty.$$

Note that the test statistic (1.1.11) applies the dimension reduction technique FPCA in detecting the change point. On the other hand, Aue *et al.* (2018) proposed a test statistic, i.e. the functional CUSUM statistic, that uses fully functional data and does not involve any dimension reduction. Furthermore, with the $L^p - m-$approximability assumption, they derived the asymptotic null distribution of the test statistic and proved the consistency of the test. As a result, the long-run covariance function (1.1.7) is used to capture the dependence structure of the observations and conduct the FPCA.

Define the functional CUSUM statistic as

$$S^0_{N,k} = \frac{1}{\sqrt{N}} \left(\sum_{i=1}^{k} X_i - \frac{k}{N} \sum_{i=1}^{N} X_i \right). \tag{1.1.13}$$

THEOREM 1.1.9 *Suppose that* $\{X_t\}$ *comes from* (1.1.9) *and* $\{Y_t\} \in L^p_{\mathcal{H}}$ *is an* $L^p - m-$*approximable sequence for some* $p > 2$. *Then,*

$$T_N := \max_{1 \le k \le N} \|S^0_{N,k}\|^2 \xrightarrow{d} \sup_{0 \le x \le 1} \sum_{l=1}^{\infty} \lambda_l B_l^2(x), \quad \text{as } N \to \infty,$$

where the λ_l *are the eigenvalues of* (1.1.7) *and the* B_l *are iid standard Brownian bridges.*

THEOREM 1.1.10 *If model* (1.1.10) *holds with* $\mu_1 - \mu_2 \neq 0$, *and if* $\{Y_t\} \in L^p_{\mathcal{H}}$ *is* $L^p - m-$*approximable for some* $p > 2$. *Then,*

$$T_N \xrightarrow{P} \infty, \quad \frac{\hat{k}^*_N}{N} \xrightarrow{P} \theta, \quad \text{as } N \to \infty,$$

where

$$\hat{k}^*_N = \min \left\{ k : \|S^0_{N,k}\| = \max_{1 \le k' \le N} \|S^0_{N,k'}\| \right\}. \tag{1.1.14}$$

1.2 Theory for long memory processes

Long memory models have been used for several decades in various disciplines such as astronomy, hydrology, finance, climatology, telecommunications and many others. For example, in the late 19th century the astronomer and mathematician Simon Newcomb found that the usual $\sigma/\sqrt{n}$-rule is not appropriate in estimating the errors of astronomical observations. As confirmed by other scholars, this is due to persistent serial correlation (see, e.g. Pearson (1902) and Jeffreys (1961)). At that time, the notion of long-range dependence was not prevalent, nor was it defined rigorously in a mathematical way. In contrast, nowadays there are numerous ways to characterize or define a long memory process. We shall introduce two popular definitions of the long-range dependence in terms of second-order stationary processes. One uses the autocovariance function, in the time-domain analysis whereas the other involves the frequency-domain concept, the spectral density. These definitions are essentially equivalent, but provide illuminating ways to view the long memory process from different perspectives. Before diving into it, we first give some preliminaries, which can be found in many classical time series textbooks, e.g. Brockwell and Davis (1991).

DEFINITION 1.2.1 If $\{X_t, t \in T\}$ is a process such that $\text{Var}(X_t) < \infty$ for each $t \in T$, then the autocovariance function $\gamma_X(\cdot, \cdot)$ of $\{X_t\}$ is defined by

$$\gamma_X(r, s) = \text{Cov}(X_r, X_s) = E[(X_r - EX_r)(X_s - EX_s)], \quad r, s \in T.$$

DEFINITION 1.2.2 The time series $\{X_t, t \in \mathbb{Z}\}$, with index set $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$, is said to be second-order stationary or weakly stationary if

(i) $E|X_t|^2 < \infty$ for all $t \in \mathbb{Z}$,

(ii) $EX_t = \mu$ is constant for all $t \in \mathbb{Z}$,

and

(iii) $\gamma_X(h) := \gamma_X(h, 0) = \text{Cov}(X_{t+h}, X_t)$ for all $t, h \in \mathbb{Z}$.

The weak stationarity assumption plays a fundamental and crucial role in time series analysis, especially when only a single series of observations is available. Roughly speaking, it imposes regularity conditions on the behavior of a time series in the sense that some basic features are preserved over time. More precisely, the first two moments of the series do not depend on time and the covariance relies only on the separation of the two time points, so that these quantities may be estimated by simple averages. In particular, the theoretical autocovariance function can be estimated by its sample counterpart.

DEFINITION 1.2.3 The sample autocovariance function of $\{x_1, \ldots, x_N\}$ is defined by

$$\widehat{\gamma}(h) = \frac{1}{N}\sum_{j=1}^{N-h}(x_{j+h} - \bar{x})(x_j - \bar{x}), \quad 0 \le h < N,$$

and $\widehat{\gamma}(h) = \widehat{\gamma}(-h), -N < h \le 0$, where $\bar{x} = \frac{1}{N}\sum_{j=1}^{N} x_j$.

Remarkably, the weak stationarity can be thought of as a "bridge" between time-domain and frequency-domain analyses. In fact, the autocovariance of a second-order stationary process admits a spectral representation, which leads to the definition of the spectral distribution function.

THEOREM 1.2.1 *If γ_X is the autocovariance function of a second-order stationary process $\{X_t\}$. Then,*

$$\gamma_X(h) = \int_{(-\pi,\pi]} e^{ihv} dF_X(v), \quad \forall\, h \in \mathbb{Z},$$

where F_X is a right-continuous, non-decreasing, bounded function on $[-\pi, \pi]$ with $F_X(-\pi) = 0$. The function F_X is called the spectral distribution function of γ_X. Furthermore, if F_X can be represented as

$$F_X(\lambda) = \int_{-\pi}^{\lambda} f_X(v)dv, \quad -\pi \le \lambda \le \pi,$$

then f_X is called the spectral density of γ_X.

In particular, if $\{X_t\}$ is real, then $\gamma_X(h) = \int_{(-\pi,\pi]} \cos(vh) dF_X(v)$, $\forall\, h \in \mathbb{Z}$ and F_X is symmetric in the sense that $F_X(\lambda) = F_X(\pi-) - F_X(-(\lambda-))$, $-\pi < \lambda < \pi$. Furthermore, if γ_X has spectral density f_X, then $f_X(\lambda) = f_X(-\lambda)$, $-\pi \le \lambda \le \pi$, in which case $\gamma_X(h) = 2\int_0^{\pi} f_X(v)\cos(vh)dv$.

The following theorem provides a useful way to compute the spectral density f from the autocovariance function γ under some regularity conditions. It also reveals the fact that f and γ are Fourier transform pairs.

THEOREM 1.2.2 *An absolutely summable complex-valued function γ defined on the integers is the autocovariance function of a stationary process if and only if*

$$f(\lambda) := \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\lambda}\gamma(n) \geq 0, \quad \forall\, \lambda \in [-\pi, \pi],$$

in which case f is the spectral density of γ.

The empirical or sample version of the spectral density f is called the periodogram. Although it is not a consistent estimator of f, it has a number of useful asymptotic properties and will be widely used in subsequent chapters.

DEFINITION 1.2.4 The periodogram of $\{X_1, \dots, X_N\}$ is defined at the Fourier frequencies $\omega_j = \frac{2\pi j}{N}$, $\omega_j \in [-\pi, \pi]$, by

$$I_X(\omega_j) = \frac{1}{2\pi N} \left| \sum_{t=1}^{N} X_t e^{-it\omega_j} \right|^2. \tag{1.2.1}$$

Since $\sum_{t=1}^{N} e^{-it\omega_j} = 0$ if $\omega_j \neq 0$, the periodogram is translation invariant, that is,

$$I_{X+c}(\omega_j) = \frac{1}{2\pi N} \left| \sum_{t=1}^{N} (X_t + c) e^{-it\omega_j} \right|^2 = I_X(\omega_j), \quad c \in \mathbb{R},\ \omega_j \neq 0.$$

1.2.1 Characterization

We shall introduce two equivalent characterizations of long-range dependence for a second-order stationary process. We start with the heuristic and intuitive definition of different types of memory from the perspective of time-domain.

DEFINITION 1.2.5 A second order stationary process $\{X_t\}$ with autocovariance function γ is said to have:

Short memory if

$$\sum_{k\in\mathbb{Z}} |\gamma(k)| < \infty \quad \text{and} \quad \sum_{k\in\mathbb{Z}} \gamma(k) > 0;$$

Long memory if

$$\sum_{k\in\mathbb{Z}} |\gamma(k)| = \infty;$$

Negative memory, or antipersistence, if

$$\sum_{k\in\mathbb{Z}} |\gamma(k)| < \infty \quad \text{and} \quad \sum_{k\in\mathbb{Z}} \gamma(k) = 0.$$

In fact, a more detailed time-domain characterization of different types of memory relies on the asymptotic behavior of the autocovariance function $\gamma(k)$, as $k \to \infty$, more precisely, the rate of decay of $\gamma(k)$ at infinity (see, Giraitis *et al.* (2012)). Recall the basic relation between γ and the spectral density f, i.e. $2\pi f(0) = \sum_{k\in\mathbb{Z}} \gamma(k)$, when γ is absolutely summable. It hints that the limit behavior of γ at infinity may determine that of f near zero frequency, and vice versa. Indeed, under some regularity conditions, Definition 1.2.5 can be rephrased by the following frequency-domain characterization. The equivalence of these two definitions can be found in Giraitis *et al.* (2012) or Beran *et al.* (2013).

DEFINITION 1.2.6 Suppose the spectral density f of a second-order stationary process $\{X_t\}$ exists, is bounded on $[\epsilon, \pi]$ for any $\epsilon > 0$, and satisfies

$$f(\lambda) \sim c_f |\lambda|^{-2d}, \quad \text{as } \lambda \to 0, \tag{1.2.2}$$

for some $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$ and $c_f > 0$. The process $\{X_t\}$ is said to have negative memory, short memory, or long memory, depending on whether $d \in \left(-\frac{1}{2}, 0\right)$, $d = 0$ or $d \in \left(0, \frac{1}{2}\right)$, where the parameter d is called the memory parameter of $\{X_t\}$.

This section ends with a brief introduction to a special continuous-time stochastic process, the fractional Brownian motion, which often serves as the limiting distribution of partial sums of dependent random variables.

DEFINITION 1.2.7 Let $0 < H < 1$ be any number. If a Gaussian process $B_H = \{B_H(t), t \in \mathbb{R}\}$, with $B_H(0) = 0$ and $EB_H(t) = 0$ for all $t \in \mathbb{R}$, has the covariance function

$$r_H(s,t) = \frac{1}{2}\left(|s|^{2H} + |t|^{2H} - |s-t|^{2H}\right), \quad s,t \in \mathbb{R},$$

then B_H is called the fractional Brownian motion (fBm) with self-similarity parameter H.

1.2.2 Asymptotic results for sums of weighted periodograms

We focus on some useful asymptotic results of periodograms and their weighted sums, which are the bedrock of proofs in later chapters and are summarized in monographs on long memory processes such as Giraitis *et al.* (2012) and Beran *et al.* (2013). However, to derive asymptotic results regarding the periodograms, we need to impose a specific linear structure on the underlying process. In fact, it can be proven that a large class of stationary processes fits into such a linear representation, due to Wold decomposition, as long as the spectral density of the process satisfies some regularity condition.

DEFINITION 1.2.8 Let $\{\zeta_j, j \in \mathbb{Z}\} \sim \mathrm{WN}(0, \sigma^2)$ be white-noise and a_k, $k = 0, 1, \ldots$, be a sequence of real numbers. A linear or a moving-average process is defined by

$$X_j = \sum_{k=0}^{\infty} a_k \zeta_{j-k}, \quad \sum_{k=0}^{\infty} a_k^2 < \infty, \quad j \in \mathbb{Z}, \tag{1.2.3}$$

where the ζ_j are called the innovations of $\{X_j\}$. The transfer function of X_j is defined as

$$A_X(\lambda) = \sum_{k=0}^{\infty} a_k e^{-ik\lambda}, \quad \lambda \in [-\pi, \pi].$$

Noticeably, with the linear structure (1.2.3), the type of memory of $\{X_j\}$ can be specified by the coefficients $\{a_k\}$. To be specific, if the a_k satisfy $\sum_{k=0}^{\infty} |a_k| < \infty$ and $\sum_{k=0}^{\infty} a_k \neq 0$, then $\{X_j\}$ has short memory. If the a_k satisfy $a_k \sim c_a k^{-1+d}$, $c_a \neq 0$, $d \in \left(0, \frac{1}{2}\right)$, as $k \to \infty$, then $\{X_j\}$ has long memory. If the a_k satisfy $a_k \sim c_a k^{-1+d}$, $c_a \neq 0$, $d \in \left(-\frac{1}{2}, 0\right)$, as $k \to \infty$, and $\sum_{k=0}^{\infty} a_k = 0$, then $\{X_j\}$ has negative memory.

Recall that the periodogram (1.2.1) is not a consistent estimate of the spectral density f_X. Nevertheless, we can show that the it is a mean consistent estimator of f_X in the sense that $|EI_X(\omega_j) - f_X(\omega_j)| \to 0$ with a known convergence rate, if an additional smoothness condition on f_X is imposed. Moreover, the convergence rate depends on the degree of smoothness of f_X. In fact, as we shall see later, such a smoothness condition is also needed in deriving delicate asymptotic results for weighted sums of periodograms. In light of this, we introduce the Lipschitz continuity.

DEFINITION 1.2.9 A function f defined on D is said to belong to the Lipschitz class $\Lambda_\alpha(D)$, $0 < \alpha \leq 1$, if for some $0 < C < \infty$, $|f(u) - f(v)| \leq C|u - v|^\alpha$, for all $u, v \in D$.

THEOREM 1.2.3 *Let either* $\Delta < a < \pi$ *or* $\Delta = a = \pi$. *If the spectral density of* $\{X_j\}$ *in* (1.2.3) *satisfies* $f_X \in \Lambda_\beta([0, a])$, $0 < \beta \leq 1$, *then*

$$\max_{0<\omega_j<\Delta} |EI_X(\omega_j) - f_X(\omega_j)| = \begin{cases} O\left(N^{-1}\log N\right), & \beta = 1, \\ O\left(N^{-\beta}\right), & 0 < \beta < 1. \end{cases}$$

Oftentimes we are interested in weighted sums of periodograms

$$Q_{N,X} = \sum_{j=1}^{\widetilde{N}} b_{N,j} I_X(\omega_j), \quad \widetilde{N} = \left\lfloor \frac{N}{2} \right\rfloor - 1,$$

where $\{b_{N,j}, j = 1, \ldots, \widetilde{N}\}$ is an array of real numbers and I_X is the periodogram of $\{X_t\}_{t=1}^N$ defined in (1.2.1). Theorem 1.2.3 alludes that likely $Q_{N,X}$ is close to the quantity $Q_{N,f} := \sum_{j=1}^{\widetilde{N}} b_{N,j} f_X(\omega_j)$ in some sense. This intuition is formalized in the following two theorems, whose proofs are discussed in Giraitis *et al.* (2012). One establishes the asymptotic normality of $Q_{N,X}$ and the other derives an upper bound of the MSE of $Q_{N,X}$. In both cases, $Q_{N,f}$ can serve as the "central value".

THEOREM 1.2.4 *Suppose that the process* $\{X_j\}$ *is of the form* (1.2.3) *with a continuous transfer function, and the innovations* ζ_j *are strengthened to be IID*$(0,1)$ *such that* $E\zeta_0^4 < \infty$*. In addition, the weights* $b_{N,j}$ *satisfy*

$$\frac{\max_{j\in\{1,\ldots,\widetilde{N}\}} |b_{N,j}|}{\left(\sum_{j=1}^{\widetilde{N}} b_{N,j}^2\right)^{1/2}} \to 0, \quad \textit{as } N \to \infty.$$

If the spectral density f_X *of* $\{X_j\}$ *satisfies* $f_X \in \Lambda_\beta([0,\pi])$ *with* $\frac{1}{2} < \beta \leq 1$ *and if for some* $C_1, C_2 \in (0,\infty)$*,*

$$f_X(\lambda) \in [C_1, C_2], \quad \forall\, \lambda \in [0,\pi],$$

then,

$$EQ_{N,X} = \sum_{j=1}^{\widetilde{N}} b_{N,j} f_X(\omega_j) + o(v_N),$$

$$v_N^{-1}\left(Q_{N,X} - \sum_{j=1}^{\widetilde{N}} b_{N,j} f_X(\omega_j)\right) \xrightarrow{d} \mathcal{N}(0,1),$$

where

$$v_N^2 = \sum_{j=1}^{\widetilde{N}} (b_{N,j} f_X(\omega_j))^2 + \textit{cum}_4(\zeta_0) \cdot \frac{1}{N}\left(\sum_{j=1}^{\widetilde{N}} b_{N,j} f_X(\omega_j)\right)^2.$$

THEOREM 1.2.5 *Let* $\{X_j\}$ *be defined as in* (1.2.3) *and the innovations* ζ_j *are strengthened to be IID*$(0,1)$ *such that* $E\zeta_0^4 < \infty$*. Assume that the spectral density of* $\{X_j\}$ *satisfies* $f_X(\lambda) = |\lambda|^{-2d} g(\lambda)$*,* $|d| < \frac{1}{2}$*, where* $g \in \Lambda_\beta([0,\pi])$ *with* $\frac{1}{2} < \beta \leq 1$ *and is bounded away from* 0 *and* ∞*. Then, for some constant* C*,*

$$E\left(Q_{N,X} - \sum_{j=1}^{\widetilde{N}} b_{N,j} f_X(\omega_j)\right)^2 \leq C \cdot B_{f,N}^2,$$

where $B_{f,N}^2 = \sum_{j=1}^{\widetilde{N}} (b_{N,j} f_X(\omega_j))^2$*.*

1.2.3 Local Whittle estimation

We discuss some estimation methods for the memory parameter d, or sometimes we consider the self-similarity (Hurst) parameter $H = d + 0.5$. In particular, we will focus on Whittle and local Whittle estimations, which are widely applied in practice and can be easily computed via many `R` packages, such as `LongMemoryTS`, `LSTS`, `longmemo` and many others. The former is a parametric approach whereas the later is used especially in semiparametric models. Both of them can be thought of as deriving from the approximation of the log-likelihood function of a zero mean Gaussian process. We start with the Whittle estimation, which is the predecessor of the LW estimation.

We consider a linear model structure similar to (1.2.3), but with subtle differences. Assume that observations $\{X_1, \ldots, X_N\}$ follow the linear model

$$X_j = \sum_{k=0}^{\infty} a_k(\boldsymbol{\theta})\zeta_{j-k}, \quad a_0(\boldsymbol{\theta}) = 1, \quad j \in \mathbb{Z}, \tag{1.2.4}$$

$$\sum_{k=0}^{\infty} a_k(\boldsymbol{\theta})^2 < \infty, \quad \boldsymbol{\theta} \in \boldsymbol{\Theta},$$

where $\{a_k(\boldsymbol{\theta})\}$ is a sequence of real numbers, $\{\zeta_j\} \sim \mathrm{WN}(0, \sigma^2)$ and $\boldsymbol{\Theta}$ is a subset of $\mathbb{R}^q$. We are interested in estimating σ^2 and $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ characterizes the linear dependence structure of $\{X_j\}$. We further assume that $\boldsymbol{X} = (X_1, \ldots, X_N)$ follows a zero mean Gaussian distribution with covariance matrix $\boldsymbol{\Sigma}_{N,\boldsymbol{\theta}}$. If we denote $\boldsymbol{\Gamma}_{N,\boldsymbol{\theta}} = \boldsymbol{\Sigma}_{N,\boldsymbol{\theta}}/\sigma^2$, then the corresponding log-likelihood function can be written as

$$\ell(\boldsymbol{\theta}, \sigma^2) = -\frac{N}{2}\log(2\pi) - N\log\sigma - \frac{1}{2}\log|\boldsymbol{\Gamma}_{N,\boldsymbol{\theta}}| - \frac{1}{2\sigma^2}\boldsymbol{X}^T\boldsymbol{\Gamma}_{N,\boldsymbol{\theta}}^{-1}\boldsymbol{X}, \tag{1.2.5}$$

where $|\boldsymbol{\Gamma}_{N,\boldsymbol{\theta}}|$ denotes the determinant of $\boldsymbol{\Gamma}_{N,\boldsymbol{\theta}}$. Multiplying (1.2.5) by $-\frac{2}{N}$ and ignoring the constant, we obtain

$$\mathcal{L}_N = \log\sigma^2 + \frac{1}{N}\log|\boldsymbol{\Gamma}_{N,\boldsymbol{\theta}}| + \frac{1}{N\sigma^2}\boldsymbol{X}^T\boldsymbol{\Gamma}_{N,\boldsymbol{\theta}}^{-1}\boldsymbol{X}. \tag{1.2.6}$$

Unfortunately, it is difficult to find the solution $(\widehat{\boldsymbol{\theta}}, \widehat{\sigma}^2)$ that minimizes (1.2.6). To resolve this problem, we shall replace the second and third terms in (1.2.6) by their approximations. Note that by Theorem 4.4.2 in Brockwell and Davis (1991), the spectral density of $\{X_j\}$ is given by

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} s_{\boldsymbol{\theta}}(\lambda), \quad s_{\boldsymbol{\theta}}(\lambda) = \left| \sum_{k=0}^{\infty} a_k(\boldsymbol{\theta}) e^{-ik\lambda} \right|^2, \quad \lambda \in [-\pi, \pi], \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

Define the integrated weighted periodogram as $Q_X(\boldsymbol{\theta}) = \int_{-\pi}^{\pi} \frac{I_X(\lambda)}{s_{\boldsymbol{\theta}}(\lambda)} d\lambda$, where I_X is the periodogram of $\{X_1, \ldots, X_N\}$. Then, it can be proven that, as stated in Giraitis *et al.* (2012) (p. 205),

$$\frac{1}{N} \log |\boldsymbol{\Gamma}_{N,\boldsymbol{\theta}}| \to 0, \quad \frac{1}{N} \boldsymbol{X}^T \boldsymbol{\Gamma}_{N,\boldsymbol{\theta}}^{-1} \boldsymbol{X} \xrightarrow{P} Q_X(\boldsymbol{\theta}).$$

Thus, we can approximate the $\mathcal{L}_N$ in (1.2.6) by

$$\mathcal{L}_N \approx \log \sigma^2 + \frac{Q_X(\boldsymbol{\theta})}{\sigma^2} := \Lambda_N(\boldsymbol{\theta}, \sigma^2), \tag{1.2.7}$$

which leads to the definition of the Whittle estimator.

DEFINITION 1.2.10 Denote by $\boldsymbol{\theta}_0$, σ_0^2 the true values of $\boldsymbol{\theta}$, σ^2, respectively. Whittle estimators of $\boldsymbol{\theta}_0$, σ_0^2 based on observations $\{X_1, \ldots, X_N\}$ are defined as

$$(\widehat{\boldsymbol{\theta}}_W, \widehat{\sigma}_W^2) = \underset{(\boldsymbol{\theta}, \sigma^2) \in \boldsymbol{\Omega}}{\arg\min}\, \Lambda_N(\boldsymbol{\theta}, \sigma^2),$$

where $\boldsymbol{\Omega} = (0, \infty) \times \boldsymbol{\Theta}$. Clearly,

$$\widehat{\boldsymbol{\theta}}_W = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\arg\min}\, Q_X(\boldsymbol{\theta}), \quad \widehat{\sigma}_W^2 = Q_X(\widehat{\boldsymbol{\theta}}_W).$$

Under some regularity conditions, Whittle estimators are consistent and asymptotically follow a normal distribution. However, to obtain these properties, we do not need Gaussianity assumptions on $\{X_j\}$. A comprehensive discussion for Whittle estimators can be found in Giraitis *et al.* (2012) and Beran *et al.* (2013).

The Whittle estimation is a parametric method, whose consistency highly relies on correct specification of the spectral density over the whole range $[-\pi, \pi]$. This is suggested by the fact that under (1.2.4), the spectral density f is known only up to an Euclidean parameter. In other words, any misspecification of f generally leads to inconsistency. Thus, it is necessary to introduce a more robust estimator, with no specific form of the spectral density being assumed. We shall focus on the LW estimation. To begin with, we recall the likelihood approximation (1.2.7). Observe that $\frac{Q_X(\boldsymbol{\theta})}{\sigma^2} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{I_X(\lambda)}{f_X(\lambda)} d\lambda$. On the other hand, the Wiener-Kolmogorov formula for the one-step prediction error gives us

$$\sigma^2 = 2\pi \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f_X(\lambda) d\lambda\right).$$

Hence, ignoring the constant, we may approximate $\mathcal{L}_N$ by

$$\mathcal{L}_{N,LW} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\log f_X(\lambda) + \frac{I_X(\lambda)}{f_X(\lambda)}\right) d\lambda.$$

The essence of the LW estimation is to specify the behavior of the spectral density near the origin, rather than $[-\pi, \pi]$, using the relation (1.2.2). To be specific, we consider low frequencies $|\lambda| \leq \omega_m$, $\frac{m}{N} \to 0$ as $N \to \infty$. In such case, the spectral density $f_X(\lambda)$ can be approximated by $c_f|\lambda|^{-2d}$ and, by symmetry, we can only focus on positive frequencies. Thus, we look at an analogue of $\mathcal{L}_{N,LW}$, defined as

$$\frac{1}{2\pi}\int_0^{\omega_m} \left\{\log\left(c_f\lambda^{-2d}\right) + \frac{I_X(\lambda)}{c_f\lambda^{-2d}}\right\} d\lambda,$$

which can be further approximated by the Riemann sum with partition consisting of low Fourier frequencies, i.e.

$$\frac{1}{2\pi}\cdot\frac{1}{m}\sum_{j=1}^{m}\left\{\log\left(c_f\omega_j^{-2d}\right) + \frac{I_X(\omega_j)}{c_f\omega_j^{-2d}}\right\}. \tag{1.2.8}$$

The local Whittle estimator of (c_f, d) is thus defined by minimizing the objective function (1.2.8). In particular, after standard calculations, the LW estimator for d can be expressed as follows.

DEFINITION 1.2.11 The local Whittle estimator for the memory parameter d is defined as

$$\hat{d}_{LW} = \underset{d \in \left(-\frac{1}{2}, \frac{1}{2}\right)}{\arg\min}\, U_N(d),$$

where

$$U_N(d) = \log\left(\frac{1}{m}\sum_{j=1}^{m} \omega_j^{2d} I_X(\omega_j)\right) - d\left(\frac{2}{m}\sum_{j=1}^{m} \log \omega_j\right),$$

and m is the number of low Fourier frequencies.

Note that the LW estimator for d is scale invariant, that is, the LW estimators based on $\{X_j\}$ and $\{cX_j\}$ are the same for all $c \neq 0$.

The consistency and the asymptotic normality of the LW estimator were established in the seminal article Robinson (1995). A good literature review regarding the LW estimation can be found in Giraitis *et al.* (2012) and Beran *et al.* (2013).

Chapter 2

Theory for the test based on level 1 scores: The simple iid case

2.1 Introduction

As mentioned at the beginning of Chapter 1, the final goal of this dissertation is to establish a test to distinguish a SRD functional time series with a change in the mean function from a certain LRD model. It can be accomplished by precisely estimating the degree of dependence inherent in the original functional observations, which is characterized by the self-similarity parameter. To be specific, the test statistic is based on a normalized version of the LW estimator, which approximately follows a standard normal distribution if the null hypothesis is true and diverges to infinity otherwise. However, it is much more complex to define and to estimate the self-similarity parameter for curve-valued observations. To address this problem, instead of inventing an estimation method suitable for functional observations, we project the functional process (after estimating and subtracting local means) onto the so-called dominant subspace (see, Li *et al.* (2020)), namely, the eigenspace spanned by the first eigenfunction, which results in a scalar process, so that we can apply the traditional LW estimation to it. In this procedure, we benefit from the property that the degree of dependence or strength of the memory signal is preserved under the projection, that is, the self-similarity parameter of the resulting scalar process, i.e. the level 1 scores (projections), is the same as that of the functional one. In other words, dealing with the projected process is just as effective as handling the original functional observations, in terms of estimating the self-similarity parameter, whereas the former is much easier to handle. Nevertheless, the latent change in the mean function of the functional observations can severely perturb and distort the LW estimation. So the first step in the test procedure is to eliminate the impact of the change point in the mean function, which can be done by estimating the location of the change point and removing piecewise means from the original process. Accordingly, we need to guarantee that this step has negligible effect on the LW estimation. However, we found that this needs a lot of efforts if the error functions have a general weak dependence structure, e.g. the $L^p - m$-approximable, partly because of the lack of asymptotic theory for the sums of weighted periodograms in such a general case. Thus, we decided to first work out a special case, where the error functions are iid and then extend the test to the general SRD case.

This chapter is organized as follows. In section 2.2 we describe the whole procedure for implementing the test. Section 2.3 contains the asymptotic results under the null and alternative hypotheses. The most salient ones are that the test statistic is asymptotically normally distributed under the null hypothesis and diverges to infinity under the LRD alternative. In Section 2.4 we carry out a small simulation study which verifies the applicability of the main asymptotic results. Section 2.5 includes the proofs of main results in Section 2.3. Note that this chapter leads to the more general methodology and theory developed in Chapter 3, which means that inevitably there will be some repetitions. The test under the assumption of independence is however simpler, and its asymptotic justification is simpler as well. In order to strike a balance between conciseness and completeness, we shall omit almost identical proofs while keeping the whole test procedure and the asymptotic results. Informative references to the proof in Chapter 3 are given.

2.2 Test procedure

Under the null hypothesis, we assume that the functional time series $\{X_t\}$ satisfies

$$X_t(u) = \begin{cases} \mu(u) + Y_t(u), & 1 \le t \le n^\star, \\ \mu(u) + \delta(u) + Y_t(u), & n^\star + 1 \le t \le N, \end{cases} \tag{2.2.1}$$

where the Y_t are iid random functions and the function δ is a change in mean level at the unknown position $n^\star$. We denote by $\hat{n}$ an estimator of $n^\star$. On the other hand, under the alternative,

$$X_t(u) = \nu(u) + Z_t(u), \quad 1 \le t \le N, \tag{2.2.2}$$

where $\{Z_t\}$ is a stationary mean zero LRD functional time series. The mean levels of the observations are unknown in each scenario, meaning that in practice μ, δ and ν need to be estimated beforehand.

The first step is to eliminate the influence of the change point in the original process. We define the residual functions R_t by removing from the observed functions X_t the estimated means before and after the estimated change point, i.e.,

$$R_t = \begin{cases} X_t - \frac{1}{\hat{n}} \sum_{s=1}^{\hat{n}} X_s, & 1 \leq t \leq \hat{n}, \\ X_t - \frac{1}{N-\hat{n}} \sum_{s=\hat{n}+1}^{N} X_s, & \hat{n}+1 \leq t \leq N. \end{cases}$$

Since the Y_t are not observable, a hypothesis test must be based on the R_t.

Next, we estimate a suitable subspace and then project the residuals onto it. Since $\{Y_t\}$ are iid, the covariance function can fully capture their dependence structure. Consider the covariance kernel of the sequence $\{Y_t\}$ defined as

$$c_Y(u, v) = E\{Y_0(u)Y_0(v)\} \tag{2.2.3}$$

and its estimator

$$\hat{c}_R(u, v) = \frac{1}{N} \sum_{t=1}^{N} R_t(u)R_t(v). \tag{2.2.4}$$

(Observe that $\overline{R}_N = \frac{1}{N} \sum_{t=1}^{N} R_t = 0$.)

We calculate the first eigenfunction $\widehat{\psi}_1^{(R)}$ of the kernel $\hat{c}_R$ and the corresponding scores $\widehat{\xi}_{1t}^{(R)} = \langle R_t, \widehat{\psi}_1^{(R)} \rangle$. Based on the $\widehat{\xi}_{1t}^{(R)}$, we compute the local Whittle estimator (LWE) $\widehat{H}_1^{(R)}$. Our test statistic is defined as

$$T_1^{(R)} = 2\sqrt{m} \left(\widehat{H}_1^{(R)} - \frac{1}{2} \right), \tag{2.2.5}$$

where m is the number of low frequencies used in the LW estimation. Recall that the LW estimator for $d = H - \frac{1}{2}$ is given in Definition 1.2.11. We reject H_0 if $T_1^{(R)} > q_{1-\alpha}$, where $q_{1-\alpha}$ is the $(1-\alpha)$th quantile of the standard normal distribution.

2.3 Asymptotic justification

2.3.1 Asymptotic justification under the null hypothesis

ASSUMPTION 2.3.1 *The functional time series* $\{X_t\}$ *follows model* (2.2.1)*, where the* Y_t *are iid mean zero random functions in* $L^2([0,1])$ *with* $E\|Y_1\|^4 < \infty$.

The level 1 scores of $\{Y_t\}$ are defined as

$$\xi_{1t}^{(Y)} = \left\langle Y_t, \psi_1^{(Y)} \right\rangle = \int_0^1 Y_t(u)\psi_1^{(Y)}(u)du, \tag{2.3.1}$$

where $\psi_1^{(Y)}$ is the first eigenfunction of c_Y. Since the Y_t are not observable, we cannot directly conduct the hypothesis testing based on $\{Y_t\}$. In other words, the LW estimation based on the scores $\{\xi_{1t}^{(Y)}\}$ is infeasible. So we must find a proper proxy or approximation for $\{\xi_{1t}^{(Y)}\}$. Not surprisingly, a good candidate is the estimated scores $\widehat{\xi}_{1t}^{(R)}$. Recall that we denote by $\widehat{H}_1^{(R)}$ the LWE based on the scores $\widehat{\xi}_{1t}^{(R)}$ computed using the residuals R_t. Our objective is thus to show that

$$2\sqrt{m}\left(\widehat{H}_1^{(R)} - \frac{1}{2}\right) \xrightarrow{d} \mathcal{N}(0,1).$$

This can be achieved by showing that the periodogram of the R_t is close to that of the Y_t with difference that is negligible with respect to the LW estimation. To ensure this, we must impose suitable assumptions, including the separation of the eigenvalues of c_Y, the consistency of the change point estimator and an increasing rate at which the number of low frequencies m increases with the sample size N.

ASSUMPTION 2.3.2 *The eigenvalues* $\lambda_j^{(Y)}$ *of the covariance function* c_Y *satisfy*

$$\lambda_1^{(Y)} > \lambda_2^{(Y)}.$$

ASSUMPTION 2.3.3 *For some* $0 < \theta < 1$, $n^\star = \lfloor N\theta \rfloor$ *and*

$$|\hat{n} - n^\star| = O_P(1). \tag{2.3.2}$$

ASSUMPTION 2.3.4 *The number of Fourier frequencies* m *satisfies the following condition:*

$$\frac{1}{m} + \frac{m(\log m)^2}{N} \to 0, \quad \text{as } N \to \infty.$$

Then we can establish the following theorem whose proof is given in Section 2.5.2.

THEOREM 2.3.1 *Under Assumptions 2.3.1–2.3.4,*

$$T_1^{(R)} = 2\sqrt{m}\left(\widehat{H}_1^{(R)} - \frac{1}{2}\right) \stackrel{d}{\to} \mathcal{N}(0,1). \tag{2.3.3}$$

A key step in proving Theorem 2.3.1 is the following result, whose proof is in Section 2.5.1.

THEOREM 2.3.2 *Under Assumptions 2.3.1 and 2.3.3,*

$$\|\widehat{C}_R - C_Y\|_{\mathcal{S}} = O_P(N^{-1/2}),$$

where C_Y *is the covariance operator of the covariance function* (2.2.3).

The following corollary to Theorem 2.3.2 follows immediately from Lemma 2.3 in Horváth and Kokoszka (2012).

COROLLARY 2.3.1 *If the Assumptions of Theorem 2.3.2 hold and, in addition, Assumption 2.3.2 holds, then*

$$\left\|\widehat{\psi}_1^{(R)} - \psi_1^{(Y)}\right\| = O_P(N^{-1/2}). \tag{2.3.4}$$

2.3.2 Consistency of the tests

In this section, we aim to establish the consistency of the test under the LRD alternative. Specifically, we shall demonstrate that the test statistic (2.2.5) diverges to infinity if the alternative hypothesis is true. Recall that we divided the discussion of the null hypothesis into two cases according to the assumptions on the error functions Y_t, i.e. iid versus weak dependence. Likewise, we continue to use a similar framework in terms of the LRD alternative. At the first glance, alterations of the null hypothesis should have no impact on the alternative hypothesis. However, in practical implementation, changes in the assumptions on the error functions are reflected by the way we estimate their dependence structure, i.e. covariance function versus long-run covariance function, and thus ultimately leading to a difference in the test procedure. Fortunately, the asymptotic results for the LRD alternative are much alike despite the discrepancy in the test procedure or in the null hypothesis, partly because the assumptions in the LRD setting are almost the same. We shall see later that in each scenario, i.e. the iid and weakly dependent cases, we assume the LRD model to have the form of a functional linear process, as in Li *et al.* (2021).

Nevertheless, there is yet another subtle difference stemming from different characterizations of the true dependence structure of the error functions Z_t. Recall that under H_0 we have proven that the sample covariance operator $\widehat{C}_R$ is close to the covariance operator C_Y in the sense that their distance has a sufficient decay rate, which ensures that replacing the level 1 scores by their estimated versions has negligible effect on the LW estimation. Similarly, under H_A we need an analogous result to guarantee the consistency of the test statistic. To accomplish this task, the definition of the covariance operator of Z_t should be consistent with the form of the

sample covariance operator $\widehat{C}_R$. We first display an assumption on the error functions Z_t.

ASSUMPTION 2.3.5 *The process* $\{Z_t\}$ *in* (2.2.2) *has the form*

$$Z_t = \sum_{j=0}^{\infty} b_j \eta_{t-j}. \tag{2.3.5}$$

The scalar impulse responses b_j *satisfy* $b_j \sim j^{H_0 - \frac{3}{2}}$ *for an unknown* $H_0 \in \left(\frac{1}{2}, 1\right)$. *The functions* $\eta_t \in L^2$ *are iid, have mean zero and satisfy* $E\|\eta_t\|^4 < \infty$.

Note that the dependence structure of the Z_t is completely specified by the coefficients $\{b_j\}$ and is independent of $\{\eta_t\}$. We consider the covariance function of Z_t defined by

$$c_Z(u,v) = E[Z_0(u)Z_0(v)]. \tag{2.3.6}$$

Note that $\sum_{j=0}^{\infty} b_j^2 < \infty$, and

$$\begin{aligned} c_Z(u,v) &= E\left[\sum_{j=0}^{\infty} b_j \eta_{-j}(u) \cdot \sum_{s=0}^{\infty} b_s \eta_{-s}(v)\right] \\ &= \sum_{j=0}^{\infty} b_j^2 \cdot E[\eta_0(u)\eta_0(v)]. \end{aligned}$$

Thus, we can conclude that c_Z is in $L^2([0,1]\times[0,1])$. In order to obtain the desirable consistency of $\widehat{C}_R$, we need to restrict the limiting behavior of the change point estimator $\hat{n}$.

ASSUMPTION 2.3.6 *The change point estimator* $\hat{n}$ *satisfies*

$$\frac{\hat{n}}{N} \xrightarrow{P} \xi,$$

for some random variable ξ *taking values in* $(0,1)$.

THEOREM 2.3.3 *Let* C_Z *be the covariance operator defined by* (2.3.6). *Under Assumptions 2.3.5–2.3.6,*

$$\|\widehat{C}_R - C_Z\|_{\mathcal{S}} = o_P(1),$$

where $\widehat{C}_R$ *is defined in Theorem 2.3.2 and* H_0 *is given in Assumption 2.3.5.*

With additional separation condition on the eigenvalues of c_Z, we can establish the consistency of the first eigenfunction, which follows immediately from Lemma 2.3 in Horváth and Kokoszka (2012).

Let $\psi_j^{(Z)}$ be the eigenfunction corresponding to the j-th largest eigenvalue $\lambda_j^{(Z)}$ of c_Z and $\xi_{1t}^{(Z)}$ be the first scores of $\{Z_t\}$, that is,

$$\int_0^1 c_Z(u,v)\psi_j^{(Z)}(v)dv = \lambda_j^{(Z)}\psi_j^{(Z)}(u), \quad \xi_{1t}^{(Z)} = \langle Z_t, \psi_1^{(Z)}\rangle.$$

ASSUMPTION 2.3.7 *The eigenvalues $\lambda_j^{(Z)}$ of the covariance function c_Z satisfy*

$$\lambda_1^{(Z)} > \lambda_2^{(Z)}.$$

COROLLARY 2.3.2 *If the Assumptions of Theorem 2.3.3 hold and, in addition, Assumption 2.3.7 holds, then,*

$$\left\|\widehat{\psi}_1^{(R)} - \psi_1^{(Z)}\right\| = o_P(1). \tag{2.3.7}$$

There is one more step before we can achieve our final goal of establishing the consistency of the test: We need to add some regularity conditions on the first scores of $\{Z_t\}$, the transfer function and the increasing rate of the number of low frequencies m.

ASSUMPTION 2.3.8 *(i) The spectral density of the first scores $\xi_{1t}^{(Z)}$ of the series $\{Z_t\}$, denoted by $f_1^{(Z)}(\lambda)$, satisfies*

$$f_1^{(Z)}(\lambda) \sim D_0\lambda^{1-2H_0}\left(1 + O\left(\lambda^{\vartheta}\right)\right),\ as\ \lambda \to 0+,$$

where D_0 is an unknown positive constant, $\vartheta \in (0,2]$ and $H_0 \in \left(\frac{1}{2},1\right)$ is the same as in Assumption 2.3.5.

(ii) The transfer function $B(\lambda) := \sum_{j=0}^{\infty} b_j e^{ij\lambda}$ is differentiable on $(0,\lambda_0)$ for some $\lambda_0 > 0$, with

$$\frac{d}{d\lambda}B(\lambda) = O\left(\frac{|B(\lambda)|}{\lambda}\right),\ as\ \lambda \to 0+,$$

where the weights b_j are defined in Assumption 2.3.5.

ASSUMPTION 2.3.9 *The number of Fourier frequencies m satisfies the following condition:*

$$\frac{1}{m} + \frac{m^{1+2\vartheta}(\log m)^2}{N^{2\vartheta}} \to 0, \quad \text{as } N \to \infty,$$

where ϑ is defined in Assumption 2.3.8.

Note that Assumption 2.3.9 is a stronger version of Assumption 2.3.4.

THEOREM 2.3.4 *Under Assumptions 2.3.5–2.3.9, $\widehat{H}^{(R)} \xrightarrow{P} H_0$.*

Since $H_0 > \frac{1}{2}$, we obtain the following corollary.

COROLLARY 2.3.3 *Under the Assumptions of Theorem 2.3.4, $T_1^{(R)} \xrightarrow{P} \infty$, where $T_1^{(R)}$ is defined in* (2.2.5).

2.4 A small simulation study

Under H_0, the functional observations X_t are generated by the model (2.2.1) with $\mu(u) = 0$ and $\{Y_t\}$ a sequence of iid standard Brownian motions on $[0, 1]$. We assume that $\delta(u) = \delta$. We consider $\delta = 0.25$ and $\delta = 0.50$. We consider two change point locations $\theta = 0.25$ and $\theta = 0.50$. The change point estimator (1.1.14) is applied in computing the test statistic.

Under H_A, we consider the functional long-range dependent process defined as follows, cf. (2.3.5),

$$X_t = \sum_{j=1}^{L} g(j) j^{H_0 - \frac{3}{2}} B_{t-j}, \quad H_0 \in \left(\frac{1}{2}, 1\right), \tag{2.4.1}$$

where the function g satisfies $g(x) \to 1$, as $x \to \infty$, and the B_t are iid standard Brownian motions. The function g characterizes the departure of an observed functional time series from the ideal case of $b_j = j^{H_0 - 3/2}$, cf. Assumption 2.3.5. We used the truncation level $L = 1,500$ because this is the largest series length we consider. We considered the following functions g:

$$g_1(j) = 1, \quad g_2(j) = \exp\left(-0.01 \cdot j - \frac{1}{2}\right) \sin\left(\frac{2\pi j}{15}\right) + 1, \tag{2.4.2}$$

whose plots are displayed in Figure 2.1.

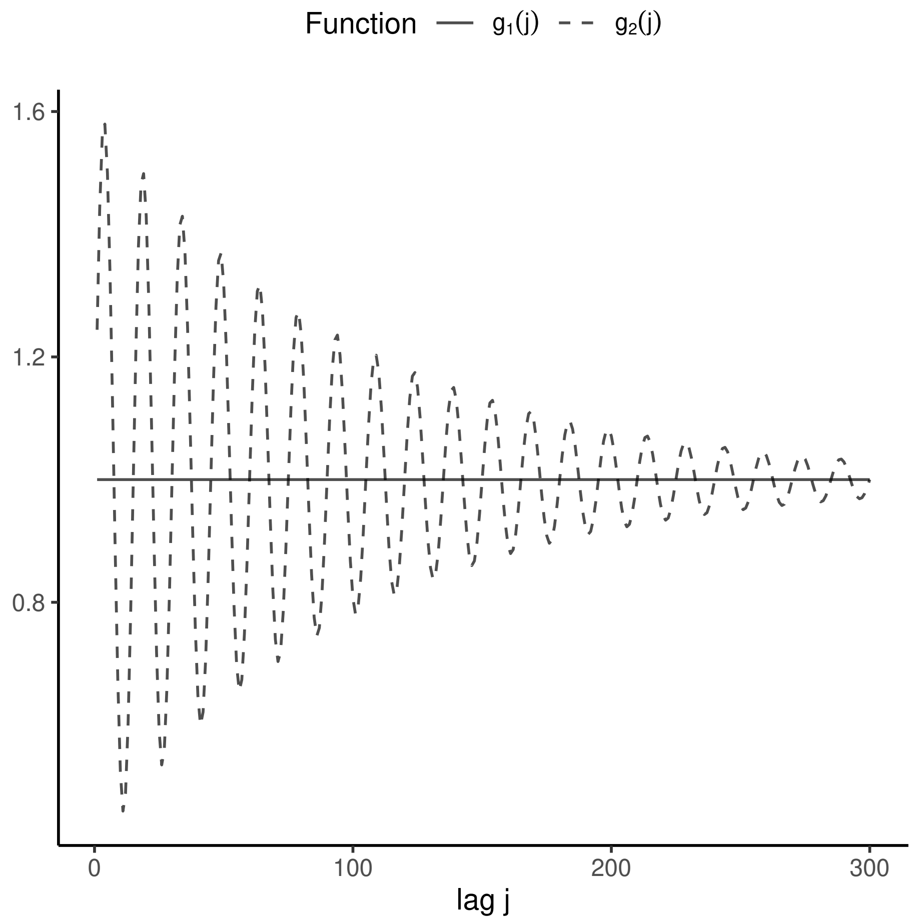

Figure 2.1: The two functions g_1, g_2 in (2.4.2).

The results in Tables 2.1 and 2.2 indicate that test achieves satisfactory size and is asymptotically consistent. The choice of $m = N^{0.6}$ has the best overall performance in terms of the empirical size. As the change point moves away from $N/2$, the empirical size tends to increase above the level for $N/2$. If the change point is at $N/4$, the increase is roughly by 0.5 to 2.0 percent, depending on the nominal size. The function g may have a severe impact on the empirical power, when the sample size is not sufficiently large or the degree of dependence is not strong enough.

Table 2.1: Empirical sizes (in percent), based on $2,000$ replications. The true change point is defined by $n^\star = \lfloor N\theta \rfloor$.

Nominal size	1.0			5.0			10.0		
N	500	1000	1500	500	1000	1500	500	1000	1500
				$\delta = 0.25, \theta = 0.50$					
$m = N^{0.55}$	1.3	0.7	1.1	4.0	3.4	3.6	6.9	6.1	6.5
$m = N^{0.60}$	1.0	0.7	1.0	4.2	2.9	3.5	7.3	6.5	7.0
$m = N^{0.65}$	0.9	0.6	0.4	3.6	3.2	2.9	6.3	5.7	6.1
				$\delta = 0.50, \theta = 0.50$					
$m = N^{0.55}$	1.5	0.9	1.2	4.7	4.2	4.0	8.0	6.8	7.3
$m = N^{0.60}$	1.1	0.7	1.1	4.9	3.5	4.0	8.1	7.1	7.8
$m = N^{0.65}$	1.0	0.8	0.6	4.1	3.6	3.3	7.1	6.5	6.9
				$\delta = 0.25, \theta = 0.25$					
$m = N^{0.55}$	1.2	1.0	1.6	5.0	4.5	4.6	7.8	7.7	8.1
$m = N^{0.60}$	1.2	0.7	1.2	4.4	4.0	4.2	8.0	7.7	8.1
$m = N^{0.65}$	1.0	0.6	0.9	4.4	3.6	3.0	7.5	6.8	7.2
				$\delta = 0.50, \theta = 0.25$					
$m = N^{0.55}$	1.8	1.4	1.8	6.1	5.5	5.4	10.0	9.8	8.9
$m = N^{0.60}$	1.9	1.2	1.5	6.5	4.6	4.9	10.7	9.2	9.5
$m = N^{0.65}$	1.6	1.2	0.9	5.4	5.1	4.0	9.2	8.4	8.4

Table 2.2: Empirical powers (in percent) for the two functions g in (2.4.1). The number of replications is the same as in Table 2.1.

Nominal size	1.0			5.0			10.0		
N	500	1000	1500	500	1000	1500	500	1000	1500
				$g_1, H_0 = 0.6$					
$m = N^{0.55}$	68.4	82.6	90.0	83.7	92.6	95.8	89.4	95.4	97.7
$m = N^{0.60}$	89.9	97.4	99.0	96.4	98.9	99.7	98.2	99.6	99.9
$m = N^{0.65}$	98.8	99.9	100.0	99.6	100.0	100.0	99.7	100.0	100.0
				$g_1, H_0 = 0.9$					
$m = N^{0.55}$	96.0	99.4	99.9	98.8	99.9	100.0	99.5	99.9	100.0
$m = N^{0.60}$	99.7	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0
$m = N^{0.65}$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
				$g_2, H_0 = 0.6$					
$m = N^{0.55}$	30.2	56.3	69.3	48.3	73.5	85.1	58.4	81.6	90.1
$m = N^{0.60}$	45.3	66.9	86.7	66.0	83.8	94.8	75.5	90.1	97.4
$m = N^{0.65}$	90.9	93.0	94.6	96.6	97.0	98.3	98.3	98.2	99.0
				$g_2, H_0 = 0.9$					
$m = N^{0.55}$	63.7	95.0	98.7	79.9	98.1	99.7	86.3	98.8	99.8
$m = N^{0.60}$	66.6	95.9	99.8	82.1	98.7	100.0	87.6	99.3	100.0
$m = N^{0.65}$	98.1	98.8	99.6	99.1	99.8	100.0	99.6	99.9	100.0

2.5 Proofs of main results and technical lemmas

Since the iid assumption imposed on the error functions Y_t is much stronger than the weak dependence assumption that we will discuss in next chapter, we just skip some details of the proofs here. Instead, we refer to the complete and coherent proofs in Section 3.6 when necessary, and focus only on the prominent parts that can be simplified in the iid scenario.

2.5.1 Proof of Theorem 2.3.2

By Theorem 3.6.1 and Assumption 2.3.3, we can immediately obtain

$$\left\|\frac{1}{N}\sum_{t=1}^{N}Y_t\right\|, \ \left\|\frac{1}{n^\star}\sum_{t=1}^{n^\star}Y_t\right\|, \ \left\|\frac{1}{\hat{n}}\sum_{t=1}^{\hat{n}}Y_t\right\| = O_P(N^{-1/2}). \tag{2.5.1}$$

Consider the series

$$\widetilde{R}_t = \begin{cases} X_t - \frac{1}{n^\star}\sum_{1\le s\le n^\star} X_s, & 1\le t\le n^\star, \\ X_t - \frac{1}{N-n^\star}\sum_{n^\star<s\le N} X_s, & n^\star+1\le t\le N, \end{cases}$$

and the kernel

$$\tilde{c}_R(u,v) = \frac{1}{N}\sum_{s=1}^{N}\widetilde{R}_s(u)\widetilde{R}_s(v).$$

Let $\widetilde{C}_R$ be the corresponding operator of $\tilde{c}_R$. Then,

$$\|\widehat{C}_R - C_Y\|_{\mathcal{S}} \le \|\widehat{C}_R - \widetilde{C}_R\|_{\mathcal{S}} + \|\widetilde{C}_R - C_Y\|_{\mathcal{S}}.$$

The claim thus follows from Lemmas 2.5.1–2.5.2.

LEMMA 2.5.1 *Under the Assumptions of Theorem 2.3.1,*

$$\|\widehat{C}_R - \widetilde{C}_R\|_{\mathcal{S}} = O_P(N^{-1/2}).$$

PROOF: Following a similar but simpler argument as in the proof of Lemma 3.6.1, we readily have

$$\begin{aligned}\|\widehat{C}_R - \widetilde{C}_R\|_{\mathcal{S}} &= \left\{ \iint \left(\hat{c}_R(u,v) - \tilde{c}_R(u,v)\right)^2 \, dudv \right\}^{1/2} \\ &= O_P(N^{-1/2}).\end{aligned}$$

∎

LEMMA 2.5.2 *Under the Assumptions of Theorem 2.3.1,*

$$\|\widetilde{C}_R - C_Y\|_{\mathcal{S}} = O_P(N^{-1/2}).$$

PROOF: Define

$$\overline{Y}_1 = \frac{1}{n^\star} \sum_{i=1}^{n^\star} Y_i, \quad \overline{Y}_2 = \frac{1}{N - n^\star} \sum_{i=n^\star+1}^{N} Y_i.$$

Then,

$$\begin{aligned}\tilde{c}_R(u,v) - c_Y(u,v) &= \frac{1}{N} \sum_{s=1}^{N} \widetilde{R}_s(u)\widetilde{R}_s(v) - E[Y_0(u)Y_0(v)] \\ &= \frac{1}{N} \sum_{s=1}^{N} \left\{ Y_s(u)Y_s(v) - E[Y_0(u)Y_0(v)] \right\} - \frac{n^\star}{N}\overline{Y}_1(u)\overline{Y}_1(v) \\ &\quad - \frac{N - n^\star}{N}\overline{Y}_2(u)\overline{Y}_2(v) \\ &= \mathcal{Y}_1^\star(u,v) - \mathcal{Y}_2^\star(u,v) - \mathcal{Y}_3^\star(u,v),\end{aligned}$$

where

$$\begin{aligned}\mathcal{Y}_1^\star(u,v) &= \frac{1}{N} \sum_{s=1}^{N} \left\{ Y_s(u)Y_s(v) - E[Y_0(u)Y_0(v)] \right\}, \\ \mathcal{Y}_2^\star(u,v) &= \frac{n^\star}{N}\overline{Y}_1(u)\overline{Y}_1(v), \\ \mathcal{Y}_3^\star(u,v) &= \frac{N - n^\star}{N}\overline{Y}_2(u)\overline{Y}_2(v).\end{aligned}$$

By Theorem 2.5 in Horváth and Kokoszka (2012),

$$\left\{\iint \mathcal{Y}_1^\star(u,v)^2 dudv\right\}^{1/2} = O_P(N^{-1/2}).$$

By Assumption 2.3.3 and (2.5.1),

$$\left\{\iint \mathcal{Y}_2^\star(u,v)^2 dudv\right\}^{1/2}, \left\{\iint \mathcal{Y}_3^\star(u,v)^2 dudv\right\}^{1/2} = O_P(N^{-1}).$$

It follows that

$$\|\widetilde{C}_R - C_Y\|_{\mathcal{S}} = \left\{\iint \left(\tilde{c}_R(u,v) - c_Y(u,v)\right)^2 dudv\right\}^{1/2} = O_P(N^{-1/2}).$$

∎

2.5.2 Proof of Theorem 2.3.1

In fact, Theorem 2.3.1 is a special case of Theorem 3.3.4, in which the random functions Y_t are allowed to be weakly dependent. It turns out that the big picture of the proofs of the two theorems is roughly the same and differences are reflected by some technical lemmas. Specifically, Lemmas 3.6.4, 3.6.5 and 3.6.14 hold under the assumption of iid errors with the unknown parameter α strengthened to be $\frac{1}{2}$. On the other hand, since the Y_t are iid, the level-1 scores defined in (2.3.1) are also iid. Consequently, we are able to apply existing asymptotic results for the sums of weighted periodograms without imposing additional linear structure on the representation of the Y_t. For these reasons, Theorem 2.3.1 follows immediately from Theorem 3.3.4, so we just focus on the discrepancies in specific lemmas while withholding the main proof.

Recall that the level 1 scores $\xi_{1t}^{(Y)}$ are defined in (2.3.1). Note that $\{\xi_{1t}^{(Y)}\}$ itself is a linear process with spectral density $f_1(\lambda) = \frac{\text{Var}(\xi_{10}^{(Y)})}{2\pi}$ and transfer function $\widetilde{B}(e^{-i\lambda}) = 1$ for all $\lambda \in [0,\pi]$. Moreover, $E|\xi_{1t}^{(Y)}|^4 \leq E\|Y_0\|^4 < \infty$ for all t. Thus, the assumptions of Theorems 6.2.3 and 6.2.5

in Giraitis *et al.* (2012) are satisfied. Following the proofs of Theorems 3.3.2 and 3.3.3, we readily have the following lemmas.

Let $I_1^{(Y)}(\omega_\ell) = \frac{1}{2\pi N}\left|\sum_{t=1}^N \xi_{1t}^{(Y)} e^{-it\omega_\ell}\right|^2$ and $G_0 = f_1(0) = \frac{\text{Var}(\xi_{10}^{(Y)})}{2\pi}$.

LEMMA 2.5.3 *Under the Assumptions of Theorem 2.3.1,*

$$\frac{1}{m}\sum_{\ell=1}^{m}\left(\frac{I_1^{(Y)}(\omega_\ell)}{G_0} - 1\right) = O_P\left(m^{-1/2}\right), \quad \textit{as } N \to \infty.$$

LEMMA 2.5.4 *Under the Assumptions of Theorem 2.3.1,*

$$\frac{1}{\sqrt{m}}\sum_{\ell=1}^{m}\nu_{\ell,m}\left(\frac{I_1^{(Y)}(\omega_\ell)}{G_0} - 1\right) \xrightarrow{d} \mathcal{N}(0,1), \quad \textit{as } N \to \infty,$$

where

$$\nu_{\ell,m} = \log \ell - \frac{1}{m}\sum_{j=1}^{m}\log j.$$

2.5.3 Proofs of Theorems 2.3.3 and 2.3.4

PROOF OF THEOREM 2.3.3: Define an auxiliary covariance function as

$$\hat{c}_Z(u,v) = \frac{1}{N}\sum_{t=1}^{N} Z_t(u)Z_t(v). \tag{2.5.2}$$

Note that

$$\begin{aligned}\hat{c}_R(u,v) - c_Z(u,v) &= \hat{c}_R(u,v) - \hat{c}_Z(u,v) + \hat{c}_Z(u,v) - c_Z(u,v) \\ &= \tilde{r}_1(u,v) + \tilde{r}_2(u,v),\end{aligned}$$

where $\tilde{r}_1(u,v) = \hat{c}_R(u,v) - \hat{c}_Z(u,v)$ and $\tilde{r}_2(u,v) = \hat{c}_Z(u,v) - c_Z(u,v)$. We shall show that

$$\iint \tilde{r}_1^2(u,v)dudv,\ \iint \tilde{r}_2^2(u,v)dudv = o_P(1).$$

Observe that

$$R_t = \begin{cases} Z_t - \overline{Z}_1, & 1 \le t \le \hat{n}, \\ Z_t - \overline{Z}_2, & \hat{n}+1 \le t \le N, \end{cases}$$

where $\overline{Z}_1 = \frac{1}{\hat{n}} \sum_{s=1}^{\hat{n}} Z_s$ and $\overline{Z}_2 = \frac{1}{N-\hat{n}} \sum_{s=\hat{n}+1}^{N} Z_s$. By some simple calculations, we obtain

$$\tilde{r}_1(u,v) = -\frac{\hat{n}}{N}\overline{Z}_1(u)\overline{Z}_1(v) - \frac{N-\hat{n}}{N}\overline{Z}_2(u)\overline{Z}_2(v),$$

and

$$\iint \tilde{r}_1^2(u,v)dudv = \left(\frac{\hat{n}}{N}\right)^2 \|\overline{Z}_1\|^4 + \left(\frac{N-\hat{n}}{N}\right)^2 \|\overline{Z}_2\|^4 + \frac{2\hat{n}(N-\hat{n})}{N^2}\langle \overline{Z}_1, \overline{Z}_2\rangle^2.$$

By Assumptions 2.3.5 and 2.3.6 and Corollary 1 in Li *et al.* (2020), we may have

$$\|\overline{Z}_1\|,\ \|\overline{Z}_2\| = O_P\left(N^{H_0-1}\right). \tag{2.5.3}$$

Thus,

$$\iint \tilde{r}_1^2(u,v)dudv = O_P\left(N^{4(H_0-1)}\right) = o_P(1).$$

On the other hand, using an analogous argument as in Li *et al.* (2020) (see supplement material p. 8–9), we can immediately obtain

$$\iint \tilde{r}_2^2(u,v)dudv = o_P(1),$$

which completes the proof of Theorem 2.3.3.

∎

PROOF OF THEOREM 2.3.4: Using Corollary 2.3.2 and exactly the same arguments as in Theorem 3.3.6, we can conclude the proof of Theorem 2.3.4.

∎

Chapter 3

Theory for the test based on level 1 scores under the assumption of weakly dependent errors [1]

3.1 Introduction

Stationary time series used to model long-range dependence (LRD), or long memory, exhibit spurious changes in mean that Benoit Mandelbrot called the Joseph effect, referring to periods of famine and plenty. This is a fundamental characteristic of all stationary long memory models, see e.g. Beran *et al.* (2013). Several approaches have been proposed to distinguish, informally or through formal statistical significance tests, whether for an observed time series an LRD model or a weakly dependent model with changes in mean is more appropriate. Accounts of such research are given in Berkes *et al.* (2006) and Baek and Pipiras (2012). The correct choice has practical consequences. For example, if an LRD model is valid, a long stretch of time series should be used for prediction or hypothesis tests to use the information in the data fully. On the other hand, if the mean changes, only the most recent data stretch should used, or residuals obtained after subtracting the changing means.

The objective of this paper is to develop a significance test aimed at distinguishing between long-range dependence and weak dependence with changes in mean in the context of functional data. Functional data analysis (FDA) offers effective tools for the analysis of curve-valued time series and other data structures where observational units are functions. Many monographs are now available, e.g. Bosq (2000), Ramsay and Silverman (2005), Ferraty and Vieu (2006), Shi and Choi (2011), Horváth and Kokoszka (2012), Hsing and Eubank (2015) and Kokoszka and Reimherr (2017). The last reference offers a general and accessible introduction to FDA, with chapters 10 and 11 providing sufficient mathematical background needed to understand this paper.

---

[1] This chapter was originally published in *Journal of Time Series Analysis* (see, https://doi.org/10.1111/jtsa.12723).

The problem and the proposed solution are described in Section 3.2. Our approach builds on the work of Baek and Pipiras (2012) and Li *et al.* (2021). Two key issues must be addressed simultaneously. Baek and Pipiras (2012) studied scalar time series and needed not be concerned with the estimation of functional scores. In the context of this paper, the local Whittle estimator cannot be applied to observations because the observations are functions. It must be applied to coefficients of projections on an optimal direction. This direction is however unknown and must be estimated. Incorporating the uncertainty inherent in the estimation process requires new asymptotic theory. Li *et al.* (2021) studied the local Whittle estimator based on projections similar to those we use, but they assumed an LRD model without a change point. We need a similar theory for functional residuals obtained after subtracting the mean functions before and after a potential change point. This also requires additional steps in the asymptotic analysis. Our work has led us to several results that could be of independent interest and be potentially useful in other contexts. They are presented in Section 3.3. In particular, Theorem 3.3.1 establishes a bound on the difference between the long-run covariance operator of a short memory functional time series and its estimator based on functional residuals. The rate of convergence is different than the usual $N^{-1/2}$ rate. It is valid under a very general weak dependence assumption that covers practically all functional time series models, including nonlinear models. The proofs of the results of Section 3.3 are complex, so we decided to place them in online Supporting Information that has no page limit and where we can present all details. The Supporting Information also describes in broad terms how computations were performed and contains additional tables. Complete and commented code is also available for downloads on the journal's website. The printed part of the paper, explains the test procedure, the asymptotic framework, shows how the test performs on simulated data and to what conclusions it leads for functional time series derived from intraday price data. The latter two aspects are presented in Sections 3.4 and 3.5, respectively. We hope that such a presentation will allow readers to quickly become acquainted with the contribution of the paper, while also providing detailed mathematical derivations for readers who wish to pursue them.

3.2 Problem formulation and the test

We assume that we observe a time series of functions $X_1, X_2, \ldots, X_N$. These functions can be defined on an abstract domain $\mathcal{U}$. If $\mathcal{U}$ is a Polish space, (complete and separable metric space), then the Hilbert space $L^2(\mathcal{U})$ is separable (has a countable orthonormal basis), and so is isomorphic to the space $L^2([0,1])$ of square integrable functions on the unit interval. For this reason, we assume that the X_t are members of the space $L^2 := L^2([0,1])$, which also provides a more concrete interpretation of the time series we study. In L^2, the inner product and the norm it generates are defined by

$$\langle f, g \rangle = \int_0^1 f(u)\bar{g}(u)du, \quad \|f\|^2 = \int_0^1 |f(u)|^2 du.$$

We assume that under the null hypothesis

$$X_t(u) = \begin{cases} \mu(u) + Y_t(u), & 1 \le t \le n^\star, \\ \mu(u) + \delta(u) + Y_t(u), & n^\star + 1 \le t \le N, \end{cases} \tag{3.2.1}$$

where the Y_t are mean zero weakly dependent functions forming a stationary sequence in L^2 and the function δ is a change in mean level at an unknown change point $n^\star$. Under the alternative,

$$X_t(u) = \nu(u) + Z_t(u), \quad 1 \le t \le N, \tag{3.2.2}$$

where $\{Z_t\}$ is a stationary mean zero LRD functional time series. The means μ, ν and the jump δ are unknown. Precise assumptions on the functions Y_t, Z_t and other objects are stated in Section 3.3. We note here that the Y_t are short-range dependent in a sense that suitable univariate projections are short range dependent processes. There are many ways of quantifying short-range or weak dependence, and we use more general ways for certain results and more specific ways for other results. The functions Z_t are long-range dependent in the sense that their projections on the univariate spaces we use to construct the test are scalar long memory processes with the Hurst (or self-similarity) parameter $H > 1/2$. The variance of the sample mean of the projections of the Y_t

behaves like N^{-1}, while for the Z_t like $N^{2H-2} >> N^{-1}$. This implies that the sample mean of an LRD process is "less stable" and these processes, even though they are stationary, exhibit behavior that suggests a changing population mean. These facts are discussed in many textbooks, e.g. in Giraitis *et al.* (2012) and Beran *et al.* (2013).

We now proceed with the description of the test we recommend. Comments on its modifications are presented at the end of this section.

Denote by $\hat{n}$ a change point estimator. We use the estimator

$$\hat{n} = \underset{1<k<N}{\operatorname{argmax}} \left\| \sum_{i=1}^{k} X_i - \frac{k}{N} \sum_{i=1}^{N} X_i \right\|^2, \tag{3.2.3}$$

where $\underset{x}{\operatorname{argmax}}\, f(x) = \inf \left\{ x : f(x) = \sup_{y} f(y) \right\}$.

A similar estimator studied by Aue *et al.* (2009) can also be used. Under (3.2.1), $\hat{n}$ estimates $n^{\star}$. Under (3.2.2) it does not estimate any change point, but must be computed as part of the procedure. We define the residual functions R_t by .

$$R_t(u) = \begin{cases} X_t(u) - \frac{1}{\hat{n}} \sum_{s=1}^{\hat{n}} X_s(u), & 1 \le t \le \hat{n}, \\ X_t(u) - \frac{1}{N-\hat{n}} \sum_{s=\hat{n}+1}^{N} X_s(u), & \hat{n}+1 \le t \le N. \end{cases}$$

Since the Y_t in (3.2.1) are not observable, a hypothesis test must be based on the R_t. The test uses eigenfunctions of the estimated long-run covariance kernel

$$\hat{c}_R(u,v) = \widehat{\gamma}_{R,0}(u,v) + \sum_{t=1}^{N-1} K\left(\frac{t}{h}\right) \{\widehat{\gamma}_{R,t}(u,v) + \widehat{\gamma}_{R,t}(v,u)\}, \tag{3.2.4}$$

where $K(x) = 1 - |x|$, $|x| \leq 1$, is the Bartlett kernel and the autocovariances $\widehat{\gamma}_{R,t}(u,v)$ are given by

$$\widehat{\gamma}_{R,t}(u,v) = \frac{1}{N}\sum_{s=t+1}^{N}\left(R_s(u) - \overline{R}_N(u)\right)\left(R_{s-t}(v) - \overline{R}_N(v)\right)$$
$$= \frac{1}{N}\sum_{s=t+1}^{N} R_s(u)R_{s-t}(v), \quad 0 \leq t \leq N-1.$$

The last equality holds because $\overline{R}_N = \frac{1}{N}\sum_{t=1}^{N} R_t = 0$. The first equality shows that FDA software that automatically subtracts the mean functions can be used.

Next, we compute the first eigenfunction $\widehat{\psi}_1^{(R)}$ of the kernel $\hat{c}_R$ and the scores $\widehat{\xi}_{1t}^{(R)} = \langle R_t, \widehat{\psi}_1^{(R)}\rangle$. Based on the $\widehat{\xi}_{1t}^{(R)}$, we compute the local Whittle estimator (LWE) $\widehat{H}_1^{(R)}$, which is described at the end of this section. The LWE of H is scale invariant, so the LWEs based on $\{\widehat{\xi}_{1t}^{(R)}\}$ and on $\{c\widehat{\xi}_{1t}^{(R)}\}$ are the same for any $c \neq 0$. Thus, the sign of $\widehat{\psi}_1^{(R)}$ has no effect on $\widehat{H}_1^{(R)}$. Our test statistic is

$$T_1^{(R)} = 2\sqrt{m}\left(\widehat{H}_1^{(R)} - \frac{1}{2}\right), \tag{3.2.5}$$

where m is the number of low frequencies used in the LW estimation. By Theorem 3.3.4, $T_1^{(R)} \xrightarrow{d} \mathcal{N}(0,1)$ under the null hypothesis. Thus, we reject H_0 if $T_1^{(R)} > q_{1-\alpha}$, where $q_{1-\alpha}$ is the $(1-\alpha)$th quantile of the standard normal distribution. The test is one-sided because the LRD corresponds to $H > 1/2$, as explained in Section 3.3.

For completeness and ease of reference, we provide the definition of the LWE for a scalar time series $\{\xi_t\}$. The partial sums of the weakly dependent sequence $\{Y_t\}$ in (3.2.1) are asymptotically self-similar with parameter $H = 1/2$, cf. Theorem 3.6.1. For the Z_t in Assumption 3.3.7, $H > 1/2$. Let

$$I_\xi(\omega_\ell) = \frac{1}{2\pi N}\left|\sum_{j=1}^{N}\xi_j e^{-ij\omega_\ell}\right|^2$$

be the periodogram at the Fourier frequencies

$$\omega_\ell = \frac{2\pi\ell}{N}, \quad \ell = 1, 2, \ldots, N.$$

The LWE of the self-similarity parameter H is defined as

$$\widehat{H} = \arg\min_{H \in \Theta} L(H), \tag{3.2.6}$$

where $\Theta = [\Delta_1, \Delta_2]$ with $0 < \Delta_1 < \Delta_2 < 1$ and

$$L(H) = \log\left(\frac{1}{m}\sum_{\ell=1}^{m} \omega_\ell^{2H-1} I_\xi(\omega_\ell)\right) - (2H-1)\frac{1}{m}\sum_{\ell=1}^{m} \log \omega_\ell, \tag{3.2.7}$$

with m denoting the number of low frequencies used in the estimation.

We conclude this section with comments on alternative testing approaches. The test described above uses only level one scores. Instead, one can use d tests based on scores of all levels up to level d. Assumptions in Section 3.3 are formulated in a way that makes such an extension obvious. The alternative of LRD would need to be formulated as in Li *et al.* (2020), and each test would detect LRD in a corresponding level. One can also use the average of the first d scores, an idea also present in Li *et al.* (2020), and use it for testing. We performed simulations and applied these alternative procedures to data examples. We found that they perform not as well as the simple procedure described above. Broadly speaking, the signal in the higher level scores becomes weaker. The empirical size and the power of the test become worse. These findings are in line with the work of Li *et al.* (2021) who argue that effective local Whittle estimation of functional time series can be achieved by using level one scores because the first principal component of a functional time series dominates its behavior. For these reasons, and to streamline the presentation and theory under the alternative, we consider only the simplest, but most effective, form of test.

3.3 Asymptotic justification

Section 3.3.1 presents fairly general assumptions and results that establish asymptotic validity of the test under the null hypothesis of short-range dependence with a change point. Section 3.3.2 does the same under the alternative of LRD. As noted at the end of Section 3.2, the assumptions and results of Section 3.3.2 could be generalized in the spirit of multiple projections considered by Li *et al.* (2020), but we think the justification based on a simpler model is sufficient to support the applicability of the test and addresses the key difficulties that already occur in a single projection.

The proofs of results stated in Section 3.3.1 are presented in Section 3.6 of Supporting Information whose Section 3.7 contains the proofs of the results of Section 3.3.2.

3.3.1 Asymptotic justification under the null hypothesis

We now formulate the assumptions on the series $\{Y_t\}$ in (3.2.1) more precisely. There are inclusions between the assumptions that will become clear in the following. We use this system of assumptions because certain intermediate results of independent value hold under weaker assumptions than the final results. Assumption 3.3.1 is a general quantification of weak dependence (short memory) of a sequence of functions. It is used by Berkes *et al.* (2013) to prove a form of the CLT for functions that we need, see Theorem 3.6.1 in online Supporting Information. Our proofs do not directly use Assumption 3.3.1, only the conclusions of Theorem 3.6.1. LRD models do not satisfy Assumption 3.3.1.

ASSUMPTION 3.3.1 *The Y_t in* (3.2.1) *satisfy*

$$E\|Y_0\|^4 < \infty, \ \ E[Y_0] = 0$$

and

$$Y_t = f(\varepsilon_t, \ldots, \varepsilon_{t-m+1}, \varepsilon_{t-m}, \varepsilon_{t-m-1}, \ldots), \tag{3.3.1}$$

where $f : S^\infty \to L^2$ is a measurable function and $\{\varepsilon_t\}$ is a sequence of iid random elements of a *measurable space S. We also require that the sequence $\{Y_t\}$ satisfies the following weak dependence condition:*

$$\left(E\|Y_t - Y_{t,m}\|^4\right)^{1/4} = O(m^{-\kappa}) \tag{3.3.2}$$

for some $\kappa > 4$, where

$$Y_{t,m} = f(\varepsilon_t, \ldots, \varepsilon_{t-m+1}, \varepsilon_{t,t-m}^{(m)}, \varepsilon_{t,t-m-1}^{(m)}, \ldots),$$

with the sequences $\{\varepsilon_{t,k}^{(m)}\}$ iid copies of ε_0, independent of $\{\varepsilon_t\}_{t=-\infty}^{\infty}$.

Next, consider the long-run covariance kernel of the sequence $\{Y_t\}$ in (3.2.1) defined by

$$c_Y(u,v) = E\{Y_0(u)Y_0(v)\} + \sum_{t=1}^{\infty} E\{Y_0(u)Y_t(v)\} + \sum_{t=1}^{\infty} E\{Y_0(v)Y_t(u)\}. \tag{3.3.3}$$

Denote by $\lambda_j^{(Y)}$ the eigenvalues of the kernel c_Y given by (3.3.3), and by $\psi_j^{(Y)}$ the corresponding eigenfunctions, i.e.

$$\int_0^1 c_Y(u,v)\psi_j^{(Y)}(v)dv = \lambda_j^{(Y)}\psi_j^{(Y)}(u).$$

The level j scores of Y_t are defined by

$$\xi_{j,t}^{(Y)} = \left\langle Y_t, \psi_j^{(Y)} \right\rangle = \int_0^1 Y_t(u)\psi_j^{(Y)}(u)du.$$

Our test is based on projections on estimators of the eigenfunctions $\psi_j^{(Y)}$. To ensure that these projections converge to well-defined objects, we must assume that the $\psi_j^{(Y)}$ are uniquely defined (up to a sign). The following assumption ensures that.

ASSUMPTION 3.3.2 *For an integer d,*

$$\lambda_1^{(Y)} > \lambda_2^{(Y)} > \cdots > \lambda_d^{(Y)} > \lambda_{d+1}^{(Y)}.$$

Our test involves estimation of the kernel c_Y. It requires using the Bartlett kernel $K(x) = 1 - |x|, |x| < 1$. However, the behavior of the test under the null hypothesis is valid for more general kernels, and these results could be used for tests with different alternatives. We therefore specify Assumption 3.3.3 which holds for the Bartlett kernel with $q = 1$ and $h = O(N^{1/3})$.

ASSUMPTION 3.3.3 *The kernel function K is symmetric and continuous, $K(0) = 1$, and $K(u) = 0$ if $|u| > c$ for some $c > 0$. There is $q > 0$ such that*

$$\lim_{x \to 0} \frac{1 - K(x)}{|x|^q} = \zeta \in (0, \infty) \tag{3.3.4}$$

and the bandwidth $h = h(N)$ in (3.2.4) *satisfies*

$$h(N) = O\left(N^{\frac{1}{1+2q}}\right). \tag{3.3.5}$$

To establish the asymptotic normality of the test statistic under the null hypothesis, we need to restrict that class of processes $\{Y_t\}$ in (3.2.1) to functional moving averages

$$Y_t = \sum_{j=0}^{\infty} A_j(\varepsilon_{t-j}), \quad A_j \in \mathcal{L}(L^2), \tag{3.3.6}$$

where the ε_j are mean zero iid random functions in L^2 and $\mathcal{L}(L^2)$ is the class of bounded linear operators defined on L^2. Denote by $\|A\|_{\mathcal{L}}$ the operator norm for $A \in \mathcal{L}(L^2)$.

ASSUMPTION 3.3.4 *Representation* (3.3.6) *holds with iid mean zero errors ε_t satisfying $E\|\varepsilon_0\|^4 < \infty$ and the operators A_j are of the form*

$$A_j(\cdot) = \sum_{k=1}^{\infty} a_{j,k} \langle \cdot, \varphi_k \rangle \psi_k^{(Y)},$$

where $\psi_1^{(Y)}, \psi_2^{(Y)}, \ldots$ *are the eigenfunctions of the long-run covariance kernel* c_Y *in* (3.3.3), $\{\varphi_k\}$ *is any orthonormal basis in* L^2 *and, for a* $\kappa > 4$,

$$\left\{ \sum_{j \geq m} \sum_{k=1}^{\infty} a_{j,k}^2 \right\}^{1/2} = O\left(m^{-\kappa}\right). \tag{3.3.7}$$

Proposition 3.3.1 shows that Assumption 3.3.4 implies Assumption 3.3.1. Assumption 3.3.4 is first used in Proposition 3.3.2 which shows that the projections we work with are scalar linear processes. The assumption of linearity is needed because there is at present no requisite theory related to the integrated periodogram and local Whittle estimation of nonlinear scalar time series.

By (3.3.7), for each j, $\sum_{k=1}^{\infty} a_{j,k}^2 < \infty$, so each A_j is a Hilbert-Schmidt operator. These operators are however restricted by requiring that they share the φ_k. The $\psi_k^{(Y)}$ are used as a basis chosen to represent them.

Our test involves estimation of the change point $n^\star$. Under the alternative, this change point does not exist, but since we do not know which hypothesis is true, it must always be estimated and incorporated into the test statistic. We impose the following commonly used assumption on the change point and its estimator.

ASSUMPTION 3.3.5 *Suppose that* $n^\star = \lfloor N\theta \rfloor \ for \ some \ 0 < \theta < 1$ *and*

$$|\hat{n} - n^\star| = O_P(1). \tag{3.3.8}$$

Condition (3.3.8) is known to hold for several good change point estimators. We refer to Berkes *et al.* (2009), Aston and Kirch (2012a), Aston and Kirch (2012b), Horváth and Rice (2014) and Aue *et al.* (2018), among others.

We begin with a key result regarding the long-run covariance kernel estimator (3.2.4), which is crucial in deriving the asymptotic distribution of the test statistic under the null hypothesis. The proof is presented in Section 3.6.1 of the Supporting Information. However, for the result to hold, we need a more traditional quantification of the weak dependence of the Y_t in (3.2.1).

Recall that the Hilbert–Schmidt norm of a function defined on $[0,1]\times[0,1]$ is defined by $\|g\|_{\mathcal{S}}^2 = \int_0^1\int_0^1 g^2(u,v)dudv$.

THEOREM 3.3.1 *Suppose* (3.2.1) *is satisfied. If Assumptions 3.3.1, 3.3.3 and 3.3.5, hold for* $\alpha = \frac{q}{1+2q}$ *with* q *defined in* (3.3.4)*, and if there exists* $\varrho > q$ *such that*

$$\sum_{t=-\infty}^{\infty} |t|^{\varrho}\|\gamma_t\|_{\mathcal{S}} < \infty, \tag{3.3.9}$$

then

$$\|\widehat{C}_R - C\|_{\mathcal{S}} = O_P(N^{-\alpha}),$$

where $\gamma_t(u,v) = E[Y_t(u)Y_0(v)]$*,* C *and* $\widehat{C}_R$ *are the covariance operators of* (3.3.3) *and* (3.2.4)*, respectively, and* $\|\cdot\|_{\mathcal{S}}$ *represents the Hilbert-Schmidt norm. Note that* $\alpha \in (0,1/2)$*.*

The following corollary to Theorem 3.3.1 follows directly from Lemma 2.3 in Horváth and Kokoszka (2012). Recall from Section 3.2 that the sign of $\widehat{\psi}_1^{(R)}$ does not affect the value of the test statistic (3.2.5). Thus, we assume throughout the paper that $\mathrm{sign}(\langle\widehat{\psi}_j^{(R)}, \psi_j^{(Y)}\rangle) = 1$.

COROLLARY 3.3.1 *If the assumptions of Theorem 3.3.1 hold, and, in addition, Assumption 3.3.2 holds, then for any* $j = 1,2,\ldots,d$*,*

$$\left\|\widehat{\psi}_j^{(R)} - \psi_j^{(Y)}\right\| = O_P(N^{-\alpha}), \tag{3.3.10}$$

where $\alpha > 0$ *is as in Theorem 3.3.1.*

The above results, and several other results proven in Section 3.6 of the Supporting Information, are valid under the very general weak dependence Assumption 3.3.1. However, to establish the asymptotic normality of the test statistic, we need Assumption 3.3.4. The following proposition shows that Assumption 3.3.4 implies Assumption 3.3.1.

PROPOSITION 3.3.1 *Suppose the functional observations Y_t are given by* (3.3.6). *If $E\|\varepsilon_0\|^4 < \infty$ and for a $\kappa > 4$,*

$$\left\{\sum_{j\geq m} \|A_j\|_{\mathcal{L}}^2\right\}^{1/2} = O\left(m^{-\kappa}\right), \tag{3.3.11}$$

then the Y_t satisfy Assumption 3.3.1.

The additional constraints on the form of linear operators A_j, ensure that the projections of observations Y_t (onto their eigenspace) will follow a linear process with iid errors and square-summable coefficients, as stated in the following proposition.

PROPOSITION 3.3.2 *If Assumption 3.3.4 holds, then Assumption 3.3.1 holds. Moreover, for each positive integer j, the projections $\xi_{jt}^{(Y)} := \langle Y_t, \psi_j^{(Y)}\rangle$ have the form*

$$\xi_{jt}^{(Y)} = \sum_{k=0}^{\infty} a_k \zeta_{t-k}, \quad \sum_{k=0}^{\infty} a_k^2 < \infty, \tag{3.3.12}$$

where $\{\zeta_t\}$ is a sequence of iid random variables with $E\zeta_0^4 < \infty$.

In order to establish asymptotic results regarding periodograms based on scores $\{\xi_{jt}^{(Y)}\}$, we need to impose some smoothness conditions on their spectral density. To be specific, we require the spectral density to be Lipschitz continuous and bounded away from 0. We write $f \in \Lambda_\beta(D)$, for some $\beta \in (0, 1]$, if for any $\omega, \omega' \in D$, $|f(\omega) - f(\omega')| \leq C|\omega - \omega'|^\beta$; $\Lambda_\beta(D)$ is the class of Lipschitz continuous functions on D.

Fix $j \geq 1$ and denote by $I_j^{(Y)}(\omega_\ell)$ the periodogram ordinates of the projections $\xi_{jt}^{(Y)}$, and by f_j their spectral density.

ASSUMPTION 3.3.6 *For $j \in \mathbb{N}^+$, suppose $f_j \in \Lambda_\beta([0, \pi])$ for some $\beta \in \left(\frac{1}{2}, 1\right]$ and*

$$\inf_{\omega\in[0,\pi]} f_j(\omega) > 0.$$

Even though Theorems 3.3.2 and 3.3.3 are technical results, we list them here because we need Assumptions 3.3.4 and 3.3.6 only to establish these two theorems. Other intermediate results in the proof of Theorem 3.3.4 can be proven without Assumptions 3.3.4 and 3.3.6.

THEOREM 3.3.2 *If Assumptions 3.3.4 and 3.3.6 hold and, as $N \to \infty$,*

$$m \to \infty \quad \textit{and} \quad m^{\beta+\frac{1}{2}} = O\left(N^{\beta}\right), \tag{3.3.13}$$

then

$$\frac{1}{m}\sum_{\ell=1}^{m}\left(\frac{I_j^{(Y)}(\omega_\ell)}{f_j(0)} - 1\right) = O_P\left(m^{-1/2}\right).$$

THEOREM 3.3.3 *Suppose the assumptions of Theorem 3.3.2 hold with* (3.3.13) *strengthened to*

$$m \to \infty \quad \textit{and} \quad m^{\beta+\frac{1}{2}}\log m = o\left(N^{\beta}\right). \tag{3.3.14}$$

Then,

$$\frac{1}{\sqrt{m}}\sum_{\ell=1}^{m}\nu_{\ell,m}\left(\frac{I_j^{(Y)}(\omega_\ell)}{f_j(0)} - 1\right) \xrightarrow{d} \mathcal{N}(0,1),$$

where

$$\nu_{\ell,m} = \log \ell - \frac{1}{m}\sum_{j=1}^{m}\log j. \tag{3.3.15}$$

THEOREM 3.3.4 *Suppose the observations follow model* (3.2.1). *Suppose Assumptions 3.3.2–3.3.6 and the following condition on m are satisfied:*

$$\frac{1}{m} + \frac{m(\log m)^2}{N^{2\alpha}} + \frac{m^{\beta+\frac{1}{2}}\log m}{N^{\beta}} \to 0, \quad \textit{as } N \to \infty,$$

where α and β are defined in Theorem 3.3.1 and Assumption 3.3.6, respectively. Then

$$T_1^{(R)} = 2\sqrt{m}\left(\widehat{H}_1^{(R)} - \frac{1}{2}\right) \xrightarrow{d} \mathcal{N}(0,1). \tag{3.3.16}$$

3.3.2 Consistency of the test

In this section, we focus on the asymptotic behavior of our test statistic $T_1^{(R)}$ under the LRD alternative, which is closely related to consistency of the LWE $\widehat{H}^{(R)}$. A great deal of work has been done regarding consistency of the LWE in the context of scalar long-range dependent time series, see e.g. Robinson (1995), Velasco (1999), Shao and Wu (2007), and Baek and Pipiras (2012). Li *et al.* (2021) study the LWE based on the estimated first scores of a functional LRD process. In our context, this estimator is based on a residual functional process obtained after removing the estimated mean function, so the results of Li *et al.* (2021) are not directly applicable, but elements of their proofs are used. In the context of scalar LRD time series, this problem was studied by Baek and Pipiras (2012). A substantial difficulty absent in the scalar case is due to the decomposition

$$\widehat{\xi}_{1t}^{(R)} = \langle R_t, \widehat{\psi}_1^{(R)} \rangle = \langle R_t, \psi_1^{(Z)} \rangle + \langle R_t, \widehat{\psi}_1^{(R)} - \psi_1^{(Z)} \rangle, \tag{3.3.17}$$

where $\psi_1^{(Z)}$ is the first eigenfunction of some proper covariance kernel of $\{Z_t\}$. The first term, $\langle R_t, \psi_1^{(Z)} \rangle$ can be handled in a similar way as the scalar residuals in Baek and Pipiras (2012). However, the second term is absent in the scalar case. To address this issue, we develop new results analogous to Theorem 3.3.1 and Corollary 3.3.1, which are used to handle the difference $\widehat{\psi}_1^{(R)} - \psi_1^{(Z)}$. These results indicate that the first eigenfunction obtained from $\hat{c}_R$ is still a consistent estimate of the true eigenfunction under the LRD alternative. Finally, we shall assert that our test is consistent. We put the proofs of all results stated in this section into Section 3.7 of the Supporting Information.

Under the alternative of LRD, we need a more direct model specification. We use the specification of Li *et al.* (2021) because they established fundamental properties of the LWE in that context, and our test is based on that estimator.

ASSUMPTION 3.3.7 *(i) The process* $\{Z_t\}$ *in* (3.2.2) *has the form*

$$Z_t = \sum_{j=0}^{\infty} b_j \eta_{t-j}. \tag{3.3.18}$$

The scalar impulse responses b_j satisfy $b_j \sim j^{H_0-\frac{3}{2}}$ for an unknown $H_0 \in \left(\frac{1}{2}, 1\right)$. The functions $\eta_t \in L^2$ are iid, have mean zero and satisfy $E\|\eta_t\|^4 < \infty$.

(ii) The transfer function $B(\lambda) := \sum_{j=0}^{\infty} b_j e^{ij\lambda}$ is differentiable on $(0, \lambda_0)$ for some $\lambda_0 > 0$, with

$$\frac{d}{d\lambda} B(\lambda) = O\left(\frac{|B(\lambda)|}{\lambda}\right), \ as\ \lambda \to 0+ .$$

ASSUMPTION 3.3.8 *The spectral density of the first scores $\xi_{1t}^{(Z)}$ of the series $\{Z_t\}$, denoted by $f_1^{(Z)}(\lambda)$, satisfies*

$$f_1^{(Z)}(\lambda) \sim D_0 \lambda^{1-2H_0} \left(1 + O\left(\lambda^{\vartheta}\right)\right), \ as\ \lambda \to 0+,$$

where D_0 is an unknown positive constant, $\vartheta \in (0, 2]$ and $H_0 \in \left(\frac{1}{2}, 1\right)$ is the same as in Assumption 3.3.7.

It follows from Assumption 3.3.5 that $\frac{\hat{n}}{N} \xrightarrow{P} \theta$, where θ is the change-point rescaled to (0,1). Under the alternative, there is no change point, so we merely assume that the ratio $\frac{\hat{n}}{N}$ has a limiting distribution on $(0, 1)$.

ASSUMPTION 3.3.9 *The change point estimator $\hat{n}$ satisfies*

$$\frac{\hat{n}}{N} \xrightarrow{P} \xi,$$

for some random variable ξ taking values in $(0, 1)$.

To prove Theorem 3.3.5, we need to use the Bartlett kernel function defined in Assumption 3.3.10. The same restriction arises in Li *et al.* (2021) who do not consider the residuals and work directly with the Z_t. However, we show in Section 3.7 that this restriction is only needed in proving Lemma 3.7.2. Assumption 3.3.11 is enough for the remaining part of proof.

ASSUMPTION 3.3.10 *The kernel function* K *in* (3.2.4) *is of the form:*

$$K(x) = \begin{cases} 1 - |x|, & |x| \leq 1, \\ 0, & |x| > 1. \end{cases} \tag{3.3.19}$$

ASSUMPTION 3.3.11 *The kernel function* K *in* (3.2.4) *is symmetric and continuous, satisfying* $K(0) = 1$*, and* $K(u) = 0$ *if* $|u| > c$ *for some* $c > 0$*.*

ASSUMPTION 3.3.12 *The bandwidth* $h = h(N)$ *in* (3.2.4) *satisfies*

$$\frac{h^{4-4H_0}}{N} + \frac{h \log N}{N} \to 0, \quad \textit{as } N \to \infty.$$

With $h = O\left(N^{1/3}\right)$, Assumption 3.3.12 is automatically satisfied for all $H_0 \in \left(\frac{1}{2}, 1\right)$ (cf. Assumption 3.3.3).

As noted above, we need to establish a result similar to Theorem 3.3.1, which guarantees a consistent estimate of the first eigenfunction. However, under the LRD setting, the covariance function in (3.3.3) does not converge in $L^2\left([0,1] \times [0,1]\right)$. Therefore, following Li *et al.* (2020), we consider the limiting covariance function of $\{Z_t\}$ defined as

$$c_Z(u,v) = \lim_{N \to \infty} N^{-2H_0} E\left[\sum_{t=1}^{N} \sum_{s=1}^{N} Z_t(u) Z_s(v)\right]. \tag{3.3.20}$$

Let C_Z be the covariance operator of c_Z. Denote by $\psi_j^{(Z)}$ the eigenfunction corresponding to the j-th largest eigenvalue $\lambda_j^{(Z)}$ of c_Z, i.e.

$$\int_0^1 c_Z(u,v) \psi_j^{(Z)}(v) dv = \lambda_j^{(Z)} \psi_j^{(Z)}(u).$$

Since $h^{1-2H_0}\widehat{C}_R$ is proportional to $\widehat{C}_R$, the eigenfunctions of $\widehat{C}_R$ are the same as those of $h^{1-2H_0}\widehat{C}_R$. Consequently, we just consider the functional principal component analysis (FPCA) on the same estimator $\widehat{C}_R$ as is in the null hypothesis. This congruity is crucial in our hypothesis testing problem. Note that we do not need to estimate parameter H_0 for the purpose of finding the eigenfunctions.

THEOREM 3.3.5 *Under Assumptions 3.3.7 (i), 3.3.9–3.3.10 and 3.3.12,*

$$\|h^{1-2H_0} \cdot \widehat{C}_R - C_Z\|_{\mathcal{S}} = o_P(1),$$

where $\widehat{C}_R$ is defined in Theorem 3.3.1 and H_0 is given in Assumption 3.3.7.

Analogously to the situation under the null hypothesis, we need to impose a separability condition on the eigenvalues $\lambda_j^{(Z)}$ to get a consistent estimator of $\psi_j^{(Z)}$.

ASSUMPTION 3.3.13 *For an integer d,*

$$\lambda_1^{(Z)} > \lambda_2^{(Z)} > \cdots > \lambda_d^{(Z)} > \lambda_{d+1}^{(Z)}.$$

Corollary 3.3.2 now follows immediately from Lemma 2.3 in Horváth and Kokoszka (2012).

COROLLARY 3.3.2 *If the assumptions of Theorem 3.3.5 hold and, in addition, Assumption 3.3.13 holds, then for any $j = 1, 2, \ldots, d$,*

$$\left\| \widehat{\psi}_j^{(R)} - \psi_j^{(Z)} \right\| = o_P(1). \tag{3.3.21}$$

THEOREM 3.3.6 *Suppose that Assumptions 3.3.7–3.3.10 and 3.3.12 are satisfied. In addition, Assumption 3.3.13 holds with $d = 1$ and m satisfies*

$$\frac{1}{m} + \frac{m^{1+2\vartheta}(\log m)^2}{N^{2\vartheta}} \to 0, \quad as\ N \to \infty,$$

where ϑ is defined in Assumption 3.3.8. Then, $\widehat{H}^{(R)} \xrightarrow{P} H_0$.

COROLLARY 3.3.3 *Under the assumptions of Theorem 3.3.6, $T_1^{(R)} \xrightarrow{P} \infty$, where $T_1^{(R)}$ is defined in* (3.2.5).

It is convenient to formulate at this point a global assumption on m that ensures the validity of our results both under the null and the alternative hypotheses.

ASSUMPTION 3.3.14 *The number of Fourier frequencies m satisfies the following condition:*

$$\frac{1}{m} + \frac{m(\log m)^2}{N^{2\alpha}} + \frac{m^{\beta+\frac{1}{2}}\log m}{N^{\beta}} + \frac{m^{1+2\vartheta}(\log m)^2}{N^{2\vartheta}} \to 0, \quad as\ N \to \infty,$$

where $\alpha = \frac{q}{1+2q}$ with q defined in (3.3.4) *and β, ϑ are defined in Assumptions 3.3.6 and 3.3.8, respectively.*

3.4 A small simulation study

Denote by $\{B_t\}$ a sequence of iid standard Brownian motions on $[0,1]$. This choice is motivated by the application in Section 3.5; price curves have a random walk type behavior.

Under H_0, the functional observations X_t are generated by the model

$$X_t = \begin{cases} Y_t, & 1 \le t \le \left\lfloor \frac{N}{2} \right\rfloor, \\ \delta + Y_t, & \left\lfloor \frac{N}{2} \right\rfloor + 1 \le t \le N, \end{cases} \tag{3.4.1}$$

where δ is a constant function. The process $\{Y_t\}$ follows an FAR(1) model

$$Y_t(u) = \int \psi(u,v) Y_{t-1}(v)dv + B_t(u), \quad u \in [0,1],$$

where the kernel function is defined as

$$\psi(u,v) = \frac{\exp\{-(u^2+v^2)/2\}}{4\int \exp(-x^2)dx}.$$

The Hilbert-Schmidt norm of the corresponding operator Ψ is

$$\|\Psi\|_{\mathcal{S}} = \left(\iint \psi^2(u,v)dudv \right)^{1/2} = \frac{1}{4}.$$

Under H_A, we consider the functional long-range dependent process defined as follows, cf. (3.3.18),

$$X_t = \sum_{j=1}^{L} g(j) j^{H_0 - \frac{3}{2}} B_{t-j}, \quad H_0 \in \left(\frac{1}{2}, 1 \right), \tag{3.4.2}$$

where the function g satisfies $g(x) \to 1$, as $x \to \infty$. We used the truncation level $L = 1,500$ because this is the largest series length we consider. We considered the following functions g:

$$g_1(j) = 1, \quad g_2(j) = \exp\left(-0.01 \cdot j - \frac{1}{2} \right) \sin\left(\frac{2\pi j}{15} \right) + 1.$$

The coefficients $g_i(j) j^{H_0 - \frac{3}{2}}$, $i = 1, 2$, are displayed in Figure 2.1. In the computation of the LWE (3.2.6), we used the Bartlett kernel (3.3.19) with $m = N^{0.6}$, which satisfies Assumption 3.3.14 with $q = 1$, $\beta = 1$ and $\vartheta = 1$.

The results in Tables 3.1 and 3.2 indicate that test achieves satisfactory size and is asymptotically consistent. We experimented with different jump functions δ, not necessarily constant; the sizes are similar. We also experimented with different functions g. If the function g deviates from the constant function too much, the power may decrease significantly. Figure 2.1 illustrates what kind of impulse response coefficients still produce satisfactory power. The empirical rejection rates are not sensitive to h, but increase as m approaches the theoretical upper limit $2/3$. As the change point moves away from $N/2$, the empirical size tends to increase above the level for $N/2$. If the change point is at $N/4$, the increase is roughly by 0.5 to 2.0 percent, depending on the nominal size. These findings are based on the tables presented in Section 3.8 of the Supporting Information.

Table 3.1: Empirical sizes (in percent), based on $2,000$ replications, with number of low frequencies $m = N^{0.6}$ and bandwidth $h = N^{0.3}$.

Nominal size		1.0	5.0	10.0
	$N = 500$	1.3	4.7	8.3
$\delta = 0.25$	$N = 1000$	0.8	4.1	7.9
	$N = 1500$	0.8	3.3	8.2
	$N = 500$	1.5	5.7	9.9
$\delta = 0.50$	$N = 1000$	1.1	4.9	9.1
	$N = 1500$	1.0	4.6	9.3

Table 3.2: Empirical powers (in percent) for the two functions g in (3.4.2) (cf. Figure 2.1). The values m, h and the number of replications are the same as in Table 3.1.

Nominal size		1.0		5.0		10.0	
		$H_0 = 0.6$	$H_0 = 0.9$	$H_0 = 0.6$	$H_0 = 0.9$	$H_0 = 0.6$	$H_0 = 0.9$
	$N = 500$	90.1	99.7	96.5	99.9	98.4	100.0
g_1	$N = 1000$	97.4	100.0	98.9	100.0	99.7	100.0
	$N = 1500$	99.1	100.0	99.7	100.0	99.9	100.0
	$N = 500$	46.1	67.1	66.4	82.9	76.1	87.9
g_2	$N = 1000$	67.5	96.0	84.3	98.7	90.3	99.4
	$N = 1500$	86.8	99.8	95.0	100.0	97.5	100.0

3.5 Application to intraday ETF data

We applied our test to nine Select Sector Exchange Traded Funds (ETFs) summarized in Table 3.3.

Table 3.3: The nine sector ETFs.

Ticker	Sector	Ticker	Sector
XLB	Materials	XLP	Consumer Staples
XLE	Energy	XLU	Utilities
XLF	Financials	XLV	Health Care
XLI	Industrials	XLY	Consumer Discretionary
XLK	Technology		

We consider two types of curves derived from intraday prices. They are defined as follows.

DEFINITION 3.5.1 Suppose $P_n(t)$, $n = 1, \ldots, N$, is the price of a financial asset at time t on trading day n. The functions

$$R_n(t) = 100[\ln P_n(t) - \ln P_n(t_0)], \quad n = 1, \ldots, N,$$

are called the cumulative intraday returns (CIDRs). The functions

$$A_n(t) = |R_n(t)|, \quad n = 1, \ldots, N,$$

are called the absolute cumulative intraday returns (ACIDRs).

The time t in Definition 3.5.1 is a continuous variable covering the time interval from 9:30 to 16:00 EST on days when the floor of the NYSE is open for the whole trading day. We consider only such days. We use QuantQuote data which are closing prices over one minute long subintervals. The 9:30 to 16:00 interval is rescaled to the unit interval, so $t \in [0, 1]$. To illustrate, CIDR and ACIDR curves for the XLF over five consecutive days are shown in Figures 3.1 and 3.2.

Intraday prices have been extensively studied over the past 30 years; a Google search shows close to two thousand references. FDA methods were used to analyze the rescaled prices in Definition 3.5.1 in Kokoszka and Reimherr (2013), Horváth *et al.* (2014) and Kokoszka *et al.* (2015), among many other papers.

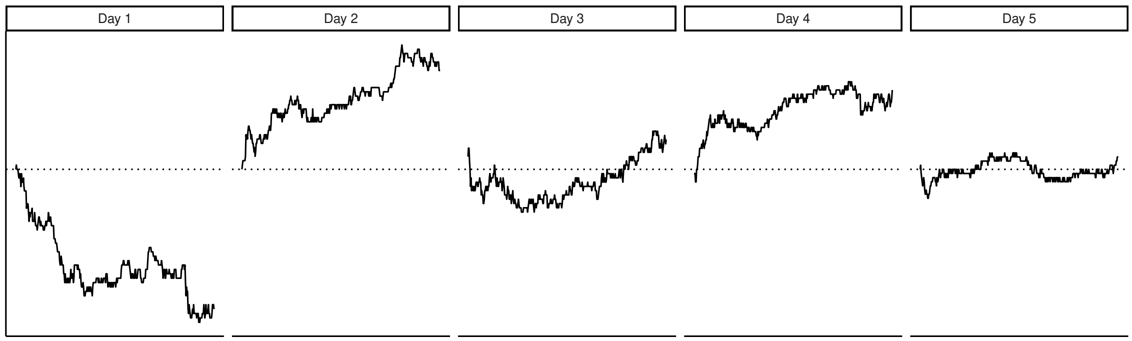

Figure 3.1: CIDRs for the XLF over 12/19/2011 - 12/23/2011

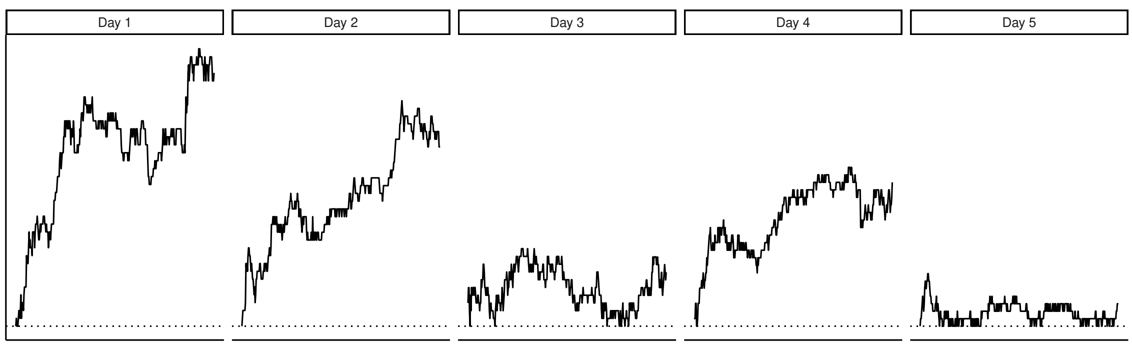

Figure 3.2: ACIDRs for the XLF over 12/19/2011 - 12/23/2011.

We can heuristically group the curves by time period, see Table 3.4, as discussed in Kokoszka *et al.* (2018) and several earlier papers cited by them. This division shows potential change points. Our test is applied to the full time period 07/05/2006 to 12/30/2011, i.e. the test is based on all $N = 1,378$ curves. It is applied using the implementation evaluated in Section 3.4, i.e. $h = N^{0.3}$, $m = N^{0.6}$.

Results of the test We now report to results of the application of our test at the five percent nominal significance level. Applied to the CIDRs, the test accepts H_0 for all ETFs except Consumer

Table 3.4: Four time periods approximately identifying the stages of the 2008 crisis.

Designation	Time span	Sample size (days)
Before	07/05/2006 - 09/28/2007	313
During	10/01/2007 - 02/27/2009	351
After 1	03/02/2009 - 07/30/2010	356
After 2	08/02/2010 - 12/30/2011	358

Table 3.5: Change-point estimates for the CIDRs and the ACIDRs of the nine sectors. The designations of four periods are given in Table 3.4.

Ticker	Change-point estimates			
	CIDRs		ACIDRs	
	$\hat{n}$	Period	$\hat{n}$	Period
XLB	669	After 1	362	During
XLE	491	During	743	After 1
XLF	669	After 1	323	During
XLI	669	After 1	497	During
XLK	260	Before	724	After 1
XLP	670	After 1	765	After 1
XLU	505	During	774	After 1
XLV	672	After 1	323	During
XLY	669	After 1	373	During

Staples. This indicates that short range dependence with a potential change point better explains the behavior of these curves than long-range dependence without change points. The LRD detected in Consumer Staples could be a false positive, but it could also be explained by the fact that staples (e.g. food and household products) have to be purchased even in times of global financial stress. The demand for such goods is less influenced by the economic situation and consequently the stock prices might not change in a dramatic way, but rather follow their long term patterns. The null hypothesis is rejected for ACIDRs for all sectors. There is thus evidence for long memory in the intraday volatility, but not in the level of the cumulative intraday returns.

Table 3.5 shows the locations of the estimated change points and locates them within the four periods listed in Table 3.4. They are all located during or right after the of the approximate period of the 2008 crisis. We emphasize that our testing framework does not provide information about the statistical significance of the change points in Table 3.5. These are the estimated change points that were used in the construction of the test statistic.

3.6 Proofs of the results in Section 3.3.1

3.6.1 Proof of Theorem 3.3.1

We state the following theorem which facilitates the proof of Theorem 3.3.1.

THEOREM 3.6.1 *Under Assumption 3.3.1,*

$$S_N(x,u) := N^{-1/2} \sum_{1\le t\le \lfloor Nx\rfloor} Y_t(u) \xrightarrow{d} \Gamma(x,u),$$

with the Gaussian process $\Gamma(x,u)$ *defined by*

$$\Gamma(x,u) = \sum_{j=1}^{\infty} \sqrt{\lambda_j^{(Y)}} W_j(x)\psi_j^{(Y)}(u),$$

where the W_j *are iid standard Brownian motions. The convergence is in the Skorokhod space* $D([0,1],L^2)$.

Theorem 3.6.1 follows from Theorem 1.1 in Berkes *et al.* (2013) which states that for every N we can define a Gaussian process $\Gamma_N(x,u)$ such that $\Gamma_N(x,u)$ and $\Gamma(x,u)$ have the same distribution, and

$$\sup_{0\le x\le 1} \int (S_N(x,u) - \Gamma_N(x,u))^2\, du = o_P(1). \tag{3.6.1}$$

Our proof will use Proposition 3.6.1 that establishes a suitable summability of autocovariances. Recall that the Hilbert–Schmidt norm of a function defined on $[0,1]\times[0,1]$ is defined by $\|g\|_{\mathcal{S}}^2 = \int_0^1\int_0^1 g^2(u,v)dudv$.

PROPOSITION 3.6.1 *If Assumption 3.3.4 holds, then there is* $\varrho \in (1,3]$ *such that*

$$\sum_{t=-\infty}^{\infty} |t|^{\varrho} \|\gamma_t\|_{\mathcal{S}} < \infty. \tag{3.6.2}$$

PROOF: Under Assumption 3.3.4, we have

$$
\begin{aligned}
\gamma_t(u,v) &= E[Y_t(u)Y_0(v)] \\
&= E\left[\sum_{j=0}^{\infty} A_j(\varepsilon_{t-j}) \sum_{l=0}^{\infty} A_l(\varepsilon_{-l})\right] \\
&= E\left[\sum_{j=0}^{\infty}\left(\sum_{k=1}^{\infty} a_{j,k}\langle\varepsilon_{t-j},\varphi_k\rangle\psi_k^{(Y)}(u)\right)\sum_{l=0}^{\infty}\left(\sum_{s=1}^{\infty} a_{l,s}\langle\varepsilon_{-l},\varphi_s\rangle\psi_s^{(Y)}(v)\right)\right] \\
&= \sum_{j=0}^{\infty}\sum_{k=1}^{\infty}\sum_{l=0}^{\infty}\sum_{s=1}^{\infty} a_{j,k}a_{l,s}\psi_k^{(Y)}(u)\psi_s^{(Y)}(v)E\left[\langle\varepsilon_{t-j},\varphi_k\rangle\langle\varepsilon_{-l},\varphi_s\rangle\right].
\end{aligned}
$$

Observe that if $t-j\neq -l$, then $E\left[\langle\varepsilon_{t-j},\varphi_k\rangle\langle\varepsilon_{-l},\varphi_s\rangle\right]=0$, and otherwise

$$E\left[\langle\varepsilon_{t-j},\varphi_k\rangle\langle\varepsilon_{t-j},\varphi_s\rangle\right] = E\left[\langle\varepsilon_0,\varphi_k\rangle\langle\varepsilon_0,\varphi_s\rangle\right] =: b_{k,s}.$$

Therefore

$$\gamma_t(u,v) = \sum_{j=0}^{\infty}\sum_{k=1}^{\infty}\sum_{s=1}^{\infty} a_{j,k}a_{j-t,s}b_{k,s}\psi_k^{(Y)}(u)\psi_s^{(Y)}(v). \tag{3.6.3}$$

By the Cauchy-Schwarz inequality,

$$|b_{k,s}| \le E\|\varepsilon_0\|^2 =: C_\varepsilon.$$

The key to proving (3.6.2), is thus showing that the sums of the products $a_{j,k}a_{j-t,s}$ in (3.6.3) are suitably small. If $t>0$, only $j\ge t$ are used. If $-t>0$, a_{j-t} with indexes $j-t\ge -t$ are used. Since $\|\gamma_{-t}\|_{\mathcal{S}} = \|\gamma_t\|_{\mathcal{S}}$ we can consider either of these cases. *We consider $-t>0$ because we then do not have to manipulate* (3.6.3) *any further.*

By (3.6.3),

$$\gamma_t^2(u,v) = \sum_{j=0}^{\infty}\sum_{k=1}^{\infty}\sum_{s=1}^{\infty} a_{j,k}a_{j-t,s}b_{k,s}\psi_k^{(Y)}(u)\psi_s^{(Y)}(v)\sum_{p=0}^{\infty}\sum_{q=1}^{\infty}\sum_{r=1}^{\infty} a_{p,q}a_{p-t,r}b_{q,r}\psi_q^{(Y)}(u)\psi_r^{(Y)}(v).$$

It follows that

$$\iint \gamma_t^2(u,v)dudv = \sum_{j=0}^{\infty}\sum_{k=1}^{\infty}\sum_{s=1}^{\infty}\sum_{p=0}^{\infty}\sum_{q=1}^{\infty}\sum_{r=1}^{\infty}\Big\{a_{j,k}a_{j-t,s}b_{k,s}a_{p,q}a_{p-t,r}b_{q,r} \cdot \int \psi_k^{(Y)}(u)\psi_q^{(Y)}(u)du \int \psi_s^{(Y)}(v)\psi_r^{(Y)}(v)dv\Big\}.$$

Since the $\psi_k^{(Y)}$ are orthonormal,

$$\iint \gamma_t^2(u,v)dudv = \sum_{j=0}^{\infty}\sum_{p=0}^{\infty}\sum_{k=1}^{\infty}\sum_{s=1}^{\infty} a_{j,k}a_{j-t,s}b_{k,s}a_{p,k}a_{p-t,s}b_{k,s}. \tag{3.6.4}$$

Hence, by (3.6.4),

$$\|\gamma_t\|_{\mathcal{S}}^2 \le C_\varepsilon^2 \sum_{j=0}^{\infty}\sum_{p=0}^{\infty}\sum_{k=1}^{\infty}\sum_{s=1}^{\infty} |a_{j,k}||a_{j-t,s}||a_{p,k}||a_{p-t,s}| = C_\varepsilon^2 \sum_{j=0}^{\infty}\sum_{k=1}^{\infty} |a_{j,k}|c_{j,k}, \tag{3.6.5}$$

where

$$c_{j,k} = \sum_{p=0}^{\infty}\sum_{s=1}^{\infty} |a_{j-t,s}||a_{p,k}||a_{p-t,s}|.$$

Applying the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}\sum_{j=0}^{\infty}\sum_{k=1}^{\infty} |a_{j,k}|c_{j,k} &\le \left\{\left(\sum_{j=0}^{\infty}\sum_{k=1}^{\infty} a_{j,k}^2\right)\left(\sum_{j=0}^{\infty}\sum_{k=1}^{\infty} c_{j,k}^2\right)\right\}^{1/2} \\ &= C_a\left(\sum_{j=0}^{\infty}\sum_{k=1}^{\infty} c_{j,k}^2\right)^{1/2},\end{aligned} \tag{3.6.6}$$

where, by (3.3.7),

$$C_a := \left(\sum_{j=0}^{\infty}\sum_{k=1}^{\infty} a_{j,k}^2\right)^{1/2} < \infty.$$

Another application of the Cauchy-Schwarz inequality yields,

$$c_{j,k}^2 \le \left(\sum_{p=0}^{\infty}\sum_{s=1}^{\infty} a_{j-t,s}^2 a_{p,k}^2\right)\left(\sum_{p=0}^{\infty}\sum_{s=1}^{\infty} a_{p-t,s}^2\right) = \widetilde{C}_t \cdot \sum_{p=0}^{\infty}\sum_{s=1}^{\infty} a_{j-t,s}^2 a_{p,k}^2,$$

where

$$\widetilde{C}_t = \sum_{p=0}^{\infty} \sum_{k=1}^{\infty} a_{p-t,k}^2. \tag{3.6.7}$$

Thus,

$$\begin{aligned} \sum_{j=0}^{\infty} \sum_{k=1}^{\infty} c_{j,k}^2 &\leq \widetilde{C}_t \sum_{j=0}^{\infty} \sum_{k=1}^{\infty} \sum_{p=0}^{\infty} \sum_{s=1}^{\infty} a_{j-t,s}^2 a_{p,k}^2 \\ &= \widetilde{C}_t \left(\sum_{p=0}^{\infty} \sum_{k=1}^{\infty} a_{p,k}^2 \right) \left(\sum_{j=0}^{\infty} \sum_{s=1}^{\infty} a_{j-t,s}^2 \right) \\ &= C_a^2 \widetilde{C}_t^2. \end{aligned} \tag{3.6.8}$$

Combining (3.6.6) and (3.6.8), we get

$$\sum_{j=0}^{\infty} \sum_{k=1}^{\infty} |a_{j,k}| c_{j,k} \leq C_a^2 \widetilde{C}_t. \tag{3.6.9}$$

Combining (3.6.5) and (3.6.9), we conclude that

$$\|\gamma_t\|_{\mathcal{S}} \leq C_\varepsilon C_a \left(\widetilde{C}_t \right)^{1/2}. \tag{3.6.10}$$

Recall that we have assumed $-t > 0$ and recall the definition of $\widetilde{C}_t$ in (3.6.7). By (3.3.7), $\left(\widetilde{C}_t \right)^{1/2} = O\left((-t)^{-\kappa}\right)$. It follows that

$$\sum_{t=-\infty}^{-1} (-t)^{\varrho} O\left((-t)^{-\kappa}\right) = O\left(\sum_{t=1}^{\infty} t^{\varrho-\kappa} \right).$$

Since $\kappa > 4$, the last sum converges for $\varrho \leq 3$.

∎

Turning to the proof of Theorem 3.3.1, consider the series

$$\widetilde{R}_t = \begin{cases} X_t - \frac{1}{n^\star}\sum_{1\le s\le n^\star} X_s, & 1 \le t \le n^\star, \\ X_t - \frac{1}{N-n^\star}\sum_{n^\star < s \le N} X_s, & n^\star + 1 \le t \le N, \end{cases}$$

and the kernel

$$\tilde{c}_R(u,v) = \widetilde{\gamma}_{R,0}(u,v) + \sum_{t=1}^{N-1} K\left(\frac{t}{h}\right)\{\widetilde{\gamma}_{R,t}(u,v) + \widetilde{\gamma}_{R,t}(v,u)\},$$

where

$$\widetilde{\gamma}_{R,t}(u,v) = \frac{1}{N}\sum_{s=t+1}^{N}\left(\widetilde{R}_s(u) - \frac{1}{N}\sum_{i=1}^{N}\widetilde{R}_i(u)\right)\left(\widetilde{R}_{s-t}(v) - \frac{1}{N}\sum_{i=1}^{N}\widetilde{R}_i(v)\right), \quad 0 \le t \le N-1.$$

Let $\widetilde{C}_R$ be the integral operator with the kernel $\tilde{c}_R(u,v)$.

At the end of this section, we show that Theorem 3.3.1 follows from Lemmas 3.6.1 and 3.6.2, which we first formulate and prove.

LEMMA 3.6.1 *Under the assumptions of Theorem 3.3.4,* $\|\widehat{C}_R - \widetilde{C}_R\|_{\mathcal{S}} = O_P\left(\frac{h}{N}\right)$.

PROOF: For a bounded linear operator Φ with kernel ϕ such that $\iint \phi^2(u,v)dudv < \infty$, we denote $\|\Phi\|_{\mathcal{S}} = \left(\iint \phi^2(u,v)dudv\right)^{1/2} = \|\phi\|_{\mathcal{S}}$. Then,

$$\begin{aligned}\|\widehat{C}_R - \widetilde{C}_R\|_{\mathcal{S}} &= \|\hat{c}_R - \tilde{c}_R\|_{\mathcal{S}} \\ &\le \|\Upsilon_1\|_{\mathcal{S}} + \|\Upsilon_2\|_{\mathcal{S}},\end{aligned}$$

where

$$\Upsilon_1(u,v) = \sum_{t=0}^{N-1} K\left(\frac{t}{h}\right)[\widehat{\gamma}_{R,t}(u,v) - \widetilde{\gamma}_{R,t}(u,v)],$$
$$\Upsilon_2(u,v) = \sum_{t=1}^{N-1} K\left(\frac{t}{h}\right)[\widehat{\gamma}_{R,t}(v,u) - \widetilde{\gamma}_{R,t}(v,u)].$$

We shall show that $\|\Upsilon_1\|_{\mathcal{S}}, \|\Upsilon_2\|_{\mathcal{S}} = O_P\left(\frac{h}{N}\right)$. We can just focus on $\|\Upsilon_1\|_{\mathcal{S}}$, because $\|\Upsilon_2\|_{\mathcal{S}}$ can be addressed in a similar way.

Suppose $\hat{n} < n^\star$, the case of $\hat{n} > n^\star$ is very similar. We rewrite R_t and $\widetilde{R}_t$ as functions of Y_t:

$$R_t = \begin{cases} Y_t - \frac{1}{\hat{n}} \sum_{s=1}^{\hat{n}} Y_s, & 1 \le t \le \hat{n}, \\ Y_t - \frac{N-n^\star}{N-\hat{n}}\delta - \frac{1}{N-\hat{n}} \sum_{s=\hat{n}+1}^{N} Y_s, & \hat{n}+1 \le t \le n^\star, \\ Y_t + \frac{n^\star-\hat{n}}{N-\hat{n}}\delta - \frac{1}{N-\hat{n}} \sum_{s=\hat{n}+1}^{N} Y_s, & n^\star + 1 \le t \le N, \end{cases} \tag{3.6.11}$$

and

$$\widetilde{R}_t = \begin{cases} Y_t - \frac{1}{n^\star} \sum_{1 \le s \le n^\star} Y_s, & 1 \le t \le n^\star, \\ Y_t - \frac{1}{N-n^\star} \sum_{n^\star < s \le N} Y_s, & n^\star + 1 \le t \le N. \end{cases} \tag{3.6.12}$$

Note that

$$R_s(u) - \widetilde{R}_s(u) = \begin{cases} \frac{1}{n^\star} \sum_{i=1}^{n^\star} Y_i(u) - \frac{1}{\hat{n}} \sum_{i=1}^{\hat{n}} Y_i(u), & 1 \le s \le \hat{n}, \\ \frac{1}{n^\star} \sum_{i=1}^{n^\star} Y_i(u) - \frac{N-n^\star}{N-\hat{n}}\delta(u) - \frac{1}{N-\hat{n}} \sum_{i=\hat{n}+1}^{N} Y_i(u), & \hat{n} < s \le n^\star \\ \frac{1}{N-n^\star} \sum_{i=n^\star+1}^{N} Y_i(u) + \frac{n^\star-\hat{n}}{N-\hat{n}}\delta(u) - \frac{1}{N-\hat{n}} \sum_{i=\hat{n}+1}^{N} Y_i(u), & n^\star < s \le N. \end{cases}$$

Using Theorem 3.6.1, we examine several typical terms and note that almost identical arguments apply to the remaining terms. Observe that

$$\frac{1}{n^\star} \sum_{i=1}^{n^\star} Y_i = N^{-1/2} \frac{N}{n^\star} N^{-1/2} \sum_{i=1}^{n^\star} Y_i = O_P\left(N^{-1/2} \frac{1}{\tau^*} \max_{0 \le x \le 1} |S_N(x, \cdot)|\right) = O_P(N^{-1/2}),$$

$$\frac{1}{\hat{n}} \sum_{i=1}^{\hat{n}} Y_i = N^{-1/2} \frac{N}{\hat{n}} N^{-1/2} \sum_{i=1}^{\hat{n}} Y_i = O_P\left(N^{-1/2} \frac{1}{\tau^*} \max_{0 \le x \le 1} |S_N(x, \cdot)|\right) = O_P(N^{-1/2}),$$

because by (3.3.8)

$$\frac{N}{\hat{n}} = \frac{1}{\frac{\hat{n}-n^\star}{N} + \frac{n^\star}{N}} \xrightarrow{P} \frac{1}{\tau^*}.$$

To deal with terms involving the jump function δ, which are the dominating terms, we observe that

$$\frac{N-n^\star}{N-\hat{n}} = \frac{1}{1+\frac{n^*-\hat{n}}{N-n^\star}} = \frac{1}{1+O_P(N^{-1})} = O_P(1).$$

We conclude that

$$\max_{1\leq s\leq N} \left\| R_s - \widetilde{R}_s \right\| = O_P(1). \tag{3.6.13}$$

More specifically, denoting

$$R_s - \widetilde{R}_s =: \begin{cases} \Delta R_1^\star, & 1 \leq s \leq \hat{n}, \\ \Delta R_2^\star, & \hat{n}+1 \leq s \leq n^\star, \\ \Delta R_3^\star, & n^\star + 1 \leq s \leq N, \end{cases} \tag{3.6.14}$$

we conclude that

$$\|\Delta R_1^\star\|,\ \|\Delta R_3^\star\| = O_P(N^{-1/2}) \quad \text{and} \quad \|\Delta R_2^\star\| = O_P(1). \tag{3.6.15}$$

We use decompositions

$$R_s(u)R_{s-t}(v) - \widetilde{R}_s(u)\widetilde{R}_{s-t}(v) = [R_s(u) - \widetilde{R}_s(u)]R_{s-t}(v) + \widetilde{R}_s(u)[R_{s-t}(v) - \widetilde{R}_{s-t}(v)].$$

Then, for $t \in \{0, 1, \ldots, N\}$, we have

$$\begin{aligned} \widehat{\gamma}_{R,t}(u,v) - \widetilde{\gamma}_{R,t}(u,v) &= \frac{1}{N}\sum_{s=t+1}^{N} [R_s(u)R_{s-t}(v) - \widetilde{R}_s(u)\widetilde{R}_{s-t}(v)] \\ &= \iota_1(u,v) + \iota_2(u,v), \end{aligned}$$

where

$$\iota_1(u,v) = \frac{1}{N}\sum_{s=t+1}^{N}[R_s(u) - \widetilde{R}_s(u)]R_{s-t}(v),$$
$$\iota_2(u,v) = \frac{1}{N}\sum_{s=t+1}^{N}\widetilde{R}_s(u)[R_{s-t}(v) - \widetilde{R}_{s-t}(v)].$$

Define the index sets

$$\mathcal{I}_1(t) = \{t+1,\dots,N\} \cap \{1,\dots,\hat{n}\},$$

$$\mathcal{I}_2(t) = \{t+1,\dots,N\} \cap \{\hat{n}+1,\dots,n^\star\},$$

$$\mathcal{I}_3(t) = \{t+1,\dots,N\} \cap \{n^\star+1,\dots,N\}.$$

Observe that by (3.6.14),

$$\iota_1(u,v) = \iota_{11}(u,v) + \iota_{12}(u,v) + \iota_{13}(u,v),$$

where

$$\iota_{11}(u,v) = \frac{1}{N}\cdot\Delta R_1^\star(u)\cdot\sum_{s\in\mathcal{I}_1(t)} R_{s-t}(v),$$
$$\iota_{12}(u,v) = \frac{1}{N}\cdot\Delta R_2^\star(u)\cdot\sum_{s\in\mathcal{I}_2(t)} R_{s-t}(v),$$
$$\iota_{13}(u,v) = \frac{1}{N}\cdot\Delta R_3^\star(u)\cdot\sum_{s\in\mathcal{I}_3(t)} R_{s-t}(v).$$

Put

$$U_t = \begin{cases} \frac{1}{\hat{n}}\sum_{s=1}^{\hat{n}} Y_s, & 1 \le t \le \hat{n}, \\ \frac{N-n^\star}{N-\hat{n}}\delta + \frac{1}{N-\hat{n}}\sum_{s=\hat{n}+1}^{N} Y_s, & \hat{n}+1 \le t \le n^\star, \\ \frac{1}{N-\hat{n}}\sum_{s=\hat{n}+1}^{N} Y_s - \frac{n^\star-\hat{n}}{N-\hat{n}}\delta, & n^\star+1 \le t \le N, \end{cases}$$

so that $R_t = Y_t - U_t$. Observe that

$$\max_{1\le t\le N} \|U_t\| = O_P(1),$$

$$\|U_t\| = \begin{cases} O_P(N^{-1/2}), & 1 \le t \le \hat{n}, \\ O_P(1), & \hat{n}+1 \le t \le n^\star, \\ O_P(N^{-1/2}), & n^\star + 1 \le t \le N, \end{cases}$$

which can be obtained by the arguments used to get (3.6.13). Thus, by (3.3.8),

$$\begin{aligned} \sum_{t=1}^{N} \|U_t\| &= \sum_{1\le t\le \hat{n}} \|U_t\| + \sum_{\hat{n}<t\le n^\star} \|U_t\| + \sum_{n^\star<t\le N} \|U_t\| \\ &= O_P(N^{1/2}). \end{aligned}$$

It follows that

$$\left\| \sum_{s\in\mathcal{I}_k(t)} U_{s-t} \right\| \le \sum_{s\in\mathcal{I}_k(t)} \|U_{s-t}\| = \begin{cases} O_P(N^{1/2}), & k = 1, 3, \\ O_P(1), & k = 2, \end{cases} \tag{3.6.16}$$

because the cardinality of $\mathcal{I}_2(t)$ satisfies $|\mathcal{I}_2(t)| \le |n^\star - \hat{n}| = O_P(1)$.

On the other hand, we know that

$$\left\| \sum_{s\in\mathcal{I}_k(t)} Y_{s-t} \right\| = \begin{cases} O_P(N^{1/2}), & k = 1, 3, \\ O_P(1), & k = 2. \end{cases} \tag{3.6.17}$$

Combining (3.6.16) and (3.6.17), we have

$$\left\| \sum_{s \in \mathcal{I}_k(t)} R_{s-t} \right\| \leq \left\| \sum_{s \in \mathcal{I}_k(t)} Y_{s-t} \right\| + \left\| \sum_{s \in \mathcal{I}_k(t)} U_{s-t} \right\| = \begin{cases} O_P(N^{1/2}), & k = 1, 3, \\ O_P(1), & k = 2. \end{cases} \tag{3.6.18}$$

Hence, by (3.6.15) and (3.6.18),

$$\|\iota_{1k}\|_{\mathcal{S}}^2 = \frac{1}{N^2} \cdot \|\Delta R_k^\star\|^2 \cdot \left\| \sum_{s \in \mathcal{I}_k(t)} R_{s-t} \right\|^2 = O_P(N^{-2}), \quad k = 1, 2, 3,$$

which implies that

$$\|\iota_1\|_{\mathcal{S}} \leq \|\iota_{11}\|_{\mathcal{S}} + \|\iota_{12}\|_{\mathcal{S}} + \|\iota_{13}\|_{\mathcal{S}} = O_P(N^{-1}).$$

Now we deal with $\|\iota_2\|_{\mathcal{S}}$. Define the index sets

$$\mathcal{K}_1(t) = \{t+1, \dots, N\} \cap \{t+1, \dots, t+\hat{n}\},$$

$$\mathcal{K}_2(t) = \{t+1, \dots, N\} \cap \{t+\hat{n}+1, \dots, t+n^\star\},$$

$$\mathcal{K}_3(t) = \{t+1, \dots, N\} \cap \{t+n^\star+1, \dots, N\},$$

and

$$\mathcal{L}_1(t) = \{1, \dots, n^\star\},$$

$$\mathcal{L}_2(t) = \{n^\star+1, \dots, N\}.$$

Notice that by (3.6.14),

$$\iota_2(u, v) = \iota_{21}(u, v) + \iota_{22}(u, v) + \iota_{23}(u, v),$$

where

$$\iota_{21}(u,v) = \frac{1}{N} \cdot \Delta R_1^\star(v) \cdot \sum_{s \in \mathcal{K}_1(t)} \widetilde{R}_s(u),$$
$$\iota_{22}(u,v) = \frac{1}{N} \cdot \Delta R_2^\star(v) \cdot \sum_{s \in \mathcal{K}_2(t)} \widetilde{R}_s(u),$$
$$\iota_{23}(u,v) = \frac{1}{N} \cdot \Delta R_3^\star(v) \cdot \sum_{s \in \mathcal{K}_3(t)} \widetilde{R}_s(u).$$

By (3.6.12),

$$\begin{aligned} \left\| \sum_{s \in \mathcal{K}_j(t)} \widetilde{R}_s \right\| &\leq \left\| \sum_{s \in \mathcal{K}_j(t)} Y_s \right\| + |\mathcal{K}_j(t) \cap \mathcal{L}_1(t)| \cdot \left\| \frac{1}{n^\star} \sum_{1 \leq i \leq n^\star} Y_i \right\| \\ &\quad + |\mathcal{K}_j(t) \cap \mathcal{L}_2(t)| \cdot \left\| \frac{1}{N - n^\star} \sum_{n^\star < i \leq N} Y_i \right\| \\ &= \begin{cases} O_P(N^{1/2}), & j = 1, 3, \\ O_P(1), & j = 2, \end{cases} \end{aligned}$$

which, together with (3.6.15), implies that

$$\|\iota_{2j}\|_{\mathcal{S}}^2 = \frac{1}{N^2} \cdot \|\Delta R_j^\star\|^2 \cdot \left\| \sum_{s \in \mathcal{K}_j(t)} \widetilde{R}_s \right\|^2 = O_P(N^{-2}), \quad j = 1, 2, 3.$$

So, we readily have

$$\|\widehat{\gamma}_{R,t} - \widetilde{\gamma}_{R,t}\|_{\mathcal{S}} \leq \|\iota_1\|_{\mathcal{S}} + \|\iota_2\|_{\mathcal{S}} = O_P(N^{-1}).$$

Without loss of generality, we assume that $K(u) = 0$ if $|u| > 1$. Since K is bounded,

$$\sum_{t=0}^{N-1} \left| K\left(\frac{t}{h}\right) \right| = \sum_{t=0}^{h} \left| K\left(\frac{t}{h}\right) \right| = O(h). \tag{3.6.19}$$

Consequently,

$$\begin{aligned}\|\Upsilon_1\|_{\mathcal{S}} &\le \sum_{t=0}^{N-1} \left| K\left(\frac{t}{h}\right) \right| \|\widehat{\gamma}_{R,t} - \widetilde{\gamma}_{R,t}\|_{\mathcal{S}} \\ &= O_P\left(\frac{h}{N}\right).\end{aligned}$$

∎

LEMMA 3.6.2 *Under the assumptions of Theorem 3.3.4,*

$$\|\widetilde{C}_R - C\|_{\mathcal{S}} = O_P\left(\sqrt{\frac{h}{N}} + h^{-q}\right).$$

PROOF: Define

$$Y_1^\star(u) = \frac{1}{n^\star} \sum_{i=1}^{n^\star} Y_i(u), \quad Y_2^\star(u) = \frac{1}{N - n^\star} \sum_{i=n^\star+1}^{N} Y_i(u).$$

Observe that

$$\tilde{c}_R(u,v) = \sum_{t=-(N-1)}^{N-1} K\left(\frac{t}{h}\right) \widetilde{\gamma}_{R,t}(u,v), \tag{3.6.20}$$

where

$$\widetilde{\gamma}_{R,t}(u,v) = \begin{cases} \frac{1}{N}\sum_{s=t+1}^{N} \left(Y_s(u) - Y_k^\star(u)\right)\left(Y_{s-t}(v) - Y_l^\star(v)\right), & 0 \le t \le N-1, \\ \frac{1}{N}\sum_{s=1-t}^{N} \left(Y_{s+t}(u) - Y_l^\star(u)\right)\left(Y_s(v) - Y_k^\star(v)\right), & -(N-1) \le t < 0, \end{cases}$$

with the convention that for $t \ge 0$,

$$Y_k^\star(u) = \begin{cases} Y_1^\star(u), & \text{if } s \in \{1, \dots, n^\star\} \cap \{t+1, \dots, N\}, \\ Y_2^\star(u), & \text{if } s \in \{n^\star+1, \dots, N\} \cap \{t+1, \dots, N\}, \end{cases}$$

and

$$Y_l^\star(v) = \begin{cases} Y_1^\star(v), & \text{if } s \in \{t+1, \dots, t+n^\star\} \cap \{t+1, \dots, N\}, \\ Y_2^\star(v), & \text{if } s \in \{t+n^\star+1, \dots, N\} \cap \{t+1, \dots, N\}. \end{cases}$$

Similarly, for $t < 0$,

$$Y_l^\star(u) = \begin{cases} Y_1^\star(u), & \text{if } s \in \{1-t, \dots, n^\star - t\} \cap \{1-t, \dots, N\}, \\ Y_2^\star(u), & \text{if } s \in \{n^\star + 1 - t, \dots, N\} \cap \{1-t, \dots, N\}, \end{cases}$$

and

$$Y_k^\star(v) = \begin{cases} Y_1^\star(v), & \text{if } s \in \{1, \dots, n^\star\} \cap \{1-t, \dots, N\}, \\ Y_2^\star(v), & \text{if } s \in \{n^\star+1, \dots, N\} \cap \{1-t, \dots, N\}. \end{cases}$$

We first show that replacing $Y_k^\star$ and $Y_l^\star$ with $EY_0 = 0$ has a negligible effect on the convergence rate of $\|\widetilde{C}_R - C\|_{\mathcal{S}}$. Define

$$\widetilde{\gamma}_t(u,v) = \begin{cases} \frac{1}{N}\sum_{s=t+1}^N Y_s(u)Y_{s-t}(v), & 0 \le t \le N-1, \\ \frac{1}{N}\sum_{s=1-t}^N Y_{s+t}(u)Y_s(v), & -(N-1) \le t < 0, \end{cases}$$

and

$$\tilde{c}(u,v) = \sum_{t=-(N-1)}^{N-1} K\left(\frac{t}{h}\right)\widetilde{\gamma}_t(u,v).$$

Let $\widetilde{C}$ be the covariance operator with kernel $\tilde{c}$. We claim that

$$\|\widetilde{C}_R - \widetilde{C}\|_{\mathcal{S}} = O_P\left(\frac{h}{N}\right). \tag{3.6.21}$$

Note that

$$\widetilde{\gamma}_{R,t}(u,v) - \widetilde{\gamma}_t(u,v)$$
$$= \begin{cases} \frac{1}{N}\sum_{s=t+1}^{N}\left[Y_k^\star(u)Y_l^\star(v) - Y_s(u)Y_l^\star(v) - Y_k^\star(u)Y_{s-t}(v)\right] & 0 \le t \le N-1, \\ \frac{1}{N}\sum_{s=1-t}^{N}\left[Y_l^\star(u)Y_k^\star(v) - Y_{s+t}(u)Y_k^\star(v) - Y_l^\star(u)Y_s(v)\right] & -(N-1) \le t < 0. \end{cases}$$

We shall show $\|\widetilde{\gamma}_{R,t} - \widetilde{\gamma}_t\|_{\mathcal{S}} = O_P(N^{-1})$ for all $|t| < N$. We will only consider $0 \le t < N$ because the case $-N < t < 0$ can be dealt in a quite similar way. Denote

$$\widetilde{\gamma}_{R,t}(u,v) - \widetilde{\gamma}_t(u,v) = \varpi_1(u,v) - \varpi_2(u,v) - \varpi_3(u,v),$$

where

$$\varpi_1(u,v) = \frac{1}{N}\sum_{s=t+1}^{N} Y_k^\star(u)Y_l^\star(v),$$
$$\varpi_2(u,v) = \frac{1}{N}\sum_{s=t+1}^{N} Y_s(u)Y_l^\star(v),$$
$$\varpi_3(u,v) = \frac{1}{N}\sum_{s=t+1}^{N} Y_k^\star(u)Y_{s-t}(v).$$

By Theorem 3.6.1 (see also Theorem 16.3 in Horváth and Kokoszka (2012), which shows that L^2-approximability is sufficient),

$$\|\bar{Y}_N\|,\ \|Y_1^\star\|,\ \|Y_2^\star\| = O_P(N^{-1/2}), \tag{3.6.22}$$

which implies

$$\|Y_k^\star\|,\ \|Y_l^\star\| = O_P(N^{-1/2}). \tag{3.6.23}$$

By the Cauchy-Schwarz inequality and (3.6.23),

$$\begin{aligned}\|\varpi_1\|_{\mathcal{S}}^2 &= \iint \frac{1}{N^2}\sum_{s=t+1}^{N}\sum_{s'=t+1}^{N} Y_k^\star(u)Y_{k'}^\star(u)Y_l^\star(v)Y_{l'}^\star(v)dudv \\ &\le \frac{1}{N^2}\sum_{s=t+1}^{N}\sum_{s'=t+1}^{N}\left(\left|\int Y_k^\star(u)Y_{k'}^\star(u)du\right|\cdot\left|\int Y_l^\star(v)Y_{l'}^\star(v)dv\right|\right) \\ &\le \frac{1}{N^2}\sum_{s=t+1}^{N}\sum_{s'=t+1}^{N}\|Y_k^\star\|\|Y_{k'}^\star\|\|Y_l^\star\|\|Y_{l'}^\star\| \\ &= O_P(N^{-2}).\end{aligned}$$

Put

$$\mathcal{J}_1(t) = \{t+1,\dots,t+n^\star\}\cap\{t+1,\dots,N\},$$

$$\mathcal{J}_2(t) = \{t+n^\star+1,\dots,N\}\cap\{t+1,\dots,N\}.$$

Then, we have

$$\begin{aligned}\|\varpi_2\|_{\mathcal{S}}^2 &= \frac{1}{N^2}\iint\left(\sum_{s\in\mathcal{J}_1(t)} Y_s(u)Y_1^\star(v) + \sum_{s\in\mathcal{J}_2(t)} Y_s(u)Y_2^\star(v)\right)^2 dudv \\ &= \frac{1}{N^2}\Bigg\{\|Y_1^\star\|^2\cdot\left\|\sum_{s\in\mathcal{J}_1(t)} Y_s\right\|^2 + \|Y_2^\star\|^2\cdot\left\|\sum_{s\in\mathcal{J}_2(t)} Y_s\right\|^2 \\ &\quad + 2\langle Y_1^\star, Y_2^\star\rangle\cdot\left\langle\sum_{s\in\mathcal{J}_1(t)} Y_s, \sum_{r\in\mathcal{J}_2(t)} Y_r\right\rangle\Bigg\} \\ &\le \frac{1}{N^2}\Bigg\{\|Y_1^\star\|^2\cdot\left\|\sum_{s\in\mathcal{J}_1(t)} Y_s\right\|^2 + \|Y_2^\star\|^2\cdot\left\|\sum_{s\in\mathcal{J}_2(t)} Y_s\right\|^2 \\ &\quad + 2\|Y_1^\star\|\|Y_2^\star\|\cdot\left\|\sum_{s\in\mathcal{J}_1(t)} Y_s\right\|\left\|\sum_{s\in\mathcal{J}_2(t)} Y_s\right\|\Bigg\} \\ &= O_P(N^{-2}).\end{aligned}$$

Using an analogous argument we can show that $\|\varpi_3\|_{\mathcal{S}}^2 = O_P(N^{-2})$. Thus, we have finished proving $\|\widetilde{\gamma}_{R,t} - \widetilde{\gamma}_t\|_{\mathcal{S}} = O_P(N^{-1})$. Note that by (3.6.19),

$$\begin{aligned}\|\widetilde{C}_R - \widetilde{C}\|_{\mathcal{S}} &= \left\| \sum_{|t|<N} K\left(\frac{t}{h}\right)(\widetilde{\gamma}_{R,t} - \widetilde{\gamma}_t) \right\|_{\mathcal{S}} \\ &\leq \sum_{|t|<N} \left| K\left(\frac{t}{h}\right) \right| \|\widetilde{\gamma}_{R,t} - \widetilde{\gamma}_t\|_{\mathcal{S}} \\ &= O_P\left(\frac{h}{N}\right).\end{aligned}$$

This completes the verification of (3.6.21). Using arguments similar to those developed in the proofs of Theorem 2.3 and Lemma 4.3 in Berkes *et al.* (2016), we now show that

$$\|\widetilde{C} - C\|_{\mathcal{S}} = O_P\left(\sqrt{\frac{h}{N}} + h^{-q}\right). \tag{3.6.24}$$

Note that

$$\begin{aligned}\|\widetilde{C} - C\|_{\mathcal{S}} &= \|\tilde{c} - E\tilde{c} + E\tilde{c} - c_Y\|_{\mathcal{S}} \\ &\leq \|\tilde{c} - E\tilde{c}\|_{\mathcal{S}} + \|E\tilde{c} - c_Y\|_{\mathcal{S}}.\end{aligned}$$

Without loss of generality, we assume $c = 1$ in Assumption 3.3.3. Simple calculations yield

$$\begin{aligned}E[\tilde{c}(u,v)] &= \sum_{t=-(N-1)}^{N-1} K\left(\frac{t}{h}\right)\left(1 - \frac{|t|}{N}\right) E\left\{Y_t(u)Y_0(v)\right\} \\ &= \sum_{t=-(N-1)}^{N-1} K\left(\frac{t}{h}\right)\gamma_t(u,v) - \frac{1}{N}\sum_{t=-(N-1)}^{N-1} K\left(\frac{t}{h}\right)|t|\gamma_t(u,v) \\ &= \sum_{t=-\infty}^{\infty} K\left(\frac{t}{h}\right)\gamma_t(u,v) - \frac{1}{N}\sum_{t=-\infty}^{\infty} K\left(\frac{t}{h}\right)|t|\gamma_t(u,v),\end{aligned}$$

where $\gamma_t(u,v) = E\left\{Y_t(u)Y_0(v)\right\}$. Note that by Assumption 3.3.3 and (3.3.9)

$$\begin{aligned}
\left\| \frac{1}{N} \sum_{t=-\infty}^{\infty} K\left(\frac{t}{h}\right)|t|\gamma_t \right\|_{\mathcal{S}} &= \left\| \frac{1}{N} \sum_{|t|\le h} K\left(\frac{t}{h}\right)|t|\gamma_t \right\|_{\mathcal{S}} \\
&\le \frac{1}{N} \sup_{x\in[-1,1]} |K(x)| \cdot h \sum_{|t|\le h} \|\gamma_t\|_{\mathcal{S}} \\
&= O\left(\frac{h}{N}\right).
\end{aligned}$$

Now fix $\varepsilon > 0$. Then we have

$$\sum_{t=-\infty}^{\infty} K\left(\frac{t}{h}\right)\gamma_t(u,v) = \sum_{t=-\infty}^{\infty} \gamma_t(u,v) + f_{N,1,\varepsilon}(u,v) + f_{N,2,\varepsilon}(u,v) - f_{N,3,\varepsilon}(u,v),$$

where

$$f_{N,1,\varepsilon}(u,v) = \sum_{|t|\le \varepsilon h} \left[K\left(\frac{t}{h}\right) - 1\right]\gamma_t(u,v),$$

$$f_{N,2,\varepsilon}(u,v) = \sum_{|t|>\varepsilon h} K\left(\frac{t}{h}\right)\gamma_t(u,v)$$

and

$$f_{N,3,\varepsilon}(u,v) = \sum_{|t|>\varepsilon h} \gamma_t(u,v).$$

By (3.3.9), we know that

$$\begin{aligned}
\|f_{N,2,\varepsilon}\|_{\mathcal{S}} &\le \sup_{x\in[-1,1]} |K(x)| \sum_{|t|>\varepsilon h} \|\gamma_t\|_{\mathcal{S}} \\
&\le \sup_{x\in[-1,1]} |K(x)| \sum_{|t|>\varepsilon h} \frac{|t|^{\varrho}}{(\varepsilon h)^{\varrho}} \|\gamma_t\|_{\mathcal{S}} \\
&\le \sup_{x\in[-1,1]} |K(x)| \sum_{t=-\infty}^{\infty} \frac{|t|^{\varrho}}{(\varepsilon h)^{\varrho}} \|\gamma_t\|_{\mathcal{S}} \\
&= O(h^{-\varrho}).
\end{aligned}$$

Similarly, $\|f_{N,3,\varepsilon}\|_{\mathcal{S}} = O(h^{-\varrho})$.

By equation (3.3.4), we have

$$\left|K\left(\frac{t}{h}\right)-1\right| = \zeta(|t|/h)^q + o\left((|t|/h)^q\right), \quad \text{as } N\to\infty.$$

It follows that

$$\begin{aligned}\|h^q f_{N,1,\varepsilon}\|_{\mathcal{S}} &\le h^q \cdot \sum_{|t|\le\varepsilon h}\left|K\left(\frac{t}{h}\right)-1\right| \|\gamma_t\|_{\mathcal{S}} \\ &= \sum_{|t|\le\varepsilon h}\{\zeta|t|^q + o(|t|^q)\}\|\gamma_t\|_{\mathcal{S}} \\ &= \zeta\sum_{|t|\le\varepsilon h}|t|^q\|\gamma_t\|_{\mathcal{S}} + o\left(\sum_{|t|\le\varepsilon h}|t|^q\|\gamma_t\|_{\mathcal{S}}\right) \\ &= O(1).\end{aligned}$$

Thus, we obtain, as in Berkes *et al.* (2016),

$$\|E\tilde{c}-c_Y\|_{\mathcal{S}} = O(h^{-\varrho}) + O(h^{-q}) = O(h^{-q}).$$

By Theorems 2.1 and 2.2 in Horváth *et al.* (2016) and Lemma 4.1 in Berkes *et al.* (2016), we deduce that

$$\|\tilde{c}-E\tilde{c}\|_{\mathcal{S}} = o_P\left(\sqrt{\frac{h}{N}}\right) + O_P\left(\sqrt{\frac{h}{N}}\right) = O_P\left(\sqrt{\frac{h}{N}}\right).$$

It follows that

$$\|\widetilde{C}-C\|_{\mathcal{S}} = O_P\left(\sqrt{\frac{h}{N}} + h^{-q}\right).$$

Combining (3.6.21) and (3.6.24), we obtain

$$\|\widetilde{C}_R - C\|_{\mathcal{S}} \le \|\widetilde{C}_R-\widetilde{C}\|_{\mathcal{S}} + \|\widetilde{C}-C\|_{\mathcal{S}} = O_P\left(\sqrt{\frac{h}{N}} + h^{-q}\right).$$

∎

PROOF OF THEOREM 3.3.1: By (3.3.5) in Assumption 3.3.3,

$$\sqrt{\frac{h}{N}} = O\left(N^{\frac{-q}{1+2q}}\right) \quad \text{and} \quad h^{-q} = O\left(N^{\frac{-q}{1+2q}}\right).$$

Thus, the convergence rate in Lemma 3.6.1 is $\|\widehat{C}_R - \widetilde{C}_R\|_{\mathcal{S}} = O_P(N^{-\frac{2q}{1+2q}})$ and the convergence rate in Lemma 3.6.2 is $\|\widetilde{C}_R - C\|_{\mathcal{S}} = O_P(N^{-\frac{q}{1+2q}})$. Since

$$\|\widehat{C}_R - C\|_{\mathcal{S}} \le \|\widehat{C}_R - \widetilde{C}_R\|_{\mathcal{S}} + \|\widetilde{C}_R - C\|_{\mathcal{S}},$$

we get in Theorem 3.3.1, $\|\widehat{C}_R - C\|_{\mathcal{S}} = O_P(N^{-\frac{q}{1+2q}})$.

∎

3.6.2 Proofs of Propositions 3.3.1 and 3.3.2

PROOF OF PROPOSITION 3.3.1: The Bernoulli shift representation is obvious, $EY_0 = 0$ is clear, and $E\|Y_0\|^4 < \infty$ follows from the argument used verify (3.3.2), to which we now proceed. Observe that

$$Y_{t,m} = \sum_{j=0}^{m-1} A_j(\varepsilon_{t-j}) + \sum_{j\ge m} A_j\left(\varepsilon_{t,t-j}^{(m)}\right),$$

so

$$Y_t - Y_{t,m} = \sum_{j\ge m} A_j(\varepsilon_{t-j}) - \sum_{j\ge m} A_j\left(\varepsilon_{t,t-j}^{(m)}\right).$$

Thus, for an absolute constant C,

$$E\|Y_t - Y_{t,m}\|^4 \le CE\left\|\sum_{j\ge m} A_j(\varepsilon_{t-j})\right\|^4.$$

Next, we use the expansion,

$$\left\|\sum_{j\ge m} A_j(\varepsilon_{t-j})\right\|^4 = \sum_{j,k,l,n\ge m} \langle A_j(\varepsilon_{t-j}), A_k(\varepsilon_{t-k})\rangle\langle A_l(\varepsilon_{t-l}), A_n(\varepsilon_{t-n})\rangle.$$

Recall that if $X, Y \in L^2$ are independent, square integrable and $EX = 0$, then $E[\langle X, Y\rangle] = 0$. Since the expectation commutes with bounded linear operators, $E[A_j(\varepsilon_{t-j})] = 0$. Therefore, as exploited in many similar arguments, the only cases of indices contributing to the expected value of the quadruple sum are:

$$j = k = l = n,\ j = k \neq l = n,\ j = l \neq k = n,\ j = n \neq k = l.$$

If $j = k = l = n$, the corresponding part of the sum is bounded from above by

$$\sum_{j \geq m} \|A_j(\varepsilon_{t-j})\|^4 \leq \|\varepsilon_0\|^4 \sum_{j \geq m} \|A_j\|_{\mathcal{L}}^4.$$

Incorporating the remaining three sums, we see that

$$\left\| \sum_{j \geq m} A_j(\varepsilon_{t-j}) \right\|^4 \leq \|\varepsilon_0\|^4 \left\{ \sum_{j \geq m} \|A_j\|_{\mathcal{L}}^4 + 3 \left(\sum_{j \geq m} \|A_j\|_{\mathcal{L}}^2 \right)^2 \right\}.$$

Thus, (3.3.2) will follow if

$$\left\{ \sum_{j \geq m} \|A_j\|_{\mathcal{L}}^4 \right\}^{1/4} = O\left(m^{-\kappa}\right) \quad \text{and} \quad \left\{ \sum_{j \geq m} \|A_j\|_{\mathcal{L}}^2 \right\}^{1/2} = O\left(m^{-\kappa}\right).$$

Recall that the norm in the space ℓ^p is defined by $\|x\|_p = \left(\sum_{j \geq 0} |x_j|^p\right)^{1/p}$, $p \geq 1$, and the function $p \mapsto \|x\|_p$ is decreasing. Thus, condition (3.3.11) is sufficient.

∎

PROOF OF PROPOSITION 3.3.2: The first claim follows from Proposition 3.3.1. To verify the second claim, observe that

$$\begin{aligned}\xi_{jt}^{(Y)} &= \sum_{k=0}^{\infty}\langle A_k(\varepsilon_{t-k}), \psi_j^{(Y)}\rangle \\ &= \sum_{k=0}^{\infty}\left\langle \sum_{l=1}^{\infty} a_{k,l}\langle\varepsilon_{t-k}, \varphi_l\rangle\psi_l^{(Y)}, \psi_j^{(Y)}\right\rangle \\ &= \sum_{k=0}^{\infty} a_{k,j}\langle\varepsilon_{t-k}, \varphi_j\rangle.\end{aligned}$$

Note that condition (3.3.7) implies that $\sum_{k=0}^{\infty} a_{k,j}^2 = O(1)$. Thus, condition (3.3.12) holds with $a_k = a_{k,j}$ and $\zeta_t = \langle\varepsilon_t, \varphi_j\rangle$.

∎

3.6.3 Proofs of Theorems 3.3.2 and 3.3.3

The asymptotic behavior of weighted sum of periodograms has been thoroughly discussed in Giraitis *et al.* (2012). Theorems 3.3.2 and 3.3.3 are like special consequences of results in their Chapter 5 and 6.

For a sequence of real numbers $\{b_{N,\ell}, \ell = 1, \ldots, [N/2]-1\}$, put

$$Q_N = \sum_{1\le\ell\le[N/2]-1} b_{N,\ell} I_j^{(Y)}(\omega_\ell), \quad \mu_N = \sum_{1\le\ell\le[N/2]-1} b_{N,\ell} f_j(\omega_\ell),$$

and

$$S_N = \sum_{1\le\ell\le[N/2]-1} b_{N,\ell}\frac{I_j^{(Y)}(\omega_\ell)}{f_j(\omega_\ell)}, \quad B_{f,N}^2 = \sum_{1\le\ell\le[N/2]-1} (b_{N,\ell} f_j(\omega_\ell))^2.$$

PROOF OF THEOREM 3.3.2: This result basically follows from Theorem 6.2.5 in Giraitis *et al.* (2012). Consider

$$b_{N,\ell} = \begin{cases} m^{-1}, & \text{if } \ell = 1, \ldots, m, \\ 0, & \text{if } \ell = m+1, \ldots, [N/2]-1. \end{cases}$$

Then, for some positive constant C,

$$Q_N = \frac{1}{m}\sum_{\ell=1}^{m} I_j^{(Y)}(\omega_\ell), \quad \mu_N = \frac{1}{m}\sum_{\ell=1}^{m} f_j(\omega_\ell),$$

and

$$B_{f,N}^2 \le \max_{\omega\in[0,\pi]} f_j(\omega)\sum_{\ell=1}^{m}\frac{1}{m^2} \le \frac{C}{m}.$$

Next, we use the decomposition

$$\begin{aligned}\frac{1}{m}\sum_{\ell=1}^{m}\left(I_j^{(Y)}(\omega_\ell) - f_j(0)\right) &= \left[\frac{1}{m}\sum_{\ell=1}^{m} I_j^{(Y)}(\omega_\ell) - \frac{1}{m}\sum_{\ell=1}^{m} f_j(\omega_\ell)\right] + \left[\frac{1}{m}\sum_{\ell=1}^{m} f_j(\omega_\ell) - f_j(0)\right] \\ &= [Q_N - \mu_N] + \left[\frac{1}{m}\sum_{\ell=1}^{m} f_j(\omega_\ell) - f_j(0)\right].\end{aligned}$$

By Theorem 6.2.5 in Giraitis *et al.* (2012), $E[Q_N - \mu_N]^2 \le \frac{C}{m}$, so $Q_N - \mu_N = O_P(m^{-1/2})$. Since $f_j \in \Lambda_\beta([0,\pi])$,

$$\begin{aligned}\sqrt{m}\left|\frac{1}{m}\sum_{\ell=1}^{m} f_j(\omega_\ell) - f_j(0)\right| &\le \frac{1}{\sqrt{m}}\sum_{\ell=1}^{m}|f_j(\omega_\ell) - f_j(0)| \\ &\le \frac{C}{\sqrt{m}}\sum_{\ell=1}^{m}\omega_\ell^\beta \\ &= O\left(\frac{m^{\beta+\frac{1}{2}}}{N^\beta}\right).\end{aligned} \tag{3.6.25}$$

The claim thus follows from condition (3.3.13).

∎

PROOF OF THEOREM 3.3.3: Put

$$b_{N,\ell} = \begin{cases}\frac{\nu_{\ell,m}}{f_j(0)}, & \text{if } \ell = 1,\dots,m, \\ 0, & \text{if } \ell = m+1,\dots,[N/2]-1.\end{cases}$$

Then, by Theorem 6.2.3 in Giraitis *et al.* (2012)

$$\mu_N = \sum_{\ell=1}^{m} \frac{\nu_{\ell,m}}{f_j(0)} f_j(\omega_\ell),$$

and

$$\upsilon_N^{-1} \sum_{\ell=1}^{m} \nu_{\ell,m} \left(\frac{I_j^{(Y)}(\omega_\ell)}{f_j(0)} - \frac{f_j(\omega_\ell)}{f_j(0)} \right) \xrightarrow{d} \mathcal{N}(0,1), \tag{3.6.26}$$

where υ_N^2 can be derived using Lemma 6.2.2 in Giraitis *et al.* (2012):

$$\upsilon_N^2 = \sum_{\ell=1}^{m} \nu_{\ell,m}^2 \frac{f_j^2(\omega_\ell)}{f_j^2(0)} + \text{cum}_4 \left(\langle Y_0, \psi_j^{(Y)} \rangle \right) \frac{1}{N} \left(\sum_{\ell=1}^{m} \nu_{\ell,m} \frac{f_j(\omega_\ell)}{f_j(0)} \right)^2 .$$

We first show that

$$\frac{\upsilon_N^2}{m} \to 1, \quad \text{as } N \to \infty. \tag{3.6.27}$$

Relation (3.6.27) will follow, once we have verified that

$$\frac{1}{m} \sum_{\ell=1}^{m} \nu_{\ell,m}^2 f_j^2(\omega_\ell) \to f_j^2(0), \tag{3.6.28}$$

and

$$\frac{1}{N} \sum_{\ell=1}^{m} \nu_{\ell,m}^2 f_j^2(\omega_\ell) \to 0. \tag{3.6.29}$$

Since f_j is bounded on $[0,\pi]$, (3.6.29) follows from (3.6.31). To verify (3.6.28), again by (3.6.31), it is enough to show that

$$\frac{1}{m} \sum_{\ell=1}^{m} \nu_{\ell,m}^2 [f_j^2(\omega_\ell) - f_j^2(0)] \to 0.$$

By Lipschitz continuity, this reduces to showing that

$$\frac{1}{N^\beta m} \sum_{\ell=1}^{m} \nu_{\ell,m}^2 \ell^\beta \to 0.$$

The above relation follows from (3.3.13) and (3.6.30) because

$$\frac{1}{N^\beta m}\sum_{\ell=1}^m \nu_{\ell,m}^2 \ell^\beta \le \frac{C}{N^\beta m}(\log m)^2 m^{1+\beta} = C\frac{m^{\beta+\frac{1}{2}}}{N^\beta}\frac{(\log m)^2}{\sqrt{m}}.$$

Combining (3.6.26) and (3.6.27), we conclude that

$$\frac{1}{\sqrt{m}}\sum_{\ell=1}^m \nu_{\ell,m}\left(\frac{I_j^{(Y)}(\omega_\ell)}{f_j(0)} - \frac{f_j(\omega_\ell)}{f_j(0)}\right) \xrightarrow{d} \mathcal{N}(0,1),$$

so to prove the claim, it suffices to observe that

$$\frac{1}{\sqrt{m}}\sum_{\ell=1}^m \nu_{\ell,m}[f_j(\omega_\ell) - f_j(0)] \to 0,$$

which follows from (3.6.25) and (3.3.14).

∎

LEMMA 3.6.3 *For the* $\nu_{\ell,m}$ *defined in* (3.3.15)*,*

$$\lim_{m\to\infty}\frac{\max_{1\le\ell\le m}|\nu_{\ell,m}|}{\left(\sum_{\ell=1}^m \nu_{\ell,m}^2\right)^{1/2}} = 0.$$

PROOF: Lemma 2 in Robinson (1995) states that

$$\frac{1}{m}\sum_{j=1}^m \log j = \log m - 1 + r_m, \quad |r_m| \le \frac{2+\log(m-1)}{m}.$$

Therefore,

$$\max_{1\le\ell\le m}|\nu_{\ell,m}| = O(\log m). \tag{3.6.30}$$

Also,

$$\frac{1}{m}\sum_{\ell=1}^m \nu_{\ell,m}^2 \sim \frac{1}{m}\sum_{\ell=1}^m \left\{\log\left(\frac{\ell}{m}\right)+1\right\}^2 \to \int_0^1 (\log x + 1)^2\, dx = 1. \tag{3.6.31}$$

∎

3.6.4 Proof of Theorem 3.3.4

The proof relies on a number of lemmas, some with long and technical proofs. To avoid disrupting the flow of arguments, the lemmas and their proofs are collected in Section 3.6.5

PROOF OF THEOREM 3.3.4: Recall the definition of $L(H)$ given in (3.2.7). Throughout the whole proof, we set $H_0 = \frac{1}{2}$. By applying the mean value theorem to $\frac{dL}{dH}$, we have

$$0 = \frac{dL(\widehat{H}^{(R)})}{dH} = \frac{dL(H_0)}{dH} + \frac{d^2L(H^*)}{dH^2}(\widehat{H}^{(R)} - H_0),$$

where H^* lies between $\widehat{H}^{(R)}$ and H_0. Hence,

$$\sqrt{m}\left(\widehat{H}^{(R)} - H_0\right) = -\sqrt{m}\frac{dL(H_0)/dH}{d^2L(H^*)/dH^2}$$

will imply the asymptotic normality once we have shown that

$$\frac{d^2L(H^*)}{dH^2} \xrightarrow{P} 4, \tag{3.6.32}$$

$$\sqrt{m}\frac{dL(H_0)}{dH} \xrightarrow{d} \mathcal{N}(0,4). \tag{3.6.33}$$

Put, for $k = 0, 1, 2,$

$$\widehat{G}_k(H) := \frac{1}{m}\sum_{\ell=1}^{m}(\log\omega_\ell)^k\omega_\ell^{2H-1}I_R(\omega_\ell). \tag{3.6.34}$$

Then, simple calculations give

$$\frac{d^2L(H^*)}{dH^2} = 4\cdot\frac{\widehat{G}_2(H^*)\widehat{G}_0(H^*) - \widehat{G}_1^2(H^*)}{\widehat{G}_0^2(H^*)}, \tag{3.6.35}$$

$$\sqrt{m}\frac{dL(H_0)}{dH} = \frac{1}{\widehat{G}_0(H_0)}\frac{2}{\sqrt{m}}\sum_{\ell=1}^{m}\nu_{\ell,m}\omega_\ell^{2H_0-1}I_R(\omega_\ell), \tag{3.6.36}$$

where $\nu_{\ell,m}$ is defined in (3.3.15).

For $k \in \mathbb{Z}^+$, define

$$\widehat{E}_k(H) = \frac{1}{m}\sum_{\ell=1}^{m}(\log \ell)^k \ell^{2H-1} I_R(\omega_\ell), \tag{3.6.37}$$

$$\widehat{F}_k(H) = \frac{1}{m}\sum_{\ell=1}^{m}(\log \ell)^k \ell^{2H-1} I_1^{(Y)}(\omega_\ell). \tag{3.6.38}$$

Note that

$$\begin{aligned}
\widehat{G}_0(H) &= \left(\frac{2\pi}{N}\right)^{2H-1} \widehat{E}_0(H), \\
\widehat{G}_1(H) &= \left(\frac{2\pi}{N}\right)^{2H-1} \widehat{E}_1(H) + \left(\frac{2\pi}{N}\right)^{2H-1} \log\left(\frac{2\pi}{N}\right) \cdot \widehat{E}_0(H), \\
\widehat{G}_2(H) &= \left(\frac{2\pi}{N}\right)^{2H-1} \widehat{E}_2(H) + 2\left(\frac{2\pi}{N}\right)^{2H-1} \log\left(\frac{2\pi}{N}\right) \cdot \widehat{E}_1(H) \\
&\quad + \left(\frac{2\pi}{N}\right)^{2H-1} \log^2\left(\frac{2\pi}{N}\right) \cdot \widehat{E}_0(H).
\end{aligned}$$

Thus,

$$\frac{d^2 L(H^*)}{dH^2} = 4 \cdot \frac{\widehat{E}_2(H^*)\widehat{E}_0(H^*) - \widehat{E}_1^2(H^*)}{\widehat{E}_0^2(H^*)}. \tag{3.6.39}$$

By Lemma 3.6.5 and Assumption 3.3.14, for $k = 0, 1, 2$,

$$\begin{aligned}
\widehat{E}_k(H_0) &= \widehat{F}_k(H_0) + \frac{1}{m}\sum_{\ell=1}^{m}(\log \ell)^k O_P\left(\frac{1}{\ell} + \frac{1}{N^\alpha}\right) \\
&= \widehat{F}_k(H_0) + O_P\left(\frac{(\log m)^3}{m} + \frac{(\log m)^2}{N^\alpha}\right) \\
&= \widehat{F}_k(H_0) + o_P(1).
\end{aligned}$$

Consequently, by Lemma 3.6.10 we conclude that $\widehat{E}_k(H^*) = \widehat{F}_k(H_0) + o_P(1)$, for $k = 0, 1, 2$. Therefore, by (3.6.39)

$$\begin{aligned}\frac{d^2L(H^*)}{dH^2} &= 4 \cdot \frac{\left[\widehat{F}_2(H_0) + o_P(1)\right]\left[\widehat{F}_0(H_0) + o_P(1)\right] - \left[\widehat{F}_1(H_0) + o_P(1)\right]^2}{\left[\widehat{F}_0(H_0) + o_P(1)\right]^2} \\ &= 4 \cdot \frac{\widehat{F}_2(H_0)\widehat{F}_0(H_0) - \widehat{F}_1^2(H_0)}{\widehat{F}_0^2(H_0)} + o_P(1). \end{aligned} \tag{3.6.40}$$

In the remainder of the argument we use a constant $G_0 = f_1(0)$, where f_1 is the spectral density of the $\xi_{1t}^{(Y)}$, cf. Lemmas 3.6.6 and 3.6.7. By (3.6.40) and Lemmas 3.6.11 and 3.6.13,

$$\begin{aligned}\frac{d^2L(H^*)}{dH^2} &= 4 \cdot \frac{G_0^2\left\{\frac{1}{m}\sum_{\ell=1}^m (\log \ell)^2 - \left(\frac{1}{m}\sum_{\ell=1}^m \log \ell\right)^2\right\} + O_P\left(\frac{(\log m)^4}{m^{1/2}}\right)}{G_0^2 + O_P\left(\frac{(\log m)^2}{m^{1/2}}\right)} + o_P(1) \\ &= 4 \cdot \frac{G_0^2(1 + o(1)) + o_P(1)}{G_0^2 + o_P(1)} + o_P(1) \\ &= 4 + o_P(1).\end{aligned}$$

We have completed the verification of (3.6.32). We now prove (3.6.33). By Lemmas 3.6.5, 3.6.9 and 3.6.14,

$$\begin{aligned}\sqrt{m}\frac{dL(H_0)}{dH} &= \frac{1}{G_0 + o_P(1)} \cdot \frac{2}{\sqrt{m}} \sum_{\ell=1}^m \nu_{\ell,m}\left\{I_1^{(Y)}(\omega_\ell) + O_P\left(\frac{1}{\ell} + \frac{1}{N^\alpha}\right)\right\} \\ &= \frac{1}{G_0}\left\{\frac{2}{\sqrt{m}}\sum_{\ell=1}^m \nu_{\ell,m} I_1^{(Y)}(\omega_\ell) + o_P(1)\right\}(1 + o_P(1)) \\ &= \left\{\frac{2}{\sqrt{m}}\sum_{\ell=1}^m \nu_{\ell,m}\left(\frac{I_1^{(Y)}(\omega_\ell)}{G_0} - 1\right) + o_P(1)\right\}(1 + o_P(1)),\end{aligned}$$

where the last equality comes from the fact that $\sum_{\ell=1}^m \nu_{\ell,m} = 0$. We complete the proof by applying Lemma 3.6.7 to the above equation.

∎

3.6.5 Lemmas used in the proof of Theorem 3.3.4

A key element in proving Theorem 3.3.4 is to decompose the periodogram $I_R(\omega_\ell)$ of the residuals R_t into terms based on the process $\{Y_t\}$ and a term involving the R_t. We first derive the required decomposition, which relies only on algebraic manipulations, and then, in Lemma 3.6.4 specify the rates of all terms in the decomposition.

As in Section 3.6.1, we assume that $\hat{n} \le n^\star$ to avoid slightly different formulas in the case of $\hat{n} > n^\star$ that can be handled by the same arguments. If $\hat{n} \le n^\star$, then

$$R_t(u) = \begin{cases} \mu(u) + Y_t(u) - \overline{X}_{1,t}(u), & t \le \hat{n}, \\ \mu(u) + Y_t(u) - \overline{X}_{2,t}(u), & \hat{n} < t \le n^\star, \\ \mu(u) + \delta(u) + Y_t(u) - \overline{X}_{2,t}(u), & n^\star < t \le N, \end{cases} \tag{3.6.41}$$

where

$$\overline{X}_{1,t}(u) = \frac{1}{\hat{n}} \sum_{1 \le s \le \hat{n}} X_s(u), \quad \overline{X}_{2,t}(u) = \frac{1}{N - \hat{n}} \sum_{\hat{n} < s \le N} X_s(u).$$

We decompose level 1 scores as

$$\widehat{\xi}_{1t}^{(R)} := \langle R_t, \widehat{\psi}_1^{(R)} \rangle = \langle R_t, \psi_1^{(Y)} \rangle + \langle R_t, \widehat{\psi}_1^{(R)} - \psi_1^{(Y)} \rangle.$$

Thus, the periodogram based on the $\widehat{\xi}_{1t}^{(R)}$ is given by

$$\begin{aligned} I_R(\omega_\ell) &:= \frac{1}{2\pi N} \left| \sum_{t=1}^{N} \widehat{\xi}_{1t}^{(R)} e^{-it\omega_\ell} \right|^2 \qquad (3.6.42) \\ &= \frac{1}{2\pi N} \left| \sum_{t=1}^{N} \langle R_t, \psi_1^{(Y)} \rangle e^{-it\omega_\ell} + \sum_{t=1}^{N} \langle R_t, \widehat{\psi}_1^{(R)} - \psi_1^{(Y)} \rangle e^{-it\omega_\ell} \right|^2. \end{aligned}$$

Using $\sum_{t=1}^{N} e^{-it\omega_\ell} = 0$ if $\omega_\ell \neq 0$ and plugging in (3.6.41) further gives

$$I_R(\omega_\ell) = \frac{1}{2\pi N} \left| \sum_{t=1}^{N} \langle Y_t, \psi_1^{(Y)} \rangle e^{-it\omega_\ell} + \langle \overline{X}_{2,t} - \overline{X}_{1,t} - \delta, \psi_1^{(Y)} \rangle \sum_{t=1}^{\hat{n}} e^{-it\omega_\ell} \right.$$
$$\left. - \sum_{t=\hat{n}+1}^{n^\star} \langle \delta, \psi_1^{(Y)} \rangle e^{-it\omega_\ell} + \sum_{t=1}^{N} \langle R_t, \widehat{\psi}_1^{(R)} - \psi_1^{(Y)} \rangle e^{-it\omega_\ell} \right|^2.$$

Note that

$$\overline{X}_{2,t}(u) - \overline{X}_{1,t}(u) - \delta(u) = \frac{\hat{n} - n^\star}{N - \hat{n}} \delta(u) + \frac{1}{N - \hat{n}} \sum_{t=1}^{N} Y_t(u) - \frac{N}{(N - \hat{n})\hat{n}} \sum_{t=1}^{\hat{n}} Y_t(u)$$

implies

$$I_R(\omega_\ell) = |x_1 + x_2 + x_3 + x_4 + y_1|^2, \tag{3.6.43}$$

where

$$\begin{aligned}
x_1 &= \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^{N} \xi_{1t}^{(Y)} e^{-it\omega_\ell}, \\
x_2 &= \frac{1}{\sqrt{2\pi N}} \left(\frac{1}{N - \hat{n}} \sum_{t=1}^{N} \xi_{1t}^{(Y)} - \frac{N}{(N - \hat{n})\hat{n}} \sum_{t=1}^{\hat{n}} \xi_{1t}^{(Y)} \right) \sum_{t=1}^{\hat{n}} e^{-it\omega_\ell}, \\
x_3 &= \frac{1}{\sqrt{2\pi N}} \frac{\hat{n} - n^\star}{N - \hat{n}} \langle \delta, \psi_1^{(Y)} \rangle \sum_{t=1}^{\hat{n}} e^{-it\omega_\ell}, \\
x_4 &= \frac{-1}{\sqrt{2\pi N}} \langle \delta, \psi_1^{(Y)} \rangle \sum_{t=\hat{n}+1}^{n^\star} e^{-it\omega_\ell}, \\
y_1 &= \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^{N} \langle R_t, \widehat{\psi}_1^{(R)} - \psi_1^{(Y)} \rangle e^{-it\omega_\ell}.
\end{aligned} \tag{3.6.44}$$

Observe that

$$|x_1|^2 = \frac{1}{2\pi N} \left| \sum_{t=1}^{N} \xi_{1t}^{(Y)} e^{-it\omega_\ell} \right|^2 =: I_1^{(Y)}(\omega_\ell) \tag{3.6.45}$$

is the periodogram of the level 1 scores of the Y_t. Our first lemma specifies the rates of decay for the remaining terms.

LEMMA 3.6.4 *Under the assumptions of Theorem 3.3.4,*

$$x_1 = O_P(1), \quad x_2 = O_P\left(\frac{1}{\ell}\right), \quad x_3 = O_P\left(\frac{1}{\ell\sqrt{N}}\right), \quad x_4 = O_P\left(\frac{1}{\sqrt{N}}\right), \tag{3.6.46}$$

and

$$y_1 = O_P(N^{-\alpha}), \tag{3.6.47}$$

where $\alpha > 0$ *is as in Theorem 3.3.1.*

PROOF: By Theorem 1.1 in Berkes *et al.* (2013),

$$N^{-1/2} \sum_{1 \le t \le \lfloor Nx \rfloor} Y_t(u) \xrightarrow{d} \Gamma(x, u), \quad \text{in } D([0,1], L^2),$$

where $\{\Gamma(x, \cdot), x \in [0,1]\}$ is a Gaussian process in L^2. Thus, using the continuous mapping theorem, we get

$$N^{-1/2} \sum_{1 \le t \le \lfloor Nx \rfloor} \xi_{1t}^{(Y)} = \left\langle N^{-1/2} \sum_{1 \le t \le \lfloor Nx \rfloor} Y_t, \psi_1^{(Y)} \right\rangle \xrightarrow{d} \left\langle \Gamma(x, \cdot), \psi_1^{(Y)} \right\rangle, \tag{3.6.48}$$

where the convergence is in $D([0,1])$; the limit $\langle \Gamma(x, \cdot), \psi_1^{(Y)} \rangle$ is a multiple of the Wiener process. Using (3.6.48) and exactly the same arguments as in Baek and Pipiras (2012) (cf. p. 141–142), we can get the convergence rates for x_2, x_3 and x_4.

The new term that arises in the functional context is y_1 given by (3.6.44). Note that

$$\begin{aligned} |y_1| &= \frac{1}{\sqrt{2\pi N}} \left| \left\langle \sum_{t=1}^{N} R_t e^{-it\omega_\ell}, \widehat{\psi}_1^{(R)} - \psi_1^{(Y)} \right\rangle \right| \\ &\le \frac{1}{\sqrt{2\pi N}} \left\| \sum_{t=1}^{N} R_t e^{-it\omega_\ell} \right\| \left\| \widehat{\psi}_1^{(R)} - \psi_1^{(Y)} \right\|. \end{aligned}$$

By (3.3.10), $\|\widehat{\psi}_1^{(R)} - \psi_1^{(Y)}\| = O_P(N^{-\alpha})$, so it suffices to show that

$$\left\| \sum_{t=1}^{N} R_t e^{-it\omega_\ell} \right\| = O_P(N^{1/2}). \tag{3.6.49}$$

Put $\overline{Y}_1 = \frac{1}{\hat{n}} \sum_{s=1}^{\hat{n}} Y_s$ and $\overline{Y}_2 = \frac{1}{N-\hat{n}} \sum_{s=\hat{n}+1}^{N} Y_s$. A simple calculation gives

$$\begin{aligned} \sum_{t=1}^{N} R_t e^{-it\omega_\ell} = \sum_{t=1}^{N} Y_t e^{-it\omega_\ell} - \overline{Y}_1 \sum_{t=1}^{\hat{n}} e^{-it\omega_\ell} - \overline{Y}_2 \sum_{t=\hat{n}+1}^{N} e^{-it\omega_\ell} \\ - \frac{N-n^\star}{N-\hat{n}} \delta \sum_{t=\hat{n}+1}^{n^\star} e^{-it\omega_\ell} + \frac{n^\star - \hat{n}}{N-\hat{n}} \delta \sum_{t=n^\star+1}^{N} e^{-it\omega_\ell}. \end{aligned}$$

Using Theorem 3.6.1, Assumption 3.3.5 and $|e^{-it\omega_\ell}| = 1$, we obtain that

$$\left\| \overline{Y}_1 \sum_{t=1}^{\hat{n}} e^{-it\omega_\ell} \right\| = O_P(\hat{n}^{1/2}) = O_P(N^{1/2}).$$

Similarly, $\left\| \overline{Y}_2 \sum_{t=\hat{n}+1}^{N} e^{-it\omega_\ell} \right\| = O_P((N-\hat{n})^{1/2}) = O_P(N^{1/2})$. By Assumption 3.3.5,

$$\left\| \frac{N-n^\star}{N-\hat{n}} \delta \sum_{t=\hat{n}+1}^{n^\star} e^{-it\omega_\ell} \right\| = O_P(1)\|\delta\| \cdot O(|n^\star - \hat{n}|) = O_P(1).$$

Similarly, $\left\| \frac{n^\star - \hat{n}}{N-\hat{n}} \delta \sum_{t=n^\star+1}^{N} e^{-it\omega_\ell} \right\| = O_P(1)$. It remains to show that

$$\left\| \sum_{t=1}^{N} Y_t e^{-it\omega_\ell} \right\| = O_P(N^{1/2}).$$

By Fubini's theorem and stationarity of $\{Y_t\}$,

$$
\begin{aligned}
E\left\{\frac{1}{N}\left\|\sum_{t=1}^{N} Y_t e^{-it\omega_\ell}\right\|^2\right\} &= \frac{1}{N}\int \sum_{t,s=1}^{N} e^{-i(t-s)\omega_\ell} E[Y_t(u)Y_s(u)]du \\
&= \frac{1}{N}\int \sum_{|h|\le N} e^{-ih\omega_\ell}(N-|h|)E[Y_0(u)Y_h(u)]du \\
&\le \sum_{h=-\infty}^{\infty}\left|\int E[Y_0(u)Y_h(u)]du\right| \\
&= \int EY_0^2(u)du + 2\sum_{h=1}^{\infty}\left|\int E[Y_0(u)Y_h(u)]du\right|.
\end{aligned}
$$

Since Y_0 is independent of $Y_{h,h}$, cf. Assumption 3.3.1,

$$
\begin{aligned}
\sum_{h=1}^{\infty}\left|\int E[Y_0(u)Y_h(u)]du\right| &= \sum_{h=1}^{\infty}\left|\int E[Y_0(u)(Y_h(u)-Y_{h,h}(u)]du\right| \\
&\le \sum_{h=1}^{\infty} E\left|\langle Y_0, Y_h - Y_{h,h}\rangle\right| \\
&\le \sum_{h=1}^{\infty} E\left\|Y_0\right\|\left\|Y_h - Y_{h,h}\right\| \\
&\le \left(E\|Y_0\|^2\right)^{1/2}\sum_{h=1}^{\infty}\left(E\left\|Y_h - Y_{h,h}\right\|^2\right)^{1/2} \\
&< \infty,
\end{aligned}
$$

where the last line follows from equation (3.3.2). Thus,

$$
\lim_{M\to\infty}\limsup_{N\to\infty} P\left(\frac{1}{N}\left\|\sum_{t=1}^{N} Y_t e^{-it\omega_\ell}\right\|^2 > M\right) = 0.
$$

This completes the verification of (3.6.49) and so (3.6.47). Finally,

$$
|x_1|^2 = \frac{1}{2\pi N}\left|\left\langle \sum_{t=1}^{N} Y_t e^{-it\omega_\ell}, \psi_1^{(Y)}\right\rangle\right|^2 \le \frac{1}{2\pi N}\left\|\sum_{t=1}^{N} Y_t e^{-it\omega_\ell}\right\|^2 = O_P(1).
$$

∎

LEMMA 3.6.5 *Under the assumptions of Theorem 3.3.4,*

$$I_R(\omega_\ell) = I_1^{(Y)}(\omega_\ell) + O_P\left(\frac{1}{\ell} + \frac{1}{N^\alpha}\right),$$

where $I_R(\omega_\ell)$ *and* $I_1^{(Y)}(\omega_\ell)$ *are defined in* (3.6.42) *and* (3.6.45)*, respectively.*

PROOF: By (3.6.43) and Lemma 3.6.4,

$$\begin{aligned} I_R(\omega_\ell) &= \left| x_1 + O_P\left(\frac{1}{\ell} + \frac{1}{N^\alpha}\right)\right|^2 \\ &= |x_1|^2 + O_P\left(\frac{1}{\ell} + \frac{1}{N^\alpha}\right) \\ &= I_1^{(Y)}(\omega_\ell) + O_P\left(\frac{1}{\ell} + \frac{1}{N^\alpha}\right). \end{aligned}$$

■

The following two lemmas are special cases of Theorems 3.3.2 and 3.3.3, respectively.

LEMMA 3.6.6 *Under the assumptions of Theorem 3.3.2,*

$$\frac{1}{m}\sum_{\ell=1}^{m}\left(\frac{I_1^{(Y)}(\omega_\ell)}{G_0} - 1\right) = O_P\left(m^{-1/2}\right), \quad \textit{as } N \to \infty.$$

LEMMA 3.6.7 *Consider the* $\nu_{\ell,m}$ *defined by* (3.3.15)*. Under the assumptions of Theorem 3.3.3,*

$$\frac{1}{\sqrt{m}}\sum_{\ell=1}^{m}\nu_{\ell,m}\left(\frac{I_1^{(Y)}(\omega_\ell)}{G_0} - 1\right) \xrightarrow{d} \mathcal{N}(0,1), \quad \textit{as } N \to \infty.$$

Let

$$\Theta = \{H : \Delta_1 < H < \Delta_2\}, \quad 0 < \Delta_1 < H_0 < \Delta_2 < 1 \tag{3.6.50}$$

be the interval over which the optimization is conducted to find the LWE. Define the following quantities:

$$T(H,H_0)=\log\frac{\widehat{G}(H_0)}{G(H_0)}-\log\frac{\widehat{G}(H)}{G(H)}-\log\left\{m^{-1}\sum_{\ell=1}^{m}\ell^{2(H-H_0)}\Big/\frac{m^{2(H-H_0)}}{2(H-H_0)+1}\right\} \quad (3.6.51)$$
$$+2(H-H_0)\left\{\frac{1}{m}\sum_{\ell=1}^{m}\log\ell-(\log m-1)\right\},$$
$$U(H,H_0)=2(H-H_0)-\log(2(H-H_0)+1), \quad (3.6.52)$$

where $G(H)=\frac{G_0}{m}\sum_{\ell=1}^{m}\omega_\ell^{2(H-H_0)}$ and $\widehat{G}(H)=\frac{1}{m}\sum_{\ell=1}^{m}\omega_\ell^{2H-1}I_R(\omega_\ell)$.

LEMMA 3.6.8 *Under the assumptions of Theorem 3.3.4,*

$$\sup_{H\in\Theta}|T(H,H_0)|=O_P\left(m^{-1/2}+\frac{\log m}{m^{2\Delta_1}}\right).$$

PROOF: Observe that

$$\sup_{\Theta}|T(H,H_0)|\le\sup_{\Theta}\left|\log\frac{\widehat{G}(H_0)}{G(H_0)}\right|+\sup_{\Theta}\left|\log\frac{\widehat{G}(H)}{G(H)}\right| \quad (3.6.53)$$
$$+\sup_{\Theta}\left|\log\left\{\frac{2(H-H_0)+1}{m}\sum_{\ell=1}^{m}\left(\frac{\ell}{m}\right)^{2(H-H_0)}\right\}\right|$$
$$+2\left|\frac{1}{m}\sum_{\ell=1}^{m}\log\ell-(\log m-1)\right|.$$

By Taylor expansion and Lemma 1 in Robinson (1995),

$$\sup_{\Theta}\left|\log\left\{\frac{2(H-H_0)+1}{m}\sum_{\ell=1}^{m}\left(\frac{\ell}{m}\right)^{2(H-H_0)}\right\}\right|=O\left(m^{-2(\Delta_1-H_0)-1}\right).$$

Since $H_0 = 1/2$, the above term is $O(m^{-2\Delta_1})$, and so is asymptotically negligible. By Lemma 2 in Robinson (1995),

$$\left| \frac{1}{m} \sum_{\ell=1}^{m} \log \ell - (\log m - 1) \right| = O\left(\frac{\log m}{m} \right).$$

Since $\Delta_1 < 1/2$, the above term is also asymptotically negligible. Hence, to complete the proof it remains to verify that

$$\sup_{\Theta} \left| \frac{\widehat{G}(H) - G(H)}{G(H)} \right| = O_P\left(m^{-1/2} + \frac{\log m}{m^{2\Delta_1}} \right). \tag{3.6.54}$$

Let

$$\frac{\widehat{G}(H) - G(H)}{G(H)} = \frac{A(H, H_0)}{B(H, H_0)},$$

where

$$A(H, H_0) = \frac{2(H - H_0) + 1}{m} \sum_{\ell=1}^{m} \left(\frac{\ell}{m} \right)^{2(H-H_0)} \left(\frac{I_R(\omega_\ell)}{G_0 \omega_\ell^{1-2H_0}} - 1 \right), \tag{3.6.55}$$

$$B(H, H_0) = \frac{2(H - H_0) + 1}{m} \sum_{\ell=1}^{m} \left(\frac{\ell}{m} \right)^{2(H-H_0)}.$$

Note that

$$\begin{aligned} \inf_{\Theta} B(H, H_0) &\geq \inf_{\Theta} \{1 - |B(H, H_0) - 1|\} \\ &= 1 - \sup_{\Theta} |B(H, H_0) - 1| \\ &= 1 + O\left(m^{-2(\Delta_1 - H_0) - 1}\right) \geq \frac{1}{2}, \end{aligned}$$

for all $m > m_0$, where m_0 is a positive integer. It remains to show

$$\sup_{\Theta} |A(H, H_0)| = O_P\left(m^{-1/2} + \frac{\log m}{m^{2\Delta_1}} \right).$$

We bound $|A(H, H_0)|$ by two terms:

$$
\begin{aligned}
|A(H,H_0)| &\le \frac{3}{m}\left|\sum_{\ell=1}^{m}\left(\frac{\ell}{m}\right)^{2(H-H_0)}\left(\frac{I_R(\omega_\ell)}{G_0\omega_\ell^{1-2H_0}}-1\right)\right| \\
&= \frac{3}{m}\Bigg|\sum_{\ell=1}^{m-1}\left(\frac{\ell}{m}\right)^{2(H-H_0)}\left(\frac{I_R(\omega_\ell)}{G_0\omega_\ell^{1-2H_0}}-1\right) \\
&\quad + \sum_{j=1}^{m}\left(\frac{I_R(\omega_j)}{G_0\omega_j^{1-2H_0}}-1\right) - \sum_{j=1}^{m-1}\left(\frac{I_R(\omega_j)}{G_0\omega_j^{1-2H_0}}-1\right)\Bigg| \\
&\le \frac{3}{m}\Bigg|\left(\frac{1}{m}\right)^{2(H-H_0)}\left(\frac{I_R(\omega_1)}{G_0\omega_1^{1-2H_0}}-1\right) \\
&\quad + \sum_{\ell=2}^{m-1}\left[\left(\frac{\ell}{m}\right)^{2(H-H_0)}\sum_{j=1}^{\ell}\left(\frac{I_R(\omega_j)}{G_0\omega_j^{1-2H_0}}-1\right)\right] \\
&\quad - \sum_{\ell=2}^{m-1}\left[\left(\frac{\ell}{m}\right)^{2(H-H_0)}\sum_{j=1}^{\ell-1}\left(\frac{I_R(\omega_j)}{G_0\omega_j^{1-2H_0}}-1\right)\right] - \sum_{j=1}^{m-1}\left(\frac{I_R(\omega_j)}{G_0\omega_j^{1-2H_0}}-1\right)\Bigg| \\
&\quad + \frac{3}{m}\left|\sum_{\ell=1}^{m}\left(\frac{I_R(\omega_\ell)}{G_0\omega_\ell^{1-2H_0}}-1\right)\right| \\
&= A_1(H,H_0) + A_2(H,H_0),
\end{aligned}
$$

where

$$
\begin{aligned}
A_1(H,H_0) &= \frac{3}{m}\left|\sum_{\ell=1}^{m-1}\left[\left(\frac{\ell}{m}\right)^{2(H-H_0)} - \left(\frac{\ell+1}{m}\right)^{2(H-H_0)}\right]\sum_{j=1}^{\ell}\left(\frac{I_R(\omega_j)}{G_0\omega_j^{1-2H_0}}-1\right)\right|, \\
A_2(H,H_0) &= \frac{3}{m}\left|\sum_{\ell=1}^{m}\left(\frac{I_R(\omega_\ell)}{G_0\omega_\ell^{1-2H_0}}-1\right)\right|.
\end{aligned}
$$

Using the elementary bound $\left|\left(1+\frac{1}{\ell}\right)^{2(H-H_0)}-1\right| \leq \frac{2}{\ell}$ for $H \in \Theta$ and $\ell > 0$, we obtain

$$\begin{aligned}
A_1(H,H_0) &= 3\left|\sum_{\ell=1}^{m-1}\left\{\left(\frac{\ell}{m}\right)^{2(H-H_0)+1}\cdot\frac{1}{\ell}\left[1-\left(1+\frac{1}{\ell}\right)^{2(H-H_0)}\right]\sum_{j=1}^{\ell}\left(\frac{I_R(\omega_j)}{G_0\omega_j^{1-2H_0}}-1\right)\right\}\right| \\
&\leq 6\sup_{H\in\Theta}\sum_{\ell=1}^{m-1}\left\{\left(\frac{\ell}{m}\right)^{2(H-H_0)+1}\cdot\frac{1}{\ell^2}\left|\sum_{j=1}^{\ell}\left(\frac{I_R(\omega_j)}{G_0\omega_j^{1-2H_0}}-1\right)\right|\right\} \\
&\leq 6\sum_{\ell=1}^{m-1}\left\{\left(\frac{\ell}{m}\right)^{2(\Delta_1-H_0)+1}\cdot\frac{1}{\ell^2}\left|\sum_{j=1}^{\ell}\left(\frac{I_R(\omega_j)}{G_0\omega_j^{1-2H_0}}-1\right)\right|\right\}.
\end{aligned}$$

By Lemma 3.6.5,

$$\frac{I_R(\omega_\ell)}{G_0}-1=\frac{I_1^{(Y)}(\omega_\ell)}{G_0}-1+O_P\left(\frac{1}{\ell}+\frac{1}{N^\alpha}\right),$$

so, since $\alpha < 1/2$,

$$\begin{aligned}
&\sum_{\ell=1}^{m-1}\left\{\left(\frac{\ell}{m}\right)^{2(\Delta_1-H_0)+1}\cdot\frac{1}{\ell^2}\left|\sum_{j=1}^{\ell}\left(\frac{I_R(\omega_j)}{G_0}-1\right)\right|\right\} \\
&\leq \sum_{\ell=1}^{m}\left\{\left(\frac{\ell}{m}\right)^{2(\Delta_1-H_0)+1}\cdot\frac{1}{\ell^2}\left|\sum_{j=1}^{\ell}\left(\frac{I_R(\omega_j)}{G_0}-1\right)\right|\right\} \\
&= \sum_{\ell=1}^{m}\left\{\left(\frac{\ell}{m}\right)^{2(\Delta_1-H_0)+1}\cdot\frac{1}{\ell^2}\left|\sum_{j=1}^{\ell}\left(\frac{I_1^{(Y)}(\omega_j)}{G_0}-1\right)+\sum_{j=1}^{\ell}O_P\left(\frac{1}{j}+\frac{1}{N^\alpha}\right)\right|\right\} \\
&= \sum_{\ell=1}^{m}\left\{\left(\frac{\ell}{m}\right)^{2(\Delta_1-H_0)+1}\cdot\frac{1}{\ell}\cdot O_P\left(\frac{1}{\ell^{1/2}}+\frac{\log\ell}{\ell}+\frac{1}{N^\alpha}\right)\right\} \\
&= \frac{1}{m^{2(\Delta_1-H_0)+1}}\sum_{\ell=1}^{m}O_P\left(\ell^{2(\Delta_1-H_0)-\frac{1}{2}}+\frac{\ell^{2(\Delta_1-H_0)}}{N^\alpha}\right),
\end{aligned}$$

where we used Lemma 3.6.6 in the penultimate line. Observe that $2(\Delta_1-H_0)-\frac{1}{2}\in\left(-\frac{3}{2},-\frac{1}{2}\right)$. If $2(\Delta_1-H_0)-\frac{1}{2}\neq -1$,

$$\sum_{\ell=1}^{m}\ell^{2(\Delta_1-H_0)-\frac{1}{2}}\sim\int_1^m x^{2(\Delta_1-H_0)-\frac{1}{2}}dx=O\left(m^{2(\Delta_1-H_0)+\frac{1}{2}}+1\right).$$

If $2(\Delta_1 - H_0) - \frac{1}{2} = -1$,

$$\sum_{\ell=1}^{m} \ell^{2(\Delta_1 - H_0) - \frac{1}{2}} \sim \int_1^m \frac{1}{x} dx = O(\log m).$$

Thus, by Assumption 3.3.14

$$\begin{aligned}
&\sum_{\ell=1}^{m-1} \left\{ \left(\frac{\ell}{m} \right)^{2(\Delta_1 - H_0)+1} \cdot \frac{1}{\ell^2} \left| \sum_{j=1}^{\ell} \left(\frac{I_R(\omega_j)}{G_0} - 1 \right) \right| \right\} \\
&= \frac{1}{m^{2(\Delta_1 - H_0)+1}} O_P \left(m^{2(\Delta_1 - H_0)+\frac{1}{2}} + \log m + \frac{m^{2(\Delta_1 - H_0)+1}}{N^\alpha} \right) \\
&= O_P \left(\frac{1}{m^{1/2}} + \frac{\log m}{m^{2(\Delta_1 - H_0)+1}} + \frac{1}{N^\alpha} \right) \\
&= O_P \left(\frac{1}{m^{1/2}} + \frac{\log m}{m^{2(\Delta_1 - H_0)+1}} \right),
\end{aligned}$$

showing that

$$A_1(H, H_0) = O_P \left(\frac{1}{m^{1/2}} + \frac{\log m}{m^{2\Delta_1}} \right).$$

A somewhat shorter argument shows that $A_2(H, H_0) = O_P(m^{-1/2})$.

∎

LEMMA 3.6.9 *Under the assumptions of Theorem 3.3.4,*

$$\widehat{G}_0(H_0) \xrightarrow{P} G_0,$$

where $\widehat{G}_0(H)$ *is given in* (3.6.34)

PROOF: Note that $\widehat{G}_0(H) = \widehat{G}(H)$. The claim follows directly from (3.6.54).

∎

LEMMA 3.6.10 *Under the assumptions of Theorem 3.3.4, we have*

$$\widehat{E}_k(H^*) - \widehat{E}_k(H_0) = o_P(1), \quad k = 0, 1, 2,$$

where $\widehat{E}_k(H)$ *is given in* (3.6.37).

PROOF: Let $\epsilon > 0$ be given. Choose N such that $2\epsilon < (\log m)^2$. Define the set $M_\epsilon = \left\{H : |H - H_0| \leq \frac{\epsilon}{(\log m)^3}\right\}$. Observe that we have the following inequalities on M_ϵ:

$$|\ell^{2(H-H_0)} - 1| \leq 2|H - H_0|(\log \ell)m^{2|H-H_0|} \leq 2|H - H_0|(\log \ell)m^{1/\log m}$$

for $\ell \in \{1, 2, \ldots, m\}$. Thus,

$$\begin{aligned}
\left|\widehat{E}_k(H) - \widehat{E}_k(H_0)\right| &= \frac{1}{m}\sum_{\ell=1}^{m}(\log \ell)^k I_R(\omega_\ell)\ell^{2H_0-1}\left|\ell^{2H_0-1} - 1\right| \\
&\leq \frac{1}{m}\sum_{\ell=1}^{m}(\log \ell)^{k+1} I_R(\omega_\ell)\ell^{2H_0-1}2|H - H_0|m^{1/\log m} \\
&= 2e|H - H_0|\widehat{E}_{k+1}(H_0) \\
&\leq 2e\epsilon(\log m)^{-3}\widehat{E}_{k+1}(H_0) \\
&\leq 2e\epsilon(\log m)^{k-2}\widehat{E}_0(H_0)
\end{aligned}$$

if $H \in M_\epsilon$. For an arbitrary $\epsilon^* > 0$, define two events

$$D_{\epsilon^*} = \left\{\left|\widehat{E}_k(H^*) - \widehat{E}_k(H_0)\right| > \epsilon^*\left(\frac{2\pi}{N}\right)^{1-2H_0}\right\}$$

and

$$\widetilde{D}_{\epsilon,\epsilon^*} = \left\{2e\epsilon(\log m)^{k-2}\widehat{E}_0(H_0) > \epsilon^*\left(\frac{2\pi}{N}\right)^{1-2H_0}\right\}.$$

Since $(D_{\epsilon^*} \cap \{H^* \in M_\epsilon\}) \subset \left(\widetilde{D}_{\epsilon,\epsilon^*} \cap \{H^* \in M_\epsilon\}\right)$, then for $k = 0, 1, 2$ we have

$$\begin{aligned}
P\left(D_{\epsilon^*}\right) &= P\left(D_{\epsilon^*} \cap \{H^* \in M_\epsilon\}\right) + P\left(D_{\epsilon^*} \cap \{H^* \in M_\epsilon^c\}\right) \\
&\leq P\left(\widetilde{D}_{\epsilon,\epsilon^*}\right) + P\left(\{H^* \in M_\epsilon^c\}\right) \\
&= P\left(\widehat{G}_0(H_0) > \frac{\epsilon^*}{2e\epsilon(\log m)^{k-2}}\right) + P\left(|H^* - H_0| > \frac{\epsilon}{(\log m)^3}\right) \\
&\leq P\left(\widehat{G}_0(H_0) > \frac{\epsilon^*}{e}(\log m)^{3-k}\right) + P\left(|H^* - H_0| > \frac{\epsilon}{(\log m)^3}\right),
\end{aligned}$$

where the last inequality holds if we choose $2\epsilon < \frac{1}{\log m}$.

We shall argue that $P\left(D_{\epsilon^*}\right) \to 0$ as $N \to \infty$. Observe that by Lemma 3.6.9,

$$\lim_{N\to\infty} P\left(\widehat{G}_0(H_0) > \frac{\epsilon^*}{e}(\log m)^{3-k}\right) = 0.$$

Next, define

$$\widetilde{M}_\epsilon = \left\{H : |H - H_0| \geq \frac{\epsilon}{(\log m)^3}\right\},$$

$$\Omega_\eta = \{H : |H - H_0| \geq \eta\}, \tag{3.6.56}$$

$$\Delta L(H, H_0) = L(H) - L(H_0), \tag{3.6.57}$$

where $L(H)$ is given in (3.2.7). Note that

$$
\begin{aligned}
P\left(|H^* - H_0| > \frac{\epsilon}{(\log m)^3}\right) &\le P\left(|H^* - H_0| \ge \frac{\epsilon}{(\log m)^3}\right) \\
&\le P\left(\left|\widehat{H}^{(R)} - H_0\right| \ge \frac{\epsilon}{(\log m)^3}\right) \\
&\le P\left(\inf_{\Theta \bigcap \widetilde{M}_\epsilon} \Delta L(H, H_0) \le 0\right) \\
&\le P\left(\inf_{\Theta \bigcap \widetilde{M}_\epsilon \bigcap \Omega_\eta} \Delta L(H, H_0) \le 0\right) \\
&\quad + P\left(\inf_{\Theta \bigcap \widetilde{M}_\epsilon \bigcap \Omega_\eta^c} \Delta L(H, H_0) \le 0\right) \\
&\le P\left(\inf_{\Theta \bigcap \Omega_\eta} \Delta L(H, H_0) \le 0\right) \\
&\quad + P\left(\inf_{\Theta \bigcap \widetilde{M}_\epsilon \bigcap \Omega_\eta^c} \Delta L(H, H_0) \le 0\right),
\end{aligned}
$$

where Θ is defined in Lemma 3.6.8. Using

$$
\inf_{\Theta \bigcap \widetilde{M}_\epsilon \bigcap \Omega_\eta^c} \Delta L(H, H_0) \ge \inf_{\Theta \bigcap \widetilde{M}_\epsilon \bigcap \Omega_\eta^c} U(H, H_0) - \sup_{\Theta \bigcap \Omega_\eta^c} |T(H, H_0)|
$$

and then applying Lemma 3.6.12, we obtain

$$
\begin{aligned}
P\left(\inf_{\Theta \bigcap \widetilde{M}_\epsilon \bigcap \Omega_\eta^c} \Delta L(H, H_0) \le 0\right) &\le P\left(\sup_{\Theta \bigcap \Omega_\eta^c} |T(H, H_0)| \ge \inf_{\Theta \bigcap \widetilde{M}_\epsilon \bigcap \Omega_\eta^c} U(H, H_0)\right) \\
&\le P\left(\sup_{\Theta \bigcap \Omega_\eta^c} |T(H, H_0)| \ge \inf_{\Theta \bigcap \widetilde{M}_\epsilon} U(H, H_0)\right) \\
&\le P\left(\sup_{\Theta \bigcap \Omega_\eta^c} |T(H, H_0)| \ge \frac{\epsilon^2}{2(\log m)^6}\right) \\
&\le P\left(\sup_{\Theta} |T(H, H_0)| \ge \frac{\epsilon^2}{2(\log m)^6}\right)
\end{aligned}
$$

and

$$P\left(\inf_{\Theta\bigcap\Omega_\eta} \Delta L(H,H_0) \leq 0\right) \leq P\left(\sup_{\Theta} |T(H,H_0)| \geq \inf_{\Theta\bigcap\Omega_\eta} U(H,H_0)\right)$$
$$\leq P\left(\sup_{\Theta} |T(H,H_0)| \geq \frac{1}{2}\eta^2\right).$$

Hence, $P\left(|H^* - H_0| > \frac{\epsilon}{(\log m)^3}\right) \to 0$ if we can show $\sup_\Theta |T(H,H_0)| = o_P((\log m)^{-6})$. Since $\left(m^{-1/2} + \frac{\log m}{m^{2\Delta_1}}\right)(\log m)^6 \to 0$, we complete the proof by applying Lemma 3.6.8.

∎

LEMMA 3.6.11 *Under the assumptions of Theorem 3.3.4, we have*

$$\widehat{F}_k(H_0) = G_0 \frac{1}{m} \sum_{\ell=1}^{m} (\log \ell)^k + O_P\left(\frac{(\log m)^2}{m^{1/2}}\right), \quad k = 0, 1, 2,$$

where $\widehat{F}_k(H)$ *is defined in* (3.6.38).

PROOF: Note that for $k = 0, 1, 2$,

$$
\begin{aligned}
\left|\widehat{F}_k(H_0) - G_0\frac{1}{m}\sum_{\ell=1}^{m}(\log \ell)^k\right| &= \left|\frac{1}{m}\sum_{\ell=1}^{m}(\log \ell)^k \left(I_1^{(Y)}(\omega_\ell) - G_0\right)\right| \\
&= \left|\frac{1}{m}\sum_{\ell=2}^{m}(\log \ell)^k \left[\sum_{r=1}^{\ell}\left(I_1^{(Y)}(\omega_r) - G_0\right)\right.\right. \\
&\qquad \left.\left. - \sum_{r=1}^{\ell-1}\left(I_1^{(Y)}(\omega_r) - G_0\right)\right]\right| \\
&= \left|\frac{1}{m}\sum_{\ell=2}^{m}(\log \ell)^k \sum_{r=1}^{\ell}\left(I_1^{(Y)}(\omega_r) - G_0\right)\right. \\
&\qquad \left. - \frac{1}{m}\sum_{\ell=1}^{m-1}(\log(\ell+1))^k \sum_{r=1}^{\ell}\left(I_1^{(Y)}(\omega_r) - G_0\right)\right| \\
&= \left|\frac{1}{m}\sum_{\ell=1}^{m-1}\left[(\log \ell)^k - (\log(\ell+1))^k\right]\sum_{r=1}^{\ell}\left(I_1^{(Y)}(\omega_r) - G_0\right)\right. \\
&\qquad \left. + \frac{1}{m}(\log m)^k \sum_{\ell=1}^{m}\left(I_1^{(Y)}(\omega_\ell) - G_0\right)\right| \\
&\leq \frac{G_0}{m}\sum_{\ell=1}^{m-1}\left|(\log \ell)^k - (\log(\ell+1))^k\right| \left|\sum_{r=1}^{\ell}\left(\frac{I_1^{(Y)}(\omega_r)}{G_0} - 1\right)\right| \\
&\qquad + \frac{G_0}{m}(\log m)^k \left|\sum_{\ell=1}^{m}\left(\frac{I_1^{(Y)}(\omega_\ell)}{G_0} - 1\right)\right|.
\end{aligned}
$$

For $k = 0$, by Assumption 3.3.14 and Lemma 3.6.6,

$$
\begin{aligned}
\left|\widehat{F}_k(H_0) - G_0\frac{1}{m}\sum_{\ell=1}^{m}(\log \ell)^k\right| &\leq \frac{G_0}{m}\left|\sum_{\ell=1}^{m}\left(\frac{I_1^{(Y)}(\omega_\ell)}{G_0} - 1\right)\right| \\
&= O_P\left(m^{-1/2}\right).
\end{aligned}
$$

We will use an elementary result that for $k = 1, 2$,

$$
\left|(\log \ell)^k - (\log(\ell+1))^k\right| \leq \frac{(\log(\ell+1))^{k-1}}{\ell}, \quad \ell > 0.
$$

By Assumption 3.3.14 and Lemma 3.6.6,

$$\begin{aligned}\left|\widehat{F}_k(H_0) - G_0 \frac{1}{m}\sum_{\ell=1}^{m}(\log \ell)^k\right| &\leq \frac{G_0}{m}\sum_{\ell=1}^{m-1}\frac{(\log(\ell+1))^{k-1}}{\ell}\left|\sum_{r=1}^{\ell}\left(\frac{I_1^{(Y)}(\omega_r)}{G_0}-1\right)\right| \\ &\quad + \frac{G_0}{m}(\log m)^k\left|\sum_{\ell=1}^{m}\left(\frac{I_1^{(Y)}(\omega_\ell)}{G_0}-1\right)\right| \\ &= O_P\left(\frac{\log m}{m^{1/2}}\right) + O_P\left(\frac{(\log m)^2}{m^{1/2}}\right) \\ &= O_P\left(\frac{(\log m)^2}{m^{1/2}}\right)\end{aligned}$$

for $k = 1, 2$.

∎

LEMMA 3.6.12 *Define the set* $N_r = \{H : |H - H_0| \geq r\}$, *where* $r \in \left(0, \frac{1}{2}\right)$. *Then we have*

$$\inf_{\Theta \cap N_r} U(H, H_0) > \frac{1}{2}r^2,$$

where $U(H, H_0)$ *is defined in* (3.6.52) *and* Θ *is defined in Lemma 3.6.8.*

PROOF: Define $f(x) = x - \log(1+x)$ for $x > -1$. Observe that $f(x)$ is decreasing on $(-1, 0)$ and increasing on $[0, \infty)$. Thus,

$$\begin{aligned}\inf_{\Theta \cap N_r} U(H, H_0) &\geq \inf_{N_r} U(H, H_0) \\ &= \min\left\{\inf_{H-H_0 \geq r} U(H, H_0), \inf_{H-H_0<-r} U(H, H_0)\right\} \\ &= \min\{2r - \log(1+2r), -2r - \log(1-2r)\} \\ &\geq \min\left\{\frac{2}{3}r^2, 2r^2\right\} \\ &> \frac{1}{2}r^2,\end{aligned}$$

where the penultimate inequality comes from the elementary results that $\log(1+x) \leq x - \frac{1}{6}x^2$ and $-\log(1-x) \geq x + \frac{1}{2}x^2$ for $x \in (0,1)$.

∎

LEMMA 3.6.13 *As $m \to \infty$,*

$$\frac{1}{m}\sum_{\ell=1}^{m}(\log \ell)^2 - \left(\frac{1}{m}\sum_{\ell=1}^{m}\log \ell\right)^2 \to 1.$$

PROOF: The result follows from the approximations (log means the natural integral)

$$\sum_{\ell=1}^{m}(\log \ell)^2 \sim \int_1^m \log^2 x \, dx = m\log^2 m - 2m\log m + 2m,$$

$$\sum_{\ell=1}^{m}\log \ell \sim \int_1^m \log x \, dx = m\log m - m.$$

∎

LEMMA 3.6.14 *Under Assumptions 3.3.5 and 3.3.14,*

$$\frac{2}{\sqrt{m}}\sum_{\ell=1}^{m}|\nu_{\ell,m}|\left(\frac{1}{\ell} + \frac{1}{N^\alpha}\right) = o(1), \tag{3.6.58}$$

where $\nu_{\ell,m}$ is defined in (3.3.15).

PROOF: To show (3.6.58), the difficulty is to show $m^{-1/2}\sum_{j=1}^{m}\frac{|\nu_{j,m}|}{j} = o(1)$. Note that

$$\begin{aligned}\sum_{j=1}^{m}\frac{|\nu_{j,m}|}{j} &= \sum_{j=1}^{m}\left|\frac{\log j}{j} - \frac{C_m}{j}\right| \\ &\leq \sum_{j=1}^{m}\frac{\log j}{j} + \sum_{j=1}^{m}\frac{C_m}{j},\end{aligned}$$

where $C_m = \frac{1}{m}\sum_{k=1}^m \log k$.

By the integral approximation, we know that

$$\sum_{j=1}^m \frac{\log j}{j} \sim \int_1^m \frac{\log x}{x} dx = \frac{(\log m)^2}{2} = O((\log m)^2)$$

and

$$\sum_{k=1}^m \log k \sim \int_1^m \log x dx = m(\log m - 1) + 1 = O(m \log m).$$

It follows that

$$m^{-1/2} \sum_{j=1}^m \frac{|\nu_{j,m}|}{j} = O\left(\frac{(\log m)^2}{\sqrt{m}}\right) = o(1).$$

Using $|\nu_{\ell,m}| \leq \log m$,

$$\frac{2}{\sqrt{m}} \sum_{\ell=1}^m \frac{|\nu_{\ell,m}|}{N^\alpha} = O\left(\frac{\sqrt{m}\log m}{N^\alpha}\right) = o(1),$$

again by Assumption 3.3.14. This completes the verification of (3.6.58).

■

3.7 Proofs of the results in Section 3.3.2

3.7.1 Proof of Theorem 3.3.5

PROOF OF THEOREM 3.3.5: Without loss of generality, we may assume that $\nu = 0$ in (3.2.2) and that the constant c in Assumption 3.3.11 is 1. Define the sample covarince function of $\{X_t\}$ as

$$\hat{c}(u,v) = \sum_{|t|<N} K\left(\frac{|t|}{h}\right) \widehat{\gamma}_t(u,v),$$

where

$$\widehat{\gamma}_t(u,v) = \begin{cases} \frac{1}{N}\sum_{s=t+1}^N \left(Z_s(u) - \overline{Z}_N(u)\right)\left(Z_{s-t}(v) - \overline{Z}_N(v)\right), & t \geq 0, \\ \frac{1}{N}\sum_{s=1-t}^N \left(Z_{s+t}(u) - \overline{Z}_N(u)\right)\left(Z_s(v) - \overline{Z}_N(v)\right), & t < 0, \end{cases}$$

$\overline{Z}_N = \frac{1}{N}\sum_{t=1}^{N} Z_t$ and the bandwidth $h(N)$ satisfies Assumption 3.3.12.

Let $\widehat{C}$ be the covariance operator of $\hat{c}(u,v)$. Note that

$$\|h^{1-2H_0}\cdot \widehat{C}_R - C_Z\|_{\mathcal{S}} \leq h^{1-2H_0}\cdot \|\widehat{C}_R - \widehat{C}\|_{\mathcal{S}} + \|h^{1-2H_0}\cdot \widehat{C} - C_Z\|_{\mathcal{S}}.$$

We only need to show that

$$\|\widehat{C}_R - \widehat{C}\|_{\mathcal{S}} = O_P\left(h\cdot N^{2H_0-2}\right) \tag{3.7.1}$$

and

$$\|h^{1-2H_0}\cdot \widehat{C} - C_Z\|_{\mathcal{S}} = o_P\left(1\right). \tag{3.7.2}$$

Observe that we can rewrite $\hat{c}_R$ as

$$\hat{c}_R(u,v) = \sum_{|t|<N} K\left(\frac{|t|}{h}\right)\widehat{\gamma}_{R,t}(u,v),$$

where

$$\widehat{\gamma}_{R,t}(u,v) = \begin{cases} \frac{1}{N}\sum_{s=t+1}^{N} R_s(u)R_{s-t}(v), & t\geq 0,\\ \frac{1}{N}\sum_{1-t}^{N} R_{s+t}(u)R_s(v), & t<0,\end{cases}$$

and R_t is given in (3.7.10).

It follows that for $t \geq 0$,

$$\begin{aligned}\widehat{\gamma}_{R,t}(u,v) - \widehat{\gamma}_t(u,v) = \frac{1}{N}\sum_{s=t+1}^{N}\Big[&\overline{Z}_{k,s}(u)\overline{Z}_{l,s-t}(v) - Z_s(u)\overline{Z}_{l,s-t}(v) - \overline{Z}_{k,s}(u)Z_{s-t}(v)\\ &+ Z_s(u)\overline{Z}_N(v) + \overline{Z}_N(u)Z_{s-t}(v) - \overline{Z}_N(u)\overline{Z}_N(v)\Big],\end{aligned}$$

and for $t < 0$,

$$\begin{aligned}\widehat{\gamma}_{R,t}(u,v) - \widehat{\gamma}_t(u,v) = \frac{1}{N}\sum_{s=1-t}^{N}\Big[&\overline{Z}_{o,s+t}(u)\overline{Z}_{p,s}(v) - Z_{s+t}(u)\overline{Z}_{p,s}(v) - \overline{Z}_{o,s+t}(u)Z_s(v)\\ &+ Z_{s+t}(u)\overline{Z}_N(v) + \overline{Z}_N(u)Z_s(v) - \overline{Z}_N(u)\overline{Z}_N(v)\Big],\end{aligned}$$

where $k, l, o, p \in \{1, 2\}$ and

$$\overline{Z}_{1,s} = \frac{1}{\hat{n}} \sum_{1 \le r \le \hat{n}} Z_r, \quad \overline{Z}_{2,s} = \frac{1}{N - \hat{n}} \sum_{\hat{n} < r \le N} Z_r. \tag{3.7.3}$$

By Corollary 1 in Li *et al.* (2020) and Assumptions 3.3.7(i) and 3.3.9, we can show that

$$\|\overline{Z}_N\|, \|\overline{Z}_{1,s}\|, \|\overline{Z}_{2,s}\| = O_P\left(N^{H_0 - 1}\right). \tag{3.7.4}$$

Thus,

$$\|\widehat{\gamma}_{R,t} - \widehat{\gamma}_t\|_{\mathcal{S}}^2 = O_P\left(N^{2(2H_0-2)}\right).$$

We then have

$$\begin{aligned}
\|\widehat{C}_R - \widehat{C}\|_{\mathcal{S}} &= \left(\iint \left(\hat{c}_R(u,v) - \hat{c}(u,v)\right)^2 du dv \right)^{1/2} \\
&= \left(\sum_{|t|<N} \sum_{|l|<N} K\left(\frac{|t|}{h}\right) K\left(\frac{|l|}{h}\right) \right. \\
&\quad \left. \cdot \iint \left[\widehat{\gamma}_{R,t}(u,v) - \widehat{\gamma}_t(u,v)\right] \left[\widehat{\gamma}_{R,l}(u,v) - \widehat{\gamma}_l(u,v)\right] du dv \right)^{1/2} \\
&\le \left(\sum_{|t|<N} \sum_{|l|<N} \left| K\left(\frac{|t|}{h}\right) \right| \left| K\left(\frac{|l|}{h}\right) \right| \|\widehat{\gamma}_{R,t} - \widehat{\gamma}_t\|_{\mathcal{S}} \|\widehat{\gamma}_{R,l} - \widehat{\gamma}_l\|_{\mathcal{S}} \right)^{1/2} \\
&= \left| \sum_{|t|<N} K\left(\frac{|t|}{h}\right) \right| \cdot O_P\left(N^{2H_0-2}\right) \\
&= O_P\left(h \cdot N^{2H_0-2}\right).
\end{aligned}$$

Next, we deal with (3.7.2).

Define

$$\widehat{\gamma}_{t1}(u,v) = \begin{cases} \frac{1}{N} \sum_{s=t+1}^{N} Z_s(u) Z_{s-t}(v), & t \ge 0, \\ \frac{1}{N} \sum_{s=1-t}^{N} Z_{s+t}(u) Z_s(v), & t < 0, \end{cases}$$

and

$$\widehat{\gamma}_{t2}(u,v) = \widehat{\gamma}_t(u,v) - \widehat{\gamma}_{t1}(u,v)$$
$$= \begin{cases} \frac{N-t}{N}\overline{Z}_N(u)\overline{Z}_N(v) - \overline{Z}_{t+1\to N}(u)\overline{Z}_N(v) - \overline{Z}_N(u)\overline{Z}_{1\to N-t}(v), & t \geq 0, \\ \frac{N+t}{N}\overline{Z}_N(u)\overline{Z}_N(v) - \overline{Z}_{1\to N+t}(u)\overline{Z}_N(v) - \overline{Z}_N(u)\overline{Z}_{1-t\to N}(v), & t < 0. \end{cases}$$

where $\overline{Z}_{i\to j} = \frac{1}{N}\sum_{r=i}^{j} Z_r$.

By (3.7.4), we can show that

$$\|\widehat{\gamma}_{t2}\|_{\mathcal{S}} \sim \|\overline{Z}_N\|^2 = O_P\left(N^{2H_0-2}\right). \tag{3.7.5}$$

Define

$$\hat{c}_i(u,v) = \sum_{|t|<N} K\left(\frac{|t|}{h}\right)\widehat{\gamma}_{ti}(u,v)$$

and $\widehat{C}_i$ be the corresponding operator for $i = 1, 2$. Note that

$$\|h^{1-2H_0}\cdot\widehat{C} - C_Z\|_{\mathcal{S}} \leq \|h^{1-2H_0}\cdot\widehat{C}_1 - C_Z\|_{\mathcal{S}} + h^{1-2H_0}\cdot\|\widehat{C}_2\|_{\mathcal{S}}.$$

By (3.7.5), we have

$$\begin{aligned} \|\widehat{C}_2\|_{\mathcal{S}} &= \left(\iint \hat{c}_2^2(u,v)dudv\right)^{1/2} \\ &= \left(\sum_{|t|<N}\sum_{|l|<N} K\left(\frac{|t|}{h}\right)K\left(\frac{|l|}{h}\right)\iint \widehat{\gamma}_{t2}(u,v)\widehat{\gamma}_{l2}(u,v)dudv\right)^{1/2} \\ &\leq \left(\sum_{|t|<N}\sum_{|l|<N} \left|K\left(\frac{|t|}{h}\right)\right|\left|K\left(\frac{|l|}{h}\right)\right| \|\widehat{\gamma}_{t2}\|_{\mathcal{S}}\|\widehat{\gamma}_{l2}\|_{\mathcal{S}}\right)^{1/2} \\ &= \left|\sum_{|t|<N} K\left(\frac{|t|}{h}\right)\right| O_P\left(N^{2H_0-2}\right) \\ &= O_P\left(h\cdot N^{2H_0-2}\right), \end{aligned}$$

and thus,

$$h^{1-2H_0} \cdot \|\widehat{C}_2\|_{\mathcal{S}} = O_P\left(\left(\frac{h}{N}\right)^{2-2H_0}\right) = o_P(1).$$

It remains to show $\|h^{1-2H_0} \cdot \widehat{C}_1 - C_Z\|_{\mathcal{S}} = o_P\,(1)$. Define $\widehat{C}_3$ and $\widehat{C}_4$ to be the covariance operators of

$$\hat{c}_3(u,v) = h^{1-2H_0} \cdot \sum_{|t|<N} K\left(\frac{|t|}{h}\right)\{\widehat{\gamma}_{t1}(u,v) - E\left[\widehat{\gamma}_{t1}(u,v)\right]\} \tag{3.7.6}$$

and

$$\hat{c}_4(u,v) = h^{1-2H_0} \cdot \sum_{|t|<N} K\left(\frac{|t|}{h}\right) E\left[\widehat{\gamma}_{t1}(u,v)\right] - c_Z(u,v). \tag{3.7.7}$$

Note that

$$h^{1-2H_0} \cdot \widehat{C}_1 - C_Z = \widehat{C}_3 + \widehat{C}_4.$$

By using Lemmas 3.7.1 and 3.7.2, we complete the proof.

■

LEMMA 3.7.1 *Recall that the covariance operator* $\widehat{C}_3$ *is defined in* (3.7.6)*. Under the assumptions of Theorem 3.3.5,*

$$\|\widehat{C}_3\|_{\mathcal{S}} = o_P(1).$$

PROOF: Using a similar argument as in Li *et al.* (2020) (see supplement material p. 8–9), we readily have

$$\begin{aligned}
\|\widehat{C}_3\|_{\mathcal{S}}^2 &= O_P\left(h^{1-2H_0}\right) + O_P\left(\frac{h^{4-4H_0}}{N}\right) + O_P\left(\left(\frac{h}{N}\right)^{4-4H_0}\right) + O_P\left(\frac{h\log N}{N}\right) \\
&= o_P(1),
\end{aligned}$$

where the last equality follows from Assumption 3.3.12.

■

LEMMA 3.7.2 *Recall that the covariance operator* $\widehat{C}_4$ *is defined in* (3.7.7)*. Suppose the assumptions of Theorem 3.3.5 are satisfied. Then,*

$$\|\widehat{C}_4\|_{\mathcal{S}} = o_P(1).$$

PROOF:

By Assumption 3.3.7, we have

$$\begin{aligned}
E\left[\widehat{\gamma}_{t1}(u,v)\right] &= \left(1 - \frac{|t|}{N}\right) E[Z_{t+1}(u)Z_1(v)] \\
&= E[Z_{t+1}(u)Z_1(v)] + O\left(\frac{|t|^{2H_0-1}}{N}\right).
\end{aligned}$$

Note that

$$\begin{aligned}
h^{1-2H_0} \cdot \sum_{|t|<N} \left| K\left(\frac{|t|}{h}\right)\right| O\left(\frac{|t|^{2H_0-1}}{N}\right) &= h^{1-2H_0} \cdot \sum_{|t|<h} \left| K\left(\frac{|t|}{h}\right)\right| O\left(\frac{|t|^{2H_0-1}}{N}\right) \\
&= h^{1-2H_0} \cdot \sup_{x\in[0,1]} |K(x)| O\left(\sum_{|t|<h} \frac{|t|^{2H_0-1}}{N}\right) \\
&= O\left(\frac{h}{N}\right) \\
&= o(1).
\end{aligned}$$

Since

$$c_Z(u,v) = \lim_{N\to\infty} N^{1-2H_0} \sum_{|t|<N} \left(1 - \frac{|t|}{N}\right) E[Z_{t+1}(u)Z_1(v)],$$

and $h \to \infty$ as $N \to \infty$, we have

$$h^{1-2H_0} \sum_{|t|<h} \left(1 - \frac{|t|}{h}\right) E[Z_{t+1}(u)Z_1(v)] - c_Z(u,v) = o(1).$$

Thus,

$$
\begin{aligned}
\hat{c}_4(u,v) &= h^{1-2H_0} \sum_{|t|<h} \left\{ K\left(\frac{|t|}{h}\right) - \left(1 - \frac{|t|}{h}\right) \right\} E[Z_{t+1}(u) Z_1(v)] + o(1) \\
&= \int_0^1 |K(x) - (1-x)| x^{2H_0-2} dx + o(1) \\
&= o(1),
\end{aligned}
$$

where the last equality comes from Assumption 3.3.10.

∎

3.7.2 Proof of Theorem 3.3.6

Proof of Theorem 3.3.6: We first present the main structure of the proof that uses several technical lemmas that are stated and proven after the main proof. They use notation introduced in the course of the main proof.

Recall that for any $\eta > 0$, the sets Θ and Ω_η are defined in (3.6.50) and (3.6.56), respectively. By (3.2.6),

$$
\begin{aligned}
P\left(\left|\widehat{H}^{(R)} - H_0\right| \geq \eta\right) &= P\left(\widehat{H}^{(R)} \in \Omega_\eta \cap \Theta\right) \\
&= P\left(\inf_{\Omega_\eta \bigcap \Theta} L(H) \leq \inf_{\Omega_\eta^c \bigcap \Theta} L(H)\right) \\
&\leq P\left(\inf_{\Omega_\eta \bigcap \Theta} \Delta L(H, H_0) \leq 0\right),
\end{aligned}
$$

where $\Delta L(H, H_0)$ is given in (3.6.57). Due to the non-uniform behavior of $L(H)$ around $H = H_0 - \frac{1}{2}$, we need to split Θ into two sets:

if $H_0 < \frac{1}{2} + \Delta_1$, define $\Theta_1 = \{H : \Delta \leq H \leq \Delta_2\}$ and $\Theta_2 = \emptyset$, where $\Delta = \Delta_1$;

if $H_0 \geq \frac{1}{2} + \Delta_1$, define $\Theta_1 = \{H : \Delta \leq H \leq \Delta_2\}$ and $\Theta_2 = \{H : \Delta_1 \leq H < \Delta\}$, where $H_0 - \frac{1}{2} < \Delta \leq H_0$.

Note that in either case, Θ_1 and Θ_2 are disjoint and $\Theta_1 \bigcup \Theta_2 = \Theta$. Thus,

$$\begin{aligned}
P\left(\left|\widehat{H}^{(R)} - H_0\right| \geq \eta\right) &\leq P\left(\inf_{\Omega_\eta \bigcap \Theta} \Delta L(H, H_0) \leq 0\right) \\
&\leq P\left(\inf_{\Omega_\eta \bigcap \Theta_1} \Delta L(H, H_0) \leq 0\right) + P\left(\inf_{\Omega_\eta \bigcap \Theta_2} \Delta L(H, H_0) \leq 0\right) \\
&\leq P\left(\inf_{\Omega_\eta \bigcap \Theta_1} \Delta L(H, H_0) \leq 0\right) + P\left(\inf_{\Theta_2} \Delta L(H, H_0) \leq 0\right).
\end{aligned}$$

We will argue that the two probabilities above converge to zero.

Recall that $T(H, H_0)$ and $U(H, H_0)$ are defined in (3.6.51) and (3.6.52). Since

$$\inf_{\Omega_\eta \bigcap \Theta_1} \Delta L(H, H_0) \geq \inf_{\Omega_\eta \bigcap \Theta_1} U(H, H_0) - \sup_{\Theta_1} |T(H, H_0)|,$$

$$P\left(\inf_{\Omega_\eta \bigcap \Theta_1} \Delta L(H, H_0) \leq 0\right) \leq P\left(\sup_{\Theta_1} |T(H, H_0)| \geq \inf_{\Omega_\eta \bigcap \Theta_1} U(H, H_0)\right).$$

Note that by Lemma 3.6.12, $\inf_{\Omega_\eta \bigcap \Theta_1} U(H, H_0) \geq \inf_{\Omega_\eta \bigcap \Theta} U(H, H_0) > \frac{1}{2}\eta^2$. Hence,

$$P\left(\inf_{\Omega_\eta \bigcap \Theta_1} \Delta L(H, H_0) \leq 0\right) \leq P\left(\sup_{\Theta_1} |T(H, H_0)| > \frac{1}{2}\eta^2\right).$$

Thus, by Lemma 3.7.3,

$$\lim_{N \to \infty} P\left(\inf_{\Omega_\eta \bigcap \Theta_1} \Delta L(H, H_0) \leq 0\right) = 0.$$

It remains to show that

$$\lim_{N \to \infty} P\left(\inf_{\Theta_2} \Delta L(H, H_0) \leq 0\right) = 0.$$

Define the following quantities:

$$p_m = \exp\left(\frac{1}{m} \sum_{\ell=1}^{m} \log \ell\right),$$

$$\widehat{D}(H, H_0) = \frac{1}{m}\sum_{\ell=1}^{m}\left(\frac{\ell}{p_m}\right)^{2(H-H_0)} \ell^{2H_0-1} I_R(\omega_\ell),$$

$$\widetilde{D}(H_0) = \frac{1}{m}\sum_{\ell=1}^{m} \ell^{2H_0-1} I_R(\omega_\ell).$$

Basic calculations give us

$$\Delta L(H, H_0) = \log\left(\frac{\widehat{D}(H, H_0)}{\widetilde{D}(H_0)}\right).$$

Note that

$$a_\ell := \inf_{\Theta_2}\left(\frac{\ell}{p_m}\right)^{2(H-H_0)} = \begin{cases} \left(\frac{\ell}{p_m}\right)^{2(\Delta-H_0)}, & 1 \le \ell \le p_m, \\ \left(\frac{\ell}{p_m}\right)^{2(\Delta_1-H_0)}, & p_m + 1 \le \ell \le m. \end{cases}$$

Hence, $\widehat{D}(H, H_0) \ge \frac{1}{m}\sum_{\ell=1}^{m} a_\ell \ell^{2H_0-1} I_R(\omega_\ell)$. It follows that

$$\begin{aligned}\left\{\inf_{\Theta_2} \Delta L(H, H_0) \le 0\right\} &= \left\{\inf_{\Theta_2} \log \widehat{D}(H, H_0) \le \log \widetilde{D}(H_0)\right\} \\ &= \left\{\log\left(\inf_{\Theta_2} \widehat{D}(H, H_0)\right) \le \log \widetilde{D}(H_0)\right\} \\ &\subset \left\{\frac{1}{m}\sum_{\ell=1}^{m}(a_\ell - 1)\ell^{2H_0-1} I_R(\omega_\ell) \le 0\right\},\end{aligned}$$

which implies that

$$P\left(\inf_{\Theta_2} \Delta L(H, H_0) \le 0\right) \le P\left(\frac{1}{m}\sum_{\ell=1}^{m}(a_\ell - 1)\ell^{2H_0-1} I_R(\omega_\ell) \le 0\right).$$

Moreover, we have, as $m \to \infty$,

$$\begin{aligned}
\frac{1}{m}\sum_{\ell=1}^{m}(a_\ell - 1) &\geq \frac{1}{m}\sum_{\ell=1}^{p_m} a_\ell - 1 \\
&\sim \frac{1}{m}\cdot\frac{1}{p_m^{2(\Delta-H_0)}}\int_1^{p_m} x^{2(\Delta-H_0)}dx - 1 \\
&\sim \frac{p_m}{m[2(\Delta-H_0)+1]} - 1 \\
&\sim \frac{1}{e[2(\Delta-H_0)+1]} - 1,
\end{aligned}$$

where the last asymptotic equivalence relation comes from the fact that $\frac{p_m}{m} \sim \frac{1}{e}$, as $m \to \infty$. According to Li *et al.* (2021) (supplement material p. 4), we may choose $\Delta \leq H_0 - \frac{1}{2} + \frac{1}{4e}$ so that $\frac{1}{m}\sum_{\ell=1}^{m}(a_\ell - 1) \geq 1$, for sufficiently large m.

It follows that

$$\begin{aligned}
P\left(\frac{1}{m}\sum_{\ell=1}^{m}(a_\ell - 1)\ell^{2H_0-1}I_R(\omega_\ell) \leq 0\right) &= P\left(\frac{1}{m}\sum_{\ell=1}^{m}(a_\ell - 1)\frac{I_R(\omega_\ell)}{G_0\omega_\ell^{1-2H_0}} \leq 0\right) \\
&\leq P\left(\frac{1}{m}\sum_{\ell=1}^{m}(a_\ell - 1)\left(\frac{I_R(\omega_\ell)}{G_0\omega_\ell^{1-2H_0}} - 1\right) \leq -1\right) \\
&\leq P\left(\left|\frac{1}{m}\sum_{\ell=1}^{m}(a_\ell - 1)\left(\frac{I_R(\omega_\ell)}{G_0\omega_\ell^{1-2H_0}} - 1\right)\right| \geq 1\right).
\end{aligned}$$

By Lemma 3.7.4, this completes the proof.

■

We now state and prove several lemmas that were used in the above proof.

LEMMA 3.7.3 *Under the assumptions of Theorem 3.3.6,*

$$\sup_{\Theta_1}|T(H, H_0)| = o_P(1).$$

PROOF: Using (3.6.53) and a similar argument as in Lemma 3.6.8, we can obtain

$$\sup_{\Theta_1} |T(H, H_0)| = \sup_{\Theta_1} \left|\log \frac{\widehat{G}(H_0)}{G(H_0)}\right| + \sup_{\Theta_1} \left|\log \frac{\widehat{G}(H)}{G(H)}\right| + O\left(\frac{1}{m^{2(\Delta - H_0)+1}}\right) + O\left(\frac{\log m}{m}\right).$$

Observe that $2(\Delta - H_0) + 1 > 0$. Therefore, using an analogous argument as in the proof of Lemma 3.6.8, it remains to show

$$\sup_{\Theta_1} |A(H, H_0)| = o_P(1), \tag{3.7.8}$$

where $A(H, H_0)$ is defined in (3.6.55).

Let $I_1^{(Z)}(\omega_\ell) = \frac{1}{2\pi N} \left|\sum_{t=1}^N \xi_{1t}^{(Z)} e^{-it\omega_\ell}\right|^2$ and $g_\ell(H_0) = D_0 \omega_\ell^{1-2H_0}$. Note that

$$\begin{aligned} \sup_{H\in\Theta_1} |A(H, H_0)| \leq & \sup_{H\in\Theta_1} \left| \frac{2(H - H_0) + 1}{m} \sum_{\ell=1}^m \left(\frac{\ell}{m}\right)^{2(H-H_0)} \left(\frac{I_R(\omega_\ell) - I_1^{(Z)}(\omega_\ell)}{g_\ell(H_0)}\right)\right| \\ & + \sup_{H\in\Theta_1} \left| \frac{2(H - H_0) + 1}{m} \sum_{\ell=1}^m \left(\frac{\ell}{m}\right)^{2(H-H_0)} \left(\frac{I_1^{(Z)}(\omega_\ell)}{g_\ell(H_0)} - 1\right)\right|. \end{aligned}$$

The convergence of the second term on the right side is established in the proof of Theorem 1(i) in Li *et al.* (2021). It remains to show that

$$\sup_{H\in\Theta_1} \left|\widetilde{A}(H, H_0)\right| = o_P(1), \tag{3.7.9}$$

where

$$\widetilde{A}(H, H_0) = \frac{2(H - H_0) + 1}{m} \sum_{\ell=1}^m \left(\frac{\ell}{m}\right)^{2(H-H_0)} \left(\frac{I_R(\omega_\ell) - I_1^{(Z)}(\omega_\ell)}{g_\ell(H_0)}\right).$$

This is done by showing that the remainder $I_R(\omega_\ell) - I_1^{(Z)}(\omega_\ell)$ is sufficiently small, and the same idea has been utilized at the beginning of Section 3.6.5. However, in the LRD case, things become a little different due to the fact that the sum of error terms $\sum_{t=1}^N Z_t$ diverges faster than that in the SRD case, i.e. $\frac{1}{N} \sum_{t=1}^N Z_t = O_P\left(N^{H_0-1}\right)$.

Next, decompose the level 1 scores based on the residuals as in (3.3.17), i.e.

$$\widehat{\xi}_{1t}^{(R)} = \langle R_t, \widehat{\psi}_1^{(R)} \rangle = \langle R_t, \psi_1^{(Z)} \rangle + \langle R_t, \widehat{\psi}_1^{(R)} - \psi_1^{(Z)} \rangle.$$

Recall that

$$R_t = \begin{cases} Z_t - \overline{Z}_{1,t}, & t = 1, \dots, \hat{n}, \\ Z_t - \overline{Z}_{2,t}, & t = \hat{n}+1, \dots, N, \end{cases} \tag{3.7.10}$$

where $\overline{Z}_{1,t}, \overline{Z}_{2,t}$ are given in (3.7.3).

Then, the periodogram based on the $\widehat{\xi}_{1t}^{(R)}$ is given by

$$\begin{aligned} I_R(\omega_\ell) &= \frac{1}{2\pi N} \left| \sum_{t=1}^{N} \widehat{\xi}_{1t}^{(R)} e^{-it\omega_\ell} \right|^2 \qquad (3.7.11) \\ &= \frac{1}{2\pi N} \Bigg| \sum_{t=1}^{N} \langle Z_t, \psi_1^{(Z)} \rangle e^{-it\omega_\ell} + \langle \overline{Z}_{2,t} - \overline{Z}_{1,t}, \psi_1^{(Z)} \rangle \sum_{t=1}^{\hat{n}} e^{-it\omega_\ell} \\ &+ \sum_{t=1}^{N} \langle R_t, \widehat{\psi}_1^{(R)} - \psi_1^{(Z)} \rangle e^{-it\omega_\ell} \Bigg|^2 \\ &= |z_1|^2 + |z_2|^2 + |z_3|^2 + z_1\overline{z}_2 + z_1\overline{z}_3 + z_2\overline{z}_1 + z_2\overline{z}_3 + z_3\overline{z}_1 + z_3\overline{z}_2, \end{aligned}$$

where

$$|z_1|^2 = I_1^{(Z)}(\omega_\ell),$$

$$z_2 = \frac{1}{\sqrt{2\pi N}} \langle \overline{Z}_{2,t} - \overline{Z}_{1,t}, \psi_1^{(Z)} \rangle \sum_{t=1}^{\hat{n}} e^{-it\omega_\ell},$$

$$z_3 = \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^{N} \langle R_t, \widehat{\psi}_1^{(R)} - \psi_1^{(Z)} \rangle e^{-it\omega_\ell},$$

and $\overline{z}_1$, $\overline{z}_2$, $\overline{z}_3$ denote the complex conjugates of z_1, z_2 and z_3.

Note that $|z_1|^2 = I_1^{(Z)}(\omega_\ell)$ is the periodogram of the level 1 scores of the Z_t and from equation

(3.16) of Theorem 2 in Robinson (1995), we have

$$\sup_{1\le\ell\le m} \frac{E|z_1|^2}{D_0\omega_\ell^{1-2H_0}} \le C,$$

for some constant C. That is,

$$z_1 = O_P\left(\left(\frac{N}{\ell}\right)^{H_0-\frac{1}{2}}\right). \tag{3.7.12}$$

Using

$$\frac{1}{N}\sum_{t=1}^{\hat{n}} e^{-it\omega_\ell} = O_P\left(\frac{1}{\ell}\right) \tag{3.7.13}$$

and $\|\overline{Z}_{1,t}\|, \|\overline{Z}_{2,t}\| = O_P\left(N^{H_0-1}\right)$, we get

$$z_2 = O_P\left(\frac{N^{H_0-\frac{1}{2}}}{\ell}\right). \tag{3.7.14}$$

Note that z_2 is negligible compared to z_1. By Lemma 3.7.5

$$z_3 = o_P\left(\left(\frac{N}{\ell}\right)^{H_0-\frac{1}{2}}\right), \tag{3.7.15}$$

so, by (3.7.11), (3.7.12) and (3.7.15),

$$\frac{\left|I_R(\omega_\ell) - I_1^{(Z)}(\omega_\ell)\right|}{g_\ell(H_0)} = O_P\left(\ell^{H_0-\frac{3}{2}}\right) + o_P(1). \tag{3.7.16}$$

Observe that $0 < 2(H-H_0)+1 < 3$ for $H \in \Theta_1$. It follows that

$$
\begin{aligned}
\left|\widetilde{A}(H,H_0)\right| &\leq \frac{3}{m}\sum_{\ell=1}^{m}\left(\frac{\ell}{m}\right)^{2(H-H_0)}\frac{\left|I_R(\omega_\ell)-I_1^{(Z)}(\omega_\ell)\right|}{g_\ell(H_0)} \\
&= \frac{1}{m^{2(H-H_0)+1}}\left\{O_P\left(\sum_{l=1}^{m}\ell^{2(H-H_0)+H_0-\frac{3}{2}}\right)+o_P\left(\sum_{l=1}^{m}\ell^{2(H-H_0)}\right)\right\} \\
&\sim \frac{1}{m^{2(H-H_0)+1}}\left\{O_P\left(m^{2(H-H_0)+H_0-\frac{1}{2}}+\log m\right)+o_P\left(m^{2(H-H_0)+1}\right)\right\} \\
&= O_P\left(m^{H_0-\frac{3}{2}}+\frac{\log m}{m^{2(H-H_0)+1}}\right)+o_P(1) \\
&= o_P(1).
\end{aligned}
$$

■

LEMMA 3.7.4 *Under the assumptions of Theorem 3.3.6,*

$$
P\left(\left|\frac{1}{m}\sum_{\ell=1}^{m}(a_\ell-1)\left(\frac{I_R(\omega_\ell)}{D_0\omega_\ell^{1-2H_0}}-1\right)\right|\geq 1\right)\rightarrow 0, \quad \textit{as } N\rightarrow\infty.
$$

PROOF:

Note that

$$
\begin{aligned}
&P\left(\left|\frac{1}{m}\sum_{\ell=1}^{m}(a_\ell-1)\left(\frac{I_R(\omega_\ell)}{D_0\omega_\ell^{1-2H_0}}-1\right)\right|\geq 1\right) \\
&\leq P\left(\left|\frac{1}{m}\sum_{\ell=1}^{m}(a_\ell-1)\left(\frac{I_R(\omega_\ell)-I_1^{(Z)}(\omega_\ell)}{D_0\omega_\ell^{1-2H_0}}\right)\right|+\left|\frac{1}{m}\sum_{\ell=1}^{m}(a_\ell-1)\left(\frac{I_1^{(Z)}(\omega_\ell)}{D_0\omega_\ell^{1-2H_0}}-1\right)\right|\geq 1\right) \\
&\leq P\left(\left|\frac{1}{m}\sum_{\ell=1}^{m}(a_\ell-1)\left(\frac{I_R(\omega_\ell)-I_1^{(Z)}(\omega_\ell)}{D_0\omega_\ell^{1-2H_0}}\right)\right|\geq \frac{1}{2}\right) \\
&\quad + P\left(\left|\frac{1}{m}\sum_{\ell=1}^{m}(a_\ell-1)\left(\frac{I_1^{(Z)}(\omega_\ell)}{D_0\omega_\ell^{1-2H_0}}-1\right)\right|\geq \frac{1}{2}\right).
\end{aligned}
$$

Following the proof of Theorem 1(i) in Li *et al.* (2021) (see supplement material p. 4), we have

$$P\left(\left|\frac{1}{m}\sum_{\ell=1}^{m}(a_\ell-1)\left(\frac{I_1^{(Z)}(\omega_\ell)}{D_0\omega_\ell^{1-2H_0}}-1\right)\right|\geq\frac{1}{2}\right)\to 0,\quad \text{as } N\to\infty.$$

Observe that $0<a_\ell\leq 1$. By (3.7.16),

$$\begin{aligned}\left|\frac{1}{m}\sum_{\ell=1}^{m}(a_\ell-1)\left(\frac{I_R(\omega_\ell)-I_1^{(Z)}(\omega_\ell)}{D_0\omega_\ell^{1-2H_0}}\right)\right| &\leq \frac{2}{m}\sum_{\ell=1}^{m}\left|\frac{I_R(\omega_\ell)-I_1^{(Z)}(\omega_\ell)}{D_0\omega_\ell^{1-2H_0}}\right| \\ &= \frac{1}{m}\sum_{\ell=1}^{m}\left[O_P\left(\ell^{H_0-\frac{3}{2}}\right)+o_P(1)\right] \\ &= O_P\left(m^{H_0-\frac{3}{2}}\right)+o_P(1) \\ &= o_P(1),\end{aligned}$$

which implies

$$P\left(\left|\frac{1}{m}\sum_{\ell=1}^{m}(a_\ell-1)\left(\frac{I_R(\omega_\ell)-I_1^{(Z)}(\omega_\ell)}{D_0\omega_\ell^{1-2H_0}}\right)\right|\geq\frac{1}{2}\right)\to 0,\quad \text{as } N\to\infty.$$

■

LEMMA 3.7.5 *Under the assumptions of Theorem 3.3.6,*

$$\frac{1}{\sqrt{2\pi N}}\sum_{t=1}^{N}\langle R_t,\widehat{\psi}_1^{(R)}-\psi_1^{(Z)}\rangle e^{-it\omega_\ell}=o_P\left(\left(\frac{N}{\ell}\right)^{H_0-\frac{1}{2}}\right).$$

PROOF: Note that by the Cauchy-Schwarz inequality,

$$\left|\frac{1}{\sqrt{2\pi N}}\sum_{t=1}^{N}\langle R_t,\widehat{\psi}_1^{(R)}-\psi_1^{(Z)}\rangle e^{-it\omega_\ell}\right|\leq\frac{1}{\sqrt{2\pi N}}\left\|\sum_{t=1}^{N}R_t e^{-it\omega_\ell}\right\|\left\|\widehat{\psi}_1^{(R)}-\psi_1^{(Z)}\right\|.$$

Observe that

$$\begin{aligned}\left\|\sum_{t=1}^{N} R_t e^{-it\omega_\ell}\right\| &= \left\|\sum_{t=1}^{N} Z_t e^{-it\omega_\ell} + \left(\overline{Z}_{2,t} - \overline{Z}_{1,t}\right)\sum_{t=1}^{\hat{n}} e^{-it\omega_\ell}\right\| \\ &\leq \left\|\sum_{t=1}^{N} Z_t e^{-it\omega_\ell}\right\| + \left|\sum_{t=1}^{\hat{n}} e^{-it\omega_\ell}\right| \left\|\overline{Z}_{2,t} - \overline{Z}_{1,t}\right\| \\ &= \left\|\sum_{t=1}^{N} Z_t e^{-it\omega_\ell}\right\| + O_P\left(\frac{N^{H_0}}{\ell}\right)\end{aligned}$$

By Theorem 2 in Robinson (1995b), we can show that uniformly for all $u \in [0,1]$,

$$E\left|\sum_{t=1}^{N} Z_t(u) e^{-it\omega_\ell}\right|^2 = O\left(\frac{N^{2H_0}}{\ell^{2H_0-1}}\right),$$

which implies that

$$\left\|\sum_{t=1}^{N} Z_t e^{-it\omega_\ell}\right\| = O_P\left(\frac{N^{H_0}}{\ell^{H_0-\frac{1}{2}}}\right),$$

and that

$$\frac{1}{\sqrt{2\pi N}}\left\|\sum_{t=1}^{N} R_t e^{-it\omega_\ell}\right\| = O_P\left(\left(\frac{N}{\ell}\right)^{H_0-\frac{1}{2}}\right). \tag{3.7.17}$$

Thus, the claim follows by combining (3.3.21) and (3.7.17).

∎

3.8 Additional tables

Table 3.6: Empirical sizes (in percent), based on $2,000$ replications, with bandwidth $h = N^{0.2}$. The true change point is defined by $n^\star = \lfloor N\theta \rfloor$.

Nominal size	1.0			5.0			10.0		
N	500	1000	1500	500	1000	1500	500	1000	1500
				$\delta = 0.25, \theta = 0.50$					
$m = N^{0.55}$	1.3	1.1	1.2	3.7	4.5	3.6	6.8	7.5	7.0
$m = N^{0.60}$	1.3	0.8	0.8	4.8	3.9	3.3	8.3	7.8	8.1
$m = N^{0.65}$	1.9	1.5	0.8	6.0	5.3	4.2	10.3	9.2	8.6
				$\delta = 0.50, \theta = 0.50$					
$m = N^{0.55}$	1.6	1.4	1.3	4.7	5.0	4.5	8.8	8.6	8.5
$m = N^{0.60}$	1.5	1.1	1.0	5.7	4.9	4.5	9.8	9.0	9.2
$m = N^{0.65}$	2.4	1.6	1.2	7.4	6.2	5.2	12.0	10.5	10.0
				$\delta = 0.25, \theta = 0.25$					
$m = N^{0.55}$	1.0	1.3	1.4	4.9	4.8	4.9	8.9	8.4	9.3
$m = N^{0.60}$	1.5	0.9	1.2	5.1	4.4	4.6	9.8	9.0	9.8
$m = N^{0.65}$	2.2	1.6	1.3	6.6	5.5	5.9	11.4	10.7	11.0
				$\delta = 0.50, \theta = 0.25$					
$m = N^{0.55}$	1.7	2.2	2.2	7.0	6.2	6.2	12.2	10.9	10.8
$m = N^{0.60}$	2.3	1.9	1.6	7.3	6.4	5.9	12.7	11.8	11.6
$m = N^{0.65}$	3.0	2.7	1.7	9.1	8.1	7.4	14.5	13.9	13.1

Table 3.7: Empirical powers (in percent) for the two functions g. The bandwidth h and the number of replications are the same as in Table 3.6.

Nominal size	1.0			5.0			10.0		
N	500	1000	1500	500	1000	1500	500	1000	1500
	$g_1, H_0 = 0.6$								
$m = N^{0.55}$	68.8	82.7	90.1	83.9	92.7	95.8	89.6	95.5	97.8
$m = N^{0.60}$	90.0	97.4	99.0	96.4	98.9	99.7	98.3	99.7	99.9
$m = N^{0.65}$	98.9	99.9	100.0	99.6	100.0	100.0	99.7	100.0	100.0
	$g_1, H_0 = 0.9$								
$m = N^{0.55}$	96.1	99.4	99.9	99.0	99.9	100.0	99.5	99.9	100.0
$m = N^{0.60}$	99.7	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0
$m = N^{0.65}$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	$g_2, H_0 = 0.6$								
$m = N^{0.55}$	30.3	56.4	69.6	48.8	73.7	85.2	58.8	81.7	90.1
$m = N^{0.60}$	45.5	67.1	86.7	66.1	84.0	94.8	75.7	90.1	97.4
$m = N^{0.65}$	91.1	93.1	94.6	96.7	97.1	98.3	98.3	98.3	99.0
	$g_2, H_0 = 0.9$								
$m = N^{0.55}$	63.9	95.1	98.8	80.2	98.1	99.7	86.8	98.9	99.8
$m = N^{0.60}$	66.8	96.0	99.8	82.4	98.7	100.0	87.7	99.4	100.0
$m = N^{0.65}$	98.1	98.8	99.6	99.2	99.8	100.0	99.6	99.9	100.0

Table 3.8: Empirical sizes (in percent), based on $2,000$ replications, with bandwidth $h = N^{0.3}$. The true change point is defined by $n^\star = \lfloor N\theta \rfloor$.

Nominal size	1.0			5.0			10.0		
N	500	1000	1500	500	1000	1500	500	1000	1500
	$\delta = 0.25, \theta = 0.50$								
$m = N^{0.55}$	1.3	1.1	1.2	3.8	4.6	3.8	7.0	7.5	7.0
$m = N^{0.60}$	1.3	0.8	0.8	4.7	4.1	3.3	8.3	7.9	8.2
$m = N^{0.65}$	2.0	1.5	0.8	6.0	5.3	4.3	10.5	9.3	8.6
	$\delta = 0.50, \theta = 0.50$								
$m = N^{0.55}$	1.6	1.4	1.4	4.7	5.1	4.6	9.0	8.6	8.5
$m = N^{0.60}$	1.5	1.1	1.0	5.7	4.9	4.6	9.9	9.1	9.3
$m = N^{0.65}$	2.4	1.6	1.2	7.5	6.3	5.2	12.0	10.7	10.1
	$\delta = 0.25, \theta = 0.25$								
$m = N^{0.55}$	1.0	1.3	1.4	4.9	4.9	5.0	8.9	8.4	9.5
$m = N^{0.60}$	1.5	0.9	1.2	5.1	4.6	4.7	9.9	9.2	9.8
$m = N^{0.65}$	2.2	1.7	1.3	6.7	5.6	5.9	11.6	10.8	11.0
	$\delta = 0.50, \theta = 0.25$								
$m = N^{0.55}$	1.8	2.2	2.2	7.3	6.4	6.4	12.4	10.9	11.0
$m = N^{0.60}$	2.3	1.9	1.7	7.4	6.7	6.1	13.0	12.0	11.7
$m = N^{0.65}$	3.1	2.7	1.8	9.4	8.3	7.4	15.0	14.2	13.5

Table 3.9: Empirical powers (in percent) for the two functions g. The bandwidth h and the number of replications are the same as in Table 3.8.

Nominal size	1.0			5.0			10.0		
N	500	1000	1500	500	1000	1500	500	1000	1500
				$g_1, H_0 = 0.6$					
$m = N^{0.55}$	69.4	82.9	90.1	84.2	92.8	96.0	89.8	95.8	97.9
$m = N^{0.60}$	90.1	97.4	99.1	96.5	98.9	99.7	98.4	99.7	99.9
$m = N^{0.65}$	98.9	99.9	100.0	99.6	100.0	100.0	99.7	100.0	100.0
				$g_1, H_0 = 0.9$					
$m = N^{0.55}$	96.5	99.4	99.9	99.0	99.9	100.0	99.5	99.9	100.0
$m = N^{0.60}$	99.7	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0
$m = N^{0.65}$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
				$g_2, H_0 = 0.6$					
$m = N^{0.55}$	30.5	56.9	69.7	49.0	73.8	85.3	59.1	81.7	90.2
$m = N^{0.60}$	46.1	67.5	86.8	66.4	84.3	95.0	76.1	90.3	97.5
$m = N^{0.65}$	91.3	93.3	94.7	96.9	97.2	98.4	98.3	98.4	99.0
				$g_2, H_0 = 0.9$					
$m = N^{0.55}$	64.5	95.1	98.9	80.4	98.1	99.7	86.9	98.9	99.8
$m = N^{0.60}$	67.1	96.0	99.8	82.9	98.7	100.0	87.9	99.4	100.0
$m = N^{0.65}$	98.1	98.8	99.6	99.2	99.8	100.0	99.6	99.9	100.0

Table 3.10: Empirical sizes (in percent), based on $2,000$ replications, with bandwidth $h = N^{0.4}$. The true change point is defined by $n^\star = \lfloor N\theta \rfloor$.

Nominal size	1.0			5.0			10.0		
N	500	1000	1500	500	1000	1500	500	1000	1500
				$\delta = 0.25, \theta = 0.50$					
$m = N^{0.55}$	1.3	1.1	1.2	3.9	4.7	3.8	7.2	7.8	7.1
$m = N^{0.60}$	1.3	0.8	0.9	4.9	4.2	3.5	8.6	8.2	8.4
$m = N^{0.65}$	2.0	1.6	0.8	6.1	5.4	4.4	10.7	9.5	8.7
				$\delta = 0.50, \theta = 0.50$					
$m = N^{0.55}$	1.6	1.4	1.4	4.9	5.3	4.6	9.1	8.7	8.6
$m = N^{0.60}$	1.5	1.1	1.1	5.7	5.0	4.7	10.2	9.3	9.7
$m = N^{0.65}$	2.4	1.6	1.2	7.5	6.3	5.2	12.3	11.0	10.2
				$\delta = 0.25, \theta = 0.25$					
$m = N^{0.55}$	1.1	1.3	1.4	5.1	5.0	5.1	9.2	8.6	9.6
$m = N^{0.60}$	1.5	1.0	1.2	5.5	4.7	4.9	10.2	9.5	10.0
$m = N^{0.65}$	2.3	1.7	1.3	6.7	5.9	5.9	12.1	10.9	11.3
				$\delta = 0.50, \theta = 0.25$					
$m = N^{0.55}$	1.8	2.3	2.2	7.5	6.7	6.5	12.5	11.2	11.4
$m = N^{0.60}$	2.4	1.9	1.8	7.8	6.8	6.3	13.5	12.3	12.0
$m = N^{0.65}$	3.1	2.9	1.9	9.7	8.7	7.6	15.5	14.7	14.2

Table 3.11: Empirical powers (in percent) for the two functions g. The bandwidth h and the number of replications are the same as in Table 3.10.

Nominal size	1.0			5.0			10.0		
N	500	1000	1500	500	1000	1500	500	1000	1500
				$g_1, H_0 = 0.6$					
$m = N^{0.55}$	70.0	83.4	90.3	84.5	93.1	96.1	90.2	95.9	98.1
$m = N^{0.60}$	90.8	97.5	99.2	96.8	99.1	99.7	98.6	99.7	99.9
$m = N^{0.65}$	99.0	99.9	100.0	99.6	100.0	100.0	99.9	100.0	100.0
				$g_1, H_0 = 0.9$					
$m = N^{0.55}$	96.7	99.5	99.9	99.1	99.9	100.0	99.5	99.9	100.0
$m = N^{0.60}$	99.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
$m = N^{0.65}$	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
				$g_2, H_0 = 0.6$					
$m = N^{0.55}$	30.9	57.9	70.0	49.8	74.1	85.5	59.6	82.4	90.2
$m = N^{0.60}$	46.7	68.3	87.4	67.7	84.6	95.2	76.7	90.5	97.5
$m = N^{0.65}$	91.8	93.6	94.9	97.2	97.5	98.5	98.4	98.6	99.0
				$g_2, H_0 = 0.9$					
$m = N^{0.55}$	65.6	95.4	99.0	81.7	98.3	99.7	87.4	98.9	99.8
$m = N^{0.60}$	68.1	96.5	99.8	83.6	98.8	100.0	88.7	99.5	100.0
$m = N^{0.65}$	98.4	98.9	99.6	99.3	99.9	100.0	99.6	99.9	100.0

3.9 Implementation details and R code

Following the descriptions of the data generating processes in Section 3.4, we need to simulate iid standard Brownian motions on the unit interval $[0, 1]$. Each Brownian motion was generated by computing the cumulative sum of independent normal random variables on the equally-spaced points $0, \frac{1}{500}, \frac{2}{500}, \ldots, 1$. As a consequence, each FAR(1) process $\{Y_t\}$ in (3.4.1) was generated at those points by standard matrix additions and multiplications. Similarly, we generated the functional long-range dependent process $\{X_t\}$ in (3.4.2). Then, we calculated the long-run covariance function described in (3.2.4) on the grid $\left\{(u, v) : u, v = 0, \frac{1}{500}, \frac{2}{500}, \ldots, 1\right\}$, which results in a square matrix of size 501×501. In computing the change point estimator (3.2.3) and the level-1 scores, integrals were approximated by the Riemann sums and the partition is defined as the aforementioned points. When we conducted the functional principal component analysis in the simulation study and in the ETF data application, we utilized the discretization technique described in Chapter 8.4.1 in Ramsay and Silverman (2005). The local Whittle estimator was computed in accordance with its definition in Section 3.2, where we chose $[\Delta_1, \Delta_2] = [0.0001, 0.9999]$ in (3.2.6), as is suggested by Li *et al.* (2021). Complete and commented code is available for downloads as a separate file.

The following code should be saved as `FTS-library.R`.

```
#################################################################
# simulate the functional AR(1) model with single change point #
#################################################################
# The error terms are iid standard Brownian motions.

FAR_CM1 <- function(K, N, delta, nstar){
  # This function simulates the FAR(1) model with single change point
  for a given sample size N.
  # It returns a numeric matrix with dimension (K+1) * N, in which
  each sample/curve is stored in each column.
```

```
# K+1 = the number of observations in the discrete version of
each curve;
# delta = change-in-mean level; It should be a numeric value.
# nstar stores the location of change point; It should be an
integer between 1 to N (both exclusive).

burn <- 100  # burn-in size
N1 <- N + burn

time <- seq(0, K)/(K)

# generate iid standard BM
S <- matrix(rnorm(K*N1, mean=0, sd=sqrt(1/K)), ncol=N1)
S <- rbind(rep(0, N1), S)
W <- apply(S, 2, cumsum)

# adjust for an additional zero row in BM
K <- K + 1

# compute the kernel function of the FAR(1) model
P <- 0.25
Tsq <- time^2
U <- matrix(rep(Tsq, K), ncol = K)
V <- t(U)
G <- (P/.7468241328)*exp(-0.5*(U+V))

# generate the FAR(1) process (without change point)
X <- matrix(0, K, N1)
for(i in 1:(N1-1)){
```

```
    X[,i+1] <- G%*%X[,i]*(1/K) + W[,i]
  }

  XX <- X[,-(1:burn)]

  # add change-in-mean level delta
  id <- seq(nstar + 1, N, by=1)
  XX[,id] <- XX[,id] + delta


  return(XX)
}

###################################################################
# simulate iid standard Brownian motions with single change point #
###################################################################

BM_CM1 <- function(K, N, delta, nstar){
  # This function simulates iid standard Brownian motions with single
  change point for a given sample size N.
  # It returns a numeric matrix with dimension (K+1) * N, in which
  each sample/curve is stored in each column.
  # K+1 = the number of observations in the discrete version of
  each curve;
  # delta = change-in-mean level; It should be a numeric value.
  # nstar stores the location of change point; It should be an
  integer between 1 to N (both exclusive).

  time <- seq(0, K)/(K)
```

```
  # generate iid standard BM
  S <- matrix(rnorm(K*N, mean=0, sd=sqrt(1/K)), ncol=N)
  S <- rbind(rep(0, N), S)
  W <- apply(S, 2, cumsum)

  # adjust for an additional zero row in BM
  K <- K + 1

  # add change-in-mean level delta
  id <- seq(nstar + 1, N, by=1)
  W[,id] <- W[,id] + delta

  return(W)
}

######################################################
# simulate functional long-range dependent processes #
######################################################

FLM_lrd_1 <- function(K, N, L = 1500, H0){
  # This function simulates functional LRD processes for a given
  sample size N.
  # The "disturbance function" is chosen as g_1 = 1 (as described
  in the paper).
  # It returns a numeric matrix with dimension (K+1) * N, in which
  each sample/curve is stored in each column.
  # K+1 = the number of observations in the discrete version of
  each curve;
```

```
  # L = truncation level of each functional moving average process;
  # H0 = self-similarity parameter; 1/2 < H0 < 1.

  Xt <- matrix(data = NA, nrow = K + 1, ncol = N) # K + 1 adjusts for
  an additional zero row

  # compute the coefficients b_j
  bj <- (1:L)^(H0 - 3/2)
  bj <- as.matrix(bj)

  # generate iid standard BMs
  S <- matrix(rnorm(K*(N + L - 1), mean=0, sd=sqrt(1/K)),
  ncol= N + L - 1)
  S <- rbind(rep(0, N + L - 1), S)
  B <- apply(S, 2, cumsum)

  # compute functional LRD processes
  for (i in 1:N){
    Xt[,i] <- B[,(i+L-1):i] %*% bj
  }

  return(Xt)
}

FLM_lrd_2 <- function(K, N, L = 1500, H0){
  # This function simulates functional LRD processes for a given
  sample size N.
  # The "disturbance function" is chosen as g_2 (as described
  in the paper).
```

```
# It returns a numeric matrix with dimension (K+1) * N, in which
each sample/curve is stored in each column.
# K+1 = the number of observations in the discrete version of
each curve;
# L = truncation level of each functional moving average process;
# H0 = self-similarity parameter; 1/2 < H0 < 1.

Xt <- matrix(data = NA, nrow = K + 1, ncol = N) #adjust for
an additional zero row

# compute the coefficients b_j
bj <- (1:L)^(H0 - 3/2)

# compute the disturbance function g_2
gj <- rep(1,L)
p = 15
gj[1:L] <- sin(2*pi/p * (1:L))
gj[1:L] <- exp(-0.01 * (1:L) - 1/2) * gj[1:L] + 1

# compute the coefficients for the BMs
bj <- gj * bj
bj <- as.matrix(bj)

# generate iid standard BMs
S <- matrix(rnorm(K*(N + L - 1), mean=0, sd=sqrt(1/K)),
ncol= N + L - 1)
S <- rbind(rep(0, N + L - 1), S)
B <- apply(S, 2, cumsum)
```

```
  # compute functional LRD processes
  for (i in 1:N){
    Xt[,i] <- B[,(i+L-1):i] %*% bj
  }

  return(Xt)
}

###########################
# change point estimation #
###########################

Dhat <- function(Xt){
  # This function computes the change point estimator (see,
  Aue et al. (2018)) for a dataset Xt
  # Xt should be a numeric matrix.

  K <- nrow(Xt); N <- ncol(Xt); N1 <- N-1
  DK <- numeric(N1)

  for(i in 2:N1){
   aa <- (rowMeans(as.matrix(Xt[,(1:i)])) -
   rowMeans(as.matrix(Xt[,-(1:i)])))^2
   aa <- sum(aa)/K
   DK[i] <- i^2*(N-i)^2/(N^2)*aa
  }

  nhat <- which.max(DK)
```

```
  return(list(nhat = nhat))

}

############################################################
# compute level-1 scores based on the residual process Rt #
############################################################

level1score <- function(Xt, nhat, h, long_run_cov){
  # This function computes the (long-run) covariance function (with
  Bartlett kernel)
  and level-1 scores based on the residual process.
  # Xt is the original data, which should be a numeric matrix.
  # nhat is the estimated change point and should be an integer.
  # h is the bandwidth parameter
  # long_run_cov should be a logical value.
  # If long_run_cov is FALSE, then the function computes the usual
  covariance kernel function and the corresponding level-1 scores;
  # otherwise, it computes the long-run covariance function with
  Bartlett kernel.
  # Note:
  # The function return a list which contains:
  # (i) The (long-run) covariance kernel function in a matrix.
  # (ii) Level-1 scores in a matrix.

  # compute the dimension of the data matrix
  K <- nrow(Xt); N <- ncol(Xt)

  # compute the residual process Rt by subtracting local means from
```

```
the original process
Rt <- matrix(NA, K, N)
X1 <- rowMeans(Xt[,(1:nhat)])
X2 <- rowMeans(Xt[,-(1:nhat)])
Rt <- cbind(Xt[,1:nhat] - X1, Xt[,-(1:nhat)] - X2)

# compute gamma_0, i.e. the covariance kernel function
gamma_0 <- matrix(NA, K, K)
gamma_0 <- Rt %*% t(Rt)/N

if (long_run_cov == FALSE){
  # compute the first eigenfunction and level-1 scores of
  the covariance kernel function
  ee <- eigen(gamma_0)
  psi1 <- ee$vectors[,1]*(1/K)^(-1/2)
  level1_scores <- Re(t(Rt)%*%psi1)*(1/K)

  return(list(cR = gamma_0,
              level1 = level1_scores))
}

# set up the kernel function K (Bartlett kernel)
h1 <- floor(h)
tt <- seq(from = 1, to = h1)
weight <- 1 - abs(tt/h)

# initialization
cR <- matrix(NA, K, K)
GamR <- array(NA, dim=c(K,K,h1 + 1)) # h1 + 1 accounts for gamma_0
```

```
GamR[,,1] <- gamma_0


# Compute sample autocovariance functions for t = 1,..., floor(h)
for(t in 1:h1){
u <- Rt[,((t+1):N)]
v <- Rt[,(1:(N-t))]
GamR[,,(t+1)] <- (u%*%t(v))/N
}

# compute the long-run covariance kernel function
cR <- gamma_0
for(i in 1:h1){
cR <- cR + weight[i] * (GamR[,,(i+1)] + t(GamR[,,(i+1)]))
}

# compute the first eigenfunction and level-1 scores
ee <- eigen(cR)
psi1 <- ee$vectors[,1]*(1/K)^(-1/2)
level1_scores <- Re(t(Rt)%*%psi1)*(1/K)


return(list(cR = cR,
            level1 = level1_scores))

}
```

```
#######################################
# compute the local Whittle estimator #
#######################################

lwe2 <- function(data, m, lowerd, upperd){
  # This function computes the local Whittle estimator for the
  fractional difference parameter d.
  # data should be a numeric vector.
  # m = number of low Fourier frequencies
  # lowerd and upperd are the lower and upper bounds of the
  optimization range in LWE.
  # The function returns a numeric vector that contains m and
  the LWE for d.

  # specify the optimization range $[\Delta_1, \Delta_2]$
  if(missing(lowerd)) {lowerd = -.5}
  if(missing(upperd)) {upperd = 1}

  N <- length(data) # sample size

  # specify the number of low Fourier frequencies
  if(missing(m)) {m = sqrt(N)}
  m <- floor(m)

  # compute DFT
  dft <- fft(data)

  # extract DFT at Fourier frequencies in (0, pi].
  M <- floor(N/2)
```

```
  dft <- dft[2:(M+1)] # exclude redundant frequencies (the first
  element is for zero frequency)

  # compute Fourier frequencies in (0, pi]
  omega <- 2*pi/N * seq(from = 1, to = M, by = 1)

  # compute the periodogram
  periodogram <- 1/(2*pi*N) * abs(dft)^2



  # compute the GPH estimator
  y <- log(periodogram)[1:m]
  nu <- log(omega[1:m]) - mean(log(omega[1:m]))
  GPH <- -sum(nu*y)/(2*sum(nu^2))

  # use GPH as the initial value if it falls into the prespecified
  optimization range
  init <- GPH*(GPH > lowerd)*(GPH < upperd) +
  .1*(1 - (GPH > lowerd)*(GPH < upperd))

  # compute the LWE
  est <- optim(par = init, lw_R, omega = omega, power = periodogram,
        m = m, method = "L-BFGS-B", lower = lowerd, upper = upperd)
  lwd <- est$par

  out <- c(m, lwd)
  names(out) <- c("m", "d_hat")
  return(out)
}
```

```
##############################
# the object function of LWE #
##############################

lw_R <- function(d, omega, power, m){
  lambda <- omega[1:m]
  Ix <- power[1:m]
  lwd <- log(mean((lambda^(2*d))*Ix)) - 2*d*mean(log(lambda))
  return(lwd)
}
```

The following code should be saved as `D_cpp1.cpp`. It contains the change point estimation function in the Rcpp version.

```
// [[Rcpp::depends(RcppArmadillo)]]
#include <RcppArmadillo.h>
#include <Rcpp.h>
using namespace arma;
using namespace Rcpp;
// [[Rcpp::export]]
List D_cpp(arma::mat Xt){
  int K = Xt.n_rows;
  int N = Xt.n_cols;
  double K1 = K;
  double N1 = N;
  arma::vec Dk(N-1);
  arma::vec RS(K);
  double rs;
  int n_hat;
  for(int k = 0; k < N - 1; k++){
    RS = pow(sum(Xt.submat(0,0,K-1,k),1)/(k+1) -
    sum(Xt.submat(0,k+1,K-1,N-1),1)/(N1-(k+1)),2);
    rs = sum(RS)/K1;
    Dk[k] = (k+1)*(k+1)*(N1-(k+1))*(N1-(k+1))*rs/(N1*N1);
    }
  n_hat = index_max(Dk) + 1;
  List output = List::create(Named("nhat") = n_hat,
                             Named("Dk") = Dk);
  return output;
  }
```

The following code computes the empirical sizes.

```
####################
# library packages #
####################
library(doParallel)
library(doRNG)

###########################
# compute empirical sizes #
###########################

size_change_point <- function(N, r = 2000, K = 500, delta, nstar,
h_pow, m_pow) {
  # This function computes the empirical Type I errors of the
  change point v.s. LRD test.
  # N = sample size, i.e. number of curves; r = number of
  replications;
  # K+1 = the number of observations in the discrete version of
  each curve;
  # delta = change-in-mean level; It should be a numeric value.
  # nstar stores the location of the change point; It should be an
  integer between 1 to N (both exclusive).

  # h_pow specifies the bandwidth h via h = N^h_pow.
  # If h_pow equals 0, then the error functions will be iid standard
  BMs. The covariance function will be used in conducting FPCA.
  # m_pow specifies the number of Fourier frequencies m via
  m = N^m_pow.
```

```
# Note:
# 1. Both h_pow and m_pow can be numeric vectors with elements
greater than 0 and less than 1.
# 2. The function returns a list which contains:
# (i) Test statistics for each combination of h and m (stored in a
3D array).
# (ii) Rejection rates for each combination of h and m at different
nominal sizes (i.e. 0.01, 0.05 and 0.10).
# (iii) Change point estimates.
# (iv) The value(s) of h and m.

# determine whether the long-run covariance function is used
in FPCA

if (length(h_pow) == 1 && h_pow == 0){
  long_run_cov <- FALSE
}
else{
  long_run_cov <- TRUE
}

# compute h and m
h <- N^h_pow
m <- N^m_pow

# critical values for different nominal sizes
cv_01 <- qnorm(1-0.01)
cv_05 <- qnorm(1-0.05)
cv_10 <- qnorm(1-0.10)
```

```
# set up cores for parallel computing
cores <- detectCores()
cl <- parallel::makeCluster(cores[1]-2,
setup_strategy = "sequential")
registerDoParallel(cl)

PP <- foreach(b=1:r, .combine='rbind') %dorng% {

  library(Rcpp)
  # The change point estimation function is written in Rcpp to
  reduce the running time.
  # This is an alternative version in "FTS-library.R" (see Dhat()).
  sourceCpp("D_cpp1.cpp")

  source("FTS-library.R")

  # generate data
  if (length(h_pow) == 1 && h_pow == 0){
    dat <- BM_CM1(K = K, N = N, delta = delta, nstar = nstar)
  }
  else{
    dat <- FAR_CM1(K = K, N = N, delta = delta, nstar = nstar)
  }

  # estimate change-point
  nhat1 <- D_cpp(dat)$nhat

  # compute level-1 scores for different bandwidth h
```

```
  level_1_scores_mat <- matrix(nrow = N, ncol = length(h))
  for (i in 1:length(h)) {
    level_1_scores <- level1score(dat, nhat1, h[i], long_run_cov)
    level_1_scores_mat[,i] <- level_1_scores$level1
  }

  # compute the test statistics for each combination of h and m
  TS_mat <- matrix(nrow = length(h), ncol = length(m))
  for (i in 1:length(h)) {
    for (j in 1:length(m)) {
      # compute the fractional difference parameter d = H - 0.5
       diff_par <- lwe2(level_1_scores_mat[,i], m = m[j],
       lowerd = -0.4999, upperd = 0.4999)[[2]]
      TS_mat[i,j] <- 2 * sqrt(m[j]) * diff_par
    }
  }

  # convert the test statistics into a long vector for the
  sake of storage
  # concatenation is by row
  TS_vec <- as.vector(t(TS_mat))

  # return the test statistics and the change point estimate
  c(TS_vec, nhat1)
}

stopCluster(cl)

# extract test statistics and change point estimates
```

```
nhat_id <- dim(PP)[2]
TS_mat <- as.matrix(PP[,-nhat_id])
nhat <- as.matrix(PP[,nhat_id])

# rejection rate at alpha = 0.01
decision_01 <- ifelse(TS_mat > cv_01, 1, 0)
rej_01 <- colMeans(decision_01)

# rejection rate at alpha = 0.05
decision_05 <- ifelse(TS_mat > cv_05, 1, 0)
rej_05 <- colMeans(decision_05)

# rejection rate at alpha = 0.10
decision_10 <- ifelse(TS_mat > cv_10, 1, 0)
rej_10 <- colMeans(decision_10)

# create an output list
out <- list("TS" = NA,
            "rej_rate_0.01" = NA,
            "rej_rate_0.05" = NA,
            "rej_rate_0.10" = NA,
            "nhat" = NA, 'm' = NA)

# store change point estimates
out$nhat <- nhat

# create matrices of rejection rates for each combination of
h and m at different nominal sizes
out$rej_rate_0.01 <- matrix(data = rej_01, nrow = length(h),
```

```
ncol = length(m), byrow = T)
out$rej_rate_0.05 <- matrix(data = rej_05, nrow = length(h),
ncol = length(m), byrow = T)
out$rej_rate_0.10 <- matrix(data = rej_10, nrow = length(h),
ncol = length(m), byrow = T)

# add row and column names
h_name <- character()
m_name <- character()
if (length(h_pow) == 1 && h_pow == 0){
  h_name <- ''
}
else{
  for (i in 1:length(h_pow)) {
    h_name[i] <- paste("h=N^", h_pow[i], sep = '')
  }
}
for (j in 1:length(m_pow)) {
  m_name[j] <- paste("m=N^", m_pow[j], sep = '')
}

for (k in 2:4){
  rownames(out[[k]]) <- h_name
  colnames(out[[k]]) <- m_name
}

# convert test statistics matrix into a 3D array
TS_array <- array(data = NA, dim = c(length(h), length(m), r),
dimnames = list(h_name, m_name))
```

```
  for (k in 1:r) {
    TS_array[,,k] <- matrix(data = TS_mat[k,], nrow = length(h),
    ncol = length(m), byrow = T)
  }

  # store test statistics
  out$TS <- TS_array

  # store value(s) of h and m
  if (!(length(h_pow) == 1 && h_pow == 0)) {
    out$h <- h_name
  }
  out$m <- m_name

  return(out)

}

rej_rates_null_iid <- function(N, r, delta, theta, m_pow){
  # nstar = floor(N * theta)
  out <- size_change_point(N = N, r = r, delta = delta,
  nstar = floor(N*theta), h_pow = 0, m_pow = m_pow)
  saveRDS(out, file = paste('null', delta, N, theta, 'iid=TRUE',
  '.rds', sep = '_' ))
}

rej_rates_null_weak_dependence <- function(N, r, delta, theta,
h_pow, m_pow){
  # nstar = floor(N * theta)
```

```
  out <- size_change_point(N = N, r = r, delta = delta,
  nstar = floor(N*theta), h_pow = h_pow, m_pow = m_pow)
  saveRDS(out, file = paste('null', delta, N, theta, 'iid=FALSE',
  '.rds', sep = '_' ))
}

sample_size <- c(500, 1000, 1500)
h_pow <- c(0.2, 0.3, 0.4)
m_pow <- c(0.55, 0.6, 0.65)

#############################
# delta = 0.25; theta = 1/2 #
#############################

for (i in sample_size) {
    set.seed(2023)
    rej_rates_null_iid(N = i, r = 2000, delta = 0.25, theta = 1/2,
    m_pow = m_pow)
    set.seed(2023)
    rej_rates_null_weak_dependence(N = i, r = 2000, delta = 0.25,
    theta = 1/2, h_pow = h_pow, m_pow = m_pow)

}

############################
# delta = 0.5; theta = 1/2 #
############################

for (i in sample_size) {
```

```
  set.seed(2023)
  rej_rates_null_iid(N = i, r = 2000, delta = 0.5, theta = 1/2,
  m_pow = m_pow)
  set.seed(2023)
  rej_rates_null_weak_dependence(N = i, r = 2000, delta = 0.5,
  theta = 1/2, h_pow = h_pow, m_pow = m_pow)

}

#############################
# delta = 0.25; theta = 1/4 #
#############################

for (i in sample_size) {
  set.seed(2023)
  rej_rates_null_iid(N = i, r = 2000, delta = 0.25, theta = 1/4,
  m_pow = m_pow)
  set.seed(2023)
  rej_rates_null_weak_dependence(N = i, r = 2000, delta = 0.25,
  theta = 1/4, h_pow = h_pow, m_pow = m_pow)

}

############################
# delta = 0.5; theta = 1/4 #
############################

for (i in sample_size) {
  set.seed(2023)
```

```
  rej_rates_null_iid(N = i, r = 2000, delta = 0.5, theta = 1/4,
  m_pow = m_pow)
  set.seed(2023)
  rej_rates_null_weak_dependence(N = i, r = 2000, delta = 0.5,
  theta = 1/4, h_pow = h_pow, m_pow = m_pow)

}
```

The following code computes the empirical powers.

```
####################
# library packages #
####################
library(doParallel)
library(doRNG)

############################
# compute empirical powers #
############################

power_FLRD <- function(N, r = 2000, K = 500, H0, model, h_pow, m_pow)
{
  # This function computes the empirical power of the change
  point v.s. LRD test.
  # N = sample size, i.e. number of curves; r = number of
  replications;
  # K+1 = the number of observations in the discrete version of
  each curve;
  # model should be either 1 or 2 (integer). It specifies which
  disturbance function g is used.
  # h_pow specifies the bandwidth h via h = N^h_pow.
  # If h_pow equals 0, then the covariance function will be used
  in conducting FPCA.
  # m_pow specifies the number of Fourier frequencies m via
  m = N^m_pow.

  # Note:
```

```
# 1. Both h_pow and m_pow can be numeric vectors with elements
greater than 0 and less than 1.
# 2. The function returns a list which contains:
# (i) Test statistics for each combination of h and m (stored
in a 3D array).
# (ii) Rejection rates for each combination of h and m at different
nominal sizes (i.e. 0.01, 0.05 and 0.10).
# (iii) Change point estimates.
# (iv) The value(s) of h and m.

# determine whether the long-run covariance function is used
in FPCA
if (length(h_pow) == 1 && h_pow == 0){
  long_run_cov <- FALSE
}
else{
  long_run_cov <- TRUE
}

# compute h and m
h <- N^h_pow
m <- N^m_pow

# critical values for different nominal sizes
cv_01 <- qnorm(1-0.01)
cv_05 <- qnorm(1-0.05)
cv_10 <- qnorm(1-0.10)

# set up cores for parallel computing
```

```
cores <- detectCores()
cl <- parallel::makeCluster(cores[1]-2,
setup_strategy = "sequential")
registerDoParallel(cl)

PP <- foreach(b=1:r, .combine='rbind') %dorng% {

  library(Rcpp)
  # The change point estimation function is written in Rcpp to
  reduce the running time.
  # This is an alternative version in "FTS-library.R" (see Dhat()).
  sourceCpp("D_cpp1.cpp")

  source("FTS-library.R")

  # generate data
  if (model == 1){
    dat = FLM_lrd_1(K = K, N = N, H0 = H0)
  }
  else{
    dat = FLM_lrd_2(K = K, N = N, H0 = H0)
  }

  # estimate change-point
  nhat1 <- D_cpp(dat)$nhat

  # compute level-1 scores for different bandwidth h
  level_1_scores_mat <- matrix(nrow = N, ncol = length(h))
  for (i in 1:length(h)) {
```

```
    level_1_scores <- level1score(dat, nhat1, h[i], long_run_cov)
    level_1_scores_mat[,i] <- level_1_scores$level1
  }

  # compute the test statistics for each combination of h and m
  TS_mat <- matrix(nrow = length(h), ncol = length(m))
  for (i in 1:length(h)) {
    for (j in 1:length(m)) {
      # compute the fractional difference parameter d = H - 0.5
      diff_par <- lwe2(level_1_scores_mat[,i], m = m[j],
      lowerd = -0.4999, upperd = 0.4999)[[2]]
      TS_mat[i,j] <- 2 * sqrt(m[j]) * diff_par
    }
  }

  # convert the test statistics into a long vector for the
  sake of storage
  # concatenation is by row
  TS_vec <- as.vector(t(TS_mat))

  # return the test statistics and the change point estimate
  c(TS_vec, nhat1)
}

stopCluster(cl)

# extract test statistics and change point estimates
nhat_id <- dim(PP)[2]
TS_mat <- as.matrix(PP[,-nhat_id])
```

```
nhat <- as.matrix(PP[,nhat_id])

# rejection rate at alpha = 0.01
decision_01 <- ifelse(TS_mat > cv_01, 1, 0)
rej_01 <- colMeans(decision_01)

# rejection rate at alpha = 0.05
decision_05 <- ifelse(TS_mat > cv_05, 1, 0)
rej_05 <- colMeans(decision_05)

# rejection rate at alpha = 0.10
decision_10 <- ifelse(TS_mat > cv_10, 1, 0)
rej_10 <- colMeans(decision_10)

# create an output list
out <- list("TS" = NA,
            "rej_rate_0.01" = NA,
            "rej_rate_0.05" = NA,
            "rej_rate_0.10" = NA,
            "nhat" = NA,
            'm' = NA)

# store change point estimates
out$nhat <- nhat

# create matrices of rejection rates for each combination of
h and m at different nominal sizes
out$rej_rate_0.01 <- matrix(data = rej_01, nrow = length(h),
ncol = length(m), byrow = T)
```

```
out$rej_rate_0.05 <- matrix(data = rej_05, nrow = length(h),
ncol = length(m), byrow = T)
out$rej_rate_0.10 <- matrix(data = rej_10, nrow = length(h),
ncol = length(m), byrow = T)

# add row and column names
h_name <- character()
m_name <- character()
if (length(h_pow) == 1 && h_pow == 0){
  h_name <- ''
}
else{
  for (i in 1:length(h_pow)) {
    h_name[i] <- paste("h=N^", h_pow[i], sep = '')
  }
}
for (j in 1:length(m_pow)) {
  m_name[j] <- paste("m=N^", m_pow[j], sep = '')
}

for (k in 2:4){
  rownames(out[[k]]) <- h_name
  colnames(out[[k]]) <- m_name
}

# convert test statistics matrix into a 3D array
TS_array <- array(data = NA, dim = c(length(h), length(m), r),
dimnames = list(h_name, m_name))
for (k in 1:r) {
```

```
    TS_array[,,k] <- matrix(data = TS_mat[k,], nrow = length(h),
    ncol = length(m), byrow = T)
  }

  # store test statistics
  out$TS <- TS_array

  # store value(s) of h and m
  if (!(length(h_pow) == 1 && h_pow == 0)) {
    out$h <- h_name
  }
  out$m <- m_name

  return(out)

}

rej_rates_alt_iid <- function(N, r, H0, model, m_pow){
  out <- power_FLRD(N = N, r = r, H0 = H0, model = model,
  h_pow = 0, m_pow = m_pow)
  saveRDS(out, file = paste('alt', model, N, H0, 'iid=TRUE',
  '.rds',sep = '_' ))
}

rej_rates_alt_weak_dependence <- function(N, r, H0, model,
h_pow, m_pow){
  out <- power_FLRD(N = N, r = r, H0 = H0, model = model,
  h_pow = h_pow, m_pow = m_pow)
  saveRDS(out, file = paste('alt', model, N, H0, 'iid=FALSE',
```

```
  '.rds',sep = '_' ))
}

sample_size <- c(500, 1000, 1500)
h_pow <- c(0.2, 0.3, 0.4)
m_pow <- c(0.55, 0.6, 0.65)

#######################
# model = 1; H0 = 0.9 #
#######################
for (i in sample_size) {

  set.seed(2023)
  rej_rates_alt_iid(N = i, r = 2000, H0 = 0.9, model = 1,
  m_pow = m_pow)

  set.seed(2023)
  rej_rates_alt_weak_dependence(N = i, r = 2000, H0 = 0.9,
  model = 1, h_pow = h_pow, m_pow = m_pow)

}

#######################
# model = 1; H0 = 0.6 #
#######################
for (i in sample_size) {

  set.seed(2023)
  rej_rates_alt_iid(N = i, r = 2000, H0 = 0.6, model = 1,
```

```
  m_pow = m_pow)

  set.seed(2023)
  rej_rates_alt_weak_dependence(N = i, r = 2000, H0 = 0.6,
  model = 1, h_pow = h_pow, m_pow = m_pow)

}

######################
# model = 2; H0 = 0.9 #
######################
for (i in sample_size) {

  set.seed(2023)
  rej_rates_alt_iid(N = i, r = 2000, H0 = 0.9, model = 2,
  m_pow = m_pow)

  set.seed(2023)
  rej_rates_alt_weak_dependence(N = i, r = 2000, H0 = 0.9, model = 2,
  h_pow = h_pow, m_pow = m_pow)

}

######################
# model = 2; H0 = 0.6 #
######################
for (i in sample_size) {

  set.seed(2023)
```

```
  rej_rates_alt_iid(N = i, r = 2000, H0 = 0.6, model = 2,
  m_pow = m_pow)

  set.seed(2023)
  rej_rates_alt_weak_dependence(N = i, r = 2000, H0 = 0.6,
  model = 2, h_pow = h_pow, m_pow = m_pow)

}
```

The following code generates Figure 1.

```
####################
# library packages #
####################
library(tidyverse)
library(ggplot2)

g_bj_plot <- function(a) {
  p = 15
  x <- 1:a
  g1 <- rep(1,a)

  g2 <- sin(2*pi/p * (1:a))
  g2 <- exp(-0.01 * (1:a) - 1/2) * g2 + 1

  bj <- (1:a)^(0.9 - 3/2)

  g1 <- g1 * bj
  g2 <- g2 * bj

  df <- data.frame(x = x, g1 = g1, g2 = g2)
  df <- df %>% pivot_longer(cols = g1:g2, names_to = 'Function',
  values_to = 'y')
  df$Function <- as.factor(df$Function)
  df %>% ggplot(aes(x = x, y = y, linetype = Function)) +
    geom_line(alpha = 0.7) +
    coord_cartesian(ylim = c(0, 0.3)) +
    labs(x = "lag j", y = '') +
```

```
    scale_linetype_manual(labels = c(expression(g[1](j)%.%
    j^{H[0] - 3/2}),
    expression(g[2](j)%.%j^{H[0] - 3/2})), values =
    c("solid",  "dashed")) +
    theme_classic() +
    theme(legend.position = ’top’)

}

pdf(file="g_bj_plot.pdf", height = 3.5)
g_bj_plot(a = 300)
dev.off()
```

The following code conducts the hypothesis tests in Section 5.

```
#############
# load data #
#############

load("9ETFs.rdata")

########################################################
# transform the raw data to the CIDRs for given period #
########################################################

fCIDR <- function(rn, date) {
  # The function converts the raw CIDR data "rn" into a matrix with
  each row representing each minute of a trading day and
  # each column representing each day between the starting and
  ending dates.
  # rn is the raw data (e.g. xlb).
  # date is a numeric vector of length 2 with the first element being
  the starting date and the second being the ending date.
  # The format for the starting/ending date is YearMonthDay (e.g.
  20060705).

  # The function returns the following objects:
  # (1) A date vector for the CIDR curves.
  # (2) CIDR curves in a numeric matrix, with each curve stored in
  each column.
```

```
  D_b <- date[1]   # the starting date
  D_e <- date[2]   # the ending date

  # select data between the starting and ending dates
  fir <- which(rn$Date == D_b & rn$Time == 93100)
  las <- which(rn$Date == D_e & rn$Time == 160000)
  rn1 <- rn[fir:las,]

  # exclude some "abnormal" trading days
  rn.sum <- aggregate(Time ~ Date, rn1, sum)
  p.ind <- rn.sum$Date[rn.sum$Time == 48962500]
  rn.df <- subset(rn1, rn1$Date %in% p.ind)

  # create dates for each CIDR curve
  YM <- matrix(rn.df$Date, 390)
  checkdate = apply(YM, 2, unique)

  # create the matrix of CIDR curves
  RN <- matrix(rn.df$CIDR, 390)

  return(list('checkdate' = checkdate,
              'RN' = RN))
}


##################
# choose periods #
##################
```

```
# choose the starting and ending dates in the study
Pall <- c(20060705, 20111230)  # all 1378 days

##########################################
# change point test for the CIDRs/ACIDRs #
##########################################

CIDR_ACIDR_test <- function(data, period = Pall, h_pow, m_pow,
curve = 'CIDR'){
  # The function conducts the change point test described in the
  main body of the paper for the CIDRs/ACIDRs of a given sector.
  # The confidence level of the test is 95%.
  # data is the raw data of a given sector (e.g. xlb).
  # period is a numeric vector of length 2 with the first element
  being the starting date and the second being the ending date.
  # The format for the starting/ending date is YearMonthDay (e.g.
  20060705).
  # h_pow is a numeric value between 0 and 1 which specifies the
  bandwidth h via h = N^h_pow.
  # m_pow is a numeric value between 0 and 1 which specifies the
  number of Fourier frequencies m via m = N^m_pow.
  # curve is a character (either 'CIDR' or 'ACIDR') that specifies
  which type of curve will be used.
  # The function returns a list containing the test statistic,
  decision, the change point estimation and curve type.

  source("FTS-library.R")

  dat0 <- data
```

```
# convert the raw data into CIDR/ACIDR curves
OPT <- fCIDR(dat0, period)
if (curve == 'CIDR') {
  curves <- OPT[[2]]
}
else {
  curves <- abs(OPT[[2]])
}

# compute sample size
N <- dim(curves)[2]

# change point estimation
n_hat <- Dhat(curves)

# compute level-1 scores
level1_scores <- level1score(curves, n_hat$nhat,
                             h = N^h_pow, long_run_cov = TRUE)
xi_t1 <- level1_scores$level1

# critical value
cv_05 = qnorm(1-0.05)

# number of Fourier frequencies
m <- N^m_pow

# local Whittle estimation for the difference parameter
d_lwe2 <- lwe2(xi_t1, m = m, lowerd = -0.4999,
```

```
            upperd = 0.4999)[[2]]

  # compute test statistic
  TS_lwe2 <- 2 * sqrt(m) * (d_lwe2)

  # decision
  reject_lwe2 <- as.character(ifelse(TS_lwe2 > cv_05, TRUE, FALSE))

  return(list('TS_lwe2' = c(TS_lwe2, reject_lwe2),
              'n_hat' =  n_hat$nhat,
              'curve' = curve))
}

nine_ETFs_table <- function(period = Pall, h_pow, m_pow,
curve = 'CIDR'){

  # The function displays test results of CIDRs/ACIDRs for all
  nine sector ETFs.
  # The confidence level of the test is 95%.
  # period is a numeric vector of length 2 with the first element
  being the starting date and the second being the ending date.
  # The format for the starting/ending date is YearMonthDay (e.g.
  20060705).
  # h_pow is a numeric value between 0 and 1 which specifies the
  bandwidth h via h = N^h_pow.
  # m_pow is a numeric value between 0 and 1 which specifies the
  number of Fourier frequencies m via m = N^m_pow.
  # curve is a character (either 'CIDR' or 'ACIDR') specifying which
  type of curve will be used.
```

```
# The function returns a list containing the test results for all
nine sector ETFs, the choice of h and m and curve type.

out <- data.frame(sectors = c('xlb', 'xle', 'xlf', 'xli', 'xlk',
'xlp', 'xlu', 'xlv', 'xly'))

# conduct change point test for all nine sector ETFs
xlb_test <- CIDR_ACIDR_test(data = xlb, period = period,
h_pow = h_pow, m_pow = m_pow, curve = curve)
xle_test <- CIDR_ACIDR_test(data = xle, period = period,
h_pow = h_pow, m_pow = m_pow, curve = curve)
xlf_test <- CIDR_ACIDR_test(data = xlf, period = period,
h_pow = h_pow, m_pow = m_pow, curve = curve)
xli_test <- CIDR_ACIDR_test(data = xli, period = period,
h_pow = h_pow, m_pow = m_pow, curve = curve)
xlk_test <- CIDR_ACIDR_test(data = xlk, period = period,
h_pow = h_pow, m_pow = m_pow, curve = curve)
xlp_test <- CIDR_ACIDR_test(data = xlp, period = period,
h_pow = h_pow, m_pow = m_pow, curve = curve)
xlu_test <- CIDR_ACIDR_test(data = xlu, period = period,
h_pow = h_pow, m_pow = m_pow, curve = curve)
xlv_test <- CIDR_ACIDR_test(data = xlv, period = period,
h_pow = h_pow, m_pow = m_pow, curve = curve)
xly_test <- CIDR_ACIDR_test(data = xly, period = period,
h_pow = h_pow, m_pow = m_pow, curve = curve)

# extract decisions
Reject_lwe2 <- c(xlb_test[[1]][2], xle_test[[1]][2],
                 xlf_test[[1]][2], xli_test[[1]][2],
```

```
                    xlk_test[[1]][2], xlp_test[[1]][2],
                    xlu_test[[1]][2], xlv_test[[1]][2],
                    xly_test[[1]][2])

  # extract change point estimation
  n_hat <- c(xlb_test$n_hat, xle_test$n_hat, xlf_test$n_hat,
             xli_test$n_hat, xlk_test$n_hat, xlp_test$n_hat,
             xlu_test$n_hat, xlv_test$n_hat, xly_test$n_hat)

  out$Reject_lwe2 <- Reject_lwe2; out$n_hat <- n_hat

  return(list('out' = out,
              'h_pow' = h_pow,
              'm_pow' = m_pow,
              'curve' = curve))
}

hpow <- 0.3
mpow <- 0.6

nine_ETFs_table(h_pow = hpow, m_pow = mpow, curve = 'CIDR')
nine_ETFs_table(h_pow = hpow, m_pow = mpow, curve = 'ACIDR')
```

The following code generates Figures 2 and 3.

```
#################################
# load data and library package #
#################################

load("9ETFs.rdata")
library(tidyverse)

#################
# choose period #
#################

P32 <- c(20100802, 20111230)

#############################################
# plot five consecutive CIDRs of sector XLF #
#############################################

# The 5 consecutive days are 12/19 - 12/23 in year 2011.

dat <- xlf
OPT <- fCIDR(dat, P32)
OT <- OPT[[2]]




# obtain the five consecutive CIDR curves in a data frame format
fiveday_CIDR <- OT[,350:354]
```

```
CIDR_5_days <- data.frame(min = 1:dim(fiveday_CIDR)[1],
                          day1 = fiveday_CIDR[,1],
                          day2 = fiveday_CIDR[,2],
                          day3 = fiveday_CIDR[,3],
                          day4 = fiveday_CIDR[,4],
                          day5 = fiveday_CIDR[,5]) %>%
  pivot_longer(cols = day1:day5, names_to = 'Day',
  values_to = 'CIDR') %>%
  mutate(Day = factor(Day, levels = c('day1', 'day2', 'day3',
  'day4', 'day5'), labels = c('Day 1', 'Day 2', 'Day 3',
  'Day 4', 'Day 5')))

# plot the five CIDRs
p_CIDRs <- CIDR_5_days %>%
  ggplot(aes(x = min, y = CIDR)) +
  geom_line() +
  geom_hline(yintercept = 0, linetype="dotted") +
  labs(x = '', y = '') +
  facet_wrap(~ Day, ncol = 5) +
  theme_classic() +
  theme(axis.text.x = element_blank(),
               axis.ticks.x = element_blank(),
               axis.text.y = element_blank(),
               axis.ticks.y = element_blank())

pdf(file="5-day-CIDRs.pdf", width = 11, height = 3.5)
p_CIDRs
dev.off()
```

```
##############################################
# plot five consecutive CIDRs of sector XLF #
##############################################

# The 5 consecutive days are 12/19 - 12/23 in year 2011.

dat <- xlf
OPT <- fCIDR(dat, P32)
ACIDR <- abs(OPT[[2]])

# obtain the five consecutive ACIDR curves in a data frame format
fiveday_ACIDR <- ACIDR[,350:354]
ACIDR_5_days <- data.frame(min = 1:dim(fiveday_ACIDR)[1],
                          day1 = fiveday_ACIDR[,1],
                          day2 = fiveday_ACIDR[,2],
                          day3 = fiveday_ACIDR[,3],
                          day4 = fiveday_ACIDR[,4],
                          day5 = fiveday_ACIDR[,5]) %>%
  pivot_longer(cols = day1:day5, names_to = 'Day',
  values_to = 'ACIDR') %>%
  mutate(Day = factor(Day, levels = c('day1', 'day2', 'day3',
  'day4', 'day5'), labels = c('Day 1', 'Day 2', 'Day 3',
  'Day 4', 'Day 5')))

# plot the five ACIDRs
p_ACIDRs <- ACIDR_5_days %>%
  ggplot(aes(x = min, y = ACIDR)) +
  geom_line() +
```

```
  geom_hline(yintercept = 0, linetype="dotted") +
  labs(x = '', y = '') +
  facet_wrap(~ Day, ncol = 5) +
  theme_classic() +
  theme(axis.text.x = element_blank(),
        axis.ticks.x = element_blank(),
        axis.text.y = element_blank(),
        axis.ticks.y = element_blank())

pdf(file="5-day-ACIDRs.pdf", width = 11, height = 3.5)
p_ACIDRs
dev.off()
```

References

Aston, J. A. D. and Kirch, C. (2012a). Detecting and estimating epidemic changes in dependent functional data. *Journal of Multivariate Analysis*, **109,** 204–220.

Aston, J. A. D. and Kirch, C. (2012b). Evaluating stationarity via change–point alternatives with applications to fMRI data. *The Annals of Applied Statistics*, **6,** 1906–1948.

Aue, A., Gabrys, R., Horváth, L. and Kokoszka, P. (2009). Estimation of a change–point in the mean function of functional data. *Journal of Multivariate Analysis*, **100,** 2254–2269.

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

Baek, C. and Pipiras, V. (2012). Statistical tests for changes in mean against long-range dependence. *Journal of Time Series Analysis*, **33,** 131–151.

Beran, J., Feng, Y., Ghosh, S. and Kulik, R. (2013). *Long Memory Processes*. Princeton University Press.

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

Berkes, I., Horváth, L., Kokoszka, P. and Shao, Q-M. (2006). On discriminating between long–range dependence and changes in mean. *The Annals of Statistics*, **34,** 1140–1165.

Berkes, I., Horváth, L. and Rice, G. (2013). Weak invariance principles for sums of dependent random functions. *Stochastic Processes and their Applications*, **123,** 385–403.

Berkes, I., Horváth, L. and Rice, G. (2016). On the asymptotic normality of kernel estimators of the long run covariance of functional time series. *Journal of Multivariate Analysis*, **144,** 150–175.

Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer, New York.

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

Giraitis, L., Koul, H. and Surgailis, D. (2012). *Large Sample Inference for Long Memory Processes*. Imperial College Press.

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

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

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

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 (B)*, **75,** 103–122.

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

Horváth, L. and Rice, G. (2014). Extensions of some classical methods in change point analysis. *Test*, **23,** 213–225.

Horváth, L., Rice, G. and Whipple, S. (2016). Adaptive bandwidth selection in the long run covariance estimator of functional time series. *Computational Statistics and Data Analysis*, **100,** 676–693.

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

Jeffreys, H. (1961). *Theory of Probability*. Clarendon Press, Oxford.

Kokoszka, P., Miao, H., Stoev, S. and Zheng, B. (2018). Risk analysis of cumulative intraday return curves. *Journal of Time Series Econometrics*, **11**.

Kokoszka, P., Miao, H. and Zhang, X. (2015). Functional dynamic factor model for intraday price curves. *Journal of Financial Econometrics*, **13,** 456–477.

Kokoszka, P. and Reimherr, M. (2013). Predictability of shapes of intraday price curves. *The Econometrics Journal*, **16,** 285–308.

Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. CRC Press.

Li, D., Robinson, P. and Shang, HL (2020). Long-range dependent curve time series. *Journal of the American Statistical Association*, **115,** 957–971.

Li, D., Robinson, P. and Shang, HL (2021). Local Whittle estimation of long-range dependence for functional time series. *Journal of Time Series Analysis*, **42,** 685–695.

Pearson, Karl (1902). On the mathematical theory of errors of judgment, with special reference to the personal equation. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, **198,** 235–299.

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

Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, **30,** 1630–1661.

Robinson, P.M. (1995b). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, **23,** 1048–1072.

Shao, X. and Wu, W. B. (2007). Local Whittle estimation of fractional integration for nonlinear processes. *Econometric Theory*, **23,** 87–127.

Shi, J. Q. and Choi, T. (2011). *Gaussian Process Regression Analysis for Functional Data*. CRC Press.

Velasco, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis*, **20,** 87–127.