STATISTICAL INFRINGEMENT FOR DEPENDENT DATA

BY

XIANYANG ZHANG

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Statistics in the Graduate College of the University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:

Associate Professor Xiaofeng Shao, Chair
Professor Annie Qu
Professor Roger Koenker
Associate Professor Yuguo Chen
Abstract

Functional data Analysis has emerged as an important area of statistics which provides convenient and informative tool for the analysis of data objects of high dimension/high resolution. In the literature, it seems that the emphasis has been placed on independent functional data or models where the covariates and errors are assumed to be independent. However, the independence assumption is often too strong to be realistic in many application especially if the data are collected sequentially over time such as climate data and high frequency financial data. Motivated by our ongoing research on the development of high-resolution climate projections through statistical downscaling, we consider the change point problem and the two sample problem for temporally dependent functional data. Specifically, in Chapter 1, we develop a self-normalization based test to test the structural stability of temporally dependent functional observations. We propose new tests to detect the differences of the covariance operators and their associated characteristics of two functional time series in Chapter 2. The self-normalization approach introduced in the first two chapters is closely linked to the fixed-$b$ asymptotic scheme in the econometrics literature. Motivated by recent studies on heteroskedasticity and autocorrelation consistent based robust inference, we propose a class of estimators for estimating the asymptotic covariance matrix of the generalized method of moments estimator in the stationary time series models in Chapter 3. Under mild conditions, we establish the first order asymptotic distribution for the Wald statistics when the smoothing parameter is held fixed. Furthermore, we derive higher order Edgeworth expansions for the finite sample distribution of the Wald statistics in the Gaussian location model under the fixed-smoothing paradigm. The results are used to justify the second order correctness of a new bootstrap method, the Gaussian dependent bootstrap, in the context of Gaussian location model. Finally, in Chapter 4, we describe an extension of the fixed-$b$ approach to the empirical likelihood estimation framework.
To my parents.
Acknowledgments

I would like to express my heartfelt gratitude to my PhD advisor, Professor Xiaofeng Shao, for his support throughout my doctoral studies. This thesis would not have been possible without his constant guidance and encouragement during my graduate study. His advice and help are always invaluable on both an academic and a personal level, for which I am truly indebted and thankful.

I am grateful to Professor Annie Qu, Professor Roger Koenker and Professor Yuguo Chen for their insightful comments and for serving on my doctoral thesis committee. I thank all the faculty members of the statistics department for providing such a resourceful and friendly environment. Further thanks go to Professor Annie Qu for her guidance and support when I worked in the Illinois Statistics Office.

I was lucky to be surrounded by many friends, who have made my life in graduate school full of joy. I am obliged to many of my friends for their enthusiasm, encouragement, and inspiration.

Finally, I would like to thank my parents for their unconditional love and constant care. To them, I dedicate this dissertation.
# Table of Contents

List of Tables .................................................. vii
List of Figures .................................................... ix

Chapter 1  Introduction .......................................... 1

Chapter 2  Testing the structural stability of temporally dependent functional observations ........................................... 4
2.1 Introduction .................................................. 4
2.2 Methodology .................................................. 6
  2.2.1 Testing for a change point in the mean function ......... 6
  2.2.2 Testing for a change point in the lag-1 autocovariance operator ......... 10
2.3 Theoretical results ......................................... 11
2.4 Numerical studies .......................................... 13
  2.4.1 Detecting the mean change for curves .................. 14
  2.4.2 Detecting the changes in the lag-1 autocovariance operator .... 17
  2.4.3 Detecting the mean change for 2-d functional observations .... 18
2.5 Applications ............................................... 19
  2.5.1 Analysis of central England temperatures .............. 19
  2.5.2 Analysis of the bias between gridded observations and GCM simulations .... 20
2.6 Discussion and conclusion .................................. 21
2.7 Proofs of the main results .................................. 22

Chapter 3  Two sample inference for temporally dependent functional data ..................................................... 38
3.1 Introduction .................................................. 38
3.2 Methodology .................................................. 40
  3.2.1 Covariance operator ................................... 41
  3.2.2 Eigenvalues and eigenfunctions ......................... 42
3.3 Theoretical results ......................................... 45
3.4 Numerical studies .......................................... 48
  3.4.1 Comparison of covariance operators ................... 48
  3.4.2 Comparison of eigenvalues and eigenfunctions .......... 50
3.5 Climate projections analysis ............................... 51
3.6 Proofs of the main results .................................. 53

Chapter 4  Fixed-smoothing asymptotics for time series ............ 71
4.1 Introduction .................................................. 71
4.2 Basic setup and assumptions ................................ 74
4.3 LRV estimators .............................................. 76
4.4 First order fixed-smoothing asymptotics ................. 79
4.5 Higher order expansions .................................... 82
  4.5.1 Expansion for the finite sample distribution of subsampling-based t-statistic .... 82
2.1 Empirical size (upper panel) and size-adjusted power (lower panel) in percentage for the SN-based test (in row (i)) and the BGHK test (in row (ii)) for independent functional data generated from BM or BB. The size-adjusted power is computed under the alternative (2.2) with $\mu_2(t) = t$ or $\mu_2(t) = \sin(t)$, and $k^* = N/2$. The sample size $N = 50, 100$, and the number of PCs $K = 1, 2, 3$. The number of Monte Carlo replications is 1000. 28

2.2 Empirical size in percentage of the SN-based test (in row (i)), the BGHK test (in row (ii)) and the HK test (in row (iii)) for temporally dependent functional data generated from ARH(1) process. The sample size $N = 50, 100$, and the number of PCs $K = 1, 2, 3$. The number of Monte Carlo replications is 1000. 29

2.3 Size-adjusted power in percentage of the SN-based test (in row (i)), the BGHK test (in row (ii)) and the HK test (in row (iii)) for temporally dependent functional data generated from ARH(1) process. The size-adjusted power is computed under the alternative (2.2) with $\mu_2(t) = t$ or $\mu_2(t) = \sin(t)$, and $k^* = N/2$. The sample size $N = 50, 100$, and the number of PCs $K = 1, 2, 3$. The number of Monte Carlo replications is 1000. 30

2.4 Empirical size (upper panel) and size-adjusted power (lower panel) in percentage of the SN-based test (in row (i)) and the tests in Horváth et al. (2010) (in row (ii) for case 1 and (iii) for case 2 ) for detecting the change point in the lag one autocovariance operator. The size-adjusted power is computed under the alternative (2.20). The sample size $N = 50, 100$, and the number of PCs $K = 1, 2, 3$. The number of Monte Carlo replications is 1000. 31

2.5 (a) Empirical size (upper panel) and size-adjusted power (lower panel) in percentage of the SN-based test for detecting the mean change of 2-d functional observations generated by tensor product. (b) Empirical size (upper panel) and size-adjusted power (lower panel) in percentage of the SN-based test for detecting the mean change of 2-d functional observations generated from the ARH(2) process. The sample size $N = 50, 100$, and the number of PCs $K = 1, 2, 3$. The number of Monte Carlo replications is 1000. The notation $(BM + t)^2$ denotes $(BM + t) \times (BM + t)$. Note: The Brownian motion is approximated by the partial sum of 100 iid standard normal variables in (a) and it is approximated by the partial sum of 60 iid standard normal variables in (b). 32

2.6 Segmentation procedure of the central England temperature data into periods with constant mean function. Note: in each iteration, $K$ is the smallest positive integer such that $\sum_{i=1}^{K} \lambda_i / \sum_{i=1}^{12} \lambda_i > 0.8$. 33

2.7 Test statistics and their $p$-values for the spatio-temporal temperature data covering a subregion of North America. Panel (a) shows the results of the tests based on the first $K$ PCs; Panel (b) shows the results of the tests based on individual PC. 33
3.1 Empirical sizes and size-adjusted powers of (i) the SN-based test, (ii) the PKM test and (iii) the CLRV test for testing the equality of the covariance operators. The nominal level is 5%. ................................................................. 65

3.2 Empirical sizes and size-adjusted powers of (i) the SN-based test, the subsampling-based test with (ii) \( l = 8 \), (iii) \( l = 12 \) and (iv) \( l = 16 \), and (v) Benko et al’s iid bootstrap based method for testing the equality of the first two eigenvalues separately (the columns with \( M = 1, 2 \)) and jointly (the column with \( M = (1, 2) \)). The nominal level is 5% and the number of replications for iid bootstrap method is 250. ............. 66

3.3 Empirical sizes of (i) the SN-based test, the subsampling-based test with (ii) \( l = 8 \), (iii) \( l = 12 \) and (iv) \( l = 16 \), and (v) Benko et al’s iid bootstrap based method for testing the equality of the first two eigenfunctions separately (the columns with \( M = 1, 2 \)) and jointly (the column with \( M = (1, 2) \)). The nominal level is 5% and the number of replications for iid bootstrap is 250. ................................................................. 67

3.4 Comparison of the eigenvalues and eigenfunctions of the covariance operators of the station observations and model outputs. ................................................................. 67

4.1 Asymptotic comparison between the first and second order approximations based on fixed-\( b \) and small-\( b \) asymptotics. ................................................................. 90

5.1 Critical value function coefficients ................................................................. 130

5.2 Critical value function coefficients ................................................................. 130
List of Figures

2.1 Size-adjusted power for detecting the mean change with different magnitude of change. Sample size $N = 50$. Note: the quantity $\delta$ measures the magnitude of change; see equation (2.19). ................................................. 34

2.2 Average daily temperature functions in the three estimated partition segments. .... 35

2.3 The first six principal components of the bias between observations and model output. The number in the title of each figure stands for the percentage of variation explained by the corresponding principal component. .................................................. 36

2.4 Average biases in two periods (1980-1990; 1991-1998) and the change in the bias. .. 37

3.1 Size-adjusted powers of the SN-based test and the subsampling-based tests for testing the equality of the first two eigenfunctions separately and jointly. ......................... 68

3.2 $p$-values for testing the nullity of lag zero cross-correlation between the station observations and model outputs at each location. The numbers 0-5 denote the ranges of the $p$-values, i.e., 0 denotes $[0.1, 1]$; 1 denotes $[0.05, 0.1]$; 2 denotes $[0.025, 0.05]$; 3 denotes $[0.01, 0.025]$; 4 denotes $[0.005, 0.01]$ and 5 denotes $[0, 0.005]$. ................. 69

3.3 The first three PCs of the station observations (left panels) and model outputs (right panels), and the associated eigenvalues and percentage of variations explained. .......... 70

4.1 Simulated values of $\Psi(t_{K-1}(1-\alpha); K)/K$ based on 500,000 replications. .......... 84

4.2 Empirical rejection probabilities for the Wald statistic with the Bartlett kernel (left panel) and QS kernel (right panel) and for the AR(1) model with $N(0,1)$ innovations 122

4.3 Empirical rejection probabilities for the Wald statistic with the Bartlett kernel (left panel) and QS kernel (right panel) and for the AR(1) model with $t(3)$ innovations 123

4.4 Empirical rejection probabilities for the Wald statistic with the Bartlett kernel (left panel) and QS kernel (right panel) and for the AR(1) model with $exp(1) - 1$ innovations 124

5.1 Coverage probabilities for the mean delivered by the BEL based on the fixed-$b$ approximation and the $\chi^2$-based approximation. The nominal level is 95% and the number of Monte Carlo replications is 1000. .................................................. 138

5.2 Coverage probabilities for the median and 5% quantile delivered by the BEL based on the fixed-$b$ approximation and the $\chi^2$-based approximation. The nominal level is 95% and the number of Monte Carlo replications is 1000. ......................... 139

5.3 Coverage probabilities for the mean delivered by the smoothed EL based on the fixed-$b$ approximation and the $\chi^2$-based approximation. Note that the corresponding kernel is $K(x) = \left(\frac{5\pi}{8}\right)^{1/2} \frac{1}{2}I_1 \left(\frac{6\pi x}{5}\right)$. The nominal level is 95% and the number of Monte Carlo replications is 1000. .................................................. 140

5.4 Coverage probabilities for the median and 5% quantile delivered by the smoothed EL based on the fixed-$b$ approximation and the $\chi^2$-based approximation. Note that the corresponding kernel is $K(x) = \left(\frac{5\pi}{8}\right)^{1/2} \frac{1}{2}I_1 \left(\frac{6\pi x}{5}\right)$. The nominal level is 95% and the number of Monte Carlo replications is 1000. ......................... 141
5.5 Coverage probabilities delivered by the BEL (left panels) and the smoothed EL (right panels) based on the fixed-\(b\) approximation and the \(\chi^2\)-based approximation. Note that the corresponding kernel for the smoothed EL is \(K(x) = \left(\frac{5\pi}{8}\right)^{1/2} \frac{1}{\pi} J_1 \left(\frac{6\pi x}{5}\right)\). The nominal level is 95\% and the number of Monte Carlo replications is 1000.
Chapter 1

Introduction

Functional data Analysis (FDA) has emerged as an important area of statistics which provides convenient and informative tool for the analysis of data objects of high dimension/high resolution. The excellent monograph by Ramsay and Silverman (2005) provides a systematic account of the existing methodologies and tools to deal with data of functional nature. In the literature, it seems that the emphasis has been placed on independent functional data or models where the covariates and errors are assumed to be independent. However, in real life this assumption is often too strong to be realistic in many application especially if the data are collected sequentially over time such as climate data and high frequency financial data. The analysis of dependent functional data is challenging, as existing methods and tools developed for independent functional data may not be applicable and dependence brings a lot of complication and difficulties to the problem. It is thus important to develop robust procedures which can accommodate the dependence within the functional observations. Specially, we shall consider the change point problem and the two sample problem for temporally dependent functional data. In Chapter 1, we develop a self-normalization (SN) based test to test the structural stability of temporally dependent functional observations. Testing for a change point in the mean of functional data has been studied in Berkes et al. (2009), but their test was developed under the independence assumption. Building on the SN-based change point test proposed in Shao and Zhang (2010) for a univariate time series, we extend the SN-based test to the functional setup by testing the constant mean of the finite dimensional eigenvectors after performing functional principal component analysis. Asymptotic theories are derived under both the null and local alternatives. Through theory and extensive simulations, our SN-based test statistic proposed in the functional setting is shown to inherit some useful properties in the univariate setup: the test is asymptotically distribution free and its power is monotonic. Furthermore, we extend the SN-based test to identify potential change points in the dependence structure of functional observations. The method is then applied to central England temperature series to detect the warming trend and to gridded temperature fields generated by global climate models to test for changes in spatial bias.
structure over time.

Motivated by the need to statistically quantify the difference between two spatiotemporal datasets that arise in climate downscaling studies, we propose new tests to detect the differences of the covariance operators and their associated characteristics of two functional time series in Chapter 2. Our two sample tests are constructed on the basis of functional principal component analysis and self-normalization, the latter of which is a new studentization technique recently developed for the inference of a univariate time series. Compared to the existing tests, our SN-based tests allow for weak dependence within each sample and it is robust to the dependence between the two samples in the case of equal sample sizes. Asymptotic properties of the SN-based test statistics are derived under both the null and local alternatives. Through extensive simulations, our SN-based tests are shown to outperform existing alternatives in size and their powers are found to be respectable. The tests are then applied to the gridded climate model outputs and interpolated observations to detect the difference in their spatial dynamics.

The self-normalization approach introduced in the first two chapters is closely linked to the fixed-$b$ asymptotic scheme considered by Kiefer, Vogelsang and their co-authors (e.g., Kiefer et al. 2000; Kiefer and Vogelsang 2005). Motivated by recent studies on heteroskedasticity and autocorrelation consistent (HAC) based robust inference, we propose a class of estimators for estimating the asymptotic covariance matrix of the generalized method of moments estimator in the stationary time series models in Chapter 3. Our proposal provides a unification of the existing smoothing parameter dependent covariance estimators, including the traditional heteroskedasticity and autocorrelation consistent covariance estimator and some recently developed estimators, such as cluster-based covariance estimator and projection-based covariance estimator. Under mild conditions, we establish the first order asymptotic distribution for the Wald statistics when the smoothing parameter is held fixed. Furthermore, we derive higher order Edgeworth expansions for the finite sample distribution of the Wald statistics in the Gaussian location model under the fixed-smoothing paradigm. In particular, we show that the error of asymptotic approximation is at the order of the reciprocal of the sample size and obtain explicit forms for the leading error terms in the expansions. The results are used to justify the second order correctness of a new bootstrap method, the Gaussian dependent bootstrap, in the context of Gaussian location model. Some simulation results are also presented to corroborate our theoretical findings.

Finally, in Chapter 4, we describe an extension of the fixed-$b$ approach introduced by Kiefer and Vogelsang (2005) to the empirical likelihood estimation framework. Under the fixed-$b$ asymptotics,
the empirical likelihood ratio statistic evaluated at the true parameter converges to a nonstandard yet pivotal limiting distribution which can be approximated numerically. The impact of the bandwidth parameter and kernel choice is reflected in the fixed-$b$ limiting distribution. Compared to the $\chi^2$-based inference procedure used by Kitamura (1997) and Smith (2011), the fixed-$b$ approach provides a better approximation to the finite sample distribution of the empirical likelihood ratio statistic; Correspondingly, as shown in our simulation studies, the confidence region based on the fixed-$b$ approach has more accurate coverage than the traditional counterpart.
Chapter 2

Testing the structural stability of temporally dependent functional observations

2.1 Introduction

Major advances in technology are enabling data collection at increasingly high resolution. These advancements challenge state-of-the-art models and methods in statistics. It has long been recognized that functional data analysis (FDA), which deals with the analysis of curves and surfaces, is an effective tool for analyzing large high resolution data sets. Systematic methods and theory have been developed for FDA mainly under the independence assumption (Ramsay and Silverman, 2002, 2005; Ferraty and Vieu, 2006), with relatively little attention paid to the analysis of dependent functional data. However, for functional data observed over time, the independence assumption is rarely satisfied in practice. This chapter aims to develop new tests to assess the structural stability of temporally dependent functional data. Our work is partially motivated by our ongoing research on the development of high-resolution climate projections through statistical downscaling, which by definition assumes a temporally stable relationship between observations and climate models. Climate change is one of the most urgent problems facing the world this century. To study climate change, scientists have relied primarily on climate projections from global/regional climate models, which are deterministic numerical models that involve systems of differential equations and produce outputs at a prespecified grid. As numerical model outputs are widely used in situations where real observations are not available, it is an important but still open question whether the properties of numerical model outputs remain stable relative to real observations over time.

To assess the structural stability, we view functional observations as a realization from a functional time series process, and test for a change point in the mean and autocovariance of the functional time series. The detection of one or multiple change points in the first or second order structure of a functional time series is itself an important problem, as failure to account for such change points could lead to invalid inference. There is a large literature and long history on change point testing in scalar or vector time series (see Csörgo and Horváth, 1997; Perron, 2006, and references therein),
but research on change point testing for functional data is very recent. Berkes et al. (2009) proposed a CUSUM-based (cumulative sum) approach to test the assumption of a common functional mean for independent functional data. Berkes et al.’s test (BGHK, hereafter) is invalid for functional time series since it does not take the temporal dependence into account. Hörmann and Kokoszka (2010) recognized the limitation of the BGHK test and modified their test by introducing a consistent long run variance (LRV) estimator. Hörmann and Kokoszka’s work has been extended recently by Aston and Kirch (2011) for weak dependent functional data with a wide class of dependence structure and two types of alternatives, namely, at most one change point and epidemic changes. However, there is a bandwidth parameter involved in both Hörmann and Kokoszka’s test (HK, hereafter) and Aston and Kirch’s test. Its selection is not addressed and the finite sample performance of their test has not been examined. To avoid choosing the bandwidth parameter, Shao and Zhang (2010) proposed a self-normalization (SN, hereafter) based test in the univariate time series setup, where an inconsistent normalization matrix is introduced to accommodate the dependence. The idea of using inconsistent normalization is not new, as it has been previously applied by Lobato (2001) and Shao (2010) to the inference in univariate time series. In this chapter, we extend the SN-based test in the univariate setup to test the structural stability of temporally dependent functional data. To our knowledge, this is the first attempt to generalize the SN idea to inference problems for functional data.

The extension of the SN concept to the functional setup is nontrivial since functional observations are collected on a space of infinite dimension and traditional methods developed for univariate/multivariate time series are not applicable in this case. To circumvent this difficulty, our method relies on the functional principal component analysis (PCA) which projects the functional data onto a space spanned by the first few principal components (PC’s). The SN-based test statistic is constructed based on the principal component scores. To accommodate the dependence, we introduce a normalization matrix which is built by taking the single change point alternative into account. The normalization matrix is inconsistent but proportional to the unknown LRV matrix, which is canceled out in the limiting null distribution of the SN-based test. The proposed test is thus asymptotically pivotal with critical values tabulated in Shao and Zhang (2010). Compared to the methods in previous studies, the SN-based test is asymptotically distribution-free and is shown to enjoy the monotonic power property in the functional setup. In addition, the SN-based test can be easily extended to detect multiple change points in the mean function and to detect a change point in the lag one autocovariance operator, the latter of which is investigated in this chapter.
To illustrate this method, the SN-based tests are used to examine the stationarity of biases in simulated spatio-temporal temperature data over a subregion of North America which consists of two sequences of functional surfaces (with spatial resolution $87 \times 35$) based on station observations and historical simulations from global climate models. Since spatially distributed temperature fields are usually viewed as smooth images by scientists, FDA is an appropriate tool for analyzing and revealing key characteristics of such a large dataset. Statistical analysis based on the SN-based test is shown to be helpful in addressing the scientific question of whether the bias between observations and model output remains stable over time.

2.2 Methodology

Mathematically, we consider functional observations $X_i(t)$, $t \in \mathcal{I}$, $i = 1, 2, \ldots, N$ defined on some compact set $\mathcal{I}$ of the Euclidian space, where $\mathcal{I}$ could be one dimensional (e.g. a curve) or multidimensional (e.g. a surface or manifold). For simplicity, we consider the Hilbert space $\mathbb{H}$ of square integrable functions defined on $\mathcal{I} = [0, 1]$ (and $\mathcal{I}^2 = [0, 1]^2$). For any $f, g \in \mathbb{H}$, the inner product between $f$ and $g$ is defined as $< f, g > = \int_{\mathcal{I}} f(t)g(t)dt$. We denote $\| \cdot \|$ as the corresponding norm, i.e., $\|f\| = < f, f >^{1/2}$. Assuming the random elements all come from the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$, we let $L^p$ be the space of real valued random variables with finite $L^p$ norm, i.e., $(\mathbb{E}|X|^p)^{1/p} < \infty$. We further define $L^p_{\mathcal{I}}$ as the space of $\mathbb{H}$-valued random variables $X$ such that $e_p(X) := (\mathbb{E}||X||^p)^{1/p} < \infty$. We then let $D[0, 1]$ be the space of functions on $[0, 1]$ which are right-continuous and have left limits, endowed with the Skorokhod topology (see Billingsley, 1999). Weak convergence in $D[0, 1]$ or more generally in the $\mathbb{R}^d$-valued function space $D^d[0, 1]$ is denoted by “$\Rightarrow$”, where $d \in \mathbb{N}$. Finally “$\rightarrow d$” denotes convergence in distribution.

2.2.1 Testing for a change point in the mean function

Given the functional observations $\{X_i(t)\}_{i=1}^N$, we are interested in testing whether the mean function remains constant over time, i.e.,

$$H_{0,1} : \mathbb{E}[X_1(t)] = \mathbb{E}[X_2(t)] = \cdots = \mathbb{E}[X_N(t)], \quad t \in \mathcal{I}. \quad (2.1)$$
Under the null, we can write $X_i(t) = \mu_1(t) + Y_i(t)$ with $E[Y_i(t)] = 0$, $i = 1, 2, \ldots, N$. Under the alternative $H_{a,1}$, we assume there is a change point in the mean function, i.e.,

$$X_i(t) = \begin{cases} 
\mu_1(t) + Y_i(t) & 1 \leq i \leq k^*; \\
\mu_2(t) + Y_i(t) & k^* < i \leq N,
\end{cases} \quad (2.2)$$

where $k^* = \lfloor N\lambda \rfloor$ is an unknown change point for some $\lambda \in (0, 1)$, $\{Y_i(t)\}$ is a zero-mean functional sequence, and $\mu_1(t) \neq \mu_2(t)$ for some $t$. To describe our methodology, we first introduce some useful notation commonly adopted in the literature of functional data; see e.g. Berkes et al. (2009). For $X(\cdot) \in L^p_H$ with $p \geq 2$, we define $c(t, s) = \text{cov}\{X(t), X(s)\}$, $t, s \in I$ as the covariance function. By Mercer’s Lemma (Riesz and Sz-Nagy, 1955), $c(t, s)$ admits the spectral decomposition,

$$c(t, s) = \sum_{j=1}^{\infty} \lambda_j \phi_j(s) \phi_j(t), \quad (2.3)$$

where $\lambda_j$ and $\phi_j$ are the eigenvalue and eigenfunction respectively. The eigenvalues are ordered so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$. Based on the Karhunen-Loève expansion (Bosq, 2000), we have $X_i(t) = E[X_i(t)] + \sum_{j=1}^{\infty} \eta_{i,j} \phi_j(t)$, where $\{\eta_{i,j}\}$ are the principal components (scores) defined by $\eta_{i,j} = \int_I \{X_i(t) - E[X_i(t)]\} \phi_j(t)dt$. A natural estimator of the covariance function $c(t, s)$ is

$$\hat{c}(t, s) = \frac{1}{N} \sum_{i=1}^{N} \{X_i(t) - \bar{X}_N(t)\} \{X_i(s) - \bar{X}_N(s)\}, \quad (2.4)$$

where $\bar{X}_N(t) = \frac{1}{N} \sum_{i=1}^{N} X_i(t)$ is the sample mean function. The eigenfunctions and corresponding eigenvalues of $\hat{c}(t, s)$ are defined by

$$\int_I \hat{c}(t, s) \hat{\phi}_j(s) ds = \hat{\lambda}_j \hat{\phi}_j(t). \quad (2.5)$$

Then the empirical scores are given by

$$\hat{\eta}_{i,j} = \int_I \{X_i(t) - \bar{X}_N(t)\} \hat{\phi}_j(t) dt, \quad i = 1, 2, \ldots, N; \quad j = 1, 2, \ldots, K,$n

where $K$ is the number of principal components we consider and is assumed to be fixed throughout. Under the null, the score vector $\eta_i = (\eta_{i,1}, \eta_{i,2}, \ldots, \eta_{i,K})'$, $i = 1, 2, \ldots, N$ has a constant mean, whereas the mean changes under the alternative. If we let $\hat{\eta} = (\hat{\eta}_{11}, \ldots, \hat{\eta}_{1K})'$ and $S_{N,\hat{\eta}}(t_1, t_2) = \int_{t_1}^{t_2} \hat{\eta}(s) ds$,
\[ \sum_{t=t_1}^{t_2} \hat{\eta}_t, \text{ for } 1 \leq t_1 \leq t_2 \leq N, \] we can then define the so-called CUSUM process as
\[ T_{N,\hat{\eta}}(k, K) := \frac{1}{\sqrt{N}} \left\{ S_{N,\hat{\eta}}(1, k) - \frac{k}{N} S_{N,\hat{\eta}}(1, N) \right\}, \quad k = 1, 2, \ldots, N. \tag{2.6} \]

To test the assumption of a common functional mean for independent and identically distributed (iid) functional data, Berkes et al. (2009) introduced a CUSUM-based test statistic which takes the following form
\[ H_{N,\hat{\eta}}(K) := \frac{1}{N^2} \sum_{j=1}^{K} \lambda_j^{-1} \sum_{k=1}^{N} \left( \sum_{i=1}^{k} \hat{\eta}_{i,j} - \frac{k}{N} \sum_{i=1}^{N} \hat{\eta}_{i,j} \right)^2. \tag{2.7} \]

It can be rewritten as
\[ H_{N,\hat{\eta}}(K) = \frac{1}{N} \sum_{k=1}^{N} T_{N,\hat{\eta}}(k, K)' \Sigma_{\eta}^{-1} T_{N,\hat{\eta}}(k, K), \tag{2.8} \]
where \( \Sigma_{\eta} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_K) \). As pointed out by Hörmann and Kokoszka (2010), the BGHK test is not applicable to functional time series because it does not take the temporal dependence of \( \eta_t \)'s into account. In the dependent case, one usually needs to estimate the LRV matrix (i.e., the spectral density function evaluated at zero frequency) of \( \eta_t \) consistently. The commonly-used lag window type estimator can be used to obtain a consistent LRV estimator \( \hat{\Sigma}_{\eta} \) (see Hörmann and Kokoszka, 2010). A test statistic can then be constructed by applying certain continuous functional \( G \) to the normalized CUSUM process \( T_{N,\hat{\eta}}([Nr], K)' \Sigma^{-1}_{\eta} T_{N,\hat{\eta}}([Nr], K), \) \( r \in [0, 1] \). In the iid case, the LRV matrix of \( \eta_t \) is simply given by \( \Sigma_{\eta} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_K) \), which can be consistently estimated by replacing each eigenvalue with its empirical estimate. From equation (2.8), it is easy to see that the BGHK test is basically a special case of this procedure with \( G(f) = \int_{I} |f(x)| dx \). For temporally-dependent functional data, the HK test statistic is asymptotically valid, but it involves a truncation lag (bandwidth parameter) in the LRV estimator, the selection of which is not addressed. In fact, the choice of the bandwidth is a difficult task in the detection problem even in the univariate setup. The bandwidth that is a fixed function of the sample size (e.g., \( N^{1/3} \), where \( N \) is the sample size) is not adaptive to the magnitude of the dependence in the series, whereas the data-dependent bandwidth could lead to nonmonotonic power (i.e., the power can decrease when the alternative gets farther away from the null) as shown in previous studies (Vogelsang, 1999; Crainiceanu and Vogelsang, 2007; Juhl and Xiao, 2009). Recently, Shao and Zhang (2010) proposed a SN-based test in the univariate time series setup, that is able to overcome the nonmonotonic power problem and has very accurate size and respectable power properties. Here we pursue an extension of the SN-based test to the functional setup.
To avoid choosing the bandwidth parameter, we define the following normalization matrix which takes the alternative into account. Let
\[
V_{N,\hat{\eta}}(k, K) := \frac{1}{N^2} \sum_{t=1}^{k} \left\{ S_{N,\hat{\eta}}(1, t) - \frac{t}{k} S_{N,\hat{\eta}}(1, k) \right\}' \left\{ S_{N,\hat{\eta}}(1, t) - \frac{t}{k} S_{N,\hat{\eta}}(1, k) \right\}
\]

\[+ \sum_{t=k+1}^{N} \left\{ S_{N,\hat{\eta}}(t, N) - \frac{N-t+1}{N-k} S_{N,\hat{\eta}}(k+1, N) \right\}' \left\{ S_{N,\hat{\eta}}(t, N) - \frac{N-t+1}{N-k} S_{N,\hat{\eta}}(k+1, N) \right\].\tag{2.9}\]

The SN-based statistic can thus be defined as
\[
G_{N,\hat{\eta}}(K) = \sup_{k=1, 2, \ldots, N-1} \left\{ T_{N,\hat{\eta}}(k, K)' V_{N,\hat{\eta}}^{-1}(k, K) T_{N,\hat{\eta}}(k, K) \right\} \tag{2.10}
\]
where \( C \) is the implicitly defined continuous mapping that corresponds to \( G_{N,\hat{\eta}}(K) \). Here the self-normalizer \( \{ V_{N,\hat{\eta}}(k, K) \}_{k=1}^{N} \) plays two roles. On the one hand, it is able to absorb the dependence without consistent estimation of the LRV matrix. This means that the resulting limiting null distribution is nuisance parameter-free. On the other hand, it is specially designed for the change point testing problem, and it has been shown very effective in eliminating the nonmonotonic power problem in the univariate time series setting by Shao and Zhang (2010). Though there is no theoretical justification of the monotonic power property of the SN-based test even in the univariate setting, the empirical power of the SN-based test is seen to be monotonic in our simulation studies (see Section 4.1). Let \( B_K(t) \) be a \( K \) dimensional vector with each component an independent standard Brownian motion. Under suitable assumptions (see Section 3), we are able to show that
\[
G_{N,\hat{\eta}}(K) \sim d G(K) := \sup_{r \in [0, 1]} \left\{ B_K(r) - r B_K(1) \right\}' V_K^{-1}(r) \left\{ B_K(r) - r B_K(1) \right\}, \tag{2.11}
\]
where \( V_K(r) = \int_{0}^{r} W_{1,K}(s, r) W_{1,K}(s, r)' ds + \int_{r}^{1} W_{2,K}(s, r) W_{2,K}(s, r)' ds \) with \( W_{1,K}(s, r) = B_K(s) - B_K(r) s/r \) for \( s \in [0, r] \) and \( W_{2,K}(s, r) = \{(B_K(1) - B_K(s)) - (1-s)/(1-r)(B_K(1) - B_K(r))\} \) for \( s \in [r, 1] \). The critical values of the nonstandard null distribution \( G(K) \) have been tabulated by Shao and Zhang (2010) for \( K = 1, 2, \ldots, 10 \) via simulations.
2.2.2 Testing for a change point in the lag-1 autocovariance operator

As an extension of the above SN-based test, we consider the problem of testing the stability of the autocovariance operator at lag one, which partially describes the dependence structure of temporally dependent functional data. Recently, Horváth et al. (2010) proposed a test for the constancy of the ARH(1) (functional autoregressive model of order one) operator against a one change point alternative. As pointed out in their paper, since the constancy of the ARH(1) operator implies the stability of the autocovariance operator at lag one, their test effectively checks whether the lag one autocovariance operator stays constant over time. Our test differs from theirs in two aspects. First, we do not assume a parametric ARH(1) model in our theory and our test can be easily extended to test for the constancy of lag $k$ autocovariance operator for $k = 1, 2, \ldots, m$, either separately or jointly. Second, our SN-based test is free of any bandwidth parameter, which is required in Horváth et al.’s work.

Without loss of generality, we assume that the functional observations $\{X_i(t)\}_{i=1}^N$ have a constant mean zero and admit the Karhunen-Loève expansion, $X_i(t) = \sum_{j=1}^{\infty} \eta_{i,j} \phi_j(t)$, $i = 1, 2, \ldots, N$. Let $R_i(\cdot) = \mathbb{E}[<X_i, \cdot > X_{i+1}]$ be the lag one autocovariance operator at time $i$. We are interested in testing the null hypothesis that

$$H_{0,2}: R_1 = \cdots = R_{N-1}$$

versus the alternative

$$H_{a,2}: R_1 = \cdots = R_{\hat{k}^*} \neq R_{\hat{k}^*+1} = \cdots = R_{N-1},$$

where the change point $\hat{k}^* = \lfloor N \hat{\lambda} \rfloor$ with $\hat{\lambda} \in (0,1)$ and $\hat{k}^* < N - 1$ is unknown. Following Horváth et al. (2010), we focus on the action of the lag one autocovariance operator $R_i$ on the space spanned by $\{\phi_1(t), \phi_2(t), \ldots, \phi_K(t)\}$, and we test the constancy of $\{R_i \phi_j, j = 1, 2, \ldots, K\}$. Based on the representation that $R_i \phi_j = \sum_{l=1}^{\infty} < R_l \phi_j, \phi_l >$, the constancy of $R_i$ implies the stability of $< R_l \phi_j, \phi_l >$, $j, l = 1, 2, \ldots, K$, which motivates us to test the stability of the vector $< R_l \phi_1, \phi_1 >, \ldots, < R_l \phi_1, \phi_K >, \ldots, < R_l \phi_K, \phi_1 >, \ldots, < R_l \phi_K, \phi_K > \in \mathbb{R}^{K^2}$, for $i = 1, 2, \ldots, N - 1$. Under $H_{0,2}$, we note that $< R_i \phi_j, \phi_l > = \mathbb{E}[< X_i, \phi_j > < X_{i+1}, \phi_l >] = \mathbb{E}[\eta_{i,j} \eta_{i+1,l}]$. Defining $\xi_i(j, l) = \eta_{i,j} \eta_{i+1,l}$ and its empirical counterpart by $\hat{\xi}_i(j, l) = \hat{\eta}_{i,j} \hat{\eta}_{i+1,l}$. We further define the vector $\xi_i = (\xi_i(1,1), \ldots, \xi_i(1,K), \ldots, \xi_i(K,1), \ldots, \xi_i(K,K))' \in \mathbb{R}^{K^2}$ and its sample counterpart $\hat{\xi}_i = (\hat{\xi}_i(1,1), \ldots, \hat{\xi}_i(1,K), \ldots, \hat{\xi}_i(K,1), \ldots, \hat{\xi}_i(K,K))'$. We aim to test the mean change of the vector $\xi_i$ based on $\hat{\xi}_i$, $i = 1, 2, \ldots, N - 1$. We define the empirical partial sum process by $S_{N,\xi}(t_1, t_2) = \sum_{t_i = t_1}^{t_2} \hat{\xi}_i$. Analogous to $T_{N,\hat{\theta}}(k,K)$ and $V_{N,\hat{\theta}}(k,K)$, we define $T_{N,\hat{\xi}}(k,K^2)$ and $V_{N,\hat{\xi}}(k,K^2)$ with
$S_{N,0}(t_1,t_2)$ replaced by $S_{N,\xi}(t_1,t_2)$. The test statistic is then given by

$$G_{N,\xi}(K^2) = \sup_{k=1,2,\ldots,N-1} \{ T_{N,\xi}(k,K^2) V_{N,\xi}^{-1}(k,K^2) T_{N,\xi}(k,K^2) \}. \quad (2.12)$$

We will show that $G_{N,\xi}(K^2)$ has the limiting null distribution $G(K^2)$ in the next section.

### 2.3 Theoretical results

In this section, we justify the asymptotic validity of the SN-based test statistic by studying its asymptotic properties under both the null and local alternatives. To this end, we adopt the dependence measure proposed in Hörmann and Kokoszka (2010), which is applicable to the temporally-dependent functional process.

**Definition 2.3.1.** Assume that $\{X_i\} \in L^p_H$ with $p > 0$ admits the following representation

$$X_i = f(\varepsilon_i, \varepsilon_{i-1}, \ldots), \quad i = 1, 2, \ldots, \quad (2.13)$$

where the $\varepsilon_i$’s are iid elements taking values in a measurable space $S$ and $f$ is a measurable function $f : S^\infty \rightarrow H$. For each $i \in \mathbb{N}$, let $\{\varepsilon_j^{(i)}\}_{j \in \mathbb{Z}}$ be an independent copy of $\{\varepsilon_j\}_{j \in \mathbb{Z}}$. The sequence $\{X_i\}$ is said to be $L^p$-$m$-approximable if

$$\sum_{m=1}^{\infty} e_p(X_m - X_m^{(m)}) < \infty, \quad (2.14)$$

where

$$X_i^{(m)} = f(\varepsilon_i, \varepsilon_{i-1}, \ldots, \varepsilon_{i-m+1}, \varepsilon_{i-m}^{(i)}, \varepsilon_{i-m-1}^{(i)}, \ldots). \quad (2.15)$$

It can be verified that a functional linear process is $L^p$-$m$-approximable when the operator coefficients satisfy certain summability conditions and the innovation sequence is in $L^p_H$ (see Proposition 2.1 in Hörmann and Kokoszka, 2010). Definition 3.3.1 is also applicable to other nonlinear functional time series models such as functional bilinear models and functional ARCH models (see Examples 2.3 and 2.4 in Hörmann and Kokoszka, 2010). For the temporally dependent functional data, the PC’s are temporally correlated. We denote by $\Sigma_{\eta,K}$ the LRV matrix of the first $K$ PC’s, i.e,

$$\Sigma_{\eta,K} = \sum_{h=-\infty}^{\infty} \mathbf{E}[\eta_0 \eta_h'] ,$$

with $\eta_i = (\eta_{i,1}, \eta_{i,2}, \ldots, \eta_{i,K})' \in \mathbb{R}^K$. Similarly we can define the LRV matrix $\Sigma_{\xi,K^2}$ for $\{\xi_i\}$. To
derive the asymptotic properties of the SN-based tests, we make the following assumption.

**Assumption 2.3.1.** Assume that \( Y_i(t) := X_i(t) - E[X_i(t)] \in L^p \) are \( L^p \)-approximable mean zero random elements. The eigenvalues of \( c(t,s) \) satisfy that \( \lambda_1 > \lambda_2 > \cdots > \lambda_K > \lambda_{K+1} > 0 \).

**Theorem 2.3.1.** Suppose that \( E[X_i(t)] \in L^2 \) and Assumption 2.3.1 holds with \( p = 4 \). Assume that \( \Sigma_{\eta,K} \) is positive definite. Then (2.11) holds under \( H_0 \).

With a similar argument, we have the following result for \( G_{N,\hat{\xi}}(K^2) \) under slightly stronger assumptions.

**Theorem 2.3.2.** Suppose that Assumption 2.3.1 holds with \( p = 8 \). Also assume that \( E[X_i(t)] = 0 \) and \( \Sigma_{\xi,K^2} \) is positive definite. Then under \( H_{0,2} \), we have

\[
G_{N,\hat{\xi}}(K^2) \to^d G(K^2). \tag{2.16}
\]

We now turn to the consistency of the proposed tests. As mentioned before, we consider the one-time shift alternative that the mean function or the lag one autocovariance operator remains constant before the change point and then becomes another constant afterward. In the case of detecting the mean change, we consider the alternative (2.2). Let

\[
\tilde{c}(t,s) = c(t,s) + \lambda(1 - \lambda)\{\mu_1(t) - \mu_2(t)\}\{\mu_1(s) - \mu_2(s)\}.
\]

It is not hard to see that \( \tilde{c}(t,s) \) is a covariance operator since it is symmetric and positive definite. Let \( \gamma_i \) and \( v_i(t) \) be the corresponding eigenvalues and eigenfunctions satisfying that \( \int_I c(t,s)v_i(s)ds = \gamma_i v_i(t) \) and \( \gamma_1 > \gamma_2 > \cdots > \gamma_K > 0 \). Set \( \Delta(t) = \mu_1(t) - \mu_2(t) \) and \( \Delta_K = (\langle \Delta, v_1 \rangle, \langle \Delta, v_2 \rangle, \ldots, \langle \Delta, v_K \rangle) \in \mathbb{R}^K \). To ensure that \( \Delta_K \neq 0 \), we suppose \( \Delta(t) \notin \text{span}\{v_1(t), v_2(t), \ldots, v_K(t)\} \) which means that the difference of the two mean functions does not belong to the orthogonal complement of the space spanned by the first \( K \) eigenfunctions of \( \tilde{c} \). Note that if \( \Delta_K = 0 \), then

\[
\int_I \tilde{c}(t,s)v_i(s)ds = \int_I c(t,s)v_i(s)ds = \gamma_i v_i(t), \quad i = 1, 2, \ldots, K,
\]

which means \( \gamma_i \) and \( v_i \) are also the eigenvalues and eigenfunctions of \( c(t,s) \). In this case, the proposed test only has trivial power against the alternative. When \( \Delta_K \neq 0 \), the following proposition shows the consistency of the SN-based test.
Proposition 2.3.3. Consider the alternative (2.2) with $\lambda \in (0, 1)$ fixed and $\Delta_K \neq 0$. Suppose that assumption 2.3.1 holds with $p = 4$, then we have $G_{N, \hat{\eta}}(K) \to \infty$ in probability.

In what follows, we consider the local alternatives where the difference of the mean functions depends on the sample size $N$. In this case, we shall use the notation $\hat{c}(N)(t, s)$, $v_i^{(N)}$, $\Delta^{(N)}(t)$ and $\Delta_K^{(N)}$ instead of $\hat{c}(t, s)$, $v_i$, $\Delta(t)$ and $\Delta_K$.

Proposition 2.3.4. Consider the alternative (2.2) where $\lambda \in (0, 1)$ is fixed, and $||\Delta^{(N)}|| = O(1) \log \log N$ with $|\Delta^{(N)}_K| = o(1)$ and $\liminf_{N \to \infty} N^{1/2} \Delta^{(N)}_K > 0$. Here $|\Delta^{(N)}_K|$ denotes the Euclidean norm of $\Delta^{(N)}_K$. Further suppose that assumption 2.3.1 holds with $p = 4$, then we have $G_{N, \hat{\eta}}(K) \to \infty$ in probability.

Here we allow $|\Delta^{(N)}_K|$ to decay to zero at a rate $(\log \log N)/N^{1/2}$ in order to have nontrivial power. The condition $||\Delta^{(N)}|| = O(1)$ implies that the projection of the change on the first $K$ PC’s takes a nonzero proportion. As a by-product of our test, the change point can be naturally estimated by

$$\hat{k}^* = \arg\max_{k=1, 2, \ldots, N-1} \left\{ T_{N, \hat{\eta}}(k, K)'\hat{V}^{-1}_{N, \hat{\eta}}(k, K)T_{N, \hat{\eta}}(k, K) \right\}. \tag{2.17}$$

When $K = 1$, we are able to show that the SN-based change point estimator is in fact consistent. However, we encountered some technical difficulty when proving the consistency result for $K > 1$.

To study the power properties of the test statistic $G_{N, \hat{\xi}}(K^2)$, we may further assume that the functional sequence comes from two stationary subsequences $\{X_i^{(1)}(t)\}_{i=-\infty}^{k^*}$ and $\{X_i^{(2)}(t)\}_{i=k^*+1}^{\infty}$ under the $H_{a, 2}$. Following the arguments presented in the proofs of Proposition 2.3.3, we can show (omitting the details) that $G_{N, \hat{\xi}}(K^2)$ is consistent under the alternative $H_{a, 2}$ with $\hat{\lambda} \in (0, 1)$ fixed.

### 2.4 Numerical studies

To demonstrate the merits of the SN-based test statistics in a finite sample, we carried out several simulation studies to investigate the size and power properties of the proposed tests: for a change in the mean function in Section 4.1, for a change in the autocovariance operator at lag one in Section 4.2, and for a mean change in the 2-d functional data (a surface) in Section 4.3. Throughout the simulations, the number of Monte Carlo replications is set to be 1000.
2.4.1 Detecting the mean change for curves

Here we investigate the finite sample properties of the SN-based test for detecting the change of mean function. First, we consider independent functional observations. We follow the simulation setup in Berkes et al. (2009), where the mean function $\mu_1(t)$ was chosen to be zero under the null hypothesis and two different cases of $Y_i(t)$ were considered, namely the trajectories of the standard Brownian motion (BM) and Brownian Bridge (BB). Under the alternative (2.2), let $\mu_2(t) = t$ or $\mu_2(t) = \sin(t)$.

The change point $k^*$ is set to be $N/2$. We generate data on a grid of $10^3$ equispaced points in $[0, 1]$, and then convert discrete data to functional observations by using B-splines with 20 basis functions. We also tried 40 and 100 basis functions with sample size $N = 50, 100$ and $K = 1, 2, 3$, and found that the number of basis functions does not affect our results much. We compare the SN-based test (i.e., (2.10)) with the BGHK test; see (2.8). The empirical size and size-adjusted power of both the SN-based test and the BGHK test are summarized in Table 2.1. Size-adjusted power is computed using finite sample critical values based on the Monte Carlo simulation under the null hypothesis.

It can be seen that the empirical size of the SN-based test is comparable with the BGHK test in all cases considered here. As the expense of accounting for dependence, the SN-based test loses some power, but the power loss is fairly moderate.

To further examine the effect of dependence on the tests, we generate a functional sequence $\{Y_i(t)\}_{i=1}^N$ from the ARH(1) model which is defined as

$$Y_i(t) = \int_I \psi(t, s)Y_{i-1}(s)ds + \varepsilon_i(t), \quad t \in I, \quad i = 1, 2, \ldots, N,$$

where $\psi(t, s)$ is the kernel function and $\{\varepsilon_i(t)\}$ is a functional innovation sequence. To ensure that the ARH(1) model has a stationary solution, we assume

$$\|\psi\|_{HS}^2 = \int_0^1 \int_0^1 \psi^2(t, s)dsdt < 1,$$

where $\| \cdot \|_{HS}$ denotes the so-called Hilbert-Schmidt norm. Following the setup in Gabrys and Kokoszka (2007), we choose two kernel functions, the Gaussian kernel,

$$\psi(t, s) = C \exp \left( -\frac{t^2 + s^2}{2} \right)$$

and the Wiener kernel,

$$\psi(t, s) = C \min(t, s),$$
in our simulations. We consider the null hypothesis (2.1) and alternative hypothesis (2.2) as in the independent case, except that \( \{Y_i(t)\} \) is now generated from the ARH(1) model. We compare the SN-based test with the BGHK test and the HK test. To implement the latter test, we have to estimate the LRV matrix of the first \( K \) scores consistently. Given a \( p \)-dimensional multivariate series \( \{u_i = (u_{i1}, \ldots, u_{ip})\}_{i=1}^{N} \), the LRV matrix can be estimated nonparametrically by

\[
\hat{\Omega} = \sum_{|j| \leq b_N} K \left( \frac{j}{b_N} \right) \hat{\Gamma}_j,
\]

where \( b_N \) is the bandwidth, \( K(\cdot) \) is the kernel function and \( \hat{\Gamma}_j \) is the sample autocovariance function at lag \( j \). Here we use the Bartlett kernel, i.e, \( k(x) = (1 - |x|)I(|x| \leq 1) \), with the data-dependent truncation lag \( b_N = 1.1447(\hat{\alpha}(1)N)^{1/3} \), where

\[
\hat{\alpha}(1) = \left\{ \frac{1}{p} \sum_{i=1}^{p} \frac{4\hat{\sigma}_i^4 \hat{\rho}_i^2}{(1 - \hat{\rho}_i)^2 (1 + \hat{\rho}_i)^2} \right\} \left\{ \frac{1}{p} \sum_{i=1}^{p} \frac{\hat{\sigma}_i^4}{(1 - \hat{\rho}_i)^4} \right\}^{-1}.
\]

(2.18)

Here \( \hat{\rho}_i \) is the least squares coefficient estimate by regression \( u_{ki} \) on \( u_{(k-1)i} \) and \( \hat{\sigma}_i^2 \) is the estimate of the innovation variance. The plug-in bandwidth formula (2.18) is suggested by Andrew (1991) and is shown to minimize the MSE of the LRV estimator when the true model is the vector autoregressive model of order one.

We report the simulation results for \( N = 50, 100, K = 1, 2, 3, ||\psi||_{HS} = 0.5 \) and BM and BB innovations in Table 2.2 and Table 2.3. Several other values of \( ||\psi||_{HS} \) (e.g. 0.3, 0.8) were also considered, but the results are not reported here to save space. From Table 2.2, we see that the size distortion of the BGHK test is severely large compared to the other two tests. This is due to the fact that it is designed only for independent functional data and is invalid in the temporally-dependent case. For the HK test, the size distortion is less severe but seems sensitive to the choice of \( K \). It tends to be oversized for small \( K \) but undersized for large \( K \). For the SN-based test, size distortion is apparent for \( N = 50 \), but improves for \( N = 100 \). The size for the SN-based test seems quite robust to the choice of \( K \). Table 2.3 presents selected results of the size-corrected powers from which several observations can be made. First, the BGHK test delivers the highest power among the three tests, which is largely due to its severe upward size distortion. Second, the power of the SN-based test is comparable to that of the HK test for \( N = 50 \) and BM innovations. Furthermore the SN-based test tends to have moderate power loss when sample size increases to 100. In the case of the BB innovations, the SN-based test is superior to the HK test in power. Overall, the severe size distortion...
of the BGHK test under weak dependence suggests its inability to accommodate dependence and thus is not recommended in testing for a change point for dependent functional data. The HK test is able to account for dependence but it is sensitive to the choice of bandwidth $b_N$ and $K$. As shown below, the data-dependent bandwidth used in the HK test could lead to nonmonotonic power.

Compared to the other two tests, the SN-based test tends to have more accurate size at the sacrifice of some power, which is consistent with the “better size but less power” phenomenon seen in the univariate setup (see Shao and Zhang, 2010).

Furthermore, we examine the monotonic power property of the SN-based test in the functional setup through simulations. In the univariate setting, the change point test, which involves LRV estimation using the data-dependent bandwidth, can exhibit nonmonotonic power (see e.g. Vogelsang, 1999; Crainiceanu and Vogelsang, 2007; Altissimo and Corradi, 2003). There are some recent studies aiming to overcome the nonmonotonic power problem in the univariate time series setup (see Juhl and Xiao, 2009; Shao and Zhang, 2010). To study the monotonic power property, we focus on the change of mean function and consider the data generating process

$$X_i(t) = Y_i(t) + \delta f(t)I\{i > N/2\}, \ i = 1, 2, \ldots, N,$$  

where $\{Y_i(t)\}$ follows the ARH(1) model with Gaussian kernel, and BM or BB innovations. The constant $\delta$ here is used to control the magnitude of change and $f(t) = t$ or $\sin(t)$. The size-adjusted power for $K = 1$ and $N = 50$ is plotted as a function of $\delta$ in Figure 2.1. Qualitatively similar results were observed for $N = 100$, but are not reported to conserve space. We compare the performance of the SN-based test with the BGHK test and the HK test. Like the univariate case, the SN-based test shows monotonic power in all situations even though it could lose moderate power to the BGHK test. Not surprisingly, due to the upward bias of the data-dependent bandwidth, the HK test exhibits nonmonotonic power, with power going to zero for relatively large changes in the mean function.

These results indicate that the nonmonotonic power issue still exists in the functional setting if one estimates the LRV matrix of scores nonparametrically using data-dependent bandwidth. In contrast, the SN-based test inherits the monotonic power property, that holds in the univariate case (Shao and Zhang, 2010).

In the univariate setting, Crainiceanu and Vogelsang (2007) and Juhl and Xiao (2009) have proposed different methods to overcome the nonmonotonic power problem in testing for a change point in mean. However, their methods both involve bandwidth parameters and their finite sample
performance is unsatisfactory as seen from the numerical comparison in Shao and Zhang (2010). For example, Crainiceanu and Vogelsang (2007) proposed to estimate the long run variance using residuals obtained under the one-break model but the size distortion is large for time series with strong dependence (e.g., AR(1) model with AR(1) coefficient 0.8). Juhl and Xiao (2009) used residuals from nonparametric regression to estimate long run variance, but they did not completely eliminate the nonmonotonic power problem (Section 4.1 of Shao and Zhang, 2010). In a sense, the two methods mentioned above were proposed to account for a possible change point in the LRV estimator. But they did not perform well in finite sample, so we expect that the extensions of these methods to functional setting will not work well, although a serious investigation is beyond the scope of the chapter.

2.4.2 Detecting the changes in the lag-1 autocovariance operator

In this subsection, we study the finite sample performance of the SN-based test for detecting the change of the autocovariance operator at lag one. Under the null, we generate functional observations from the mean zero ARH(1) model with Gaussian kernel and \( \| \psi \|_{HS} = 0.3 \). Under the alternative, we consider the following data-generating process,

\[
\begin{align*}
Y_i(t) &= \int_0^1 \psi_1(t, s)Y_{i-1}(s)ds + \epsilon_i(t), \quad i = 1, 2, \ldots, N/2; \\
Y_i(t) &= \int_0^1 \psi_2(t, s)Y_{i-1}(s)ds + \epsilon_i(t), \quad i = N/2 + 1, N/2 + 2, \ldots, N,
\end{align*}
\]

where \( \psi_1(s, t) \) and \( \psi_2(s, t) \) are both Gaussian kernels with \( \| \psi_1 \|_{HS} = 0.3 \) and \( \| \psi_2 \|_{HS} = 0.8 \). The SN-based test is compared with the tests proposed in Horváth et al. (2010) for detecting the stability of the ARH(1) model. Formally, Horváth et al.’s tests can be written as

\[
I_{N, \xi} = \frac{1}{N - 1} \sum_{k=1}^{N-1} T_{N, \xi}(k, K^2) \tilde{\Sigma}_\xi^{-1}(k, K^2) T_{N, \xi}(k, K^2),
\]

where \( \tilde{\Sigma}_\xi(k, K^2) \) is a consistent estimator of the LRV matrix of \( \{ \xi_i \} \). We define \( \hat{D}_k \) as the nonparametric LRV matrix estimator computed from \( \{ \xi_1, \xi_2, \ldots, \xi_k \} \) by using the Bartlett kernel with bandwidth given by (2.18). Similarly, we define the LRV matrix estimator \( \hat{D}_{N-k} \) based on \( \{ \xi_{k+1}, \xi_{k+2}, \ldots, \xi_{N-1} \} \). Following Horváth et al. (2010), we consider two different ways of estimating the LRV matrix here: 1) \( \tilde{\Sigma}_\xi(k, K^2) = \frac{k}{N-1} \hat{D}_k + \frac{N-k-1}{N-1} \hat{D}_{N-k} \); 2) \( \tilde{\Sigma}_\xi(k, K^2) = \hat{D}_N \) for all \( 1 \leq k \leq N - 1 \), and we denote the resulting tests by HHK1 and HHK2. We present the empirical size and size-corrected power for
It can be clearly seen that the size distortion for HHK1 test is substantial, especially for \( N = 50 \) and \( K = 2, 3 \). The HHK2 test performs relatively well but tends to be undersized when \( K \) increases. Compared to HHK1 and HHK2 tests, the size performance of SN-based test is quite satisfactory. For the size-adjusted power, we find that the HHK1 test is the most powerful in all cases, presumably due to its upward size distortion. The HHK2 test has reasonable power for \( K = 1 \) while the power could drop dramatically as \( K \) increases for small sample size. The finding here agrees generally with the results in Horváth et al. (2010) which shows that the HHK2 test is conservative for large \( K \). The SN-based test delivers moderate power and the power seems robust to \( K \). Overall, the simulation results clearly suggest a trade-off between the size distortion and power loss for the SN-based test, which has been found to be the case in the univariate setup.

### 2.4.3 Detecting the mean change for 2-d functional observations

Here, we perform a simulation study to demonstrate the validity of the SN-based test for detecting a mean function change in 2-d functional data. For simplicity, we focus on a rectangular region though our test can be applied to functional data on a region of irregular shape. Under the null, we generate 2-d functional observations \( \{Y_i(s_1, s_2)\}_{i=1}^N \) in the following two ways:

1. \( Y_i(s_1, s_2) = X_i^{(1)}(s_1)X_i^{(2)}(s_2) \), where \( \{X_i^{(1)}\} \) and \( \{X_i^{(2)}\} \) are mutually independent and contain possibly dependent continuous random processes respectively. Here we choose \( \{X_i^{(j)}(s)\}, j = 1, 2 \) to be BM, BB and ARH(1) process with Gaussian kernel with \(||\psi||_{HS} = 0.5 \) and BM innovations.

2. The sequence \( \{Y_i(s, t)\} \) follows the ARH\(_2\)(1) model defined on \([0, 1]^2\), that is,

   \[
   Y_{i+1}(s_1, s_2) = \int_0^1 \int_0^1 \psi(s_1, s_2, u_1, u_2)Y_i(u_1, u_2)du_1du_2 + \varepsilon_{i+1}(s_1, s_2),
   \]

   where \( \psi(s_1, s_2, u_1, u_2) = C \exp\{(s_1^2 + s_2^2 + u_1^2 + u_2^2)/2\} \) and \( \varepsilon_i(s_1, s_2) = X_i^{(1)}(s_1)X_i^{(2)}(s_2) \) is the tensor product of independent BM or BB.

We suppose the data is collected on a grid of \( 20 \times 20 \) equally spaced points on \([0, 1]^2\). The discrete observations are smoothed by the thin plate spline (see Wahba, 1990). In order to obtain the eigenvalues and eigenfunctions, the 2-d functional data is discretized to a fine grid of \( 30 \times 30 \) equally spaced points. The data matrix corresponding to each observation is then concatenated into a single
long vector. Hence the functional eigenanalysis problem for 2-d functional data is converted to an approximately equivalent matrix eigenanalysis task as in the one dimensional case (see Ramsay and Silverman, 2005 for more details). To illustrate the power properties of the SN-based test, we consider the following alternatives:

1. \( Y_i(s_1, s_2) = \{ X_i^{(1)}(s_1) + f(s_1) \} \{ X_i^{(2)}(s_2) + f(s_2) \} \) with \( f(s) = s \) or \( f(s) = \sin(s) \).

2. \( Y_i(s_1, s_2) = Z_i(s_1, s_2) + g(s_1, s_2) \), where \( Z_i(s_1, s_2) \) is the aforementioned ARH(1) process, and \( g(s_1, s_2) = s_1 s_2 \) or \( g(s_1, s_2) = \sin(s_1) \sin(s_2) \).

The change point \( k^* \) is set to be \( N/2 \). The selected simulation results are summarized in Table 2.5. For the data generated by tensor product, the SN-based test is conservative when \( N = 50 \) and the size becomes closer to the nominal level as \( N \) increases to 100. For the ARH(1) process, the SN-based test is oversized, the size distortion diminishes and the power appreciates as sample size increases. It is also interesting to note that the special covariance structure of BB tends to give us more power, as we have seen before. In conclusion, the SN-based test delivers satisfactory size and reasonable power in the 2-d setting.

2.5 Applications

In this section, we consider two empirical datasets, namely, the single-point time series of central England temperature record and a spatio-temporal gridded dataset consisting of the bias between observed and model-simulated annual average temperature covering a subregion of North America (latitude: 34.25°N–51.25°N; longitude: 77.25°W–120.25°W) obtained from a coupled atmosphere-ocean general circulation model (AOGCM) and interpolated station observations.

2.5.1 Analysis of central England temperatures

We first apply the SN-based test to detect the change point in the functional mean of the central England temperature record that has been previously studied in Berkes et al. (2009). This data set represents the longest continuous thermometer-based temperature record on earth, consisting of 228 years (1780-2007) of average daily temperatures in central England (see Parker et al., 1992). Following Berkes et al. (2009), we view the data as 228 curves with 365 measurements on individual curve. The discrete observations were registered as functional data by using 12 \( B \)-spline basis functions. To compute \( G_{N, \hat{\eta}}(K) \), we choose the smallest \( K \) such that \( \sum_{i=1}^{K} \hat{\lambda}_i / \sum_{i=1}^{12} \hat{\lambda}_i > 0.8 \)
following Berkes et al. (2009). If the test indicates any change point, then it is estimated by $\hat{k}^*$ (see (2.17)). The procedure is repeated until each segment has a constant mean function. The results are summarized in Table 2.6. The two change points, 1927 and 1993, detected by the SN-based test are fairly close to the change points, 1925 and 1992, identified in Berkes et al. (2009). The SN-based test suggests the mean function is stable over the period from 1780 to 1927, whereas Berkes et al.'s test detected two more change points, 1807 and 1849. However, the change at 1849 is not as obvious relative to the changes at 1925 and 1992 according to Figure 2 in Berkes et al. (2009). Of course, since it is not known whether there is a change point at 1849, either our SN-based test fails to reject due to its relatively lower power or Berkes et al.'s test falsely rejects due to its large upward size distortion. Nevertheless, our results suggest that the evidence for supporting the one change point in 1849 is weak. Figure 2.2 plots the mean function in each partition segments suggested by the SN-based test. It clearly shows the warming trend of the central England temperature. As mentioned in Berkes et al. (2009), although it is not realistic to believe that the change happens abruptly in one year, in practice this modeling assumption is useful in identifying a potential trend of change.

\subsection{Analysis of the bias between gridded observations and GCM simulations}

Next, we apply the SN-based test to a gridded spatio-temporal temperature data set covering a subregion of North America. The data set comes from two separate sources: gridded observations generated from interpolation of station records (HadCRU), and gridded simulations generated by an AOGCM (NOAA GFDL CM2.1). Both datasets provide daily average temperature for the same 19-year period, 1980-1998 (see Delworth et al., 2006; Brohan et al., 2006). Each surface is viewed as a 2-d functional datum. We aim to test whether the bias or difference between station observations and the model outputs is stable over the examined period (1980-1998). The data is first transformed to the same resolution ($87 \times 35$) by bilinear interpolations. The bias is then computed by taking the difference of the two surfaces for each year and is converted to functional data through the thin plate spline. Figure 2.3 presents the first six PCs of the bias, which summarize the major patterns of the variability. The first PC which explains 60.4\% of the total variation clearly dominates other types of variations. Although the first PC is negative over the whole region, it places more weight at the center of the domain than at the boundary. This indicates that a great amount of variability over a year will be found by the relatively heavy weights over the central region, which is relatively far from the ocean, with less contribution from the border area, which is close to the ocean. This is
consistent with the physical reality that temperature variability is much higher over land than over ocean, and is much more sensitive to the land surface parameterization scheme used by the global model. Applying the SN-based test to test the mean change, we tabulate the results of the tests based on the first $K$ PCs in Panel (a) of Table 2.7 and the results of the tests based on individual PCs in Panel (b) of Table 2.7. From Panel (a), we notice that when $K = 1$ or 2, the SN-based test does not detect any significant change points at the usual 5% significance level. When $K = 3$, the SN-based test with a $p$-value in the range $(0.005, 0.01)$ suggests that there is a change point at 1990. The results in Panel (b) are consistent with the finding from Panel (a) in that the test based on the third PC indicates a significant change point at 1990 but the tests based on other PCs do not detect any significant change points. The change in the bias is shown in Figure 2.4 by comparing the difference of the average biases in two periods (1980-1990; 1991-1998). It can be seen from Figure 2.4 that the bias tends to increase in the northern region while it decreases in the southern area. It is also worth noting that the pattern of the change of the bias seems to be similar to that of the third PC as plotted in Figure 2.3.

2.6 Discussion and conclusion

In this article, SN-based tests have been developed to detect a change point in the functional mean and the lag-1 autocovariance operator of temporally dependent functional observations. The test statistic is constructed based on the estimated finite dimensional scores and the estimation effect turns out to be asymptotically negligible due to the special form of the SN-based statistic. The limiting null distribution of the SN-based statistic is nonstandard and its critical values have been tabulated by Shao and Zhang (2010). Compared to the existing tests developed for independent/dependent functional data, the SN-based test has some appealing features: 1) it is easy to implement and does not involve any bandwidth parameter; 2) it is shown to enjoy the monotonic power properties in the functional context; 3) our test, developed for temporally-dependent functional data, inherits the “better size but less power” property of the SN-based test in the univariate setup. The finite sample performance is quite good and stable with respect to $K$. It can be readily applied to temporally-dependent functional curves and functional surfaces.

Our test statistic currently detects only one change point, but it can be further extended to the multiple change point alternative in a straightforward manner; see Shao and Zhang (2010) section 2.3 for a discussion in the univariate setting. Furthermore, the SN-based test still requires a user-
chosen parameter $K$ which also appears in related work by Berkes et al. (2009) and Hörmann and Kokoszka (2010). In practice, $K$ can be chosen by $K = \inf \{ J : \sum_{i=1}^{J} \hat{\lambda}_i / \sum_{i=1}^{m} \hat{\lambda}_i > \alpha \}$, where $m$ is the number of basis functions in smoothing and $\alpha$ is a pre-specified number, say 85%. If the goal is to infer the low frequency behavior of the functional data, this ac-hoc method of choosing $K$ should be practically useful and informative. On the other hand, if some high frequency behavior of the functional data is also of interest, then it is wise to let $K$ dependent on the sample size $N$, which requires a modification of our asymptotic theory. This is beyond the scope of this article and will be investigated in the future.

### 2.7 Proofs of the main results

We first introduce some useful notation. For $1 \leq t_1 \leq t_2 \leq N$ and $A = (a_1, a_2, \ldots, a_N) \in \mathbb{R}^{K \times N}$, let $S_N(t_1, t_2, A) = S_{N,A}(t_1, t_2) = \sum_{i=t_1}^{t_2} a_i$ with $a_i = (a_{i1}, a_{i2}, \ldots, a_{iK})' \in \mathbb{R}^K$. Let $\hat{\beta}_j = \int_0^1 Y_i(t) \hat{\phi}_j(t) dt$, $\hat{\beta}_i = (\hat{\beta}_{i1}, \ldots, \hat{\beta}_{iK})'$ and $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_N) \in \mathbb{R}^{K \times N}$. Define $\eta = (\eta_1, \eta_2, \ldots, \eta_N) \in \mathbb{R}^{K \times N}$ and its empirical counterparts $\hat{\eta} = (\hat{\eta}_1, \hat{\eta}_2, \ldots, \hat{\eta}_N) \in \mathbb{R}^{K \times N}$. Similarly, we can define $\xi \in \mathbb{R}^{K^2 \times (N-1)}$ and its empirical counterparts $\hat{\xi}$ by replacing $\xi_i$ with $\hat{\xi}_i$; see section 2.2. Let $| \cdot |$ be the Euclidian norm of a vector and $\| \cdot \|_M$ the matrix norm $\| A \|_M = \sup_{|x| \leq 1} |Ax|$ for a matrix $A$. Denote $C$ a generic constant which could be different from line to line.

**Proof of Theorem 2.3.1.** Define

$$G_{N,\eta}(K) = C(N^{-1/2} S_N(1, \lceil Nr \rceil, \eta), r \in [0,1]).$$

Set $\hat{c}_i = \text{sign}(\langle \phi_i, \hat{\phi}_i \rangle)$ and $\hat{C} = \text{diag}(\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_K)$. Following the arguments in Theorem 5.1 of Hörmann and Kokoszka (2010), we can derive that $\{ \eta_i \}$ is $L^2$-$m$-approximable. Hence by Theorem A.2 of Hörmann and Kokoszka (2010), we have

$$N^{-1/2} S_N(1, \lceil Nr \rceil, \eta) \Rightarrow \sum_{i=1}^{1/2} B_K(r).$$

Applying the continuous mapping theorem, we have that $G_{N,\eta}(K) \rightarrow^d G(K)$. Note that $\hat{\beta}_{i,j} = \hat{\beta}_{i,j} - (1/N) \sum_{i=1}^{N} \hat{\beta}_{i,j}$, for $j = 1, 2, \ldots, K$. Under the $H_{0,1}$, it is easy to see that both $T_{N,\eta}(k, K)$ and $V_{N,\eta}(k, K)$ remain the same if $\hat{\eta}_i$ is replaced by $\hat{\beta}_i$. Because of the quadratic form of $G_{N,\hat{\eta}}(K)$ and the simple fact that $\hat{C}^2 = I_K$, the statistic $G_{N,\hat{\eta}}(K)$ does not change if $\hat{\eta}_i$ is replaced by $\hat{C}_j \hat{\beta}_i$. Based
on the result that
\[
\sup_{r \in [0,1]} \frac{1}{\sqrt{N}} \left| S_N(1, [Nr], \eta) - S_N(1, [Nr], \hat{C}_{\eta}) \right| = o_p(1),
\]

which was stated in the proof of Theorem 5.1 of Hörmann and Kokoszka (2010) (see equation A.9 therein), it is straightforward to see that the difference between \(G_{N,\hat{\eta}}(K)\) and \(G_{N,\eta}(K)\) is asymptotically negligible. Therefore the proof is complete.

\[\Box\]

**Proof of Theorem 2.3.2.** Recall that \(\xi_i = (\eta_i, 1, \eta_{i+1,1}, \ldots, \eta_i, 1, K, \ldots, \eta_i, K, \eta_{i+1,1}, \ldots, \eta_i, K, \eta_{i+1,K})'\).

Define \(\eta_{i,j}^{(m)} = \int_{\mathcal{I}} X_i^{(m)}(t) \phi_j(t) dt\), where \(X_i^{(m)}\) is the \(m\)-dependent approximation of \(X_i\) (see Definition 3.3.1), and let \(\xi_i^{(m)}\) be the counterpart of \(\xi_i\) by replacing \(\eta_{i,j}\) with \(\eta_{i,j}^{(m)}\) in \(\xi_i\). Let \(\gamma(t, s) = E[X_1(t)X_2(s)]\) be the lag-1 autocovariance function. We first see that

\[
E|\xi_1 - \xi_1^{(m)}|^2 = E \sum_{j=1}^{K} \sum_{l=1}^{K} (\eta_{1,j} - \eta_{1,j}^{(m)})(\eta_{2,l} - \eta_{2,l}^{(m)})^2 \\
\leq 2 \left[ E \left\{ \sum_{j=1}^{K} \sum_{l=1}^{K} (\eta_{1,j} - \eta_{1,j}^{(m)})^2 \right\} + E \left\{ \sum_{j=1}^{K} \sum_{l=1}^{K} (\eta_{2,l} - \eta_{2,l}^{(m)})^2 (\eta_{1,j}^{(m)})^2 \right\} \right] \\
= 2(I_1 + I_2),
\]

Note that \(E(\eta_{2,l}^4) = E \left\{ \int_{\mathcal{I}} X_2(t) \phi_l(t) dt \right\}^4 \leq E \left\{ \int_{\mathcal{I}} X_2^2(t) dt \right\}^2 = E||X_2||^4 < \infty\), where we have used the orthonormal property of \(\phi_l(t)\). By the Cauchy-Schwarz inequality, we have

\[
I_1 \leq \left[ E \left\{ \sum_{j=1}^{K} (\eta_{1,j} - \eta_{1,j}^{(m)})^2 \right\}^{2/2} \right] ^{1/2} \left\{ E \left( \sum_{l=1}^{K} \eta_{2,l}^2 \right) ^{2/2} \right\} ^{1/2} \\
\leq C \left\{ \sum_{j=1}^{K} E(\eta_{1,j} - \eta_{1,j}^{(m)})^4 \right\} ^{1/2} \left( \sum_{l=1}^{K} E\eta_{2,l}^4 \right) ^{1/2} \leq C \left\{ \sum_{j=1}^{K} E(\eta_{1,j} - \eta_{1,j}^{(m)})^4 \right\} ^{1/2} \]

\[
= C \left\{ \sum_{j=1}^{K} \left( \int_{\mathcal{I}} (X_1(t) - X_1^{(m)}(t)) \phi_j(t) dt \right)^4 \right\} ^{1/2} = C \left( E||X_1 - X_1^{(m)}||^4 \right)^{1/2}
\]

Similarly we get \(I_2 \leq C \left( E||X_1 - X_1^{(m)}||^4 \right)^{1/2}\). It follows that \(\left( E|\xi_1 - \xi_1^{(m)}|^2 \right)^{1/2} \leq Ce_4(X_1 - X_1^{(m)})\), which yields

\[
\sum_{m=1}^{\infty} \left( E|\xi_1 - \xi_1^{(m)}|^2 \right)^{1/2} \leq C \sum_{m=1}^{\infty} e_4(X_1 - X_1^{(m)}) < \infty.
\]

Again using Theorem A.2 of Hörmann and Kokoszka (2010), we know that \(N^{-1/2} \{ S_N(1, [Nr], \xi) - \}

23
\[ |N_r E[\xi_1]| \Rightarrow \Sigma_{\xi,K}^{1/2} B_{K^2}(r). \]

Let

\[ G_{N,\xi}(K^2) = C(N^{-1/2} S_N(1, |N_r|, \xi), r \in [0, 1]). \]

By the continuous mapping theorem we have that \( G_{N,\xi}(K^2) \to_d G(K^2) \). Because of the form of \( G_{N,\xi}(K^2) \), it is sufficient to show that

\[
\sup_{r \in [0,1]} \frac{1}{\sqrt{N}} \left| S_N(1, |N_r|, \xi) - S_N(1, |N_r|, \hat{\xi}\hat{\xi}) \right|
\]

\[
- \left\{ E S_N(1, |N_r|, \xi) - E S_N(1, |N_r|, \hat{\xi}\hat{\xi}) \right\} = o_p(1),
\]

where \( \hat{\xi} = \text{diag}(\hat{c}_1 \hat{c}_1, \ldots, \hat{c}_1 \hat{c}_K, \hat{c}_2 \hat{c}_1, \ldots, \hat{c}_K \hat{c}_K) \). The claim follows provided that for any \( 1 \leq j, l \leq K \),

\[
I_{j,l} = \frac{1}{\sqrt{N}} \sup_{r \in [0,1]} \left| \sum_{i=1}^{|N_r|} \left\{ \eta_{i,j} \eta_{i+1,l} - E(\eta_{i,j} \eta_{i+1,l}) - \hat{c}_j \hat{\eta}_{i,j} \hat{\eta}_{i+1,l} + E(\hat{c}_j \hat{\eta}_{i,j} \hat{\eta}_{i+1,l}) \right\} \right|
\]

\[
= \frac{1}{\sqrt{N}} \sup_{r \in [0,1]} \left| \int_I \int_I \sum_{i=1}^{|N_r|} \{ X_i(t)X_{i+1}(s) - \gamma(t,s) \} \hat{u}_{j,l}(t,s) dt ds \right| = o_p(1),
\]

where \( \hat{u}_{j,l}(t,s) = \phi_j(t)\phi_l(s) - \hat{c}_j \hat{\phi}_j(t)\hat{\phi}_l(s) \). Observing that \( ||\hat{u}_{j,l}||^2 \leq 2(||\phi_j - \hat{c}_j \hat{\phi}_j||^2 + ||\phi_l - \hat{c}_j \hat{\phi}_l||^2) \)

and using the fact that \( ||\phi_j - \hat{c}_j \hat{\phi}_j|| = O_p(N^{-1/2}) \) (see Theorem 3.2 in H"ormann and Kokoszka, 2010), we derive \( ||\hat{u}_{j,l}|| = O_p(N^{-1/2}) \). Since

\[
I_{j,l} \leq \frac{||\hat{u}_{j,l}||}{\sqrt{N}} \sup_{r \in [0,1]} \left[ \int_I \int_I \left\{ \sum_{i=1}^{|N_r|} (X_i(t)X_{i+1}(s) - \gamma(t,s)) \right\}^2 dt ds \right]^{1/2},
\]

the conclusion follows from Lemma 2.7.1. \( \square \)

**Lemma 2.7.1.** Assume that \( X_i \in L^8_{m_0} \) and \( \{X_i\} \) is \( L^8-m \)-approximable. Then under \( H_{0,2} \), we have

\[
\frac{1}{N^2} \sup_{r \in [0,1]} \left[ \int_I \int_I \left\{ \sum_{i=1}^{|N_r|} (X_i(t)X_{i+1}(s) - \gamma(t,s)) \right\}^2 dt ds \right] = o_p(1).
\]

**Proof of Lemma 2.7.1.** Let \( Z_i(t,s) = X_i(t)X_{i+1}(s) - \gamma(t,s) \). We first show that the process \( \{Z_i(t,s)\} \)
is $L^4$-m-approximable in the Hilbert space of integrable functions defined on $[0,1]^2$. Note that

$$E\|Z_1 - Z_1^{(m)}\|^4 = E\left\{ \int_I \int_I (X_1(t)X_2(s) - X_1^{(m)}(t)X_2^{(m)}(s))^2 dt ds \right\}^2$$

$$\leq C E\left\{ \|X_1\|^2 \|X_2 - X_2^{(m)}\|^2 + \|X_2^{(m)}\| \|X_1 - X_1^{(m)}\|^2 \right\}^2$$

$$\leq C \left\{ (E\|X_1\|^8)^{1/2} (E\|X_2 - X_2^{(m)}\|^8)^{1/2} \right\}^2$$

$$+ (E\|X_2\|^8)^{1/2} (E\|X_1 - X_1^{(m)}\|^8)^{1/2} \right\}.$$ 

Thus we get $e_4(Z_1 - Z_1^{(m)}) \leq C\{e_8(X_1 - X_1^{(m)}) + e_8(X_2 - X_2^{(m)})\}$, which, along with the assumption that $\{X_i\}$ is $L^k$-m-approximable, implies $\{Z_i\}$ is $L^4$-m-approximable. The rest of the proof essentially follows the argument in the proof of Theorem 5.1 in Hörmann and Kokoszka (2010). We omit the details here.

Proof of Proposition 2.3.3. Define $\hat{\Delta}_K = ( < \Delta, \hat{\phi}_1 > , < \Delta, \hat{\phi}_2 > , \ldots , < \Delta, \hat{\phi}_K > )'$. Let $\alpha_{ij} = \int_I Y_i(t)v_j(t) dt$, $\alpha_i = (\alpha_{i1}, \ldots, \alpha_{iK})'$ and $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)$. Following lemma A.1 in Berkes et al. (2009), we have that under the local alternatives,

$$\hat{c}(t,s) = \frac{1}{N} \sum_{i=1}^N Y_i(t) - \hat{Y}_N(t) \{ Y_i(s) - \hat{Y}_N(s) \} + \frac{k^*(N - k^*)}{N^2} \Delta(t) \Delta(s) + r_N(t,s),$$

where

$$r_N(t,s) = \left( \frac{N - k^*}{N^2} \right)^k \sum_{i=1}^{k^*} \left[ \{ Y_i(t) - \hat{Y}_N(t) \} \Delta(t) + \{ Y_i(s) - \hat{Y}_N(s) \} \Delta(s) \right]$$

$$+ \frac{k^*}{N^2} \sum_{i=k^*+1}^N \left[ \{ Y_i(t) - \hat{Y}_N(t) \} \Delta(t) + \{ Y_i(s) - \hat{Y}_N(s) \} \Delta(s) \right],$$

is the remainder term. It can be shown that $||r_N(t,s)||^2 = \int_I \int_I r_N^2(t,s) dt ds$ is $o_p(1)$ by using the ergodic theorem for $L^p$-m-approximable process $Y_i(t)$ (note that $Y_i(t)$ admits the ergodic representation $Y_i(t) = f(\varepsilon_i(t), \varepsilon_{i-1}(t), \ldots)$). It follows that $||\hat{c}(t,s) - \hat{c}(t,s)|| = o_p(1)$. By Lemma 4.3 in Bosq (2000), we get $||\hat{c}_i \hat{\phi}_i - v_i|| = o_p(1)$ for all $1 \leq i \leq K$. Using this fact and similar arguments in the proof of Theorem 5.1 in Hörmann and Kokoszka (2010), we get the following two results:

$$N^{-1/2} S_N(1, \lfloor Nr \rfloor, \alpha) \text{satisfies the functional central limit theorem;} \quad (2.22)$$

$$\sup_{r \in [0,1]} \frac{1}{\sqrt{N}} \left| S_N(1, \lfloor Nr \rfloor, \hat{C} \hat{\beta}) - S_N(1, \lfloor Nr \rfloor, \alpha) \right| = o_p(\log \log N), \quad (2.23)$$
which imply that
\[ \sup_{r \in [0,1]} \frac{1}{\sqrt{N}} \left| S_N(1, [Nr], \hat{\mathbf{C}} \hat{\beta}) \right| = o_p(\log \log N). \]  
(2.24)

Simple calculation shows that
\[ T_{N,\hat{\beta}}(k^*, K) = \frac{1}{\sqrt{N}} \left( \sum_{i=1}^{k^*} \hat{\beta}_i - \frac{k^*}{N} \sum_{i=1}^{N} \hat{\beta}_i \right) + \frac{k^*(N-k^*)}{N^{3/2}} \Delta_K. \]

Notice that \(|\hat{\mathbf{C}} \Delta_K - \Delta_K| = o_p(1)\) provided that \(|\hat{\epsilon}_i \hat{\phi}_i - v_i| = o_p(1)\). By (4.32), we get
\[ |\hat{\mathbf{C}} T_{N,\hat{\beta}}(k^*, K)| = \left| \frac{1}{\sqrt{N}} \left( \sum_{i=1}^{k^*} \hat{\beta}_i - \frac{k^*}{N} \sum_{i=1}^{N} \hat{\beta}_i \right) + \frac{k^*(N-k^*)}{N^{3/2}} (\hat{\mathbf{C}} \Delta_K - \Delta_K) \right. 
\left. + \frac{k^*(N-k^*)}{N^{3/2}} \Delta_K \right| = O_p(N^{1/2}). \]  
(2.25)

On the other hand, it is not hard to see that
\[ \left\| \frac{1}{N^2} \sum_{t=1}^{k^*} \left\{ \hat{\mathbf{C}} S_{N,\hat{\beta}}(1, t) - \frac{t}{k^*} \hat{\mathbf{C}} S_{N,\hat{\beta}}(1, k^*) \right\} \right\|_M \left\| \hat{\mathbf{C}} S_{N,\hat{\beta}}(1, t) - \frac{t}{k^*} \hat{\mathbf{C}} S_{N,\hat{\beta}}(1, k^*) \right\|_M \]
\[ = \left\| \frac{1}{N^2} \sum_{t=1}^{k^*} \left\{ S_{N}(1, t, \hat{\mathbf{C}} \hat{\beta}) - \frac{t}{k^*} S_{N}(1, k^*, \hat{\mathbf{C}} \hat{\beta}) \right\} \left\| S_{N}(1, t, \hat{\mathbf{C}} \hat{\beta}) - \frac{t}{k^*} S_{N}(1, k^*, \hat{\mathbf{C}} \hat{\beta}) \right\|_M \right. 
\left. \leq \frac{K}{N^2} \sum_{t=1}^{k^*} \left\| S_{N}(1, t, \hat{\mathbf{C}} \hat{\beta}) - \frac{t}{k^*} S_{N}(1, k^*, \hat{\mathbf{C}} \hat{\beta}) \right\|^2 = o_p((\log \log N)^2). \]  
(2.26)

Here we used the inequality that \(\|ab\|_M \leq K \max_{1 \leq i,j \leq K} \{a_ib_j\} \leq K|a||b|\) for \(a, b \in \mathbb{R}^K\). Similarly, we can prove the same result for the second term in \(\hat{\mathbf{C}} V_{N,\hat{\beta}}(k^*, K)\hat{\mathbf{C}}'\) (see 2.9). Therefore we have \(\|\hat{\mathbf{C}} V_{N,\hat{\beta}}(k^*, K)\hat{\mathbf{C}}'\| = o_p((\log \log N)^2)\). Along with the fact that \(\|\hat{\mathbf{C}} T_{N,\hat{\beta}}(k^*, K)\| = O_p(N^{1/2})\), we get
\[ G_{N,\hat{\beta}}(K) \geq T_{N,\hat{\beta}}(k^*, K) V_{N,\hat{\beta}}^{-1}(k^*, K) T_{N,\hat{\beta}}(k^*, K) \]
\[ \geq |T_{N,\hat{\beta}}(k^*, K)|^2/||V_{N,\hat{\beta}}(k^*, K)||_M, \]
where the right hand side diverges to infinity as \(N \rightarrow +\infty\). Note the second inequality comes from the fact that \(a' A^{-1} a \geq |a|^2/||A||_M\) for a vector \(a \in \mathbb{R}^K\) and an invertible matrix \(A \in \mathbb{R}^{K \times K}\). \(\Box\)

**Proof of Proposition 2.3.4.** Note that \(||\hat{\epsilon}^{(N)}(t, s) - c(t, s)|| = o(1)\) under the assumption that \(||\Delta^{(N)}|| = o(1)\). Following the arguments in proof of Proposition 2.3.3, we can show that \(||\hat{c}(t, s) - c(t, s)|| = o_p(1)\) and hence \(||\hat{\epsilon}_i \hat{\phi}_i - \phi_i|| = o_p(1)\) for all \(1 \leq i \leq K\). Using similar arguments in the proof of
Theorem 5.1 in Hörmann and Kokoszka (2010), we get

\[
\sup_{r \in [0,1]} \frac{1}{\sqrt{N}} \left| S_N(1, \lfloor Nr \rfloor, \hat{\beta}) \right|
\leq \sup_{r \in [0,1]} \frac{1}{\sqrt{N}} \left| S_N(1, \lfloor Nr \rfloor, \hat{\beta}) - S_N(1, \lfloor Nr \rfloor, \eta) \right|
+ \sup_{r \in [0,1]} \frac{1}{\sqrt{N}} \left| S_N(1, \lfloor Nr \rfloor, \eta) \right| = o_p(\log \log N).
\]

Note that \(|\hat{\mathbf{C}} \hat{\Delta}_N - \Delta_K^{(N)}| = o_p(\|\Delta_K^{(N)}\|)\) provided that \(\|\Delta_K^{(N)}\| = O(\|\Delta_K^{(N)}\|)\) and \(\|\hat{c}_i \phi_i - v_i^{(N)}\| \leq \|\hat{c}_i \phi_i - \phi_i\| + \|\phi_i - v_i^{(N)}\| = o_p(1)\). Under the assumption that \(\liminf_{N \to \infty} N^{1/2} |\Delta_K^{(N)}| > 0\), we have \(|\hat{\mathbf{C}} T_N, \hat{\eta}(k^*, K) | = O_p(N^{1/2} |\Delta_K^{(N)}|)\) and \(|\hat{\mathbf{C}} V_N, \hat{\eta}(k^*, K) \hat{\mathbf{C}}' M| = o_p((\log \log N)^2)\) (see 2.25 and 2.26). Therefore we get

\[
G_{N, \hat{\eta}}(K) \geq |T_{N, \hat{\eta}}(k^*, K)|^2 / \|V_{N, \hat{\eta}}(k^*, K)\|_M,
\]

where the right hand side diverges to infinity as \(N \to +\infty\). \qed
Table 2.1: Empirical size (upper panel) and size-adjusted power (lower panel) in percentage for the SN-based test (in row (i)) and the BGHK test (in row (ii)) for independent functional data generated from BM or BB. The size-adjusted power is computed under the alternative (2.2) with $\mu_2(t) = t$ or $\mu_2(t) = \sin(t)$, and $k^* = N/2$. The sample size $N = 50, 100$, and the number of PCs $K = 1, 2, 3$. The number of Monte Carlo replications is 1000.

<table>
<thead>
<tr>
<th></th>
<th>$K = 1$</th>
<th>$K = 2$</th>
<th>$K = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10% 5% 1%</td>
<td>10% 5% 1%</td>
<td>10% 5% 1%</td>
</tr>
<tr>
<td>$N = 50$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>10.7 5.7 0.7</td>
<td>9.6 3.7 0.7</td>
<td>10.8 5.2 1.4</td>
</tr>
<tr>
<td>(ii)</td>
<td>10.0 5.3 1.2</td>
<td>10.3 5.0 0.8</td>
<td>10.9 5.5 1.0</td>
</tr>
<tr>
<td>BB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>7.5 3.8 0.8</td>
<td>8.2 4.6 1.1</td>
<td>10.7 6.0 1.3</td>
</tr>
<tr>
<td>(ii)</td>
<td>10.6 5.4 0.8</td>
<td>10.9 5.1 1.1</td>
<td>10.5 5.2 1.2</td>
</tr>
<tr>
<td>$N = 100$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>9.9 5.1 1.1</td>
<td>9.2 4.3 0.5</td>
<td>9.1 4.6 0.7</td>
</tr>
<tr>
<td>(ii)</td>
<td>10.4 5.4 0.5</td>
<td>10.3 4.5 0.6</td>
<td>9.5 3.8 0.6</td>
</tr>
<tr>
<td>BB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>10.0 5.1 1.3</td>
<td>8.4 3.5 0.7</td>
<td>9.9 4.7 0.7</td>
</tr>
<tr>
<td>(ii)</td>
<td>9.6 5.2 0.9</td>
<td>9.3 4.9 0.6</td>
<td>9.1 4.1 0.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$K = 1$</th>
<th>$K = 2$</th>
<th>$K = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10% 5% 1%</td>
<td>10% 5% 1%</td>
<td>10% 5% 1%</td>
</tr>
<tr>
<td>$N = 50$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM, $t$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>77.6 64.5 44.9</td>
<td>71.7 58.4 39.4</td>
<td>67.4 51.7 23.8</td>
</tr>
<tr>
<td>(ii)</td>
<td>89.5 79.8 48.9</td>
<td>83.6 73.7 48.9</td>
<td>77.8 65.4 38.8</td>
</tr>
<tr>
<td>BB, $t$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>99.8 99.4 95.6</td>
<td>100 100 99.6</td>
<td>100 100 99.9</td>
</tr>
<tr>
<td>(ii)</td>
<td>100 100 99.7</td>
<td>100 100 100</td>
<td>100 100 100</td>
</tr>
<tr>
<td>BM, $\sin(t)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>70.0 57.7 38.9</td>
<td>62.1 48.3 29.1</td>
<td>56.0 41.4 17.0</td>
</tr>
<tr>
<td>(ii)</td>
<td>82.1 71.9 39.4</td>
<td>74.4 61.4 36.4</td>
<td>66.9 52.4 28.7</td>
</tr>
<tr>
<td>BB, $\sin(t)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>99.3 98.1 89.7</td>
<td>100 99.6 96.9</td>
<td>100 99.9 99.4</td>
</tr>
<tr>
<td>(ii)</td>
<td>99.9 99.7 97.6</td>
<td>100 100 100</td>
<td>100 100 100</td>
</tr>
<tr>
<td>$N = 100$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM, $t$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>96.9 89.9 70.8</td>
<td>92.9 87.4 73.0</td>
<td>90.9 84.0 66.7</td>
</tr>
<tr>
<td>(ii)</td>
<td>99.3 98.4 95.5</td>
<td>99.1 97.9 94.0</td>
<td>98.5 96.8 91.2</td>
</tr>
<tr>
<td>BB, $t$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>100 99.9 99.6</td>
<td>100 100 100</td>
<td>100 100 100</td>
</tr>
<tr>
<td>(ii)</td>
<td>100 100 100</td>
<td>100 100 100</td>
<td>100 100 100</td>
</tr>
<tr>
<td>BM, $\sin(t)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>92.7 84.2 62.7</td>
<td>87.1 78.7 59.2</td>
<td>83.9 73.8 52.1</td>
</tr>
<tr>
<td>(ii)</td>
<td>98.4 95.8 89.6</td>
<td>96.3 93.5 86.6</td>
<td>95.2 90.9 78.0</td>
</tr>
<tr>
<td>BB, $\sin(t)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>99.9 99.7 98.8</td>
<td>100 100 100</td>
<td>100 100 100</td>
</tr>
<tr>
<td>(ii)</td>
<td>100 100 100</td>
<td>100 100 100</td>
<td>100 100 100</td>
</tr>
</tbody>
</table>
Table 2.2: Empirical size in percentage of the SN-based test (in row (i)), the BGHK test (in row (ii)) and the HK test (in row (iii)) for temporally dependent functional data generated from ARH(1) process. The sample size $N = 50, 100$, and the number of PCs $K = 1, 2, 3$. The number of Monte Carlo replications is 1000.

<table>
<thead>
<tr>
<th></th>
<th>$K = 1$</th>
<th></th>
<th></th>
<th>$K = 2$</th>
<th></th>
<th></th>
<th>$K = 3$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>$N = 50$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaussian, BM</td>
<td>(i) 15.2</td>
<td>10.3</td>
<td>3.9</td>
<td>15.2</td>
<td>8.4</td>
<td>2.4</td>
<td>14.5</td>
<td>8.0</td>
<td>2.2</td>
</tr>
<tr>
<td>(ii)</td>
<td>44.1</td>
<td>32.2</td>
<td>16.2</td>
<td>37.5</td>
<td>25.0</td>
<td>12.4</td>
<td>32.7</td>
<td>23.0</td>
<td>11.4</td>
</tr>
<tr>
<td>(iii)</td>
<td>17.7</td>
<td>9.2</td>
<td>0.6</td>
<td>11.1</td>
<td>3.2</td>
<td>0.3</td>
<td>4.9</td>
<td>1.1</td>
<td>0.0</td>
</tr>
<tr>
<td>Gaussian, BB</td>
<td>(i) 17.3</td>
<td>10.6</td>
<td>3.1</td>
<td>14.0</td>
<td>7.1</td>
<td>2.5</td>
<td>14.5</td>
<td>8.2</td>
<td>2.3</td>
</tr>
<tr>
<td>(ii)</td>
<td>42.7</td>
<td>32.3</td>
<td>13.8</td>
<td>36.1</td>
<td>25.5</td>
<td>10.0</td>
<td>34.9</td>
<td>23.2</td>
<td>9.6</td>
</tr>
<tr>
<td>(iii)</td>
<td>19.8</td>
<td>8.8</td>
<td>0.2</td>
<td>11.1</td>
<td>2.6</td>
<td>0.0</td>
<td>6.3</td>
<td>1.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Wiener, BM</td>
<td>(i) 16.0</td>
<td>10.4</td>
<td>4.0</td>
<td>16.1</td>
<td>9.2</td>
<td>3.0</td>
<td>16.0</td>
<td>9.8</td>
<td>2.9</td>
</tr>
<tr>
<td>(ii)</td>
<td>46.4</td>
<td>33.6</td>
<td>16.6</td>
<td>40.2</td>
<td>26.9</td>
<td>12.6</td>
<td>36.6</td>
<td>25.0</td>
<td>10.2</td>
</tr>
<tr>
<td>(iii)</td>
<td>17.5</td>
<td>8.4</td>
<td>0.5</td>
<td>10.9</td>
<td>2.8</td>
<td>0.1</td>
<td>6.1</td>
<td>0.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Wiener, BB</td>
<td>(i) 17.0</td>
<td>10.4</td>
<td>2.9</td>
<td>13.3</td>
<td>7.3</td>
<td>2.2</td>
<td>15.4</td>
<td>9.7</td>
<td>2.2</td>
</tr>
<tr>
<td>(ii)</td>
<td>42.8</td>
<td>31.0</td>
<td>14.3</td>
<td>37.9</td>
<td>26.7</td>
<td>11.2</td>
<td>36.4</td>
<td>23.9</td>
<td>10.0</td>
</tr>
<tr>
<td>(iii)</td>
<td>19.0</td>
<td>8.9</td>
<td>0.2</td>
<td>11.2</td>
<td>3.3</td>
<td>0.0</td>
<td>6.5</td>
<td>1.8</td>
<td>0.0</td>
</tr>
<tr>
<td>$N = 100$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaussian, BM</td>
<td>(i) 13.3</td>
<td>7.8</td>
<td>2.0</td>
<td>11.7</td>
<td>5.7</td>
<td>1.2</td>
<td>11.7</td>
<td>6.1</td>
<td>1.2</td>
</tr>
<tr>
<td>(ii)</td>
<td>51.2</td>
<td>35.9</td>
<td>16.4</td>
<td>39.7</td>
<td>27.9</td>
<td>11.6</td>
<td>34.9</td>
<td>24.1</td>
<td>9.7</td>
</tr>
<tr>
<td>(iii)</td>
<td>15.2</td>
<td>7.4</td>
<td>0.4</td>
<td>11.6</td>
<td>3.9</td>
<td>0.2</td>
<td>7.2</td>
<td>2.1</td>
<td>0.0</td>
</tr>
<tr>
<td>Gaussian, BB</td>
<td>(i) 11.6</td>
<td>6.7</td>
<td>1.6</td>
<td>10.9</td>
<td>4.9</td>
<td>1.1</td>
<td>11.5</td>
<td>7.1</td>
<td>1.2</td>
</tr>
<tr>
<td>(ii)</td>
<td>46.7</td>
<td>33.0</td>
<td>13.9</td>
<td>35.9</td>
<td>25.1</td>
<td>10.2</td>
<td>36.4</td>
<td>25.8</td>
<td>11.4</td>
</tr>
<tr>
<td>(iii)</td>
<td>16.1</td>
<td>8.0</td>
<td>1.4</td>
<td>12.4</td>
<td>5.3</td>
<td>0.3</td>
<td>10.0</td>
<td>3.5</td>
<td>0.1</td>
</tr>
<tr>
<td>Wiener, BM</td>
<td>(i) 13.7</td>
<td>7.8</td>
<td>2.1</td>
<td>11.7</td>
<td>5.8</td>
<td>1.3</td>
<td>12.9</td>
<td>7.1</td>
<td>1.3</td>
</tr>
<tr>
<td>(ii)</td>
<td>52.2</td>
<td>37.2</td>
<td>17.5</td>
<td>43.8</td>
<td>29.7</td>
<td>12.8</td>
<td>38.3</td>
<td>26.1</td>
<td>11.7</td>
</tr>
<tr>
<td>(iii)</td>
<td>15.3</td>
<td>7.4</td>
<td>0.5</td>
<td>11.6</td>
<td>4.1</td>
<td>0.2</td>
<td>7.5</td>
<td>2.6</td>
<td>0.0</td>
</tr>
<tr>
<td>Wiener, BB</td>
<td>(i) 11.9</td>
<td>6.4</td>
<td>1.9</td>
<td>10.4</td>
<td>5.6</td>
<td>1.2</td>
<td>12.0</td>
<td>7.8</td>
<td>1.3</td>
</tr>
<tr>
<td>(ii)</td>
<td>45.1</td>
<td>32.0</td>
<td>13.5</td>
<td>38.5</td>
<td>27.5</td>
<td>12.8</td>
<td>37.9</td>
<td>27.3</td>
<td>11.9</td>
</tr>
<tr>
<td>(iii)</td>
<td>16.4</td>
<td>7.6</td>
<td>1.5</td>
<td>13.3</td>
<td>5.9</td>
<td>0.5</td>
<td>10.8</td>
<td>3.7</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Table 2.3: Size-adjusted power in percentage of the SN-based test (in row (i)), the BGHK test (in row (ii)) and the HK test (in row (iii)) for temporally dependent functional data generated from ARH(1) process. The size-adjusted power is computed under the alternative (2.2) with $\mu_2(t) = t$ or $\mu_2(t) = \sin(t)$, and $k^* = N/2$. The sample size $N = 50, 100$, and the number of PCs $K = 1, 2, 3$.

The number of Monte Carlo replications is 1000.

<table>
<thead>
<tr>
<th></th>
<th>$N = 50$</th>
<th>$N = 100$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K = 1$</td>
<td>$K = 2$</td>
</tr>
<tr>
<td></td>
<td>10% 5% 1%</td>
<td>10% 5% 1%</td>
</tr>
<tr>
<td>Gaussian, BM, $t$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>35.7 22.5 5.0</td>
<td>48.6 34.7 11.4</td>
</tr>
<tr>
<td>(ii)</td>
<td>44.0 25.0 11.8</td>
<td>46.2 28.2 11.8</td>
</tr>
<tr>
<td>(iii)</td>
<td>36.1 21.6 7.4</td>
<td>46.9 32.4 7.6</td>
</tr>
<tr>
<td>Gaussian, BB, $t$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>75.3 66.8 33.2</td>
<td>99.3 96.9 84.9</td>
</tr>
<tr>
<td>(ii)</td>
<td>89.9 83.0 69.4</td>
<td>99.8 90.7 97.6</td>
</tr>
<tr>
<td>(iii)</td>
<td>65.1 40.6 11.0</td>
<td>79.6 53.1 23.3</td>
</tr>
<tr>
<td>Gaussian, BM, $\sin(t)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>31.8 18.5 4.1</td>
<td>36.3 24.8 7.4</td>
</tr>
<tr>
<td>(ii)</td>
<td>38.3 20.7 8.9</td>
<td>36.7 21.3 7.7</td>
</tr>
<tr>
<td>(iii)</td>
<td>31.6 19.2 5.1</td>
<td>37.0 25.0 4.4</td>
</tr>
<tr>
<td>Gaussian, BB, $\sin(t)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>75.3 66.8 33.2</td>
<td>99.3 96.9 84.9</td>
</tr>
<tr>
<td>(ii)</td>
<td>89.9 83.0 69.4</td>
<td>99.8 90.7 97.6</td>
</tr>
<tr>
<td>(iii)</td>
<td>65.1 40.6 11.0</td>
<td>79.6 53.1 23.3</td>
</tr>
<tr>
<td>Wiener, BM, $t$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>34.3 20.0 4.7</td>
<td>35.5 25.1 7.8</td>
</tr>
<tr>
<td>(ii)</td>
<td>42.6 23.8 11.0</td>
<td>42.1 26.2 13.3</td>
</tr>
<tr>
<td>(iii)</td>
<td>32.6 21.4 6.0</td>
<td>33.8 21.4 6.0</td>
</tr>
<tr>
<td>Wiener, BB, $t$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>76.3 68.1 39.2</td>
<td>95.4 89.5 73.1</td>
</tr>
<tr>
<td>(ii)</td>
<td>89.4 83.0 71.6</td>
<td>99.4 98.8 94.9</td>
</tr>
<tr>
<td>(iii)</td>
<td>63.5 41.3 13.1</td>
<td>72.8 45.5 9.8</td>
</tr>
<tr>
<td>Wiener, BM, $\sin(t)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>30.1 17.7 3.7</td>
<td>29.7 21.1 5.0</td>
</tr>
<tr>
<td>(ii)</td>
<td>37.1 19.6 8.3</td>
<td>34.5 20.8 9.2</td>
</tr>
<tr>
<td>(iii)</td>
<td>29.4 18.5 4.7</td>
<td>28.3 16.8 2.7</td>
</tr>
<tr>
<td>Wiener, BB, $\sin(t)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>69.2 59.7 32.5</td>
<td>88.4 77.9 58.4</td>
</tr>
<tr>
<td>(ii)</td>
<td>83.8 76.6 59.9</td>
<td>96.6 92.5 84.1</td>
</tr>
<tr>
<td>(iii)</td>
<td>61.1 41.4 13.7</td>
<td>70.3 44.6 11.5</td>
</tr>
<tr>
<td>Wiener, BM, $t$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>51.4 36.8 17.4</td>
<td>55.1 43.3 18.1</td>
</tr>
<tr>
<td>(ii)</td>
<td>63.4 47.3 29.8</td>
<td>69.9 57.9 35.4</td>
</tr>
<tr>
<td>(iii)</td>
<td>57.9 44.8 19.6</td>
<td>58.0 45.0 20.0</td>
</tr>
<tr>
<td>Wiener, BB, $t$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>95.1 89.7 70.5</td>
<td>100 99.9 95.8</td>
</tr>
<tr>
<td>(ii)</td>
<td>99.2 98.2 91.4</td>
<td>100 100 100</td>
</tr>
<tr>
<td>(iii)</td>
<td>96.6 87.6 38.1</td>
<td>99.0 87.7 44.9</td>
</tr>
<tr>
<td>Wiener, BM, $\sin(t)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>44.3 31.2 13.7</td>
<td>44.5 32.9 12.3</td>
</tr>
<tr>
<td>(ii)</td>
<td>56.9 45.2 24.4</td>
<td>58.3 45.1 26.0</td>
</tr>
<tr>
<td>(iii)</td>
<td>52.1 37.3 15.9</td>
<td>47.1 33.2 12.7</td>
</tr>
<tr>
<td>Wiener, BB, $\sin(t)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>92.2 85.1 60.2</td>
<td>99.6 97.2 89.1</td>
</tr>
<tr>
<td>(ii)</td>
<td>98.3 95.8 84.8</td>
<td>100 99.7 97.4</td>
</tr>
<tr>
<td>(iii)</td>
<td>93.9 84.2 40.7</td>
<td>97.8 86.1 47.3</td>
</tr>
</tbody>
</table>
Table 2.4: Empirical size (upper panel) and size-adjusted power (lower panel) in percentage of the SN-based test (in row (i)) and the tests in Horváth et al. (2010) (in row (ii) for case 1 and (iii) for case 2 ) for detecting the change point in the lag one autocovariance operator. The size-adjusted power is computed under the alternative (2.20). The sample size $N = 50, 100,$ and the number of PCs $K = 1, 2, 3$. The number of Monte Carlo replications is 1000.

<table>
<thead>
<tr>
<th></th>
<th>$K = 1$</th>
<th></th>
<th>$K = 2$</th>
<th></th>
<th>$K = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>$N = 50$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>11.0</td>
<td>5.8</td>
<td>1.6</td>
<td>8.6</td>
<td>3.8</td>
</tr>
<tr>
<td>(ii)</td>
<td>16.4</td>
<td>9.5</td>
<td>2.2</td>
<td>33.6</td>
<td>21.8</td>
</tr>
<tr>
<td>(iii)</td>
<td>10.6</td>
<td>3.0</td>
<td>0.3</td>
<td>4.5</td>
<td>1.4</td>
</tr>
<tr>
<td>BB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>8.5</td>
<td>4.1</td>
<td>0.8</td>
<td>8.4</td>
<td>4.4</td>
</tr>
<tr>
<td>(ii)</td>
<td>15.8</td>
<td>9.3</td>
<td>2.2</td>
<td>32.2</td>
<td>21.6</td>
</tr>
<tr>
<td>(iii)</td>
<td>9.4</td>
<td>3.8</td>
<td>0.4</td>
<td>4.5</td>
<td>1.3</td>
</tr>
<tr>
<td>$N = 100$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>9.2</td>
<td>5.4</td>
<td>1.4</td>
<td>8.5</td>
<td>4.4</td>
</tr>
<tr>
<td>(ii)</td>
<td>14.1</td>
<td>7.1</td>
<td>1.5</td>
<td>16.9</td>
<td>10.6</td>
</tr>
<tr>
<td>(iii)</td>
<td>10.0</td>
<td>3.8</td>
<td>0.7</td>
<td>6.4</td>
<td>2.7</td>
</tr>
<tr>
<td>BB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>9.4</td>
<td>5.0</td>
<td>0.8</td>
<td>8.2</td>
<td>3.0</td>
</tr>
<tr>
<td>(ii)</td>
<td>14.0</td>
<td>7.1</td>
<td>1.6</td>
<td>20.4</td>
<td>12.1</td>
</tr>
<tr>
<td>(iii)</td>
<td>10.7</td>
<td>5.1</td>
<td>0.7</td>
<td>8.9</td>
<td>2.7</td>
</tr>
<tr>
<td>$N = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>8.9</td>
<td>4.3</td>
<td>1.1</td>
<td>10.3</td>
<td>5.4</td>
</tr>
<tr>
<td>(ii)</td>
<td>11.4</td>
<td>6.2</td>
<td>1.4</td>
<td>15.3</td>
<td>7.8</td>
</tr>
<tr>
<td>(iii)</td>
<td>9.4</td>
<td>5.6</td>
<td>1.0</td>
<td>9.1</td>
<td>3.4</td>
</tr>
<tr>
<td>BB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>10.0</td>
<td>5.4</td>
<td>1.3</td>
<td>8.6</td>
<td>4.5</td>
</tr>
<tr>
<td>(ii)</td>
<td>12.3</td>
<td>5.2</td>
<td>1.3</td>
<td>14.7</td>
<td>7.5</td>
</tr>
<tr>
<td>(iii)</td>
<td>10.4</td>
<td>3.9</td>
<td>0.7</td>
<td>8.3</td>
<td>3.0</td>
</tr>
<tr>
<td>$N = 50$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>27.3</td>
<td>17.2</td>
<td>5.2</td>
<td>22.5</td>
<td>14.5</td>
</tr>
<tr>
<td>(ii)</td>
<td>37.8</td>
<td>26.6</td>
<td>12.5</td>
<td>36.7</td>
<td>26.0</td>
</tr>
<tr>
<td>(iii)</td>
<td>27.0</td>
<td>15.6</td>
<td>2.6</td>
<td>11.8</td>
<td>5.4</td>
</tr>
<tr>
<td>BB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>33.4</td>
<td>21.7</td>
<td>10.0</td>
<td>23.2</td>
<td>14.1</td>
</tr>
<tr>
<td>(ii)</td>
<td>37.2</td>
<td>27.6</td>
<td>13.3</td>
<td>32.1</td>
<td>24.2</td>
</tr>
<tr>
<td>(iii)</td>
<td>32.9</td>
<td>17.6</td>
<td>2.1</td>
<td>17.0</td>
<td>5.4</td>
</tr>
<tr>
<td>$N = 100$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>46.6</td>
<td>29.0</td>
<td>9.8</td>
<td>35.6</td>
<td>22.1</td>
</tr>
<tr>
<td>(ii)</td>
<td>60.0</td>
<td>47.8</td>
<td>23.6</td>
<td>56.7</td>
<td>44.2</td>
</tr>
<tr>
<td>(iii)</td>
<td>50.1</td>
<td>32.8</td>
<td>5.3</td>
<td>20.5</td>
<td>7.9</td>
</tr>
<tr>
<td>BB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>44.7</td>
<td>30.7</td>
<td>12.9</td>
<td>32.9</td>
<td>24.6</td>
</tr>
<tr>
<td>(ii)</td>
<td>55.5</td>
<td>42.2</td>
<td>20.8</td>
<td>46.8</td>
<td>36.9</td>
</tr>
<tr>
<td>(iii)</td>
<td>45.5</td>
<td>26.2</td>
<td>6.7</td>
<td>14.2</td>
<td>5.3</td>
</tr>
<tr>
<td>$N = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>70.2</td>
<td>57.8</td>
<td>26.8</td>
<td>54.0</td>
<td>38.0</td>
</tr>
<tr>
<td>(ii)</td>
<td>87.9</td>
<td>74.8</td>
<td>46.4</td>
<td>78.9</td>
<td>69.0</td>
</tr>
<tr>
<td>(iii)</td>
<td>80.5</td>
<td>54.7</td>
<td>21.5</td>
<td>39.8</td>
<td>23.9</td>
</tr>
<tr>
<td>BB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i)</td>
<td>65.0</td>
<td>47.9</td>
<td>20.9</td>
<td>45.9</td>
<td>30.1</td>
</tr>
<tr>
<td>(ii)</td>
<td>83.3</td>
<td>72.2</td>
<td>45.5</td>
<td>69.4</td>
<td>58.1</td>
</tr>
<tr>
<td>(iii)</td>
<td>74.2</td>
<td>63.3</td>
<td>29.8</td>
<td>37.2</td>
<td>19.1</td>
</tr>
</tbody>
</table>
Table 2.5: (a) Empirical size (upper panel) and size-adjusted power (lower panel) in percentage of the SN-based test for detecting the mean change of 2-d functional observations generated by tensor product. (b) Empirical size (upper panel) and size-adjusted power (lower panel) in percentage of the SN-based test for detecting the mean change of 2-d functional observations generated from the ARH_2(1) process. The sample size \( N = 50, 100, \) and the number of PCs \( K = 1, 2, 3. \) The number of Monte Carlo replications is 1000. The notation \((BM + t)^2\) denotes \((BM + t) \times (BM + t).\) Note: The Brownian motion is approximated by the partial sum of 100 iid standard normal variables in (a) and it is approximated by the partial sum of 60 iid standard normal variables in (b).

<table>
<thead>
<tr>
<th></th>
<th>( K = 1 )</th>
<th>( K = 2 )</th>
<th>( K = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>( N = 50 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM \times BM</td>
<td>6.8</td>
<td>3.6</td>
<td>0.7</td>
</tr>
<tr>
<td>BB \times BB</td>
<td>8.4</td>
<td>4.4</td>
<td>0.9</td>
</tr>
<tr>
<td>ARH(1) \times ARH(1)</td>
<td>10.2</td>
<td>4.4</td>
<td>0.4</td>
</tr>
<tr>
<td>( N = 100 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM \times BM</td>
<td>10.2</td>
<td>4.9</td>
<td>1.0</td>
</tr>
<tr>
<td>BB \times BB</td>
<td>7.6</td>
<td>3.7</td>
<td>1.0</td>
</tr>
<tr>
<td>ARH(1) \times ARH(1)</td>
<td>9.6</td>
<td>5.0</td>
<td>0.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>( K = 1 )</th>
<th>( K = 2 )</th>
<th>( K = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>( N = 50 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(BM + t)^2</td>
<td>62.4</td>
<td>47.0</td>
<td>20.7</td>
</tr>
<tr>
<td>(BB + t)^2</td>
<td>99.8</td>
<td>99.1</td>
<td>95.5</td>
</tr>
<tr>
<td>(BM + sin(t))^2</td>
<td>52.2</td>
<td>37.8</td>
<td>14.8</td>
</tr>
<tr>
<td>(BB + sin(t))^2</td>
<td>99.2</td>
<td>97.1</td>
<td>89.0</td>
</tr>
<tr>
<td>(ARH(1) + t)^2</td>
<td>31.1</td>
<td>22.1</td>
<td>10.1</td>
</tr>
<tr>
<td>(ARH(1) + sin(t))^2</td>
<td>26.0</td>
<td>17.4</td>
<td>7.2</td>
</tr>
<tr>
<td>( N = 100 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(BM + t)^2</td>
<td>77.6</td>
<td>67.8</td>
<td>41.6</td>
</tr>
<tr>
<td>(BB + t)^2</td>
<td>100</td>
<td>100</td>
<td>99.6</td>
</tr>
<tr>
<td>(BM + sin(t))^2</td>
<td>67.9</td>
<td>55.4</td>
<td>28.5</td>
</tr>
<tr>
<td>(BB + sin(t))^2</td>
<td>100</td>
<td>100</td>
<td>98.6</td>
</tr>
<tr>
<td>(ARH(1) + t)^2</td>
<td>43.7</td>
<td>31.1</td>
<td>15.3</td>
</tr>
<tr>
<td>(ARH(1) + sin(t))^2</td>
<td>35.5</td>
<td>25.2</td>
<td>10.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>( K = 1 )</th>
<th>( K = 2 )</th>
<th>( K = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>( N = 50 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM \times BM, t</td>
<td>14.8</td>
<td>9.5</td>
<td>2.9</td>
</tr>
<tr>
<td>BB \times BB, t</td>
<td>13.1</td>
<td>7.0</td>
<td>1.5</td>
</tr>
<tr>
<td>( N = 100 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM \times BM</td>
<td>10.4</td>
<td>5.9</td>
<td>1.4</td>
</tr>
<tr>
<td>BB \times BB</td>
<td>12.7</td>
<td>6.7</td>
<td>1.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>( K = 1 )</th>
<th>( K = 2 )</th>
<th>( K = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>( N = 50 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM \times BM, t</td>
<td>31.6</td>
<td>21.5</td>
<td>7.3</td>
</tr>
<tr>
<td>BB \times BB, t</td>
<td>99.9</td>
<td>99.7</td>
<td>98.6</td>
</tr>
<tr>
<td>BM \times BM, sin(t)</td>
<td>25.3</td>
<td>16.4</td>
<td>5.1</td>
</tr>
<tr>
<td>BB \times BB, sin(t)</td>
<td>99.0</td>
<td>98.1</td>
<td>94.0</td>
</tr>
<tr>
<td>( N = 100 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM \times BM</td>
<td>58.0</td>
<td>41.6</td>
<td>21.1</td>
</tr>
<tr>
<td>BB \times BB</td>
<td>100</td>
<td>100</td>
<td>99.7</td>
</tr>
<tr>
<td>BM \times BM, sin(t)</td>
<td>46.4</td>
<td>30.6</td>
<td>12.8</td>
</tr>
<tr>
<td>BB \times BB, sin(t)</td>
<td>99.9</td>
<td>99.7</td>
<td>97.7</td>
</tr>
</tbody>
</table>
Table 2.6: Segmentation procedure of the central England temperature data into periods with constant mean function. Note: in each iteration, $K$ is the smallest positive integer such that $\sum_{i=1}^{K} \hat{\lambda}_i / \sum_{i=1}^{2} \hat{\lambda}_i > 0.8$.

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Segment</th>
<th>$K$</th>
<th>$G_{N,\hat{g}}(K)$</th>
<th>$p$-value</th>
<th>Estimated change-point $k^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1780–2007</td>
<td>8</td>
<td>559.4</td>
<td>(0.001, 0.005)</td>
<td>1927</td>
</tr>
<tr>
<td>2</td>
<td>1780–1927</td>
<td>8</td>
<td>173.1</td>
<td>(0.1, 1)</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td>1928–2007</td>
<td>8</td>
<td>323.9</td>
<td>(0.025, 0.05)</td>
<td>1993</td>
</tr>
<tr>
<td>4</td>
<td>1928–1993</td>
<td>7</td>
<td>49.2</td>
<td>(0.1, 1)</td>
<td>—</td>
</tr>
<tr>
<td>5</td>
<td>1994–2007</td>
<td>5</td>
<td>153.0</td>
<td>(0.05, 0.1)</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 2.7: Test statistics and their $p$-values for the spatio-temporal temperature data covering a subregion of North America. Panel (a) shows the results of the tests based on the first $K$ PCs; Panel (b) shows the results of the tests based on individual PC.

<table>
<thead>
<tr>
<th>(a)</th>
<th>$K$</th>
<th>$G_{N,\hat{g}}(K)$</th>
<th>$p$-value</th>
<th>Estimated change-point $k^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>25.2</td>
<td>(0.1, 1)</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>34.4</td>
<td>(0.1, 1)</td>
<td>—</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>160.5</td>
<td>(0.005, 0.01)</td>
<td>1990</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>182.7</td>
<td>(0.01, 0.025)</td>
<td>1991</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>218.2</td>
<td>(0.025, 0.05)</td>
<td>1991</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>221.9</td>
<td>(0.025, 0.05)</td>
<td>1991</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b)</th>
<th>PC</th>
<th>$G_{N,\hat{g}}(K)$</th>
<th>$p$-value</th>
<th>Estimated change-point $k^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1</td>
<td></td>
<td>25.2</td>
<td>(0.1, 1)</td>
<td>—</td>
</tr>
<tr>
<td>PC2</td>
<td></td>
<td>10.0</td>
<td>(0.1, 1)</td>
<td>—</td>
</tr>
<tr>
<td>PC3</td>
<td></td>
<td>93.7</td>
<td>(0.001, 0.005)</td>
<td>1990</td>
</tr>
<tr>
<td>PC4</td>
<td></td>
<td>3.1</td>
<td>(0.1, 1)</td>
<td>—</td>
</tr>
<tr>
<td>PC5</td>
<td></td>
<td>2.5</td>
<td>(0.1, 1)</td>
<td>—</td>
</tr>
<tr>
<td>PC6</td>
<td></td>
<td>3.6</td>
<td>(0.1, 1)</td>
<td>—</td>
</tr>
</tbody>
</table>
Figure 2.1: Size-adjusted power for detecting the mean change with different magnitude of change. Sample size $N = 50$. Note: the quantity $\delta$ measures the magnitude of change; see equation (2.19).
Figure 2.2: Average daily temperature functions in the three estimated partition segments.
Figure 2.3: The first six principal components of the bias between observations and model output. The number in the title of each figure stands for the percentage of variation explained by the corresponding principal component.
Figure 2.4: Average biases in two periods (1980-1990; 1991-1998) and the change in the bias.
Chapter 3

Two sample inference for temporally dependent functional data

3.1 Introduction

Functional data analysis (FDA) which deals with the analysis of curves and surfaces has received considerable attention in the statistical literature during the last decade (Ramsay and Silverman, 2002, 2005; Ferraty and Vieu, 2006). This chapter falls into a sub-field of functional data analysis: inference for temporally dependent functional data. Specifically, we focus on testing the equality of the second order structures (e.g., the covariance operators and their associated eigenvalues and eigenfunctions) of two temporally dependent functional sequences. Our work is partially motivated by our ongoing collaboration with atmospheric scientists on the development and assessment of high-resolution climate projections through statistical downscaling. Climate change is one of the most urgent problems facing the world this century. To study climate change, scientists have relied primarily on climate projections from global/regional climate models, which are numerical models that involve systems of differential equations and produce outputs at a prespecified grid. As numerical model outputs are widely used in situations where real observations are not available, it is an important but still open question whether the numerical model outputs are able to mimic/capture the spatial and temporal dynamics of the real observations. To partly answer this question, we view the spatio-temporal model outputs and real observations as realizations from two temporally dependent functional time series defined on the two dimensional space and test the equality of their second order structures which reflects their spatial dynamics/dependence.

Two sample inference for functional data has been investigated by a few researchers. Fan and Lin (1998), Cuevas et al. (2004) and Horváth et al. (2011) developed the tests for the equality of mean functions. Benko et al. (2009), Panaretos et al. (2010), Fremdt et al. (2011), and Kraus and Panaretos (2012) proposed tests for the equality of the second order structures. All the above-mentioned works assumed the independence between the two samples and/or independence within each sample. However, the assumption of independence within the sample is often too strong to be
realistic in many applications, especially if data are collected sequentially over time. For example, the independence assumption is questionable for the climate projection data considered in this chapter, as the model outputs and real station observations are simulated or collected over time and temporal dependence is expected. Furthermore the dependence between numerical model outputs and station observations is likely because the numerical models are designed to mimic the dynamics of real observations. See Section 3.5 for empirical evidence of their dependence. In this chapter, we develop new tests that are able to accommodate weak dependence between and within two samples. Our tests are constructed on the basis of functional principal component analysis (FPCA) and the recently developed self-normalization (SN) method (Shao, 2010), the latter of which is a new studentization technique for the inference of a univariate time series.

FPCA attempts to find the dominant modes of variation around an overall trend function and has been proved a key technique in the context of FDA. The use of FPCA in the inference of temporally dependent functional data can be found in Gabrys and Kokoszka (2007), Hörmann and Kokoszka (2010), Horváth et al. (2011) among others. To account for the dependence, the existing inference procedure requires a consistent estimator of the long run variance (LRV) matrix of the principal component scores or consistent estimator of the LRV operator. However, there is a bandwidth parameter involved in the LRV estimation and its selection has not been addressed in the functional setting. The same issue appears when one considers the block bootstrapping and subsampling schemes (Lahiri, 2003; Politis et al., 1999), since these techniques also require the selection of a smoothing parameter, such as the block length in the moving block bootstrap, and the window width in the subsampling method (see e.g., Politis and Romano, 2010; McMurry and Politis, 2011). Since the finite sample performance can be sensitive to the choice of these tuning parameters and the bandwidth choice can involve certain degree of arbitrariness, it is desirable to use inference methods that are free of bandwidth parameters. To this end, we build on the bandwidth-free SN method (Shao, 2010) recently developed in the univariate time series setup, and propose SN-based tests in the functional setting by using recursive estimates obtained from functional data samples.

The generalization of the SN method from univariate to functional time series was first done in Zhang et al. (2011) where the focus was on testing the structure stability of temporally dependent functional data. Here we extend the SN method to test the equality of the second order properties of two functional time series, which is rather different and new techniques and results are needed. To study the asymptotic properties of the proposed test statistics, we establish functional central limit theorems for the recursive estimates of quantities associated with the second order properties.
of the functional time series which seems unexplored in the literature and are thus of independent interest. Based on the functional central limit theorem, we show that the SN-based test statistics have pivotal limiting distributions under the null and are consistent under the local alternatives. From a methodological viewpoint, this seems to be the first time that the SN method is extended to the two sample problem. Compared to most of the existing methods which assumed the independence between the two samples and/or independence within each sample, the SN method not only allows for unknown dependence within each sample but also allows for unknown dependence between the two samples when the sample sizes of the two sequences are equal.

3.2 Methodology

We shall consider temporally dependent functional processes \( \{(X_i(t), Y_i(t)), t \in \mathcal{I}\}_{i=1}^{+\infty} \) defined on some compact set \( \mathcal{I} \) of the Euclidian space, where \( \mathcal{I} \) can be one dimensional (e.g., a curve) or multidimensional (e.g., a surface or manifold). For simplicity, we consider the Hilbert space \( \mathbb{H} \) of square integrable functions with \( \mathcal{I} = [0, 1] \) (or \( \mathcal{I} = [0, 1]^2 \)). For any functions \( f, g \in \mathbb{H} \), the inner product between \( f \) and \( g \) is defined as \( \int_{\mathcal{I}} f(t)g(t)dt \) and \( \| \cdot \| \) denotes the inner product induced norm. Assume the random elements all come from the same probability space \((\Omega, \mathcal{A}, \mathbb{P})\). Let \( L^p \) be the space of real valued random variables with finite \( L^p \) norm, i.e., \( (E|X|^p)^{1/p} < \infty \). Further we denote \( L^p_{\mathbb{H}} \) the space of \( \mathbb{H} \) valued random variables \( X \) such that \( (E\|X\|^p)^{1/p} < \infty \).

Given two sequences of temporally dependent functional observations, \( \{X_i(t)\}_{i=1}^{N_1} \) and \( \{Y_i(t)\}_{i=1}^{N_2} \) defined on a common region \( \mathcal{I} \), we are interested in comparing their second order properties. Suppose that the functional time series are second order stationary. We assume that \( E[X_i(t)] = E[Y_i(t)] = 0 \). The result can be easily extended to the situation with nonzero mean functions. Define \( C_X = E[<X_i, \cdot>] \) and \( C_Y = E[<Y_i, \cdot>] \) as the covariance operators of the two sequences respectively. For the convenience of presentation, we shall use the same notation for the covariance operator and the associated covariance function. Denote by \( \{\phi^X_j\}_{j=1}^{\infty} \) and \( \{\lambda^X_j\}_{j=1}^{\infty} \) the eigenfunctions and eigenvalues of \( C_X \). Analogous quantities are \( \{\phi^Y_j\}_{j=1}^{\infty} \) and \( \{\lambda^Y_j\}_{j=1}^{\infty} \) for the second sample. Denote by \( |v| \) the Euclidean norm of a vector \( v \in \mathbb{R}^p \). Let \( \text{vech}(\cdot) \) be the operator that stacks the columns below the diagonal of a symmetric \( m \times m \) matrix as a vector with \( m(m+1)/2 \) components. Let \( D[0, 1] \) be the space of functions on \([0, 1]\) which are right-continuous and have left limits, endowed with the Skorokhod topology (see Billingsley, 1999). Weak convergence in \( D[0, 1] \) or more generally in the \( \mathbb{R}^m \)-valued function space \( D^m[0, 1] \) is denoted by “\( \Rightarrow \)” , where \( m \in \mathbb{N} \) and convergence in
distribution is denoted by \( \sim \rightarrow^d \). Define \( [a] \) the integer part of \( a \in \mathbb{R} \), and \( \delta_{ij} = 1 \) if \( i = j \) and \( \delta_{ij} = 0 \) if \( i \neq j \). In what follows, we shall discuss the tests for comparing the three quantities \( C_X, \phi_X^i \) and \( \lambda_X^i \) with \( C_Y, \phi_Y^i \) and \( \lambda_Y^i \), respectively.

### 3.2.1 Covariance operator

Consider the problem of testing the hypothesis \( H_{1,0} : C_X = C_Y \) versus the alternative \( H_{1,a} : C_X \neq C_Y \) (in the operator norm sense) for two mean zero stationary functional time series \( \{ X_i(t) \}_{i=1}^{N_1} \) and \( \{ Y_i(t) \}_{i=1}^{N_2} \). Let \( N = N_1 + N_2 \). Throughout the chapter, we assume that

\[
N_1/N \rightarrow \gamma_1, \quad N_2/N \rightarrow \gamma_2, \quad \text{as} \min(N_1, N_2) \rightarrow +\infty,
\]

where \( \gamma_1, \gamma_2 \in (0, 1) \) and \( \gamma_1 + \gamma_2 = 1 \). Define the one-dimensional operator \( \mathcal{X}_i = \langle X_i, \cdot \rangle > X_i = X_i \otimes X_i \) and \( \mathcal{Y}_j = \langle Y_j, \cdot \rangle > Y_j = Y_j \otimes Y_j \). Let \( \hat{C}_{XY} \) be the empirical covariance operator based on the pooled samples, i.e.,

\[
\hat{C}_{XY} = \frac{1}{N_1 + N_2} \left( \sum_{i=1}^{N_1} \mathcal{X}_i + \sum_{i=1}^{N_2} \mathcal{Y}_i \right). \tag{3.1}
\]

Denote by \( \{ \hat{\lambda}_{XY}^i \} \) and \( \{ \hat{\phi}_{XY}^i \} \) the corresponding eigenvalues and eigenfunctions. The population counterpart of \( \hat{C}_{XY} \) is then given by \( \hat{C}_{XY} = \gamma_1 C_X + \gamma_2 C_Y \) whose eigenvalues and eigenfunctions are denoted by \( \{ \hat{\lambda}^i \} \) and \( \{ \hat{\phi}^i \} \) respectively. Further let \( \hat{C}_{X,m} = \frac{1}{m} \sum_{i=1}^{m} \mathcal{X}_i \) be the sample covariance operator based on the subsample \( \{ X_i(t) \}_{i=1}^{m} \) with \( 2 \leq m \leq N_1 \). Define \( \{ \hat{\phi}_{X,m}^i \}_{i=1}^{m} \) and \( \{ \hat{\lambda}_{X,m}^i \}_{i=1}^{m} \) the eigenfunctions and eigenvalues of \( \hat{C}_{X,m} \) respectively, i.e.,

\[
\int_T \hat{C}_{X,m}(t,s)\hat{\phi}_{X,m}^i(s)ds = \hat{\lambda}_{X,m}^i \hat{\phi}_{X,m}^i(t), \tag{3.2}
\]

and \( \int_T \hat{\phi}_{X,m}^i(t)\hat{\phi}_{X,m}^j(t)dt = \delta_{ij} \). Similarly, quantities \( \hat{C}_{Y,m} \), \( \{ \hat{\phi}_{Y,m}^i \}_{i=1}^{N_2} \) and \( \{ \hat{\lambda}_{Y,m}^i \}_{i=1}^{N_2} \) are defined for the second sample with \( 2 \leq m' \leq N_2 \). To introduce the SN-based test, we define the recursive estimates

\[
\hat{c}_{k}^{i,j} = \langle (\hat{C}_{X,[kN_1/N]} - \hat{C}_{Y,[kN_2/N]}), \hat{\phi}_{XY}^i, \hat{\phi}_{XY}^j \rangle, \quad 2 \leq k \leq N, \quad 1 \leq i, j \leq K,
\]

which estimate the difference of the covariance operators on the space spanned by \( \{ \hat{\phi}^j \}_{j=1}^{K} \). Here \( K \) is a user-chosen number, which is held fixed in the asymptotics. Denote by \( \hat{\alpha}_k = \text{vech}(C_k) \) with \( C_k = (c_{k}^{i,j})_{i,j=1}^{K} \). In the independent and Gaussian case, Panaretos et al. (2010) proposed the
following test (hereafter, the PKM test),

\[ T_{N_1,N_2} = \frac{N_1N_2}{2N} \sum_{i=1}^{K} \sum_{j=1}^{K} \left( \frac{\mathcal{C}_{i,j}}{\hat{\theta}_i \hat{\theta}_j} \right)^2, \quad \hat{\theta}_j = \frac{1}{N} \left\{ \sum_{i=1}^{N_1} (X_i, \hat{\phi}_{X_i}^j >)^2 + \sum_{i=1}^{N_2} (Y_i, \hat{\phi}_{Y_i}^j >)^2 \right\}, \]

which converges to \( \chi^2_{(K+1)K/2} \) under the null. To take the dependence into account, we introduce the SN matrix

\[ V_{SN,N}^{(1)}(d) = \frac{1}{N^2} \sum_{k=1}^{N} k^2 (\hat{\alpha}_k - \hat{\alpha}_N)(\hat{\alpha}_k - \hat{\alpha}_N)', \tag{3.3} \]

with \( d = (K + 1)K/2 \). The SN-based test statistic is then defined as,

\[ G_{SN,N}^{(1)}(d) = N\hat{\alpha}_N' (V_{SN,N}^{(1)}(d))^{-1} \hat{\alpha}_N. \tag{3.4} \]

Notice that the PKM test statistic can also be written as a quadratic form of \( \hat{\alpha}_N \) but with a different normalization matrix that is only applicable to the independent and Gaussian case. The special form of the SN-based test statistic makes it robust to the dependence within each sample and also the dependence between the two samples when their sample sizes are equal. We shall study the asymptotic behavior of \( G_{SN,N}^{(1)}(d) \) under the weak dependence assumption in Section 3.3.

### 3.2.2 Eigenvalues and eigenfunctions

In practice, it is also interesting to infer how far the marginal distributions of two sequences of stationary functional time series coincide/differ and quantify the difference. By the Karhunen-Loève expansion (Bosq, 2000), we have

\[ X_i(t) = \sum_{j=1}^{+\infty} \sqrt{\lambda_X^j} \beta_{X_i,j} \phi_X^j(t), \quad Y_i(t) = \sum_{j=1}^{+\infty} \sqrt{\lambda_Y^j} \beta_{Y_i,j} \phi_Y^j(t), \]

where \( \beta_{X_i,j} = \int_I X_i(t) \phi_X^j(t) dt \) and \( \beta_{Y_i,j} = \int_I Y_i(t) \phi_Y^j(t) dt \) are the principal components (scores), which satisfy that \( E[\beta_{X_i,j} \beta_{X_i,j'}] = \delta_{jj'} \) and \( E[\beta_{Y_i,j} \beta_{Y_i,j'}] = \delta_{jj'} \). The problem is then translated into testing the equality of the functional principal components (FPC’s) namely the eigenvalues and eigenfunctions. For a prespecified positive integer \( M \), we denote the vector of the first \( M \) eigenvalues by \( \lambda_X^{1:M} = (\lambda_X^1, \ldots, \lambda_X^M) \) and \( \lambda_Y^{1:M} = (\lambda_Y^1, \ldots, \lambda_Y^M) \). Further define \( \phi_X^{1:M} = (\phi_X^1, \ldots, \phi_X^M) \) and \( \phi_Y^{1:M} = (\phi_Y^1, \ldots, \phi_Y^M) \) the first \( M \) eigenfunctions of the covariance operators \( C_X \) and \( C_Y \) respectively. Since the eigenfunctions are determined up to a sign, we assume that \( < \phi_X^j, \phi_Y^j > \geq 0 \) in order for the comparison to be meaningful. We aim to test the null hypothesis \( H_{2,0} : \lambda_X^{1:M} = \lambda_Y^{1:M} \) and
$H_{3.0} : \phi_{X}^{1 \cdot M} = \phi_{Y}^{1 \cdot M}$ versus the alternatives that $H_{2.a} : \lambda_{X}^{1 \cdot M} \neq \lambda_{Y}^{1 \cdot M}$ and $H_{3.a} : \phi_{X}^{1 \cdot M} \neq \phi_{Y}^{1 \cdot M}$ (in the $L^2$ norm sense). The problem of comparing the FPC’s of two independent and identically distributed (iid) functional sequences has been considered in Benko et al. (2009), where the authors proposed an iid bootstrap method which seems not applicable to the dependent case. The block bootstrap based method is expected to be valid in the dependence case but the choice of the block size seems to be a difficult task in the current setting. To accommodate the dependence and avoid the bandwidth choice, we adopt the SN idea.

Recall the recursive estimates of the eigenvalues $\hat{\lambda}_{X,m}$ and $\hat{\lambda}_{Y,m}$, which are calculated based on the subsamples $\{X_i(t)\}_{i=1}^m$ and $\{Y_i(t)\}_{i=1}^{m'}$. Let $\hat{\theta}_k = \hat{\lambda}_{X,m}^{(kN_1/N)} - \hat{\lambda}_{Y,m}^{(kN_2/N)}$ and $\hat{\theta}_k = (\hat{\theta}_1, \ldots, \hat{\theta}_M)'$ with $[N\epsilon] \leq k \leq N$ for some $\epsilon \in (0,1]$, which is held fixed in the asymptotics. We consider the trimmed SN-based test statistic

$$G_{SN,N}^{(2)}(M) = N^3 \hat{\theta}_N^{-1} \sum_{k=[N\epsilon]}^N k^2 (\hat{\theta}_k - \hat{\theta}_N)(\hat{\theta}_k - \hat{\theta}_N)' \hat{\theta}_N. \quad (3.5)$$

The trimmed version of the SN-based test statistic is proposed out of technical consideration when the functional observations lie on an infinite dimensional space. It can be seen from the proof in the appendix that the trimming is not required when functional data lie on a finite dimensional space; see Remark 3.6.1.

**Remark 3.2.1.** To compare the difference between the eigenvalues, one may also consider their ratios. Define $\hat{\tilde{\zeta}}_k = (\hat{\lambda}_{X,m}^{(kN_1/N)}/\hat{\lambda}_{Y,m}^{(kN_2/N)}) \cdots (\hat{\lambda}_{X,m}^{(N \cdot kN_1/N)}/\hat{\lambda}_{Y,m}^{(N \cdot kN_2/N)})'$ for $k = [N\epsilon], \ldots, N$. An alternative SN-based test statistic is given by

$$\tilde{G}_{SN,N}^{(2)}(M) = N(\hat{\tilde{\zeta}}_N - I_M)' \left( \frac{1}{N^2} \sum_{k=[N\epsilon]}^N k^2 (\hat{\tilde{\zeta}}_k - \hat{\tilde{\zeta}}_N)(\hat{\tilde{\zeta}}_k - \hat{\tilde{\zeta}}_N)' \right)^{-1} (\hat{\tilde{\zeta}}_N - I_M), \quad (3.6)$$

where $I_M$ is a $M$-dimensional vector of all ones. Since the finite sample improvement by using $\tilde{G}_{SN,N}^{(2)}(M)$ is not apparent, we do not further investigate the properties of $\tilde{G}_{SN,N}^{(2)}(M)$.

We now turn to the problem of testing the equality of the eigenfunctions. To proceed, we let

$$\tilde{\nu}_j = (\hat{\phi}_{X,Y}^{j+1}, \hat{\phi}_{X,Y}^{j+2}, \ldots, \hat{\phi}_{X,Y}^p) \quad (3.7)$$

be a vector of $p - j$ orthonormal basis functions for $j = 1, 2, \ldots, M$ with $M \leq p$ and $p$ being a user chosen number. Recall that $\hat{\phi}_{X,m}^j(t)$ and $\hat{\phi}_{Y,m}^j(t)$ are the $j$th eigenfunctions of the empirical
can be shown that the recursive estimates in the appendix. Under suitable assumptions as given in the next section, it can be close to 0. In this case, one remedy is to consider alternative basis functions, e.g., (3.20) and φ

\[ \phi \]

and the projection vectors

\[ \hat{\eta}_k = (\hat{\phi}_X^{1},\ldots,\hat{\phi}_X^{+1},\hat{\phi}_{XY}^{0},\ldots,\hat{\phi}_{XY}^{N}) \]

for some 0 < \( \epsilon < 1 \).

**Remark 3.2.2.** It is worth noting that \( G_{SN,N}^{(3)}(M_0) \) is designed for testing the equality of the first \( M \) eigenfunctions. Suppose we are interested in testing the hypothesis for a particular eigenfunction, i.e., the null \( \phi_X^i = \phi_Y^i \) versus the alternative \( \phi_X^i \neq \phi_Y^i \). We can consider the basis functions

\[ \hat{\nu}_j = (\hat{\phi}_{XY}^{1},\ldots,\hat{\phi}_{XY}^{+1},\hat{\phi}_{XY}^{0}) \]

and the projection vector \( \hat{\eta}_k = (\hat{\phi}_X^{1},\hat{\phi}_X^{2},\ldots,\hat{\phi}_X^{N},\hat{\phi}_{XY}^{1},\ldots,\hat{\phi}_{XY}^{N})' \). The SN-based test statistic can then be constructed in a similar manner. We also note that when \( \phi_X^i \neq \phi_Y^i \) and \( \phi_X^i = \phi_Y^i \) for \( i \neq j \), the choice of \( \hat{\nu}_j \) may result in trivial power because \( \phi_X^i - \phi_Y^i \) for \( i \neq j \) can be close to 0. In this case, one remedy is to consider alternative basis functions, e.g., (3.20) and (3.21) as suggested in the simulation.

**Remark 3.2.3.** The choice of the basis functions \( \hat{\nu}_j \) is motivated by the Bahadur representation of the recursive estimates in the appendix. Under suitable assumptions as given in the next section, it can be shown that

\[ \langle \hat{\phi}_X^k, \phi \rangle = \langle \phi_X, \phi \rangle - \frac{1}{k} \sum_{i=1}^{k} \left( \sum_{s \neq a} \frac{\beta_{X,s}\beta_{X,a}}{\lambda_X - \lambda_X^a} \langle \phi_X, \phi \rangle \right) + R_{X,k} \]

(3.9)
with $R^k_{X,k}$ being the remainder term and $\phi \in L^2(I)$. The second term on the RHS of (3.9) plays a key role in determining the limiting distribution of the SN-based test statistic. When $\phi = \phi^j_X$ with $j \neq a$, the linear term reduces to $\frac{1}{k} \sum_{i=1}^k \frac{\beta_{X,i} \beta_{X,ia}}{\lambda_X - \lambda_X^i}$, which satisfies the functional central limit theorem under suitable weak dependence assumption. Notice that the linear term vanishes when $\phi = \phi^a_X$ and the asymptotic distribution of the projection vector is degenerate. It is also worth noting that the linear terms in the Bahadur representations of $\langle \hat{\phi}^a_X, k, \phi^j_X \rangle$ and $\langle \hat{\phi}^j_X, k, \phi^a_X \rangle$ are opposite of each other which suggests that when testing the eigenfunctions jointly, the basis functions should be chosen in a proper way so that the asymptotic covariance matrix of the projection vector, i.e., $\hat{\eta}_k$ is nondegenerate.

### 3.3 Theoretical results

To study the asymptotic properties of the proposed statistics, we adopt the dependence measure proposed in Hörmann and Kokoszka (2010), which is applicable to the temporally dependent functional process. There are also other weak dependence measures (e.g., mixing) or specific processes (e.g., functional linear processes) suitable for the asymptotic analysis of functional time series (see Bosq, 2000), we decide to use Hörmann and Kokoszka’s $L^p$-m-approximating dependence measure for its broad applicability to linear and nonlinear functional processes as well as its theoretical convenience and elegance.

**Definition 3.3.1.** Assume that $\{X_i\} \in L^p_\mathbb{H}$ with $p > 0$ admits the following representation

$$X_i = f(\varepsilon_i, \varepsilon_{i-1}, \ldots), \ i = 1, 2, \ldots, \ (3.10)$$

where the $\varepsilon_i$’s are iid elements taking values in a measurable space $S$ and $f$ is a measurable function $f : S^\infty \rightarrow \mathbb{H}$. For each $i \in \mathbb{N}$, let $\{\varepsilon_j^{(i)}\}_{j \in \mathbb{Z}}$ be an independent copy of $\{\varepsilon_j\}_{j \in \mathbb{Z}}$. The sequence $\{X_i\}$ is said to be $L^p$-m-approximable if

$$\sum_{m=1}^{\infty} (E||X_m - X_m^{(m)}||^p)^{1/p} < \infty, \ (3.11)$$

where $X_i^{(m)} = f(\varepsilon_i, \varepsilon_{i-1}, \ldots, \varepsilon_{i-m+1}, \varepsilon_{i-m}, \varepsilon_{i-m-1}, \ldots)$. 

45
Define $B_q(r)$ as a $q$-dimensional vector of independent Brownian motions. For $\epsilon \in [0, 1)$, we let

$$W_q(\epsilon) = B_q(1)^t J_q(\epsilon)^{-1} B_q(1), \quad \text{where} \quad J_q(\epsilon) = \int_\epsilon^1 (B_q(r) - r B_q(1))(B_q(r) - r B_q(1))^t \, dr.$$ 

The critical values of $W_q := W_q(0)$ have been tabulated by Lobato (2001). In general, the quantiles of $W_q(\epsilon)$ can be obtained via simulation. To derive the asymptotic properties of the proposed tests, we make the following assumptions.

**Assumption 3.3.1.** Assume we make the following assumptions.

**Assumption 3.3.2.** Assume that Assumption 3.3.2 allows dependence between $G$ and $\Lambda$.

**Assumption 3.3.3.** Assume $\lambda_1^X > \lambda_2^X > \cdots > \lambda_{m_0+1}^X$ and $\lambda_1^Y > \cdots > \lambda_{m_0+1}^Y$, for some positive integer $m_0 \geq 2$.

Note that Assumption 3.3.2 allows dependence between $\{X_i(t)\}$ and $\{Y_i(t)\}$, which is weaker than Assumption 3.3.1. To investigate the asymptotic properties of $G_{SN,N}^{(1)}(d)$ under the local alternatives, we consider the local alternative $H_{1,a} : C_X - C_Y = L \hat{C}/\sqrt{N}$ with $\hat{C}$ being a Hilbert-Schmidt operator, where $L$ is a nonzero constant. Define $\Delta = (\prec \hat{C} \hat{\phi}^t, \hat{\phi}^j >) \in \mathbb{R}^{K \times K}$ as the projection of $\hat{C}$ onto the space spanned by $\{\hat{\phi}^1, \hat{\phi}^2, \ldots, \hat{\phi}^K\}$ and assume that $\text{vech}(\Delta) \neq 0 \in \mathbb{R}^d$. The following theorem states the asymptotic behaviors of $G_{SN,N}^{(1)}(d)$ under the null and the local alternatives.

**Theorem 3.3.1.** Suppose Assumptions 3.3.1, 3.3.3 hold with $m_0 \geq K$. Further assume that the asymptotic covariance matrices $\Lambda^*_d(\Lambda^*_d)^t$ given in Lemma 3.6.3 is positive definite. Then under $H_{1,0}$, $G_{SN,N}^{(1)}(d) \overset{d}{\to} W_d$ and under $H_{1,a}$, $\lim_{|L| \to +\infty} \lim_{N \to +\infty} G_{SN,N}^{(1)}(d) = +\infty$. Furthermore, if $\gamma_1 = \gamma_2$, then the conclusion also holds with Assumption 3.3.1 replaced by Assumption 3.3.2.

It is seen from Theorem 3.3.1 that $G_{SN,N}^{(1)}(d)$ has pivotal limiting distributions under the null and they are consistent under the local alternatives as $L \to +\infty$. To study the asymptotics of $G_{SN,N}^{(2)}(M)$ and $G_{SN,N}^{(3)}(M_0)$, we introduce some notation. Let $\omega_{X_i}^{j,k} = \beta_{X_i,j} \beta_{X_i,k}$ and $r_{X_i}^{j,k,j',k'}(h) = E[(\omega_{X_i}^{j,k} - \delta_{jk} \lambda_j)(\omega_{X_i}^{j',k'} - \delta_{j'k'} \lambda_{j'})]$ be the cross-covariance function between $\omega_{X_i}^{j,k}$ and $\omega_{X_i}^{j',k'}$ at lag $h$. Set $r_{X_i}^{j,k}(h) := r_{X_i}^{j,k,j,k}(h)$. Define $v_{X_i}^{j,k} = \omega_{X_i}^{j,k} - E[\omega_{X_i}^{j,k}] = \omega_{X_i}^{j,k} - \delta_{jk} \lambda_j$. Analogous quantities $r_{Y}^{j,k,j',k'}(h)$ and $v_{Y}^{j,k}$ can be defined for the second sample. We make the following assumption to facilitate our derivation.
Assumption 3.3.4. Suppose that
\[
\sum_{j,k} \sum_{j',k'} \left( \sum_{h=-\infty}^{+\infty} |r^{j,k}_X(h)| \right)^2 < +\infty, \quad \sum_{j,k} \sum_{h=-\infty}^{+\infty} |r^{j,k}_X(h)| < +\infty, \tag{3.12}
\]
and
\[
\sum_{j,k} \sum_{j',k'} \sum_{s_1,s_2 \in \mathbb{Z}} \left| \text{cum}(v^{j,k}_X, v^{j,k'}_{X_{s_1}}, v^{j',k'}_{X_{s_2}}) \right| < \infty, \tag{3.13}
\]
The summability conditions also hold for the second sample \{Y_i(t)\}.

Assumption 3.3.4 is parallel to the summability condition considered in Benko et al. (2009) (see Assumption 1 therein) for iid functional data. It is not hard to verify the above assumption for Gaussian linear functional process (see e.g., Bosq, 2000), as demonstrated in the following proposition.

Proposition 3.3.2. Consider the linear process \(X_i(t) = \sum_{j=0}^{+\infty} b_j \varepsilon_{i-j}(t)\), where \(\varepsilon_j(t) = \sum_{i=1}^{+\infty} \sqrt{\lambda_i} z_{i,j} \phi_i(t)\) with \(\{z_{i,j}\}\) being a sequence of independent standard normal random variables across both index \(i\) and \(j\). Let \(\pi(h) = \sum_{i,j} \lambda_j < \infty\) and \(\sum_{h} |\pi(h)| < \infty\). Then Assumption 3.3.4 holds for \(\{X_i(t)\}\).

Theorem 3.3.3. Suppose Assumptions 3.3.1, 3.3.3, 3.3.4 hold with \(m_0 \geq M\) and the asymptotic covariance matrix \(\tilde{\Lambda}_M \tilde{\Lambda}_M^t\) given in Lemma 3.6.5 is positive definite. Then under \(H_{2,0}\), we have \(G^{(2)}_{SN,N}(M) \rightarrow^d W_M(\varepsilon)\). Under the local alternative \(H_{2,a} : \lambda^{1:M}_X - \lambda^{1:M}_Y = \frac{L}{\sqrt{N}} \tilde{\lambda}\) with \(\tilde{\lambda} \neq 0 \in \mathbb{R}^M\), we have \(\lim_{|L| \rightarrow \infty} \lim_{N \rightarrow +\infty} G^{(2)}_{SN,N}(M) = +\infty\).

In order to study the asymptotic properties of \(G^{(3)}_{SN,N}(M_0)\) under the null and local alternative, we further make the following assumption.

Assumption 3.3.5. Let \(\tilde{\beta}^{(m)}_{X_{i,j}} = \int X_i^{(m)}(t) \phi^j_X(t) dt\), where \(X_i^{(m)}\) is the \(m\)-dependent approximation of \(X_i(t)\) (see definition 3.3.1). Suppose one of the following conditions holds:
\[
\sum_{m=1}^{+\infty} \sum_{j=1}^{+\infty} \left\{ E \left( \beta_{X_{i,j}} - \tilde{\beta}_{X_{i,j}}^{(m)} \right)^4 \right\}^{1/4} < \infty, \quad \sum_{j=1}^{+\infty} \left( E \tilde{\beta}_{X_{i,j}}^{(1)} \right)^{1/4} < \infty, \tag{3.14}
\]
or
\[
\sum_{m=1}^{+\infty} \sum_{j=1}^{+\infty} \left| \phi^j_X \phi^j_X \right| < +\infty, \quad 2 \leq j \leq p. \tag{3.15}
\]
The same condition holds for the second sample \(\{Y_i(t)\}\).
Theorem 3.3.4. Suppose Assumptions 3.3.1, 3.3.3, 3.3.4 and 3.3.5 hold with $m_0 \geq M$ and the asymptotic covariance matrix $\tilde{\Lambda}_M^T \tilde{\Lambda}_M$ given in Lemma 3.6.7 is positive definite. Then under $H_{3.0}$, we have $G_{SN,N}^{(3)}(M_0) \rightarrow^d W_{M_0}(\epsilon)$.

Proposition 3.3.5. Define $\tilde{\Delta}$ by replacing $\hat{\phi}_{X,N}^j$, $\hat{\phi}_{Y,N}^j$ and $\hat{\phi}_{XY}^j$ with $\phi_X^j$, $\phi_Y^j$ and $\tilde{\phi}^j$ in the definition of $\hat{\eta}_N$. Consider the local alternative $H_{3,a}: \Delta = L \tilde{\psi}/\sqrt{N}$ with $\tilde{\psi} \neq 0 \in \mathbb{R}^{M_0}$. Suppose Assumptions 3.3.1, 3.3.3, 3.3.4 and 3.3.5 hold with $m_0 \geq M$ and the asymptotic covariance matrix $\tilde{\Lambda}_M^T \tilde{\Lambda}_M$ given in Lemma 3.6.7 is positive definite. Then we have $\lim_{|L| \rightarrow \infty} \lim_{N \rightarrow +\infty} G_{SN,N}^{(3)}(M_0) = +\infty$ under $H_{3,a}$.

It is worth noting that the conclusions in Theorem 3.3.3, Theorem 3.3.4 and Proposition 3.3.5 also hold with Assumption 3.3.1 replaced by Assumption 3.3.2 and $\gamma_1 = \gamma_2$. Finally, we point out that condition (3.14) can be verified for Gaussian linear functional process as shown in the following proposition.

Proposition 3.3.6. Consider the Gaussian linear process in Proposition 3.3.2. Assume that $\sum_{j=1}^\infty \sqrt{\lambda_j} < \infty$ and $\sum_{m=1}^\infty (\sum_{j=m}^\infty b_j^2)^{1/2} < \infty$. Then Assumption 3.3.4 and condition (3.14) are satisfied for $\{X_i(t)\}$.

3.4 Numerical studies

We conduct a number of simulation experiments to assess the performance of the proposed SN-based tests in comparison with the alternative methods in the literature. We generate functional data on a grid of $10^3$ equispaced points in $[0, 1]$, and then convert discrete observations into functional objects by using B-splines with 20 basis functions. We also tried 40 and 100 basis functions and found that the number of basis functions does not affect our results much. Throughout the simulations, we set the number of Monte Carlo replications to be 1000 except for the iid bootstrap method in Benko et al. (2009), where the number of replications is only 250 because of high computational cost.

3.4.1 Comparison of covariance operators

To investigate the finite sample properties of $G_{SN,N}^{(1)}(d)$ for dependent functional data, we modify the simulation setting considered in Panaretos et al. (2010). Formally, we consider the model,

$$\sum_{j=1}^3 \left\{ \xi_{j+1}^i \sqrt{2} \sin(2\pi jt) + \xi_{j+2}^i \sqrt{2} \cos(2\pi jt) \right\}, \quad i = 1, 2, \ldots, \quad t \in [0, 1].$$

(3.16)
where the coefficients \( \xi_i = (\xi_{i,1}^1, \xi_{i,2}^1, \xi_{i,3}^1, \xi_{i,1}^2, \xi_{i,2}^2, \xi_{i,3}^2)' \) are generated from a VAR process,

\[
\xi_i = \rho \xi_{i-1} + \sqrt{1 - \rho^2} e_i,
\]

with \( e_i \in \mathbb{R}^6 \) being a sequence of iid normal random variables with mean zero and covariance matrix

\[
\Sigma_e = \frac{1}{1 + \mu^2} \text{diag}(v) + \frac{\mu^2}{1 + \mu^2} \mathbf{1}_6 \mathbf{1}_6'.
\]

We generate two independent functional time series \( \{X_i(t)\} \) and \( \{Y_i(t)\} \) from (3.16) with \( \rho = 0.5 \) and \( \mu = 1 \). We compare the SN-based test with the PKM test which is designed for independent Gaussian process, and the traditional test which is constructed based on a consistent LRV estimator (denoted by CLRV), i.e.,

\[
G_{\text{CL},N}(d) = N \hat{\alpha} N \hat{\Sigma}_\alpha^{-1} \hat{\alpha} N,
\]

where \( \hat{\Sigma}_\alpha \) is a lag window LRV estimator with Bartlett kernel and data dependent bandwidth (see Andrews, 1991). We report the simulation results for \( N_1 = N_2 = 100, 200, K = 1, 2, 3, 4 \) and various values of \( v \) in Table 3.1. Results in scenario A show that the size distortion of all the three tests increases as \( K \) gets larger. The SN-based test has the best size compared to the other two tests. The PKM test is severely oversized due to the fact that it does not take the dependence into account. It is seen from the table that the CLRV test also has severe size distortion especially for large \( K \), which is presumably due to the poor estimation of the LRV matrix of \( \hat{\alpha} N \) when the dimension is high. Under the alternatives, we report the size-adjusted power which is computed using finite sample critical values based on the simulation under the null model where we assume that both \( \{X_i(t)\} \) and \( \{Y_i(t)\} \) are generated from (3.16) with \( \rho = 0.5, \mu = 1 \) and \( v = v_X \). From scenarios B-D in Table 3.1, we observe that the PKM is most powerful which is largely due to its severe upward size distortion. The SN-based test is less powerful compared to the other two tests but the power loss is generally moderate in most cases. Furthermore, we present the results when choosing \( K \) by

\[
K^* = \text{argmin} \left\{ 1 \leq J \leq 20 : \frac{\sum_{i=1}^{20} \hat{\lambda}_{iXY}^i}{\sum_{i=1}^{20} \hat{\lambda}_{iXY}^i} > 85\% \right\}.
\]

An alternative way of choosing \( K \) is to consider the penalized fit criteria (see Panaretos et al., 2010, for the details). We notice that the performance of all the three tests based on automatic choice \( K^* \) is fairly close to the performance when \( K = 4 \) in most cases. To sum up, the SN-based test provides the best size under the null and has reasonable power under different alternatives considered here, which is consistent with the “better size but less power” phenomenon seen in the univariate setup (Lobato, 2001; Shao, 2010).
3.4.2 Comparison of eigenvalues and eigenfunctions

In this subsection, we study the finite sample performance of the SN-based test for testing the equality of the eigenvalues and eigenfunctions. We consider the data generating process,

\[
\sum_{j=1}^{2} \left\{ \xi_{j,1}^i \sqrt{2} \sin(2\pi j t + \delta_j) + \xi_{j,2}^i \sqrt{2} \cos(2\pi j t + \delta_j) \right\}, \quad i = 1, 2, \ldots, \ t \in [0, 1],
\]

where \( \xi_i^* = (\xi_{1,1}^i, \xi_{1,2}^i, \xi_{2,1}^i, \xi_{2,2}^i)' \) is a 4-variate VAR process (3.17) with \( e_i \in \mathbb{R}^4 \) being a sequence of iid normal random variables with mean zero and covariance matrix \( \Sigma_e = \frac{1}{1+\mu^2} \text{diag}(v) + \frac{\mu^2}{1+\mu^2} 1_41_4' \).

We set \( \rho = 0.5 \) and \( \mu = 0 \). Under \( H_{2,0} \) (or \( H_{2,a} \)), \( \{X_i(t)\} \) and \( \{Y_i(t)\} \) are generated independently from (3.19) with \( \delta_1 = \delta_2 = 0 \) and \( \nu_X = \nu_Y \) (or \( \nu_X \neq \nu_Y \)). Notice that the eigenvalues of \( \{X_i(t)\} \) and \( \{Y_i(t)\} \) are given respectively by \( \nu_X \) and \( \nu_Y \) when \( \delta_1 = \delta_2 = 0 \). Under \( H_{3,0} \) and \( H_{3,a} \), we generate \( \{X_i(t)\} \) and \( \{Y_i(t)\} \) independently from (3.19) with \( \nu_X = \nu_Y, \delta_X = \delta_Y = \delta \) and \( \delta_{X,1} = \delta_{Y,1} = \delta \), \( \delta_{X,2} = \delta_{Y,2} = 0 \), where \( \delta = 0 \) under the null and \( \delta \neq 0 \) under the alternatives. We aim to test the equality of the first two eigenvalues and eigenfunctions separately and jointly. Because functional data are finite dimensional, we implement the untrimmed version of the SN-based tests, i.e., \( \epsilon = 0 \). To further assess the performance of the SN-based test, we compare our method with the subsampling approach with several choices of subsampling widths and the iid bootstrap method in Benko et al. (2009). Suppose \( N_1 = N_2 = N_0 \). Let \( l \) be the subsampling width and \( \lambda_{\text{sub},i}^j = \lambda_{\text{sub},X,i}^j - \lambda_{\text{sub},Y,i}^j, \ i = 1, 2, \ldots, s_{N_0}(l) = [N_0/l], \) where \( \lambda_{\text{sub},X,i}^j \) and \( \lambda_{\text{sub},Y,i}^j \) are estimates of the \( j \)th eigenvalues based on the \( i \)th nonoverlapping subsamples \( \{X_k(t)\}_{k=(i-1)l+1}^{il} \) and \( \{Y_k(t)\}_{k=(i-1)l+1}^{il} \) respectively. The subsampling variance estimate is given by \( \sigma_{\text{sub},j}^2 = \frac{1}{s_{N_0}(l)} \sum_{i=1}^{s_{N_0}(l)} \left( \lambda_{\text{sub},i}^j - \frac{1}{s_{N_0}(l)} \sum_{i=1}^{s_{N_0}(l)} \lambda_{\text{sub},i}^j \right)^2, \) and the test statistic based on the subsampling variance estimate for testing the equality of the \( j \)th eigenvalue is defined as \( G_{\text{sub},N} = N_0(\hat{\lambda}_{X,N_0}^j - \hat{\lambda}_{Y,N_0}^j)^2/\sigma_{\text{sub},j}^2. \) Since the data-dependent rule for choosing the subsampling width is not available in the current setting, we tried \( l = 8, 12, 16 \) for \( N_0 = 48, 96. \) For testing the equality of eigenvalues jointly and equality of the eigenfunctions, we shall consider a multivariate version of the subsampling-based test statistic which can be defined in a similar fashion. Table 3.2 summarizes the simulation results for testing the eigenvalues with various values of \( \nu. \) From scenario A, we see that performance of the SN-based test under the null is satisfactory while the size distortion of the subsampling-based method is quite severe and is sensitive to the choice of block size \( l. \) It is also not surprising to see that the iid bootstrap method has obvious size distortion as it does not take the dependence into account. Under the alternatives (scenarios B-D), we report the size-adjusted power by using the simulated critical values as described.
in previous subsection. When the sample size is 48, the SN-based method delivers the highest power
among the tests and it tends to have some moderate power loss when the sample size increases to
96. On the other hand, the subsampling method is sensitive to the choice of subsampling width and
its power tends to decrease when a larger subsampling width is chosen.

To test the equality of the first two eigenfunctions, we implement the SN-based test and the
subsampling-based test with the basis functions,

\[ \hat{\nu}_j^* = (\hat{\phi}_{1XY} + \hat{\phi}_{2XY}, \ldots, \hat{\phi}_{j-1XY} + \hat{\phi}_{j+1XY}, \hat{\phi}_{j}^1 + \hat{\phi}_{j}^2, \ldots, \hat{\phi}_{pXY} + \hat{\phi}_{j}^p), \quad 1 \leq j \leq 2, \quad p = 4, \quad (3.20) \]

for testing individual eigenfunction and

\[ \hat{\nu}_{j}^{**} = (\hat{\phi}_{j+1XY}^1 + \hat{\phi}_{j+1XY}^2, \hat{\phi}_{j}^1, \hat{\phi}_{j}^2, \ldots, \hat{\phi}_{pXY} + \hat{\phi}_{j}^p), \quad 1 \leq j \leq 2, \quad p = 4, \quad (3.21) \]

for testing the first two eigenfunctions jointly. The tests with the above basis functions tend to
provide similar sizes but higher powers as compared to the tests with the basis functions \( \hat{\nu}_i \) in our
simulation study. The basis functions \( \hat{\nu}_j^* \) is constructed by adding the same estimated eigenfunction
\( \hat{\phi}_{j}^1 \) to each component of \( \hat{\nu}_j \), and the associated SN-based test is expected to be asymptotically
valid in view of the Bahadur representation (3.9). The simulation results are summarized in Table
3.3 and Figure 3.1 which present the sizes of the SN-based test, the subsampling-based test and
the iid bootstrap method, and the size adjusted powers of the former two respectively. It is seen
from Table 3.3 that the sizes of the SN-based test are accurate while the subsampling-based test
is apparently size-distorted. It is somewhat surprising to see that the iid bootstrap provides better
sizes compared to the subsampling-based approach which is designed for dependent data. Figure 3.1
plots the (size-adjusted) power functions of the SN-based test and the subsampling-based test which
are monotonically increasing on \( \delta \). When \( N_1 = N_2 = 48 \), the SN-based test delivers the highest
power in most cases. The subsampling-based test with a small subsampling width becomes most
powerful when sample size increases to 96. Overall, the SN-based test is preferable as it provides
quite accurate size under the null and has respectable power under the alternatives.

### 3.5 Climate projections analysis

We apply the SN-based test to a gridded spatio-temporal temperature dataset covering a subregion
of North America. The dataset comes from two separate sources: gridded observations generated
from interpolation of station records (HadCRU), and gridded simulations generated by an AOGCM (NOAA GFDL CM2.1). Both datasets provide monthly average temperature for the same 19-year period, 1980-1998. Each surface is viewed as a two dimensional functional datum. The yearly average data have been recently analyzed in Zhang et al. (2011), where the goal is to detect a possible change point of the bias between the station observations and model outputs. In this chapter, we analyze the monthly data from 1980 to 1998, which includes 228 functional images for each sequence. We focus on the second order properties and aim to test the equality of the eigenvalues and eigenfunctions of the station observations and model outputs. To perform the analysis, we first remove the seasonal mean functions from the two functional sequences. At each location, we have two time series from the demeaned functional sequences. We apply the SN-based test developed in Shao (2010) to test whether their cross-correlation at lag zero is equal to zero. The \( p \)-values of these tests are plotted in Figure 3.2. The result tends to suggest that the dependence between the station observations and model outputs may not be negligible at certain regions as the corresponding \( p \)-values are extremely small. The two sample tests introduced in this chapter are useful in this case because they are robust to such dependence.

We perform FPCA on the demeaned sequences. Figure 3.3 plots the first three PC's of the station observations and model outputs. We then apply the SN-based tests \( G^{(2)}_{SN,N}(M) \) and \( G^{(3)}_{SN,N}(M_0) \) (with \( p = 3 \)) to the demeaned sequences, which yields the results summarized in Table 3.4. It is seen from the table that the first two eigenvalues of the station observations and model outputs may be the same, at least statistical significance is below the 10% level, while there is a significant difference between their third eigenvalue. The SN-based tests also suggest that there are significant differences of the first and second PCs between the station observations and model outputs as the corresponding \( p \)-values are less than 5% while the difference between the third PCs is not significant at the 10% level; compare Figure 3.3. We also tried the basis functions \( \nu_j^* \) and \( \nu_j^{**} \) for \( G^{(3)}_{SN,N}(M_0) \) (see 3.20 and 3.21), which leads to the same conclusion. To sum up, our results suggest that the second order properties of the station observations and model outputs may not be the same.

In climate projection studies, the use of numerical models outputs has become quite common nowadays because of advances in computing power and efficient numerical algorithms. As mentioned in Jun et al. (2008), “Climate models are evaluated on how well they simulate the current mean climate state, how they can reproduce the observed climate change over the last century, how well they simulate specific processes, and how well they agree with proxy data for very different time periods in the past.” Furthermore, different institutions produce different model outputs based on
different choices of parametrizations, model components, as well as initial and boundary conditions. Thus there is a critical need to assess the discrepancy/similarity between numerical model outputs and real observations, as well as among various model outputs. The two sample tests proposed here can be used towards this assessment at a preliminary stage to get a quantitative idea of the difference, followed by a detailed statistical characterization using sophisticated spatio-temporal modeling techniques (see e.g., Jun et al., 2008). In particular, the observed significance level for each test can be used as a similarity index that measures the similarity between numerical model outputs and real observations, and may be used to rank model outputs. A detailed study along this line would be interesting, but is beyond the scope of this article.

3.6 Proofs of the main results

Lemma 3.6.1. Under Assumption 3.3.2, \{X_i\} and \{Y_i\} are both \(L^2\)-approximable sequences in \(H_{HS}\).

Proof. Let \(X_1^{(m)} = X_1^{(m)} \otimes X_1^{(m)}\), where \(X_1^{(m)}\) is the \(m\)-dependent approximation of \(X_1\). We have

\[
||X_1 - X_1^{(m)}||_{HS} = \left( \int_I \int_I (X_1(t)X_1(s) - X_1^{(m)}(t)X_1^{(m)}(s))^2 dt ds \right)^{1/2}
\]

\[
\leq \sqrt{2} ||X_1 - X_1^{(m)}||(||X_1|| + ||X_1^{(m)}||),
\]

which implies that \((E||X_1 - X_1^{(m)}||_{HS})^{1/2} \leq c(E||X_1 - X_1^{(m)}||^4)^{1/4}\). The same arguments apply to \{Y_i\}. The conclusion follows by noting the fact that \{X_i(t)\} and \{Y_i(t)\} are both \(L^4\)-approximable.

Lemma 3.6.2. Suppose Assumptions 3.3.2-3.3.3 hold. Then we have \(NE||\hat{C}_{XY} - \tilde{C}||_{HS}^2 < c\) and

\[
\limsup_{N \to \infty} NE||\hat{\lambda}_{XY} - \hat{\lambda}||^2 < \infty, \quad \limsup_{N \to \infty} NE||\hat{c}_{XY} \hat{\phi}_{XY} - \tilde{\phi}^j||^2 < \infty,
\]

for \(1 \leq j \leq K\), where \(c\) is a finite constant that does not depend on \(N\), and \(\hat{c}_{XY} = \text{sign}(\hat{\phi}_{XY}, \hat{\phi}^j)\).

Proof. Note that \{X_i(t)\} and \{Y_i(t)\} are both \(L^4\)-approximable. The conclusion follows from Theorem 3.1 and Theorem 3.2 in Hörmann and Kokoszka (2010).

Let \(\hat{R}_{X,k} = \langle (X_k - C_X)\hat{\phi}^j, \hat{\phi}^j \rangle_{1 \leq i, j \leq K}\) and \(\hat{R}_{Y,k} = \langle (Y_k - C_Y)\tilde{\phi}^j, \tilde{\phi}^j \rangle_{1 \leq i, j \leq K}\) with \(1 \leq k \leq N\). Define the empirical estimates \(\hat{R}_{X,k}\) and \(\hat{R}_{Y,k}\) by replacing \(\tilde{\phi}^j\) with \(\hat{c}_{XY} \hat{\phi}_{XY}^j\) in the
definition of $\tilde{R}_{X,k}$ and $\tilde{R}_{Y,k}$. In Lemmas 3.6.3 and 3.6.4 below, we prove an invariance principle for the partial sum process of $\{\tilde{R}_{X,k}, \tilde{R}_{Y,k}\}$ and show that the estimation effect caused by replacing $\tilde{\phi}_i^j$ with $\tilde{c}_{XY}^j \tilde{\phi}_{XY}^j$ is asymptotically negligible.

**Lemma 3.6.3.** Under Assumption 3.3.1, we have

$$
\frac{1}{\sqrt{N}} \left( \frac{1}{\gamma_1} \sum_{k=1}^{[Nr]} \text{vech}(\tilde{R}_{X,k}) - \frac{1}{\gamma_2} \sum_{k=1}^{[Nr]} \text{vech}(\tilde{R}_{Y,k}) \right) \Rightarrow \Lambda_d^* B_d(r). \tag{3.22}
$$

When $\gamma_1 = \gamma_2$, (3.22) holds under Assumption 3.3.2.

**Proof.** Define $\tilde{R}_{X,k}^{(m)}$ and $\tilde{R}_{Y,k}^{(m)}$ by replacing $X_i$ and $Y_i$ with the $m$—dependent approximations $X_i^{(m)}$ and $Y_i^{(m)}$ in the definition of $\tilde{R}_{X,k}$ and $\tilde{R}_{Y,k}$ respectively. Consider the joint process $(\text{vech}(\tilde{R}_{X,k})', \text{vech}(\tilde{R}_{Y,k})')'$ and observe that

$$
E \left| (\text{vech}(\tilde{R}_{X,k})', \text{vech}(\tilde{R}_{Y,k})')' - (\text{vech}(\tilde{R}_{X,k})^{(m)}', \text{vech}(\tilde{R}_{Y,k})^{(m)}')' \right|^2 \\
\leq c \sum_{1 \leq i \leq j \leq K} \left\{ E(\langle \mathcal{X}_i - \mathcal{X}_i^{(m)}, \tilde{\phi}_i^j \rangle^2) + E(\langle \mathcal{Y}_i - \mathcal{Y}_i^{(m)}, \tilde{\phi}_i^j \rangle^2) \right\} \\
\leq c(E(||\mathcal{X}_i - \mathcal{X}_i^{(m)}||_2^2) + E(||\mathcal{Y}_i - \mathcal{Y}_i^{(m)}||_2^2),
$$

which implies that the joint process is $L^2$-$m$-approximable. By Theorem A.1 of Aue et al. (2009), we can establish the functional central limit theorem for the joint processes, i.e.,

$$
\frac{1}{\sqrt{N}} \sum_{k=1}^{[Nr]} \begin{pmatrix} \text{vech}(\tilde{R}_{X,k}) \\ \text{vech}(\tilde{R}_{Y,k}) \end{pmatrix} \Rightarrow \begin{pmatrix} \Sigma_{11}^* & 0 \\ \Sigma_{21}^* & \Sigma_{22}^* \end{pmatrix} \begin{pmatrix} B_d^{(1)}(r) \\ B_d^{(2)}(r) \end{pmatrix}, \quad N_0 \rightarrow +\infty,
$$

where $B_d^{(1)}(r)$ and $B_d^{(2)}(r)$ are two independent $d$—dimensional Brownian motions. It is easy to see the conclusion holds when $\{X_i(t)\}$ and $\{Y_i(t)\}$ are independent. If $\gamma_1 = \gamma_2$, the continuous mapping theorem from $D^{2d}[0,1]$ to $D^d[0,1]$ yields that,

$$
\frac{1}{\sqrt{N}} \left( \frac{1}{\gamma_1} \sum_{k=1}^{[Nr]} \text{vech}(\tilde{R}_{X,k}) - \frac{1}{\gamma_2} \sum_{k=1}^{[Nr]} \text{vech}(\tilde{R}_{Y,k}) \right) \Rightarrow \frac{1}{\gamma_1} (\Sigma_{11}^* - \Sigma_{21}^*) B_d^{(1)}(\gamma_1 r) - \frac{1}{\gamma_1} \Sigma_{22}^* B_d^{(2)}(\gamma_1 r) \\
=^d \Lambda_d^* B_d(r),
$$

where $\Lambda_d^*$ is a lower triangular matrix such that $\Lambda_d^*(\Lambda_d^*)' = \frac{1}{\gamma_1} ((\Sigma_{11}^* - \Sigma_{21}^*)(\Sigma_{11}^* - \Sigma_{21}^*)' + \Sigma_{22}^*(\Sigma_{22}^*)')$.

\[\square\]
Lemma 3.6.4. Suppose Assumptions 3.3.2-3.3.3 hold, we have

\[
\frac{1}{\sqrt{N}} \sup_{r \in [0,1]} \left\{ \left| \sum_{k=1}^{[N,r]} \{ \text{vech}(\tilde{R}_{X,k}) - \text{vech}(\tilde{R}_{Y,k}) \} \right| + \left| \sum_{k=1}^{[N,r]} \{ \text{vech}(\tilde{R}_{Y,k}) - \text{vech}(\tilde{R}_{Y,k}) \} \right| \right\} = o_p(1). \tag{3.23}
\]

Proof. Define \( Z_{X,k}(t, s) = X_k(t)X_k(s) - CX(t, s) \) and \( Z_{X,k}^{(m)}(t, s) = X_k^{(m)}(t)X_k^{(m)}(s) - CX(t, s) \). We aim to show that for each \( 1 \leq i, j \leq K \),

\[
\frac{1}{\sqrt{N}} \sup_{r \in [0,1]} \left\{ \sum_{k=1}^{[N,r]} \int_I \int_I Z_{X,k}(t, s) \left( \tilde{\phi}_i(t)\tilde{\phi}_j(s) - \tilde{\phi}_{XY}\tilde{\phi}_{XY}(t)\tilde{\phi}_{XY}(s) \right) dtds \right\} = o_p(1).
\]

From Lemma 3.6.2 and the proof of Theorem 3.4 in Zhang et al. (2011), the result follows from the fact that

\[
\frac{1}{N^2} \max_{1 \leq k \leq N_1} \left[ \int_I \int_I \left\{ \sum_{i=1}^{k} Z_{X,i}(t, s) \right\}^2 dtds \right] = o_p(1). \tag{3.24}
\]

Under Assumption 3.3.2, it is straightforward to show that \( \{Z_{X,k}(t, s)\} \) is \( L^2-m \)-approximable. Let \( g(t, s) = E[Z_{X,1}(t, s)]^2 \). By the Cauchy-Schwarz inequality, we get

\[
\int_I \int_I g(t, s)dtds = E[|Z_{X,1}(t, s)|^2] + 2 \sum_{r=1}^{+\infty} \int_I \int_I (E[Z_{X,1}(t, s)]^2)^{1/2}(E[Z_{X,r+1}(t, s) - Z_{X,r+1}^{(r)}(t, s)]^2)^{1/2}dtds.
\]

\[
\leq E[|Z_{X,1}(t, s)|^2] + 2(E[|Z_{X,1}(t, s)|^2]^{1/2} \sum_{r=1}^{+\infty} (E[|Z_{X,r+1}(t, s) - Z_{X,r+1}^{(r)}(t, s)|^2]^{1/2}) < \infty.
\]

Since \( Z_{X,1} \) and \( Z_{X,r+1}^{(r)} \) are independent, we have

\[
E \left( \sum_{i=1}^{N_1} Z_{X,i}(t, s) \right)^2 = \sum_{|r| < N_1} (N_1 - |r|)E[Z_{X,1}(t, s)Z_{X,r+1}(t, s)]
\]

\[
\leq N_1 E[Z_{X,1}(t, s)]^2 + 2 \sum_{0 < r < N_1} (N_1 - |r|)E[Z_{X,1}(t, s)(Z_{X,r+1}(t, s) - Z_{X,r+1}^{(r)}(t, s))]
\]

\[
\leq N_1 E[Z_{X,1}(t, s)]^2 + 2N_1 \sum_{r=1}^{+\infty} (E[Z_{X,1}(t, s)]^2)^{1/2}(E[Z_{X,r+1}(t, s) - Z_{X,r+1}^{(r)}(t, s)]^2)^{1/2} = N_1g(t, s).
\]
the fact that \( \hat{\Delta} = \left( \begin{array}{c} \hat{\alpha}_N \end{array} \right) \) also holds under Assumption 3.3.2 provided that in \( \hat{\alpha}_i \) where \( \hat{R}_{X,l}^{i,j} \) and \( \hat{R}_{Y,l}^{i,j} \) are the \((i,j)\)th elements of \( \hat{R}_{X,l} \) and \( \hat{R}_{Y,l} \). Notice that the SN-based test statistic can be written as,

\[
G^{(1)}_{SN,N}(d) = N \hat{\alpha}^{-1}_N V_{SN,N}(d) \hat{\alpha}_N \\
= N (\alpha_N^* + \text{vech}(\hat{\Delta}))' \left( \frac{1}{N^2} \sum_{k=1}^{N} \hat{\alpha}_k^2 (\alpha_k^* - \alpha_N^*) (\alpha_k^* - \alpha_N^*)' \right)^{-1} (\alpha_N^* + \text{vech}(\hat{\Delta})),
\]

where \( \hat{\Delta} = (\langle CX - CY \rangle \hat{\phi}_{XY}^t \hat{\phi}_{XY}^t, \hat{\phi}_{XY}^t \hat{\phi}_{XY}^t \rangle)_{i,j=1}^{K} \) which is a matrix of zeros under the null. Using the fact that \( ||\hat{\phi}_{XY}^t \hat{\phi}_{XY}^t - \hat{\phi}_i^t|| = o_p(1) \), we have \( \sqrt{N} \hat{\Delta} = L \hat{\Delta} + o_p(L) \) under the local alternative. By Lemma 3.6.3, Lemma 3.6.4 and the continuous mapping theorem, we have \( G^{(1)}_{SN,N}(d) \rightarrow d W_d \) under \( H_{1,0} \) and \( \lim_{|L| \rightarrow +\infty} \lim_{N \rightarrow +\infty} G^{(1)}_{SN,N}(d) = +\infty \) under \( H_{1,a} \). In view of Lemma 3.6.3, the conclusion also holds under Assumption 3.3.2 provided that \( \gamma_1 = \gamma_2 \).

Proof of Theorem 3.3.1. Define \( \hat{\alpha}_i^* \) by replacing \( \hat{c}_k^{i,j} \) with \( \frac{1}{[kN_1/N]} \sum_{l=1}^{[kN_1/N]} \hat{R}_{X,l}^{i,j} - \frac{1}{[kN_2/N]} \sum_{l=1}^{[kN_2/N]} \hat{R}_{Y,l}^{i,j} \) in \( \hat{\alpha}_i \) where \( \hat{R}_{X,l}^{i,j} \) and \( \hat{R}_{Y,l}^{i,j} \) are the \((i,j)\)th elements of \( \hat{R}_{X,l} \) and \( \hat{R}_{Y,l} \). Define the \( m \)-dimensional linear operator \( \Pi_{X,m} x = \sum_{i=1}^{m} < \phi_i^j, x > \phi_i^j \). Consider the symmetric operator \( \Pi_{X,m} \hat{C}_{X,k} \Pi_{X,m} \) as well as its eigenvalues and eigenfunctions denoted by \( \hat{\gamma}_{X,k} \geq \hat{\gamma}_{X,k}^2 \geq \cdots, \hat{\gamma}_{X,k}^m \), respectively. Define the \( m \times m \) matrix with \((j,l)\)th entry being \( \frac{1}{k} \sum_{l=1}^{k} w_{X,j}^l \), where \( \omega_{X,j}^l = \beta_{X,j} - \beta_{X,l} \). Suppose the eigenvalues and eigenvectors of \( \hat{\Sigma}_{X,k}^m \) are denoted correspondingly by \( \lambda^a(\hat{\Sigma}_{X,k}^m) \) and \( \xi^a(\hat{\Sigma}_{X,k}^m) \) with \( 1 \leq a \leq m \). It is not hard to see that

\[
\hat{\gamma}_{X,k}^m = \lambda^a(\hat{\Sigma}_{X,k}^m), \quad \hat{\gamma}_{X,k}^m(t) = \xi^a(\hat{\Sigma}_{X,k}^m) \Phi_{X}^m(t),
\]

where \( \Phi_{X}^m(t) = (\phi_1^X(t), \phi_2^X(t), \ldots, \phi_m^X(t)) \). By Lemma A.1 of Kneip and Utikal (2001), we obtain for every \( 1 \leq a \leq m \),

\[
\hat{\gamma}_{X,k}^m - \lambda^a_X = \frac{1}{k} \sum_{i=1}^{k} (\beta_{X,a} - \lambda^a_X) + R_{X,k}^m,
\]

which implies (3.24). The proof is completed by applying the same argument to the second sample.
where
\[ |R_{X,k}^{m,a}| \leq \frac{6 \sup_{|c| = 1} e'(\Sigma_{X,k}^m - \Sigma_{X}^m)^2 e}{\min_{s \neq a} |\lambda_X - \lambda_X^a|} \]
and \( \Sigma_{X}^m = \text{diag}(\lambda_X^1, \ldots, \lambda_X^m) \). The same arguments can be applied directly to the second sample. Let \( \hat{\tau}_k^m = (\hat{\tau}_{X,k}^m, \hat{\tau}_{Y,k}^m)' = (\hat{\tau}_{X,k,1}^m, \ldots, \hat{\tau}_{X,k,m}^m, \hat{\tau}_{Y,k,1}^m, \ldots, \hat{\tau}_{Y,k,m}^m)' \), \( \lambda = (\lambda_{X}^1, \lambda_{X}^1, \lambda_{Y}^1)' \) and \( U^i = ((U_X^i)', (U_Y^i)')' = (\beta_{X,i,1}^2 - \lambda_X^1, \ldots, \beta_{X,i,m}^2 - \lambda_X^m, \beta_{Y,i,1}^2 - \lambda_Y^1, \ldots, \beta_{Y,i,m}^2 - \lambda_Y^m)' \). Then we have

\[
\hat{\tau}_k^m - \lambda = \frac{1}{k} \sum_{i=1}^{k} \begin{pmatrix} U_X^i \\ U_Y^i \end{pmatrix} + R_k^m, \tag{3.25}
\]

with \( R_k^m = ((R_{X,k}^m)', (R_{Y,k}^m)')' = (R_{X,k,1}^m, \ldots, R_{X,k,m}^m, R_{Y,k,1}^m, \ldots, R_{Y,k,m}^m)' \).

**Lemma 3.6.5.** Under Assumption 3.3.1 or Assumption 3.3.2 with \( \gamma_1 = \gamma_2 \), we have

\[
\frac{1}{\sqrt{N}} \left( \frac{1}{\gamma_1} \sum_{i=1}^{[N_1 \tau]} U_X^i - \frac{1}{\gamma_2} \sum_{i=1}^{[N_2 \tau]} U_Y^i \right) \Rightarrow D \hat{\Lambda}_M B_M(r). \tag{3.26}
\]

**Proof.** Suppose \( \beta_{X,i,k}^{(v)} \) is the principle component associated with the \( v \)-dependent approximation sequence \( \{X_i^{(v)}(t)\} \). For every \( 1 \leq k \leq M \),

\[
E \left| \beta_{X,i,k}^2 - (\beta_{X,i,k}^{(v)})^2 \right|^2 = E \left| \left( \int X_i(t) \phi_{X,k}^2(t) dt \right)^2 - \left( \int X_i^{(v)}(t) \phi_{X,k}(t) dt \right)^2 \right|^2
\]

\[
= E \left| \left( \int (X_i(t) - X_i^{(v)}(t)) \phi_{X,k}^2(t) dt \right) \left( \int (X_i(t) + X_i^{(v)}(t)) \phi_{X,k}(t) dt \right) \right|^2
\]

\[
\leq E \|X_i - X_i^{(v)}\|^2 \|X_i + X_i^{(v)}\|^2
\]

\[
\leq (E \|X_i - X_i^{(v)}\|^4)^{1/2} (E \|X_i + X_i^{(v)}\|^4)^{1/2}
\]

\[
\leq c (E \|X_i - X_i^{(v)}\|^4)^{1/2}.
\]

The same argument can be applied to \( \{\beta_{Y,i,k}^{(v)}\} \) which implies that the process \( \{(U_X^i, U_Y^i)\} \) is \( L^2 \)-\( m \)-approximable and hence satisfies the functional central limit theorem. The rest of the proof is analogous to the proof of Lemma 3.6.3. \( \square \)

**Lemma 3.6.6.** Under Assumptions 3.3.2-3.3.4, we have

\[
N^{-1} \sum_{t=1}^{N_1} \sup_{m \in \mathbb{N}} |R_{X,k}^m|^2 = O_p(1), \tag{3.27}
\]
and \(N^{1/2} \sup_{m \in \mathbb{N}} |R_{X,N}^m| = o_p(1)\) as \(N \to \infty\). The same conclusion also holds for the remainder term \(\{R_{Y,t}^m\}\).

**Proof.** Recall that \(v_{i,k}^j = \omega_{X_i}^j - E[\omega_{X_i}^j]\). Note first that for any \(1 \leq a \leq M\) with \(M\) fixed and \(1 \leq k \leq N\), we have

\[
E \sup_{m \in \mathbb{N}} |R_{X,t}^{m,a}|^2 \leq l^2 E \sup_{m \in \mathbb{N}} \left\{ \frac{6 \sup_{|e| \equiv 1} e' (S_{X,t}^m - \Sigma_{X,t}^m) e}{\min_{s \neq a} |\lambda_s^X - \lambda_i^X|} \right\}^2 \leq c l^2 E \sup_{m \in \mathbb{N}} \left\{ \sum_{j,k=1}^m \left( \frac{1}{l} \sum_{i=1}^l \beta_{X,j} \beta_{X,k} - \delta_{jk} \lambda_i^X \right)^2 \right\} \leq c l^2 \sum_{j,k=1}^m \frac{1}{l} \sum_{i=1}^l \beta_{X,j} \beta_{X,k} - \delta_{jk} \lambda_i^X \lambda_i^X, \]

where we have used the inequality that \(\sup_{|e| \equiv 1} e'B^2e \leq \text{tr}(BB') = \sum_{j,k=1}^m b_{jk}^2\) for a symmetric matrix \(B = (b_{jk})_{j,k=1}^m\). Using the fact that

\[
E[v_{X_1}^{j,k} v_{X_2}^{j,k} v_{X_3}^{j,k'} v_{X_4}^{j,k'}] = \text{cum}(v_{X_1}^{j,k}, v_{X_2}^{j,k}, v_{X_3}^{j,k'}, v_{X_4}^{j,k'}) + r_{X}^{j,k}(i_2 - i_1)v_{X}^{j,k'}(i_4 - i_3) + r_{X}^{j,k}(i_3 - i_1)v_{X}^{j,k'}(i_4 - i_2) + r_{X}^{j,k}(i_3 - i_2)v_{X}^{j,k'}(i_4 - i_1),
\]

with \(r_{X}^{j,k}(h) = r_{X}^{j,k}(h)\), we obtain \(\sup_{m \in \mathbb{N}} |R_{X,t}^{m,a}|^2 \leq c (I_1 + I_2 + I_3 + I_4)\). Under Assumption 3.3.4, we have

\[
|I_1| \leq \frac{1}{l^2} \sum_{j,k} \sum_{j',k'} \sum_{i_1 + i_2 + i_3 + i_4 \in \mathbb{Z}} |\text{cum}(v_{X_1}^{j,k}, v_{X_2}^{j,k}, v_{X_3}^{j,k'}, v_{X_4}^{j,k'})| < \frac{c}{7}.
\]

Further, it is not hard to see that

\[
|I_2| \leq c \left( \sum_{j,k} \sum_{h=-\infty}^{+\infty} |r_{X}^{j,k}(h)| \right)^2 < +\infty, \quad |I_3| \leq c \left( \sum_{j,k} \sum_{j',k'} \sum_{h=-\infty}^{+\infty} |r_{X}^{j,k}(h)| \right)^2 < +\infty,
\]

for \(s = 3, 4\). Therefore, we get \(N^{-1} \sum_{t=1}^{N_1} \sup_{m \in \mathbb{N}} |R_{X,t}^{m,a}|^2 = O_p(1)\). Using similar arguments above, we have

\[
E[N^{1/2} \sup_{m \in \mathbb{N}} |R_{X,N_1}^m|] \leq c N^{1/2} E \left( \frac{1}{N_1} \sum_{i=1}^{N_1} v_{X_i}^{j,k} \right)^2 \leq c N^{1/2} \frac{1}{N_1} \sum_{j,k} \sum_{h=-\infty}^{+\infty} |r_{X}^{j,k}(h)| \to 0.
\]

The conclusion follows by using the same arguments for the second sample. \(\square\)

**Proof of Theorem 3.3.3.** Let \(\hat{\lambda}_k = (\hat{\lambda}_{X,k}^1, \hat{\lambda}_{Y,k}^1)' = (\hat{\lambda}_{X,k}^1, \ldots, \hat{\lambda}_{X,k}^M, \hat{\lambda}_{Y,k}^1, \ldots, \hat{\lambda}_{Y,k}^M)'\). Using the Ba-
From (3.29), we know that, for 1 ≤ j ≤ M, we have

\[ \frac{|N_r|}{\sqrt{N}} (\hat{\lambda}_{[N_r]} - \lambda) = \frac{|N_r|}{\sqrt{N}} (\hat{\tau}^{m}_{[N_r]} - \lambda) + \frac{|N_r|}{\sqrt{N}} (\hat{\lambda}_{[N_r]} - \hat{\tau}^{m}_{[N_r]}) \]

\[ = \frac{1}{\sqrt{N}} \sum_{i=1}^{|N_r|} \left( U_i X_i^T \right) + \frac{|N_r|}{\sqrt{N}} R_{[N_r]} + \frac{|N_r|}{\sqrt{N}} (\hat{\lambda}_{[N_r]} - \hat{\tau}^{m}_{[N_r]}). \] (3.28)

By Lemma 3.2 in Hörmann and Kokoszka (2010), we know that

\[ \sup_{|N_r| \leq k \leq N} |\hat{\lambda}_{X,k} - \lambda_{X,k}'| \leq c \sup_{|N_r| \leq k \leq N} ||\hat{C}_{X,k} - C_X||_{HS}, \]

for 1 ≤ j ≤ M. Recall that \( Z_{X,k}(t,s) = X_k(t)X_k(s) - C_X(t,s) \). By the arguments in the proof of Lemma 3.6.4, we can show that

\[ E \sup_{|N_r| \leq k \leq N} ||\hat{C}_{X,k} - C_X||^2_{HS} \leq E \max_{|N_r| \leq k \leq N} \left\{ \frac{1}{k} \sum_{i=1}^k Z_{X,i}(t,s) \right\}^2 \int \int \text{dtds} \]

\[ \leq \frac{1}{e^2 N^2} E \max_{1 \leq k \leq N} \left\{ \sum_{i=1}^k Z_{X,i}(t,s) \right\}^2 \int \int \text{dtds} = o(1). \] (3.29)

Hence we know that \( \sup_{|N_r| \leq k \leq N} |\hat{\lambda}_{X,k} - \lambda_{X,k}'| = o_p(1) \), which implies that the event \( E_X := \{ \hat{\lambda}_{X,k} > \hat{\lambda}_{X,k}' \} \) for all \( |N_r| \leq k \leq N \} \) occurs with probability tending to 1. The same arguments apply to the second sample. Let \( \Pi_{X,\infty} = \sum_{j=1}^{+\infty} <\phi_{X,j}^2, \cdot > \phi_{X,j}^2 \). By the properties of Hilbert-Schmidt norm, we have

\[ \sup_{r \in [e,1]} ||\Pi_{X,m} \hat{C}_{X,[N_r]} \Pi_{X,m} - \hat{C}_{X,[N_r]}||_{HS} \]

\[ \leq \sup_{r \in [e,1]} \left( ||\Pi_{X,m} - \Pi_{X,\infty}||_{HS} \right) \left( ||\Pi_{X,m} \hat{C}_{X,[N_r]} \Pi_{X,m}||_{HS} + ||\hat{C}_{X,[N_r]} (\Pi_{X,m} - \Pi_{X,\infty})||_{HS} \right) \]

\[ \leq 2 \sup_{r \in [e,1]} ||\hat{C}_{X,[N_r]} (\Pi_{X,m} - \Pi_{X,\infty})||_{HS} = 2 \left( \sup_{r \in [e,1]} ||\Pi_{X,m} - \Pi_{X,\infty}||_{HS} \right)^{1/2} \]

\[ = 2 \left( \sup_{r \in [e,1]} \left\{ \frac{1}{m+1} \sum_{j=m+1}^{+\infty} (\hat{\lambda}_{X,j,[N_r]} - \hat{\phi}_{X,j,[N_r]}^2)^2 \right\}^{1/2} \right) \]

\[ = 2 \left( \sup_{r \in [e,1]} \left\{ \frac{1}{m+1} \sum_{j=m+1}^{+\infty} \left( \hat{\lambda}_{X,j,[N_r]}^2 - \phi_{X,j,[N_r]}^2 \right) < \phi_{X,j,[N_r]}^2 \right\}^{1/2} \right). \]

From (3.29), we know \( P(\sup_{|N_r| \leq k \leq N} ||\hat{C}_{X,k}||_{HS} \leq 2 ||C_X||_{HS}) \to 1 \). On the event \( \{ \sup_{|N_r| \leq k \leq N} ||\hat{C}_{X,k}||_{HS} \leq \)}
2\|C_X\|_{HS}, we get
\[
\sup_{r \in [r,1]} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} (\lambda^i_{X,[N_r]})^2 \leq \sup_{r \in [r,1]} \sum_{j=1}^{\infty} (\lambda^j_{X,[N_r]})^2 \leq 2\|C_X\|^2_{HS},
\]
which implies that \( \sup_{r \in [r,1]} \|\Pi_{X,m}\hat{C}_{X,[N_r]}\Pi_{X,m} - \hat{C}_{X,[N_r]}\|_{HS} \rightarrow 0 \) as \( m \rightarrow +\infty \) for any fixed \( N \).

Then on the event \( \cap_{1 \leq j \leq M} (E^i_X \cap E^j_Y) \cap \{ \sup_{|N_r| \leq k \leq N} \|\hat{C}_{X,k}\|_{HS} \leq 2\|C_X\|_{HS} \}, \) we have
\[
\limsup_{m \rightarrow +\infty} \sup_{r \in [r,1]} |\hat{\lambda}_{[N_r]}| - \hat{\tau}^m_{[N_r]} = 0.
\]

Recall that \( \hat{\theta}_k = (\hat{\theta}_1, \ldots, \hat{\theta}_k) \) with \( \hat{\theta}_k = \hat{\lambda}^i_{X,[kN_1/N]} - \hat{\lambda}^j_{Y,[kN_2/N]} \). Take \( \limsup_{m \rightarrow +\infty} \) elementwise in (328), we get
\[
\frac{|N_r|}{\sqrt{N}} (\hat{\lambda}_{[N_r]} - \lambda) = \frac{1}{\sqrt{N}} \sum_{i=1}^{[N_r]} \left( \begin{array}{c} U_X^i \\ U_Y^i \end{array} \right) + \frac{|N_r|}{\sqrt{N}} \limsup_{m \rightarrow +\infty} R^m_{X,[N_r]},
\]
which implies that
\[
\frac{k}{\sqrt{N}} \{ \hat{\theta}_k - (\hat{\lambda}^i_{X} - \hat{\lambda}^j_{Y}) \} = \frac{1}{\sqrt{N}} \left\{ k \sum_{i=1}^{[kN_1/N]} U_X^i - k \sum_{i=1}^{[kN_2/N]} U_Y^i \right\} + \frac{k}{\sqrt{N}} \left( \limsup_{m \rightarrow +\infty} R^m_{X,[kN_1/N]} - \limsup_{m \rightarrow +\infty} R^m_{Y,[kN_2/N]} \right),
\]
with \( N_1 \leq k \leq N \). Notice that
\[
\left| \limsup_{m \rightarrow +\infty} R^m_{X,[kN_1/N]} - \limsup_{m \rightarrow +\infty} R^m_{Y,[kN_2/N]} \right| \leq \sup_{m} |R^m_{X,[kN_1/N]}| + \sup_{m} |R^m_{Y,[kN_2/N]}|.
\]

By Lemma 3.6.5 and Lemma 3.6.6, the assumptions in Theorem 2.1 of Shao (2010) are satisfied. Thus by similar arguments, we get \( G_{SN,N}^{(2)}(M) \rightarrow^d W_M(\varepsilon) \) under \( H_{2,0} \) and \( \lim_{|L| \rightarrow +\infty} \lim_{N \rightarrow +\infty} G_{SN,N}^{(2)}(M) = +\infty \) under \( H_{2,a} \).

\[ \square \]

**Remark 3.6.1.** Suppose \( \{X(t)\} \) are on a finite dimensional space spanned by \( \{\phi^1_X, \ldots, \phi^{m_0}_X\} \). Then it is easy to see that \( \hat{C}_{X,k} = \Pi_{X,m}\hat{C}_{X,k}\Pi_{X,m} \) for \( 2 \leq k \leq N_1 \) and \( m \geq m_0 \), which implies that \( \hat{\lambda}_{[N_r]} = \hat{\tau}^m_{[N_r]} \) for \( r \in (0,1] \). As seen from the proof of Theorem 3.3.3, trimming is not required in this case as the arguments go through without approximating the compact operator \( \hat{C}_{X,k} \) with a sequence of finite rank operators.
Proof of Proposition 3.3.2. Note that $X_m(t) = \sum_{j=0}^{\infty} b_j z_{m-j}(t) = \sum_{i=1}^{\infty} \sqrt{\lambda_i} \left( \sum_{j=0}^{\infty} b_j z_{i,m-j} \right) \phi_i(t)$, where $z_{k,j}$’s are independent standard normal random variables. Hence we get

$$u_{X_m}^{jk} = \sqrt{\lambda_j \lambda_k} \left( \sum_{i=0}^{\infty} b_i z_{j,m-i} \right) \left( \sum_{i=0}^{\infty} b_i z_{k,m-i} \right) = \sqrt{\lambda_j \lambda_k} \sum_{i,i'=0}^{\infty} b_i b_i' z_{j,m-i} z_{k,m-i'},$$

which implies that $r^{jk,j'k'}_{X}(h) = \sqrt{\lambda_j \lambda_k \lambda_j' \lambda_k'} \sum_{i,i_2,i_3,i_4} b_1 b_2 b_3 b_4 \text{cov}(z_{j,m-i_1} z_{k,m-i_2}, z_{j',m-i_3} z_{k',m+h-i_4}).$

We then have

$$\sum_{j,k} \sum_{h=\infty}^{+\infty} \left( \sum_{j,k}^{+\infty} |r^{jk,j'k'}_{X}(h)| \right)^2 \leq \sum_{j,k} 2 \lambda_j \lambda_k \sum_h \left( \sum_i b_i z_{h} \right)^2 \lambda_j \lambda_k \sum_h \left( \sum_i b_i z_{h} \right)^2 < \infty,$$

and $\sum_{j,k} \sum_{h=\infty}^{+\infty} |r^{jk}_{X}(h)| \leq \left( \sum_{j,k} \lambda_j + \sum_j \lambda_j^2 \right) \sum_h \left( \sum_i b_i z_{h} \right)^2 < \infty$, where we have used the fact $\text{cov}(X_1 X_2, X_3 X_4) = \text{cov}(X_1, X_3) \text{cov}(X_2, X_4) + \text{cov}(X_1, X_4) \text{cov}(X_2, X_3)$ for 4-variate normal random variable $(X_1, X_2, X_3, X_4)$. Notice that

$$\text{cum}(v_{X_0}^{jk}, v_{X_{i_1}}^{jk}, v_{X_{i_2}}^{j'k'}, v_{X_{i_3}}^{j'k'}) = \lambda_j \lambda_k \lambda_j' \lambda_k' \prod_{l_1,\ldots,l_8}^{+\infty} b_{i_t},$$

$$\text{cum}(z_{j,m-i_1} z_{k,m-i_2}, z_{j,m+i_1-i_3} z_{k,m+i_1-i_4}, z_{j',m+i_2-i_5} z_{k',m+i_2-i_6}, z_{j',m+i_3-i_7} z_{k',m+i_3-i_8}).$$

We may decompose the cumulant using equation (2.3.7) in Brillinger (1975). Some typical terms are the following:

$$\text{cov}(z_{j,m-i_1} z_{k,m+i_1-i_4}) \text{cov}(z_{k,m-l_2} z_{j',m+i_2-l_5} z_{k',m+i_3-l_6}),$$

$$\text{cov}(z_{j,m-i_1} z_{k,m+i_1-i_4}) \text{cov}(z_{j',m+i_2-l_5} z_{k',m+i_3-l_6}).$$

It is not hard to see that each term is bounded provided that

$$\sum_{i_1,i_2,i_3} |\pi(i_1)\pi(i_2)\pi(i_3-i_1)\pi(i_3-i_2)| \leq c \sum_{i_1,i_2,i_3} |\pi(i_1)\pi(i_2)\pi(i_3-i_1)| \leq c \left( \sum_h |\pi(h)| \right)^3 < \infty.$$

To prove Theorem 3.3.4, we begin with the Bahadur representation of the recursive estimates of the eigenfunctions. By Lemma A of Kneip and Utikal (2001), we get

$$\xi^a(S_{X,k}^m - e^{m,a} - \tilde{S}_{X,k}^m(S_{X,k} - \sum X)^m e^{m,a} + \tilde{R}_{X,k}^m, \tilde{R}_{X,k}^m).$$
where

\[ |\tilde{R}^{m,a}_{X,k}| \leq \frac{6 \sup_{|\xi| = 1} e'(\tilde{\Sigma}^m_{X,k} - \Sigma^m_X)^2 e}{\min_{s \neq a} |\lambda^s_X - \lambda^a_X|^2}, \]

\[ S^m_X = \sum_{s \neq a} \frac{1}{X^s_X - X^a_X} \epsilon^{m,s}(\epsilon^{m,s})', \] and \( \epsilon^{m,1}, \epsilon^{m,2}, \ldots, \epsilon^{m,m} \) are \( m \)-dimensional unit vectors. It follows that

\[ \hat{\gamma}^{m,a}_{X,k}(t) - \phi^a_X(t) = (\xi^a(\hat{\Sigma}^m_{X,k}) - \epsilon^{m,a}) \Phi^{m,a}_X(t) \]

\[ = - \sum_{s \neq a} \frac{1}{X^s_X - X^a_X} \left\{ \frac{1}{k} \sum_{i=1}^k \beta_{X,i,s} \beta_{X,i,a} \right\} \phi^s_X(t) + \Phi^{m,a}_X(t)' \tilde{R}^{m,a}_{X,k}. \]

Let \( \nu_l = (\hat{\phi}_l^{a+1}, \ldots, \hat{\phi}_l^p) \) and further define \( \tilde{a}^{m,a}_X = (\hat{\phi}_X, \hat{\phi}_k^{a+1}, \ldots, \hat{\phi}_k^p)' \), \( \tilde{\Gamma}^{m,a}_{X,k} = (\hat{\gamma}^{m,a}_{X,k}, \hat{\phi}_k^{a+1}, \ldots, \hat{\phi}_k^p)' \), and \( \tilde{\Psi}^{m,a}_X = (\hat{\phi}_X, \hat{\phi}_k^{a+1}, \ldots, \hat{\phi}_k^p)' \) with \( \tilde{\Psi}^{m,a}_X = \tilde{\Psi}^a_X \). We thus get

\[ \tilde{\Gamma}^{m,a}_{X,k} - \tilde{\Psi}^{m,a}_X = - \sum_{1 \leq s \neq a \leq m} \left\{ \frac{1}{k} \sum_{i=1}^k \beta_{X,i,s} \beta_{X,i,a} \right\} \tilde{\Psi}^{s,a}_X + \tilde{\Psi}^{m,a}_X (m, a) \tilde{R}^{m,a}_{X,k}, \]

where \( \tilde{\Psi}^{m,a}_X (m, a) = (\tilde{\Psi}^{1,a}_X, \ldots, \tilde{\Psi}^{m,a}_X) \in \mathbb{R}^{(p-a) \times m} \). Note that on the event \( \cap_{1 \leq j \leq M} (E^j_X \cap E^j_Y) \), we have

\[ |\tilde{a}^{m,a}_X - \tilde{\Gamma}^{m,a}_{X,k}| \leq c |\hat{\phi}^{m,a}_X| \to 0, \quad \text{as } m \to \infty, \]

for \( N \leq k \leq N \). Then we have

\[ \tilde{a}^{m,a}_X - \tilde{\Psi}^{m,a}_X = \frac{1}{k} \sum_{i=1}^k \left\{ - \sum_{1 \leq s \neq a < \infty} \frac{\tilde{\Psi}^{s,a}_X \nu^{m,a}_{X,i}}{X^s_X - X^a_X} \right\} + \limsup_{m \to \infty} \tilde{\Psi}^{m,a}_X (m, a) \tilde{R}^{m,a}_{X,k}. \]

Under Assumption 3.3.4, it is not hard to show that \( N^{-1} \sum_{i=1}^{N^i} \sup_{m \in \mathbb{N}} |\tilde{\Psi}^{m,a}_X (m, a) \tilde{R}^{m,a}_{X,i}|^2 = O_p(1) \) and \( N^{1/2} \sup_{m \in \mathbb{N}} |\tilde{\Psi}^{m,a}_X (m, a) \tilde{R}^{m,a}_{X,i}| = o_p(1) \) for \( 1 \leq a \leq M \) by using similar arguments in the proof of Lemma 3.6.6. Defining analogous quantities for the second sample with the subscript \( Y \), we can obtain the same result for the second sample.

**Lemma 3.6.7.** Define \( \tilde{h}^{a}_{X,i} = - \sum_{s \neq a} \left( \frac{\tilde{\Psi}^{s,a}_X w^{s,a}}{X^s_X - X^a_X} \right) \) and \( \tilde{h}^{a}_{Y,i} = - \sum_{s \neq a} \left( \frac{\tilde{\Psi}^{s,a}_Y w^{s,a}}{X^s_Y - X^a_Y} \right) \). Let \( \tilde{h}_i = (\hat{h}^{X}_{X,i}, \hat{h}^{X}_{Y,i}') = (\hat{h}^{X}_{X,i}, \ldots, \hat{h}^{X}_{X,i}, \hat{h}^{Y}_{X,i}, \ldots, \hat{h}^{Y}_{X,i})' \). Suppose Assumptions 3.3.1 holds. Then under the null, we have

\[ \frac{1}{\sqrt{N}} \left\{ \frac{1}{\gamma_1} \sum_{i=1}^{N^i} \hat{h}_{X,i} - \frac{1}{\gamma_2} \sum_{i=1}^{N^i} \hat{h}_{Y,i} \right\} \Rightarrow \tilde{\Lambda}_M B_{M_0}(r). \] (3.30)

Under Assumption 3.3.5 and the local alternative, (3.30) also holds. We have the same conclusion.
with Assumption 3.3.1 replaced by Assumption 3.3.2 and \( \gamma_1 = \gamma_2 \).

Proof of Lemma 3.6.7. We only prove the result under condition (3.14) as the derivation under (3.15) is straightforward (see Remark 3.6.2). Under condition (3.14), we have for \( a + 1 \leq j \leq p \),

\[
\left\{ E \left| \sum_{1 \leq s \neq a < \infty} \frac{< \phi_X^s, \tilde{\phi}^j >}{\lambda_X - \lambda_X^a} (\beta_{X,1,s} \beta_{X,1,a} - \beta_{X,1,s}^{(v)} \beta_{X,1,a}^{(v)}) \right|^2 \right\}^{1/2} \\
\leq c \sum_{1 \leq s \neq a < \infty} \left| < \phi_X^s, \tilde{\phi}^j > \right| \left\{ E \left( \beta_{X,1,s} \beta_{X,1,a} - \beta_{X,1,s}^{(v)} \beta_{X,1,a}^{(v)} \right)^2 \right\}^{1/2} \\
\leq c \sum_{1 \leq s < \infty} \left\{ E \left( \beta_{X,1,s} - \beta_{X,1,s}^{(v)} \right)^4 \right\}^{1/4} + c \left( E \left| X_1 - X_1^{(v)} \right|^4 \right)^{1/4} \sum_{1 \leq s < \infty} (E \beta_{X,1,s}^4)^{1/4},
\]

which implies that

\[
\sum_{v=1}^{\infty} \left\{ E \left| \sum_{1 \leq s \neq a < \infty} \frac{< \phi_X^s, \tilde{\phi}^j >}{\lambda_X - \lambda_X^a} (\beta_{X,1,s} \beta_{X,1,a} - \beta_{X,1,s}^{(v)} \beta_{X,1,a}^{(v)}) \right|^2 \right\}^{1/2} \\
\leq c \sum_{v=1}^{\infty} \sum_{s=1}^{\infty} \left\{ E \left( \beta_{X,1,s} - \beta_{X,1,s}^{(v)} \right)^4 \right\}^{1/4} + c \sum_{s=1}^{\infty} (E \beta_{X,1,s}^4)^{1/4} < \infty.
\]

Applying the same arguments to the second sample, we show that the sequence \( \{ \tilde{h}_k \} \) is \( L^2 \)-m-approximable. The conclusion follows from the arguments in the proof of Lemma 3.6.3.

**Remark 3.6.2.** Notice that under the assumption \( \phi_X^j = \tilde{\phi}_X^j = \phi^j \) for \( 2 \leq j \leq p \), \( \tilde{h}_{X,i} = - \left( \frac{w_{X,1}^{a+1}}{\lambda_{X,1} - \lambda_X^a}, \ldots, \frac{w_{X,p}^{a+1}}{\lambda_{X,p} - \lambda_X^a} \right) \) and the result in Lemma 3.6.7 can be established without the summability condition (3.14). In general, if we have \( \sum_{s=1}^{+\infty} | < \phi_X^s, \phi^j > | < +\infty \) and \( \sum_{s=1}^{+\infty} | < \phi_Y^s, \phi^j > | < +\infty \), for \( 2 \leq j \leq p \), then condition (3.14) can be dropped by noting that

\[
E \left( \beta_{X,1,s} \beta_{X,1,a} - \beta_{X,1,s}^{(v)} \beta_{X,1,a}^{(v)} \right)^2 \leq c \left( E \left| X_1 - X_1^{(v)} \right|^4 \right)^{1/4}.
\]

Proof of Theorem 3.3.4 and Proposition 3.3.5. Define \( \tilde{\Psi}_X^{s,a} \) and \( \tilde{h}_k \) by replacing \( v_i \) with \( \tilde{v}_i \) in \( \tilde{\Psi}_X^{s,a} \) and \( \tilde{h}_k \). We shall first show that

\[
\frac{1}{\sqrt{N}} \sup_{r \in [0,1]} \left\{ \sum_{i=1}^{[N_1 r]} \left| \tilde{h}_{X,i} - \tilde{h}_{X,i} \right| + \sum_{i=1}^{[N_2 r]} \left( \tilde{h}_{Y,i} - \tilde{h}_{Y,i} \right) \right\} = o_p(1), \quad (3.31)
\]
which suggests that the estimation effect caused by using $\hat{\phi}_{XY}$ instead of $\tilde{\phi}_{i}$ is asymptotically negligible. Using the fact that $||\hat{\phi}_{i}^{j} - \tilde{\phi}_{i}^{j}|| = O_{p}(1/\sqrt{N})$ and the Cauchy-Schwarz inequality, we have

$$\frac{1}{\sqrt{N}} \sup_{r \in [0,1]} \sum_{i=1}^{[N_{r}^{1}]} \sum_{s \neq a} \left( \tilde{\Psi}_{s,a}^{i} - \hat{\Psi}_{s,a}^{i} \right) \left( \lambda_{X}^{i} - \lambda_{X}^{a} \right) w_{s}^{i} \leq c \left( \sum_{s \neq a} \left( \tilde{\Psi}_{s,a}^{i} - \hat{\Psi}_{s,a}^{i} \right) \left( \lambda_{X}^{i} - \lambda_{X}^{a} \right) w_{s}^{i} \right)^{1/2}. \quad (3.32)$$

Under Assumptions 3.3.4, we can show that $E \left( \sum_{i=1}^{N_{1}^{1}} w_{s}^{i} \right) \leq N_{1} \tilde{g}_{s}$ with $\sum_{s=1}^{+\infty} \tilde{g}_{s} < +\infty$. Note that

$$\sum_{s \neq a} \max_{1 \leq k \leq N_{1}^{1}} \left( \sum_{i=1}^{k} w_{s}^{i} \right)^{2} \leq (\log \log 4N_{1})^{2}N_{1} \sum_{s \neq a} \tilde{g}_{s} \leq c(\log \log 4N_{1})^{2}N_{1}. \quad (3.33)$$

Then it is not hard to see the RHS of (3.32) is of order $o_{p}(1)$. The same arguments for the second sample imply (3.31). The rest of the proof follows similar arguments in the proofs of Theorem 3.3.3 and Lemma 3.6.7. We omit the details. \hfill \Box

Proof of Proposition 3.3.6. Note first that the assumptions in Proposition 3.3.2 are satisfied provided that $\sum_{j=1}^{+\infty} \sqrt{\lambda_{j}} < +\infty$ and $\sum_{m=1}^{+\infty} (\sum_{j=m}^{+\infty} b_{j}^{2})^{1/2} < +\infty$. Because $\beta_{X_{1},j}^{(m)} = \sqrt{\lambda_{j}} \left( \sum_{l=0}^{m-1} b_{l} z_{j,1-l} + \sum_{l=m}^{+\infty} b_{l} z_{j,1-l} \right)$, where $\{z_{j,k}^{(1)}\}$ is a sequence of independent copies of $\{z_{j,k}\}$, we have

$$\sum_{m=1}^{+\infty} \sum_{j=1}^{+\infty} \left( E \left( \beta_{X_{1},j}^{(m)} \right) \right) = \frac{1}{4} \sum_{j=1}^{+\infty} \sqrt{\lambda_{j}} \sum_{m=1}^{+\infty} \left( E \left( \sum_{l=m}^{+\infty} b_{l} (z_{j,1-l} - z_{j,1-l}^{(1)}) \right) \right) \leq \sum_{j=1}^{+\infty} \sqrt{\lambda_{j}} \sum_{m=1}^{+\infty} \left( \sum_{l=m}^{+\infty} b_{l}^{2} \right)^{1/2} < +\infty.$$

Using similar arguments, we can show that $\sum_{j=1}^{+\infty} \left( E \beta_{X_{1},j}^{(1)} \right)^{1/4} < +\infty.$ \hfill \Box
Table 3.1: Empirical sizes and size-adjusted powers of (i) the SN-based test, (ii) the PKM test and (iii) the CLRV test for testing the equality of the covariance operators. The nominal level is 5%.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$N_1 = N_2$</th>
<th>$K$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>$K^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A $\mathbf{v}_X = (12, 7, 0.5, 9, 5, 0.3)$</td>
<td>100</td>
<td>(i)</td>
<td>4.3</td>
<td>5.7</td>
<td>6.8</td>
<td>8.7</td>
<td>8.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ii)</td>
<td>14.5</td>
<td>20.9</td>
<td>22.9</td>
<td>32.2</td>
<td>32.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
<td>9.1</td>
<td>12.9</td>
<td>20.9</td>
<td>39.8</td>
<td>38.8</td>
</tr>
<tr>
<td></td>
<td>$\mathbf{v}_Y = (12, 7, 0.5, 9, 5, 0.3)$</td>
<td>200</td>
<td>(i)</td>
<td>4.7</td>
<td>5.7</td>
<td>4.6</td>
<td>7.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ii)</td>
<td>12.8</td>
<td>20.6</td>
<td>26.7</td>
<td>34.7</td>
<td>34.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
<td>6.9</td>
<td>9.6</td>
<td>14.5</td>
<td>25.2</td>
<td>25.1</td>
</tr>
<tr>
<td>B $\mathbf{v}_X = (14, 7, 0.5, 6, 5, 0.3)$</td>
<td>100</td>
<td>(i)</td>
<td>19.1</td>
<td>23.6</td>
<td>17.7</td>
<td>14.2</td>
<td>14.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ii)</td>
<td>27.6</td>
<td>37.7</td>
<td>31.6</td>
<td>22.9</td>
<td>23.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
<td>27.0</td>
<td>33.8</td>
<td>23.0</td>
<td>20.5</td>
<td>20.5</td>
</tr>
<tr>
<td></td>
<td>$\mathbf{v}_Y = (8, 7, 0.5, 6, 5, 0.3)$</td>
<td>200</td>
<td>(i)</td>
<td>31.2</td>
<td>37.7</td>
<td>30.4</td>
<td>21.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ii)</td>
<td>39.1</td>
<td>61.6</td>
<td>51.7</td>
<td>44.2</td>
<td>44.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
<td>37.6</td>
<td>57.0</td>
<td>44.7</td>
<td>30.1</td>
<td>30.1</td>
</tr>
<tr>
<td>C $\mathbf{v}_X = (12, 7, 0.5, 9, 3, 0.3)$</td>
<td>100</td>
<td>(i)</td>
<td>5.5</td>
<td>10.9</td>
<td>30.8</td>
<td>62.4</td>
<td>62.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ii)</td>
<td>4.7</td>
<td>16.1</td>
<td>57.3</td>
<td>94.6</td>
<td>94.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
<td>5.5</td>
<td>13.4</td>
<td>42.3</td>
<td>79.4</td>
<td>79.2</td>
</tr>
<tr>
<td></td>
<td>$\mathbf{v}_Y = (12, 7, 0.5, 3, 9, 0.3)$</td>
<td>200</td>
<td>(i)</td>
<td>5.3</td>
<td>10.0</td>
<td>45.7</td>
<td>90.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ii)</td>
<td>6.4</td>
<td>13.0</td>
<td>67.8</td>
<td>99.9</td>
<td>99.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
<td>6.1</td>
<td>12.5</td>
<td>60.5</td>
<td>99.9</td>
<td>99.9</td>
</tr>
<tr>
<td>D $\mathbf{v}_X = (12, 7, 0.5, 9, 5, 0.3)$</td>
<td>100</td>
<td>(i)</td>
<td>6.1</td>
<td>8.3</td>
<td>28.3</td>
<td>80.1</td>
<td>75.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ii)</td>
<td>5.5</td>
<td>14.6</td>
<td>47.2</td>
<td>100.0</td>
<td>94.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
<td>6.9</td>
<td>12.3</td>
<td>37.2</td>
<td>95.7</td>
<td>90.6</td>
</tr>
<tr>
<td></td>
<td>$\mathbf{v}_Y = (12, 7, 0.5, 0.5, 0.3)$</td>
<td>200</td>
<td>(i)</td>
<td>5.7</td>
<td>8.9</td>
<td>39.7</td>
<td>96.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ii)</td>
<td>6.4</td>
<td>14.5</td>
<td>53.6</td>
<td>100.0</td>
<td>99.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
<td>6.0</td>
<td>12.9</td>
<td>47.7</td>
<td>100.0</td>
<td>99.3</td>
</tr>
</tbody>
</table>

Note: Under the alternatives, we simulate the size-adjusted critical values by assuming that both $\{X_i\}$ and $\{Y_i\}$ are generated from (3.16) with $\rho = 0.5$, $\mu = 1$ and $\mathbf{v} = \mathbf{v}_X$. 
Table 3.2: Empirical sizes and size-adjusted powers of (i) the SN-based test, the subsampling-based test with (ii) $l = 8$, (iii) $l = 12$ and (iv) $l = 16$, and (v) Benko et al’s iid bootstrap based method for testing the equality of the first two eigenvalues separately (the columns with $M = 1, 2$) and jointly (the column with $M = (1, 2)$). The nominal level is 5% and the number of replications for iid bootstrap method is 250.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$N_1 = N_2$</th>
<th>$M$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbf{A}$</td>
<td>$v_X = (10, 0.5, 5, 0.3)$</td>
<td>48</td>
<td>(i)</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(ii)</td>
<td>24.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(iii)</td>
<td>21.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(iv)</td>
<td>21.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(v)</td>
<td>11.2</td>
</tr>
<tr>
<td>$v_Y = (10, 0.5, 5, 0.3)$</td>
<td>96</td>
<td>(i)</td>
<td>5.2</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(ii)</td>
<td>19.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(iii)</td>
<td>16.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(iv)</td>
<td>16.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(v)</td>
<td>14.4</td>
</tr>
<tr>
<td>$\mathbf{B}$</td>
<td>$v_X = (20, 0.5, 5, 0.3)$</td>
<td>48</td>
<td>(i)</td>
<td>25.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(ii)</td>
<td>24.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(iii)</td>
<td>19.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(iv)</td>
<td>14.1</td>
</tr>
<tr>
<td>$v_Y = (10, 0.5, 5, 0.3)$</td>
<td>96</td>
<td>(i)</td>
<td>48.4</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(ii)</td>
<td>58.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(iii)</td>
<td>50.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(iv)</td>
<td>53.8</td>
</tr>
<tr>
<td>$\mathbf{C}$</td>
<td>$v_X = (10, 0.5, 5, 0.3)$</td>
<td>48</td>
<td>(i)</td>
<td>6.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(ii)</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(iii)</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(iv)</td>
<td>6.1</td>
</tr>
<tr>
<td>$v_Y = (10, 0.5, 1, 0.3)$</td>
<td>96</td>
<td>(i)</td>
<td>4.7</td>
<td>91.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(ii)</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(iii)</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(iv)</td>
<td>5.4</td>
</tr>
<tr>
<td>$\mathbf{D}$</td>
<td>$v_X = (20, 0.5, 5, 0.3)$</td>
<td>48</td>
<td>(i)</td>
<td>27.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(ii)</td>
<td>25.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(iii)</td>
<td>20.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(iv)</td>
<td>14.9</td>
</tr>
<tr>
<td>$v_Y = (10, 0.5, 1, 0.3)$</td>
<td>96</td>
<td>(i)</td>
<td>55.3</td>
<td>87.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(ii)</td>
<td>54.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(iii)</td>
<td>48.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(iv)</td>
<td>50.1</td>
</tr>
</tbody>
</table>

Note: Under the alternatives, we simulate the size-adjusted critical values by assuming that both $\{X_i\}$ and $\{Y_i\}$ are generated from (3.16) with $\rho = 0.5$, $\mu = 0$ and $\nu = \nu_X$. 

66
Table 3.3: Empirical sizes of (i) the SN-based test, the subsampling-based test with (ii) \( l = 8 \), (iii) \( l = 12 \) and (iv) \( l = 16 \), and (v) Benko et al’s iid bootstrap based method for testing the equality of the first two eigenfunctions separately (the columns with \( M = 1, 2 \)) and jointly (the column with \( M = (1, 2) \)). The nominal level is 5\% and the number of replications for iid bootstrap is 250.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( N_1 = N_2 )</th>
<th>( M )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( v_X = (10, 0.5, 5, 0.3) )</td>
<td>48</td>
<td>(i)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ii)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iv)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(v)</td>
</tr>
<tr>
<td>( v_Y = (8, 0.5, 4, 0.3) )</td>
<td>96</td>
<td>(i)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ii)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iv)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(v)</td>
</tr>
<tr>
<td>( v_X = (4, 0.5, 2, 0.3) )</td>
<td>48</td>
<td>(i)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ii)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iv)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(v)</td>
</tr>
<tr>
<td>( v_Y = (2, 0.5, 1, 0.3) )</td>
<td>96</td>
<td>(i)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(ii)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iii)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(iv)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(v)</td>
</tr>
</tbody>
</table>

Table 3.4: Comparison of the eigenvalues and eigenfunctions of the covariance operators of the station observations and model outputs.

<table>
<thead>
<tr>
<th>( K )</th>
<th>( G_{SN,N}^{(2)}(M) )</th>
<th>( p )-value</th>
<th>( G_{SN,N}^{(3)}(M_0) )</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.8</td>
<td>(0.1, 1)</td>
<td>126.4</td>
<td>(0.025, 0.05)</td>
</tr>
<tr>
<td>2</td>
<td>5.4</td>
<td>(0.1, 1)</td>
<td>295.4</td>
<td>(0, 0.005)</td>
</tr>
<tr>
<td>3</td>
<td>119.9</td>
<td>(0.005, 0.01)</td>
<td>34.2</td>
<td>(0.1, 1)</td>
</tr>
<tr>
<td>4</td>
<td>326.2</td>
<td>(0.005, 0.01)</td>
<td>318.0</td>
<td>(0.005, 0.01)</td>
</tr>
</tbody>
</table>

Note: The first three rows show the results for testing individual eigencomponent, and the last row shows the results for testing the first three eigenvalues and eigenfunctions jointly.
Figure 3.1: Size-adjusted powers of the SN-based test and the subsampling-based tests for testing the equality of the first two eigenfunctions separately and jointly.
Figure 3.2: $p$-values for testing the nullity of lag zero cross-correlation between the station observations and model outputs at each location. The numbers 0-5 denote the ranges of the $p$-values, i.e., 0 denotes [0.1, 1]; 1 denotes [0.05, 0.1]; 2 denotes [0.025, 0.05]; 3 denotes [0.01, 0.025]; 4 denotes [0.005, 0.01] and 5 denotes [0, 0.005].
Figure 3.3: The first three PCs of the station observations (left panels) and model outputs (right panels), and the associated eigenvalues and percentage of variations explained.
Chapter 4

Fixed-smoothing asymptotics for time series

4.1 Introduction

Many economic and financial applications involve time series data with autocorrelation and heteroskedasticity properties. Often the unknown dependence structure is not the chief object of interest but the inference on the parameter of interest involves the estimation of unknown dependence. In stationary time series models estimated by generalized method of moments (GMM), robust inference is typically accomplished by consistently estimating the asymptotic covariance matrix, which is proportional to the long run variance (LRV) matrix of the estimating equations or moment conditions defining the estimator, using a kernel smoothing method. In the econometrics and statistics literatures, the bandwidth parameter/truncation lag involved in the kernel smoothing method is assumed to grow slowly with sample size in order to achieve consistency. The inference is conducted by plugging in a covariance matrix estimator that is consistent under heteroskedasticity and autocorrelation. This approach dates back to Newey and West (1987) and Andrews (1991). Recently, Kiefer and Vogelsang (2005) (KV, hereafter) developed an alternative first order asymptotic theory for the HAC (heteroskedasticity and autocorrelation consistent) based robust inference, where the proportion of the bandwidth involved in the HAC estimator to the sample size $T$, denoted as $b$, is held fixed in the asymptotics. Under the fixed-$b$ asymptotics, the HAC estimator converges to a nondegenerate yet nonstandard limiting distribution. The tests based on the fixed-$b$ asymptotic approximation were shown to enjoy better finite sample properties than the tests based on the small-$b$ asymptotic theory under which the HAC estimator is consistent and the limiting distribution of the studentized statistic admits a standard form, such as standard normal or $\chi^2$ distribution. Using the higher order Edgeworth expansions, Jansson (2004), Sun et al. (2008) and Sun (2010) rigorously proved that the fixed-$b$ asymptotics provides a high order refinement over the traditional small-$b$ asymptotics in the Gaussian location model. Sun et al. (2008) also provided an interesting decision theoretical justification for the use of fixed-$b$ rules in econometric testing. For non-Gaussian linear
processes, Gonçalves and Vogelsang (2011) obtained an upper bound on the convergence rate of the error in the fixed-$b$ approximation and showed that it can be smaller than the error of the normal approximation under suitable assumptions.

Since the seminal contribution by KV, there has been a growing body of work in econometrics and statistics to extend and expand the fixed-$b$ idea in the inference for time series data. For example, Sun (2012) developed a procedure for hypothesis testing in time series models by using the nonparametric series method. The basic idea is to project the time series onto a space spanned by a set of fourier basis functions (see Phillips, 2005; Müller, 2007, for early developments) and construct the covariance matrix estimator based on the projection vectors with the number of basis functions held fixed. Also see Sun (2011) for the use of a similar idea in the inference of the trend regression models. Ibragimov and Müller (2010) proposed a subsampling based $t$-statistic for robust inference where the unknown dependence structure can be in the temporal, spatial or other forms. In their paper, the number of non-overlapping blocks is held fixed. The $t$-statistic based approach was extended by Bester et al. (2011) to the inference of spatial and panel data with group structure. In the context of misspecification testing, Chen and Qu (2012) proposed a modified $M$ test of Kuan and Lee (2006) which involves dividing the full sample into several recursive subsamples and constructing a normalization matrix based on them. In the statistical literature, Shao (2010) developed the self-normalized approach to inference for time series data that uses an inconsistent LRV estimator based on recursive subsample estimates. The self-normalized method is an extension of Lobato (2001) from the sample autocovariances to more general approximately linear statistics and it coincides with KV’s fixed-$b$ approach in the inference of the mean of a stationary time series by using the Bartlett kernel and letting $b = 1$. Although the above inference procedures are proposed in different settings and for different problems and data structures, they share a common feature in the sense that the underlying smoothing parameters in the asymptotic covariance matrix estimators such as the number of basis functions, the number of cluster groups and the number of recursive subsamples, play a similar role as the bandwidth in the HAC estimator. Throughout the chapter, we shall call these asymptotics, where the smoothing parameter (or function of smoothing parameter) is held fixed, the fixed-smoothing asymptotics. In contrast, when the smoothing parameter grows with respect to sample size, we use the term increasing-domain asymptotics. At some places the terms fixed-$K$ (or fixed-$b$) and increasing-$K$ (or small-$b$) asymptotics are used to follow the convention in the literature.

In this chapter, we introduce a general class of estimators for estimating the LRV matrix in
the inference of stationary time series models estimated by GMM. Our proposal includes the traditional lag window type (or HAC) covariance estimator, the projection-based covariance estimator, the cluster-based covariance estimator and the blockwise recursive subsampling-based covariance estimator as special cases. The general covariance estimator considered here involves projecting the original data onto a space spanned by a sequence of basis functions (not necessarily orthogonal), where the number of basis functions $K$ plays a key role in determining asymptotic properties of the estimator. Under the fixed-$K$ asymptotics, we show that the Wald statistic based on the general LRV estimator converges to an (approximate) $F$ distribution with a scale constant depending only on $K$ and the number of restrictions being tested. Thus our result provides a unification of the various recently proposed fixed-smoothing inference procedures in the first order sense.

Furthermore, we derive higher order expansions of the finite sample distributions of the subsampling-based $t$-statistic and the Wald statistic with HAC covariance estimator when the underlying smoothing parameters are held fixed, under the framework of the Gaussian location model. Specifically, we show that the error in the rejection probability (ERP, hereafter) is of order $O(1/T)$ under the fixed-smoothing asymptotics. Under the assumption that the eigenfunctions of the kernel in the HAC estimator have zero mean and other mild assumptions, we derive the leading error term of order $O(1/T)$ under the fixed-smoothing framework. These results are similar to those obtained under the fixed-$b$ asymptotics (see Sun et al., 2008), but are stronger in the sense that we are able to derive the exact form of the leading error term with order $O(1/T)$. The explicit form of the leading error term in the approximation provides a clear theoretical explanation for the empirical findings in the literature regarding the direction and magnitude of size distortion for time series with various degrees of dependence. To the best of our knowledge, this is the first time that the leading error terms are made explicit through the higher order Edgeworth expansion under the fixed-smoothing asymptotics. It is also worth noting that our nonstandard argument differs from that in Jansson (2004) and Sun et al. (2008), and it may be of independent theoretical interest and be useful for future follow-up work.

Finally, we propose a novel bootstrap method for time series, the Gaussian dependent bootstrap, which is able to mimic the second order properties of the original time series and produces a Gaussian bootstrap sample. For the Gaussian location model, we show that the inference based on the Gaussian dependent bootstrap is more accurate than the first order approximation under the fixed-smoothing asymptotics. This seems to be the first time a bootstrap method is shown to be second order correct under the fixed-smoothing asymptotics; see Gonçalves and Vogelsang (2011) for a
recent attempt for the moving block bootstrap in the non-Gaussian setting.

We now introduce some notation. For a vector \( x = (x_1, x_2, \ldots, x_{q_0}) \in \mathbb{R}^{q_0} \), we let \( ||x|| = (\sum_{i=1}^{q_0} x_i^2)^{1/2} \) be the Euclidean norm. For a matrix \( A = (a_{ij})_{i,j=1}^{q_0} \in \mathbb{R}^{q_0 \times q_0} \), denote by \( ||A||_2 = \sup_{||x|| = 1} ||Ax|| \) the spectral norm and \( ||A||_{\infty} = \max_{1 \leq i,j \leq q_0} |a_{ij}| \) the max norm. Denote by \( \lfloor a \rfloor \) the integer part of a real number \( a \). Let \( L^2[0,1] \) be the space of square integrable functions on \([0,1]\). Denote by \( D[0,1] \) the space of functions on \([0,1]\) which are right continuous and have left limits, endowed with the Skorokhod topology (see Billingsley, 1999). Denote by “\( \Rightarrow \)” weak convergence in the \( \mathbb{R}^{q_0} \)-valued function space \( D^{q_0}[0,1] \), where \( q_0 \in \mathbb{N} \). Denote by “\( \rightarrow^d \)” and “\( \rightarrow^p \)” convergence in distribution and convergence in probability, respectively. The notation \( N(\mu, \Sigma) \) is used to denote the multivariate normal distribution with mean \( \mu \) and covariance \( \Sigma \). Let \( \chi_k^2 \) be a random variable following \( \chi^2 \) distribution with \( k \) degrees of freedom and \( G_k \) be the corresponding distribution function.

4.2 Basic setup and assumptions

In linear and nonlinear models with moment conditions, it is standard to employ GMM to estimate the model parameters. We follow the GMM setup as described in KV. Consider a \( d \times 1 \) vector of parameters \( \theta \in \Theta \subseteq \mathbb{R}^d \) of interest, where \( \Theta \) is the parameter space. Denote \( \theta_0 \) the true parameter of \( \theta \) which is an interior point of \( \Theta \). Let \( y_t \) denote a vector of observed data and assume the moment conditions

\[
E[f(y_t, \theta)] = 0, \quad t = 1, 2, \ldots, T
\]

hold if and only if \( \theta = \theta_0 \), where \( f(\cdot) \) is \( m \times 1 \) vector of functions with \( m \geq d \) and \( \text{rank}(E[\partial f(y_t, \theta_0)/\partial \theta']) = d \). When \( m > d \), the parameter \( \theta \) is over-identified with the degree of over-identification \( v = m - d \).

Define the partial sum \( g_t(\theta) = T^{-1} \sum_{j=1}^{t} f(y_j, \theta) \). Then the GMM estimator of \( \theta_0 \) is given by

\[
\hat{\theta}_T = \arg\min_{\theta \in \Theta} g_T(\theta)' W_T g_T(\theta),
\]

where \( W_T \) is a \( m \times m \) semi-positive definite weighting matrix. Further define

\[
G_t(\theta) = (G_{t1}(\theta), \ldots, G_{tm}(\theta))' = \frac{\partial g_t(\theta)}{\partial \theta'} = \frac{1}{T} \sum_{j=1}^{t} \frac{\partial f(y_j, \theta)}{\partial \theta'}.
\]
Using the mean value theorem for each element of $g_T$, we have $g_T(\hat{\theta}_T) = g_T(\theta_0) + \tilde{G}_T(\hat{\theta}_T - \theta_0)$, where $\tilde{G}_T = (G_{T1}(\hat{\theta}_{T1}), \ldots, G_{Tm}(\hat{\theta}_{Tm}))'$ and $\hat{\theta}_{Tj}$ is between $\theta_0$ and $\hat{\theta}_T$ for each $1 \leq j \leq m$. Note that $G_T(\hat{\theta}_T)'W_Tg_T(\hat{\theta}_T) = 0$ by the first order condition, which implies that

$$G_T(\hat{\theta}_T)'W_Tg_T(\hat{\theta}_T) = G_T(\hat{\theta}_T)'W_Tg_T(\hat{\theta}_T) = 0.$$ 

Solving the above equation, we have

$$T^{1/2}(\hat{\theta}_T - \theta_0) = -(G_T(\hat{\theta}_T)'W_T\tilde{G}_T)^{-1}G_T(\hat{\theta}_T)'W_T(T^{1/2}g_T(\theta_0)).$$

To derive the asymptotic distribution of $\hat{\theta}_T$, we make the following high-level assumptions as KV and Sun (?).

**Assumption 4.2.1.** $\hat{\theta}_T \to^p \theta_0$.

**Assumption 4.2.2.** $T^{1/2}g_{Tj}(\theta_0) \Rightarrow \Delta W_m(r)$ where

$$\Delta \Delta' = \Omega = \sum_{j=-\infty}^{+\infty} E[f(y_t, \theta_0)f(y_{t-j}, \theta_0)'],$$

and $W_m(r)$ is a $m$--dimensional vector of independent standard Brownian motions.

**Assumption 4.2.3.** $\tilde{G}_T \to^p G_0$ uniformly for all $\hat{\theta}_{Tj}$ between $\hat{\theta}_T$ and $\theta_0$, where $G_0 = E[\partial f(y_j, \theta_0)/\partial \theta']$ and $1 \leq j \leq m$.

**Assumption 4.2.4.** The weighting matrix $W_T$ is symmetric and semi-positive definite such that $W_T \to^p W_0$ and $G_0'W_0G_0$ is positive definite.

Under Assumptions 4.2.1-4.2.4, it is easy to see that

$$T^{1/2}(\hat{\theta}_T - \theta_0) \to^d - (G_0'W_0G_0)^{-1}G_0'W_0\Delta W_m(1) =^d N(0, V_0),$$

where “$=^d$” denotes “equal in distribution” and the asymptotic covariance matrix

$$V_0 := (G_0'W_0G_0)^{-1}G_0'W_0\Omega W_0G_0(G_0'W_0G_0)^{-1}.$$ 

To make inference on $\theta_0$, we have to estimate $G_0$, $W_0$ and the LRV matrix $\Omega$. Under the above assumptions, $G_0$ and $W_0$ can be consistently estimated by their sample counterparts $G_T(\hat{\theta}_T)$ and

75
respectively. It remains to estimate the LRV matrix \( \Omega \). In the next section, we introduce a general class of estimators for \( \Omega \) and \( V_0 \).

### 4.3 LRV estimators

To present the idea, we focus on the hypothesis testing problem that \( H_0 : r(\theta_0) = 0 \) versus the alternative that \( H_a : r(\theta_0) \neq 0 \), where \( r(\theta) \) is a \( p \times 1 \) continuously differentiable function with the first order derivative matrix \( R(\theta) = \partial r(\theta)/\partial \theta' \) and \( p \leq d \). Let

\[
\hat{V}_T = (G_T(\hat{\theta}_T)'W_T G_T(\hat{\theta}_T))^{-1}(G_T(\hat{\theta}_T)'W_T \hat{\Omega}_T W_T G_T(\hat{\theta}_T))(G_T(\hat{\theta}_T)'W_T G_T(\hat{\theta}_T))^{-1},
\]

be an estimator of \( V_0 \), where \( \hat{\Omega}_T \) is the LRV estimate of \( \Omega \). The Wald statistic for testing \( H_0 \) against \( H_a \) is defined as

\[
F_T = Tr(\hat{\theta}_T)' \hat{D}_T^{-1} r(\hat{\theta}_T)/p, \tag{4.3}
\]

where \( \hat{D}_T = R(\hat{\theta}_T) \hat{V}_T R(\hat{\theta}_T)' \). The widely used lag window type LRV estimator is given by

\[
\hat{\Omega}_T = \frac{1}{T} \sum_{i=1}^{T} \sum_{j=1}^{T} K \left( \frac{i-j}{bT} \right) f(y_i, \hat{\theta}_T)f(y_j, \hat{\theta}_T)', \tag{4.4}
\]

where \( K(\cdot) \) is a kernel function and \( b \) is the proportion of the truncation lag to the sample size. By setting

\[
\hat{u}_i = R(\hat{\theta}_T)(G_T(\hat{\theta}_T)'W_T G_T(\hat{\theta}_T))^{-1} G_T(\hat{\theta}_T)'W_T f(y_i, \hat{\theta}_T),
\]

we have

\[
\hat{D}_T = \frac{1}{T} \sum_{i=1}^{T} \sum_{j=1}^{T} K \left( \frac{i-j}{bT} \right) \hat{u}_i \hat{u}_j'.
\]

When \( K(\cdot) \) is semi-positive definite, by Mercer’s theorem, we have the spectral decomposition,

\[
K(r-t) = \sum_{j=1}^{+\infty} \lambda_j \phi_j(r)\phi_j(t), \quad 0 \leq r, t \leq 1/b, \tag{4.5}
\]

where \( \{\lambda_j\} \) and \( \{\phi_j\} \) are the eigenvalues and orthonormal eigenfunctions corresponding to the kernel function respectively. We thus have the representation,

\[
\hat{D}_T = \sum_{s=1}^{K} \lambda_s \left\{ \frac{1}{\sqrt{T}} \sum_{i=1}^{T} \phi_s \left( \frac{i}{bT} \right) \hat{u}_i \right\} \left\{ \frac{1}{\sqrt{T}} \sum_{j=1}^{T} \phi_s \left( \frac{j}{bT} \right) \hat{u}_j' \right\},
\]
with \( K = +\infty \). In the traditional asymptotics, \( b \) goes to zero as \( T \) increases which is referred as the small-\( b \) asymptotics. When \( b \in (0,1] \) is held fixed, it corresponds to the fixed-\( b \) asymptotics in KV. As pointed out in some recent studies (see e.g., Bester et al., 2011; Sun, 2011, 2012; Chen and Qu, 2012), \( K \) can also be held as a fixed positive integer, which can lead to a more accurate first order approximation. In light of these recent findings, we introduce a general class of estimators to estimate the LRV matrix. With a slight abuse of notation, we let \( \{\phi_s(t)\}_{s=1}^K \) be a sequence of linearly independent functions in \( L^2[0,1/b] \) and \( \{\lambda_j\} \) be a sequence of nonnegative weights such that \( \sum_{j=1}^K \lambda_j = 1 \). A set of elements \( \{\psi_i\}_{i=1}^K \) in a real valued vector space is called linearly independent if and only if \( \sum_{i=1}^K a_i \psi_i = 0 \) \( \Rightarrow \) \( a_i = 0 \) for \( i = 1,2,\ldots,K \). Here \( 0 \) denotes the null element in the vector space. Note that \( \lambda_j \)'s in (4.5) are nonnegative when we consider semi-positive definite kernels in (4.4). Further let \( V_s = \frac{1}{bT} \sum_{i=1}^T \phi_s \left( \frac{i}{bT} \right) \hat{u}_i \), be the normalized inner product between \( \{\hat{u}_i\}_{i=1}^T \) and \( \{\phi_s(i/(bT))\}_{i=1}^T \). Define \( R = (R_{ij})_{i,j=1}^K \) with \( R_{ij} = \int_0^1 \hat{\phi}_s(t/b)\hat{\phi}_s(t/b)dt \), where \( \hat{\phi}_s(t/b) = \phi_s(t/b) - \int_0^1 \phi_s(t/b)dt \), and \( L = (L_{ij})_{i,j=1}^K \) an upper triangular matrix based on the Cholesky decomposition of \( R^{-1} \), i.e., \( L'L = R^{-1} \). Define \( V = (V_1', V_2', \ldots, V_K')' \) and

\[
V^* = (V_1^*, V_2^*, \ldots, V_K^*)' = (L \otimes I_p)V,
\]

where \( V_i^* = \sum_{j=1}^K L_{ij}V_j \) for \( 1 \leq i \leq K \). Then the general LRV estimator is given by

\[
\hat{D}_T = \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^T \left\{ \sum_{s=1}^K \lambda_s \sum_{m=1}^K L_{sm} \phi_m \left( \frac{i}{bT} \right) \sum_{l=1}^K L_{tl} \hat{\phi}_l \left( \frac{j}{bT} \right) \right\} \hat{u}_i \hat{u}_j',
\]

and the test statistic based on the general LRV estimator is defined as,

\[
F_T = [\sqrt{Tr(\hat{\theta}_T)^{-1}D_T}] [\sqrt{Tr(\hat{\theta}_T)}]/p.
\]

The matrix \( R \) is introduced for orthogonalization so that the limiting distribution of the test statistic \( F_T \) does not depend on the basis functions. Note that the choice of \( R \) is not unique (See Example 4.3.3). In what follows, we shall show that the recently developed nonparametric series covariance estimator (Sun, 2011, 2012), the recursive subsampling-based covariance estimator (Chen and Qu, 2012) and the cluster covariance estimator (CCE) (Bester et al., 2011) are all special cases of the general LRV estimator. Throughout Examples 4.3.1-4.3.3, we set \( b = 1 \) and \( \lambda_j = 1/K \) for \( j = 1,2,\ldots,K \).

**Lemma 4.3.1.** Let \( \{\phi_s(t)\}_{s=1}^K \) be a sequence of orthonormal basis functions with \( \int_0^1 \phi_s(t)dt = 0 \).
Then we have $R = I_{K \times K}$ and $\hat{D}_T = \frac{1}{T} \sum_{j=1}^{K} V_j V_j'$, where $V_i = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \phi_{ij}(i/T) \hat{u}_i$. When $\phi_s(t) = \sqrt{2} \sin(2\pi s t)$ (or $\phi_s(t) = \sqrt{2} \cos(2\pi s t)$), $s = 1, 2, \ldots, K$, it is straightforward to see that the LRV estimator corresponds to the series estimator considered in Sun (2011, 2012). In this case, the LRV estimator involves projecting the data onto a set of orthonormal basis and using the sample variance of the projection vectors, namely $\hat{D}_T$.

**Lemma 4.3.2.** For any fixed $K$ with $K \leq T$, we consider the basis function $\phi_s(t) = I\{0 < t \leq s/(K + 1)\}$, $s = 1, 2, \ldots, K$, where $I$ denotes the indicator function. Simple calculation gives us $R_{ij} = \int_0^1 \hat{\phi}_i(t) \hat{\phi}_j(t) dt = \min(i, j)/(K + 1) - (ij)/(K + 1)^2$, and $\hat{D}_T = \frac{1}{K} \sum_{s=1}^{K} V_s^* V_s'$, where

$$V_s^* = \sqrt{\frac{K + 1}{T}} \left( \sqrt{\frac{s + 1}{s}} \sum_{i=1}^{\lfloor \frac{T}{s+1} \rfloor} \hat{u}_i - \sqrt{\frac{s}{s + 1}} \sum_{i=1}^{\lfloor \frac{T}{s+1} \rfloor} \hat{u}_i \right),$$

with $s = 1, 2, \ldots, K$ and $V_{K+1} = 0$. Therefore, the general LRV estimator reduces to the recursive subsampling-based estimator in Chen and Qu (2012), where the idea is to divide the full sample into $K + 1$ recursive subsamples and construct a normalization matrix based on the subsamples.

**Lemma 4.3.3.** Let $\{A_j\}_{j=1}^{K}$ be a partition of the unit intervals $[0, 1]$ with $K > p$. Suppose $A_j$ is a finite union of disjoint intervals in $[0, 1]$. Let $\phi_s(t) = I(t \in A_s)$, $s = 1, 2, \ldots, K$. If we set $R_{ij} = \int_0^1 \phi_i(t) \phi_j(t) dt$, then $L = \text{diag}(1/\sqrt{|A_1|}, 1/\sqrt{|A_2|}, \ldots, 1/\sqrt{|A_K|})$, where $|A|$ denotes the Lebesgue measure of the set $A$. Further assume $|A_1| = |A_2| = \cdots = |A_K| = 1/K$, then we have

$$\hat{D}_T = \frac{1}{T} \sum_{i=1}^{T} \sum_{j=1}^{T} \sum_{s=1}^{K} I(i/T \in A_s)I(j/T \in A_s) \hat{u}_i \hat{u}_j' = \frac{1}{T} \sum_{i=1}^{T} \sum_{j=1}^{T} I(i, j \in \text{the same group}) \hat{u}_i \hat{u}_j',$$

where $i$ is in group $s$ if and only if $i/T \in A_s$, $s = 1, 2, \ldots, K$. In this case, the general LRV estimator is the same as the CCE considered in Bester et al. (2011), where the idea is to utilize the group structure in the observations and construct a covariance estimator based on the parameter estimates in each group. Using similar arguments in Sun (2012), we can show that

$$\frac{1}{\sqrt{T}} \sum_{i=1}^{[Tr]} \hat{u}_i \Rightarrow \Lambda B_p(r),$$

78
where $\Lambda$ is an invertible matrix such that

$$\Lambda \Lambda' = R(\theta_0)(G_0'W_0G_0)^{-1}G_0'W_0\Omega W_0G_0(G_0'W_0G_0)^{-1}R'(\theta_0)$$

and $B_p(r)$ denotes a $p$-dimensional vector of independent Brownian bridges. It implies that

$$\frac{1}{\sqrt{T}} \sum_{i \in \text{sth group}} \hat{u}_i \sim d \int_{A_s} dB_p(r) = \frac{1}{\sqrt{K}} \Lambda (Z_s - \bar{Z}),$$

and

$$\hat{D}_T \sim d \frac{1}{K} \sum_{s=1}^{K} (Z_s - \bar{Z})(Z_s - \bar{Z})' \Lambda',$$

where $(Z'_1, Z'_2, \ldots, Z'_K)' \sim N(0, I_K \otimes I_p)$ and $\bar{Z} = \sum_{s=1}^{K} Z_s / K$. When $p = 1$, it is well known that

$$\sum_{s=1}^{K} (Z_s - \bar{Z})^2 \sim d K - 1,$$

which implies $\sqrt{F_T} \sim d \sqrt{\frac{K-1}{K} |t_{K-1}|}$ under the $H_0$. Note that $\sqrt{\frac{K-1}{K} F_T} \sim d \sqrt{\frac{K-1}{K} |t_{K-1}|}$ under the subsampling-based $t$-statistic in Ibragimov and Müller (2010) when we consider a location model and $r(\theta_0) = \theta_0 - \theta^*$ for a specific value $\theta^*$. When $p > 1$, we have $F_T \sim d \frac{K}{K-p} F_{p,K-p}$. It is worth noting that the choice of $R = (R_{ij})$ with $R_{ij} = \int_0^1 \hat{\phi}_i(t) \hat{\phi}_j(t) dt$ is also valid. In this case, the limiting distribution of $F_T$ would be a scaled $F$ distribution with $p$ numerator and $K - p + 1$ denominator degrees of freedom [see Proposition 4.4.1].

**Remark 4.3.1.** For the subsampling-based inference, Assumption 4.2.2 can be relaxed by the assumptions which guarantee the finite dimensional convergence of

$$\left( \frac{1}{\sqrt{|G_i|}} \sum_{i \in G_i} \hat{u}_i, \ldots, \frac{1}{\sqrt{|G_K|}} \sum_{i \in G_K} \hat{u}_i \right).$$

Here $G_i$ is the set index for the $i$th group and $| \cdot |$ denotes the cardinality. When heteroscedasticity is present across different groups, the $t$-statistic tends to be conservative (see Ibragimov and Müller (2010)).

### 4.4 First order fixed-smoothing asymptotics

In what follows, we consider the first order fixed-smoothing asymptotics of the test statistic $F_T$ based on the general LRV estimator under the null hypothesis and local alternatives. To emphasize the dependence on the smoothing parameter $K$, we shall use the notation $F_T(K)$ instead of $F_T$.

**Proposition 4.4.1.** Suppose $p \leq K < \infty$ and $b \in (0, 1]$ are both fixed. Let $R = (R_{ij})_{i,j=1}^{K}$ with
$R_{ij} = \int_0^1 \tilde{\phi}_i(t/b) \tilde{\phi}_j(t/b) dt$ in the general LRV estimator. Further assume that $\phi_j(t)$ is continuously differentiable almost everywhere for $j = 1, 2, \ldots, K$. Under Assumptions 4.2.1-4.2.4 and $H_0$, we have

$$F_T(K) \to^d Q_{p,K} := U_p' D_p^{-1} U_p/p,$$

(4.8)

where $D_p = \sum_{j=1}^K \lambda_j \eta_j \eta_j'$, $\{\eta_j\}_{j=1}^K$ and $U_p$ are independent and identically distributed (iid) as $N(0, I_p)$. In particular, if $\lambda_j = 1/K$ for $j = 1, 2, \ldots, K$, we get

$$F_T(K) \to^d \frac{K}{K - p + 1} F_{p,K-p+1},$$

(4.9)

Remark 4.4.1. When the weights $\lambda_j$’s are not equal and $p = 1$, $D_p$ is a weighted sum of independent $\chi^2_1$ random variables. The limiting null distribution $Q_{p,K}$ can be further approximated by a scaled $F$ distribution with the parameters chosen properly to match the first two moments (see Sun, 2010). Compared to Sun (2012), we do not make the assumption that $\int_0^1 \phi_i(t) dt = 0$ and we allow the basis functions to be non-orthonormal (see Example 4.3.2). It is also worth noting that the above results hold when $\phi_s(t) = I(t \in A_s)$ with $A_s$ being a finite union of disjoint intervals in $[0, 1]$.

Proposition 4.4.2. Consider the local alternatives $H'_a : r(\theta_0) = c/\sqrt{T}$ with $c \neq 0 \in \mathbb{R}^p$. Under the same assumptions in Proposition 4.4.1 with $\lambda_j = 1/K$, we have

$$F_T(K) \to^d \frac{K}{K - p + 1} F_{p,K-p+1,c'}(R(\theta_0) V_0 R(\theta_0)' - 1),$$

where $F_{a,b,\delta}$ denotes the noncentral $F$ distribution with degrees of freedom $a$ and $b$, and noncentral parameter $\delta$.

The Proposition shows that the test $F_T(K)$ has non-trivial power against the local alternatives of order $1/\sqrt{T}$ and it is seen to be consistent if $||c|| \to +\infty$ as $T \to +\infty$.

Proof of Proposition 4.1. Define $S_t(\hat{\theta}_T) = \frac{1}{T} \sum_{i=1}^T \hat{u}_i$. Using the continuous mapping theorem, we can show that

$$\sqrt{T} S_{[Tr]}(\hat{\theta}_T) = \frac{1}{\sqrt{T}} \sum_{i=1}^{[Tr]} \hat{u}_i \Rightarrow \Lambda B_p(r) := d \Lambda(W_p(r) - r W_p(1)),$$

where $\Lambda$ is invertible such that $\Lambda \Lambda' = R(\theta_0)(G_0' W_0 G_0)^{-1} G_0' W_0 \Omega W_0 G_0 (G_0' W_0 G_0)^{-1} R(\theta_0)'$ and $W_p(r)$ is a $p$-dimensional vector of independent Brownian motions. Using summation by parts,
we get
\[
V_s = \frac{1}{bT} \sum_{i=1}^{T-1} \left[ \phi_s(t/(bT)) - \phi_s((t+1)/(bT)) \right] \sqrt{T} S_i(\hat{\theta}_T) + \sqrt{T} \phi_s(1/b) S_T(\hat{\theta}_T),
\]
where the last term disappears by recalling the fact that \( G_T(\hat{\theta}_T)' W_T g_T(\hat{\theta}_T) = 0 \). By the continuous mapping theorem, we have
\[
\begin{pmatrix}
V_1 \\
\vdots \\
V_K \\
\sqrt{Tr(\hat{\theta}_T)}
\end{pmatrix}
\xrightarrow{d}
\begin{pmatrix}
-\frac{\Lambda}{b} \int_0^1 \phi_1'(r/b) B_p(r) dr \\
\vdots \\
-\frac{\Lambda}{b} \int_0^1 \phi_K'(r/b) B_p(r) dr \\
\Lambda W_p(1)
\end{pmatrix}
\xrightarrow{d}
\begin{pmatrix}
\Lambda \int_0^1 \phi_1(r/b) dW_p(r) \\
\vdots \\
\Lambda \int_0^1 \phi_K(r/b) dW_p(r) \\
\Lambda W_p(1)
\end{pmatrix}.
\]
Here we are using the fact that
\[
-\frac{\Lambda}{b} \int_0^1 \phi'_s(r/b) B_p(r) dr = \Lambda \int_0^1 \phi_s(r/b) dB_p(r) = \Lambda \int_0^1 \phi_s(r/b) - \int_0^1 \phi_s(r/b) dr dW_p(r)
\]
\[= \Lambda \int_0^1 \hat{\phi}_s(r/b) dW_p(r),
\]
for \( 1 \leq s \leq K \). It is not hard to see that
\[
\text{Cov} \left( \int_0^1 \hat{\phi}_s(r/b) dW_p(r), \int_0^1 dW_p(r) \right) = 0
\]
and
\[
\text{Cov} \left( \int_0^1 \hat{\phi}_s(r/b) dW_p(r), \int_0^1 \hat{\phi}_t(r/b) dW_p(r) \right) = R_{st} I_p,
\]
for \( 1 \leq s, t \leq K \), which implies
\[
V = (V'_1, V'_2, \ldots, V'_K, \sqrt{Tr(\hat{\theta}_T)})' \xrightarrow{d} N(0, \tilde{R} \otimes \Lambda \Lambda'), \text{ where } \tilde{R} = \begin{pmatrix} R & 0 \\ 0 & 1 \end{pmatrix}.
\]
We thus get \( V^* = (L \otimes I_p)V \xrightarrow{d} N(0, LRL' \otimes \Lambda \Lambda') = d N(0, I_K \otimes \Lambda \Lambda') \). In other words, \( V^* \) is free of the effect of the basis functions asymptotically. Recall that \( \tilde{D}_T = \sum_{s=1}^K \lambda_s V'_s V'_s \), it is not hard to see that
\[
F_T(K) = (\Lambda^{-1} \sqrt{Tr(\hat{\theta}_T)})' \{ \Lambda^{-1} \tilde{D}_T(\Lambda^{-1})' \}^{-1} (\Lambda^{-1} \sqrt{Tr(\hat{\theta}_T)})/p \xrightarrow{d} U'_p D_p^{-1} U'_p/p,
\]
where \( D_p = \sum_{j=1}^K \lambda_j \eta_j \eta'_j \) and \( \{ \eta_j \}_{j=1}^K \) and \( U_p \) are iid with distribution \( N(0, I_p) \). When \( \lambda_j = 1/K, j =...
it is straightforward to see that $F_T(K) \xrightarrow{d} \frac{1}{K-p+1} F_{p,K-p+1}$.

Proof of Proposition 4.2. Notice that $\sqrt{T} \theta_T \xrightarrow{d} N(c, \Lambda \Lambda')$ under the local alternatives. The result follows from the arguments in the proof of Proposition 4.1 and Theorem 5.2.2 in Anderson (2003).

4.5 Higher order expansions

This chapter is partially motivated by recent studies on the ERP for the Gaussian location model by Jansson (2004) and Sun et al. (2008), who showed that the ERP is of order $O(1/T)$ under the fixed-$b$ asymptotics, which is smaller than the ERP under the small-$b$ asymptotics. A natural question is to what extent the ERP result can be extended to the recently proposed fixed-smoothing based inference methods under the fixed-smoothing asymptotics. Following Jansson (2004) and Sun et al. (2008), we focus on the inference of the mean of a univariate stationary Gaussian time series or equivalently a Gaussian location model. We conjecture that the higher order terms in the asymptotic expansion under the Gaussian assumption will also show up in the general expansion without the Gaussian assumption.

4.5.1 Expansion for the finite sample distribution of subsampling-based $t$-statistic

We first investigate the Edgeworth expansion of the finite sample distribution of subsampling-based $t$-statistic (Ibragimov and Müller, 2010). Here we treat the subsampling-based $t$-statistic and other cases separately, because the $t$-statistic corresponds to a different choice of normalization factor (compare with the Wald statistic in Section 4.5.2). Given the observations $\{X_1, X_2, \ldots, X_T\}$ from a Gaussian stationary time series, we divide the sample into $K$ approximately equal sized groups of consecutive observations. The observation $X_i$ is in the $j$-th group if and only if $i \in \mathcal{M}_j = \{s \in \mathbb{Z} : (j-1)T/K < s \leq jT/K\}$, $j = 1, 2, \ldots, K$. Define the sample mean of the $k$-th group as

$$\hat{\mu}_k = \frac{1}{|\mathcal{M}_k|} \sum_{i \in \mathcal{M}_k} X_i, \quad k = 1, 2, \ldots, K,$$

where $|\cdot|$ denotes the cardinality of a finite set. Let $\hat{\theta} = (\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_K)'$, $\bar{\mu}_n = \frac{1}{K} \sum_{i=1}^K \hat{\mu}_i$ and $S_n^2 = \frac{1}{K-1} \sum_{i=1}^K (\hat{\mu}_i - \bar{\mu}_n)^2$. Then the subsampling-based $t$-statistic for testing the null hypothesis
\( H_0 : \mu = \mu_0 \) versus the alternative \( H_a : \mu \neq \mu_0 \), is given by
\[
T_K = \frac{\sqrt{K}(\bar{\mu}_n - \mu_0)}{S_n} = \frac{\sqrt{K}(\bar{\mu}_n - \mu_0)}{\left\{ \frac{1}{K-1} \sum_{i=1}^{K} (\hat{\mu}_i - \bar{\mu}_n)^2 \right\}^{1/2}}.
\]
(4.10)

Our goal here is to develop an Edgeworth expansion of \( P(|T_K| \leq x) \) when \( K \) is fixed and sample size \( T \to \infty \). It is not hard to see that the distribution of \( T_K \) is symmetric, so it is sufficient to consider \( P(|T_K| \leq x) \) since \( P(T_K \leq x) = \frac{1+P(|T_K| \leq x)}{2} \) for any \( x \geq 0 \). Denote by \( t_k \) a random variable following \( t \) distribution with \( k \) degrees of freedom. The following theorem gives the higher order expansion under the Gaussian assumption.

**Theorem 4.5.1.** Assume that \( \{X_i\} \) is a stationary Gaussian time series satisfying that \( \sum_{h=\infty}^{+\infty} \gamma_X(h) > 0 \) and \( \sum_{h=-\infty}^{+\infty} h^2 |\gamma_X(h)| < \infty \). Further suppose that \( |M_1| = |M_2| = \cdots = |M_K| \) and \( K \) is fixed. Then under \( H_0 \), we have
\[
\sup_{x \in (0, +\infty)} |P(|T_K| \leq x) - \Psi(x; K)| = O(1/T^2),
\]
(4.11)
where \( \Psi(x; K) = P(|t_{K-1}| \leq x) - \frac{B}{\sqrt{\pi T}} \Upsilon(x; K) \) with
\[
\Upsilon(x; K) = -K^2 P(|t_{K-1}| \leq x) + (K + 1)E \left[ \chi^2_{K-1} G_1 \left( \frac{\chi^2_{K-1} x^2}{K-1} \right) \right] - E \left[ \chi^2_{K-1} G_1 \left( \frac{(K-1)\chi^2_{K-1}}{x^2} \right) \right] + 1,
\]
and \( B = \sum_{h=-\infty}^{+\infty} |h| \gamma_X(h) \).

We present the proof of Theorem 4.5.1 in Section 4.9, which requires some nonstandard arguments. From the above expression, we see that the leading error term is of order \( O(1/T) \) and the magnitude and direction of the error depend upon \( B/\sigma^2 \), which is related to the second order properties of time series, and \( \Upsilon(x; K) \), which is independent of the dependence structure of \( \{X_i\} \) and can be approximated numerically for given \( x \) and \( K \). Figure 4.1 plots the approximated values of \( \Upsilon(t_{K-1}(1-\alpha); K)/K \) for different \( K \) and \( \alpha \), where \( t_{K-1}(1-\alpha) \) denotes the 100(1 - \( \alpha \))% quantile of the \( t \) distribution with \( K - 1 \) degrees of freedom. It can be seen from Figure 4.1 that \( \Upsilon(t_{K-1}(1-\alpha); K)/K \) increases rapidly for \( K < 10 \) and it becomes stable for relatively large \( K \). For each \( K \geq 2 \), \( \Upsilon(t_{K-1}(1-\alpha); K)/K \) is an increasing function of \( \alpha \). In the simulation work of Ibragimov and Müller (2010) (see Figure 2 therein), they found that the size of the subsampling-based \( t \)-test is relatively robust to the correlations if \( K \) is small (say \( K = 4 \) in their simulation). This finding is in
fact supported by our theory. For $K \leq 4$, the magnitude of $\Upsilon(x; K)$ is rather small, so the leading error term is small across a range of correlations. As $K$ increases, the first order approximation deteriorates, which is reflected in the increasing magnitude of $\Upsilon(t_{K-1}(1-\alpha); K)$ with respect to $K$.

Notice that $\Upsilon(t_{K-1}(1-\alpha); K)$ is always positive and $\sigma^2 > 0$ by assumption, so the sign of the leading error term, i.e., $-\frac{B}{2\sigma^2 T} \Upsilon(x; K)$ is determined by $B$. When $B > 0$ (e.g., AR(1) process with positive coefficient), the first order based inference tends to be oversized and conversely it tends to be undersized when $B < 0$ (e.g., MA(1) process with negative coefficient). Some simulations for AR(1) and MA(1) models in the Gaussian location model support these theoretical findings. We decide not to report these results to conserve space. Given the sample size $T$, the size distortion for the first order based inference may be severe if the ratio $B/\sigma^2$ is large. For example, this is the case for AR(1) model, $X_t = \rho X_{t-1} + \varepsilon_t$, as the correlation $\rho$ gets closer to 1. As indicated by Figure 4.1, we show in the following proposition that $\Upsilon(t_{K-1}(1-\alpha); K)/K$ converges as $K \to +\infty$.

**Proposition 4.5.2.** As $K \to +\infty$, we have $\Upsilon(x; K)/K = 2x^2G'(x^2) + O(1/K)$, for any fixed $x \in \mathbb{R}$.

![Figure 4.1: Simulated values of $\Upsilon(t_{K-1}(1-\alpha); K)/K$ based on 500,000 replications.](image)

Under the local alternative $H'_a: \mu = \mu_0 + (\delta \sigma)/\sqrt{T}$ with $\delta \neq 0$, we can derive a similar expansion for $T_K$ with $K$ fixed. Formally let $Z$ be a random variable following the standard normal distribution and $S_{K-1} = \sqrt{\chi^2_{K-1}/(K-1)}$ with the $\chi^2_{K-1}$ distribution being independent with $Z$. Then the quantity $t_{K-1, \delta} = (Z + \delta)/S_{K-1}$ follows a noncentral $t$ distribution with noncentral parameter $\delta$. Define $e_1(x) = E[I\{|t_{K-1, \delta}| > x\} Z^2]$ and $e_2(x) = E[I\{|t_{K-1, \delta}| > x\} \chi^2_{K-1}]$. Then under the local
alternative, we have

$$P(|T_K| \leq x) = P(|t_{K-1,\delta}| \leq x) - \frac{B}{2\sigma^2 T} \Upsilon_\delta(x; K) + O(1/T^2),$$

where $\Upsilon_\delta(x; K) = K^2 P(|t_{K-1,\delta}| > x) - e_1(x) - (K+1)e_2(x)$. For fixed $\delta$, $P(|t_{K-1,\delta}| > t_{K-1}(1-\alpha))$ is a monotonic increasing function of $K$. Unreported numerical study shows that $\Upsilon_\delta(t_{K-1}(1-\alpha); K)$ is roughly monotonic with respect to $K$ for $\delta \in (0, 4]$, which suggests that larger $K$ tends to deliver more power when $B > 0$. Combined with the previous discussion, we see that the choice of $K$ leads to a trade-off between the size distortion and power loss.

**Remark 4.5.1.** Theorem 4.5.1 gives the ERP and the exact form of the leading error term under the fixed-$K$ asymptotics. The higher order expansion derived here is based on an expansion of the density function of $(\hat{\mu}_1, \ldots, \hat{\mu}_K)$ which is made possible by the Gaussian assumption. Extension to the general GMM setting without the Gaussian assumption may require a different strategy in the proof. Expansion for a distribution function or equivalently characteristic function has been used in the higher order expansion of the finite sample distribution under the Gaussian assumption (see e.g., Velasco and Robinson, 2001; Sun et al., 2008). With $K$ fixed in the asymptotics, the leading term of the variance of the LRV estimator is captured by the first order fixed-$K$ limiting distribution and the leading term of the bias of the LRV estimator is reflected in the leading error term $-\frac{B}{2\sigma^2 T} \Upsilon(x, K)$. Specifically, let $\Sigma_T = (\sigma_{ij})_{i,j=1}^K$ with $\sigma_{ij} = q \text{Cov}(\hat{\mu}_i, \hat{\mu}_j)$. Then the leading error term captures the difference between $\Sigma_T$ and $\sigma^2 I_K$ and the effect of the off-diagonal elements $\sigma_{ij}$ with $|i - j| > 1$ is of order $O(1/T^2)$ and thus is not reflected in the leading term.

**Remark 4.5.2.** When the number of groups $K$ grows slowly with the sample size $T$, the Edgeworth expansion for $T_K$ was developed for $P(T_K \leq x)$ in Lahiri (2007, 2010) under the general non-Gaussian setup. The expansion given here is different from the usual Edgeworth expansion under the increasing-domain asymptotics in terms of the form and the convergence rate. Using the same argument, we can show that under the fixed-$K$ asymptotics, the leading error term in the expansion of $P(T_K \leq x)$ is of order $O(1/T)$ under the Gaussian assumption. In the non-Gaussian case, we conjecture that the order of the leading error term is $O(1/\sqrt{T})$, which is due to the effect of the third and fourth order cumulants.

The higher order Edgeworth expansion results in Sun et al. (2008) suggest that the fixed-$b$ based approximation is a refinement of the approximation provided by the limiting distribution derived under the small-$b$ asymptotics. In a similar spirit, it is natural to ask if the fixed-$K$ based
approximation refines the first order approximation under the increasing-\(K\) asymptotics. To address this question, we consider the expansion under the increasing-domain asymptotics, where \(K\) grows slowly with the sample size \(T\).

**Proposition 4.5.3.** Under the same conditions in Theorem 4.5.1 but with \(\lim_{T \to \infty} (1/K + K/T) = 0\), we have

\[
P(|T_K| \leq x) = G_1(x^2) + \frac{1}{K-1} x^4 G'_1(x^2) - \frac{BK}{T \sigma^2} x^2 G'_1(x^2) + O(1/T).
\]

(4.12)

**Remark 4.5.3.** Since

\[
P(|t_{K-1}| \leq x) = G_1(x^2) + \frac{1}{K-1} x^4 G'_1(x^2) + O(1/K^2)
\]

(see e.g., Sun, 2012), we know that the fixed-\(K\) based approximation captures the first two terms in (4.12), whereas the increasing-\(K\) based approximation (i.e., \(\chi^2\)) only captures the first term. In view of Proposition 4.5.2, it is not hard to see that

\[
\Psi(x; K) = G_1(x^2) + \frac{1}{K-1} x^4 G'_1(x^2) - \frac{BK}{T \sigma^2} x^2 G'_1(x^2) + O(1/K^2) + O(1/T),
\]

which implies that the fixed-\(K\) based expansion is able to capture all the three terms in (4.12) as the smoothing parameter \(K \to \infty\) with \(T^{1/3} = o(K)\). Loosely speaking, this suggests that the fixed-\(K\) based expansion holds for a broad range of \(K\) and it gets close to the corresponding increasing-\(K\) based expansion when \(K\) is large.

### 4.5.2 Fixed-\(b\) expansion

Consider a semi-positive definite bivariate kernel \(G(\cdot, \cdot)\) which satisfies the spectral decomposition

\[
G(r, t) = \sum_{j=1}^{+\infty} \lambda_j \phi_j(r) \phi_j(t), \quad 0 \leq r, t \leq 1,
\]

(4.13)

where \(\{\phi_j\}\) are the eigenfunctions and \(\{\lambda_j\}\) are the eigenvalues which are in a descending order, i.e., \(\lambda_1 \geq \lambda_2 \geq \cdots \geq 0\). Suppose we have the observations \(\{X_1, X_2, \ldots, X_T\}\) from a stationary Gaussian time series with mean \(\mu\) and autocovariance function \(\gamma_X(i - j) = E[(X_i - \mu)(X_j - \mu)]\). The LRV estimator based on the kernel \(G(\cdot, \cdot)\) and bandwidth \(S_T = bT\) with \(b \in (0, 1]\) is given by

\[
\hat{D}_{T,b} = \frac{1}{T} \sum_{i=1}^{T} \sum_{j=1}^{T} G \left( \frac{i}{bT}, \frac{j}{bT} \right) (X_i - \bar{X}_T)(X_j - \bar{X}_T).
\]
where $\bar{X}_T = \sum_{i=1}^{T} X_i / T$ is the sample mean. For the convenience of presentation, we set $b = 1$. See Remark 4.5.4 for the case $b \in (0, 1)$. To illustrate the idea, we define the projection vectors $\xi_j = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} \phi_j^0(i/T) X_i$ with $\phi_j^0(t) = \phi_j(t) - \frac{1}{T} \sum_{i=1}^{T} \phi_j(i/T)$ for $j = 1, 2, \ldots$. Here the dependence of $\xi_j$ on $T$ is suppressed to simplify the notation. Following Sun (2012), we limit our attention to the case $\int_0^1 \phi_j(t) dt = 0$ (e.g., Fourier basis and Haar wavelet basis). For any semi-positive definite kernel $\tilde{G}(\cdot, \cdot)$, we can define the demeaned kernel,

$$\tilde{G}(r, t) = \tilde{G}(t) - \int_0^1 \tilde{G}(s, t) ds - \int_0^1 \tilde{G}(r, p) dp + \int_0^1 \int_0^1 \tilde{G}(s, p) ds dp.$$

Suppose $\tilde{G}(\cdot, \cdot)$ admits the spectral decomposition $\tilde{G}(r, t) = \sum_{i=1}^{\infty} \tilde{\lambda}_i \tilde{\phi}_i(r) \tilde{\phi}_i(t)$ with $\{\tilde{\phi}_i\}$ and $\{\tilde{\lambda}_i\}$ being the eigenfunctions and eigenvalues respectively. Notice that

$$\int_0^1 \int_0^1 \tilde{G}(r, t) dr dt = \sum_{i=1}^{\infty} \tilde{\lambda}_i \left( \int_0^1 \tilde{\phi}_i(t) dt \right)^2 = 0,$$

which implies $\int_0^1 \tilde{\phi}_i(t) dt = 0$ whenever $\lambda_i > 0$, i.e., the eigenfunctions of the demeaned kernel $\tilde{G}(\cdot, \cdot)$ are all mean zero. Based on the spectral decomposition (4.13) of $G(\cdot, \cdot)$, the LRV estimator with $b = 1$ can be rewritten as

$$\hat{D}_{T,1} = \frac{1}{T} \sum_{i=1}^{T} \sum_{j=1}^{T} G \left( \frac{i}{T}, \frac{j}{T} \right) (X_i - \bar{X}_T)(X_j - \bar{X}_T) = \sum_{i=1}^{\infty} \lambda_i \xi_i^2.$$

We focus on testing the null hypothesis $H_0 : \mu = \mu_0$ versus the alternative $H_a : \mu \neq \mu_0$. Define a sequence of random variables

$$F_T(K) = \frac{\xi_0^2}{\sum_{j=1}^{K} \lambda_j \xi_j^2}, \quad K = 1, \ldots, \infty$$

with $\xi_0 = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} (X_i - \mu_0)$. The Wald test statistic with HAC estimator is given by $F_T(\infty) = \xi_0^2 / \hat{D}_{T,1}$. Let $\{v_i\}_{i=0}^{+\infty}$ be a sequence of independent and identically distributed (i.i.d.) standard normal random variables. Further define $F(K) := F(v; K) = \frac{\xi_0^2}{\sum_{j=1}^{K} \lambda_j \xi_j^2}$ and

$$\mathbb{E}_T(x; K) = \frac{1}{2\sigma^2} \sum_{i=0}^{K} (\text{var}(\xi_i) - \sigma^2) E[(v_i^2 - 1)1\{F(v; K) \leq x\}], \quad K = 1, \ldots, \infty$$

with $\sigma^2 = \sum_{h=-\infty}^{+\infty} \gamma_X(h)$ being the LRV. The following theorem establishes the asymptotic expansion of the finite sample distribution of $F_T(K)$ with $1 \leq K \leq \infty$. 87
The second derivatives of the eigenfunctions \( \{\phi_i^{(2)}(\cdot)\}_{i=1}^{+\infty} \) exist. Further assume that the eigenfunctions are mean zero and satisfy that

\[
\sup_{1 \leq i \leq J} |\phi_i^{(j)}(t)| < C J^j \text{ for } j = 0, 1, 2, J \in \mathbb{N}, \text{ and some constant } C \text{ which does not depend on } j \text{ and } J;
\]

(2) The eigenvalues \( \lambda_n = O(1/n^a) \), for some \( a > 19 \).

Under the assumption that \( \{X_t\} \) is a stationary Gaussian time series with \( \sigma^2 = \sum_{h=-\infty}^{+\infty} \gamma_X(h) > 0 \) and \( \sum_{h=-\infty}^{+\infty} h^2 |\gamma_X(h)| < \infty \), and the null hypothesis \( H_0 \), we have \( \sup_{x \in [0, +\infty)} |\varphi_T(x; K)| = O(1/T) \) and

\[
\sup_{x \in [0, +\infty)} |P(F_T(K) \leq x) - P(F(K) \leq x) - \mathcal{N}_T(x; K)| = o(1/T), \tag{4.15}
\]

for any \( 1 \leq K \leq +\infty \).

The proof of Theorem 4.5.4 is based on the arguments of the proof of Theorem 4.5.1 given in Section 4.9 and the truncation argument. The technical details are provided in Section 4.10. For \( K < \infty \), Theorem 4.5.4 shows that the \( O(1/T) \) ERP rate can be extended to the Wald statistic with series variance estimator (Sun, 2012). When \( K = \infty \), Theorem 4.5.4 gives the asymptotic expansion of the Wald test statistic \( F_T(\infty) \) which is of particular interest. The leading error term \( \mathcal{N}_T(x; \infty) \) reflects the departure of \( \{\xi_j\}_{j=1}^{+\infty} \) from the i.i.d. standard normal random variables \( \{\eta_j\}_{j=1}^{+\infty} \). Specifically, the form of \( \mathcal{N}_T(x; \infty) \) suggests that the leading error term captures the difference between the LRV and the variances of \( \xi_j \)’s which are not exactly the same across \( i = 0, 1, 2, \ldots \). By the orthogonality assumption, the covariance between \( \xi_i \) and \( \xi_j \) with \( i \neq j \) is of smaller order and hence is not reflected in the leading term. Assume \( \int_0^1 G(r, r) dr = \sum_{j=1}^{+\infty} \lambda_j = 1 \). As seen from Theorem 4.5.4, the bias of the LRV estimator (i.e., \( \sum_{i=1}^{+\infty} \lambda_i (\text{var}(\xi_i) - \sigma^2)) \) is reflected in the leading error term \( \mathcal{N}_T(x; \infty) \), which is a weighted sum of the relative difference of \( \text{var}(\xi_i) \) and \( \sigma^2 \). Note that the difference \( \text{var}(\xi_i) - \sigma^2 \) relies on the second order properties of the time series and the eigenfunctions of \( G(\cdot, \cdot) \), and the weight \( E[(\eta_i^2 - 1)I(\mathcal{F}(\infty) \leq x)] \) which depends on the eigenvalues of \( G(\cdot, \cdot) \) is of order \( O(\lambda_i) \), as seen from the arguments used in the proof of Theorem 4.5.4.

In the econometrics and statistics literatures, the bivariate kernel \( G(\cdot, \cdot) \) is usually defined through a semi-positive definite univariate kernel \( K(\cdot) \) i.e., \( G(r, t) = K(r-t) \). In what follows, we make several remarks regarding this special case.

**Remark 4.5.4.** For \( 0 < b \leq 1 \), we define \( G_b(\cdot, \cdot) = G(\cdot/b, \cdot/b) \). If \( G(\cdot, \cdot) \) is semi-positive definite on \([0, 1/b]^2\), then \( G_b(\cdot, \cdot) \) satisfies the spectral decomposition \( G_b(r, t) = \sum_{j=1}^{+\infty} \lambda_j \phi_{j,b}(r) \phi_{j,b}(t) \) with
0 \leq r, t \leq 1. The eigencomponets of $G_b(r, t)$ can be obtained by solving a homogenous Fredholm integral equation of the second kind, where the solutions can be approximated numerically when analytical solutions are unavailable. When $G(r, t) = K(r-t)$, it was shown in Knessl and Keller (1991) that under suitable assumptions on $K(\cdot)$, $\lambda_{j,b} = b \int_{-\infty}^{+\infty} K(r)dr - (\pi^2 j^2 b^2 / 2) \int_{-\infty}^{+\infty} r^2 K(r)dr + o(b^3)$ and $\phi_{j,b} \approx \sqrt{2} \sin(\pi j x)$ for $x$ bounded away from 0 and 1 as $b \to 0$, which implies that $\lambda_{M,b}/\lambda_{1,b} \to 1$ for any fixed $M \in \mathbb{N}$ and $b \to 0$. Our result can be extended to the case where $b < 1$ if the assumptions in Theorem 4.5.4 hold for $\{\lambda_{j,b}\}$ and $\{\phi_{j,b}\}$. It is also worth noting that our result is established under different assumptions as compared to Theorem 6 in Sun et al. (2008), where the bivariate kernel is defined as $G(r, t) = K(r-t)$ and the technical assumption $b < 1/(16 \int_{-\infty}^{+\infty} |K(r)|dr)$ is required, which rules out the case $b = 1$ for most kernels. Here we provide an alternative way of proving the $O(1/T)$ ERP when the eigenfunctions are mean zero. Furthermore, we provide the exact form of the leading error term which has not been obtained in the literature.

Remark 4.5.5. The assumption on the eigenvalues is satisfied by the bivariate kernel defined through the QS kernel and the Daniel kernel with $0 < b \leq 1$, and the Tukey-Hanning kernel with $b = 1$ because these kernels are analytical on the corresponding regions and their eigenvalues decay exponentially fast (see Little and Reade, 1984). However, the assumption does not hold for the Bartlett kernel because the decay rate of its eigenvalues is of order $O(1/n^2)$. For the demeaned Tukey-Hanning kernel with $b = 1$, we have that the eigenfunctions $\phi_1(t) = \sqrt{\tau} \cos \pi t$ and $\phi_2(t) = \frac{\sin \pi t - 2/\pi}{\sqrt{1/2 - 4/\pi^2}}$ with eigenvalues $\lambda_1 = 0.25$, $\lambda_2 = 0.0474$, and $\lambda_j = 0$ for $j \geq 3$. It is not hard to construct a kernel that satisfies the conditions in Theorem 4.5.4. For example, one can consider the kernel $K(r-t) = \sum_{j=1}^{+\infty} \lambda_j \{\cos(2\pi jr) \cos(2\pi j t) + \sin(2\pi jr) \sin(2\pi j t)\} = \sum_{j=1}^{+\infty} \lambda_j \cos(2\pi j(x-t))$ with $\sum_{j=1}^{+\infty} \lambda_j = 1$ and $\lambda_j = O(1/j^{1+\epsilon})$ for some $\epsilon > 0$. Then the asymptotic expansion (4.15) holds for the Wald statistic based on the difference kernel $G(r, t) = K(r-t)$.

Define the Parzen characteristic exponent

$$q = \max \left\{q_0 : q_0 \in \mathbb{Z}^+, g_{q_0} = \lim_{x \to 0} \frac{1 - K(x)}{|x|^{q_0}} < \infty \right\}.$$ 

For the Bartlett kernel $q$ is 1; For the Parzen and QS kernels, $q$ is equal to 2. Let $c_1 = \int_{-\infty}^{+\infty} K(x)dx$ and $c_2 = \int_{-\infty}^{+\infty} K^2(x)dx$. Further define $\mathcal{F}_b(\infty)$ and $\mathcal{F}_{T,b}(x; \infty)$ with $\phi_j$ and $\lambda_j$ being replaced with $\phi_{j,b}$ and $\lambda_{j,b}$ in the definition of $\mathcal{F}(\infty)$ and $\mathcal{F}(x; \infty)$. We summarize the first and second order approximations for the distribution of studentized sample mean in the Gaussian location model based on both fixed-$b$ and small-$b$ asymptotics in Table 4.1 above. The formulae for the second
order approximation under the small-\(b\) asymptotics is from Velasco and Robinson (2011).

**Remark 4.5.6.** A few remarks are in order regarding Table 4.1. First of all, it is worth noting that \(P (\mathcal{F}_b(\infty) \leq x) = G_1(x) + (c_2 G_1''(x)x^2 - c_1 G_1'(x)x)b + O(b^2)\) as \(b \to 0\) in Sun et al. (2008), which suggests that the fixed-\(b\) limiting distribution captures the first two terms in the higher order asymptotic expansion under the small-\(b\) asymptotics and thus provides a better approximation than the \(\chi^2\) approximation. Secondly, it is interesting to compare the second order asymptotic expansions under the fixed-\(b\) asymptotics and small-\(b\) asymptotics. We show in Proposition 4.5.5 that the higher order expansion under fixed-\(b\) asymptotics is consistent with the corresponding higher order expansion under small-\(b\) asymptotics as \(b\) approaches zero.

Because our fixed-\(b\) expansion is established under the assumption that the eigenfunctions have mean zero, we shall consider the Wald statistic \(F_T(\infty)\) based on the demeaned kernel \(\tilde{G}_b(r, t) = K_b(r - t) - \int_0^1 K_b(s - t)ds - \int_0^1 K_b(r - p)dp + \int_0^1 \int_0^1 K_b(s - p)dsdp\) with \(K_b(\cdot) = K(\cdot/b)\) and \(b \in (0, 1]\). Let \{\(\tilde{\phi}_{j,b}\)\} and \{\(\tilde{\lambda}_{j,b}\)\} be the corresponding eigenfunctions and eigenvalues of \(\tilde{G}_b(\cdot, \cdot)\).

**Proposition 4.5.5.** Suppose \(K(\cdot) : \mathbb{R} \to [0, 1]\) is symmetric, semi-positive definite, piecewise smooth with \(K(0) = 1\) and \(\int_0^{+\infty} xK(x)dx < \infty\). The Parzen characteristic exponent of \(K\) is no less than one. Further assume that

\[
\sup_{k \in \mathbb{N}} \left| \sum_{i=1}^k \tilde{\lambda}_{i,b}(\text{var}(\tilde{\xi}_{i,b}) - \sigma^2) \right| = O \left( \sum_{i=1}^\infty \tilde{\lambda}_{i,b}(\text{var}(\tilde{\xi}_{i,b}) - \sigma^2) \right), \tag{4.16}
\]

as \(b + 1/(bT) \to 0\), where \(\tilde{\xi}_{i,b}\) is defined by replacing \(\phi_j\) with \(\tilde{\phi}_{j,b}\) in the definition of \(\xi_i\). Then under the assumption that \(\sigma^2 > 0\) and \(\sum_{h=-\infty}^{+\infty} h^2 |\gamma_X(h)| < \infty\), we have

\[
\mathcal{R}_T,b(x; \infty) = -\frac{g_q \sum_{h=-\infty}^{+\infty} |h|^{q} \gamma_X(h)}{\sigma^2 (bT)^q} G_1'(x)x(1 + o(1)) + O(1/T),
\]

for fixed \(x \in \mathbb{R}\), as \(b \to 0\) and \(bT \to +\infty\).

In Proposition 4.5.5, the condition (4.16) is not primitive and it requires that the bias for the LRV estimators based on the kernel \(\tilde{G}_{k,b}(r, t) = \sum_{i=1}^k \tilde{\lambda}_{i,b} \tilde{\phi}_{i,b}(r) \tilde{\phi}_{j,b}(t)\) is at the same or smaller
order of the bias for the LRV estimator based on $\hat{G}_b(r, t)$. This condition simplifies our technical arguments and it can be verified through a case-by-case study. As shown in proposition 4.5.5, the fixed-$b$ expansion is consistent with the small-$b$ expansion as $b$ approaches zero and it is expected to be more accurate in terms of approximating the finite sample distribution when $b$ is relatively large. Overall speaking, the above result suggests that the fixed-$b$ expansion provides a good approximation to the finite sample distribution which holds for a broad range of $b$.

4.6 Gaussian Dependent Bootstrap

Given the higher order expansions presented in Section 4.5, it seems natural to investigate if bootstrapping can help to improve the first order approximation. Though the higher order corrected critical values can also be obtained by direct estimation of the leading error term, it involves estimation of the eigencomponents of the kernel function and a choice of truncation number for the leading error term $\aleph_T(x; \infty)$ (see 4.14) besides estimating the second order properties of the time series. Therefore it is rather inconvenient to implement this analytical approach because numerical or analytical calculation of eigencomponents can be quite involved, the truncation number and the bandwidth parameter used in estimating second order properties are both user-chosen numbers and it seems difficult to come up with good rules about their (optimal) choice in the current context. By contrast, the bootstrap procedure proposed below, which involves only one user-chosen number, aims to estimate the leading error term in an automatic fashion and the computational cost is moderate given current high computing power.

To present the idea, we again limit our attention to the univariate Gaussian location model. Consider a consistent estimate of the covariance matrix of $\{X_i\}_{i=1}^T$ which takes the form $\hat{\Xi}(\omega; l) \in \mathbb{R}^{T \times T}$ with the $(i, j)$th element given by $\omega_l(i-j)\hat{\gamma}_X(|i-j|)$ for $i, j = 1, 2, \ldots, T$, where $\omega$ is a kernel function with $\omega_l(\cdot) = \omega(\cdot/l)$ and $\hat{\gamma}_X(h) = \frac{1}{T}\sum_{i=1}^{T-h}(X_i - \bar{X}_T)(X_i+h - \bar{X}_T)$ for $h = 0, 1, 2, \ldots, T-1$. Estimating the covariance matrix of a stationary time series has been investigated by a few researchers. See Wu and Pourahmadi (2009) for the use of a banded sample covariance matrix and McMurry and Politis (2010) for a tapered version of the sample covariance matrix. In what follows, we shall consider the Bartlett kernel, i.e., $\omega(x) = (1 - |x|)I\{|x| < 1\}$, which guarantees to yield a semi-positive definite estimates, i.e., $\hat{\Xi}(\omega; l) \succeq 0$.

We now introduce a simple bootstrap procedure which can be shown to be second order correct. Suppose $X_1^*, \ldots, X_T^*$ is the bootstrap sample generated from $N(0, \hat{\Xi}(\omega; l))$. It is easy to see that
X_i’s are stationary and Gaussian conditional on the data. This is why we name this bootstrap method “Gaussian Dependent Bootstrap”. There is a large literature on bootstrap for time series; see Lahiri (2003) for a review. However, most of the existing bootstrap methods do not deliver a conditionally normally distributed bootstrap sample. Since our higher order results are obtained under the Gaussian assumption, we need to generate Gaussian bootstrap sample in order for our expansion results to be useful.

Denote by $T^*_K$ the bootstrapped subsampling $t$-statistic obtained by replacing $(X_1 - \mu_0, X_2 - \mu_0, \ldots, X_T - \mu_0)$ with $(X_1^*, X_2^*, \ldots, X_T^*)$. Define the bootstrapped projection vectors $\xi_0^* = \frac{1}{\sqrt{T}} \sum_{j=1}^{T} X_j^*$ and $\xi_j^* = \frac{1}{\sqrt{T}} \sum_{i=1}^{T} \phi_j^0(i/T) X_i$ for $j = 1, \ldots, L$. Let $P^*$ be the bootstrap probability measure conditional on the data. The following theorems state the second order accuracy of the Gaussian dependent bootstrap in the univariate Gaussian location model.

**Theorem 4.6.1.** For the Gaussian location model, under the same conditions in Theorem 4.5.1 and $1/l + l^3/T \to 0$, we have

$$\sup_{x \in [0, +\infty)} |P(|T_K| \leq x) - P^*(|T_K^*| \leq x)| = o_p(1/T). \quad (4.17)$$

**Theorem 4.6.2.** For the Gaussian location model, under the assumptions in Theorem 4.5.4 and that $1/l + l^3/T \to 0$, we have

$$\sup_{x \in [0, +\infty)} |P(F_T(\infty) \leq x) - P^*(F_T^*(\infty) \leq x)| = o_p(1/T), \quad (4.18)$$

where $F_T^*(\infty) = \frac{(\xi_0^*)^2}{\sum_{j=1}^{L} \lambda_j (\xi_j^*)^2}$ with $\{\lambda_j\}_{j=1}^{L}$ given in (4.13). Note that $F_T(\infty) = (\xi_0^*)^2/\hat{D}_{T,1}$, where $\hat{D}_{T,1} = T^{-1} \sum_{i,j=1}^{T} \mathcal{G}(i/T, j/T)(X_i - \bar{X}_T^*)(X_j - \bar{X}_T^*)$ and $\bar{X}_T^*$ is the bootstrap sample mean.

**Remark 4.6.1.** The higher order terms in the small-$b$ expansion and the increasing-$K$ expansion (see Table 4.1 and Proposition 4.5.3) depend on the second order properties only through the quantities $\sum_{h=-\infty}^{+\infty} |h|^k \gamma_X(h)$ for $k = 0, 1, \ldots, q$. It suggests that the Gaussian dependent bootstrap also preserves the second order accuracy under the increasing-domain asymptotics provided that $\sum_{h=-\infty}^{+\infty} |h|^{q+1} \gamma_X(h) < \infty$. A rigorous proof is omitted due to space limitation.

The bootstrap-based autocorrelation robust testing procedures have been well studied in both econometrics and statistics literatures under the increasing-domain asymptotics. In the statistical literature, Lahiri (1996) showed that for the studentized $M$-estimator, the ERP of the moving block bootstrap (MBB)-based one-sided testing procedure is of order $o_p(T^{-1/2})$ which provides an
asymptotic refinement to the normal approximation. Under the framework of the smooth function model, Götze and Künsch (1996) showed that the ERP for the MBB-based one-sided test is of order $O_p(T^{-3/4+\epsilon})$ for any $\epsilon > 0$ when the HAC estimator is constructed using the truncated kernel. Note that in the latter paper, the HAC estimator used in the studentized bootstrap statistic needs to take a different form from the original HAC estimator to achieve the higher order accuracy. Also see Lahiri (2007) for a recent contribution. In the econometric literature, the Edgeworth analysis for the block bootstrap has been conducted by Hall and Horowitz (1996), Andrews (2002) and Inoue and Shintani (2006), among others, in the GMM framework. Within the increasing-domain asymptotic framework, it is still unknown whether the bootstrap can achieve an ERP of $o_p(1/T)$ when a HAC covariance matrix estimator is used for studentization (see Härdle, Horowitz and Kreiss, 2003). Note that Hall and Horowitz (1996) and Andrews (2002) obtained the $o_p(1/T)$ results for symmetrical tests but they assumed the uncorrelatedness of the moment conditions after finite lags. Note that all the above results were obtained under the non-Gaussian assumption.

Within the fixed-smoothing asymptotic framework, Jansson (2004) established that the error of the fixed-$b$ approximation to the distribution of two-sided test statistic is of order $O(\log(T)/T)$ for the Gaussian location model and the case $b = 1$, which was further refined by Sun et al. (2008) by dropping the $\log(T)$ term. In the non-Gaussian setting, Gonçalves and Vogelsang (2010) showed that the fixed-$b$ approximation to the distribution of one-sided test statistic has an ERP of order $o(T^{-1/2+\epsilon})$ for any $\epsilon > 0$ when all moments exist. The latter authors further showed that the MBB (with iid bootstrap as a special case) is able to replicate the fixed-$b$ limiting distribution and thus provides more accurate approximation than the normal approximation. However, because the exact form of the leading error term was not obtained in their studies, their results seem not directly applicable to show the higher order accuracy of bootstrap under the fixed-$b$ asymptotics. Using the asymptotic expansion results developed in Section 4.5, we show that the Gaussian dependent bootstrap can achieve an ERP of order $o_p(1/T)$ under the Gaussian assumption. This appears to be the first result that shows the higher order accuracy of bootstrap under the fixed-smoothing asymptotics. Our result also provides a positive answer to the open question mentioned in Härdle, Horowitz and Kreiss (2003) that whether the bootstrap can achieve an ERP of $o_p(1/T)$ in the dependence case when a HAC covariance matrix estimator is used for studentization. It is worth noting that our result is established for the symmetrical distribution functions under the fixed-smoothing asymptotics and the Gaussian assumption. It seems that in general the ERP of order $o_p(1/T)$ cannot be achieved under the increasing-domain asymptotics or for the non-Gaussian case. In the
supplementary material, we provide some simulation results which demonstrate the effectiveness of the proposed Gaussian dependent bootstrap in both Gaussian and non-Gaussian settings. The MBB is expected to be second order accurate, as seen from its empirical performance, but a rigorous theoretical justification seems very difficult. Finally, we mention that it is an important problem to choose \( l \). For a given criterion, the optimal \( l \) presumably depends on the second order property of the time series in a sophisticated fashion. Some of the rules proposed for block-based bootstrap (see Lahiri, 2003 chapter 7) may still work, but a serious investigation is beyond the scope of this article.

4.7 Simulation study

We conduct a small simulation study to compare and contrast the finite sample performance of the small-\( b \) approximation, fixed-\( b \) approximation, MBB and Gaussian dependent bootstrap (GDB). Following the setup in Gonçalves and Vogelsang (2011), we consider the AR(1) model,

\[
y_t = \rho y_{t-1} + \sqrt{1-\rho^2} \varepsilon_t, \quad t = 1, 2, \ldots, T,
\]

with \( \{\varepsilon_t\} \) being a sequence of iid \( N(0,1) \), \( t(3) \) or \( \exp(1) - 1 \) random variables. Consider the Wald statistic based on the HAC estimator with the Bartlett kernel and QS kernel for testing the null hypothesis \( E[y_t] = 0 \) versus the alternative that \( E[y_t] \neq 0 \) at 5% nominal level. Throughout the simulation we set \( T = 50 \) and the number of Monte Carlo replications to be 1000. The bootstrap tests are based on 1000 replications for each sample. We implement the MBB in a ‘naive’ fashion as described in Gonçalves and Vogelsang (2011). The simulation results for \( b = 0.04, 0.06, 0.08, 0.1, 0.2, \ldots, 1 \) and \( \rho = -0.7, 0, 0.5, 0.9 \) are summarized in Figures 4.2-4.4. We present the results for GDB with \( l = 5, 10 \) and MBB with block size equal to 5 and 10. It is seen from the figures that the GDB is more accurate than the small-\( b \) asymptotic approximation in most cases and improvement is often substantial especially for large \( b \). In the dependent cases (e.g., \( \rho = -0.7, 0.5 \) and 0.9), the GDB tends to provide a refinement over the fixed-\( b \) approximation for a proper bandwidth which is consistent with our theoretical findings. The improvement is apparent when the dependence is strong and \( b \) is small. In addition it is interesting to note that the GDB not only provides an improvement when the innovations are Gaussian but also in the case of \( t(3) \) distributed fat tailed innovations and \( \exp(1) - 1 \) distributed skewed innovation. The performance of GDB and MBB is in general comparable. MBB delivers slightly better size in most cases when the dependence is positive. When \( \rho = 0.9 \), the MBB with block size 10 apparently outperforms all the other methods for all three cases, suggesting that
with a proper choice of block size, the MBB is capturing not only the asymptotic bias and variance of long run variance estimator but also the higher order moments. Since the GDB only captures the second order properties, it is not surprising that it can be inferior to MBB in some cases. Overall, the simulation results are consistent with those in Gonçalves and Vogelsang (2011), and demonstrate the effectiveness of the proposed Gaussian dependent bootstrap in the Gaussian setting. The simulation results also suggest that our procedure may be useful in some non-Gaussian settings, though it can hardly be justified theoretically. The moving block bootstrap is expected to be second order accurate under the fixed-smoothing asymptotics, as seen from its empirical performance, but a rigorous theoretical justification seems very difficult.

4.8 Conclusion

In this chapter, we derive the Edgeworth expansions of the subsampling-based $t$-statistic and the Wald statistic with HAC estimator in the Gaussian location model. Our work differs from the existing ones in two important aspects: (i) the expansion is derived under the fixed-smoothing asymptotics and the ERP of order $O(1/T)$ is shown for a broad class of fixed-smoothing inference procedures; (ii) We obtain an explicit form for the leading error term, which is unavailable in the literature. An in-depth analysis of the behavior of the leading error term when the smoothing parameter grows with sample size (i.e., $K \to \infty$ in the subsampling $t$-statistic or $b \to 0$ in the Wald statistic with the HAC estimator) shows the consistency of our results with the expansion results under the increasing-domain asymptotics. Building on these expansions, we further propose a new bootstrap method, the Gaussian dependent bootstrap, which provides a higher order correction than the first order fixed-smoothing approximation.

We mention a few directions that are worthy of future research. Firstly, it would be interesting to relax the Gaussian assumption in all the expansions we obtained in this chapter. For non-Gaussian time series, Edgeworth expansions have been obtained by Götze and Künsch (1996), Lahiri (2007, 2010), among others, for studentized statistics of a smooth function model under weak dependence assumption, but their results were derived under the increasing-smoothing asymptotics. For the location model and studentized sample mean, the extension to the non-Gaussian case may require an expansion of the corresponding characteristic function, which involves calculation of the high order cumulants under the fixed-smoothing asymptotics. The detailed calculation of the high order terms can be quite involved and challenging. We conjecture that under the fixed-smoothing asymptotics,
the leading error term in the expansion of its distribution function involves the third and fourth order cumulants, which reflects the non-Gaussianness, and the order of the leading error term is $O(T^{-1/2})$ instead of $O(T^{-1})$. Secondly, we expect that our expansion results will be useful in the optimal choice of the smoothing parameter, the kernel and its corresponding eigenvalues and eigenfunctions, for a given loss function. The optimal choice of the smoothing parameter has been addressed in Sun et al. (2008) using the expansion derived under the increasing-smoothing asymptotics. As the finite sample distribution is better approximated by the corresponding fixed-smoothing based approximations at either first or second order than its increasing-smoothing counterparts, the fixed-smoothing asymptotic theory proves to be more relevant in terms of explaining the finite sample results (see Gonçalves and Vogelsang, 2010). Therefore, it might be worth reconsidering the choice of the optimal smoothing parameter under the fixed-smoothing asymptotics. Thirdly, we restrict our attention to the Gaussian location model when deriving the higher order expansions. It would be interesting to extend the results to the general GMM setting. A recent attempt by Sun (2010) for the HAC based inference seems to suggest this is feasible. Finally, under the fixed-smoothing asymptotics, the second correctness of the moving block bootstrap for studentized sample mean, although suggested by the simulation results, is still an open but challenging topic for future research.

4.9 Proof of Theorem 4.5.1

Consider the $K + 1$ dimensional multivariate normal density function which takes the form

$$f(y, \Sigma) = (2\pi)^{-\frac{K+1}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}y'\Sigma^{-1}y\right).$$

We assume the $(i, j)$th element and the $(j, i)$th element of $\Sigma$ are functionally unrelated. The results can be extended to the case where symmetric matrix elements are considered functionally equal (see e.g., McCulloch, 1982). In the following, we use $\otimes$ to denote the Kronecker product in matrix algebra and use vec to denote the operator that transforms a matrix into a column vector by stacking the columns of the matrix one underneath the other. For a vector $y \in \mathbb{R}^{l \times 1}$ whose elements are differential functions of a vector $x \in \mathbb{R}^{k \times 1}$, we define $\frac{\partial y}{\partial x}$ to be a $k \times l$ matrix with the $(i, j)$th element being $\frac{\partial y_j}{\partial x_i}$. The notation $u \asymp v$ represents $u = O(v)$ and $v = O(u)$. We first present the following lemmas whose proofs are given in the next section.
Lemma 4.9.1.  
\[ \frac{\partial f}{\partial \text{vec}(\Sigma)}(y, \Sigma) = \frac{f(y, \Sigma)}{2} \{ (\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1}) \}. \]

Lemma 4.9.2.  
\[ \frac{\partial^2 f}{\partial \text{vec}(\Sigma) \partial \text{vec}(\Sigma)}(y, \Sigma) = \frac{1}{4} \left( (\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1}) \right) \left\{ (\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1}) \right\} f(y, \Sigma) - \frac{1}{2} \left( (\Sigma^{-1}yy') \otimes \Sigma^{-1} + \Sigma^{-1} \otimes (\Sigma^{-1}yy') - \Sigma^{-1} \otimes \Sigma^{-1} \right) f(y, \Sigma). \]

Lemma 4.9.3. Let \( \{ \Sigma_T \} \subset \mathbb{R}^{(K+1) \times (K+1)} \) be a sequence of positive definite matrices with \( K+1 \leq T \). If \( K \) is fixed with respect to \( T \) and \( \| \Sigma_T - \Sigma \|_2 = O(1/T) \) for a positive definite matrix \( \Sigma \), then we have \( \| \Sigma_T^{-1} - \Sigma^{-1} \|_2 = O(1/T) \).

Lemma 4.9.4. Let \( \tilde{\Sigma}_T(y) \) be a \( (K+1) \times (K+1) \) positive symmetric matrix which depends on \( y \in \mathbb{R}^{K+1} \). Assume that \( \sup_{y \in \mathbb{R}^{K+1}} \| \tilde{\Sigma}_T(y) - \Sigma \|_2 \leq \| \Sigma_T - \Sigma \|_2 = O(1/T) \) for a positive definite matrix \( \Sigma \). Let \( R_T = \Sigma_T - \Sigma \). If \( K \) is fixed with respect to \( T \), we have \( \int_{y \in \mathbb{R}^{K+1}} \| \text{vec}(R_T)' \frac{\partial^2 f}{\partial \text{vec}(\Sigma) \partial \text{vec}(\Sigma)}(y, \tilde{\Sigma}_T(y)) \text{vec}(R_T) \| dy = O(1/T^2) \).

Proof of Theorem 4.5.1. For the convenience of our presentation, we ignore the functional symmetry of the covariance matrix in the proof. With some proper modifications, we can extend the results to the case where the functional symmetry is taken into consideration. Let \( |M_1| = |M_2| = \cdots = |M_K| = q \). Define \( Y_i = \sqrt{q}(\mu_i - \mu_0) \), and \( \bar{Y} = \frac{1}{K} \sum_{i=1}^{K} Y_i \) and \( S_Y = \frac{1}{K-1} \sum_{i=1}^{K} (Y_i - \bar{Y})^2 \) as the sample mean and sample variance of \( \{ Y_i \}_{i=1}^{K} \) respectively. Note that \( T_K(Y) = \sqrt{K}\bar{Y}/S_Y \), where \( Y = (Y_1, Y_2, \ldots, Y_K)' \). Simple algebra yields that

\[ \sigma_{ij} := \text{Cov}(Y_i, Y_j) = \sum_{h=1-q}^{q-1} \left( \frac{q - |h|}{q} \right) \gamma_X(h - (j - i)q). \]

Notice that \( Y \) follows a normal distribution with mean zero and covariance matrix \( \Sigma_T \), where \( \Sigma_T = (\sigma_{ij})_{i,j=1}^{K} \). The density function of \( Y \) is given by,

\[ f(y, \Sigma_T) = (2\pi)^{-K/2} |\Sigma_T|^{-1/2} \exp \left( -\frac{1}{2} y' \Sigma_T^{-1} y \right). \]

97
Under the assumption $\sum_{h=-\infty}^{+\infty} h^2 |\gamma_X(h)| < \infty$, it is straightforward to see that $||\Sigma_T - \sigma^2 I_K||_2 = O(1/T)$. Taking a Taylor expansion of $f(y, \Sigma_T)$ around elements of the matrix $\sigma^2 I_K$, we have

$$f(y, \Sigma_T) = f(y, \sigma^2 I_K) + \left\{ \frac{\partial f}{\partial \text{vec}(\Sigma)}(y, \sigma^2 I_K) \right\} \text{vec}(\Sigma_T - \sigma^2 I_K)$$

$$+ \text{vec}(\Sigma_T - \sigma^2 I_K)' \frac{\partial^2 f}{\partial \text{vec}(\Sigma)\text{vec}(\Sigma)}(y, \Sigma_T(y))\text{vec}(\Sigma_T - \sigma^2 I_K),$$

where $\sup_{y \in \mathbb{R}^K} ||\Sigma_T(y) - \sigma^2 I_K||_2 \leq ||\Sigma_T - \sigma^2 I_K||_2 = O(1/T)$. By Lemma 4.10.1 and Lemma 4.10.4, we get

$$\frac{\partial f}{\partial \text{vec}(\Sigma)}(y, \sigma^2 I_K) = f(y, \sigma^2 I_K) \left\{ -\frac{1}{2\sigma^2} \text{vec}(I_K) + \frac{1}{2\sigma^4} y \otimes y \right\},$$

and

$$\int_{y \in \mathbb{R}^K} \text{vec}(\Sigma_T - \sigma^2 I_K)' \frac{\partial^2 f}{\partial \text{vec}(\Sigma)\text{vec}(\Sigma)}(y, \Sigma_T(y))\text{vec}(\Sigma_T - \sigma^2 I_K) \, dy = O\left( \frac{1}{T^2} \right), \quad (4.20)$$

which imply that

$$f(y, \Sigma_T) = f(y, \sigma^2 I_K) \left\{ 1 - \frac{1}{2\sigma^2} \sum_{i=1}^{K} (\sigma_{ii} - \sigma^2) \right\}$$

$$+ \frac{1}{2\sigma^4} f(y, \sigma^2 I_K) \sum_{i=1}^{K} \sum_{j=1}^{K} (\sigma_{ij} - \sigma^2 \delta_{ij}) y_i y_j + R(y)$$

$$= g(y, \sigma^2 I_K) + R(y),$$

where $g$ denotes the major term, $R(y)$ is the remainder term and $\delta_{ij} = I\{i = j\}$ is the kronecker’s delta. Define $\Psi(x; K) = \int_{|T_K(y)| > x} g(y, \sigma^2 I_K) \, dy$. By (4.20), we see that

$$\sup_{x \in \mathbb{R}} \left| \int_{|T_K(y)| > x} f(y, \Sigma_T) \, dy - \Psi(x; K) \right| \leq \int_{\mathbb{R}} |R(y)| \, dy = O(1/T^2).$$

It follows from some simple calculation that

$$\Psi(x; K) = \left\{ 1 - \frac{1}{2\sigma^2} \sum_{i=1}^{K} (\sigma_{ii} - \sigma^2) \right\} P(|t_{K-1}| > x) + \frac{1}{2\sigma^2} (J_1 + J_2),$$

where

$$J_1 = \sum_{i=1}^{K} (\sigma_{ii} - \sigma^2) E[I\{|T_K(v)| > x\} v_i^2], \quad J_2 = \sum_{i \neq j} \sigma_{ij} E[I\{|T_K(v)| > x\} v_i v_j].$$
Here \( \{v_i\}_{i=1}^K \) are iid standard normal random variables and \( \tilde{T}_K(v) = \sqrt{K \bar{v}}/S_v \) is the \( t \) statistic based on \( \{v_i\} \) with \( \bar{v} = \frac{1}{K} \sum_{i=1}^K v_i \) and \( S_v^2 = \frac{1}{K-1} \sum_{i=1}^K (v_i - \bar{v})^2 \). Let \( U = K \bar{v}^2 \) and \( D = (K - 1)S_v^2 \). Then \( U \sim \chi^2_1 \), \( D \sim \chi^2_{K-1} \), and \( U \) and \( D \) are independent. We define that

\[
E[I(\tilde{T}_K(v) > x) v_i^2] = \frac{1}{K} E[I(\tilde{T}_K(v) > x) \sum_{i=1}^K v_i^2]
\]

\[
= \frac{1}{K} E[I(\tilde{T}_K(v) > x) U] + \frac{1}{K} E[I(\tilde{T}_K(v) > x) D]
\]

\[
= \frac{1}{K} E \left[ U \Gamma_{K-1} \left( \frac{(K-1)U}{x^2} \right) \right] + \frac{1}{K} E \left[ D - DG \left( \frac{Dx^2}{K-1} \right) \right],
\]

and

\[
E[I(\tilde{T}_K(v) > x) v_i v_j] = \frac{1}{K(K-1)} E[I(\tilde{T}_K(v) > x) \sum_{i \neq j} v_i v_j]
\]

\[
= \frac{1}{K-1} E[I(\tilde{T}_K(v) > x) U] - \frac{1}{K(K-1)} E[I(\tilde{T}_K(v) > x) \sum_{i=1}^K v_i^2]
\]

\[
= \frac{1}{K} E \left[ U \Gamma_{K-1} \left( \frac{(K-1)U}{x^2} \right) \right] - \frac{1}{K(K-1)} E \left[ D - DG \left( \frac{Dx^2}{K-1} \right) \right].
\]

We then have

\[
P(\|T_K\| > x) = \Psi(x; K) + O(1/T^2)
\]

\[
= \{1 - \alpha\} P(\|t_{K-1}\| > x) + \beta E \left[ U \Gamma_{K-1} \left( \frac{(K-1)U}{x^2} \right) \right]
\]

\[
+ \tau \left\{ K - 1 - E \left[ D \right] \right\} + O(1/T^2),
\]

uniformly for \( x \in \mathbb{R} \), where the coefficients are given by

\[
\alpha = \frac{1}{2\sigma^2} \sum_{i=1}^K (\sigma_{ii} - \sigma^2) = -\frac{K^2B}{2\sigma^2T} + O(1/T^2),
\]

\[
\beta = \frac{1}{2K\sigma^2} \sum_{i=1}^K \sum_{j=1}^K (\sigma_{ij} - \delta_{ij}\sigma^2) = -\frac{B}{2\sigma^2T} + O(1/T^2),
\]

and

\[
\tau = \frac{1}{2K\sigma^2} \sum_{i=1}^K (\sigma_{ii} - \sigma^2) - \frac{1}{2K(K-1)\sigma^2} \sum_{i \neq j} \sigma_{ij} = -\frac{(K+1)B}{2\sigma^2T} + O(1/T^2).
\]

The conclusion thus follows from equation (4.21). \( \square \)
4.10  Proofs of the other main results

Lemma 4.10.1 and Lemma 4.10.2 below are straightforward consequences of matrix calculus (see e.g., Vetter, 1973, Brewer, 1978, and Turkington, 2002).

Lemma 4.10.1.

\[ \frac{\partial f}{\partial \text{vec}(\Sigma)}(y, \Sigma) = \frac{f(y, \Sigma)}{2} \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1}) \} . \]

Proof. By matrix calculus, we get

\[ \frac{\partial f}{\partial \text{vec}(\Sigma)}(y, \Sigma) = (2\pi)^{-\frac{K+1}{2}} \{ \exp \left( -\frac{1}{2} y'\Sigma^{-1}y \right) \frac{\partial |\Sigma|^{-\frac{1}{2}}}{\partial \text{vec}(\Sigma)} + |\Sigma|^{-\frac{1}{2}} \frac{\partial}{\partial \text{vec}(\Sigma)} \exp \left( -\frac{1}{2} y'\Sigma^{-1}y \right) \} \]

\[ = (2\pi)^{-\frac{K+1}{2}} \left\{ -\frac{1}{2} |\Sigma|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} y'\Sigma^{-1}y \right) \text{vec}(\Sigma^{-1}) \right. \]

\[ + \left. \frac{1}{2} |\Sigma|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} y'\Sigma^{-1}y \right) (\Sigma^{-1}y) \otimes (\Sigma^{-1}y) \right\} \]

\[ = \frac{f(y, \Sigma)}{2} \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1}) \} , \]

where we have used the formulas \( \frac{\partial X^{-1}b}{\partial \text{vec}(X)} = -X^{-1}b \otimes (X^{-1})'a \) and \( \frac{\partial X^m}{\partial \text{vec}(X)} = m|X|^{m-1} \frac{\partial X}{\partial \text{vec}(X)} = m|X|^m \text{vec}((X^{-1})') \) (see Theorem 4.3 and Theorem 4.19 in Turkington, 2002).

Lemma 4.10.2.

\[ \frac{\partial^2 f}{\partial \text{vec}(\Sigma) \partial \text{vec}(\Sigma)}(y, \Sigma) = \frac{1}{4} \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1}) \} \{ (\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1}) \}' \frac{f(y, \Sigma)}{2} \]

\[ - \frac{1}{2} \{(\Sigma^{-1}yy'(\Sigma^{-1}) - \Sigma^{-1} + \Sigma^{-1} \otimes (\Sigma^{-1}yy'(\Sigma^{-1}) - \Sigma^{-1} \otimes \Sigma^{-1}) \} f(y, \Sigma) \].

Proof. From Lemma 4.10.1, we have

\[ \frac{\partial^2 f}{\partial \text{vec}(\Sigma) \partial \text{vec}(\Sigma)}(y, \Sigma) = \frac{\partial}{\partial \text{vec}(\Sigma)} \left( \frac{f(y, \Sigma)}{2} \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1}) \} \right) \]

\[ = \left( \frac{\partial}{\partial \text{vec}(\Sigma)} \frac{f(y, \Sigma)}{2} \right) \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1}) \}' \]

\[ + \frac{f(y, \Sigma)}{2} \frac{\partial}{\partial \text{vec}(\Sigma)} \left( \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1}) \} \right) = I_1 + I_2 . \]

Again from Lemma 4.10.1, it is not hard to see that

\[ I_1 = \frac{f(y, \Sigma)}{4} \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1}) \} \{(\Sigma^{-1}y) \otimes (\Sigma^{-1}y) - \text{vec}(\Sigma^{-1}) \} ' . \]
In view of Lemma 4.3 in Turkington (2002), we have
\[
\frac{\partial \text{vec}(\Sigma^{-1}yy')}{\partial \text{vec}(\Sigma)} = \frac{\partial \text{vec}(\Sigma^{-1})y}{\partial \text{vec}(\Sigma)} (y'\Sigma^{-1} \otimes I_{K+1}) + \frac{\partial \text{vec}(y'\Sigma^{-1})}{\partial \text{vec}(\Sigma)} (I_{K+1} \otimes y'\Sigma^{-1}).
\]

Also by Theorem 4.3 in Turkington (2002), we get
\[
\frac{\partial \text{vec}(\Sigma^{-1}yy')}{\partial \text{vec}(\Sigma)} = -\Sigma^{-1}y \otimes \Sigma^{-1}; \quad \frac{\partial \text{vec}(y'\Sigma^{-1})}{\partial \text{vec}(\Sigma)} = -\Sigma^{-1} \otimes \Sigma^{-1}y,
\]
which implies that
\[
\frac{\partial \text{vec}(\Sigma^{-1}yy'\Sigma^{-1})}{\partial \text{vec}(\Sigma)} = -(\Sigma^{-1}yy'\Sigma^{-1}) \otimes \Sigma^{-1} - \Sigma^{-1} \otimes (\Sigma^{-1}yy'\Sigma^{-1}).
\]

Further by Theorem 4.2 in Turkington (2002), we obtain \(\frac{\partial \text{vec}(\Sigma^{-1})}{\partial \text{vec}(\Sigma)} = -\Sigma^{-1} \otimes \Sigma^{-1}\). The conclusion thus follows directly from the above derivation.

\[\square\]

**Lemma 4.10.3.** Let \(\{\Sigma_T\} \subset \mathbb{R}^{(K+1) \times (K+1)}\) be a sequence of positive definite matrices with \(K + 1 \leq T\). If \(K\) is fixed with respect to \(T\) and \(\|\Sigma_T - \Sigma\|_2 = O(1/T)\) for a positive definite matrix \(\Sigma\), then we have
\[
\|\Sigma_T^{-1} - \Sigma^{-1}\|_2 = O(1/T).
\]

**Proof.** Let \(\Sigma_T = \Sigma + R_T\) with \(\|R_T\|_2 = O(1/T)\). For sufficiently large \(T\), we have \(\|\Sigma^{-1}R_T\|_2 \leq \|\Sigma^{-1}\|_2 \|R_T\|_2 < 1\). By the last equation at p. 355 of Horn and Johnson (1986), we have
\[
\|\Sigma_T^{-1} - \Sigma^{-1}\|_2 \leq \frac{\|\Sigma^{-1}\|_2^2 \|R_T\|_2}{1 - \|\Sigma^{-1}\|_2 \|R_T\|_2} = O(1/T).
\]

\[\square\]

**Lemma 4.10.4.** Let \(\tilde{\Sigma}_T(y)\) be a \((K + 1) \times (K + 1)\) positive symmetric matrix which depends on \(y \in \mathbb{R}^{K+1}\). Assume that \(\sup_{y \in \mathbb{R}^{K+1}} \|\tilde{\Sigma}_T(y) - \Sigma\|_2 \leq \|\Sigma_T - \Sigma\|_2 = O(1/T)\) for a positive definite matrix \(\Sigma\). Let \(R_T = \Sigma_T - \Sigma\). If \(K\) is fixed with respect to \(T\), we have
\[
\int_{y \in \mathbb{R}^{K+1}} \left| \text{vec}(R_T)' \frac{\partial^2 f}{\partial \text{vec}(\Sigma) \partial \text{vec}(\tilde{\Sigma}_T(y))} \text{vec}(R_T) \right| dy = O(1/T^2).
\]

**Proof.** Let \(\tilde{R}_T(y) = \tilde{\Sigma}_T(y) - \Sigma\). Note that \(\sup_{y \in \mathbb{R}^{K+1}} \|\Sigma^{-1}\tilde{R}_T(y)\|_2 \leq \|\Sigma^{-1}\|_2 \sup_{y \in \mathbb{R}^{K+1}} \|\tilde{R}_T(y)\|_2 \leq \|\Sigma^{-1}\|_2^2 \|\Sigma_T - \Sigma\|_2 < 1\), for large enough \(T\). By using the same arguments in Lemma 4.10.3, we have \(\sup_{y \in \mathbb{R}^{K+1}} \|\tilde{\Sigma}_T^{-1}(y) - \Sigma^{-1}\|_2 = O(1/T)\). Therefore, when \(T\) is sufficiently large, we have
Proof of Proposition 4.5.2. Note first that
\[ \Upsilon(x; K)/K = -KP(|t_{K-1}| \leq x) + \frac{K + 1}{K} E \left[ \chi_{K-1}^2 G_1 \left( \frac{\chi_{K-1}}{K-1} x^2 \right) \right] + O(1/K). \]

Using the fact that \( P(|t_{K-1}| \leq x) = G_1(x^2) + \frac{1}{K-1} x^4 G_1''(x^2) + O(1/K^2) \), we get
\[
\Upsilon(x; K)/K = -KG_1(x^2) - \frac{K}{K-1} x^4 G_1''(x^2) + \frac{K + 1}{K} E \left[ \chi_{K-1}^2 \left\{ G_1(x^2) \left( \frac{\chi_{K-1}}{K-1} - 1 \right) x^2 G_1'(x^2) + \frac{1}{2} \left( \frac{\chi_{K-1}}{K-1} - 1 \right)^2 x^4 G_1'(x^2) \right\} \right] + O(1/K)
= 2x^2 G_1'(x^2) + O(1/K).
\]

Proof of Proposition 4.5.3. Recall that \( q = T/K \) is assumed to be an integer. Using the notation in
the proof of Theorem 4.5.1, let \( S_Y^2 = \frac{1}{K} \sum_{i=1}^{K} (Y_i - \bar{Y})^2 = \frac{1}{K} \{ \sum_{i=1}^{K} Y_i^2 - K(\bar{Y})^2 \} \). Notice that

\[
\text{cov}(Y) = \begin{pmatrix}
\sigma^2 - B/q & B/(2q) & 0 & \ldots & 0 \\
B/(2q) & \sigma^2 - B/q & B/(2q) & \ldots & 0 \\
& \ddots & \ddots & \ddots & \ddots \\
0 & \ldots & B/(2q) & \sigma^2 - B/q
\end{pmatrix}_{K \times K}
\]

\[+ O(1/q^2) l_K l'_K \]

\[= \sigma^2 I_K + \frac{B}{2q} \begin{pmatrix}
-2 & 1 & 0 & \ldots & 0 \\
1 & -2 & 1 & \ldots & 0 \\
& \ddots & \ddots & \ddots & \ddots \\
0 & \ldots & 1 & -2
\end{pmatrix}_{K \times K} + O(1/q^2) l_K l'_K \]

\[= \sigma^2 I_K + \frac{B}{2q} M + O(1/q^2) l_K l'_K, \]

where \( l'_K = (1, 1, \ldots, 1)_{1 \times K} \) and the summation of all the \( O(1/q^2) \) is of order \( O(K/q^2) \). Because

\[
E[Y_i^2] = \sum_{h=1-q}^{q-1} \left( \frac{q - |h|}{q} \right) \gamma_X(h) = \sigma^2 - B/q + O(1/q^2),
\]

and

\[
E[Y_i Y_j] = \frac{1}{K^2} \sum_{i,j=1}^{K} E[Y_i Y_j] = \frac{1}{K^2} \{ K \sigma^2 - B/q + O(K/q^2) \} = \sigma^2 / K + O(1/K^2) + O(1/Kq^2),
\]

we obtain

\[
E[S_Y^2] - \sigma^2 = \frac{K}{K-1} \{ \sigma^2 - B/q - \sigma^2 / K + o(1/T) \} - \sigma^2 = -B/q + O(1/T).
\]

Consider the covariance matrix of \( \tilde{Y}' = (Y_1 - \bar{Y}, Y_2 - \bar{Y}, \ldots, Y_K - \bar{Y}) \). It is easy to see that \( \tilde{Y} = (I_K - l_K l'_K / K) Y = H_K Y \), where \( H_K = I_K - l_K l'_K / K \) is an idempotent matrix. Ignoring the \( O(1/q^2) \) order term in \( \text{cov}(Y) \), we have

\[
(c_{ij})_{i,j=1}^{K} := \text{cov}(\tilde{Y}) = H_K \text{cov}(Y) H_K \approx H_K \{ \sigma^2 I_K + B M/(2q) \} H_K
\]

\[= \sigma^2 H_K + \frac{B}{2q} H_K M H_K = \sigma^2 H_K + \frac{B}{2q} \left( M - \frac{1}{K} A - \frac{2}{K^2} l_K l'_K \right), \]
where

\[
A = \begin{pmatrix}
\begin{array}{cccccc}
-2 & -1 & -1 & \ldots & -2 \\
-1 & 0 & 0 & \ldots & -1 \\
\vdots & & & & \vdots \\
-1 & 0 & 0 & \ldots & -1 \\
-2 & -1 & -1 & \ldots & -2
\end{array}
\end{pmatrix}_{K \times K}.
\]

Since \( \hat{Y} \) is Gaussian, we get

\[
E[S_Y^4] = \frac{1}{(K-1)^2} \sum_{i,j=1}^{K} E[(Y_i - \hat{Y})^2(Y_j - \hat{Y})^2] = \frac{1}{(K-1)^2} \sum_{i,j=1}^{K} (c_{ii}c_{jj} + 2c_{ij}^2),
\]

where \( c_{ii} = \left(1 - \frac{1}{K}\right)\sigma^2 - \frac{B}{q} + O(1/T) \) and \( c_{ij} = -\frac{1}{K}\sigma^2 + \frac{B}{2q} \mathbb{I}(|i-j| = 1) + O(1/T) \), for \( i \neq j \). It implies that

\[
\sum_{i,j=1}^{K} c_{ij}^2 = \sum_{i=1}^{K} c_{ii}^2 + \sum_{|i-j|=1} c_{ij}^2 + \sum_{|i-j|>1} c_{ij}^2 = K \left(1 - \frac{1}{K}\right)^2 \sigma^4 + \frac{KB^2}{q^2} - \frac{2(K-1)B}{q} \sigma^2 + 2(K-1)
\]

\[
\left(\frac{\sigma^4}{K^2} + \frac{B^2}{4q^2} - \frac{\sigma^2 B}{Kq}\right) + \frac{(K-1)(K-2)}{K^2} \sigma^4 + O(1/q)
\]

\[
= (K-1)\sigma^4 + O(K/q),
\]

and

\[
\sum_{i,j=1}^{K} c_{ii}c_{jj} = K^2c_{11}^2 + O(K/q) = (K-1)^2\sigma^4 - \frac{2BK(K-1)\sigma^2}{q} + O(K/q).
\]

Therefore we get

\[
E[S_Y^4] = \frac{K + 1}{K-1} \sigma^4 - \frac{2BK\sigma^2}{(K-1)q} + O(1/T),
\]

which implies

\[
\text{var}(S_Y^2) = \frac{K + 1}{K-1} \sigma^4 - \frac{2BK\sigma^2}{(K-1)q} - (\sigma^2 - B/q)^2 + O(1/T) = \frac{2\sigma^4}{K-1} + O(1/T).
\]

Let \( X = (X_1, X_2, \ldots, X_T)' \), \( \hat{\mu}_{GLS} = (l'_T \text{cov}(X)^{-1} l_T)^{-1} l'_T \text{cov}(X)^{-1} X \) and \( \sigma^2_{GLS} = T\text{var}(\hat{\mu}_{GLS}) = T(l'_T \text{cov}(X)^{-1} l_T)^{-1} \). Note that \( \hat{\mu}_{GLS} - \mu_0 \) is independent of \( S_Y \) and \( \sigma^2_{GLS} = \sigma^2 + O(1/T) \) (see
Proof of Lemma 4.10.5. For \( x \), using similar arguments in Lemma 1 of Sun (2011), we have

\[
P(|T_K| \leq x) = P \left( \frac{T(\hat{\mu}_{GLS} - \mu_0)^2}{\sigma^2_{GLS}} \leq \frac{x^2}{\alpha^2} \right) + O(1/T)\]

\[
= E[G_1(S_\nu^T x^2/\sigma^2)] + O(1/T) \]

\[
= G_1(x^2) + \frac{x^2}{\sigma^2} G'_1(x^2) E[S_\nu^T - \sigma^2] + \frac{x^4 G''_1(x^2)}{2\sigma^4} E[(S_\nu^T - \sigma^2)^2] + O(1/T) \]

\[
= G_1(x^2) - \frac{BK}{T\sigma^2} x^2 G'_1(x^2) + \frac{1}{K-1} x^2 G''_1(x^2) + O(1/T). \]

\[\square\]

4.10.1 Proof of the main results in Section 4.5.2

We first establish a high order expansion for Wald statistic based on the kernel \( G_{k,1}(r, t) = \sum_{i=1}^k \lambda_i \phi_i(r) \phi_j(t) \) in Lemma 4.10.6 below. Let \( \xi = (\xi_0, \xi_1, \ldots, \xi_K) \) with \( \xi_0 = \frac{1}{\sqrt{T}} \sum_{i=1}^T (X_i - \mu_0) \) and \( \xi_j = \frac{1}{\sqrt{T}} \sum_{i=1}^T \phi_j(i/T) X_i \) for \( j = 1, 2, \ldots, K \), and \( \Sigma_\xi \) be the covariance matrix of \( \xi \). Define \( Q_j(x) = P(F(J) \leq x) \) for \( 1 \leq J \leq \infty \). We present the following lemma regarding the convergence rate of \( \Sigma_\xi \) for the basis functions \( \{ \phi_j(t) \}_{j=1}^K \) without the mean zero and orthogonality assumption. Define \( R = (R_{ij})_{i,j=1}^K \) with \( R_{ij} = \int_0^1 \phi_i(t) \phi_j(t) dt \), where \( \phi_j(t) = \phi_j(t) - \int_0^1 \phi_j(t) dt \), and \( \tilde{R} = \text{diag}(1, R) = (\tilde{R}_{i,j})_{i,j=0}^K \).

Lemma 4.10.5. Assume the basis functions \( \{ \phi_j(t) \}_{j=1}^K \) are bounded with finite discontinuous points and satisfy \( \sup_{\alpha \in [0,1]} \left\{ \frac{1}{\alpha} \int_0^{1-\alpha} \phi_x(s) \{ \tilde{\phi}_r(x + \alpha) - \tilde{\phi}_r(x) \} dx \right\} + \frac{1}{\alpha} \int_0^1 \phi_x(s) \{ \tilde{\phi}_r(x - \alpha) - \tilde{\phi}_r(x) \} dx \right\} < \infty \), for \( 1 \leq s, r \leq K \). If \( \sum_{h=-\infty}^{+\infty} h^2 |\gamma_X(h)| < \infty \) and \( K \) is fixed, then we have \( ||\Sigma_\xi - \sigma^2 \tilde{R}||_\infty = O(1/T) \).

Proof of Lemma 4.10.5. For \( s = 1, 2, \ldots, K \), we have

\[
\text{cov}(\xi_0, \xi_s) = \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^T \gamma_X(j-i) \phi_s^0 \left( \frac{j}{T} \right) = \frac{1}{T} \sum_{h=1-\infty}^{T-1} \gamma_X(h) \sum_{1 \leq i, h+i \leq T} \phi_s^0 \left( \frac{h+i}{T} \right). \]

Simple algebra gives us

\[
\frac{1}{T} \sum_{1 \leq i, h+i \leq T} \phi_s^0 \left( \frac{h+i}{T} \right) = \begin{cases} \frac{1}{T} \sum_{i=1}^T \phi_s(i/T) - \frac{1}{T} \sum_{i=1}^h \phi_s(i/T), & h > 0; \\ \frac{1}{T} \sum_{i=1}^T \phi_s(i/T) - \frac{1}{T} \sum_{i=T-|h|+1}^T \phi_s(i/T), & h < 0. \end{cases} \]
It implies that
\[
\text{cov}(\xi_0, \xi_s) = \frac{1}{T} \int_0^1 \phi_s(t) dt \sum_{h=-\infty}^{+\infty} |h| \gamma_X(h) - \frac{1}{T} \sum_{0<h<T} \gamma_X(h) \left\{ \sum_{i=1}^h \phi_s(i/T) + \sum_{i=T-h+1}^T \phi_s(i/T) \right\} + O(1/T^2).
\]

(4.22)

Note that the second term on the right hand side of (4.22) is of order \(O(1/T)\) because the basis functions \(\{\phi_s(t)\}\) are bounded. Consider the covariance between \(\xi_s\) and \(\xi_r\) with \(1 \leq s, r \leq K\).

Straightforward calculation yields
\[
\text{cov}(\xi_s, \xi_r) = \frac{1}{T} \sum_{i=1}^T \sum_{j=1}^T \phi_s^0 \left( \frac{i}{T} \right) \phi_r^0 \left( \frac{j}{T} \right) \gamma_X(i-j)
\]
\[
= \frac{1}{T} \sum_{h=1}^{T-1} \sum_{1 \leq j, j+h \leq T} \phi_s^0 \left( \frac{j+h}{T} \right) \phi_r^0 \left( \frac{j}{T} \right) \gamma_X(h) + \frac{1}{T} \sum_{h=1}^{T-1} \sum_{r \leq j, j+h \leq T} \phi_s^0 \left( \frac{j+h}{T} \right) \phi_r^0 \left( \frac{j}{T} \right) \gamma_X(h)
\]
\[
+ \gamma_X(0) \frac{1}{T} \sum_{j=1}^T \phi_s^0 \left( \frac{j}{T} \right) \phi_r^0 \left( \frac{j}{T} \right) = I_1 + I_2 + I_3.
\]

Notice that
\[
\frac{1}{T} \sum_{1 \leq j \leq T} \phi_s^0 \left( \frac{j}{T} \right) \phi_r^0 \left( \frac{j}{T} \right) = \int_0^1 \phi_s(t) \phi_r(t) dt + C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t)) = R_{sr} + C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t)),
\]

(4.23)

where \(C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t))\) is of order \(O(1/T)\). It is not hard to see that
\[
I_1 = \frac{1}{T} \sum_{h=1}^{T-1} \gamma_X(h) \left\{ \sum_{j=1}^{T-h} \phi_s^0 \left( \frac{j}{T} \right) \phi_r^0 \left( \frac{j+h}{T} \right) - \phi_s^0 \left( \frac{j}{T} \right) \right\} = J_{1,T},
\]
\[
I_2 = \frac{1}{T} \sum_{h=1}^{T-1} \gamma_X(h) \left\{ \sum_{j=1}^{h} \phi_r^0 \left( \frac{j}{T} \right) \phi_s^0 \left( \frac{j+h}{T} \right) - \sum_{j=T-h+1}^T \phi_r^0 \left( \frac{j}{T} \right) \phi_s^0 \left( \frac{j}{T} \right) \right\},
\]
and
\[
I_3 = \frac{1}{T} \sum_{h=1}^{T-1} \gamma_X(h) \left\{ \sum_{j=1}^{h} \phi_r^0 \left( \frac{j}{T} \right) \phi_s^0 \left( \frac{j}{T} \right) - \sum_{j=1}^{h} \phi_r^0 \left( \frac{j}{T} \right) \phi_s^0 \left( \frac{j}{T} \right) \right\} = J_{2,T}.
\]
Using (4.23), we have
\[
\text{cov}(\xi_s, \xi_r) = \left\{ \begin{array}{l}
\sigma^2 - \sum_{|h| \geq T} \gamma_X(h) \{ R_{sr} + C_T(\hat{\phi}_s(t), \hat{\phi}_r(t)) \} - \frac{1}{T} \sum_{h=1}^{T-1} \gamma_X(h) \left\{ \sum_{j=1}^{h} \phi^0_r \left( \frac{j}{T} \right) \phi^0_s \left( \frac{j}{T} \right) \right\} \\
+ \sum_{j=T-h+1}^{T} \phi^0_r \left( \frac{j}{T} \right) \phi^0_s \left( \frac{j}{T} \right) \right\} + J_{1,T} + J_{2,T}.
\end{array} \right.
\]

Under the assumption that \( \sup_{a \in [0,1]} |\frac{1}{\alpha} \int_0^{1-\alpha} \tilde{\phi}_r(x)(\tilde{\phi}_s(x + \alpha) - \tilde{\phi}_s(x))dx| < \infty \), it is straightforward to see that
\[
|J_{1,T}| \leq \frac{1}{T} \sum_{h=1}^{T-1} |h\gamma_X(h)| \sup_{1 \leq h \leq T} \left| \frac{1}{h} \sum_{j=1}^{T-h} \phi^0_r \left( \frac{j}{T} \right) \left\{ \phi^0_s \left( \frac{j + h}{T} \right) - \phi^0_s \left( \frac{j}{T} \right) \right\} \right|
\leq C \frac{1}{T} \sum_{h=1}^{T-1} |h\gamma_X(h)| \left\{ \sup_{a \in [0,1]} \left| \frac{1}{\alpha} \int_0^{1-\alpha} \tilde{\phi}_r(x)(\tilde{\phi}_s(x + \alpha) - \tilde{\phi}_s(x))dx \right| \right\} .
\]

which implies that \( J_{1,T} = O(1/T) \). The same argument applies to \( J_{2,T} \). The proof is then complete.

The assumption regarding the basis functions in Lemma 4.10.5 is mild. If \( \{ \phi_j(t) \}_{j=1}^K \) are lipschitz continuous of order one, then the assumption is satisfied.

**Lemma 4.10.6.** Suppose \( \sigma^2 > 0 \) and the basis functions \( \{ \phi_j(t) \}_{j=1}^K \) are mean zero and orthogonal. Under the assumptions in Lemma 4.10.5 and \( H_0 \), we have \( \sup_{x \in [0,1]} |\mathbb{R}_T(x; K)| = O(1/T) \) and
\[
\sup_{x \in [0,1]} |P(F_T(K) \leq x) - Q_K(x) - \mathbb{R}_T(x; K)| = O(1/T^2),
\]
with \( K \) fixed and \( T \to \infty \).

**Proof of Lemma 4.10.6.** Note that when \( \{ \phi_j(t) \}_{j=1}^K \) are mean zero and orthogonal, we have \( \hat{R} = I_{K+1} \). It follows directly from Lemma 4.10.5 that \( \sup_{x \in [0,\infty)} |\mathbb{R}_T(x; K)| = O(1/T) \). To show the second part, we first note that under the Gaussian assumption, the density function of \( \xi \) is given by \( f(u, \Sigma_\xi) = (2\pi)^{-(K+1)/2} \Sigma_\xi^{-1/2} \exp \left( -\frac{1}{2} u' \Sigma_\xi^{-1} u \right) \). Taking a Taylor expansion of the density function \( f(u, \Sigma_\xi) \) around the covariance matrix \( \sigma^2 I_{k+1} \), we get
\[
f(u, \Sigma_\xi) = f(u, \sigma^2 I_{K+1}) + \frac{\partial f}{\partial \text{vec}(\Sigma)}(u, \sigma^2 I_{K+1}) \text{vec}(\Sigma_\xi - \sigma^2 I_{k+1}) + R_T(u).
\]
By Lemma 4.10.4, the remainder term $R_T(u)$ satisfies that $\int_{\mathbb{R}^{K+1}} |R_T(u)| dv = O(1/T^2)$. Following Lemma 4.10.1, we have $\frac{\partial f}{\partial \text{vec}(\Sigma)}(u, \sigma^2 I_{K+1}) = f(u, \sigma^2 I_{K+1}) \left\{ \frac{1}{\sigma^2} u \otimes u - \frac{1}{\sigma^2} \text{vec}(I_{K+1}) \right\}$, which implies that

$$P(F_T(K) \leq x) = Q_K(x) \left\{ 1 - \frac{1}{2\sigma^2} \sum_{i=0}^{K} (\text{var}(\xi_i) - \sigma^2) \right\} + \zeta_T(x),$$

where $\zeta_T(x) = \frac{1}{2\sigma^2} \int_{\mathcal{F}(u; K) \leq x} f(u, \sigma^2 I_{K+1}) u' \otimes u' \text{vec}(\Sigma - \sigma^2 I_{K+1}) du + \int_{\mathcal{F}(u; K) \leq x} R_T(u) du$ and $\mathcal{F}(u; K) = \sum_{j=1}^{u_0} \lambda_j u_j^2$. By letting $v = u/\sigma$ and noting that $E[I\{\mathcal{F}(v; K) \leq x\} v_s v_r] = 0$ for $s \neq r$, we obtain

$$\zeta_T(x) = \frac{1}{2\sigma^2} \sum_{i=0}^{K} E[I\{\mathcal{F}(v; K) \leq x\} v_i^2 (\text{var}(\xi_i) - \sigma^2) + \int_{\mathcal{F}(u; K) \leq x} R_T(u) du,$$

where $v = (v_0, v_1, \ldots, v_K)$ is a $(K+1)$-dimensional vector of i.i.d. standard normal random variables. Therefore, we get

$$\sup_{x \in [0, +\infty)} |P(F_T(K) \leq x) - Q_K(x) - R_T(x; K)| = \sup_{x \in [0, +\infty)} \left| \int_{\mathcal{F}(u; K) \leq x} R_T(u) du \right| \leq \int_{\mathbb{R}^{K+1}} |R_T(u)| dv = O(1/T^2),$$

which completes the proof. 

\[\square\]

**Lemma 4.10.7.** Let $\{\Sigma_{T, J+1}\} \subset \mathbb{R}^{(J+1) \times (J+1)}$ be an array of positive definite matrices with $J + 1 \leq T$. Assume that $||\Sigma_{T, J+1} - \Sigma_{J+1}||_\infty = O(1/T)$ for a sequence of positive definite matrices $\{\Sigma_j\}_{j=1}^\infty$ with $\sup_j ||\Sigma_{J+1}^{-1}||_2 < \infty$. If $J$ satisfies that $1/J + J^2/T \to 0$, then we have $||\Sigma_{T, J+1}^{-1} - \Sigma_{J+1}^{-1}||_\infty = O(J^2/T)$.

**Proof.** Let $\Sigma_{T, J+1} = \Sigma_{J+1} + R_T, J+1$. For sufficiently large $T$, we have $||\Sigma_{J+1}^{-1} R_T, J+1||_2 \leq (J + 1)||\Sigma_{J+1}^{-1}||_2 ||R_T, J+1||_\infty < 1$, where we are using the fact that $||R_T, J+1||_2 \leq (J + 1)||R_T, J+1||_\infty$. It follows that

$$||\Sigma_{T, J+1}^{-1} - \Sigma_{J+1}^{-1}||_\infty \leq ||\Sigma_{T, J+1}^{-1} - \Sigma_{J+1}^{-1}||_2 \leq (J + 1) \frac{||\Sigma_{J+1}^{-1}||_2^2 ||R_T, J+1||_\infty}{1 - ||\Sigma_{J+1}^{-1} R_T, J+1||_2} = O(J^2/T).$$

\[\square\]
Lemma 4.10.8. Let $\tilde{\Sigma}_{T,j+1}(y)$ be a $(J + 1) \times (J + 1)$ positive definite matrix which depends on $y \in \mathbb{R}^{J+1}$, and $\Sigma_{T,j+1}$ and $\Sigma_j = \sigma^2 I_j$ satisfy the assumptions in Lemma 4.10.7. Assume that $\sup_{y \in \mathbb{R}^{J+1}} ||\tilde{\Sigma}_{T,j+1}(y) - \sigma^2 I_{J+1}|| \leq ||\Sigma_{T,j+1} - \sigma^2 I_{J+1}||_{\infty} = O(J/T)$. Let $R_{T,j+1} = \Sigma_{T,j+1} - \sigma^2 I_{J+1}$. If $J = o(T^{1/6})$, we have

$$\int_{y \in \mathbb{R}^{J+1}} \left| \frac{\partial^2 f}{\partial \text{vec}(\Sigma) \partial \text{vec}(\Sigma)}(y, \tilde{\Sigma}_{T,j+1}(y)) \text{vec}(R_{T,j+1}) \right| dy = o(1/T).$$

Proof. Let $R_{T,j+1}(y) = \tilde{\Sigma}_{T,j+1}(y) - \sigma^2 I_{J+1}$. Note first that $\sup_{y \in \mathbb{R}^{J+1}} ||R_{T,j+1}(y)/\sigma^2||_2 \leq (J + 1) \sup_{y \in \mathbb{R}^{J+1}} ||R_{T,j+1}(y)||_{\infty}/\sigma^2 \leq (J + 1)||\Sigma_{T,j+1} - \sigma^2 I_{J+1}||_{\infty}/\sigma^2 < 1$, for large enough $T$. Following the arguments in Lemma 4.10.7, we know that

$$\sup_{y \in \mathbb{R}^{J+1}} ||\tilde{\Sigma}_{T,j+1}(y) - \sigma^{-2} I_{J+1}||_2 \leq \frac{C(J + 1)||\Sigma_{T,j+1} - \sigma^2 I_{J+1}||_{\infty}}{1 - (J + 1)||\Sigma_{T,j+1} - \sigma^2 I_{J+1}||_{\infty}/\sigma^2} = O(J^2/T).$$

Choose $r = J^3/T$. Then we have

$$y' \left( \tilde{\Sigma}_{T,j+1}^{-1}(y) - \frac{1}{(1 + r)\sigma^2} I_{J+1} \right) y = y' \left( \tilde{\Sigma}_{T,j+1}^{-1}(y) - \frac{1}{\sigma^2} I_{J+1} \right) y + \frac{r}{(r+1)\sigma^2} ||y||^2 \\
\geq \left( \frac{r}{(r+1)\sigma^2} - ||\tilde{\Sigma}_{T,j+1}^{-1}(y) - I_{J+1}/\sigma^2||_2 \right) ||y||^2 \geq 0,$$

when $T$ is sufficiently large. On the other hand, we have

$$\sup_{y \in \mathbb{R}^{J+1}} ||\tilde{\Sigma}_{T,j+1}^{-1}(y)||_2 \leq \sup_{y \in \mathbb{R}^{J+1}} ||\tilde{\Sigma}_{T,j+1}^{-1}(y)||_{2}^{J+1} \leq \left( \frac{1}{\sigma^2} + \frac{CJ^2}{T} \right)^{J+1} \leq \left( \frac{1}{(r+1)\sigma^2} I_{J+1} \right) \left( 1 + r + \frac{C(r+1)J^2\sigma^2}{T} \right)^{J+1} \leq \frac{1}{(r+1)\sigma^2} I_{J+1} \left( 1 + C\sigma^2 \right)^{(1/r)(J+1)r} \leq C \left[ \frac{1}{(r+1)\sigma^2} I_{J+1} \right].$$
The above arguments imply that \( f(y, \bar{\Sigma}_{T,J+1}(y)) \leq C f(y, (1+r)\sigma^2 I_{J+1}) \) for all \( y \). Therefore we get

\[
\int_{y \in \mathbb{R}^{J+1}} \left| \text{vec}(R_{T,J+1})' \frac{\partial^2 f}{\partial \text{vec}(\Sigma) \partial \text{vec}(\Sigma)}(y, \bar{\Sigma}_{J+1}(y)) \text{vec}(R_{T,J+1}) \right| dy \\
\leq C \int_{y \in \mathbb{R}^{J+1}} \left| \text{vec}(R_{T,J+1})' \left\{ (\bar{\Sigma}_{J+1}(y)) \otimes (\bar{\Sigma}_{J+1}^{-1}(y)) - \text{vec}(\bar{\Sigma}_{J+1}^{-1}(y)) \{ (\bar{\Sigma}_{J+1}^{-1}(y)) \otimes (\bar{\Sigma}_{J+1}^{-1}(y)) \} \right\} \text{vec}(R_{T,J+1}) \right| f(y, (1+r)\sigma^2 I_{J+1}) dy \\
+ C \int_{y \in \mathbb{R}^{J+1}} \left| \text{vec}(R_{T,J+1})' \left\{ (\bar{\Sigma}_{J+1}(y)) y y' (\bar{\Sigma}_{J+1}^{-1}(y)) \otimes (\bar{\Sigma}_{J+1}^{-1}(y)) + \bar{\Sigma}_{J+1}^{-1}(y) \otimes (\bar{\Sigma}_{J+1}^{-1}(y)) y y' (\bar{\Sigma}_{J+1}^{-1}(y)) \right\} \right| \text{vec}(R_{T,J+1}) \right| f(y, (1+r)\sigma^2 I_{J+1}) dy \leq C J^6 / T^2 = o(1/T),
\]

where the first inequality in the last row comes from the fact that \( \sup_{y \in \mathbb{R}^{J+1}} \| \bar{\Sigma}_{J+1}^{-1}(y) - \sigma^{-2} I_{J+1} \|_{\infty} \leq \sup_{y \in \mathbb{R}^{J+1}} \| \bar{\Sigma}_{J+1}^{-1}(y) - \sigma^{-2} I_{J+1} \|_2 = O(J^2 / T) \).

**Lemma 4.10.9.** Recall that \( Q_J(x) = P(\mathcal{F}(J) \leq x) \) for \( 1 \leq J \leq \infty \). We have

\[
\sup_{x \in [0, +\infty)} |Q_J(x) - Q_\infty(x)| = O \left( \sum_{j=J+1}^{\infty} \lambda_j \right). \tag{4.26}
\]

**Proof.** Let \( \{v_j\}_{j=0}^{+\infty} \) be a sequence of i.i.d. standard normal random variables. Define \( U(J) = \sum_{j=1}^{J} \lambda_j v_j^2 \), \( V(J) = \sum_{j=J+1}^{\infty} \lambda_j v_j^2 \) and \( Q_J = v_J^2 / U(J) \) for \( 2 \leq J \leq \infty \). For any \( x \in [0, +\infty) \) and large enough \( J \) with \( J \geq 3 \), we have,

\[
|Q_J(x) - Q_\infty(x)| = |E[E[I\{Q_J \leq x\}|U(J)] - E[E[I\{Q_\infty \leq x\}|U(\infty)]]| \\
= |E[G_1(xU(J))] - E[G_1(xU(\infty))]| \\
= |E[G_1(xU(J) + xV(J))] - E[G_1(xU(J))]| \\
= |E[xV(J)G_1'(xU(J))]| = \left| E \left[ \frac{V(J)}{U(J)} xU(J)G_1'(xU(J)) \right] \right| \\
\leq CE \left[ \frac{V(J)}{U(J)} \right] \leq CE[V(J)] E \left[ \frac{1}{U(J)} \right] \leq C \sum_{j=J+1}^{\infty} \lambda_j,
\]

where \( U(J) \leq \hat{U}(J) \leq U(J) + V(J) \) and \( C \) does not depend on \( x \). Note that we are using the mean value theorem, and the facts that \( E[1/U(J)] \leq E \left[ 1/(\lambda_3 \hat{\lambda}_3^3) \right] < \infty \) and \( \sup_{x \in [0, +\infty)} |xG_1'(x)| < \infty \).

**Lemma 4.10.10.** Let \( V_T(J) = \sum_{j=J+1}^{\infty} \lambda_j \xi_j^2 \). Assume that \( \sup_{1 \leq i \leq \infty} \sup_{t \in [0,1]} \phi_i(t) < \infty \) and \( \{X_i\} \) is a stationary Gaussian time series. Then we have \( EV_T^2(J) = O((\sum_{j=J+1}^{\infty} \lambda_j)^2) \).
Proof. Let \( \sigma_{ij} = \gamma_X(i - j) \). For \( i, j \geq J + 1 \), we have

\[
E[\xi_i^2 \xi_j^2] = \frac{1}{T^2} \sum_{i_1,i_2=1}^{T} \sum_{j_1,j_2=1}^{T} \phi_{i_1}^0(i_1/T)\phi_{i_2}^0(i_2/T)\phi_j^0(j_1/T)\phi_j^0(j_2/T)E[(X_{i_1} - \mu)(X_{i_2} - \mu)(X_{j_1} - \mu)(X_{j_2} - \mu)]
\]

\[= \frac{1}{T^2} \sum_{i_1,i_2=1}^{T} \sum_{j_1,j_2=1}^{T} \phi_{i_1}^0(i_1/T)\phi_{i_2}^0(i_2/T)\phi_j^0(j_1/T)\phi_j^0(j_2/T)(\sigma_{i_1,i_2}\sigma_{j_1,j_2} + \sigma_{i_1,j_1}\sigma_{i_2,j_2} + \sigma_{i_1,j_2}\sigma_{i_2,j_1})
\]

\[= I_{1,T} + I_{2,T} + I_{3,T}.
\]

For the first term, we have

\[I_{1,T} = \left( \frac{1}{T} \sum_{i_1,i_2=1}^{T} \phi_{i_1}^0(i_1/T)\phi_{i_2}^0(i_2/T)\sigma_{i_1,i_2} \right) \left( \frac{1}{T} \sum_{j_1,j_2=1}^{T} \phi_j^0(j_1/T)\phi_j^0(j_2/T)\sigma_{j_1,j_2} \right) = L_{1,T}L_{2,T}.
\]

Note that

\[|L_{1,T}| = \frac{1}{T} \sum_{h=1}^{T-1} \sum_{1 \leq i_1,j_1+h \leq T} \phi_{i_1}^0(i_1/T)\phi_{i_1}^0(h/T)\gamma_X(h) \leq C \sum_{h=-\infty}^{+\infty} \gamma_X(h)\frac{1}{T} \sum_{1 \leq i_1 \leq T} |\phi_{i_1}^0(i_1/T)|
\]

\[\leq C \sum_{h=-\infty}^{+\infty} \gamma_X(h),
\]

which implies that \( |I_{1,T}| \leq C(\sum_{h=-\infty}^{+\infty} \gamma_X(h))^2 \). Similar arguments apply to the other terms \( I_{2,T} \) and \( I_{3,T} \). We then have \( \sup_{J+1 \leq i,j \leq T} E[\xi_i^2 \xi_j^2] < C \). Therefore, we obtain

\[E[V_T(J)^2] = \sum_{i=J+1}^{\infty} \sum_{j=J+1}^{\infty} \lambda_i \lambda_j E[\xi_i^2 \xi_j^2] \leq C \left( \sum_{i=J+1}^{\infty} \lambda_i \right)^2.
\]

\[\square
\]

Lemma 4.10.11. Assume the eigenfunctions are continuously differentiable, mean zero and uniformly bounded, and \( \sum_{j=1}^{\infty} \lambda_j < \infty \). Suppose that \( \{X_i\} \) is a stationary Gaussian time series with \( \sum_{h=-\infty}^{+\infty} h^2 |\gamma_X(h)| < \infty \). When \( 1/J + J/T \to 0 \), we have

\[\sup_{x \in [0, +\infty)} |P(F_T(J) \leq x) - P(F_T(\infty) \leq x)| = O \left( \left( \sum_{j=J+1}^{\infty} \lambda_j \right)^{1/3} \right).
\]

Recall that \( F_T(J) = \frac{\xi_J^2}{\sum_{j=1}^{J} \lambda_j \xi_j^2} \) for \( J = 1, 2, \ldots, \infty \).
Proof. Let \( R_T(J) = F_T(J) - F_T(\infty) \) and \( \frac{\xi_T^2 V_j(J)}{(\sum_{j=1}^{\infty} \lambda_j \xi_j^2)} \). For any \( \delta > 0 \), we have

\[
P(F_T(\infty) \leq x - \delta) - P(|R_T(J)| \geq \delta) \leq P(F_T(J) \leq x) \leq P(F_T(\infty) \leq x + \delta) + P(|R_T(J)| \geq \delta). \tag{4.27}
\]

Observe that

\[
P(|R_T(J)| \geq \delta) \leq \frac{E|R_T(J)|}{\delta} \leq \frac{(E[V_j^2(J)])^{1/2}}{\delta} \left( E \left[ \frac{\xi_j^4}{(\sum_{j=1}^{\infty} \lambda_j \xi_j^2)^4} \right] \right)^{1/2}.
\]

Choose a fixed \( J_0 \geq 9 \), denote by \( \hat{\Sigma}_{T,J_0+1} \) the covariance matrix of \((\xi_0, \ldots, \xi_{J_0})\). By Lemma 4.10.5, we know that \( ||\hat{\Sigma}_{T,J_0+1} - \sigma^2 I_{J_0+1}|| \leq (J_0 + 1)||\hat{\Sigma}_{T,J_0+1} - \sigma^2 I_{J_0+1}|| \leq O(1/T) \). For large enough \( T \), we have \( ||\hat{\Sigma}_{T,J_0+1}|| \leq 2 \sigma^2 \). Let \( \tilde{\lambda} = \min(1, \frac{1}{2\sigma^2}) > 0 \), we know that \( \hat{\Sigma}_{T,J_0+1} - \tilde{\lambda} I_{J_0+1} \) is semi-positive definite, i.e., for any \( x \in \mathbb{R}^{J_0+1}, x^T \hat{\Sigma}_{T,J_0+1} x \geq \tilde{\lambda} x^T x \). Using similar arguments in Lemma 4.10.3, we know that \( ||\hat{\Sigma}_{T,J_0+1}|| \leq ||\hat{\Sigma}_{T,J_0+1}|| \leq 2/(\sigma^2)^{J_0+1} \) for large enough \( T \). For any \( J \geq J_0 \), we have

\[
E \left[ \frac{\xi_0^4}{(\sum_{j=1}^{J_0} \lambda_j \xi_j^2)^4} \right] \leq \frac{\xi_0^4}{E \left[ \frac{\lambda_j^4}{(\sum_{j=1}^{J_0} \lambda_j \xi_j^2)^4} \right]} \leq \frac{1}{(2\pi)^{J_0+1/2}||\Sigma_{T,J_0+1}||^{1/4}} \int_{w \in \mathbb{R}^{J_0+1}} \frac{w_0^4}{\lambda_j^4 (\sum_{j=1}^{J_0} w_j^2)^4} \exp(-\tilde{\lambda} w^T w/2) dw \leq CE[(\chi_2^2)^2] E[(1/\chi_2^4)] < \infty,
\]

where \( w = (w_0, w_1, \ldots, w_{J_0}) \) and \( \chi_m^2 \) denotes a chi-square random variable with \( m \) degrees of freedom. By Lemma 4.10.10, we obtain

\[
P(|R_T(J)| \geq \delta) \leq C \left( \sum_{j=J_0+1}^{\infty} \lambda_j / \delta \right). \tag{4.28}
\]

In what follows, we show that \( \sup_{x \in [0,\infty)} |P(F_T(\infty) \leq x + \delta) - P(F_T(\infty) \leq x)| \leq C \sqrt{\delta} \) for any \( \delta > 0 \). Let \( X = (X_1, X_2, \ldots, X_T)' \), \( l_T = (1, 1, \ldots, 1)' \), \( X^* = X - l_T \mu_0 \) and \( \Omega_T = \text{cov}(X) \). Then the GLS estimate of \( \mu \) is given by \( \hat{\mu}_{GLS} = (l_T' \Omega_T^{-1} l_T)^{-1} l_T' \Omega_T^{-1} X \) and \( \hat{\mu}_{OLS} - \mu_0 = \hat{\mu}_{GLS} - \mu_0 + \frac{1}{T} l_T' \tilde{X} \), where \( \tilde{X} = (l_T - l_T l_T' \Omega_T^{-1} l_T)^{-1} l_T' \Omega_T^{-1} X^* \). The following facts which can be found in Sun et al.(2008) play an important role in the proof presented below: (1) \( \hat{\mu}_{GLS} - \mu_0 \) is independent of \( \tilde{X} \); (2) \( \hat{\mu}_{GLS} - \mu_0 \) is independent of \( X - l_T \hat{\mu}_{OLS} \). Notice that \( \hat{D}_T = \sum_{j=1}^{\infty} \lambda_j \hat{\xi}_j^2 = \frac{1}{T} (X - l_T \hat{\mu}_{OLS})' G_T (X - l_T \hat{\mu}_{OLS}) \) with \( G_T = (G(i/T, j/T))_{i,j=1} \). Then \( \hat{\mu}_{GLS} - \mu_0 \) is also independent of \( \hat{D}_T \). Define \( \sigma^2_{GLS} = T \text{Var}(\hat{\mu}_{GLS}) = T(l_T' \Omega_T^{-1} l_T)^{-1} \). Denote by \( \Phi_{norm} \) and \( \phi_{norm} \) the cumulative distribution function and
density function of the standard normal distribution. Therefore, we get

$$P(F_T(\infty) \leq x) = P\left(\frac{\sqrt{T}(\hat{\mu}_{OLS} - \mu_0)}{\sqrt{D_T}} \leq \sqrt{x}\right) - 1 = 2P\left(\frac{\sqrt{T}(\hat{\mu}_{OLS} - \mu_0)}{\sqrt{D_T}} \leq \sqrt{x}\right) - 1$$

$$= 2P\left(\frac{\sqrt{T}(\hat{\mu}_{GLS} - \mu_0)}{\sigma_{GLS}} \leq \sqrt{x}\right) - 1$$

$$= 2E\left[\Phi_{norm}\left(\sqrt{x}\right) - \sqrt{x}\right] - 1,$$

which implies that for $x, \delta \geq 0$ with $x - \delta \geq 0$,

$$\left|P(F_T(\infty) \leq x + \delta) - P(F_T(\infty) \leq x)\right|$$

$$= 2\left|E\left[\Phi_{norm}\left(\frac{\sqrt{x + \delta}D_T}{\sigma_{GLS}} - \sqrt{x}\right)\right] - E\left[\Phi_{norm}\left(\frac{\sqrt{x}D_T}{\sigma_{GLS}} - \sqrt{x}\right)\right]\right|$$

$$= 2\left|E\left[\Phi_{norm}\left(\frac{\sqrt{x + \delta}D_T}{\sigma_{GLS}} - \sqrt{x}\right)\right] - E\left[\Phi_{norm}\left(\frac{\sqrt{x}D_T}{\sigma_{GLS}} - \sqrt{x}\right)\right]\right|$$

$$\leq C\sqrt{\delta}E\left[\left|\sqrt{x + \delta}D_T\right|\right] < C\sqrt{\delta}(D_T)^{1/2}/\sigma_{GLS} < C\sqrt{\delta},$$

(4.29)

with $\sqrt{xD_T}/\sigma_{GLS} \leq a^* \leq \sqrt{(x + \delta)D_T}/\sigma_{GLS}$ or $\sqrt{(x - \delta)D_T}/\sigma_{GLS} \leq a^* \leq \sqrt{xD_T}/\sigma_{GLS}$. Here we are using the fact that $\sigma^2_{GLS} = \sigma^2 + O(1/T)$ and $E[D_T]$ is uniformly bounded for all $T$. Choosing $\delta = (\sum_{j=J+1}^{\infty} \lambda_j)^{2/3}$, the conclusion follows in view of (4.27), (4.28) and (4.29).

Lemma 4.10.12. Under the assumptions in Theorem 4.5.4, we have $||\Sigma_{\xi,J+1} - \sigma^2I_{J+1}||_\infty = O(J/T)$ with $J \leq T$, where $\Sigma_{\xi,J+1}$ denotes the covariance matrix of $(\xi_0, \xi_1, \ldots, \xi_J)$.

Proof of Lemma 4.10.12. Using the arguments in Lemma 4.10.5, we have for any $1 \leq s \leq J$,

$$|\text{cov}(\xi_0, \xi_s)| \leq C\left|\frac{1}{T^2} \sum_{i=1}^{T} \phi_s(i/T) + \frac{1}{T} \sum_{0 < h < T} \gamma_X(h) \left\{ \sum_{i=1}^{h} \phi_s(i/T) + \sum_{i=T-h+1}^{T} \phi_s(i/T) \right\} \right| \leq C/T,$$

where $C$ is a generic constant which does not depend on $s$. Again by the arguments in Lemma 4.10.5, we have

$$|\text{cov}(\xi_s, \xi_r) - \sigma^2\delta_{sr}| \leq \sum_{h=1}^{T-1} \left|\sum_{|h| \geq T} \gamma_X(h)C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t)) + \sum_{|h| \geq T} \gamma_X(h)\delta_{sr} \right|$$

$$+ \frac{1}{T} \sum_{h=1}^{T-1} \gamma_X(h) \left\{ \sum_{j=1}^{h} \phi_0^0 \left( \frac{j}{T} \right) \phi_0^0 \left( \frac{j}{T} \right) + \sum_{j=T-h+1}^{T} \phi_0^0 \left( \frac{j}{T} \right) \phi_0^0 \left( \frac{j}{T} \right) \right\}$$

$$+ |J_{1,T}| + |J_{2,T}|, \quad 1 \leq s, r \leq J,$$
where \(J_{1,T}, J_{2,T}\) and \(C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t))\) are defined in the proof of Lemma 4.10.5. By the Trapezoidal rule and the assumption that \(\sup_{1 \leq i \leq J} \sup_{t \in [0,1]} |\phi''_s(t)| < CJ^2\), we have

\[
|C_T(\tilde{\phi}_s(t), \tilde{\phi}_r(t))| \leq C(J^2/T^2 + 1/T),
\]

which implies that \(|\text{cov}(\xi_s, \xi_r) - \sigma^2 \delta_{sr}| \leq CJ/T + |J_{1,T}| + |J_{2,T}|\) for \(J \leq T\). By the mean value theorem and the assumption that \(\sup_{1 \leq i \leq J} \sup_{t \in [0,1]} |\phi'_s(t)| < CJ\), we get

\[
|J_{1,T}| \leq \frac{1}{T} \sum_{h=1}^{T-1} |\gamma_X(h)| \left| \sum_{j=1}^{T-h} |\phi_0^0(j/T)\{\phi_s(j + h/T) - \phi_s(j/T)\}| \right| \\
\leq \frac{CJ}{T} \sum_{h=1}^{T-1} |\gamma_X(h)| \left( \frac{1}{T} \sum_{j=1}^{T-h} |\phi_0^0(j/T)| \right) \leq \frac{CJ}{T}.
\]

Using the same argument for \(J_{2,T}\), we get \(|\text{cov}(\xi_s, \xi_r) - \sigma^2 \delta_{sr}| \leq CJ/T\), which completes the proof.

**Proof of Theorem 4.5.4.** Suppose \(J = o(T^{1/6})\). By Lemma 4.10.12, we know \(|\Sigma_{J+1} - \sigma I_{J+1}|_\infty = O(J/T)\). Using Lemma 4.10.8 and similar arguments in the proof of Lemma 4.10.6, we can show that

\[
\sup_{x \in \mathbb{R}} |P(F_T(J) \leq x) - Q_J(x) - R_T(x; J)| = o(1/T),
\]

where \(R_T(x; J) = \frac{1}{\sqrt{T}} \sum_{i=0}^{J} (\text{var}(\xi_i) - \sigma^2)E[(v_i^2 - 1)I\{F(v; J) \leq x\}]\) with \(v = (v_0, v_1, \ldots, v_J) \sim N(0, I_{J+1})\). Next, we show that \(R_T(x; J)\) converges uniformly as \(J \to +\infty\). Note first that

\[
\sup_{x \in [0, +\infty)} |R_T(x; J + p) - R_T(x; J)| \leq \sup_{x \in [0, +\infty)} \left| \frac{1}{2\sigma^2} \sum_{i=J+1}^{J+p} (\text{var}(\xi_i) - \sigma^2)E[(v_i^2 - 1)I\{F(v; J + p) \leq x\}] \right| \\
+ \sup_{x \in [0, +\infty)} \left| \frac{1}{2\sigma^2} \sum_{i=1}^{J} (\text{var}(\xi_i) - \sigma^2)E[(v_i^2 - 1)I\{F(v; J + p) \leq x\} - I\{F(v; J) \leq x\}] \right| \\
+ \sup_{x \in [0, +\infty)} \left| \frac{1}{2\sigma^2}(\text{var}(\xi_0) - \sigma^2)E[(v_0^2 - 1)I\{F(v; J + p) \leq x\} - I\{F(v; J) \leq x\}] \right| = I_1 + I_2 + I_3,
\]

for any \(J, p \in \mathbb{Z}^+\). In view of (4.30) and (4.31), we have

\[
|\text{var}(\xi_s) - \sigma^2| < C(i/T + i^2/T^2),
\]

where
for $1 \leq i < \infty$. Hence we get, for sufficiently large $J$,

$$I_1 \leq \frac{1}{2\sigma^2} \sup_{x \in [0, +\infty)} \sum_{i=J+1}^{J+p} \left( \left( \text{var}(\xi_i) - \sigma^2 \right) E \left[ (v_i^2 - 1) G_1 \left( x \sum_{j=1}^{J+p} \lambda_j v_j^2 \right) \right] \right)$$

$$\leq \frac{C}{T} \sup_{x \in [0, +\infty)} \sum_{i=J+1}^{J+p} \left( i + \frac{i^2}{T} \right) \left( \left( \text{var}(\xi_i) - \sigma^2 \right) E \left[ (v_i^2 - 1) G_1 \left( x \sum_{j=1}^{J+p} \lambda_j v_j^2 \right) \right] \right)$$

$$\leq \frac{C}{T} \sup_{x \in [0, +\infty)} \sum_{i=J+1}^{J+p} \left( i + \frac{i^2}{T} \right) \left( \left( v_i^2 - 1 \right) \left\{ G_1 \left( x \sum_{j \neq i} \lambda_j v_j^2 \right) + \lambda_i v_i^2 \right\} \right)$$

$$\leq \frac{C}{T} \sup_{x \in [0, +\infty)} \sum_{i=J+1}^{J+p} \left( i + \frac{i^2}{T} \right) \left( \left( v_i^2 - 1 \right) \left( \sum_{j \neq i} \lambda_j v_j^2 \right) \right) \leq \frac{C}{T} \sum_{i=J+1}^{J+p} \left( i + \frac{i^2}{T} \right) \lambda_i E \left( \left( \frac{\sum_{j \neq i} \lambda_j v_j^2 + \alpha \sum_{j \neq i} \lambda_j v_j^2}{\sum_{j \neq i} \lambda_j v_j^2} \right) \right)$$

$$\leq \frac{C}{T} \sum_{i=J+1}^{J+p} \left( i + \frac{i^2}{T} \right) \lambda_i \leq \frac{C}{T} \sum_{i=J+1}^{J+p} \left( i + \frac{i^2}{T} \right) \leq \frac{C}{T} \sum_{i=J+1}^{J+p} \left( i + \frac{i^2}{T} \right) = O \left( \frac{1}{J-a+2} \right).$$

(4.33)

where $y_i = x(\sum_{j \neq i} \lambda_j v_j^2 + \alpha \lambda_i v_i^2)$ for some $0 \leq \alpha_i \leq 1$. On the other hand, we get

$$I_2 \leq \frac{CJ}{T} \sup_{x \in [0, +\infty)} \sum_{i=1}^{J} \left( \left( \text{var}(\xi_i) - \sigma^2 \right) E \left[ (v_i^2 - 1) G_1 \left( x \sum_{j=1}^{J+p} \lambda_j v_j^2 \right) - G_1 \left( x \sum_{j=1}^{J} \lambda_j v_j^2 \right) \right] \right)$$

$$\leq \frac{CJ}{T} \sup_{x \in [0, +\infty)} \sum_{i=1}^{J} \left( \left( v_i^2 - 1 \right) \left( \sum_{j=1}^{J+p} \lambda_j v_j^2 \right) \right) \leq \frac{CJ^2}{T} \left( \sum_{j=J+1}^{J+p} \lambda_j \right) = O \left( \frac{1}{J-a+3} \right).$$

Finally using the Cauchy-Schwarz inequality and similar arguments in Lemma 4.10.9, we know

$$I_3 \leq \frac{C}{T} \left( E[(v_0^2 - 1)^2] \right)^{1/2} \sup_{x \in [0, +\infty)} \left\{ E[\{I(\mathcal{F}(v; J + p) \leq x\} - I(\mathcal{F}(v; J) \leq x)\}^2] \right\}^{1/2}$$

$$\leq \frac{C}{T} \sup_{x \in [0, +\infty)} \left\{ E[\{I(\mathcal{F}(v; J + p) \leq x\} - I(\mathcal{F}(v; J) \leq x)\}] \right\}^{1/2} \leq \frac{C}{T} \left( \sum_{j=J}^{+\infty} \lambda_j \right)^{1/2} = O \left( \frac{J^{(a-1)/2}}{T} \right).$$

Therefore, it is straightforward to see that $\sup_{x \in [0, +\infty]} |\mathcal{R}_T(x; J) - \mathcal{R}_T(x; \infty)| = O \left( \frac{J^{(a-1)/2}}{T} \right)$ and
By the Taylor expansion, we have

$$\sup_{x \in [0, \infty)} |\mathcal{N}_T(x; \infty)| = O(1/T),$$

which imply that

$$\sup_{x \in [0, \infty)} |P(F_T(J) \leq x) - Q_J(x) - \mathcal{N}_T(x; \infty)| = o(1/T),$$

for $J = o(T^{1/6})$. Let $J = T^{1/6}/\log(T)$ and note that $(\sum_{j=j+1}^{\infty} \lambda_j)^{1/3} = o(1/T)$. The proof is completed in view of Lemma 4.10.9 and Lemma 4.10.11.

Proof of Proposition 4.5.5. Under the assumption that $\sup_{x \in \mathbb{R}} |\mathcal{K}(x)| \leq 1$ and $\int_{-\infty}^{+\infty} |\mathcal{K}(x)| dx < \infty$, we have

$$\sum_{j=1}^{+\infty} (\tilde{\lambda}_{j,b})^2 = \int_0^1 \int_0^1 \mathcal{G}_b^2(r, t) dr dt \leq \sup_{t \in [0,1]} \int_0^1 \mathcal{G}_b^2(r, t) dr \leq 4 \sup_{t \in [0,1]} \int_0^1 |\mathcal{G}_b(r, t)| dr$$

$$\leq 16 \sup_{t \in [0,1]} \int_{-t}^{1-t} |\mathcal{K}_b(r)| dr \leq 16 \int_{-\infty}^{+\infty} |\mathcal{K}_b(r)| dr \leq Cb,$$

and $\tilde{\lambda}_{1,b} \leq (\int_0^1 \int_0^1 \mathcal{G}_b^2(r, t) dr dt)^{1/2} \leq C\sqrt{b}$. Suppose $\{\tilde{a}_i\}$ is a sequence of random variables such that $0 \leq \tilde{a}_i \leq 1$. Using the fact that $\sum_{j=1}^{+\infty} \tilde{\lambda}_{j,b} = \int_0^1 \mathcal{G}_b(r, r) dr = 1 + O(b)$, we get

$$\sup_i E \left( \sum_{j \neq i}^{+\infty} (\tilde{\lambda}_{j,b})^2 + \tilde{a}_i \tilde{\lambda}_{i,b} v_i^2 - 1 \right)^2 \leq \sup_i \left\{ \sum_{j \neq i}^{+\infty} (\tilde{\lambda}_{j,b})^2 + (\tilde{\lambda}_{i,b})^2 E(\tilde{a}_i v_i^2 - 1)^2 \right\} + O(b) \leq Cb.$$

(4.34)

By the Taylor expansion, we have

$$\mathcal{N}_{T,b}(x; \infty) = \frac{1}{2\sigma^2} \sum_{i=1}^{+\infty} (\text{var}(\tilde{\xi}_{i,b}) - \sigma^2) E[|v_i^2 - 1| I(|v_i; \infty) \leq x)] + O(1/T)$$

$$= \frac{1}{2\sigma^2} \sum_{i=1}^{+\infty} (\text{var}(\tilde{\xi}_{i,b}) - \sigma^2) E \left[ (v_i^2 - 1) G_1 \left( x \sum_{j=1}^{+\infty} \tilde{\lambda}_{j,b} v_j^2 \right) \right] + O(1/T)$$

$$= \frac{x^2}{\sigma^2} \sum_{i=1}^{+\infty} (\text{var}(\tilde{\xi}_{i,b}) - \sigma^2) E \left[ G_1' \left( x \sum_{j \neq i}^{+\infty} \tilde{\lambda}_{j,b} v_j^2 \right) \right]$$

$$+ \frac{x^2}{4\sigma^2} \sum_{i=1}^{+\infty} (\text{var}(\tilde{\xi}_{i,b}) - \sigma^2) E \left[ v_i^4 (v_i^2 - 1) G_1'' \left( x \sum_{j \neq i}^{+\infty} \tilde{\lambda}_{j,b} v_j^2 + a_i \tilde{\lambda}_{i,b} v_i^2 \right) \right] + O(1/T)$$

$$= I_{1T,b} + I_{2T,b} + O(1/T),$$

where $0 \leq a_i \leq 1$. Let $A_{i,b} = E \left[ G_1' \left( x \sum_{j \neq i}^{+\infty} \tilde{\lambda}_{j,b} v_j^2 \right) \right]$, $B_{i,b} = \tilde{\lambda}_{i,b} (\text{var}(\tilde{\xi}_{i,b}) - \sigma^2)$, $C_{i,b} = \sum_{j=1}^{i} B_{j,b}$ and $S_{N,b} = \sum_{i=1}^{N} A_{i,b} B_{i,b}$. Using summation by parts, we have $S_{N,b} = A_{N,b} C_{N,b} - \sum_{i=1}^{N-1} (A_{i+1,b} -$
Note that \( \{A_{i,b}\}_{i=1}^{\infty} \) is a nonincreasing sequence and \( \lim_{b \to 0} \sup_i A_{i,b} = G_1'(x) \) as seen from (4.34). Let \( \hat{D}_{T,b} \) be defined by replacing \( \phi_j \) and \( \lambda_j \) with \( \tilde{\phi}_{j,b} \) and \( \tilde{\lambda}_{j,b} \) in the definition of \( \hat{D}_T \). It is not hard to see that as \( b + 1/(bT) \to 0 \),

\[
\lim_{N \to +\infty} A_{N,b} C_{N,b} = \sigma^2 G'(x) \left( \frac{E[\hat{D}_{T,b}]/\sigma^2 - \sum_{j=1}^{+\infty} \tilde{\lambda}_{j,b}}{\hat{D}(bT)^q} \right) (1 + o(1))
= - \frac{G'(x)g_q \sum_{j=-\infty}^{+\infty} |h|^q \gamma X(h)}{(bT)^q} (1 + o(1)) + O(1/T),
\]

where we have used the fact \( E[\hat{D}_{T,b}]/\sigma^2 - \sum_{j=1}^{+\infty} \tilde{\lambda}_{j,b} = - \frac{g_q \sum_{h=-\infty}^{+\infty} |h|^q \gamma X(h)}{(bT)^q} (1 + o(1)) + O(1/T) \), which can be proved by using similar arguments in the proof of Lemma 2 in Sun et al. (2008).

On the other hand, observe that | \( \sum_{i=1}^{N-1} (A_{i+1,b} - A_{i,b}) C_{i,b} \) | \( \leq \sup_{i \in N} C_{i,b} | \sum_{i=1}^{N-1} (A_{i,b} - A_{i+1,b}) \) \( \leq \sup_{i \in N} C_{i,b} | (A_{1,b} - \lim_{N \to +\infty} A_{N,b}) | = o(\lim_{N \to +\infty} C_{N,b}) \) as \( b + 1/(bT) \to 0 \), for all \( N \). Hence we get

\[
I_{1T,b} = - \frac{x G'(x) g_q \sum_{h=-\infty}^{+\infty} |h|^q \gamma X(h)}{(bT)^q} (1 + o(1)) + O(1/T).
\]

Define \( H_{i,b} = \hat{\lambda}_{i,b} E \left[ v_i^t (v_i^2 - 1) G_1'(x) \left( \sum_{j \neq i} \hat{\lambda}_{j,b} v_j^2 + a_i \hat{\lambda}_{i,b} v_i^2 \right) \right] \) and \( \tilde{S}_{N,b} = \sum_{i=1}^{+\infty} H_{i,b} B_{i,b} \). Again using summation by parts, we obtain \( \tilde{S}_{N,b} = H_{N,b} C_{N,b} - \sum_{i=1}^{N-1} (H_{i+1,b} - H_{i,b}) C_{i,b} \). By (4.34), we can show that \( \sup_i |H_{i,b}/\hat{\lambda}_{i,b} - 12G_1''(x)| = O(\sqrt{b}) \). Therefore, we get \( \lim_{N \to +\infty} C_{N,b} H_{N,b} = o(\lim_{N \to +\infty} C_{N,b}) \) and

\[
\left| \sum_{i=1}^{N-1} (H_{i+1,b} - H_{i,b}) C_{i,b} \right| \leq \sup_{i \in N} C_{i,b} \left\{ \sum_{i=1}^{N-1} (|H_{i+1,b} - 12\hat{\lambda}_{i+1,b} G_1''(x)| + 12G_1''(x) |\hat{\lambda}_{i,b} - \hat{\lambda}_{i+1,b}|)
+ |12\hat{\lambda}_{i,b} G_1''(x) - H_{i,b}| \right\} = O \left( \sqrt{b} \lim_{N \to +\infty} C_{N,b} \right).
\]

The conclusion follows from the above arguments by noting that \( I_{2T,b} = o(I_{1T,b}) \). \( \square \)

### 4.10.2 Proof of the main results in Section 4.6

**Lemma 4.10.13.** Let \( \omega_l(x) = (1 - |x/l|) I\{|x/l| < 1\} \). Suppose that \( m^3/l^2 + (ml)^3/T + 1/m \to 0 \) and \( \sum_{h=-\infty}^{+\infty} h^2 |\gamma X(h)| < \infty \). Then under the Gaussian assumption, we have

\[
\sup_{0 \leq k \leq m} \left| \sum_{h=1-l}^{l-1} g_{k,T}(h) \omega_l(h) \gamma X(h) - \sum_{h=1-T}^{T-1} g_{k,T}(h) \gamma X(h) \right| = O_p(\sqrt{m^3/l^2 + (ml)^3/T}),
\]

117
where \(|g_k,h| \leq C(k|h| + |h| + 1)| \) for \(0 \leq k \leq m\) and \(|h| \leq T\), and the constant \(C\) does not depend on \(k\) and \(h\).

Proof of Lemma 4.10.13. Note first that for any \(\epsilon > 0\),

\[
P \left( \sup_{0 \leq k \leq m} \left| \sum_{h=1}^{l-1} g_k,h(\omega_i)\gamma_X(h) - \sum_{h=1}^{T-1} g_k,T(\gamma_X(h) \right| > \epsilon \right) \\
\leq \frac{m}{2} \left( \sum_{h=1}^{l-1} g_k,h(\omega_i)\gamma_X(h) - \sum_{h=1}^{T-1} g_k,T(\gamma_X(h) \right| > \epsilon \right)^2 \\
\leq \frac{1}{\epsilon^2} \sum_{k=0}^{m} \left( \sum_{h=1}^{l-1} g_k,h(\omega_i)\gamma_X(h) - \sum_{h=1}^{T-1} g_k,T(\gamma_X(h) \right| + \frac{Cm^3}{\epsilon^2}.
\]

Let \(z_i = X_i - E[X_i]\) and \(w_i = z_i z_i - \gamma_X(h)\). Simple calculation yields that

\[
\left| \sum_{h=1}^{l-1} g_k,h(\omega_i)\gamma_X(h) - \gamma_X(h) \right| = \left| \sum_{h=1}^{l-1} g_k,h(\omega_i)\gamma_X(h) - \gamma_X(h) \right| + C(k + 1)/l
\]

\[
\leq \left| \sum_{h=1}^{l-1} g_k,h(\omega_i) \left\{ \frac{1}{T} \sum_{i=1}^{T-|h|} w_i \right\} \right| + \left| \sum_{h=1}^{l-1} g_k,T(\omega_i) \left\{ \frac{T-|h|}{T} \right\} \right| + C(k + 1)/l := I_{1T} + I_{2T} + I_{3T} + C(k + 1)/l,
\]

which implies that

\[
E \left( \sum_{h=1}^{l-1} g_k,h(\omega_i)\gamma_X(h) - \gamma_X(h) \right|^2 \leq C(EI_{1T}^2 + EI_{2T}^2 + EI_{3T}^2 + (k + 1)^2/\epsilon^2).
\]

We proceed to derive the order of \(EI_{1T}^2\). Notice that

\[
EI_{1T}^2 = \frac{1}{T^2} \sum_{h_1,h_2=1-l}^{l-1} \sum_{i_1=1}^{T-|h_1|} \sum_{i_2=1}^{T-|h_2|} \text{cov}(w_{i_1|h_1}, w_{i_2|h_2}) g_k,h(\gamma_X(h_1)) g_k,T(\gamma_X(h_2)) \omega_i(\gamma_X(h_2))
\]

\[
\leq \frac{C(k + 1)^2}{T^2} \sum_{h_1,h_2=1-l}^{l-1} \sum_{i_1=1}^{T-|h_1|} \sum_{i_2=1}^{T-|h_2|} \left( |h_1| + 1 |h_2| + 1 \right) |\gamma_X(i_1 - i_2)\gamma_X(i_1 - i_2 + |h_1| - |h_2|) |
\]

\[
+ \frac{C(k + 1)^2}{T^2} \sum_{h_1,h_2=1-l}^{l-1} \sum_{i_1=1}^{T-|h_1|} \sum_{i_2=1}^{T-|h_2|} \left( |h_1| + 1 |h_2| + 1 \right) (|i_1 - i_2 - |h_2|) |\gamma_X(i_1 - i_2 + |h_1|) |
\]

\[
\leq \frac{C(k + 1)^2}{T^2} \sum_{h_1,h_2=1-l}^{l-1} \sum_{s=1-T}^{T-1} (T-s) |(h_1| + 1 |h_2| + 1) |\gamma_X(s)\gamma_X(s + |h_1| - |h_2|) |
\]

\[
+ \frac{C(k + 1)^2}{T^2} \sum_{h_1,h_2=1-l}^{l-1} \sum_{s=1-T}^{T-1} (T-s) |(h_1| + 1 |h_2| + 1) |\gamma_X(s - |h_2|)\gamma_X(s + |h_1|) |
\]

\[= J_{1T} + J_{2T}.\]
Then we get

\[
J_{1,T} \leq \frac{C(k + 1)^2}{T} \sum_{h_1, h_2 = 1-l}^{l-1} \frac{C(k + 1)^2}{T} \sum_{|s| = -\infty}^{+\infty} |\gamma_X(s)| \sum_{|s + h_1| - |h_2|}^{l-1} |\gamma_X(s + |h_1| - |h_2|)|
\]

and

\[
J_{2,T} \leq \frac{C(k + 1)^2}{T} \sum_{|s| = -\infty}^{+\infty} |\gamma_X(s)| \sum_{|s + h_1| - |h_2|}^{l-1} \frac{C(k + 1)^2}{T} \sum_{h_1, h_2 = 1-l}^{l-1} |\gamma_X(s + |h_1| - |h_2|)|
\]

It implies that \( E_{1T}^2 \leq \frac{C(k + 1)^2}{T} \). Applying similar arguments to \( I_{2T} \) and \( I_{3T} \), we get \( E_{2T}^2 \leq C(k + 1)^2 l^4 T^2 \) and \( E_{3T}^2 \leq C(k + 1)^2 l^4 T^2 \). Note the constant \( C \) above does not depend on \( m \) by the assumption. We then have

\[
P \left( \sup_{0 \leq k \leq m} \left| \sum_{h = 1-l}^{l-1} g_{k,T}(h) \omega_1(h) \gamma_X(h) - \sum_{h = 1-T}^{T-1} g_{k,T}(h) \gamma_X(h) \right| > \epsilon \right) \leq \frac{C}{\epsilon^2} (m^3 l^2 + (ml)^3 / T) \to 0.
\]

\[\square\]

**Proof of Theorem 4.6.2.** We choose \( m \) so that \( m^3 l^2 + (ml)^3 / T + 1/m \to 0 \) (e.g., \( l \asymp T^{1/5} \) and \( m \asymp T^{2/15 - \epsilon} \) for some \( \epsilon > 0 \)). From equation (4.24) in Lemma 4.10.5, we know that

\[
\text{var}(\xi_i) - \sigma^2 - (\text{var}(\xi_i^*) - \hat{\sigma}^2) = \frac{1}{T} \left\{ \sum_{h = 1-T}^{T-1} g_{i,T}(h) \gamma_X(h) - \sum_{h = 1-l}^{l-1} g_i(h) \omega_1(h) \gamma_X(h) \right\} - \sum_{|h| \geq T} \gamma_X(h)
\]
where $\hat{\sigma}^2 = \sum_{h=1}^{l-1} \omega_l(h) \hat{\gamma}_X(h)$ and $g_{0,T}(h) = -|h|$, 

$$g_{i,T}(h) = T C_T(\phi_i(s), \phi_i(t)) - \left[ \sum_{j=1}^{h} \left\{ \phi_i^0 \left( \frac{j}{T} \right) \right\}^2 + \sum_{j=T+1}^{T+h-1} \left\{ \phi_i^0 \left( \frac{j}{T} \right) \right\}^2 \right] I\{h \geq 1\}$$

$$+ \left[ \sum_{j=1}^{T-h} \phi_i^0 \left( \frac{j}{T} \right) \left\{ \phi_i^0 \left( \frac{j+h}{T} \right) - \phi_i^0 \left( \frac{j}{T} \right) \right\} \right] I\{h \geq 1\}$$

$$+ \left[ \sum_{j=T+1}^{T} \phi_i^0 \left( \frac{j}{T} \right) \left\{ \phi_i^0 \left( \frac{j+h}{T} \right) - \phi_i^0 \left( \frac{j}{T} \right) \right\} \right] I\{h \leq -1\},$$

for $1 \leq i \leq m$. Note that $\sup_{1 \leq i \leq m} |T C_T(\phi_i(s), \phi_i(t))| \leq C$. It is not hard to see that $|g_{i,T}(h)| \leq C(|ih| + |h| + 1)$ for $0 \leq i \leq m$. By Lemma 4.10.13, we know

$$\sup_{0 \leq i \leq m} |\text{var}(\xi_i) - \sigma^2 - \text{var}^*(\xi_i^*) + \hat{\sigma}^2| = O_p\left(\frac{\sqrt{m^3/l^2} + (m!/l)}{T}\right).$$

Since the bootstrap sample is normally distributed and $\sum_{h=1}^{l-1} h^2 \omega_l(h)|\hat{\gamma}_X(h)|$ is bounded in probability in view of the fact that $\sum_{h=-\infty}^{\infty} h^2 \omega_l(h)E|\hat{\gamma}_X(h)| < \infty$, Theorem 4.5.4 is also applicable to the bootstrap sample, i.e.,

$$\sup_{x \in [0, \infty)} |P(F_T^*(\infty) \leq x) - Q_\infty(x) - X_T^*(x; \infty)| = o_p(1/T),$$

where $X_T^*(x; \infty) = \frac{1}{2\pi} \sum_{i=0}^{\infty} (\text{var}^*(\xi_i^*) - \hat{\sigma}^2)E[(v_i^2 - 1)I\{F(v; \infty) \leq x\}]$. It is not hard to show that $\hat{\sigma}^2 - \sigma^2 = O_p(\sqrt{1/T + 1/l^2})$. Note that $\text{var}^*(\xi_i^*) - \hat{\sigma}^2 = \frac{1}{T} \sum_{h=1}^{l-1} g_{i,T}(h) \omega_l(h) \hat{\gamma}_X(h)$, which implies that $\sup_{1 \leq i \leq m} \frac{|\text{var}^*(\xi_i^*) - \hat{\sigma}^2|}{\sqrt{1/T + 1/l^2}} = O_p(1)$ (see e.g., 4.32). Using the arguments in (4.33), we can show that

$$\sup_{x \in [0, \infty)} \left| \frac{1}{2\sigma^2} \sum_{i=m+1}^{\infty} (\text{var}(\xi_i) - \text{var}^*(\xi_i^*) + \hat{\sigma}^2 - \sigma^2)E[(v_i^2 - 1)I\{F(v; \infty) \leq x\}] \right| = O_p\left(\frac{1}{T^{m^{a-2}}}\right).$$
Thus we get

$$
\sup_{x \in [0, +\infty)} |N_T(x; \infty) - N_T^*(x; \infty)| \leq \sup_{x \in [0, +\infty)} \left| \frac{1}{2\sigma^2} \sum_{i=0}^{\infty} (\text{var}(\xi_i) - \text{var}^*(\xi_i^*) + \hat{\sigma}^2 - \sigma^2) E[(v_i^2 - 1)I\{F(v; \infty) \leq x\}] \right|
$$

$$
+ \sup_{x \in [0, +\infty)} \left| \frac{1}{2\sigma^2} \sum_{i=1}^{\infty} (\text{var}^*(\xi_i^*) - \hat{\sigma}^2) E[(v_i^2 - 1)I\{F(v; \infty) \leq x\}] \right|
$$

$$
\leq \frac{1}{2\sigma^2} \sup_{1 \leq i \leq m} |\text{var}(\xi_i) - \sigma^2 - \text{var}^*(\xi_i^*) + \hat{\sigma}^2| \sup_{x \in [0, +\infty)} \left| \sum_{i=1}^{m} E[(v_i^2 - 1)I\{F(v; \infty) \leq x\}] \right| + O_p \left( \frac{\sqrt{l/T + 1/l^2}}{T} \right)
$$

$$
= O_p \left( \frac{\sqrt{m^2/l^2 + (ml)^{1/2}}}{T} \right) + O_p \left( \frac{\sqrt{l/T + 1/l^2}}{T} \right) + O_p \left( \frac{1}{T m^{a-2}} \right).
$$

It then follows that

$$
\sup_{x \in [0, +\infty)} |P(F_T(\infty) \leq x) - P(F_T^*(\infty) \leq x)| \leq \sup_{x \in [0, +\infty)} |N_T(x; \infty) - N_T^*(x; \infty)| + o_p(1/T) = o_p(1/T).
$$

**Proof of Theorem 4.6.1.** The proof is similar to those of Lemma 4.10.13 and Theorem 4.6.2. The details are omitted.
Figure 4.2: Empirical rejection probabilities for the Wald statistic with the Bartlett kernel (left panel) and QS kernel (right panel) and for the AR(1) model with $N(0,1)$ innovations
Figure 4.3: Empirical rejection probabilities for the Wald statistic with the Bartlett kernel (left panel) and QS kernel (right panel) and for the AR(1) model with t(3) innovations
Figure 4.4: Empirical rejection probabilities for the Wald statistic with the Bartlett kernel (left panel) and QS kernel (right panel) and for the AR(1) model with exp(1) – 1 innovations
Chapter 5

Fixed-$b$ asymptotics for blockwise empirical likelihood

5.1 Introduction

Empirical likelihood (EL) (Owen, 1988, 1990) is a nonparametric technique for conducting inference for parameters in nonparametric settings. EL has been studied extensively in the statistics and econometrics literature (see Owen, 2001, Kitamura, 2006, and Chen and Van Keilegom, 2009, for comprehensive reviews). One striking property of EL is the nonparametric version of Wilks’ theorem which states that the EL ratio statistic evaluated at the true parameter converges to a $\chi^2$ limiting distribution. This property was first demonstrated for the mean parameter by Owen (1990) and was further extended to the estimating equation framework by Qin and Lawless (1994). However, Wilks’ phenomenon fails to hold for stationary time series because the dependence within the observations is not taken into account in EL. Kitamura (1997) proposed the blockwise empirical likelihood (BEL) which is able to accommodate the dependence of the data, and Wilks’ theorem continues to hold for the BEL ratio statistic under suitable weak dependence assumptions. The BEL can be viewed as a special case of the generalized empirical likelihood (GEL) with smoothed moment conditions (Smith, 2011).

The performance of BEL and its variations (Nordman, 2009; Smith, 2011) can depend crucially on the choice of the bandwidth parameter for which no sound guidance seems provided in the literature. Recently, Kiefer and Vogelsang (2005) proposed the so-called fixed-$b$ asymptotic theory in the heteroscedasticity-autocorrelation robust (HAR) testing context. It was found that the asymptotic distribution obtained by treating the bandwidth as a fixed proportion (say $b$) of the sample size, provides a better approximation to the sampling distribution of the studentized test statistic than the traditional $\chi^2$-based approximation. See Jansson (2004), Sun, Phillips and Jin (2008), and Zhang and Shao (2013) for rigorous theoretical justifications. The fixed-$b$ approach has the advantage of accounting for the effect of the bandwidth and the kernel, as different bandwidth parameters and kernels correspond to different limiting (null) distributions (also see Shao and Politis, 2013,
for a recent extension to the subsampling and block bootstrap context). The main thrust of the present chapter is the development of a new asymptotic theory in the BEL estimation framework by adopting the fixed-\(b\) approach. We consider the problem in the moment condition model (Qin and Lawless, 1994; Smith, 2011) which is a fairly general framework used by both statisticians and econometricians. Under the fixed-\(b\) asymptotic framework, we show that the asymptotic null distribution of the EL ratio statistic evaluated at the true parameter is nonstandard yet pivotal and it can be approximated numerically. It is interesting to note that the fixed-\(b\) limiting distribution coincides with the \(\chi^2\) distribution as \(b\) gets close to zero. We also illustrate the idea in the GEL estimation framework and demonstrate the usefulness of the fixed-\(b\) approach through simulation studies.

We use the following notation throughout this chapter. Let \(D[0,1]\) be the space of functions on \([0,1]\) which are right-continuous and have left limits, endowed with the Skorokhod topology (see Billingsley, 1999). Weak convergence in \(D[0,1]\) or more generally in the \(\mathbb{R}^m\)-valued function space \(D^m[0,1]\) is denoted by “\(\Rightarrow\)”, where \(m \in \mathbb{N}\). Convergence in probability and convergence in distribution are denoted by “\(\rightarrow_p\)” and “\(\rightarrow_d\)” respectively. Let \(C[0,1]\) be the space of continuous functions on \([0,1]\). Denoted by \(\lfloor a \rfloor\) the integer part of \(a \in \mathbb{R}\).

### 5.2 Methodology

#### 5.2.1 Empirical likelihood

Suppose we are interested in the inference of a \(p\)-dimensional parameter vector \(\theta\), which is identified by a set of moment conditions. Denote by \(\theta_0\) the true parameter of \(\theta\) which is an interior point of a compact parameter space \(\Theta \subseteq \mathbb{R}^p\). Let \(\{y_t\}_{t=1}^n\) be a sequence of \(\mathbb{R}^l\)-valued stationary time series and assume the moment conditions

\[
E[f(y_t, \theta_0)] = 0, \quad t = 1, 2, \ldots, n, \tag{5.1}
\]

hold, where \(f(y, \theta) : \mathbb{R}^{l+p} \to \mathbb{R}^k\) is a map which is differentiable with respect to \(\theta\) and \(\text{rank}(E[\partial f(y_t, \theta_0)/\partial \theta']) = p\) with \(k \geq p\). To deal with time series data, we consider the smoothed moment conditions introduced by Smith (2011),

\[
\hat{f}_{tn}(\theta) = \frac{1}{S_n} \sum_{s=t-n}^{t-1} K\left(\frac{s}{S_n}\right) f(y_{t-s}, \theta), \tag{5.2}
\]
where $K(\cdot)$ is a kernel function and $S_n = bn$ with $b \in (0, 1)$ is the bandwidth parameter. Note that smoothing of the moment conditions induces a heteroskedasticity and autocorrelation consistent (HAC) covariance estimator of the long run variance matrix of $\{f(y_t, \theta)\}_{t=1}^n$. To proceed, we introduce some notation. Let $f_t(\theta) = f(y_t, \theta)$ and $\tilde{f}_n(\theta) = \sum_{t=1}^n f_{tn}(\theta)/n$, where $f_{tn}(\theta)$ is defined in (5.2).

Consider the profile empirical log-likelihood function based on the smoothed moment restrictions,

$$L_n(\theta) = \sup \left\{ \sum_{t=1}^n \log(p_t) : p_t \geq 0, \sum_{t=1}^n p_t = 1, \sum_{t=1}^n p_t f_{tn}(\theta) = 0 \right\}. \tag{5.3}$$

Standard Lagrange multiplier arguments imply that the maximum is attained when

$$p_t = \frac{1}{n \{1 + \lambda' f_{tn}(\theta)\}}, \quad \text{with} \quad \sum_{t=1}^n \frac{f_{tn}(\theta)}{1 + \lambda' f_{tn}(\theta)} = 0.$$

The maximum empirical likelihood estimate (MELE) is then given by $\hat{\theta}_{el} = \arg\max_{\theta \in \Theta} L_n(\theta)$. Following Kitamura (2006), the empirical log-likelihood function can also be derived by considering the dual problem (see e.g., Borwein and Lewis, 1991),

$$L_n(\theta) = \min_{\lambda \in \mathbb{R}^k} - \sum_{t=1}^n \log(1 + \lambda' f_{tn}(\theta)) - n \log n, \tag{5.4}$$

where $\log(x) = -\infty$ for $x < 0$. Expression (5.4) has a natural connection with the generalized empirical likelihood (GEL) and it also facilitates our theoretical derivation under the fixed-$b$ asymptotics.

To introduce the fixed-$b$ approach, we define the empirical log-likelihood ratio function

$$elr(\theta) = 2 \max_{\lambda \in \mathbb{R}^k} \sum_{t=1}^n \log(1 + \lambda' f_{tn}(\theta))/S_n, \tag{5.5}$$

for $\theta \in \Theta$ and $S_n = bn$. Under the traditional small-$b$ asymptotics, i.e., $b + 1/(nb) \to 0$ as $n \to \infty$, and suitable weak dependence assumptions (see e.g., Smith, 2011), it can be shown that

$$elr(\theta_0) = n \tilde{f}_n(\theta_0)' \left( b \sum_{t=1}^n f_{tn}(\theta_0) f_{tn}(\theta_0)' \right)^{-1} \tilde{f}_n(\theta_0) + o_p(1) \to^d \frac{\kappa_1^2}{\kappa_2} \chi^2_k,$$

where $\kappa_1 = \int_{-\infty}^{+\infty} K(x) dx$ and $\kappa_2 = \int_{-\infty}^{+\infty} K^2(x) dx$ (assuming that $\kappa_1, \kappa_2 < \infty$). However, the $\chi^2$-based approximation can be poor especially when the dependence is strong and the bandwidth parameter is large. To derive the fixed-$b$ limiting distribution, we make the following high level
assumption which is standard in the moment condition models.

**Assumption 5.2.1.** Assume that \( \sum_{s=1}^{|nr|} f_i(\theta_0)/\sqrt{n} \Rightarrow \Lambda W_k(r) \) for \( r \in [0,1] \), where \( \Lambda \Lambda' = \Omega = \sum_{j=-\infty}^{+\infty} \Gamma_j \) with \( \Gamma_j = \mathbb{E} f_{i+j}(\theta_0)f_i(\theta_0)' \) and \( W_k(r) \) is a \( k \)-dimensional vector of independent standard Brownian motions.

Assumption 5.2.1 can be verified under suitable moment and weak dependence assumptions on \( f_i(\theta_0) \) (see e.g., Phillips, 1987). For the kernel function, we shall assume the following.

**Assumption 5.2.2.** The kernel \( K : \mathbb{R} \rightarrow [-c_0,c_0] \) for some \( 0 < c_0 < \infty \), is piecewise continuously differentiable.

Next, we fix \( b \in (0,1) \), where \( b = S_n/n \). Using summation by parts, continuous mapping theorem and It\'s formula, it is not hard to show that for \( t = \lfloor nr \rfloor \) with \( r \in [0,1] \),

\[
\sqrt{n} f_{in}(\theta_0) = \frac{\sqrt{n}}{S_n} \sum_{s=t-n}^{t-1} K(s/S_n) f_{i-s}(\theta_0) \Rightarrow \Lambda D_k(r;b)/b,
\]

where \( D_k(r;b) = \int_0^1 K((r-s)/b) dW_k(s) \). Let \( C^{\otimes k}[0,1] = \{(f_1, f_2, \ldots, f_k) : f_i \in C[0,1] \} \). For any \( g \in C^{\otimes k}[0,1] \), define the nonlinear functional \( G_{cl}(g) = \max_{\lambda \in \mathbb{R}^k} \int_0^1 \log(1 + \lambda' g(t)) dt \). We show in the appendix that the functional \( G_{cl}(\cdot) \) is continuous under the sup norm. Therefore, by the continuous mapping theorem, we obtain the following result which characterizes the asymptotic behavior of \( elr(\theta_0) \).

**Theorem 5.2.1.** Suppose Assumptions 5.2.1-5.2.2 hold. As \( n \to +\infty \) and \( b \) is held fixed, we have

\[
elr(\theta_0) \rightarrow^d U_{el,k}(b;K) := \frac{2}{b} \max_{\lambda \in \mathbb{R}^k} \int_0^1 \log \left( 1 + \lambda' \int_0^1 K((r-s)/b) dW_k(s) \right) dr.
\]

Theorem 5.2.1 shows that the fixed-\( b \) limiting distribution of \( elr(\theta_0) \) is nonstandard yet pivotal for a given bandwidth and kernel, and its critical values can be obtained via simulation or iid bootstrap (because the bootstrapped sample satisfies the functional central limit theorem). Let \( u_{el,k}(b;K;1-\alpha) \) be the \( 100(1-\alpha)\% \) quantile of \( U_{el,k}(b;K)/(1-b) \). Given \( b \in (0,1) \), a \( 100(1-\alpha)\% \) confidence region for the parameter \( \theta_0 \) is then given by

\[
CI(1-\alpha; b) = \left\{ \theta \in \mathbb{R}^p : \frac{elr(\theta)}{1-b} \leq u_{el,k}(b;K;1-\alpha) \right\}.
\]

When \( K(x) = I(x \geq 0) \), we have \( D_k(r;b) = W_k(r) \) and \( A := \{ \lambda \in \mathbb{R}^k : \min_{r \in [0,1]} (1 + \lambda' D_k(r;b)) \geq 0 \} = \{ \lambda \in \mathbb{R}^k : \min_{r \in [0,1]} (1 + \lambda W_k(r)) \geq 0 \} \). By Lemma 1 of Nordman et al. (2012), we know
That the set \( A \) is bounded with probability one, which implies that \( P(U_{c.t.k}(b;K) = \infty) = 0 \). We conjecture that the probability \( P(U_{c.t.k}(b;K) = \infty) \) can be positive for particular \( K(\cdot) \) and \( b \in (0, 1) \).

In our simulation, the critical values are calculated based on the cases where \( U_{c.t.k}(b;K) < \infty \) (when \( b \) is close to zero, the probability \( P(U_{c.t.k}(b;K) = \infty) \) is rather small as seen from our unreported simulation results). In fact, the nonstandard limiting distribution also provides some insights on how likely the origin is not contained in the convex hull of \( \{f_{tn}(\theta_0)\}_{t=1}^n \), when the sample size \( n \) is large.

**Remark 5.2.1.** To capture the dependence within the observations, one may employ the commonly used blocking technique which was first applied to the EL by Kitamura (1997). To illustrate the idea, we consider the fully overlapping smoothed moment condition which is given by \( f_{tn}(\theta) = \frac{1}{m} \sum_{j=t}^{t+m-1} f(y_t, \theta) \) with \( t = 1, 2, \ldots, n - m + 1 \) and \( m = \lfloor nb \rfloor \) for \( b \in (0, 1) \). Under suitable weak dependence assumption, we have \( \sqrt{n}f_{tn}(\theta_0) \Rightarrow \{W_k(r + b) - W_k(r)\}/b \) for \( t = \lfloor nr \rfloor \). Using similar arguments in Theorem 5.2.1, we can show that

\[
elr(\theta_0) \to^d U_{c.t.k}(b) := \frac{2}{b} \max_{\lambda \in \mathbb{R}^k} \int_0^{1-b} \log(1 + \lambda'\{W_k(r + b) - W_k(r)\})dr.
\]

We generate the critical values of \( U_{c.t.k}(b)/(1 - b) \) for \( b \) from 0.01 to 0.3 with the spacing being 0.01 and further approximate the critical values by a cubic function of \( b \) following the practice of Kiefer and Vogelsang (2005). The estimates of the coefficients of the corresponding cubic functions are given in Table 5.1. Similarly, we summarize the critical values of \( U_{c.t.k}(b;K)/(1 - b) \) with \( K(x) = (\frac{\pi x}{8})^{1/2} \frac{1}{\pi} J_1(\frac{6\pi x}{5}) \) for \( b \) from 0.01 to 0.2 in Table 5.2, where \( J_1(\cdot) \) denotes the Bessel function of the first kind.

**Remark 5.2.2.** A natural question to ask here is whether the fixed-\( b \) asymptotics is consistent with the traditional small-\( b \) asymptotics when \( b \) is close to zero. In what follows, we provide an affirmative answer to this question by showing that \( U_{c.t.k}(b;K) \) converges to a scaled \( \chi^2_k \) distribution as \( b \to 0 \). We shall assume that \( K \) satisfies certain regularity conditions (see Assumption 2.2 in Smith, 2011).

Using the Taylor expansion and some standard arguments for EL, it is not hard to show that

\[
U_{c.t.k}(b;K) = \int_0^1 D_k(r, b)dr \left( \int_0^1 D_k(r, b)D_k(r, b)dr \right)^{-1} \int_0^1 D_k(r, b)dr/b + o_p(1).
\]
Define the semi-positive definite kernel
\[ K \] normalized partial sum of 1000 iid standard normal random variables and the number of Monte Carlo
\[ R \] estimated coefficients and multiple
\[ \eta \]  
\[ \beta \]

Table 5.1: Critical value function coefficients

<table>
<thead>
<tr>
<th>( u_{\epsilon,1}(b; 0.90) )</th>
<th>( a_0 )</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
<th>( a_3 )</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.661</td>
<td>6.547</td>
<td>12.819</td>
<td>-8.329</td>
<td>0.9984</td>
<td></td>
</tr>
<tr>
<td>3.917</td>
<td>5.819</td>
<td>34.483</td>
<td>-14.192</td>
<td>0.9976</td>
<td></td>
</tr>
<tr>
<td>6.593</td>
<td>4.631</td>
<td>231.740</td>
<td>-484.131</td>
<td>0.9855</td>
<td></td>
</tr>
<tr>
<td>4.827</td>
<td>-1.521</td>
<td>212.177</td>
<td>-477.642</td>
<td>0.9883</td>
<td></td>
</tr>
<tr>
<td>6.586</td>
<td>-18.806</td>
<td>469.034</td>
<td>-1076.779</td>
<td>0.9945</td>
<td></td>
</tr>
<tr>
<td>9.928</td>
<td>-36.918</td>
<td>1028.429</td>
<td>-2530.245</td>
<td>0.9906</td>
<td></td>
</tr>
<tr>
<td>6.424</td>
<td>-1.099</td>
<td>405.193</td>
<td>-1072.778</td>
<td>0.9962</td>
<td></td>
</tr>
<tr>
<td>7.783</td>
<td>3.125</td>
<td>560.737</td>
<td>-1552.979</td>
<td>0.9909</td>
<td></td>
</tr>
<tr>
<td>13.108</td>
<td>22.209</td>
<td>1080.359</td>
<td>-3330.819</td>
<td>0.9565</td>
<td></td>
</tr>
</tbody>
</table>

Note: the critical value \( u_{\epsilon,1}(b; 1 - \alpha) \) is approximated by a cubic function \( a_0 + a_1 b + a_2 b^2 + a_3 b^3 \) of \( b \). The estimated coefficients and multiple \( R^2 \) are reported. The Brownian motion is approximated by a normalized partial sum of 1000 iid standard normal random variables and the number of Monte Carlo replication is 5000.

Table 5.2: Critical value function coefficients

<table>
<thead>
<tr>
<th>( u_{\epsilon,1}(b; \kappa; 0.90) )</th>
<th>( a_0 )</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
<th>( a_3 )</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.324</td>
<td>8.243</td>
<td>112.149</td>
<td>-155.159</td>
<td>0.9992</td>
<td></td>
</tr>
<tr>
<td>5.116</td>
<td>-9.585</td>
<td>533.935</td>
<td>-1459.216</td>
<td>0.9989</td>
<td></td>
</tr>
<tr>
<td>9.612</td>
<td>-85.001</td>
<td>2237.407</td>
<td>-6595.415</td>
<td>0.9955</td>
<td></td>
</tr>
<tr>
<td>5.650</td>
<td>8.652</td>
<td>799.757</td>
<td>-2928.350</td>
<td>0.9930</td>
<td></td>
</tr>
<tr>
<td>7.709</td>
<td>-23.367</td>
<td>1833.534</td>
<td>-6752.999</td>
<td>0.9868</td>
<td></td>
</tr>
<tr>
<td>11.708</td>
<td>-54.068</td>
<td>4501.733</td>
<td>-17748.759</td>
<td>0.9802</td>
<td></td>
</tr>
<tr>
<td>6.860</td>
<td>48.990</td>
<td>1373.570</td>
<td>-6288.066</td>
<td>0.9804</td>
<td></td>
</tr>
<tr>
<td>7.656</td>
<td>105.088</td>
<td>1714.933</td>
<td>-8718.140</td>
<td>0.9731</td>
<td></td>
</tr>
<tr>
<td>4.963</td>
<td>505.832</td>
<td>346.210</td>
<td>-9711.193</td>
<td>0.9502</td>
<td></td>
</tr>
</tbody>
</table>

Note: the critical value \( u_{\epsilon,1}(b; \kappa; 1 - \alpha) \) is approximated by a cubic function \( a_0 + a_1 b + a_2 b^2 + a_3 b^3 \) of \( b \). The estimated coefficients and multiple \( R^2 \) are reported. The Brownian motion is approximated by a normalized partial sum of 1000 iid standard normal random variables and the number of Monte Carlo replication is 5000.

Under Assumption 5.2.2, we first derive that
\[
\frac{1}{b} \int_0^1 D_k(r, b) dr = \int_0^1 \int_0^1 K((r - s)/b) dr dW_k(s)/b = \int_0^1 \int_{-s/b}^{(1-s)/b} K(t) dt dW_k(s) \to^d \kappa_1 W_k(1).
\]

Define the semi-positive definite kernel \( K^\alpha(r, s) = \int_0^1 K((t - r)/b) K((t - s)/b) dt/(bn_2) \). Then we have \( V_k(b) = \frac{1}{b} \int_0^1 D_k(r; b) D_k(r; b) dr = \kappa_2 \int_0^1 \int_0^1 K^\alpha(r, s) dW_k(r) dW_k(s) = \kappa_2 \sum_{j=1}^{+\infty} \lambda_{j,b} \eta_j \eta_j^\prime \), where \( \{\eta_j\}_{j=1}^{+\infty} \) is a sequence of iid random vectors following \( N_k(0, I_k) \), and \( \lambda_{j,b} \) are the eigenvalues associated with the kernel \( K^\alpha(r, s) \). Note that
\[
E\{V_k(b)\}/\kappa_2 = I_k \int_0^1 K^\alpha(r, r) dr = \frac{I_k}{\kappa_2 b} \int_0^1 \int_0^1 K^2((t - r)/b) dt dr \to I_k.
\]

Let \( \eta_j = (\eta_{j1}, \ldots, \eta_{jk}) \) and denote by \( V_k^{(l,m)}(b) \) the \((l, m)\)th element of \( V_k(b) \) with \( 1 \leq l, m \leq k \).
Using the fact that \( \sum_{j=1}^{+\infty} \lambda_{j,b}^2 = \int_0^1 \int_0^1 \{K_*(r,s)\}^2 dr ds \to 0 \) as \( b \to 0 \) (see e.g., Sun, 2010), we get

\[
E\{V_k^{(l,m)}(b)/\kappa_2\}^2 = \sum_{j=1}^{+\infty} \sum_{j'=1}^{+\infty} \lambda_{j,b} \lambda_{j',b} E\eta_j \eta_{j'} \eta_j' \eta_{j'}' \to 0 \quad \text{as} \quad b \to 0,
\]

which implies that \( V_k(b) \to^\mu \kappa_2 I_k \). Therefore, we have \( U_{el,k}(b,K) \to^d \frac{\kappa_1^2}{\kappa_2} \lambda_k^2 \) as \( b \to 0 \). Compared to the \( \chi^2 \)-approximation, the fixed-\( b \) limiting distribution which captures the choice of the kernel and the bandwidth is expected to provide better approximation to the finite sample distribution of the BEL ratio statistic at the true parameter when \( b \) is relatively large.

5.2.2 Generalized empirical likelihood

In this subsection, we extend the fixed-\( b \) approach to the Generalized empirical likelihood (GEL) estimation framework (Newey and Smith, 2004). To describe the GEL, we let \( \rho \) be a concave function defined on an open set \( I \) which contains the origin. Set \( \rho(x) = -\infty \) for \( x / \in I \), and let \( \rho_j(x) = \partial^j \rho(x)/\partial x^j \) and \( \rho_j = \rho_j(0) \) for \( j = 0,1,2 \). We normalize \( \rho \) so that \( \rho_1 = \rho_2 = -1 \). Consider a set \( \Pi_n(\theta) = \{ \lambda : \lambda' f_{tn}(\theta) \in I, t = 1,2,\ldots,n \} \). The GEL estimator is defined as the solution to a saddle point problem,

\[
\hat{\theta}_{gel} = \arg\min_{\theta \in \Theta} \sup_{\lambda \in R^k} \hat{P}(\theta, \lambda) = \arg\min_{\theta \in \Theta} \sup_{\lambda \in \Pi_n(\theta)} \hat{P}(\theta, \lambda),
\]

where \( \hat{P}(\theta, \lambda) = J_n \sum_{i=1}^{n} \{\rho(\lambda' f_{tn}(\theta)) - \rho_0\} \). And the GEL ratio function is given by

\[
gelr(\theta) = \frac{2}{\lambda} \sup_{\lambda \in R^k} \hat{P}(\theta, \lambda).
\] (5.9)

The GEL estimator includes a number of special cases which have been well studied in the statistics and econometrics literature. The EL, exponential tilting (ET) and continuous updating (CUE) are all special cases of the GEL. For example, \( \rho(x) = \log(1-x) \) and \( I = (-\infty,1) \) for EL, \( \rho(x) = -e^x \) and \( I = R \) for ET, and \( \rho(x) = -(1+x)^2/2 \) and \( I = R \) for CUE. More generally, members of the Cressie-Read power divergence family of discrepancies discussed by Imbens et al. (1998) are included in the GEL class with \( \rho(x) = -(1+\gamma x)^{(\gamma+1)/\gamma} / \gamma + 1 \) (see Newey and Smith, 2004).

Define the nonlinear functional \( G_{gel}(f) = \max_{\lambda \in R^k} \int_0^1 \{\rho(\lambda' g(t)) - \rho_0\} dt \) for \( g \in C^{\otimes k}[0,1] \). Sup-
pose $\rho(\cdot)$ is strictly concave and is twice continuously differentiable. Then under suitable assumptions, it can be shown that $G_{gel}(\cdot)$ is a continuous functional under the sup norm. Since the argument follows from that presented in the appendix with a minor modification, we skip the details (see Remark 5.4.1). Therefore, we have

$$gel(\theta_0) \rightarrow_d U_{\rho,k}(b; K) := \frac{2}{b} \max_{\lambda \in \mathbb{R}^k} \int_0^1 \left\{ \lambda' \int_0^1 K((r-s)/b) dW_k(s) - \rho_0 \right\} dr.$$  

The GEL-based confidence region for the parameter $\theta_0$ is then defined as

$$\widetilde{CI}(1-\alpha; b) = \left\{ \theta \in \mathbb{R}^p : \frac{gel(\theta)}{1-b} \leq u_{\rho,k}(b; K; 1-\alpha) \right\}, \quad (5.10)$$

where $u_{\rho,k}(b; K; 1-\alpha)$ is the $100(1-\alpha)$% quantile of $U_{\rho,k}(b; K)/(1-b)$ which can again be obtained via simulation or iid bootstrap.

5.3 Numerical studies

We conduct two sets of simulation studies to compare and contrast the finite sample performance of the inference procedure based on the fixed-$b$ approximation and the BEL of Kitamura (1997) and Smith (2011).

5.3.1 Mean and quantiles

Consider two time series models, namely the AR(1) model $Y_t = \rho Y_{t-1} + \epsilon_t$ with $\rho = -0.5, 0.2, 0.5, 0.8$, and the AR(2) model, $Y_t = \frac{5}{6} Y_{t-1} - \frac{1}{6} Y_{t-2} + \epsilon_t$, the latter of which was used in Chen and Wong (2009), where the focus was to compare the finite sample coverages of the quantile delivered by BEL. In both models, $\{\epsilon_t\}$ is a sequence of iid standard normal random variables. We focus on the inference for mean, median and the 5% quantile. In the mean case, we have $f(y_t, \theta) = y_t - \theta$. For the $q$-th quantile, we consider the moment condition $f_q(y_t, \theta) = \int_{-\infty}^{(\theta-y_t)/h} K(x) dx - q$, where $K(\cdot)$ is a $r$-th order window which satisfies that

$$\int u^j K(u) du = \begin{cases} 1, & j = 0; \\ 0, & 1 \leq j \leq r - 1; \\ \kappa_0, & j = r, \end{cases}$$
for some integer \( r \geq 2 \) and \( h \) is a bandwidth such that \( h \to 0 \) as \( n \to +\infty \). Note that when \( h = 0 \), we have \( f_q(y_t, \theta) = I(y_t \leq \theta) - q \). To accommodate the dependence, we consider the BEL with fully overlapping moment conditions i.e., \( f_{tn}(\theta) = \frac{1}{m} \sum_{j=t}^{t+m-1} f(y_t, \theta) \) for \( t = 1, 2, \ldots, n - m + 1 \) and \( m = \lfloor nb \rfloor \) with \( b \in (0, 1) \). For the purpose of comparison, we also consider the smoothed EL with the kernel \( K(x) = \left( \frac{5\pi}{8} \right)^{1/2} \frac{1}{2} J_1 \left( \frac{6\pi x}{5} \right) \), where \( J_1(\cdot) \) is the Bessel function of the first kind. Note that the HAC covariance estimator induced by using \( K(\cdot) \) is essentially the same as the nonparametric long run variance estimator with the Quadratic spectral kernel (see Example 2.3 of Smith, 2011). The sample sizes considered are \( n = 100 \) and 400, and \( b \) is chosen from 0.02 to 0.2. To draw inference for the quantiles, we employed the second order Epanechnikov window with bandwidth \( h = cn^{-1/4} \) for \( c = 0, 1 \), following Chen and Wong (2009). The coverage probabilities and corresponding interval widths for the mean and quantiles delivered by the fixed-\( b \) approximation and the \( \chi^2 \)-based approximation are depicted in Figures 5.1-5.4.

In the mean case, undercoverage occurs for both the fixed-\( b \) calibration and the \( \chi^2 \)-based approximation when the dependence is positive, and it gets more severe as the dependence strengthens. Inference based on the fixed-\( b \) calibration provides uniformly better coverage probabilities in all cases considered here and it is quite robust to the choice of \( b \). The improvement is significant especially for large bandwidth. On the other hand, the fixed-\( b \) based interval is slightly wider than the \( \chi^2 \)-based interval. For negative dependence, i.e. \( \rho = -0.5 \), the fixed-\( b \) calibration tends to provide overcoverage, but the improvement over the \( \chi^2 \)-based approximation can still be observed for relatively large \( b \). These findings are consistent with the intuition that the larger \( b \) is, the more accurate the fixed-\( b \) based approximation provides relative to the \( \chi^2 \)-based approximation used by Kitamura (1997) and Smith (2011). The results for the median and 5\% quantile are qualitatively similar to those in the mean case. The choice of \( h = 1 \) tends to provide slightly shorter interval widths as compared to the unsmoothed counterpart, i.e. \( h = 0 \) in some cases (see Chen and Wong, 2009). A comparison of Figure 5.1 versus Figure 5.3 (Figure 5.2 versus Figure 5.4) shows that the coverage probabilities for the EL based on the kernel \( K(x) \) are generally closer to the nominal level than the BEL counterpart while the corresponding interval widths are wider. This phenomenon is consistent with the finding that QS kernel provides better coverage but wider interval widths as compared to the Bartlett kernel in Kiefer and Volgesang (2005) under the GMM framework. Our unreported simulation results also demonstrate the usefulness of the fixed-\( b \) calibration under the GEL estimation framework. The results for ET are available upon request.
5.3.2 Time series regression

We consider the stylized linear regression model with an intercept and a regressor \( x_t \): \( y_t = \beta_1 + \beta_2 x_t + u_t \) for \( 1 \leq t \leq n \), where \( \{x_t\} \) and \( \{u_t\} \) are generated independently from an AR(1) model with common coefficient \( \tilde{\rho} \). We set the true parameter \( \beta_0 = (\beta_{01}, \beta_{02}) = (0, 0) \) and choose \( \tilde{\rho} \in \{0.2, 0.5, 0.8\} \). Suppose we are interested in constructing confidence contour for \( \beta_0 \). We consider the stylized linear regression model with an intercept and a regressor \( x_t \). Consider the moment conditions \( f_t(\beta) = (u_t(\beta), x_t u_t(\beta), x_{t-1} u_t(\beta), x_{t-2} u_t(\beta)) \) with \( u_t(\beta) = y_t - \beta_1 - \beta_2 x_t \) and \( 3 \leq t \leq n \). We report the coverage probabilities for the BEL and the smoothed EL with kernel \( K(x) \) based on the fixed-\( b \) approximation and the \( \chi^2 \)-based approximation in Figure 5.5. As the dependence strengthens, both fixed-\( b \) calibration and \( \chi^2 \)-based approximations deteriorate. The coverage probabilities obtained from the fixed-\( b \) calibration are consistently closer to the nominal level, and the improvement is significant for large bandwidths. In contrast, the coverage probabilities based on the \( \chi^2 \) approximation is severely downward biased for relatively large \( b \) which can be very harmful for practical use.

To sum up, the fixed-\( b \) approximation provides a uniformly better approximation to the sampling distribution of the EL ratio statistic for a wide range of \( b \) and it tends to deliver more accurate coverage probability in confidence interval construction and size in testing. From a practical viewpoint, the choice of the bandwidth parameter has a great impact on the finite sample performance of the EL ratio statistic and it would be interesting to consider the optimal bandwidth under the fixed-\( b \) paradigm in future research.

5.4 Proofs of the main results

Define the set of functions \( Q = \{g = (g_1, g_2, \ldots, g_k) \in C^{\otimes k}[0,1] : g_i \text{'s are linearly independent}\} \) and let \( G_{el}(g) = \max_{\lambda \in \mathbb{R}^k} \int_0^1 \log(1 + \lambda' g(t))dt \) be a nonlinear functional from \( C^{\otimes k}[0,1] \) to the real line \( \mathbb{R} \), where \( \log(x) = -\infty \) for \( x < 0 \). We shall prove in the following that \( G_{el}(g) \) is a continuous map for functions in \( Q \) under the sup norm (Seijo and Sen, 2011). For any \( g \in C^{\otimes k}[0,1] \), we define \( H_g = \{\lambda \in \mathbb{R}^k : \min_{t \in [0,1]} (1 + \lambda' g(t)) \geq 0\} \) and \( L_g(\lambda) = -\int_0^1 \log(1 + \lambda' g(t))dt \). It is straightforward to show that \( L_g(\lambda) \) is strictly convex for \( g \in Q \) on the set \( H_g \). We also note that \( H_g \) is a closed convex set, which contains a neighborhood of the origin. Let \( \lambda_g = \arg\max_{\lambda \in \mathbb{R}^k} \int_0^1 \log(1 + \lambda' g(t))dt \) be the maximizer of \( -L_g(\lambda) \).

We first show that \( G_{el}(g) < \infty \) if and only if \( H_g \) is bounded. If \( G_{el}(g) = \infty \), then \( \lambda_g \) cannot be finite, which implies that \( H_g \) is unbounded. On the other hand, suppose \( H_g \) is unbounded. Note that \( H_g = \cap_{t \in [0,1]} \{\lambda \in \mathbb{R}^k : \lambda' g(t) \geq -1\} \) which is the intersection of a set of closed half-
spaces. The recession cone of $H_g$ is then given by $0^+H_g = \cap_{t \in [0,1]} \{ \lambda \in \mathbb{R}^k : \lambda'g(t) \geq 0 \}$ (see Section 8 of Rockafellar, 1970). By Theorem 8.4 of Rockafellar (1970), there exists a nonzero vector $	ilde{\lambda} \in 0^+H_g$, and the set $\{ t \in [0,1] : \tilde{\lambda}'g(t) > 0 \}$ has positive Lebesgue measure because of the linearly independence of $g$. We have $G_{el}(g) \geq -L_g(a\tilde{\lambda})$ for any $a > 0$, where $-L_g(a\tilde{\lambda}) \to \infty$ as $a \to \infty$. Thus we get $G_{el}(g) = \infty$.

Next, we consider the case $G_{el}(g) = \infty$. Following the discussion above, there exists $\tilde{\delta}$ such that the set $\mathcal{B} := \{ t \in [0,1] : \tilde{\lambda}'g(t) > \tilde{\delta} \}$ has Lebesgue measure $\Lambda(\mathcal{B}) > 0$. For any $A_0 > 0$, we choose $\epsilon_0 \in (0,1)$ and large enough $a > 0$ so that

$$\Lambda(\mathcal{B}) \log(1 + a\tilde{\delta} - \epsilon_0) + \log(1 - \epsilon_0) > A_0.$$ 

For any $f \in Q$ with $||f - g|| := \sup_{t \in [0,1]} |f(t) - g(t)| \leq \epsilon_0/||\tilde{\lambda}||$, we have

$$\int_0^1 \log(1 + a\tilde{\lambda}'f(t))dt = \int_{\mathcal{B}} \log(1 + a\tilde{\lambda}'(f(t) - g(t)) + a\tilde{\lambda}'g(t))dt$$

$$+ \int_{\mathcal{B}^c} \log(1 + a\tilde{\lambda}'(f(t) - g(t)) + a\tilde{\lambda}'g(t))dt$$

$$\geq \Lambda(\mathcal{B}) \log(1 + a\tilde{\delta} - \epsilon_0) + \log(1 - \epsilon_0) > A_0.$$ 

In what follows, we turn to the case $G_{el}(g) < \infty$, i.e., $H_g$ is bounded as shown before.

**Case 1:** we first consider the case that $G_{el}(g) < \infty$, i.e., $H_g$ is bounded as shown before.

Given any $\epsilon > 0$, we shall first show that $\sup_{\lambda \in \tilde{B}(\lambda_g; \tau)} |L_f(\lambda) - L_g(\lambda)| < \epsilon$ for any $f \in Q$ with $||f - g|| < \tilde{\delta}(\epsilon)$, where $0 < \tilde{\delta}(\epsilon) < \delta$. Because $G_{el}(g) < \infty$, we have $\int_0^1 \log(1 + \lambda'g(t))dt < \infty$ for any $\lambda \in \tilde{B}(\lambda_g; \tau)$. Simple algebra yields that

$$\left| \int_0^1 \log(1 + \lambda'f(t))dt - \int_0^1 \log(1 + \lambda'g(t))dt \right| \leq \max \left\{ \log(1 + M\tilde{\delta}(\epsilon)/c'), \log(1 + M\tilde{\delta}(\epsilon)/c) \right\},$$

(5.11)

where $M = ||\lambda_g|| + \tau$. The RHS of (5.11) can be made arbitrarily small for sufficiently small $\tilde{\delta}(\epsilon)$. Therefore we get $\sup_{\lambda \in \tilde{B}(\lambda_g; \tau)} |L_f(\lambda) - L_g(\lambda)| < \epsilon$ for small enough $\tilde{\delta}(\epsilon)$, which implies that
$|G_{el}(g) - \sup_{\lambda \in B(\lambda_g; \tau)} \int_0^1 \log(1 + x'f(t))dt| < \epsilon$. Next, we show that there exists a local maxima of $-L_f(\lambda)$ in $\bar{B}(\lambda_g; \tau)$. Suppose $\epsilon$ is sufficiently small and choose $0 < \xi < \tau$ such that $-L_g(\lambda_g) > \max_{\lambda \in \bar{B}(\lambda_g; \tau) \cap B(\lambda_g; \xi)} -L_g(\lambda) + 2\epsilon$, where $\bar{B}(\lambda_g; \xi) = \{\lambda \in \mathbb{R}^k : |\lambda - \lambda_g| < \xi\}$. Thus we get

$$\max_{\lambda \in \bar{B}(\lambda_g; \tau) \cap B(\lambda_g; \xi)} -L_f(\lambda) \leq \max_{\lambda \in \bar{B}(\lambda_g; \tau) \cap B(\lambda_g; \xi)} -L_g(\lambda) + \epsilon$$

$$< -L_g(\lambda_g) - \epsilon \leq -L_f(\lambda_g) \leq \max_{\lambda \in \bar{B}(\lambda_g; \xi)} -L_f(\lambda).$$

Because $f \in Q$, $L_f(\lambda)$ is strictly convex. Hence, the local maxima is also the global maxima, which implies that $|G_{el}(g) - G_{el}(f)| < \epsilon$.

**Case 2**: We now consider the case $\min_{x \in [0,1]} (1 + 4'g(t)) = 0$. For any $0 < \delta^* < \delta^{**} < 1$, let $H_g(\delta^*) = \{(1 - \delta^*)\lambda : \lambda \in H_g\}$ and $H_f(\delta^{**}) = \{(1 - \delta^{**})\lambda : \lambda \in H_f\}$. There exists a small enough $\delta > 0$ such that for any $f \in Q$ with $||f - g|| < \delta$, $H_f(\delta^{**}) \subseteq H_g(\delta^*) \subseteq \bar{H}_f \cap \bar{H}_g$. By the continuity of $L_g(\lambda)$, we know for any $\epsilon > 0$, there exists a $\delta^* > 0$ such that when $|\lambda - \lambda_g| \leq \delta^*|\lambda_g|$, $-L_g(\lambda_g) < -L_g(\lambda) + \epsilon/4$. By the construction of $H_g(\delta^*)$, we have

$$-L_g(\lambda_g) < -L_g((1 - \delta^*)\lambda_g) + \epsilon/4 \leq \sup_{\lambda \in H_g(\delta^*)} -L_g(\lambda) + \epsilon/4.$$

Using similar arguments in the first case and the boundness of $H_g$, we can show that

$$\left| \sup_{\lambda \in H_g(\delta^*)} -L_g(\lambda) - \sup_{\lambda \in H_g(\delta^*)} -L_f(\lambda) \right| < \epsilon/8,$$

for sufficiently small $\delta$. Furthermore, when $\lambda_f \in H_g(\delta^*)$, we have $-L_f(\lambda_f) = \sup_{\lambda \in H_g(\delta^*)} -L_f(\lambda)$. When $\lambda_f \notin H_g(\delta^*)$, by the convexity of $L_f(\lambda)$, we get $L_f((1 - \delta^{**})\lambda_f) \leq (1 - \delta^{**})L_f(\lambda_f)$, which implies that

$$-L_f(\lambda_f) \leq L_f((1 - \delta^{**})\lambda_f) \leq \frac{\sup_{\lambda \in H_g(\delta^*)} -L_f(\lambda)}{1 - \delta^{**}} \leq \frac{\sup_{\lambda \in H_g(\delta^*)} -L_g(\lambda) + \epsilon/8}{1 - \delta^{**}} \leq \sup_{\lambda \in H_g(\delta^*)} -L_g(\lambda) + \epsilon/4$$

$$< \sup_{\lambda \in H_g(\delta^*)} -L_f(\lambda) + \epsilon/2.$$
for small enough $\delta^{**}$ (e.g., $\delta^{**} < \min(1/3, \frac{1}{\max_{y \in [0,1]} f_{ty,n}(y)})$). Thus we have

$$|G_{cl}(f) - G_{cl}(g)| \leq \left| -L_f(\lambda_f) - \sup_{\lambda \in H_\delta(\delta^*)} -L_f(\lambda) \right| + \left| -L_g(\lambda_g) - \sup_{\lambda \in H_\delta(\delta^*)} -L_g(\lambda) \right|$$

$$\left| \sup_{\lambda \in H_\delta(\delta^*)} -L_g(\lambda) - \sup_{\lambda \in H_\delta(\delta^*)} -L_f(\lambda) \right| < \epsilon.$$ 

Combining the above arguments, we show that the map $G_{cl}$ is continuous under the sup norm.

Next, we consider the limiting process $D_k(r; b) = \int_0^1 K((r - s)/b)dW_k(s)$ with $b \in (0, 1)$ being fixed in the asymptotics. Because the components of $D_k(r; b)$ are mutually independent, we have $P(\alpha'D_k(r; b) = 0 \text{ for some } \alpha \in \mathbb{R}^k) = 0$ which implies that $P(D_k(r; b) \in Q) = 1$. Under the assumptions in Theorem 5.2.1, the set \{ $\lambda : \min_{r \in [0,1]} (1 + \lambda'D_k(r; b)) \geq 0$ \} is compact and convex almost surely (note the convexity and closeness of the set follow directly from its definition). Using summation by parts, we get

$$\sqrt{n}f_{tn}(\theta_0) = \frac{1}{b\sqrt{n}} \sum_{s=t-n}^{t-1} \mathcal{K} \left( \frac{s}{S_n} \right) f_{t-s}(\theta_0) = \frac{1}{b\sqrt{n}} \sum_{s=1}^{n} \left\{ \mathcal{K} \left( \frac{t-s}{S_n} \right) - \mathcal{K} \left( \frac{t-s-1}{S_n} \right) \right\} \sum_{k=1}^{s} f_k(\theta_0).$$

By the continuous mapping theorem and Itô’s formula, we obtain

$$\sqrt{n}f_{tn}(\theta_0) \Rightarrow^d \Lambda \left\{ \frac{1}{b} \mathcal{K} \left( \frac{r-1}{b} \right) W_k(1) + \frac{1}{b^2} \int_0^1 \mathcal{K}' \left( \frac{r-s}{b} \right) W_k(s)ds \right\} = ^d \Lambda D_k(r; b)/b,$$

for $t = \lfloor nr \rfloor$ with $r \in [0, 1]$. Finally, by the continuous mapping theorem, we get

$$\text{clr}(\theta_0) = \frac{2}{b} \max_{\lambda \in \mathbb{R}^k} \sum_{t=1}^{n} \log(1 + \tilde{\lambda}'\sqrt{n}b\Lambda^{-1}f_{tn}(\theta_0))/n, \quad \tilde{\lambda} = \Lambda'\lambda/\sqrt{n}b,$$

$$-^dU_{cl,k}(b; \mathcal{I}) := \frac{2}{b} \max_{\lambda \in \mathbb{R}^k} \int_0^1 \log(1 + \tilde{\lambda}'D_k(r; b)) dr.$$ 

**Remark 5.4.1.** For ET and CUE, we have $\mathcal{I} = \mathbb{R}$. Given any $g \in Q$ with $G_{gel}(g) < \infty$, we have $H_g = \{ \lambda \in \mathbb{R}^k : \lambda'g(t) \in \mathcal{I}, \text{for all } t \in [0,1] \} = \mathbb{R}^k$ and $\lambda_g < \infty$. Therefore, $\lambda_g$ is an interior point of $H_g$ and the arguments in Case 1 can be applied to show the continuity of $G_{gel}(-)$ at $g$. 

137
Figure 5.1: Coverage probabilities for the mean delivered by the BEL based on the fixed-$b$ approximation and the $\chi^2$-based approximation. The nominal level is 95% and the number of Monte Carlo replications is 1000.
Figure 5.2: Coverage probabilities for the median and 5% quantile delivered by the BEL based on the fixed-$b$ approximation and the $\chi^2$-based approximation. The nominal level is 95% and the number of Monte Carlo replications is 1000.
Figure 5.3: Coverage probabilities for the mean delivered by the smoothed EL based on the fixed-
$b$ approximation and the $\chi^2$-based approximation. Note that the corresponding kernel is $K(x) = \left(\frac{5\pi}{8}\right)^{1/2} \frac{1}{x} J_1 \left(\frac{6\pi x}{5}\right)$. The nominal level is 95% and the number of Monte Carlo replications is 1000.
Figure 5.4: Coverage probabilities for the median and 5% quantile delivered by the smoothed EL based on the fixed-\(b\) approximation and the \(\chi^2\)-based approximation. Note that the corresponding kernel is \(K(x) = \left(\frac{\pi x}{8}\right)^{1/2} \frac{1}{2} J_1 \left(\frac{2\pi x}{5}\right)\). The nominal level is 95% and the number of Monte Carlo replications is 1000.
Figure 5.5: Coverage probabilities delivered by the BEL (left panels) and the smoothed EL (right panels) based on the fixed-$b$ approximation and the $\chi^2$-based approximation. Note that the corresponding kernel for the smoothed EL is $K(x) = \left(\frac{5\pi}{8}\right)^{1/2} \frac{1}{2} J_1 \left(\frac{9\pi x}{5}\right)$. The nominal level is 95% and the number of Monte Carlo replications is 1000.


