Non-stationary functional time series and functional machine learning : inference and applications

Chen, Yichao

2021

Chen, Y. (2021). Non-stationary functional time series and functional machine learning : inference and applications. Doctoral thesis, Nanyang Technological University, Singapore.
<https://hdl.handle.net/10356/151903>

<https://hdl.handle.net/10356/151903>

<https://doi.org/10.32657/10356/151903>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 23:31:49 SGT

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**NON-STATIONARY FUNCTIONAL TIME SERIES
AND FUNCTIONAL MACHINE LEARNING:
INFERENCE AND APPLICATIONS**

CHEN YICHAO

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

2021

NON-STATIONARY FUNCTIONAL TIME SERIES AND FUNCTIONAL MACHINE LEARNING: INFERENCE AND APPLICATIONS

CHEN YICHAO

SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of
Doctor of Philosophy

2021

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me except where otherwise stated in this thesis. The thesis work has not been submitted for a degree or professional qualification to any other university or institution. I declare that this thesis is written by myself and is free of plagiarism and of sufficient grammatical clarity to be examined. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

7 Jan 2021

Date

CHEN YICHAO

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it of sufficient grammatical clarity to be examined. To the best of my knowledge, the thesis is free of plagiarism and the research and writing are those of the candidate's except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

7 Jan 2021

.....
Date

.....
PUN CHI SENG

Authorship Attribution Statement

This thesis contains material from 1 paper published in the following peer-reviewed journal where I am listed as the first author.

Chapter 2 is published as: Chen, Y. and Pun C. S. A bootstrap-based KPSS test for functional time series. Journal of Multivariate Analysis, 174:104535 (2019). DOI:10.1016/j.jmva.2019.104535.

The contributions of the co-authors are as follows:

- Prof. Pun suggested the study area, proposed the research topic, provided meritorious instruction throughout the study of this topic, and offered incredible help on revising the manuscript draft.
- I prepared the draft of the manuscript and conducted all simulation studies, under the supervision of Prof. Pun.
- I derived all theoretical results and performed all simulation studies under guidance of Prof. Pun.
- I drafted the initial version of the paper, which was edited and finalized by Prof. Pun.

7 Jan 2021

Date

CHEN YICHAO

Abstract

Functional time series analysis is important in the research of functional data, mainly in finance, also be involved in biology, medicine and many other areas. Stationary functional time series have many good properties that allow us to do the analysis and prediction with many existing methods. However, many of the functional time series data in our real world are nonstationary. Some of them may have a trend or just be random walks. For example, most of the stock price curves are nonstationary functional time series, which are difficult to perform analysis directly on these curves. In view of this, functional KPSS test, functional unit root test and functional co-integration test can be important and have broad prospects of research and application.

This thesis consists of two parts. In first part, we study the theory of statistical inference in nonstationary functional time series, including functional KPSS test and functional unit root test. In this part, our main constructions consist of two parts as follows:

(a) We examine bootstrap methods to construct a generalized KPSS test for functional time series. Bootstrap-based functional testing provides an intuitive and efficient estimation of the distribution of the generalized KPSS test statistic and is capable of achieving non-trivial powers against many alternative hypothesis. We derive the asymptotic distribution of the simple bootstrap-based KPSS test statistic for functional time series, which proves the bootstrap validity of average. Simulation studies are then conducted to examine the performance of the proposed KPSS tests in small and moderate sample sizes. The results demonstrate that the bootstrap-based functional KPSS test has good empirical size and power. Moreover, its implementation is more

efficient than the existing KPSS test for functional time series.

(b) We propose a unit root test for functional time series. We derive a new analytical framework for non-stationary functional time series. Specifically, for the proposed test statistic, we derive its limit distribution under the null hypothesis of a random walk and its asymptotic behavior of alternative hypotheses of trend stationary, weakly dependent stationary, and autoregressive stationary models. For the trend stationary model as an alternative, a theoretical derivation of the test consistency is provided, while for other two alternatives, a combination of theoretical and experimental validation on the statistical power of the test is presented. Simulation studies are conducted to justify the theories and the desirable finite-sample performance of the proposed functional unit root test. The proposed test is also applied to real data of intraday stock price curves and the test results are plausible.

If we test our functional data as i.i.d. or stationary data, we can consider to apply machine learning to analyze the functional data. In second part:

(c) We do an application on the functional data in chemistry. We apply different machine learning methods on the prediction of metallic nanoparticle size and size distribution from the functional data generated by the localized surface plasmon resonances (LSPRs). All of these predictions receive reliable results.

Acknowledgement

The four years of my PhD life are coming to an end. Looking back on these four years, there is so much to say. Whether happy and satisfied, or regretful and confused emotions are slowly intertwined in my mind. But in any case, this is the time. I wish to thank the people who accompanied me through this journey.

I first give my heartfelt thanks to my supervisor, Professor Pun Chi Seng. From the beginning of receiving me as his PhD student when I was confused in my last year of college, to every time's patient guidance to me during these four years. He showed me how to learn, to explore and to express from a novice in research; he helped me revise and deliver the papers, and even patiently guide my every question on language; he talked with me about the past and the dream, often for a duration of several hours; and finally, he guided me to finish my thesis. Many times in the past four years, I have also known that I often fell into a lot of meaningless impulses and many times not to do my best. But his patience to me always touched me, so that I can finally cross the confusion and finish the meaningful journey.

I would like to thank Professor Pan Guangming, who recommended me and gave me the opportunity to continue my study here. He also gave me a lot of care and guidance in the four years. I am also thankful to the other committee member Prof. Feng Qu, and many teachers who gave me dedicated instructions during my PhD life, especially to Prof. Bei Xiaohui and Prof. Wu Guohua.

I would also express my sincere thanks to group members: Ye Zi, Lei Qian and Wu Weiyin, to visiting scholars and friends in NTU who have been with me and gave me a lot help: Yang Yubo, Chen Hongbin, Gao Yang, Huang Rui, Peng Mengjiao, Zhang

Zhixiang, Yu Hao, Wang Lei, Liu Wenjie, Sheng Changtao, Wang Haoyang and Yang Zhiguo, and to my best brothers from primary school: Li Xin and Li Run. It was you who made my four-year journey always full of color and temperature.

Last but not least, I give the greatest thanks to my parents. Their love and care to me make me come to the present without any concern.

The company and help from these important people are great honors in my heart. I want to thank you again. I hope myself to be a more mature and excellent person in the future, and live up to your expectations.

Contents

Abstract	1
Acknowledgement	3
List of Figures	8
List of Tables	11
1 Introduction	15
1.1 Functional Time Series Analysis and Inference	15
1.2 History of Unit Root Test	16
1.3 Bootstrap Method on the Functional KPSS test	17
1.4 Machine Learning on Nanoparticle Size Prediction	18
1.5 Outline	19
2 A Bootstrap-based KPSS Test for Functional Time Series	21
2.1 Problem Formulation and Preliminary Results	22
2.1.1 The Problem	22
2.1.2 Preliminary Asymptotic Results of the FKPSS Test	24
2.2 Asymptotic Theory of the Bootstrap Test Statistic	26
2.2.1 Bootstrap-based FKPSS Testing	26
2.2.2 Bootstrap Validity on Average	27

2.3	Simulation Studies	31
2.3.1	Simulation Setup	33
2.3.2	Simulation Results for the Independent Errors	35
2.3.3	Simulation Results for the Stationary Errors	38
2.4	Conclusion	41
2.5	Proofs	43
2.5.1	Preliminary Results	43
2.5.2	Main Results in Section 2.2.2	50
3	Functional Unit Root Test	57
3.1	Problem Formulation and Test Statistic	58
3.1.1	The Problem	58
3.1.2	Test Statistic of Functional Unit Root Test	60
3.2	Asymptotic Theory of the Functional Unit Root Test	61
3.2.1	Limit Distribution of the Test Statistic under the Null Hypothesis	61
3.2.2	Asymptotic Behavior of the Test Statistic under Alternative Hypotheses	63
3.2.3	Summary of the Asymptotic Tests	66
3.3	Simulation Studies	66
3.3.1	Details of Simulation and Implementation of FURT	66
3.3.2	Simulation Results: Empirical Sizes and Powers	69
3.3.3	Trends of Largest Eigenvalues under H_1^{ws} and H_1^{ar1}	71
3.4	Application of FURT to Intraday Stock Price Curves	72
3.5	Conclusion	76
3.6	Proof of the theorems in Section 3.2.1	77
3.6.1	Preliminary Results	77
3.6.2	Proof of Theorem 3.2.1	78
3.7	Proofs of the theorems in Section 3.2.2	78
3.7.1	Proof of Theorem 3.2.2	78
3.7.2	Proof of Theorem 3.2.3	79

3.7.3	Proof of Theorem 3.2.4	82
3.7.4	Proof of Theorem 3.2.5	83
4	Functional Machine Learning Application on Nanoparticle Size and Size	
	Distributions Prediction	87
4.1	Introduction of Chemical Experimental Preparation	88
4.1.1	Preparation of Small Au NS Seeds	88
4.1.2	Preparation of Large Gold Nanospheres	89
4.1.3	SEM Characterization of the Synthesised Au NS	89
4.2	Introduction of machine learning methods	90
4.2.1	Linear regression	90
4.2.2	Random Forest Regression	92
4.2.3	XGBoost Regression	92
4.2.4	Neural Network	93
4.3	Feature Engineering	94
4.3.1	Introduction of Feature Engineering Methods	94
4.3.2	Input Data Preparation	96
4.4	Testing Procedure and Results Analysis	99
4.4.1	Testing Procedure	99
4.4.2	Linear Regression	100
4.4.3	Random Forest	105
4.4.4	XGBoost	107
4.4.5	Neural Network	111
4.4.6	Comparison of the Results between different Methods	113
4.4.7	Analysis of the Results	115
4.5	Conclusion	121
5	Conclusion and Future Work	123
	Bibliography	125

List of Figures

2.1	A simple example of simulated data under different hypotheses when $N = 20$. Figures in the first row show the simulated data under H_0^{ts} (left) and H_A^{ts+rw} (right) respectively. Figures in the second row show the simulated data under H_A^{ur} (left) and H_A^{cp} (right) respectively. . . .	36
3.1	Plots of the largest eigenvalues $\hat{\lambda}_1$ of $\hat{c}_N(t, s)$ under H_1^{ws} (left) and H_1^{ar1} (right) with bandwidth $h = N^{\frac{1}{5}}$ (blue), $N^{\frac{2}{5}}$ (red) against the sample size N	72
3.2	Plots of $N\hat{\lambda}_1$ under H_1^{ws} (upper) and H_1^{ar1} (lower) with bandwidth $h = N^{\frac{1}{5}}$ (left, blue), $N^{\frac{2}{5}}$ (right, red) against the sample size N	73
3.3	p -values of FURT of 50 intraday stock price curves. The left (resp.) panel presents the p -values of the left-tailed (resp. right-tailed) tests. The horizontal lines of 5% and 10% are added to indicate that the point below the line suggest a rejection of FURT.	74
3.4	Price curves of 8x8 Inc. from 2 Jan 2020 to 28 Feb 2020.	75
3.5	Difference of price curves of 8x8 Inc. between adjacent days from 2 Jan 2020 to 28 Feb 2020.	76
4.1	Measurements of Au NS.	90

List of Figures

4.2	LSPR of Au NSs with mean particle size = 21.72nm and standard deviation = 1.59nm.	91
4.3	Relation between the particle size (nm) and position of the first peak. .	95
4.4	Curve of experimental data (hollow points) for particle size=57.83nm, and fitted Gaussian function when choosing 40 (red), 80 (green) and 120 (blue) points between the peak respectively.	97
4.5	Curve of experimental data (hollow points) for particle size=128.70nm, 139.71nm, 159.99nm and 184.45nm, and fitted Gaussian functions (red) respectively.	98
4.6	3d graphs between particle size, peak position and FWHM of experimental data (black) and estimated surface from B-spline linear regression (red).	116
4.7	3d graphs between deviation, peak position and FWHM of experimental data (black) and estimated surface from B-spline linear regression (blue).	117
4.8	3d graph between particle size, peak position and FWHM of experimental data (black) and estimated surface from random forest (red). .	118
4.9	3d graph between deviation, peak position and FWHM of experimental data (black) and estimated surface from random forest (blue).	118
4.10	3d graph between particle size, peak position and FWHM of experimental data (black) and estimated surface from XGBoost (red).	119
4.11	3d graph between deviation, peak position and FWHM of experimental data (black) and estimated surface from XGBoost (blue).	119
4.12	3d graph between particle size, peak position and FWHM of experimental data (black) and estimated surface from neural network (red). .	120
4.13	3d graph between deviation, peak position and FWHM of experimental data (black) and estimated surface from neural network (blue).	120

List of Tables

2.1	Empirical sizes (%) and powers (%) for the different FKPSS tests for FTS with normally distributed errors, different sample sizes $N = 30, 50, 70, 100, 150, 200$ and different significance levels $\alpha = 1\%, 5\%$.	37
2.2	Empirical sizes (%) and powers (%) for the different FKPSS tests for FTS with t_5 -distributed errors, different sample sizes $N = 30, 50, 70, 100, 150, 200$ and different significance levels $\alpha = 1\%, 5\%$	39
2.3	Empirical sizes (%) and powers (%) for the different FKPSS tests for FTS with $F(10, 8)$ -distributed errors, different sample sizes $N = 30, 50, 70, 100, 150, 200$ and different significance levels $\alpha = 1\%, 5\%$.	40
2.4	Averaged power increase rates (%) from iKY to SB with different sample sizes $N = 30, 50, 70, 100, 150, 200$, different significance levels $\alpha = 1\%, 5\%$ and different kinds of sample noises.	41

2.5	Empirical sizes (%) and powers (%) for the different FKPSS tests for FT-S with AR(1) errors, different sample sizes $N = 30, 50, 70, 100, 150, 200$ and different significance levels $\alpha = 1\%, 5\%$	42
3.1	Empirical sizes (%) and powers (%) for FURT for H_0^{ur} versus H_1^{ts} . . .	69
3.2	Empirical sizes (%) and powers (%) for FURT for H_0^{ur} versus H_1^{ws} or H_1^{ar1}	70
4.1	Estimated error of particle size (%) and deviation (nm) of linear regression with Fourier coefficients.	100
4.2	Estimated error of particle size (%) and deviation (nm) of B-spline linear regression with peak position or FWHM individually.	102
4.3	Estimated error of particle size (%) B-spline linear regression with both peak position (first column) and FWHM (first row).	102
4.4	Estimated error of deviation (nm) B-spline linear regression with both peak position (first column) and FWHM (first row).	103
4.5	Estimated error of particle size (%) B-spline linear regression with both first peak position (first column) and second peak position (first row), for data with two peaks only.	103
4.6	Estimated error of deviation (nm) B-spline linear regression with both first peak position (first column) and second peak position (first row), for data with two peaks only.	104

4.7	Estimated error of particle size (%) B-spline linear regression with both first FWHM (first column) and second FWHM (first row), for data with two peaks only.	104
4.8	Estimated error of deviation (nm) B-spline linear regression with both first FWHM (first column) and second FWHM (first row), for data with two peaks only.	104
4.9	Estimated error of particle size (%) and deviation (nm) of random forest with Fourier coefficients.	105
4.10	Estimated error of particle size (%) and deviation (nm) of random forest with peak position and FWHM.	106
4.11	Estimated error of particle size (%) and deviation (nm) of random forest with two peak positions and FWHMs, for data with two peaks only. .	107
4.12	Estimated error of particle size (%) and deviation (nm) of XGBoost with Fourier coefficients.	108
4.13	Estimated error of particle size (%) and deviation (nm) of XGBoost with peak position and FWHM.	109
4.14	Estimated error of particle size (%) and deviation (nm) of XGBoost with two peak positions and FWHMs, for data with two peaks only. .	110

4.15	Estimated error of particle size (%) and deviation (nm) of neural network with Fourier coefficients, which contains 3 hidden layers with different neurons on each layer.	112
4.16	Estimated error of particle size (%) and deviation (nm) of neural network with two peak positions and FWHMs for data with two peaks only, which contains 3 hidden layers with different neurons on each layer.	113
4.17	Best estimated results of different machine learning methods from Fourier coefficients.	113
4.18	Best estimated results of different machine learning methods from first peak position and FWHM.	114
4.19	Best estimated results of different machine learning methods from two peak positions and FWHMs, for data with two peaks only.	114

Introduction

1.1 Functional Time Series Analysis and Inference

Functional time series (FTS) data emerge in many applications such as the time series of intra-day stock (index) price curves (see Shang (2017a)), age-specific medical data (see Erbas et al. (2007)), intra-day environmental data (see Shang (2017b)), and intra-day electricity price curves (see Liebl (2013)). Moreover, applications of FTS are not limited to the measurement frequency of curves and can be extended to any frequency of interest (e.g., monthly, yearly, or even hourly or minute-by-minute), which provides new insights for high-dimensional data analysis. Over the past two decades, the theoretical developments of functional (dependent) data analysis have become popular, as using the functional representation of the collated data provides a more flexible and precise way to investigate the relation between data on different measured periods; see Ramsay and Silverman (2005) and Horváth and Kokoszka (2012). Early FTS models were functional autoregressive (AR) models; see Bosq (2000) and Kargin and Onatski (2008). However, recent studies such as Aue et al. (2015); Hörmann et al. (2015); Hörmann and Kokoszka (2010); Hörmann and Kokoszka (2012) have extended these to a non-linear dependence structure.

Consequently, hypothesis testing for the various assumptions of FTS models has

received considerable research attention including Gabrys et al. (2010) for error correlation, Horváth et al. (2010) for stability, Kokoszka and Reimherr (2013) for the order of an autoregression, Horváth et al. (2014); Kokoszka and Young (2016) for stationarity, Constantinou et al. (2017) for the separability of space-time domains, Aue et al. (2018) for structural break detection, Górecki et al. (2018) for normality, and Hörmann et al. (2018) for periodicity, to name a few.

Just like the assumption in traditional research of time series, stationarity is an important condition to the research of FTS. Therefore, the stationarity of FTS is always an important research area in recent years. However, many common functional time series data in our real lives are always non-stationary, such as stock price curves and environmental data. In fact, most of them will be rejected in the stationarity test, see Horváth et al. (2014). Because of this, the non-stationarity test is needed for these kinds of data. Unit root test and co-integration test were paid the most attention. KPSS test (trend stationary test) is also an alternative to the Dickey-Fuller unit root test.

1.2 History of Unit Root Test

The unit root model has long been used in econometrics, including growth convergence theories, purchasing power parity hypothesis, and estimation of exchange rates and future interest rates. Testing for unit roots has been intensively studied since the seminal work of Dickey and Fuller (1979, 1981) for the scalar time series. Subsequently, Bhargava (1986); Dickey et al. (1986) develop the theory of unit root tests, especially their contribution to the deformation of unit root models, and Phillips and Perron (1988) develops the Phillips–Perron test by making non-parametric correction to the statistic of the Dickey–Fuller test. In the last two decades, the studies of unit root tests were extended to panel data. For example, Maddala and Wu (1999) conducts a comparative study of unit root tests under different null and alternative hypotheses; Levin et al. (2002) provides a systematic theory of panel unit root tests; Im et al. (2003) establishes a unit root test for dynamic heterogeneous panels; Chang (2004) construct a bootstrap-based

unit root test for panel data, which makes the tests with small sample more efficient; Chang (2012) invents an equi-squared-sum-based approach for testing panel unit roots; Moon and Perron (2012) uses the false discovery rate in classification between unit root and stationary models; Westerlund and Larsson (2015) analyzes the local asymptotic power of panel unit root tests; and Zhang and Chan (2018) proposes a Portmanteau-type test for unit roots. Recently, Zhang et al. (2018) extends the study to high-dimensional unit root tests, whose statistic is constructed with the largest eigenvalue of the kernel of the integrated operator. Chapter 3 is to extend the sequence of studies on unit root tests to FTS data.

1.3 Bootstrap Method on the Functional KPSS test

Most testing procedures of FTS test rely on the limiting distribution of the corresponding test statistic, which requires a large sample size to validate the tests. However, while such an asymptotic approximation is common in the literature on functional testing, we may only be able to obtain relatively small samples in practice.

As a practical alternative, the bootstrap-based functional testing procedure has thus been introduced for functional data. Benko et al. (2009) and Paparoditis and Sapatinas (2016) propose two bootstrap tests for testing the equality of two sample mean functions using different test statistics. The simple bootstrap (SB) method adopted in these studies is initiated by Efron (1979); Efron and Tibshirani (1993). In terms of statistical estimation, the SB method is applied to functional data by Cuevas et al. (2006) and to stationary FTS by Paparoditis (2018); Shang (2018). Pilavakis et al. (2017) extend the use of bootstrap methods for independent functional data to weakly dependent functional data by applying the moving block bootstrap (MBB) method of Künsch (1989). In the bootstrap-based functional testing of FTS, the stationarity assumption plays an important role. Although the bootstrap unit root test is available in a multivariate setting (see Chang (2004); Chang and Park (2003)), the application of bootstrap methods to non-stationary FTS remains a problem.

We take a step toward overcoming this challenge by investigating the bootstrap methods used to test trend-stationary FTS, also known as the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test (see Kwiatkowski et al. (1992)). The KPSS test, introduced as an alternative to the Dickey–Fuller unit root test, has been found to have better utility in econometrics; see Dickey and Fuller (1979, 1981); Lee and Schmidt (1996). Following the time-domain approach in Kwiatkowski et al. (1992), Horváth et al. (2014) and Kokoszka and Young (2016) derive the limiting distributions of a generalized KPSS test statistic for FTS under the stationarity and trend-stationarity hypotheses, respectively. In Chapter 2, we bridge this research gap by applying bootstrap methods to the functional KPSS (FKPSS) test in the case of small or moderate sample sizes. We validate our proposed bootstrap FKPSS test theoretically and numerically through simulations with various alternative hypotheses. Our theoretical results shed light on the validity of bootstrap methods for non-stationary FTS, while our simulation results show that the bootstrap-based FKPSS test has good empirical sizes and powers in various settings.

1.4 Machine Learning on Nanoparticle Size Prediction

Establishing structure-to-property relationships lies at the heart of nanomaterial informatics. These correlations provide a fundamental framework to select and design specific nanostructures for targeted applications in sensing, therapeutics, light harvesting, and electronics. A current bottleneck in characterizing these nanoscopic particles lies in the need to employ a combination of microscopic and spectroscopic techniques which cannot achieve high throughput capabilities. For instance, elucidating the shape, size, and optical properties of metallic nanoparticles require using electron microscopy and extinction spectroscopy, respectively. While extinction spectroscopy is relatively simple to perform, imaging nanoparticles using electron microscopy is a much more tedious process. Aside from the multistep workflow to prepare liquid dispersions of metallic nanoparticles for electron microscopy, the amount of time spent on imaging

and post-imaging analyses increases disproportionately with increasing samples. Consequently, the entire nanoparticle characterization workflow is prolonged and further impedes the execution of downstream experiments.

To enhance the efficiency of materials characterization, various machine learning algorithms have been developed to speed up image analyses. These algorithms enable automated recognition of target nanoparticle shapes as well as statistical tabulation of nanoparticle size and size distributions. However, by focusing predominantly on post-imaging analyses, these approaches still necessitate the use of electron microscopic characterization and does not provide significant simplification to the overall workflow of materials characterization. Streamlining the nanomaterial characterization process is thus key to boost the overall research productivity. Herein, in Chapter 4, we introduce machine learning to modernize the materials characterization workflow, employing algorithms to quantitatively predict nanoparticle size and size distributions based on their optical properties from the functional data generated by LSPRs.

1.5 Outline

We end this chapter with an outline of the entire thesis.

In the first part, Chapters 2 and 3, we construct two non-stationary FTS, details of each chapter are as follows.

In Chapter 2, we construct a functional KPSS test based on simple bootstrap method and moving block bootstrap method, which solves the existing problem that many FTS tests require large sample size to insure a high accuracy in the test. Section 2.1 states the problem and introduces the preliminary results of the FKPSS test. Section 2.2 proposes a bootstrap-based functional testing procedure and proves the bootstrap validity on average. The proof of the main result (Theorem 2.2.1) is documented in 2.5. Then, we present simulation studies in Section 2.3, in which we used the model parameters calibrated to a real dataset and compared the proposed bootstrap tests with the Monte Carlo test in Kokoszka and Young (2016). Finally, Section 2.4 concludes.

In Chapter 3, we construct a functional unit root test, which fills the gap of unit root test in FTS model. Section 3.1 states the problem and the statistic of the FURT. Section 3.2 presents the limit distribution of the test statistic under the null hypothesis and the asymptotic behavior of the test statistic under alternative hypotheses. The proof of this section is placed in 3.6 and 3.7. Then, we conduct simulation studies in Section 3.3 and an empirical study of intraday stock price curves in Section 3.4. Finally, Section 3.5 concludes.

Next the second part, in Chapter 4, we focuses on the application of machine learning methods to predict the nanoparticle size and size distribution from functional experimental data of LSPRs. Section 4.1 gives a brief introduction of the generation of experimental data and states the problem. In Section 4.2, we briefly introduce the history and characteristics of different machine learning methods. Section 4.3 introduces two feature engineering methods we use and describes the details of feature extraction. The procedure of prediction, the comparison and analysis of results are presented in Section 4.4. Section 4.5 concludes.

The last part of the thesis, that is, Chapter 5, concludes all the works which were done in this thesis, and gives a brief introduction of the future work, mainly about the functional co-integration test.

A Bootstrap-based KPSS Test for Functional Time Series [†]

In this chapter, we will apply bootstrap method to build an advanced version of functional KPSS test. It is inspired mainly by the construction of functional stationary test in Horváth et al. (2014) and functional KPSS test in Kokoszka and Young (2016). These two tests both require a large sample size, at least 200 for many alternative hypothesis, for a relatively good power. Meanwhile, inspired by the bootstrap-based unit root test of panel data in Chang (2004) and bootstrap-based test of equality of mean functions and covariance operator in Paparoditis and Sapatinas (2016), we consider using bootstrap method to optimize the existing functional KPSS test.

The difficulty of bootstrap theory is that we need to study of the probability space not on the probability space from original sample data. It is a new probability space conditional on the original probability space, and the two probability spaces are also interactive. In fact, because of the feature of original functional KPSS test statistic structure (from partial sum of estimated noises), it is very difficult for us to apply existing limit theorem directly on statistic, as in Paparoditis and Sapatinas (2016). We are inspired from Berkes et al. (2013), try to break down the core issue into many

[†]This chapter is based on the paper Chen and Pun (2019):“Chen, Y. & Pun, C. S. (2019). A bootstrap-based KPSS test for functional time series. *J. Multivariate Anal.* **174**, 104535.”

steps. Many existing theorems based on basic probability space can be extended to bootstrap conditional probability space, see the Theorem 2.5.2 for details. Certainly, it is based on the condition of *i.i.d.* noises. This is the premise of our simple bootstrap. Meanwhile, we have also considered the moving block bootstrap method for dependent noises. These methods are all better than the original functional KPSS test in Kokoszka and Young (2016) in the simulation study, especially when the sample size is small. Such a good result of bootstrap-based test completes our initial vision.

2.1 Problem Formulation and Preliminary Results

2.1.1 The Problem

Suppose that time series data of length N are collected in functional form (curve), denoted by $\mathcal{X}_N = \{X_n(t)\}_{n=1}^N$, where the index t is the argument of the curves. Some of the most natural and common FTS are intra-day price curves and yield curves, for which t is the time within a trading day (measured in seconds or minutes) and time to maturity (measured in years), respectively. Without loss of generality, we assume that the curves are being normalized and defined on the unit interval $[0, 1]$. Throughout this chapter, all the random functions and functional parameters are defined on a Hilbert space $L^2([0, 1])$ of all square-integrable real-valued functions on $[0, 1]$ with an inner product $\langle f, g \rangle := \int_0^1 f(t)g(t)dt$ for $f, g \in L^2([0, 1])$. To ease the notational burden, we suppress the limits of the integral, that is $\int f(t)dt := \int_0^1 f(t)dt$.

Following Kwiatkowski et al. (1992), we consider the null hypothesis of trend stationarity, which is stated as follows:

$$H_0^{ts} : X_n(t) = \mu(t) + n\xi(t) + \eta_n(t), \quad t \in [0, 1], \text{ for } n \in \{1, \dots, N\}, \quad (2.1)$$

where $\mu(t)$ and $\xi(t)$ are deterministic functional parameters of the intercept and the slope of the trend, respectively and $\{\eta_n(t)\}_{n=1}^N$ are identically distributed random functions defined on $(\Omega, \mathcal{F}, \Pr)$ satisfying $E(\eta_n) = 0$ and $E(\|\eta_n\|^{2+\delta}) < \infty$ for some $\delta > 0$, in which $E(\cdot)$ is the expectation operator defined on $(\Omega, \mathcal{F}, \Pr)$ and $\|\eta_n\| :=$

$\{\int \eta_n^2(t)dt\}^{1/2}$. Moreover, we assume the errors $\{\eta_n(t)\}_{n=1}^N$ to be Bernoulli shifts¹ as in Kokoszka and Young (2016).

We aim to devise a significance test with the null hypothesis H_0^{ts} using the bootstrap approach, initiated in Efron (1979); Efron and Tibshirani (1993); Künsch (1989). Since the bootstrap samples are generated regardless of the true data-generating mechanism, our bootstrap-based tests are expected to have non-trivial powers against many alternative hypotheses, including

trend stationary plus random walk, considered in Kokoszka and Young (2016):

$$H_A^{ts+rw} : X_n(t) = \mu(t) + n\xi(t) + \sum_{i=1}^n u_i(t) + \eta_n(t), \quad t \in [0, 1], \text{ for } n \in \{1, \dots, N\}, \quad (2.2)$$

where $\{u_i(t)\}_{i=1}^N$ are i.i.d. random functions satisfying $E(u_i) = 0$ and $E(\|u_i\|^2) < \infty$ for any i ;

unit root, the initiative of Kwiatkowski et al. (1992):

$$H_A^{ur} : X_n(t) = X_{n-1}(t) + \eta_n(t), \quad t \in [0, 1], \text{ for } n \in \{2, \dots, N\}; \quad (2.3)$$

change point, considered in Horváth et al. (2014):

$$H_A^{cp} : X_n(t) = \mu(t) + \delta(t)I_{[n > \lfloor N\tau \rfloor]} + \eta_n(t), \quad t \in [0, 1], \text{ for } n \in \{1, \dots, N\} \quad (2.4)$$

with some $0 < \tau < 1$ and fixed functional parameter $\delta(t)$ and I is the indicator function.

Both the null and the alternative hypotheses suppose that the functions $\mu(t)$ and $\xi(t)$ are unknown and have to be estimated before testing. Among the statistical tests of FTS

¹Specifically, we assume that $\eta_n = f(\epsilon_n, \epsilon_{n-1}, \dots)$ for some measurable function $f : \Omega^\infty \rightarrow L^2$ and independent and identically distributed (i.i.d.) functions ϵ_n , $\forall n$ defined on $(\Omega, \mathcal{F}, \text{Pr})$. The sequence $\{\eta_n(t)\}_{n=-\infty}^\infty$ can be approximated by l -dependent sequences $\{\eta_{n,l}(t)\}_{n=-\infty}^\infty$ that $\sum_{l=1}^\infty (E\|\eta_n - \eta_{n,l}\|^{2+\delta})^{1/\kappa} < \infty$ for some $\kappa > 2 + \delta$, where $\eta_{n,l}$ is defined by $\eta_{n,l} = f(\epsilon_n, \dots, \epsilon_{n-l+1}, \epsilon_{n-l}^*, \epsilon_{n-l-1}^*, \dots)$ with the ϵ^* 's being the independent copies of ϵ_0 and independent of the ϵ 's.

in the literature, Horváth et al. (2014) consider the stationarity hypothesis against H_0^{ts} and Kokoszka and Young (2016) consider H_0^{ts} against H_A^{ts+rw} ((2.1) versus (2.2)). They derive the limiting distributions of a FKPSS test statistic that generalizes Kwiatkowski et al. (1992) under different null hypotheses, which provides a benchmark for our bootstrap-based testing in the large-sample case.

2.1.2 Preliminary Asymptotic Results of the FKPSS Test

In this subsection, we introduce some notations and estimators of the functional parameters, which are components of the FKPSS test statistic based on original samples and its limiting distribution. Hereafter, we emphasize $\mathcal{X}_N$ as the collection of the original samples to distinguish it from the bootstrap samples in the next section.

We define the long-run covariance function of the errors $\{\eta_n\}$ as

$$c(t, s) = E\{\eta_0(t)\eta_0(s)\} + \sum_{n=1}^{\infty} [E\{\eta_0(t)\eta_n(s)\} + E\{\eta_0(s)\eta_n(t)\}], \quad (2.5)$$

where $\eta_0(t)$ is identically distributed as $\{\eta_n(t)\}_{n=1}^N$. From Lemma 2.2 in Berkes et al. (2013), the infinite sum in the definition of $c(t, s)$ is well defined on $L^2([0, 1] \times [0, 1])$, namely $\iint c^2(t, s)dsdt < \infty$. Moreover, the bivariate function $c(t, s)$ is positive-definite and therefore there exist eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ and the corresponding orthonormal eigenfunctions $\phi_1(t), \phi_2(t), \dots, t \in [0, 1]$ satisfying $\lambda_i \phi_i(t) = \int c(t, s)\phi_i(s)ds$ for $i \geq 1$. The eigen-decomposition of $c(t, s)$ plays a crucial role in the FKPSS tests.

Next, we elaborate on the FKPSS test statistic within the framework of Kokoszka and Young (2016). To estimate the error $\eta_n(t)$ from the samples, we first estimate the functional parameters $\xi(t)$ and $\mu(t)$ under H_0^{ts} using the least squares method. The least squares estimator of $\xi(t)$ is

$$\hat{\xi}(t) = \frac{1}{s_N} \sum_{n=1}^N \left(n - \frac{N+1}{2}\right) X_n(t), \quad s_N = \sum_{n=1}^N \left(n - \frac{N+1}{2}\right)^2 = \frac{N(N^2-1)}{12}, \quad (2.6)$$

and the least squares estimator of $\mu(t)$ is

$$\hat{\mu}(t) = \bar{X}_N(t) - \hat{\xi}(t) \left(\frac{N+1}{2}\right), \quad \bar{X}_N(t) = \frac{1}{N} \sum_{n=1}^N X_n(t). \quad (2.7)$$

Consequently, the estimates of the errors $\eta_n(t)$ are given for $n \in \{1, \dots, N\}$ by

$$\begin{aligned}\hat{\eta}_n(t) &= X_n(t) - \hat{\mu}(t) - n\hat{\xi}(t) = (X_n(t) - \bar{X}_N(t)) - \hat{\xi}(t) \left(n - \frac{N+1}{2}\right) \\ &= X_n(t) + \left\{ \left(n - \frac{N+1}{2}\right) \frac{6}{N-1} - 1 \right\} \bar{X}_N(t) \\ &\quad - \frac{12}{N(N^2-1)} \left(n - \frac{N+1}{2}\right) \sum_{i=1}^N iX_i(t).\end{aligned}\tag{2.8}$$

for $n = 1, 2, \dots, N$. Then, a suitable FKPSS test statistic of FTS is given by

$$R_N = \iint Z_N^2(x, t) dt dx, \quad Z_N(x, t) = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \hat{\eta}_n(t).\tag{2.9}$$

Note that Z_N is the partial sum of the estimates of the errors.

As discussed by Kokoszka and Young (2016), in the scalar case, the KPSS test statistic in Kwiatkowski et al. (1992) is equal to the test statistic R_N in (2.9) with a scaling factor of a consistent estimate of the long-run (scalar) variance. In the FTS case, the information of the long-run covariance function, in terms of the eigenvalues, is absorbed into the limiting distribution. Hence, the construction of (2.9) shares the same view as the classic KPSS test in Kwiatkowski et al. (1992). The limiting distribution of the test statistic R_N is established by Kokoszka and Young (2016) and analogous to Kwiatkowski et al. (1992), it is characterized by a second-level Brownian bridge $V(x)$, which is defined as follows:

$$V(x) = W(x) + (2x - 3x^2)W(1) + (-6x + 6x^2) \int_0^1 W(y) dy, \quad x \in [0, 1], \tag{2.10}$$

where $W(x)$ is a standard Wiener process on $[0, 1]$.

Theorem 2.1.1 (Theorem 3.1 in Kokoszka and Young (2016)). Under the null hypothesis H_0^{ts} in (2.1), we have

$$R_N \xrightarrow{d} R_\infty := \sum_{i=1}^{\infty} \lambda_i \int V_i^2(x) dx$$

as $N \rightarrow \infty$, where $\lambda_1, \lambda_2, \dots$ are the eigenvalues of the long-run covariance function $c(t, s)$ in (2.5) and $V_1, V_2, \dots$ are the i.i.d. second-level Brownian bridges defined by (2.10).

The existing FKPSS test is constructed based on the limiting distribution in Theorem 2.1.1. Specifically, the FKPSS test with significance level α makes the following decision: we reject H_0^{ts} if $R_N > d_{N,\alpha}$, where $d_{N,\alpha}$ is the critical value of the test given by $\Pr_{H_0}(R_N > d_{N,\alpha}) = \alpha$. In Kokoszka and Young (2016), the critical value $d_{N,\alpha}$ is approximated by the $(1 - \alpha)$ -quantile of R_∞ . However, such a test may only be valid when the number of observations N is sufficiently large. In this study, we consider a practical alternative to bootstrap-based functional testing, which works well even when N is small. The established Theorem 2.1.1 serves as a benchmark when we validate the proposed bootstrap-based testing procedure.

2.2 Asymptotic Theory of the Bootstrap Test Statistic

In this section, we introduce a practical FKPSS testing procedure with the bootstrap method and validate its consistency in estimating the critical values of interest. We propose different bootstrap schemes for two cases: independent errors and stationary errors. To choose the appropriate scheme to use, we conduct a preliminary test of the error correlation as in Gabrys et al. (2010).

2.2.1 Bootstrap-based FKPSS Testing

We propose two types of bootstrap schemes according to the dependence of the errors; these only differ in Step 2 of the following five-step procedure. If we know that the errors are independent, as stated in Step 2 (SB), we adopt an SB scheme following Efron (1979). Otherwise, in Step 2 (MBB), we instead adopt an MBB scheme following Künsch (1989). Our bootstrap-based functional testing procedure is illustrated as follows.

Step 1. Compute the error estimates $\{\hat{\eta}_n(t)\}_{n=1}^N$ for the observed FTS data $\mathcal{X}_N$ by (2.8).

Step 2. (SB) *If the errors are independent*, generate bootstrap errors $\{\eta_n^*(t)\}_{n=1}^N$ by

resampling the error estimates $\{\hat{\eta}_n(t)\}_{n=1}^N$ with replacement.

(MBB) *If the errors are stationary*, set a length of block $b = b(N)$ and split the estimated errors $\{\hat{\eta}_n(t)\}_{n=1}^N$ into $N - b + 1$ blocks of length b :

$$\{\hat{\eta}_1(t), \hat{\eta}_2(t), \dots, \hat{\eta}_b(t)\}, \{\hat{\eta}_2(t), \hat{\eta}_3(t), \dots, \hat{\eta}_{b+1}(t)\}, \dots, \\ \{\hat{\eta}_{N-b+1}(t), \hat{\eta}_{N-b+2}(t), \dots, \hat{\eta}_N(t)\}.$$

Then, draw $\lceil N/b \rceil$ blocks at random with replacement and align these $\lceil N/b \rceil$ blocks in the order they were picked. The first N samples of the series are defined as the bootstrap errors $\{\eta_n^*(t)\}_{n=1}^N$.

Step 3. Generate bootstrap functional pseudo-observations $\mathcal{X}_N^* = \{X_n^*(t)\}_{n=1}^N$ according to H_0^{ts} :

$$X_n^*(t) = \hat{\mu}(t) + n\hat{\xi}(t) + \eta_n^*(t).$$

Step 4. Compute the error estimates for the bootstrap pseudo-observations $\mathcal{X}_N^*$, denoted by $\{\hat{\eta}_n^*(t)\}$, by (2.8). Note that $\hat{\eta}_n^* \neq \eta_n^*$ in general. Then, a bootstrap replicate of the KPSS test statistic R_N is given by (2.9), with $\{\hat{\eta}_n(t)\}$ replaced by $\{\hat{\eta}_n^*(t)\}$, denoted by R_N^* .

Step 5. Repeat Steps 2 to 4 B times and obtain a set of B bootstrap replicates of R_N , denoted by $\{R_N^{*(b)}\}_{b=1}^B$. Then, the bootstrap estimate of the critical value $d_{N,\alpha}$ is given by the $(1 - \alpha)$ -percentile of $\{R_N^{*(b)}\}_{b=1}^B$, denoted by $\hat{d}_{N,\alpha}^*$. We reject H_0^{ts} if $R_N > \hat{d}_{N,\alpha}^*$.

2.2.2 Bootstrap Validity on Average

In this section, we provide the bootstrap consistency results for the SB method. To the best of our knowledge, the existing theoretical results of bootstrap-based functional tests are mostly restricted to independent observations and the SB; see Benko et al. (2009); Paparoditis and Sapatinas (2016). Pilavakis et al. (2017) investigate the K -sample mean problem for weakly dependent observations using the stationary bootstrap. Both

Paparoditis and Sapatinas (2016) and Pilavakis et al. (2017) adopt the form of the L^2 -type test statistic proposed by Horváth et al. (2013) and the bootstrap central limit theorem in Politis and Romano (1994) to show the bootstrap validity. However, the test statistic we considered in (2.9) is a Lagrange multiplier test statistic, whose bootstrap replicate does not fit into the analytical frameworks of the aforementioned studies.

To show the consistency results of the bootstrap FKPSS test statistic, we consider the SB method and prove the asymptotic bootstrap validity in a broader (on average) sense, where the averaged (random) limiting bootstrap distribution is consistent with the limiting distribution of the sample test statistic; see Cavaliere and Georgiev (2017) and Bücher and Kojadinovic (2018). According to Lemma 2.3 in Bücher and Kojadinovic (2018), we have the following (heuristic) equivalent relation:

$$\begin{aligned} & \text{“Conditional weak convergence of } R_N^* \text{ (conditional bootstrap validity)”} \\ \iff & \text{“Unconditional weak convergence of } R_N^* \text{ (bootstrap validity on average)”} \\ + & \text{“}\bar{\mathbb{E}}[f\{R_N^{*(1)}\}f\{R_N^{*(2)}\}] - \bar{\mathbb{E}}[f\{R_N^{*(1)}\}]\bar{\mathbb{E}}[f\{R_N^{*(2)}\}] \rightarrow 0, \text{ as } N \rightarrow \infty\text{”} \end{aligned} \quad (2.11)$$

for all bounded and Lipschitz-continuous f , where $R_N^{*(1)}$ and $R_N^{*(2)}$ are any two different bootstrap replicates of R_N . $\bar{\mathbb{E}}(\cdot)$ is the expectation operator defined in the general (unconditional) probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}}_r)$, which will be introduced in detail in this section.

Using the probabilistic arguments in Bücher and Kojadinovic (2018), we aim to prove the unconditional weak convergence of the bootstrap replicate of the FKPSS test statistic R_N^* (to R_∞), which is less involved but loses conditional inference. Recall the verifiable point in Cavaliere and Georgiev (2017) that although a conditional inference of the bootstrap is not warranted, the bootstrap may still be asymptotically valid in a broader sense. Such an unconditional weak convergence result is interesting in its own right.

The major mathematical merit of this section is to extend the limit theorems for a Lagrange multiplier statistic in the (empirical) probability space to the bootstrap counterpart in an enlarged (general) probability space that nests the randomness of the bootstrapping as well. To this end, we need to extend the results in Berkes et al. (2013)

to the general probability space. Some innovative lemmas and theorems for proving the main theorem (Theorem 2.2.1) are provided in Appendix 2.5.

Since we aim to prove the bootstrap validity for the SB method, we impose the i.i.d. assumption and a moment condition on the errors:

Assumption 2.2.1. The errors $\{\eta_n\}_{n=1}^N$ are i.i.d. random functions satisfying $E(\eta_n) = 0$ and $E(\|\eta_n\|^4) < \infty$.

To proceed, we introduce the bootstrap probability space as in Park (2002) and Chang (2004). Consider a realization $\mathcal{X}_\infty := \{X_1(t), X_2(t), \dots\}$ defined on the probability space $(\Omega, \mathcal{F}, \Pr)$. We define the bootstrap probability space $(\Omega^*, \mathcal{F}^*, \Pr^*)$ conditional on the given realization $\mathcal{X}_\infty$. The conditional expectation operator with respect to $\Pr^*$ is denoted by $E^*(\cdot)$. For each realization $\mathcal{X}_\infty$, the bootstrap errors $\{\eta_n^*(t)\}_{n=1}^N$, which are sampled from $\{\hat{\eta}_n(t)\}_{n=1}^N$ and defined in Section 2.2.1, are regarded as random variables on $(\Omega^*, \mathcal{F}^*, \Pr^*)$ for any sample size N . Specifically, the randomness of $\{\eta_n^*(t)\}_{n=1}^N$ comes from bootstrapping, conditional on $\{\hat{\eta}_n(t)\}_{n=1}^N$. Moreover, under the conditional probability measure $\Pr^*$, the bootstrap errors $\{\eta_n^*(t)\}_{n=1}^N$ become i.i.d. samples from the empirical distribution of $\{\hat{\eta}_n(t)\}_{n=1}^N$. Now, $\{R_N^*\}$ can be viewed as a sequence of random variables defined on $(\Omega^*, \mathcal{F}^*, \Pr^*)$ that facilitates the later analyses on the N -asymptotics of R_N^* . For a sequence of random variables $\{Q_N\}$ on $(\Omega^*, \mathcal{F}^*, \Pr^*)$, we denote by $Q_N \stackrel{a.s.}{=} o_{\Pr^*}(1)$ (resp. $Q_N \stackrel{\Pr}{=} o_{\Pr^*}(1)$) if $\Pr^*(|Q_N| > \epsilon) \rightarrow 0$ almost surely (resp. in probability $\Pr$) for any $\epsilon > 0$ when $N \rightarrow \infty$. Moreover, we denote the weak convergence of conditional law by $Q_N \xrightarrow{d^*} Q$ a.s. (resp. in probability $\Pr$) if $E^*\{f(Q_N)\} \xrightarrow{a.s.} E\{f(Q)\}$ (resp. $E^*\{f(Q_N)\} \xrightarrow{\Pr} E\{f(Q)\}$) when $N \rightarrow \infty$ for any bounded and continuous function f , where the limiting random variable Q has a distribution independent of the original sample realization.

Next, we introduce a general (unconditional) probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\Pr})$, which essentially combines $(\Omega, \mathcal{F}, \Pr)$ and $(\Omega^*, \mathcal{F}^*, \Pr^*)$. More specifically, $\bar{\Omega} = \Omega \times \Omega^*$, $\bar{\mathcal{F}} = \mathcal{F} \otimes \mathcal{F}^*$ (tensor-product σ -field), and for any event $\bar{f} \in \bar{\mathcal{F}}$, $\bar{\Pr}$ is defined as $\bar{\Pr}(\bar{f}) := E\{\Pr^*(\bar{f})\}$. It is easy to verify that $\bar{\Pr}$ is a probability measure. Moreover, the conditional probability measure $\Pr^*$ can be interpreted as $\Pr^*(\cdot) = \bar{\Pr}(\cdot | \mathcal{X}_\infty)$ and

for any event $f \in \mathcal{F}$, we have $\bar{\text{Pr}}(f) = \text{Pr}(f)$. The expectation operator with respect to $\bar{\text{Pr}}$ is denoted by $\bar{\text{E}}$ and we have $\bar{\text{E}}(\cdot) = \text{E}\{\text{E}^*(\cdot)\}$. Similarly, we use $\xrightarrow{\bar{\text{Pr}}}$ and $\xrightarrow{\bar{d}}$ to denote the convergence in probability and in distribution, respectively for random variables defined on $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\text{Pr}})$.

For the random variables $\{X_n\}_{n=1}^\infty$ on $(\Omega, \mathcal{F}, \text{Pr})$, we have

$$\bar{\text{Pr}}(|X_n| > \epsilon) = \text{E}\{\text{Pr}^*(|X_n| > \epsilon)\} = \text{E}\{\mathbf{1}_{(|X_n| > \epsilon)}(\omega)\} = \text{Pr}(|X_n| > \epsilon) \quad \text{for any } \epsilon > 0.$$

Hence, we have the following equivalent relation: $X_n = o_{\bar{\text{Pr}}}(1) \Leftrightarrow X_n = o_{\text{Pr}}(1)$. Similar arguments yield that $X_n = o_{\bar{\text{Pr}}}(1) \Leftrightarrow X_n = o_{\text{Pr}}(1)$, $X_n = \mathcal{O}_{\bar{\text{Pr}}}(1) \Leftrightarrow X_n = \mathcal{O}_{\text{Pr}}(1)$ and $X_n \xrightarrow{d} X \Leftrightarrow X_n \xrightarrow{\bar{d}} X$.

Now, we are ready to present our main theorem. Theorem 2.2.1 shows the weak convergence of the test statistic R_N^* with the SB scheme and validates the bootstrap consistency.

Theorem 2.2.1. Under the null hypothesis H_0^{ts} in (2.1) and Assumption 2.2.1, we have

$$R_N^* \xrightarrow{\bar{d}} R_\infty = \sum_{i=1}^{\infty} \lambda_i \int V_i^2(x) dx$$

as $N \rightarrow \infty$, where R_∞ , $\{\lambda_i\}$, and $\{V_i\}$ are given in Theorem 2.1.1.

Appendix 2.5 provides the proof of Theorem 2.2.1. In this proof, we develop many novel results on the convergence in the unconditional probability space. Theorems 2.1.1 and 2.2.1 show that the bootstrap test statistic R_N^* has the same limiting distribution as R_N , which validates the consistency between using the critical values of R_N^* and R_N . As discussed by Kokoszka and Young (2016), R_N under H_A^{ts+rw} increases at order N^2 and the critical value $d_{N,\alpha}$ increases at an order not greater than N . From Slutsky's theorem, the power of our bootstrap-based FKPSS test thus approaches unity.

This study considers the bootstrap validity only for the SB method, leaving the validity of the MBB an open question. Indeed, certain properties of the SB method play an important role in our proof. For example, in Lemma 2.5.1, the first inequality in the proof follows from Lemma 2.3 of Berkes et al. (2013), where we require $\{\bar{\eta}_{n,K}^*\}_{n=1}^N$

to be i.i.d. random variables on the bootstrap probability space and $E^*\{\bar{\eta}_{1,K}^*(t)\} = 0$. However, these conditions do not hold under the MBB method when the block length $b > 1$ irrespective of whether the noises of the original sample data are independent. Therefore, more novel mathematical tools are required to extend our theoretical results to the MBB method, which is beyond the scope of this study. However, we still include the MBB in our experiments to examine its performance numerically.

At last, we focus on the condition (2.11), which implies that $R_N^{*(1)}$ and $R_N^{*(2)}$ are independent of $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\text{Pr}})$ in the N -asymptotic sense. Hence, it is worth investigating the joint statistical properties of $R_N^{*(1)}$ and $R_N^{*(2)}$. To proceed in this research direction, the following proposition shows a special case of (2.11) with $f(x) = x$.

Proposition 2.2.1. Let $R_N^{*(1)}$ and $R_N^{*(2)}$ be any two different bootstrap replicates of R_N based on the same $\mathcal{X}_N$, generated by Steps 1–4 in Section 2.2.1. Under the null hypothesis H_0^{ts} in (2.1) and Assumption 2.2.1, we have

$$\bar{E}\{R_N^{*(1)}R_N^{*(2)}\} - \bar{E}\{R_N^{*(1)}\}\bar{E}\{R_N^{*(2)}\} \rightarrow 0, \text{ as } N \rightarrow \infty. \quad (2.12)$$

Proposition 2.2.1 asserts that $R_N^{*(1)}$ and $R_N^{*(2)}$ are uncorrelated on $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\text{Pr}})$ in the N -asymptotic sense. The key mechanism behind the proof of (2.12) or even (2.11) relies on the form of the test statistic considered.

2.3 Simulation Studies

In this section, we present the simulation studies conducted to examine the finite-sample performance of our proposed bootstrap-based FKPSS tests, whose procedure is stated in Section 2.2.1. We also compare the bootstrap-based FKPSS test with the benchmark FKPSS test proposed by Kokoszka and Young (2016), whose implementation is based on the limiting distribution in Theorem 2.1.1 and the Monte Carlo method.

To resemble the real-data analysis, we adopt empirical data on the monthly US Treasury yield curves from January 1970 to December 2002 in R package ‘fcds’ (Shang and Hyndman (2013)), denoted by $\{X_n(t)\}_{n=1}^{396}$. These curves have been normalized

and defined on $[0, 1]$. Then, we simulate multiple sets of functional data of length N , denoted by $\{\widetilde{X}_n(t)\}_{n=1}^N$, under different hypotheses with the model parameters calibrated to the real data. We consider different sample sizes ($N = 30, 50, 70, 100, 150$, and 200) to investigate the effect of sample size on the FKPSS tests. For each hypothesis in (2.1)–(2.4) and each sample size N , we generate 1000 sets of the simulated data and evaluate the testing performance based on the empirical size and power of the bootstrap-based and Kokoszka and Young (2016)’s FKPSS tests. In all computations, we use 240 equidistant points to estimate all functional data (i.e., the curve values at $t_i = i/240$, $i = 1, \dots, 240$). To illustrate the proposed SB and MBB schemes in Section 2.2.1, we split our studies into two cases according to the dependence of the errors.

To compute the benchmark results, we strictly follow Kokoszka and Young (2016)’s guideline to implement their tests using the Monte Carlo method, including the choice of smoothing bandwidth $h = N^{2/5}$ and kernel function:

$$K(t) = \begin{cases} 1, & 0 \leq |t| < 0.1, \\ 1.1 - |t|, & 0.1 \leq |t| < 1.1, \\ 0, & |t| \geq 1.1. \end{cases}$$

As suggested by Kokoszka and Young (2016), when the errors are independent, we may only estimate the covariance function $E\{\eta_0(t)\eta_0(s)\}$ rather than $c(t, s)$ in the FKPSS test, whose results are labeled “iKY” in this chapter. This can be compared with our SB results, as both are based on Assumption 2.2.1. However, the dependence of the errors may not be certain in advance, forcing us to use the general FKPSS test that estimates the long-run covariance function $c(t, s)$ in (2.5), whose results are labeled “KY.” This can be compared with our MBB results. Hence, in our first study with independent errors (Section 2.3.2), two pairs of tests are compared, namely “SB versus iKY” and “MBB versus KY,” while in our second study with stationary errors (Section 2.3.3), we compare MBB with KY.

Our bootstrap schemes set the number of bootstrap replicates as $B = 200^2$. For the

²Our preliminary tests show that increasing B (from 200) does not improve the results significantly

MBB scheme, we adopt the block length³ of $b = \lceil (2N)^{1/3} \rceil$, which produces robust results for our studies. To see the impacts of the block length and smoothing bandwidth on test performance, we also include the results with $b = \lceil N^{2/5} \rceil$ and $h = (2N)^{1/3}$ in our comparisons.

2.3.1 Simulation Setup

In the simulations, we use the model parameters estimated with the real data $\{X_n(t)\}_{n=1}^{396}$ under the corresponding hypothesis, while the simulated errors $\{\tilde{\eta}_n(t)\}_{n=1}^N$ are generated as additive processes with the convention $\tilde{\eta}_n(0) \equiv 0$. The error simulation is illustrated as follows. Let $\hat{\sigma}_i^2$ be the sample variance of the estimated error increments $\{\hat{\eta}_n(t_i) - \hat{\eta}_n(t_{i-1})\}_{n=1}^{396}$ for $i = 1, \dots, 240$. Section 2.3.2 considers three types of independent errors, while Section 2.3.3 considers stationary errors following the AR(1) model.

Specifically, in Section 2.3.2, we set $\tilde{\eta}_n(t_i) - \tilde{\eta}_n(t_{i-1}) = \kappa_i \zeta_n$, $\forall i, n$, where $\{\zeta_n\}_{n=1}^N$ are i.i.d. samples from the distributions of (a) standard normal, (b) Student's t with 5 degrees of freedom (df), (c) de-meaned F with 10 and 8 degrees of freedom, and κ_i is a scaling parameter to enforce the variance of $\tilde{\eta}_n(t_i) - \tilde{\eta}_n(t_{i-1})$ to be $\hat{\sigma}_i^2$ for a fixed i . We remark that (a) and (b) are symmetric distributions and (c) is a non-symmetric distribution.

In Section 2.3.3, we generate the stationary errors as follows for $n \in \{2, \dots, N\}$:

$$\begin{cases} \tilde{\eta}_1(t_i) - \tilde{\eta}_1(t_{i-1}) &= Z_{1,i}, \\ \tilde{\eta}_n(t_i) - \tilde{\eta}_n(t_{i-1}) &= 0.5\{\tilde{\eta}_{n-1}(t_i) - \tilde{\eta}_{n-1}(t_{i-1})\} + Z_{n,i}, \end{cases}$$

where $Z_{1,i} \sim \mathcal{N}(0, \hat{\sigma}_i^2)$, $Z_{n,i} \sim \mathcal{N}(0, \frac{3}{4}\hat{\sigma}_i^2)$ for $n \geq 2$, and $\{Z_{n,i}\}_{i \in \{1, \dots, 240\}}^{n \in \{1, \dots, N\}}$ are independent.

and does not provide new insights for this study. We use $B = 200$ to reduce the processing time.

³Although the theoretical study about the block length in the MBB of the FKPSS test statistic is unexplored and beyond the scope of this study, please refer to the following studies: Lahiri (1999), (for parameter estimation), Zvingelis (2003) (for coverage probability), and Pilavakis et al. (2017) (for K -sample functional tests), which suggested $b^{-1} + bN^{-1/2} = o(1)$. We choose $b \propto N^{1/3}$ following Politis and White (2004) and the scaling factor $2^{1/3}$ is prefixed to guarantee a significant number of blocks and stable results.

The generation of the simulated data under different hypotheses is summarized as follows.

Under H_0^{ts} in (2.1), we estimate $\hat{\xi}(t)$, $\hat{\mu}(t)$, and $\{\hat{\eta}_n(t)\}_{n=1}^{396}$ using (2.6), (2.7), and (2.8), respectively. Then, we simulate the errors $\{\tilde{\eta}_n(t)\}_{n=1}^N$ based on $\{\hat{\eta}_n(t)\}_{n=1}^{396}$. The simulated data under H_0^{ts} are constructed as

$$\widetilde{X}_n^{ts}(t) := \hat{\mu}(t) + n\hat{\xi}(t) + \tilde{\eta}_n(t), \quad t \in [0, 1], \quad n \in \{1, \dots, N\}.$$

Under H_A^{ts+rw} in (2.2), we estimate $\hat{\xi}(t)$, $\hat{\mu}(t)$, and $\{\hat{\eta}_n(t)\}_{n=1}^{396}$ using (2.6), (2.7), and (2.8), respectively. Then, we simulate the errors $\{\tilde{\eta}_n(t)\}_{n=1}^N$ based on $\{\hat{\eta}_n(t)\}_{n=1}^{396}$. Following the framework of Kokoszka and Young (2016), we generate the increments $u_i(t)$ as follows:

$$u_i(t) = N_{i1} \sin(\pi t) + N_{i2} \sin(2\pi t), \quad t \in [0, 1], \quad i \in \{1, 2, \dots, N\},$$

where N_{ij} , $i \in \{1, \dots, N\}$, $j = 1, 2$ are i.i.d. standard normal random variables.

The simulated data under H_A^{ts+rw} are constructed as

$$\widetilde{X}_n^{ts+rw}(t) := \hat{\mu}(t) + n\hat{\xi}(t) + \sum_{i=1}^n u_i(t) + \tilde{\eta}_n(t), \quad t \in [0, 1], \quad n \in \{1, 2, \dots, N\}.$$

Under H_A^{ur} in (2.3), we estimate the errors $\hat{\eta}_n(t) := X_{n+1}(t) - X_n(t)$, $n \in \{1, \dots, 395\}$.

Then, we simulate the errors $\{\tilde{\eta}_n(t)\}_{n=1}^{N-1}$ based on $\{\hat{\eta}_n(t)\}_{n=1}^{395}$. The simulated data under H_A^{ur} are constructed as

$$\widetilde{X}_1^{ur}(t) := X_1(t) \quad \text{and} \quad \widetilde{X}_{n+1}^{ur}(t) := \widetilde{X}_n^{ur}(t) + \tilde{\eta}_n(t), \quad t \in [0, 1], \quad n \in \{1, \dots, N-1\}.$$

Under H_A^{cp} in (2.4), we first define the change point as follows. For every $j \in \{1, \dots, 395\}$, we divide the dataset $\{X_n(t)\}_{n=1}^{396}$ into two datasets: $\mathcal{X}_1^{(j)} := \{X_n(t)\}_{n=1}^j$ and $\mathcal{X}_2^{(j)} := \{X_n(t)\}_{n=j+1}^{396}$. For $j \in \{1, \dots, 395\}$, we estimate $\hat{\mu}_k^{(j)}(t)$, $k = 1, 2$ using the sample mean of the functional data $\mathcal{X}_k^{(j)}$ and the errors by

$$\hat{\eta}_n^{(j)}(t) = \begin{cases} X_n(t) - \hat{\mu}_1^{(j)}(t), & 1 \leq n \leq j, \\ X_n(t) - \hat{\mu}_2^{(j)}(t), & j+1 \leq n \leq 396. \end{cases}$$

We choose the change point index as $j^* := \arg \min_j \sum_{n=1}^{396} \int \{\hat{\eta}_n^{(j)}(t)\}^2 dt$. Then, we simulate the errors $\{\tilde{\eta}_n(t)\}_{n=1}^N$ based on $\{\hat{\eta}_n^{(j^*)}(t)\}_{n=1}^{396}$. The simulated data under H_A^{cp} are constructed as

$$\tilde{X}_n^{cp}(t) := \hat{\mu}_1^{(j^*)}(t) + \{\hat{\mu}_2^{(j^*)}(t) - \hat{\mu}_1^{(j^*)}(t)\} I_{(n > \lfloor \frac{Nj^*}{396} \rfloor)} + \tilde{\eta}_n(t), \quad t \in [0, 1], n \in \{1, 2, \dots, N\}.$$

Here are some examples of simulated data under different hypotheses in Figure 2.1. The curves with gradually changing colors from yellow to gray and then to blue show the functional time series in the time order $n = 1$ to 20. It is not difficult to find the differences between them. In the figure of simulated data under H_0^{ts} , an upward tendency over time is very clear. In the figures of simulated data under H_1^{ts+rw} and H_1^{ur} , the data are relatively messy because of the term of functional random walk. In the figure of data under H_1^{cp} , the change function is easy to distinguish between the two sets of functional data.

2.3.2 Simulation Results for the Independent Errors

Tables 2.1, 2.2, and 2.3 present the simulation results for errors that are normally distributed, Student's t -distributed with five degrees of freedom, and F -distributed with 10 and eight degrees of freedom, respectively. The first columns of these tables indicate the hypothesis (model) under which the data are simulated and the second columns indicate the test. We restate that iKY and KY represent the original FKPSS test in Kokoszka and Young (2016), with independent and dependent errors of sample data respectively. In parentheses next to MBB (resp. KY) is the block length b (resp. bandwidth h).

For each N , two significance levels, $\alpha = 1\%$ and 5% , are considered. The results, displayed in percentages, are the rejection frequencies by the tests in 1000 simulations. Hence, the results for H_0^{ts} correspond to the empirical sizes and those for the alternative hypotheses correspond to the powers of the tests.

Table 2.1 shows that the empirical sizes of the SB for every N and α are close to the desirable significance levels, which implies that the distributions of the test statistics R_N

Figure 2.1 A simple example of simulated data under different hypotheses when $N = 20$. Figures in the first row show the simulated data under H_0^{ts} (left) and H_A^{ts+rw} (right) respectively. Figures in the second row show the simulated data under H_A^{ur} (left) and H_A^{cp} (right) respectively.

and R_N^* under the null hypothesis coincide even for a small sample size N . Moreover, we see that SB is better than iKY in terms of achieving the desirable significance levels and higher power values when N is less than 100, especially $\alpha = 1\%$. We illustrate this point in greater detail later. When N is large, SB and iKY perform equally well, which is consistent with the requirement of the FKPSS test constructed based on the limiting distribution of the test statistic. The bootstrap-based FKPSS test provides a practical

Table 2.1

Empirical sizes (%) and powers (%) for the different FKPSS tests for FTS with normally distributed errors, different sample sizes $N = 30, 50, 70, 100, 150, 200$ and different significance levels $\alpha = 1\%, 5\%$.

N		30		50		70		100		150		200	
α		1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%
H_0^{ts}	SB	1.3	5.3	2.0	3.8	1.1	5.0	1.4	5.5	1.6	5.6	1.4	6.1
	iKY	0.4	3.5	1.0	3.8	1.0	4.6	0.6	4.4	1.0	4.9	1.3	4.9
	MBB $((2N)^{1/3})$	1.4	6.2	1.2	5.4	0.9	4.5	0.6	4.9	1.3	6.2	1.3	5.8
	MBB $(N^{2/5})$	1.4	6.2	1.2	5.4	0.9	4.5	0.7	5.3	1.5	5.1	1.2	6.2
	KY $((2N)^{1/3})$	0.0	3.4	0.0	4.1	0.4	3.6	0.3	3.2	0.4	4.3	1.0	4.0
	KY $(N^{2/5})$	0.0	3.3	0.0	5.1	0.0	4.4	0.1	3.2	0.2	4.0	0.5	4.4
H_A^{ts+rw}	SB	25.9	48.1	63.5	81.3	85.9	94.6	97.0	99.0	99.8	99.9	100.0	100.0
	iKY	23.2	48.5	61.3	83.4	85.6	94.8	97.0	99.2	99.6	100.0	100.0	100.0
	MBB $((2N)^{1/3})$	15.0	38.0	37.9	64.5	55.0	80.6	76.0	92.1	88.6	97.7	94.3	98.7
	MBB $(N^{2/5})$	15.0	38.0	37.9	64.5	55.0	80.6	71.9	90.2	85.8	95.6	91.0	98.0
	KY $((2N)^{1/3})$	0.0	23.5	9.8	51.7	31.1	66.5	55.5	84.7	80.7	92.6	88.2	97.5
	KY $(N^{2/5})$	0.0	26.3	7.4	48.6	30.8	66.1	51.7	80.8	71.0	91.5	85.1	94.2
H_A^{ur}	SB	77.2	88.8	93.7	98.7	96.8	99.8	99.2	99.8	100.0	100.0	100.0	100.0
	iKY	76.2	88.6	91.5	97.4	97.0	99.8	99.2	99.8	99.8	100.0	100.0	100.0
	MBB $((2N)^{1/3})$	31.7	59.6	40.3	65.5	43.4	71.6	61.3	82.9	70.3	86.5	78.2	90.5
	MBB $(N^{2/5})$	31.7	59.6	40.3	65.5	43.4	71.6	56.0	77.8	65.2	85.5	73.5	88.5
	KY $((2N)^{1/3})$	0.0	31.0	8.5	41.0	21.8	59.8	36.6	68.2	54.5	79.5	65.4	86.2
	KY $(N^{2/5})$	0.0	26.3	5.5	43.8	19.7	53.3	34.3	64.3	47.7	77.3	58.2	83.5
H_A^{cp}	SB	8.5	19.1	12.5	33.4	19.4	40.3	36.1	59.4	56.2	78.4	77.5	91.1
	iKY	3.9	17.4	10.9	26.0	15.9	42.4	30.9	60.8	55.2	79.2	70.3	90.8
	MBB $((2N)^{1/3})$	3.5	14.7	5.5	21.1	6.8	29.5	17.7	44.8	30.9	67.3	48.8	80.6
	MBB $(N^{2/5})$	3.5	14.7	5.5	21.1	6.8	29.5	16.0	43.0	30.0	64.8	46.1	78.9
	KY $((2N)^{1/3})$	0.0	7.5	0.1	14.7	1.5	22.4	7.9	37.6	20.6	61.2	37.9	76.7
	KY $(N^{2/5})$	0.0	7.6	0.1	14.4	1.3	21.6	6.9	36.5	17.2	58.4	32.2	74.4

alternative when N is small or moderate.

The empirical sizes and powers in Table 2.1 show that the MBB test uniformly

outperforms the FKPSS test in Kokoszka and Young (2016) under all alternative hypotheses H_A^{ts+rw} , H_A^{ur} , and H_A^{cp} . In particular, with the H_A^{ur} as in the classic KPSS test (Kwiatkowski et al. (1992)), the powers are satisfactory even when N is small and approach unity as N increases. Further, with different block lengths b and bandwidths h , the results are similar and our conclusion remains. In general, our suggestion of $b = \lceil (2N)^{1/3} \rceil$ performs slightly better than using $b = \lceil N^{2/5} \rceil$ and using $h = (2N)^{1/3}$ is slightly better than using $h = N^{2/5}$, which was suggested by Kokoszka and Young (2016). Our simulation results confirm that the bootstrap-based FKPSS test performs well with a small sample size and requires fewer samples to achieve satisfactory power than under the benchmark, even for H_A^{cp} .

Tables 2.2 and 2.3 show that the bootstrap-based FKPSS test is robust to alternative symmetric and asymmetric error distributions. The conclusions are analogous to those in Table 2.1. Even when N is large, our proposed bootstrap-based FKPSS test generally improves the power of the existing FKPSS test.

To better compare the aggregated performance of SB and iKY, Table 2.4 reports the averaged power increase rate from iKY to SB over the three alternative hypotheses considered, namely $\frac{1}{3} \sum_{H_A} \left(\frac{Power_{H_A}(SB)}{Power_{H_A}(iKY)} - 1 \right)$. SB provides a significant improvement in power when N is small. As expected, their deviations are reduced when N increases. If we consider the significance level $\alpha = 1\%$ (resp. 5%), the overall averaged increase rate is about 10% (resp. 5%) for the symmetric distributions (normal and Student's t). However, when the error distribution is asymmetric, SB is better than iKY in only a few cases of small N and is comparable with iKY for most cases. This finding shows that SB has its own limitations and that the performance of the FKPSS test in Kokoszka and Young (2016) is stable regardless of the error distribution.

2.3.3 Simulation Results for the Stationary Errors

In this section, we examine the performance of the bootstrap-based FKPSS tests when the errors are stationary. Table 2.5 shows the empirical sizes and powers of MBB and KY for the FTS data with AR(1) errors.

Table 2.2

Empirical sizes (%) and powers (%) for the different FKPSS tests for FTS with t_5 -distributed errors, different sample sizes $N = 30, 50, 70, 100, 150, 200$ and different significance levels $\alpha = 1\%, 5\%$.

N		30		50		70		100		150		200	
α		1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%
H_0^{ts}	SB	0.7	4.7	1.0	6.2	0.8	5.7	1.1	4.4	1.5	4.8	0.9	5.9
	iKY	0.6	4.2	0.9	3.8	0.8	4.7	0.8	5.7	0.7	6.5	0.8	5.0
	MBB $((2N)^{1/3})$	1.5	4.2	0.5	5.0	0.8	6.3	1.2	5.4	0.9	3.7	1.6	4.6
	MBB $(N^{2/5})$	1.5	4.2	0.5	5.0	0.8	6.3	0.6	5.1	1.2	4.7	1.0	5.9
	KY $((2N)^{1/3})$	0.0	3.7	0.0	3.3	0.0	3.8	0.0	4.5	0.4	4.9	0.5	4.6
	KY $(N^{2/5})$	0.0	3.0	0.1	3.9	0.1	3.7	0.1	3.0	0.2	3.7	0.4	5.5
H_A^{ts+rw}	SB	24.6	49.3	68.0	82.9	85.7	95.3	96.1	99.2	99.7	99.9	100.0	100.0
	iKY	26.3	48.6	58.9	83.3	87.3	94.0	96.9	98.6	99.9	99.9	100.0	100.0
	MBB $((2N)^{1/3})$	19.4	40.4	36.3	65.4	57.3	79.4	78.2	93.0	88.7	96.7	92.6	98.5
	MBB $(N^{2/5})$	19.4	40.4	36.3	65.4	57.3	79.4	73.2	90.9	86.9	96.9	91.9	98.2
	KY $((2N)^{1/3})$	0.0	25.7	12.2	52.7	32.9	68.1	56.6	83.8	79.3	93.5	88.0	96.7
	KY $(N^{2/5})$	0.0	22.5	6.1	50.8	30.2	66.4	52.0	80.3	72.8	92.0	81.7	95.4
H_A^{ur}	SB	77.5	88.7	90.8	98.8	97.9	99.4	99.6	99.9	99.9	100.0	100.0	100.0
	iKY	74.6	88.7	91.8	98.2	97.0	99.3	99.0	100.0	99.8	100.0	100.0	100.0
	MBB $((2N)^{1/3})$	35.1	55.8	38.3	67.0	46.9	71.6	63.1	80.5	73.6	88.0	78.6	90.8
	MBB $(N^{2/5})$	35.1	55.8	38.3	67.0	46.9	71.6	56.3	78.5	66.2	84.0	75.3	88.4
	KY $((2N)^{1/3})$	0.0	31.2	9.4	44.3	24.2	57.2	38.9	68.8	55.4	80.2	65.7	86.1
	KY $(N^{2/5})$	0.0	29.3	5.3	41.2	22.5	54.2	34.6	62.4	50.8	76.6	58.6	81.3
H_A^{cp}	SB	8.2	18.2	12.6	32.3	21.0	44.2	37.4	57.7	57.4	81.2	75.6	89.5
	iKY	3.8	20.5	11.6	31.1	19.0	43.9	33.4	61.4	57.7	82.4	74.9	89.6
	MBB $((2N)^{1/3})$	4.1	15.2	6.3	22.6	9.8	26.2	17.7	48.2	31.4	67.2	52.2	78.4
	MBB $(N^{2/5})$	4.1	15.2	6.3	22.6	9.8	26.2	14.0	43.3	29.2	64.6	46.4	78.9
	KY $((2N)^{1/3})$	0.0	7.5	0.0	14.5	1.0	26.2	6.1	42.2	23.9	59.9	41.9	79.1
	KY $(N^{2/5})$	0.0	7.5	0.0	13.9	0.7	24.9	4.8	40.2	19.8	57.8	34.6	77.0

Table 2.5 shows that in this case, the empirical sizes of MBB also tend to the desirable significance levels but at a slower rate than KY. However, the powers of MBB are much higher than those of KY and are only comparable when N is very large. We see

Table 2.3

Empirical sizes (%) and powers (%) for the different FKPSS tests for FTS with $F(10, 8)$ -distributed errors, different sample sizes $N = 30, 50, 70, 100, 150, 200$ and different significance levels $\alpha = 1\%, 5\%$.

N		30		50		70		100		150		200	
α		1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%
H_0^{ts}	SB	0.4	3.3	1.5	3.3	0.6	4.1	1.1	4.4	1.4	3.6	0.8	4.8
	iKY	0.7	4.9	1.0	3.7	0.9	4.7	0.7	5.5	0.6	4.0	0.7	4.9
	MBB $((2N)^{1/3})$	1.6	6.1	1.2	5.7	1.1	5.0	1.0	3.7	0.5	5.4	1.1	3.8
	MBB $(N^{2/5})$	1.6	6.1	1.2	5.7	1.1	5.0	0.8	4.6	1.1	5.3	0.9	5.7
	KY $((2N)^{1/3})$	0.0	3.8	0.1	2.9	0.1	4.2	0.1	4.7	0.2	3.8	0.6	3.3
	KY $(N^{2/5})$	0.0	3.9	0.0	2.8	0.0	4.4	0.1	4.9	0.2	3.5	0.5	3.4
H_A^{ts+rw}	SB	36.8	58.0	70.0	83.5	87.5	94.1	95.1	99.2	99.4	99.8	99.9	99.9
	iKY	37.9	61.5	72.4	86.4	86.3	95.7	96.5	99.2	99.6	99.8	99.9	100.0
	MBB $((2N)^{1/3})$	22.5	49.0	40.3	66.8	59.1	81.6	77.6	90.5	89.0	97.2	93.6	99.0
	MBB $(N^{2/5})$	22.5	49.0	40.3	66.8	59.1	81.6	70.6	91.1	85.3	95.8	89.6	97.9
	KY $((2N)^{1/3})$	0.0	29.9	10.2	54.3	36.8	71.4	60.6	83.8	78.1	93.8	88.3	96.6
	KY $(N^{2/5})$	0.0	29.1	9.4	50.7	29.7	67.2	51.1	81.4	72.9	90.0	82.6	96.3
H_A^{ur}	SB	76.7	89.5	93.3	97.6	77.7	99.6	99.8	99.9	100.0	100.0	100.0	100.0
	iKY	75.3	91.4	93.0	97.1	97.5	99.6	99.3	100.0	100.0	100.0	100.0	100.0
	MBB $((2N)^{1/3})$	32.0	57.3	37.8	68.5	45.8	70.8	63.1	82.6	72.8	87.9	79.8	92.2
	MBB $(N^{2/5})$	32.0	57.3	37.8	68.5	45.8	70.8	55.5	76.7	68.4	85.4	73.4	89.6
	KY $((2N)^{1/3})$	0.0	31.3	8.7	45.5	23.5	58.5	37.3	71.3	53.5	78.6	64.9	87.0
	KY $(N^{2/5})$	0.0	31.9	5.9	43.2	19.9	55.7	34.0	67.8	47.8	75.1	58.4	83.0
H_A^{cp}	SB	10.6	24.9	21.1	41.7	28.1	51.8	45.5	69.0	66.3	82.5	80.1	91.8
	iKY	8.7	27.5	21.6	39.8	30.4	54.7	48.9	70.3	70.9	83.3	81.3	91.7
	MBB $((2N)^{1/3})$	1.3	14.7	3.9	25.6	9.3	35.9	20.9	51.2	42.8	74.5	58.3	83.1
	MBB $(N^{2/5})$	1.3	14.7	3.9	25.6	9.3	35.9	19.3	49.6	34.2	73.1	55.0	83.2
	KY $((2N)^{1/3})$	0.0	7.2	0.1	18.8	1.9	31.5	7.5	50.1	28.9	71.3	50.5	84.1
	KY $(N^{2/5})$	0.0	7.0	0.0	16.8	1.2	30.5	6.1	46.5	21.9	66.1	42.4	80.9

that using the block length of $b = \lceil (2N)^{1/3} \rceil$ yields better performance; however, there is no confirmative suggestion for the choice of the bandwidth h . These results, together with those presented in the previous section, further confirm that the bootstrap-based

Table 2.4

Averaged power increase rates (%) from iKY to SB with different sample sizes $N = 30, 50, 70, 100, 150, 200$, different significance levels $\alpha = 1\%, 5\%$ and different kinds of sample noises.

N	30		50		70		100		150		200	
α	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%
i.i.d. normal	43.6	3.1	6.9	9.1	7.4	-1.7	5.6	0.8	0.7	-0.4	3.4	0.1
i.i.d. t_5	37.7	-3.3	7.7	1.3	3.2	0.7	3.9	-1.8	-0.2	-0.5	0.3	0.0
i.i.d. $F(10, 8)$	6.9	-5.7	-1.8	0.6	-2.0	-2.3	-2.6	-0.6	-2.2	-0.3	-0.5	0.0

FKPSS test is a better alternative in practice when N is small or moderate.

2.4 Conclusion

In this study, we proposed a bootstrap-based FKPSS test and validated its efficiency theoretically and numerically. The theoretical results contribute to the literature on bootstrapping non-stationary FTS data. In particular, when the errors of sample data are independent, our simulations demonstrate the good empirical size and power of the bootstrap-based FKPSS test even when the sample size is small. The MBB also shows a good empirical power when the errors are dependent, though some of the empirical sizes are indifferent. Although the existing FKPSS test (Kokoszka and Young (2016)) performs well theoretically when the sample size is large, the ability to collect a huge amount of functional data is restricted in practice. Hence, the bootstrap-based FKPSS test proposed in this study serves as a practical alternative.

One interesting future research direction would be to prove (or disprove) the conditional weak convergence of R_N^* or (2.11). Inspired by the recent study by Pilavakis et al. (2017), one line of research could be to study MBB for different functional test statistics, including the FKPSS test statistic.

Table 2.5

Empirical sizes (%) and powers (%) for the different FKPSS tests for FTS with AR(1) errors, different sample sizes $N = 30, 50, 70, 100, 150, 200$ and different significance levels $\alpha = 1\%, 5\%$.

N		30		50		70		100		150		200	
α		1%	5%	1%	5%	1%	5%	1%	5%	1%	5%	1%	5%
H_0^{ts}	MBB $((2N)^{1/3})$	4.8	16.4	3.6	13.7	2.2	11.1	2.9	9.1	2.0	9.7	1.9	9.2
	MBB $(N^{2/5})$	4.8	16.4	3.6	13.7	2.2	11.1	2.5	12.2	2.5	9.3	1.8	10.7
	KY $((2N)^{1/3})$	0.0	4.8	0.1	6.2	0.5	5.8	0.3	5.9	0.8	6.4	1.2	5.9
	KY $(N^{2/5})$	0.0	4.8	0.0	6.3	0.4	5.5	0.3	5.7	0.6	6.1	1.0	5.6
H_A^{ts+rw}	MBB $((2N)^{1/3})$	17.2	46.4	28.6	60.9	47.8	69.7	70.8	87.6	82.7	94.4	92.0	97.5
	MBB $(N^{2/5})$	17.2	46.4	28.6	60.9	47.8	69.7	64.7	84.3	81.2	94.1	87.7	97.1
	KY $((2N)^{1/3})$	0.0	18.5	4.6	41.5	20.5	60.0	46.9	74.2	73.9	88.9	84.9	96.0
	KY $(N^{2/5})$	0.0	29.1	9.4	50.7	29.7	67.2	55.1	81.4	72.9	90.0	82.6	96.3
H_A^{ur}	MBB $((2N)^{1/3})$	43.5	68.4	43.4	69.7	49.8	76.0	65.7	82.8	73.6	88.9	76.4	91.5
	MBB $(N^{2/5})$	43.5	68.4	43.4	69.7	49.8	76.0	54.0	81.9	65.4	86.3	72.1	89.2
	KY $((2N)^{1/3})$	0.0	36.5	10.9	45.9	24.3	63.1	39.7	72.3	55.2	80.4	66.4	86.9
	KY $(N^{2/5})$	0.0	36.5	8.1	44.3	22.1	60.0	34.9	67.4	48.9	77.1	58.0	82.1
H_A^{cp}	MBB $((2N)^{1/3})$	6.8	23.9	6.9	23.5	7.4	24.6	12.7	33.0	18.4	40.1	25.6	49.7
	MBB $(N^{2/5})$	6.8	23.9	6.9	23.5	7.4	24.6	11.1	31.1	16.1	37.2	21.6	45.2
	KY $((2N)^{1/3})$	0.0	7.7	0.2	11.1	1.5	15.6	4.2	22.9	7.7	31.9	13.7	39.1
	KY $(N^{2/5})$	0.0	8.4	0.1	12.1	1.0	15.4	3.5	23.2	6.9	29.9	11.3	39.5

2.5 Proofs

In this appendix, to ease the notational burden, we set $k_n = n - (N + 1)/2$ for $n \in \{1, \dots, N\}$, suppress the curve arguments of the functional data, and recall that we use f to represent $\int_0^1$. Following the framework of Kokoszka and Young (2016), we define the partial sums of original and bootstrap sample errors as follows:

$$V_N(x, t) := \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \eta_n(t), \quad V_N^*(x, t) := \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \eta_n^*(t), \quad 0 \leq t, x \leq 1. \quad (2.13)$$

2.5.1 Preliminary Results

Before we present the proof of Theorem 2.2.1, we will first obtain some preliminary results. The following two theorems play a crucial role in the subsequent derivation. Theorem 2.5.1 is a well-known theorem; see Billingsley (1968) for its proof. Theorem 2.5.2 is one of the main results in this study and it resembles Theorem 1.1 in Berkes et al. (2013) except that on unconditional probability space, which nests the randomness from bootstrapping.

Theorem 2.5.1. Suppose Z_N, Y_N and Y are random variables taking values in a separable metric space with distance function ρ . If $Y_N \xrightarrow{d} Y$ and $\rho(Z_N, Y_N) \xrightarrow{p} 0$, then $Z_N \xrightarrow{d} Y$.

Theorem 2.5.2. Under the null hypothesis H_0^{ts} in (2.1) and Assumption 2.2.1,

1. $\sum_{i=1}^{\infty} \lambda_i < \infty$, where $\lambda_1, \lambda_2, \dots$ are eigenvalues of long-run covariance function $c(t, s)$ in (2.5).
2. We can define a sequence of Gaussian processes $\Gamma_N(x, t)$ such that:

$$\{\Gamma_N(x, t), 0 \leq x, t \leq 1\} \xrightarrow{\bar{d}} \{\Gamma(x, t), 0 \leq x, t \leq 1\}$$

and

$$\kappa_N = \sup_{0 \leq x \leq 1} \|V_N^*(x, \cdot) - \Gamma_N(x, \cdot)\| = o_{\bar{P}_T}(1),$$

where $\Gamma(x, t) = \sum_{i=1}^{\infty} \lambda_i^{1/2} W_i(x) \phi_i(t)$.

The first claim in Theorem 2.5.2, i.e., $\sum_{i=1}^{\infty} \lambda_i < \infty$, has been proved in Lemma 2.2 in Berkes et al. (2013). Thus, we only prove the second claim of Theorem 2.5.2. To this end, we introduce the following Lemmas 2.5.1-2.5.5, which resemble Lemmas 2.4-2.7 in Berkes et al. (2013).

Lemma 2.5.1. Under the null hypothesis H_0^{ts} in (2.1) and Assumption 2.2.1, for all $x > 0$,

$$\lim_{K \rightarrow \infty} \limsup_{N \rightarrow \infty} \bar{\Pr} \left(\max_{1 \leq k \leq N} \left\| \frac{1}{\sqrt{N}} \sum_{n=1}^k \bar{\eta}_{n,K}^* \right\| > x \right) = 0, \quad (2.14)$$

where $\bar{\eta}_{n,K}^*(t) = \sum_{l=K+1}^{\infty} \langle \eta_n^*, \phi_l \rangle \phi_l(t)$, and $\phi_1, \phi_2, \dots$ are eigenfunctions of $c(t, s)$ in (2.5).

Proof of Lemma 2.5.1. Notice that

$$\begin{aligned} \Pr^* \left(\max_{1 \leq k \leq N} \left\| \frac{1}{\sqrt{N}} \sum_{n=1}^k \bar{\eta}_{n,K}^* \right\|^2 > x \right) &\leq \frac{1}{x} \mathbb{E}^* \left\{ \left\| \frac{1}{\sqrt{N}} \sum_{n=1}^N \bar{\eta}_{n,K}^* \right\|^2 \right\} \leq \frac{1}{x} \mathbb{E}^* \left\{ \|\bar{\eta}_{1,K}^*\|^2 \right\} \\ &= \frac{1}{x} \frac{1}{N} \sum_{n=1}^N \left(\sum_{l=K+1}^{\infty} \langle \hat{\eta}_n, \phi_l \rangle^2 \right), \end{aligned}$$

where the first inequality follows from Lemma 2.3 in Berkes et al. (2013), with

$$\mathbb{E}^* \{ \bar{\eta}_{1,K}^*(t) \} = \sum_{l=K+1}^{\infty} \langle \mathbb{E}^*(\eta_1^*), \phi_l \rangle \phi_l(t) = 0$$

and

$$\begin{aligned} \mathbb{E}^* \{ \|\bar{\eta}_{1,K}^*\|^2 \} &= \sum_{l=K+1}^{\infty} \mathbb{E}^* (\langle \eta_1^*, \phi_l \rangle^2) = \frac{1}{N} \sum_{n=1}^N \sum_{l=K+1}^{\infty} \langle \hat{\eta}_n, \phi_l \rangle^2 \leq \frac{1}{N} \sum_{n=1}^N \sum_{l=1}^{\infty} \langle \hat{\eta}_n, \phi_l \rangle^2 \\ &= \frac{1}{N} \sum_{n=1}^N \|\hat{\eta}_n\|^2 < \infty \end{aligned}$$

when N and original samples $\mathcal{X}_N$ are fixed. Since

$$\begin{aligned} \mathbb{E}(\langle \hat{\eta}_n, \phi_l \rangle^2) &= \mathbb{E} \left\{ \left\langle (\eta_n - \bar{\eta}_N) - (\hat{\xi} - \xi)k_n, \phi_l \right\rangle^2 \right\} = \mathbb{E} \left\{ \left\langle \eta_n - \sum_{i=1}^N \left(\frac{1}{N} + \frac{k_n k_i}{s_N} \right) \eta_i, \phi_l \right\rangle^2 \right\} \\ &= \left(1 - \frac{1}{N} - \frac{k_n^2}{s_N} \right)^2 \lambda_l + \sum_{1 \leq i \leq N, i \neq n} \left(\frac{1}{N} + \frac{k_n k_i}{s_N} \right)^2 \lambda_l, \end{aligned}$$

where we have used $\hat{\eta}_n = (\eta_n - \bar{\eta}_N) - (\hat{\xi} - \xi)k_n$ with $\bar{\eta}_N = (1/N) \sum_{n=1}^N \eta_n$ under H_0^{ts} , we have

$$\frac{1}{N} \sum_{n=1}^N \mathbb{E}\{\langle \hat{\eta}_n, \phi_l \rangle^2\} = \left\{ \frac{1}{N} \sum_{n=1}^N \left(1 - \frac{1}{N} - \frac{k_n^2}{s_N}\right)^2 + \frac{1}{N} \sum_{1 \leq i \neq n \leq N} \left(\frac{1}{N} + \frac{k_n k_i}{s_N}\right)^2 \right\} \lambda_l =: C_N \lambda_l.$$

By noting that $k_n k_i / s_N = \mathcal{O}(1/N)$ for any n and i , it is easy to see that $C_N = \mathcal{O}(1)$.

Therefore,

$$\mathbb{E} \left\{ \Pr^* \left(\max_{1 \leq k \leq N} \left\| \frac{1}{\sqrt{N}} \sum_{n=1}^k \bar{\eta}_{n,K}^* \right\|^2 > x \right) \right\} \leq \frac{\mathcal{O}(1)}{x} \sum_{l=K+1}^{\infty} \lambda_l \quad (2.15)$$

and the right-hand side of (2.15) tends to zero as $K \rightarrow \infty$. Then, (2.14) follows. $\square$

Lemma 2.5.2. Under the null hypothesis H_0^{ts} in (2.1) and Assumption 2.2.1, for any constant vector $(t_1, t_2, \dots, t_K) \in R^K$

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \langle \eta_n^*, \sum_{i=1}^K t_i \phi_i \rangle \xrightarrow{\bar{d}} \sum_{i=1}^K t_i \lambda_i^{\frac{1}{2}} W_i(x) \quad \text{as } N \rightarrow \infty,$$

where $W_1, W_2, \dots, W_K$ are i.i.d. standard Wiener processes.

Proof of Lemma 2.5.2. We first show that

$$\frac{1}{\hat{\sigma}_N \sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \langle \eta_n^*, \sum_{i=1}^K t_i \phi_i \rangle \xrightarrow{d^*} W(x) \quad a.s., \quad (2.16)$$

where $\hat{\sigma}_N = \mathbb{E}^*(\langle \eta_n^*, \sum_{i=1}^K t_i \phi_i \rangle^2)$ and $W(x)$ is a standard Wiener process i.i.d. as $\{W_i(x)\}_{i=1}^K$. By Theorem 2.2 in Park (2002), we only need to show that there exists $r > 2$ such that

$$N^{1-\frac{r}{2}} \mathbb{E}^*(|\langle \eta_n^*, \sum_{i=1}^K t_i \phi_i \rangle|^r) = N^{-\frac{r}{2}} \sum_{n=1}^N |\langle \hat{\eta}_n, \sum_{i=1}^K t_i \phi_i \rangle|^r \xrightarrow{a.s.} 0.$$

We set $r = 4$. Due to $(a + b)^4 \leq 8(a^4 + b^4)$, we ought to show that

$$N^{-2} \sum_{n=1}^N \langle \eta_n - \bar{\eta}_N, \sum_{i=1}^K t_i \phi_i \rangle^4 \xrightarrow{a.s.} 0 \quad (2.17)$$

and

$$N^{-2} \sum_{n=1}^N \langle (\hat{\xi} - \xi)k_n, \sum_{i=1}^K t_i \phi_i \rangle^4 \xrightarrow{a.s.} 0. \quad (2.18)$$

To show (2.17), again, we have

$$\frac{1}{N} \sum_{n=1}^N \langle \eta_n - \bar{\eta}_N, \sum_{i=1}^K t_i \phi_i \rangle^4 \leq \frac{1}{N} \sum_{n=1}^N 8 \langle \eta_n, \sum_{i=1}^K t_i \phi_i \rangle^4 + \frac{1}{N} \sum_{n=1}^N 8 \langle \bar{\eta}_N, \sum_{i=1}^K t_i \phi_i \rangle^4.$$

By the strong law of large numbers, we have

$$\frac{1}{N} \sum_{n=1}^N 8 \langle \eta_n, \sum_{i=1}^K t_i \phi_i \rangle^4 \xrightarrow{a.s.} 8E(\langle \eta_0, \sum_{i=1}^K t_i \phi_i \rangle^4),$$

and

$$\frac{1}{N} \sum_{n=1}^N 8 \langle \bar{\eta}_N, \sum_{i=1}^K t_i \phi_i \rangle^4 = 8 \langle \bar{\eta}_N, \sum_{i=1}^K t_i \phi_i \rangle^4 \xrightarrow{a.s.} 8(E \langle \eta_0, \sum_{i=1}^K t_i \phi_i \rangle)^4 = 0.$$

Then by the Cauchy-Schwarz inequality, we have

$$E(\langle \eta_0, \sum_{i=1}^K t_i \phi_i \rangle^4) \leq E(\|\eta_0\|^4) \|\sum_{i=1}^K t_i \phi_i\|^4 = (\sum_{i=1}^K t_i^2)^2 E(\|\eta_0\|^4) < \infty \quad (2.19)$$

Therefore, $N^{-2} \sum_{n=1}^N \langle \eta_n - \bar{\eta}_N, \sum_{i=1}^K t_i \phi_i \rangle^4 \xrightarrow{a.s.} 0$. To show (2.18), we further notice that

$$\begin{aligned} N^{-2} \sum_{n=1}^N \langle (\hat{\xi} - \xi) k_n, \sum_{i=1}^K t_i \phi_i \rangle^4 &\leq \left| \max_{1 \leq i, j \leq N} \left(\frac{k_i k_j}{s_N} \right)^4 \right| N^{-2} \sum_{j=1}^N N^3 \sum_{n=1}^N \langle \eta_n, \sum_{i=1}^K t_i \phi_i \rangle^4 \\ &= \mathcal{O}(N^{-1}) \frac{1}{N} \sum_{n=1}^N \langle \eta_n, \sum_{i=1}^K t_i \phi_i \rangle^4 \xrightarrow{a.s.} 0 \end{aligned}$$

since $(1/N) \sum_{n=1}^N \langle \eta_n, \sum_{i=1}^K t_i \phi_i \rangle^4 \xrightarrow{a.s.} E(\langle \eta_0, \sum_{i=1}^K t_i \phi_i \rangle^4) < \infty$ by (2.19). Therefore, we have proved both (2.17) and (2.18) and thus (2.16) follows.

Now, we show that

$$\frac{1}{\hat{\sigma}_N \sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \langle \eta_n^*, \sum_{i=1}^K t_i \phi_i \rangle \xrightarrow{\bar{d}} W(x). \quad (2.20)$$

Let $Q_N = (1/\hat{\sigma}_N \sqrt{N}) \sum_{n=1}^{\lfloor Nx \rfloor} \langle \eta_n^*, \sum_{i=1}^K t_i \phi_i \rangle$ and $Q = W(x)$. Then (2.16) implies that for any bounded and continuous function f , $E^*\{f(Q_N)\} \rightarrow E\{f(Q)\}$ almost surely. Since f is bounded, $E[E^*\{f(Q_N)\}]$ is also bounded. By the bounded convergence theorem, we have

$$E[E^*\{f(Q_N)\}] \rightarrow E[E\{f(Q)\}],$$

which implies that for any bounded and continuous function f , $\bar{E}\{f(Q_N)\} \rightarrow \bar{E}\{f(Q)\}$, and thus (2.20) follows.

Next, we show that

$$\hat{\sigma}_N^2 \xrightarrow{\text{Pr}} \sum_{i=1}^K t_i^2 \lambda_i. \quad (2.21)$$

Notice that $\hat{\sigma}_N^2 = (1/N) \sum_{n=1}^N \langle \hat{\eta}_n, \sum_{i=1}^K t_i \phi_i \rangle^2$ and

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N \langle \eta_n - \bar{\eta}_N, \sum_{i=1}^K t_i \phi_i \rangle^2 &= \frac{1}{N} \sum_{n=1}^N \langle \eta_n, \sum_{i=1}^K t_i \phi_i \rangle^2 - \langle \bar{\eta}_N, \sum_{i=1}^K t_i \phi_i \rangle^2 \\ &\xrightarrow{a.s.} E(\langle \eta_0, \sum_{i=1}^K t_i \phi_i \rangle^2) - (E\langle \eta_0, \sum_{i=1}^K t_i \phi_i \rangle)^2 = E(\langle \eta_0, \sum_{i=1}^K t_i \phi_i \rangle^2). \end{aligned}$$

Since for $i \neq j$

$$\begin{aligned} E(\langle \eta_0, \phi_i \rangle \langle \eta_0, \phi_j \rangle) &= E \left\{ \int \eta_0(s) \phi_i(s) ds \int \eta_0(t) \phi_j(t) dt \right\} \\ &= E \left\{ \iint \eta_0(s) \phi_i(s) \eta_0(t) \phi_j(t) ds dt \right\} \\ &= \iint E\{\eta_0(s) \eta_0(t)\} \phi_i(s) \phi_j(t) ds dt = \iint C(s, t) \phi_i(s) \phi_j(t) ds dt \\ &= \int \lambda_i \phi_i(t) \phi_j(t) dt = 0, \end{aligned}$$

we have $E(\langle \eta_0, \sum_{i=1}^K t_i \phi_i \rangle^2) = \sum_{i=1}^K E(\langle \eta_0, t_i \phi_i \rangle^2) = \sum_{i=1}^K t_i^2 \lambda_i$. Hence, we only need to prove

$$\hat{\sigma}_N^2 - (1/N) \sum_{n=1}^N \langle \eta_n - \bar{\eta}_N, \sum_{i=1}^K t_i \phi_i \rangle^2 \xrightarrow{\text{Pr}} 0.$$

To this end, we note that

$$\begin{aligned} &E(|\hat{\sigma}_N^2 - \frac{1}{N} \sum_{n=1}^N \langle \eta_n - \bar{\eta}_N, \sum_{i=1}^K t_i \phi_i \rangle^2|) \\ &= E\{|\frac{1}{N} \sum_{n=1}^N \langle (\hat{\xi} - \xi) k_n, \sum_{i=1}^K t_i \phi_i \rangle^2 - 2 \frac{1}{N} \sum_{n=1}^N \langle \eta_n - \bar{\eta}_N, \sum_{i=1}^K t_i \phi_i \rangle \langle (\hat{\xi} - \xi) k_n, \sum_{j=1}^K t_j \phi_j \rangle|\} \\ &= E\{\frac{1}{N s_N} (\sum_{n=1}^N \langle k_n \eta_n, \sum_{i=1}^K t_i \phi_i \rangle)^2\} = \frac{1}{N} E(\langle \eta_0, \sum_{i=1}^K t_i \phi_i \rangle^2) = \frac{1}{N} \sum_{i=1}^K t_i^2 \lambda_i = o(1). \end{aligned}$$

Therefore, (2.21) is proved and consequently by (2.20) and Theorem 2.5.1, we have

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \langle \eta_n^*, \sum_{i=1}^K t_i \phi_i \rangle \xrightarrow{\mathcal{D}} \left(\sum_{i=1}^K t_i^2 \lambda_i \right)^{1/2} W(x).$$

Finally, we prove that

$$\left(\sum_{i=1}^K t_i^2 \lambda_i \right)^{1/2} W(x) \stackrel{d}{=} \sum_{i=1}^K t_i \lambda_i^{1/2} W_i(x), \quad (2.22)$$

where $W, W_1, \dots, W_K$ are i.i.d. standard Wiener processes.

For any $x_1 \in [0, 1]$, we have

$$\left(\sum_{i=1}^K t_i^2 \lambda_i \right)^{1/2} W(x_1) \stackrel{d}{=} \sum_{i=1}^K t_i \lambda_i^{1/2} W_i(x_1) \sim N(0, \sum_{i=1}^K t_i^2 \lambda_i x_1).$$

For any $x_1, x_2 \in [0, 1]$ and $x_1 \leq x_2$, we need to prove that

$$\left(\left(\sum_{i=1}^K t_i^2 \lambda_i \right)^{1/2} W(x_1), \left(\sum_{i=1}^K t_i^2 \lambda_i \right)^{1/2} W(x_2) \right) \stackrel{d}{=} \left(\sum_{i=1}^K t_i \lambda_i^{1/2} W_i(x_1), \sum_{i=1}^K t_i \lambda_i^{1/2} W_i(x_2) \right). \quad (2.23)$$

From Corollary 1 of Theorem 5.1 in Billingsley (1968), it suffices to prove that

$$\begin{aligned} & \left(\left(\sum_{i=1}^K t_i^2 \lambda_i \right)^{1/2} W(x_1), \left(\sum_{i=1}^K t_i^2 \lambda_i \right)^{1/2} W(x_2) - \left(\sum_{i=1}^K t_i^2 \lambda_i \right)^{1/2} W(x_1) \right) \\ & \stackrel{d}{=} \left(\sum_{i=1}^K t_i \lambda_i^{1/2} W_i(x_1), \sum_{i=1}^K t_i \lambda_i^{1/2} W_i(x_2) - \sum_{i=1}^K t_i \lambda_i^{1/2} W_i(x_1) \right). \end{aligned}$$

The claim above is straightforward as the components on each side are independent. Similarly, we can show the similar results for arbitrary number of points x 's in $[0, 1]$. Thus we can conclude that $(\sum_{i=1}^K t_i^2 \lambda_i)^{1/2} W(x)$ has the same finite-dimensional distribution as $\sum_{i=1}^K t_i \lambda_i^{1/2} W_i(x)$. Then for any $x_1, x_2 \in [0, 1]$, since $W(x_1) - W(x_2) \sim \mathcal{N}(0, |x_2 - x_1|)$, we have

$$\begin{aligned} \mathbb{E}\{|\left(\sum_{i=1}^K t_i^2 \lambda_i\right)^{1/2} W(x_2) - \left(\sum_{i=1}^K t_i^2 \lambda_i\right)^{1/2} W(x_1)|^4\} &= \left(\sum_{i=1}^K t_i^2 \lambda_i\right)^2 (x_1 - x_2)^2 \mathbb{E}\{W(1)^4\} \\ &= 3 \left(\sum_{i=1}^K t_i^2 \lambda_i\right)^2 (x_1 - x_2)^2. \end{aligned}$$

By Theorem 12.3 in Billingsley (1968), $(\sum_{i=1}^K t_i^2 \lambda_i)^{1/2} W(x)$ is tight. (2.22) is then proved by Theorem 15.1 in Billingsley (1968), which implies

$$\left(\sum_{i=1}^K t_i^2 \lambda_i \right)^{1/2} W(x) \stackrel{d}{=} \sum_{i=1}^K t_i \lambda_i^{1/2} W_i(x).$$

Therefore, Lemma 2.5.2 follows from (2.20)-(2.22). $\square$

Lemma 2.5.3. Under the null hypothesis H_0^{ts} in (2.1) and Assumption 2.2.1, the K -dimensional stochastic process

$$\left(\frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \langle \eta_n^*, \phi_1 \rangle, \dots, \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \langle \eta_n^*, \phi_K \rangle \right)$$

converges weakly, as $N \rightarrow \infty$, in $D^K[0, 1]$ to

$$\left(\lambda_1^{1/2} W_1(x), \dots, \lambda_K^{1/2} W_K(x) \right),$$

where $\{W_i\}_{i=1}^K$ are i.i.d. Wiener processes.

Proof of Lemma 2.5.3. This lemma follows from Lemma 2.5.2 and the Cramér–Wold theorem. $\square$

By the Skorokhod–Dudley–Wichura theorem (see p.47 in Shorack and Wellner (1986)), we obtain the following corollary from Lemma 2.5.3.

Corollary 2.5.1. Under the null hypothesis H_0^{ts} in (2.1) and Assumption 2.2.1, for every positive integer N , there exist K i.i.d. Wiener processes $\{W_{i,N}\}_{i=1}^K$ such that for each i ,

$$\sup_{0 \leq x \leq 1} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \langle \eta_n^*, \phi_i \rangle - \lambda_i^{1/2} W_{i,N}(x) \right| \xrightarrow{\bar{P}_T} 0 \quad \text{as } N \rightarrow \infty.$$

Lemma 2.5.4. Under the null hypothesis H_0^{ts} in (2.1) and Assumption 2.2.1, for $\{W_{i,N}\}_{i=1}^K$ defined in Corollary 2.5.1, we have

$$\sup_{0 \leq x \leq 1} \int \left\{ \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \eta_{n,K}^*(t) - \sum_{i=1}^K \lambda_i^{1/2} W_{i,N}(x) \phi_i(t) \right\}^2 dt \xrightarrow{\bar{P}_T} 0 \quad \text{as } N \rightarrow \infty,$$

where $\eta_{n,K}^*(t) = \sum_{i=1}^K \langle \eta_n^*, \phi_i \rangle \phi_i(t)$.

Proof of Lemma 2.5.4. Since

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \eta_{n,K}^*(t) - \sum_{i=1}^K \lambda_i^{1/2} W_{i,N}(x) \phi_i(t) = \sum_{i=1}^K \left\{ \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \langle \eta_n^*, \phi_i \rangle - \lambda_i^{1/2} W_{i,N}(x) \right\} \phi_i(t)$$

and by Corollary 2.5.1, we have

$$\begin{aligned}
& \sup_{0 \leq x \leq 1} \int \left\{ \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \eta_{n,K}^*(t) - \sum_{i=1}^K \lambda_i^{\frac{1}{2}} W_{i,N}(x) \phi_i(t) \right\}^2 dt \\
&= \sup_{0 \leq x \leq 1} \sum_{i=1}^K \left\{ \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \langle \eta_n^*, \phi_i \rangle - \lambda_i^{\frac{1}{2}} W_{i,N}(x) \right\}^2 \\
&\leq \sum_{i=1}^K \sup_{0 \leq x \leq 1} \left\{ \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \langle \eta_n^*, \phi_i \rangle - \lambda_i^{\frac{1}{2}} W_{i,N}(x) \right\}^2 \xrightarrow{\bar{\mathbb{P}}_T} 0 \quad \text{as } N \rightarrow \infty.
\end{aligned}$$

□

Lemma 2.5.5. Under the null hypothesis H_0^{ts} in (2.1) and Assumption 2.2.1, we have

$$\sup_{0 \leq x \leq 1} \int \left\{ \sum_{i=K+1}^{\infty} \lambda_i^{\frac{1}{2}} W_i(x) \phi_i(t) \right\}^2 dt \xrightarrow{\bar{\mathbb{P}}_T} 0 \quad \text{as } K \rightarrow \infty,$$

where $W_1, W_2, \dots$ are i.i.d. Wiener processes.

Proof of Lemma 2.5.5. The proof of Lemma 2.5.5 follows directly from Lemma 2.7 in Berkes et al. (2013). □

Proof of Theorem 2.5.2. Now, we are ready to prove Theorem 2.5.2.

From Lemmas 2.5.1, 2.5.4 and 2.5.5, for each N , we can define i.i.d. Wiener processes $\{W_{i,N}\}_{i=1}^{\infty}$ such that

$$\sup_{0 \leq x \leq 1} \int \left\{ \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \eta_n^*(t) - \sum_{i=1}^{\infty} \lambda_i^{\frac{1}{2}} W_{i,N}(x) \phi_i(t) \right\}^2 dt \xrightarrow{\bar{\mathbb{P}}_T} 0 \quad \text{as } N \rightarrow \infty.$$

We denote $\sum_{i=1}^{\infty} \lambda_i^{1/2} W_{i,N}(x) \phi_i(t)$ by $\Gamma_N(x, t)$. Since the distribution of $W_{i,N}$, $i = 1, 2, \dots$ does not depend on N , we have $\{\Gamma_N(x, t), 0 \leq x, t \leq 1\} \stackrel{\bar{d}}{=} \{\Gamma(x, t), 0 \leq x, t \leq 1\}$. Then Theorem 2.5.2 follows. □

2.5.2 Main Results in Section 2.2.2

Using Theorem 2.5.2, we can deduce some lemmas that are useful in proving Theorem 2.2.1. Lemmas 2.5.6 and 2.5.7 below follow directly from Lemmas 5.3 and 5.4 and Lemma 5.6 in Kokoszka and Young (2016), respectively; hence the proofs are omitted.

Lemma 2.5.6. We have

$$\frac{N^{3/2}}{s_N} \sum_{n=1}^N k_n \eta_n^*(t) = \frac{1}{N^{-3}s_N} \left\{ \left(\frac{N-1}{2N} \right) V_N^*(1, t) - \frac{1}{N} \sum_{k=1}^{N-1} V_N^* \left(\frac{k}{N}, t \right) \right\},$$

and

$$\begin{aligned} Z_N^*(x, t) &= V_N^*(x, t) - \frac{\lfloor Nx \rfloor}{N} V_N^*(1, t) \\ &\quad - \frac{1}{2} \frac{1}{N^{-3}s_N} \left\{ \left(\frac{N-1}{2N} \right) V_N^*(1, t) - \frac{1}{N} \sum_{k=1}^{N-1} V_N^* \left(\frac{k}{N}, t \right) \right\} \frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right). \end{aligned}$$

Lemma 2.5.7. Let $\Gamma^0(x, t) = \Gamma(x, t) + (2x - 3x^2)\Gamma(1, t) + (-6x + 6x^2) \int \Gamma(y, t) dy$, where Γ is defined with Wiener processes $W_1, W_2, \dots$ in Theorem 2.5.2. Then, we have

$$\int \|\Gamma^0(x, \cdot)\|^2 dx = \sum_{i=1}^{\infty} \lambda_i \int V_i^2(x) dx,$$

where V_i is defined by (2.10) with W replaced by W_i .

Lemma 2.5.8. For V_N^* defined in (2.13), we define Γ_N by Theorem 2.5.2. Then we have

$$\int \left\{ \frac{1}{N} \sum_{k=1}^{N-1} V_N^* \left(\frac{k}{N}, t \right) - \int_0^1 \Gamma_N(y, t) dy \right\}^2 dt \xrightarrow{\mathbb{P}_T} 0.$$

Proof of Lemma 2.5.8. Since V_N^* is a stepwise function in x with jumps at $y := k/N$ for $k \in \mathbb{Z}$, we have

$$\frac{1}{N} \sum_{k=1}^{N-1} V_N^* \left(\frac{k}{N}, t \right) = \frac{1}{N} \sum_{k=0}^{N-1} V_N^* \left(\frac{k}{N}, t \right) = \int_0^1 V_N^*(y, t) dy.$$

For any t , by Theorem 2.5.2, we have

$$\begin{aligned} \left\| \int V_N^*(y, \cdot) dy - \int \Gamma_N(y, \cdot) dy \right\| &\leq \int \|V_N^*(y, \cdot) - \Gamma_N(y, \cdot)\| dy \\ &\leq \int \sup_{0 \leq y \leq 1} \|V_N^*(y, \cdot) - \Gamma_N(y, \cdot)\| dy \\ &\leq \kappa_N = o_{\mathbb{P}_T}(1), \end{aligned}$$

where κ_N is defined in Theorem 2.5.2. Then, the result follows. $\square$

Lemma 2.5.9. Let Z_N^* be Z_N in (2.9) with $\hat{\eta}_n$ replaced by $\hat{\eta}_n^*$ and Γ_N^0 be Γ^0 in Lemma 2.5.7 with Γ replaced by Γ_N defined in Theorem 2.5.2. Then, we have

$$\sup_{0 \leq x \leq 1} \|Z_N^*(x, \cdot) - \Gamma_N^0(x, \cdot)\| \xrightarrow{\bar{\mathbb{P}}_T} 0.$$

Proof of Lemma 2.5.9. We first rewrite Z_N^* as follows by Lemma 2.5.6:

$$Z_N^*(x, t) = V_N^*(x, t) + a_N(x)V_N^*(1, t) + b_N(x)\frac{1}{N} \sum_{k=1}^{N-1} V_N^*\left(\frac{k}{N}, t\right),$$

where

$$a_N(x) = \left\{ \frac{1}{2N^{-3}s_N} \left(\frac{N-1}{2N} \right) - 1 \right\} \frac{\lfloor Nx \rfloor}{N} - \frac{1}{2N^{-3}s_N} \left(\frac{N-1}{2N} \right) \left(\frac{\lfloor Nx \rfloor}{N} \right)^2$$

and

$$b_N(x) = \frac{1}{2N^{-3}s_N} \frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right).$$

Since

$$\begin{aligned} \|Z_N^*(x, \cdot) - \Gamma_N^0(x, \cdot)\| &\leq \|V_N^*(x, \cdot) - \Gamma_N(x, \cdot)\| + \|a_N(x)V_N^*(1, \cdot) - (2x - 3x^2)\Gamma_N(1, \cdot)\| \\ &\quad + \left\| b_N(x)\frac{1}{N} \sum_{k=1}^{N-1} V_N^*\left(\frac{k}{N}, \cdot\right) - (-6x + 6x^2) \int \Gamma_N(y, \cdot) dy \right\|, \end{aligned}$$

it suffices to prove

$$\sup_{0 \leq x \leq 1} \|V_N^*(x, \cdot) - \Gamma_N(x, \cdot)\| \xrightarrow{\bar{\mathbb{P}}_T} 0, \quad (2.24)$$

$$\sup_{0 \leq x \leq 1} \|a_N(x)V_N^*(1, \cdot) - (2x - 3x^2)\Gamma_N(1, \cdot)\| \xrightarrow{\bar{\mathbb{P}}_T} 0, \quad (2.25)$$

and

$$\sup_{0 \leq x \leq 1} \left\| b_N(x)\frac{1}{N} \sum_{k=1}^{N-1} V_N^*\left(\frac{k}{N}, \cdot\right) - (-6x + 6x^2) \int \Gamma_N(y, \cdot) dy \right\| \xrightarrow{\bar{\mathbb{P}}_T} 0. \quad (2.26)$$

The first one (2.24) follows from Theorem 2.5.2. To show (2.25), we note that

$$\begin{aligned} \|a_N(x)V_N^*(1, \cdot) - (2x - 3x^2)\Gamma_N(1, \cdot)\| &\leq |a_N(x)| \|V_N^*(1, \cdot) - \Gamma_N(1, \cdot)\| \\ &\quad + |a_N(x) - (2x - 3x^2)| \|\Gamma_N(1, \cdot)\|. \end{aligned}$$

Hence, we ought to prove that

$$\sup_{N \geq 1} \left\{ \sup_{0 \leq x \leq 1} |a_N(x)| \right\} < \infty, \quad \sup_{0 \leq x \leq 1} |a_N(x) - (2x - 3x^2)| \rightarrow 0, \quad \|\Gamma_N(1, \cdot)\| = \mathcal{O}_{\bar{\mathbb{P}}_T}(1), \quad (2.27)$$

$$\|V_N^*(1, \cdot) - \Gamma_N(1, \cdot)\| \xrightarrow{\bar{\mathbb{P}}_T} 0. \quad (2.28)$$

The three claims (2.27) follow from the proof of Lemma 5.7 in Kokoszka and Young (2016) and (2.28) follows from Theorem 2.5.2. To show (2.26), similarly, we note that

$$\begin{aligned} & \left\| b_N(x) \frac{1}{N} \sum_{k=1}^{N-1} V_N^* \left(\frac{k}{N}, \cdot \right) - (-6x + 6x^2) \int \Gamma_N(y, \cdot) dy \right\| \\ & \leq |b_N(x)| \left\| \frac{1}{N} \sum_{k=1}^{N-1} V_N^* \left(\frac{k}{N}, \cdot \right) - \int \Gamma_N(y, \cdot) dy \right\| + |b_N(x) - (-6x + 6x^2)| \left\| \int \Gamma_N(y, \cdot) dy \right\|. \end{aligned}$$

It suffices to prove that

$$\sup_{N \geq 1} \left\{ \sup_{0 \leq x \leq 1} |b_N(x)| \right\} < \infty, \quad \sup_{0 \leq x \leq 1} |b_N(x) - (-6x + 6x^2)| \rightarrow 0, \quad \left\| \int \Gamma_N(y, \cdot) dy \right\| = \mathcal{O}_{\bar{\mathbb{P}}_T}(1), \quad (2.29)$$

$$\left\| \frac{1}{N} \sum_{k=1}^{N-1} V_N^* \left(\frac{k}{N}, \cdot \right) - \int \Gamma_N(y, \cdot) dy \right\| \xrightarrow{\bar{\mathbb{P}}_T} 0. \quad (2.30)$$

The three claims (2.29) follow from the proof of Lemma 5.7 in Kokoszka and Young (2016) and (2.30) follows from Lemma 2.5.8. $\square$

Proof of Theorem 2.2.1. Now, we are ready to prove Theorem 2.2.1.

By Lemma 2.5.9, we have

$$\rho\{Z_N^*(x, \cdot), \Gamma_N^0(x, \cdot)\} := \sup_{0 \leq x \leq 1} \|Z_N^*(x, \cdot) - \Gamma_N^0(x, \cdot)\| \xrightarrow{\bar{\mathbb{P}}_T} 0.$$

By Theorem 2.5.2, we have $\Gamma_N^0(x, t) \stackrel{\bar{d}}{=} \Gamma^0(x, t)$. Thus Theorem 2.5.1 suggests that $Z_N^*(x, t) \xrightarrow{\bar{d}} \Gamma_N^0(x, t)$.

Moreover, by Lemma 2.5.7, we have

$$\iint \{\Gamma_N^0(x, t)\}^2 dx dt \stackrel{\bar{d}}{=} \sum_{i=1}^{\infty} \lambda_i \int_0^1 V_i^2(x) dx.$$

Therefore, by continuous mapping theorem, we obtain

$$R_N^* = \iint \{Z_N^*(x, t)\}^2 dx dt \xrightarrow{\bar{d}} \sum_{i=1}^{\infty} \lambda_i \int_0^1 V_i^2(x) dx,$$

which completes the proof of Theorem 2.2.1. $\square$

Proof of Proposition 2.2.1. Under Assumption 2.2.1, consider two bootstrap test statistics $R_N^{*(1)}$ and $R_N^{*(2)}$ from the same realization. Since

$$R_N^{*(1)} = \frac{1}{N^2} \sum_{k=1}^{N-1} \int \left\{ \sum_{n=1}^k \hat{\eta}_n^{*(1)}(t) \right\}^2 dt = \frac{1}{N^2} \sum_{k=1}^{N-1} \sum_{n=1}^k \sum_{m=1}^k \int \hat{\eta}_n^{*(1)}(t) \hat{\eta}_m^{*(1)}(t) dt,$$

we have

$$\bar{E}\{R_N^{*(1)}\} = \frac{1}{N^2} \sum_{k=1}^{N-1} \sum_{n=1}^k \sum_{m=1}^k \int \bar{E}\{\hat{\eta}_n^{*(1)}(t) \hat{\eta}_m^{*(1)}(t)\} dt.$$

For $1 \leq n, m \leq N-1$,

$$\begin{aligned} & \bar{E}\{\hat{\eta}_n^{*(1)}(t) \hat{\eta}_m^{*(1)}(t)\} \\ &= \bar{E}\left[\left\{\eta_n^{*(1)}(t) - \sum_{i=1}^N \left(\frac{1}{N} + \frac{k_n k_i}{s_N}\right) \eta_i^{*(1)}(t)\right\} \left\{\eta_m^{*(1)}(t) - \sum_{j=1}^N \left(\frac{1}{N} + \frac{k_m k_j}{s_N}\right) \eta_j^{*(1)}(t)\right\}\right] \\ &= \bar{E}\left\{\eta_n^{*(1)}(t) \eta_m^{*(1)}(t) - \eta_n^{*(1)}(t) \sum_{j=1}^N \left(\frac{1}{N} + \frac{k_m k_j}{s_N}\right) \eta_j^{*(1)}(t) - \eta_m^{*(1)}(t) \sum_{i=1}^N \left(\frac{1}{N} + \frac{k_n k_i}{s_N}\right) \eta_i^{*(1)}(t) \right. \\ & \quad \left. + \sum_{i=1}^N \sum_{j=1}^N \left(\frac{1}{N} + \frac{k_n k_i}{s_N}\right) \left(\frac{1}{N} + \frac{k_m k_j}{s_N}\right) \eta_i^{*(1)}(t) \eta_j^{*(1)}(t) \right\}. \end{aligned}$$

Note that

$$\begin{aligned} \bar{E}[\eta_n^{*(1)}(t) \eta_m^{*(1)}(t)] &= E[E^*\{\eta_n^{*(1)}(t) \eta_m^{*(1)}(t)\}] = E\left\{\frac{1}{N} \sum_{i=1}^N \hat{\eta}_i(t) \hat{\eta}_i(t)\right\} \\ &= E\left[\frac{1}{N} \sum_{i=1}^N \left\{\eta_i(t) - \sum_{j=1}^N \left(\frac{1}{N} + \frac{k_j k_i}{s_N}\right) \eta_j(t)\right\}^2\right] = \left(1 - \frac{2}{N}\right) E\{\eta_0^2(t)\} \end{aligned}$$

and $\bar{E}\{\eta_n^{*(1)}(t) \eta_m^{*(1)}(t)\} = E[E^{*(1)}(t) \{\eta_n^{*(1)}(t) \eta_m^{*(1)}(t)\}] = 0$ for $n \neq m$. Then, for $n \neq m$ we have

$$\bar{E}\{\hat{\eta}_n^{*(1)}(t) \hat{\eta}_m^{*(1)}(t)\} = -\left(\frac{1}{N} + \frac{k_m k_n}{s_N}\right) \left(1 - \frac{2}{N}\right) E\{\eta_0^2(t)\}.$$

Similarly,

$$\bar{E}\{\hat{\eta}_n^{*(1)}(t) \hat{\eta}_n^{*(1)}(t)\} = \left\{1 - \left(\frac{1}{N} + \frac{k_m k_n}{s_N}\right)\right\} \left(1 - \frac{2}{N}\right) E\{\eta_0^2(t)\}.$$

Then simple calculation yields

$$\bar{E}\{R_N^{*(1)}\} = \frac{1}{N^2} \sum_{k=1}^{N-1} \sum_{n=1}^k \sum_{m=1}^k \int \bar{E}\{\hat{\eta}_n^{*(1)}(t) \hat{\eta}_m^{*(1)}(t)\} dt$$

$$\begin{aligned}
&= \frac{1}{N^2} \sum_{k=1}^{N-1} \left\{ k - \sum_{n=1}^k \sum_{m=1}^k \left(\frac{1}{N} + \frac{k_n k_m}{s_N} \right) \right\} \left(1 - \frac{2}{N} \right) E(\|\eta_0\|^2) \\
&= \frac{1}{N^2} \left\{ \frac{N(N-1)}{2} - \frac{1}{N} \frac{(N-1)N(2N-1)}{6} - \frac{1}{s_N} \sum_{k=1}^{N-1} \left(\sum_{n=1}^k k_n \right)^2 \right\} \left(1 - \frac{2}{N} \right) E(\|\eta_0\|^2) \\
&\rightarrow \left\{ \frac{1}{2} - \frac{1}{3} - \left(-\frac{2}{5} \right) \right\} E(\|\eta_0\|^2) = \frac{17}{30} E(\|\eta_0\|^2), \text{ as } N \rightarrow \infty.
\end{aligned}$$

Similarly, we can obtain the same limit for $\bar{E}\{R_N^{*(2)}\}$ and thus

$$\bar{E}\{R_N^{*(1)}\} \bar{E}\{R_N^{*(2)}\} \rightarrow \left(\frac{17}{30} \right)^2 \{E(\|\eta_0\|^2)\}^2, \quad \text{as } N \rightarrow \infty. \quad (2.31)$$

Since $R_N^{*(1)}$ and $R_N^{*(2)}$ are independent on the bootstrap probability space, we have

$$\bar{E}\{R_N^{*(1)} R_N^{*(2)}\} = E[E^*\{R_N^{*(1)} R_N^{*(2)}\}] = E[E^*\{R_N^{*(1)}\} E^*\{R_N^{*(2)}\}].$$

Analogous to the previous derivation for $\bar{E}[R_N^{*(1)}]$, we have

$$\begin{aligned}
E^*\{R_N^{*(1)}\} &= \frac{1}{N^2} \left\{ \frac{N(N-1)}{2} - \frac{1}{N} \frac{(N-1)N(2N-1)}{6} - \frac{1}{s_N} \sum_{k=1}^{N-1} \left(\sum_{n=1}^k k_n \right)^2 \right\} \\
&\quad * \left\{ \frac{1}{N} \sum_{i=1}^N \int \hat{\eta}_i(t) \hat{\eta}_i(t) dt \right\}.
\end{aligned}$$

Hence, we obtain

$$\begin{aligned}
&E[E^*\{R_N^{*(1)}\} E^*\{R_N^{*(2)}\}] \\
&= \left[\frac{1}{N^2} \left\{ \frac{N(N-1)}{2} - \frac{1}{N} \frac{(N-1)N(2N-1)}{6} - \frac{1}{s_N} \sum_{k=1}^{N-1} \left(\sum_{n=1}^k k_n \right)^2 \right\} \right]^2 \\
&\quad * E\left[\left\{ \frac{1}{N} \sum_{i=1}^N \int \hat{\eta}_i(t) \hat{\eta}_i(t) dt \right\}^2 \right].
\end{aligned}$$

We note that

$$\left[\frac{1}{N^2} \left\{ \frac{N(N-1)}{2} - \frac{1}{N} \frac{(N-1)N(2N-1)}{6} - \frac{1}{s_N} \sum_{k=1}^{N-1} \left(\sum_{n=1}^k k_n \right)^2 \right\} \right]^2 \rightarrow \left(\frac{17}{30} \right)^2$$

and

$$E\left[\left\{ \frac{1}{N} \sum_{i=1}^N \int \hat{\eta}_i(t) \hat{\eta}_i(t) dt \right\}^2 \right] = \frac{1}{N^2} E\left[\left\{ \sum_{i=1}^N \int (\eta_i(t) - \sum_{p=1}^N \left(\frac{1}{N} + \frac{k_p k_i}{s_N} \right) \eta_p(t)) dt \right\}^2 \right]$$

$$\begin{aligned}
&= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \iint \mathbb{E} \left[\left\{ \eta_i(t) - \sum_{p=1}^N \left(\frac{1}{N} + \frac{k_p k_i}{s_N} \right) \eta_p(t) \right\}^2 \left\{ \eta_j(s) - \sum_{q=1}^N \left(\frac{1}{N} + \frac{k_q k_j}{s_N} \right) \eta_q(s) \right\}^2 \right] dt ds \\
&= \frac{1}{N^2} \sum_{i=1}^N C_i^2 \mathbb{E}(\|\eta_0\|^4) + \frac{1}{N^2} \left\{ \left(\sum_{i=1}^N C_i \right)^2 - \sum_{i=1}^N C_i^2 \right\} \{ \mathbb{E}(\|\eta_0\|^2) \}^2 \\
&\quad + \sum_{i \neq j} \left\{ \sum_{t=1}^N \left(\frac{1}{N} + \frac{k_t k_i}{s_N} \right) \left(\frac{1}{N} + \frac{k_t k_j}{s_N} \right) - \left(\frac{1}{N} + \frac{k_i k_j}{s_N} \right) \right\} \iint [\mathbb{E} \{ \eta_0(t) \eta_0(s) \}]^2 dt ds \\
&\rightarrow \{ \mathbb{E}(\|\eta_0\|^2) \}^2,
\end{aligned}$$

where

$$C_i = 1 + \sum_{j=1}^N \left(\frac{1}{N} + \frac{k_j k_i}{s_N} \right)^2 - 2 \left(\frac{1}{N} + \frac{k_i^2}{s_N} \right).$$

Therefore, we have

$$\bar{\mathbb{E}} \{ R_N^{*(1)} R_N^{*(2)} \} \rightarrow \left(\frac{17}{30} \right)^2 \{ \mathbb{E}(\|\eta_0\|^2) \}^2, \quad \text{as } N \rightarrow \infty. \quad (2.32)$$

By (2.31) and (2.32), the result (2.12) follows. $\square$

Functional Unit Root Test

As mentioned in Section 1.2, in this Chapter, we propose a functional unit root test (FURT) statistic based on the partial sum of estimated errors. Theoretically, we derive the limit distribution of the proposed statistic under the null hypothesis of integration of order one and the asymptotic behavior of the statistic under three different alternative hypotheses, namely trend stationary, weakly dependent stationary, and autoregressive stationary models. The test consistency is validated theoretically and numerically. We also apply the proposed FURT to intraday price curves of 50 randomly chosen US stocks.

Our study appears to be the first on functional unit root tests. To see the major mathematical contribution of this chapter, we realize that most existing FTS tests leverage a core function convergence theorem, namely Theorem 1.1 in Berkes et al. (2013), to develop the asymptotic theories of the test statistics under stationary or non-stationary hypotheses. The key to applying the theorem is to include the demeaning of estimated errors such that the mean will dominate the limit distribution of the statistic. However, when such a construction is extended to testing for unit root model, the partial sum of estimated errors becomes the difference between the last and first samples of FTS, which is not dominated by a mean term. Thus, Theorem 1.1 in Berkes et al. (2013) becomes unavailable. In this chapter, we manage to get rid of the reliance of the theorem and innovate a new analytical framework that is based on some basic limit theories.

3.1 Problem Formulation and Test Statistic

3.1.1 The Problem

Suppose that N time series data are available in functional form (i.e. each is a curve), denoted by $\mathcal{X}_N = \{X_n(t)\}_{n=1}^N$, where the index t is the argument of the curves. Without loss of generality, we assume that the curves are being normalized and defined on the unit interval $t \in [0, 1]$. Throughout this chapter, all the random functions and functional parameters are defined on a Hilbert space $L^2([0, 1])$ of all square-integrable real-valued functions on $[0, 1]$ with an inner product $\langle f, g \rangle := \int_0^1 f(t)g(t)dt$ for $f, g \in L^2([0, 1])$. To ease the notational burden, we suppress the limits of the integral, i.e. $\int f(t)dt := \int_0^1 f(t)dt$. We study the problem of our interest on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, on which we define $\mathbb{E}[\cdot]$ as the expectation operator.

With the definition of integrated FTS, we consider the null hypothesis of FTS being integration of order one ($I(1)$) as follows:

Unit root non-stationary:

$$H_0^{ur} : X_n(t) = \mu(t) + \sum_{i=1}^n \eta_i(t), \quad t \in [0, 1], \quad n = 1, \dots, N, \quad (3.1)$$

where the intercept μ is a deterministic functional parameter and the noises $\{\eta_n\}_{n=1}^N$ are a subsequence of Bernoulli shifts $\{\eta_n\}_{n=-\infty}^{\infty}$, which are defined below.

Definition 3.1.1 (Bernoulli shifts). The sequence $\{\eta_n\}_{n=-\infty}^{\infty}$ forms a Bernoulli shift-s if $\eta_n = f(\varepsilon_n, \varepsilon_{n-1}, \dots)$ for some measurable function $f : \Omega^\infty \mapsto L^2$ and independent and identically distributed (i.i.d.) functions $\varepsilon_n, -\infty < n < \infty$. Moreover, $\mathbb{E}[\eta_0(t)] = 0$ for all t and $\mathbb{E}\|\eta_0\|^{2+\delta} < \infty$ for some $0 < \delta < 1$, and the sequence $\{\eta_n\}_{n=-\infty}^{\infty}$ can be approximated by l -dependent sequences $\{\eta_{n,l}\}_{n=-\infty}^{\infty}$ such that $\sum_{l=1}^{\infty} (\mathbb{E}\|\eta_n - \eta_{n,l}\|^{2+\delta})^{1/\kappa} < \infty$ for some $\kappa > 2 + \delta$, where $\eta_{n,l}$ is defined by $\eta_{n,l} = f(\varepsilon_n, \dots, \varepsilon_{n-l+1}, \varepsilon_{n-l}^*, \varepsilon_{n-l-1}^*, \dots)$ with the ε^* 's being the independent copies of ε_0 and independent of any $\varepsilon_n, -\infty < n < \infty$.

For example, i.i.d. or autoregressive distributed noises $\{\eta_n\}_{n=1}^N$ with $\mathbb{E}[\eta_0(t)] = 0$ for all t and $\mathbb{E}\|\eta_0\|^{2+\delta} < \infty$ for some $0 < \delta < 1$ are the ordinary Bernoulli shifts.

All the μ and $\{\eta_n\}_{n=1}^N$ presented in this chapter will follow the same definition if there is no further notification. With the noises process defined above, our framework is general enough to nest most stationary time series models in practice; see Hörmann and Kokoszka (2010); Wu (2005) for the investigation of weak dependence and Pötscher and Prucha (1997) for its applications in econometrics.

For testing the unit root model, a general alternative hypothesis may be that H_0^{ur} does not hold. However, it is common to consider the following specific alternative hypotheses to reveal practical insights. Note that the consideration of the alternative hypothesis does not influence the construction of our test statistic and its limiting distribution but the testing rule, characterized by the critical value, will be determined accordingly to achieve a desirable power. The alternative hypotheses we consider are listed as follows:

Trend stationary:

$$H_1^{ts} : X_n(t) = \mu(t) + n\xi(t) + \eta_n(t), \quad t \in [0, 1], \quad n = 1, \dots, N, \quad (3.2)$$

which is considered in Kokoszka and Young (2016), where the slope $\xi \in L^2([0, 1])$ is a deterministic functional parameter. To be distinguished from the weakly dependent stationary model, we impose that $\|\xi\| > 0$.

Weakly dependent stationary:

$$H_1^{ws} : X_n(t) = \mu(t) + \eta_n(t), \quad t \in [0, 1], \quad n = 1, \dots, N, \quad (3.3)$$

which is considered in Horváth et al. (2014). Under this alternative hypothesis, we impose that the difference of the noises $\Delta\eta_n(t) = \eta_n(t) - \eta_{n-1}(t)$, $n = 2, \dots, N$ cannot be part of (another) Bernoulli shifts, e.g., i.i.d. or moving average noises are not allowed. This is to avoid the situation that (3.3) may be indifferent from the unit root non-stationary model under H_0^{ur} .

Autoregressive (of order 1, i.e. AR(1)) stationary:

$$H_1^{ar1} : X_n(t) - \mu(t) = \Psi(X_{n-1}(t) - \mu(t)) + \eta_n(t), \quad t \in [0, 1], \quad n = 2, \dots, N, \quad (3.4)$$

which is considered in Horváth et al. (2010), where Ψ is an integral operator with the kernel function $\psi(s, t)$ satisfying $\iint \psi^2(s, t) ds dt < 1$.

We remark that for the hypotheses above, the μ and ψ are unknown and the μ 's are estimated differently according to the corresponding hypothesis.

3.1.2 Test Statistic of Functional Unit Root Test

Given a set of observed FTS data $\{X_n(t)\}_{n=1}^N$, we consider the estimated noises under the null hypothesis H_0^{ur}

$$\hat{\eta}_n(t) = X_{n+1}(t) - X_n(t) \quad (3.5)$$

for $n \in \{1, 2, \dots, N-1\}$. Then a suitable functional unit root test (FURT) statistic is defined by

$$R_N = \iint Z_N^2(x, t) dt dx, \quad Z_N(x, t) = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \hat{\eta}_n(t). \quad (3.6)$$

Note that Z_N is the partial sum of estimated noises. The construction of the test statistic (3.6) is quite straightforward and belongs to the same class of “partial sum of estimated errors”-based construction as in Horváth et al. (2014); Kokoszka and Young (2016), while there are two points which should be paid attention.

1. The $\hat{\eta}_n(t)$ in (3.5) estimates $\eta_{n-1}(t)$ in (3.1) and we do not estimate $\eta_1(t)$ in (3.1). It is motivated by the fact that the estimate of the intercept $\mu(t)$ in (3.1) is usually inaccurate, especially for the financial and econometric data. Thus, we get rid of the estimation of $\eta_1(t)$ while it does not affect the property of our proposed test, noting that we are developing an asymptotic test.

2. The $Z_N(x, t)$ in (3.6) contains $\hat{\eta}_{\lfloor Nx \rfloor}(t)$ and thus probably $X_{N+1}(t)$, which is not in the observed FTS data. In fact, the test statistic R_N in (3.6) can be rewritten as

$$R_N = \frac{1}{N} \iint [X_{\lfloor Nx \rfloor + 1}(t) - X_1(t)]^2 dt dx = \frac{1}{N^2} \sum_{n=2}^N \int [X_n(t) - X_1(t)]^2 dt, \quad (3.7)$$

from which, we can see that the evaluation of the test statistic R_N does not involve $X_{N+1}(t)$. Despite of the simpler expression of R_N in (3.7), we still opt for the expression in (3.6) because it facilitates later asymptotic analyses and provides more intuitive uses of some existing results.

A key difference between our statistic (3.6) and the ones in Horváth et al. (2014); Kokoszka and Young (2016) is that we do not de-mean in the functional residual (3.5). Thanks to the special structure of $I(1)$, such a construction is sensible and as shown in the later context, it is appealing in theory and practice.

3.2 Asymptotic Theory of the Functional Unit Root Test

In this section, we aim to identify the limiting distribution of the test statistic R_N in (3.6) under the unit root non-stationary null hypothesis H_0^{ur} and study its asymptotic behavior under alternative hypotheses. Throughout the chapter, $W_1, W_2, \dots$ are denoted as independent standard Wiener processes on $[0, 1]$.

3.2.1 Limit Distribution of the Test Statistic under the Null Hypothesis

To derive the limit distribution of the test statistic R_N under H_0^{ur} , we first introduce and need to impose a regularity condition on the long-run covariance function of the noises $\{\eta_n\}$, which is defined as

$$c(t, s) = \mathbb{E}[\eta_0(t)\eta_0(s)] + \sum_{n=1}^{\infty} (\mathbb{E}[\eta_0(t)\eta_n(s)] + \mathbb{E}[\eta_0(s)\eta_n(t)]), \quad (3.8)$$

where $\eta_0(t)$ is i.i.d. as $\{\eta_n(t)\}_{n=1}^N$. By Lemma 2.2 in Berkes et al. (2013), the infinite sum in the definition of $c(t, s)$ is well-defined on $L^2([0, 1] \times [0, 1])$, i.e. $\iint c^2(t, s)dsdt < \infty$. By the Cauchy–Schwarz inequality, we also have

$$\left(\int c(t, t)dt \right)^2 \leq \int c^2(t, t)dt = \iint c^2(t, s)\mathbf{1}_{\{s=t\}}dsdt \leq \iint c^2(t, s)dsdt < \infty$$

and thus $\int c(t, t)dt < \infty$, which will be used in the later analyses. Moreover, the bivariate function $c(t, s)$ is positive-definite and therefore there exist eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ and the corresponding orthonormal eigenfunctions $\phi_1(t), \phi_2(t), \dots$, $t \in [0, 1]$ satisfying $\lambda_i \phi_i(t) = \int c(t, s) \phi_i(s) ds$ for $i \geq 1$. The eigen-decomposition of $c(t, s)$ plays a crucial role in many FTS tests, see Horváth et al. (2014); Kokoszka and Young (2016).

The following theorem shows the limit distribution of test statistic under $I(1)$ null hypothesis:

Theorem 3.2.1. Under the null hypothesis H_0^{ur} (3.1), we have

$$R_N \xrightarrow{d} R_\infty = \sum_{i=1}^{\infty} \lambda_i \int W_i^2(x) dx \quad \text{as } N \rightarrow \infty,$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ are the eigenvalues of the long-run covariance function $c(t, s)$ in (3.8).

The proof of Theorem 3.2.1, documented in Appendix 3.6, is quite straightforward. It is noteworthy that the limit distribution R_∞ depends on the Wiener processes instead of Brownian bridges as in Horváth et al. (2014); Kokoszka and Young (2016) mainly because the construction of our test statistic gets rid of the de-meaning step. From Theorem 3.6.2, we also know that $\sum_{i=1}^{\infty} \lambda_i < \infty$ under null hypothesis, which implies the test statistic tends to a finite distribution in probability one.

In implementation, we can use a kernel estimator $\hat{c}_N(t, s)$, e.g. proposed in Horváth et al. (2013), as a proxy of the long-run covariance function $c(t, x)$:

$$\hat{c}_N(t, s) = \hat{\gamma}_0(t, s) + \sum_{n=1}^{N-1} K\left(\frac{n}{h(N)}\right) (\hat{\gamma}_n(t, s) + \hat{\gamma}_n(s, t)), \quad \text{where } \hat{\gamma}_n(t, s) = \frac{1}{N} \sum_{j=n+1}^N \hat{\eta}_j(t) \hat{\eta}_{j-n}(s) \quad (3.9)$$

with $\hat{\eta}_n(t)$ given in (3.5) for $n = 1, 2, \dots, N$, in which the kernel function K and the smoothing bandwidth $h(N)$ should satisfy the followings:

- $K(0) = 1$, $K(u) = 0$ for $u > c$, and K is continuous on $[0, c]$ with some $c > 0$;
- $h(N) \rightarrow \infty$ but $h(N)/N \rightarrow 0$ as $N \rightarrow \infty$.

Once the kernel K and bandwidth $h(N)$ are specified, we can approximate the limit distribution R_∞ in Theorem 3.2.1 with

$$\hat{R}_\infty^{(M)} := \sum_{i=1}^M \hat{\lambda}_i \int W_i^2(x) dx, \quad (3.10)$$

where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq 0$ are the estimated eigenvalues from the kernel estimator $\hat{c}_{N-1}(t, s)$ in (3.9) and M is a large number. In this study, we choose $M = N$ as N is considered as a large number in our study.

Under the null hypothesis H_0^{ur} (3.1), since $\hat{\eta}_i(t) = \eta_{i+1}(t)$, it is proved in Horváth et al. (2013, Section 5) that

$$\iint [\hat{c}_N(t, s) - c(t, s)]^2 dt ds \xrightarrow{\mathbb{P}} 0 \quad \text{as } N \rightarrow \infty.$$

From the lemma in Dunford and Schwartz (1988, p.1091), because $\hat{c}_N(t, s)$ tends to $c(t, s)$, so $\hat{\lambda}_i$ tends to λ_i for every i . We can conclude that $\hat{R}_\infty^{(N)}$ is a consistent estimate of R_∞ under H_0^{ur} . Thus, the critical value of R_N can be deduced based on $\hat{R}_\infty^{(N)}$.

3.2.2 Asymptotic Behavior of the Test Statistic under Alternative Hypotheses

To validate an asymptotic test and to achieve a desirable testing power, it is crucial to understand the asymptotic behavior of the test statistic under the alternative hypothesis. In general, it is analytically challenging due to the inconsistency between the construction of the test statistic and the model under the alternative hypothesis. However, we manage to analyze our test statistic for three main classes of alternative hypotheses. The proofs of the results in this section are documented in Appendix 3.7.

Alternative Hypothesis H_1^{ts}

For the alternative hypothesis of the trend stationary model H_1^{ts} (3.2), our analysis is thorough in the sense that a combination of Theorems 3.2.2 and 3.2.3 below shows the asymptotic power of one.

We first examine the asymptotic behavior of the test statistic under H_1^{ts} .

Theorem 3.2.2. Under alternative hypothesis H_1^{ts} in (3.2), we have

$$\frac{1}{N}R_N \xrightarrow{\mathbb{P}} \|\xi\|^2 \quad \text{as } N \rightarrow \infty,$$

where ξ is the deterministic functional parameter in (3.2).

By Theorem 3.2.2 and noting that $\|\xi\|^2 > 0$ is independent of N , we have that under H_1^{ts} , the test statistic R_N tends to infinite in probability in the order of N .

Next, we perform asymptotic analysis of the estimated limit distribution $\hat{R}_\infty$ under H_1^{ts} .

Theorem 3.2.3. Under alternative hypothesis H_1^{ts} in (3.2), we have

$$\frac{1}{N}\hat{\lambda}_1 \xrightarrow{\mathbb{P}} 0 \quad \text{as } N \rightarrow \infty,$$

where $\hat{\lambda}_1$ is the largest eigenvalue of the estimated long-run covariance function $\hat{c}_N(t, s)$ defined in (3.9).

Theorems 3.2.2 and 3.2.3 imply that under H_1^{ts} , the largest eigenvalue of estimated long-run covariance function $\hat{c}_N(t, s)$ tends to infinite always slower than R_N in probability. By the result of Dunford and Schwartz (1988, p.1091), we know that $\hat{R}_\infty^{(N)}$ and the critical value (lower-tailed α -quantile) of $\hat{R}_\infty^{(N)}$, denoted by $\hat{d}_{N,\alpha}$, evolve slower than R_N in N (in probability). In other words, we can conclude that under H_1^{ts} ,

$$\mathbb{P}(R_N > \hat{d}_{N,1-\alpha}) \rightarrow 1 \quad \text{as } N \rightarrow \infty,$$

which implies that the H_0^{ur}/H_1^{ts} right-tailed test is consistent.

Alternative Hypotheses H_1^{ws} and H_1^{ar1}

For the alternative hypotheses of the weakly dependent stationary model H_1^{ws} (3.3) and AR(1) stationary model H_1^{ar1} (3.4), we manage to evaluate the order of the test statistic R_N while the order of the largest eigenvalue of the estimated long-run covariance function $\hat{c}_N(t, s)$, $\hat{\lambda}_1$, or $\hat{R}_\infty^{(N)}$ remains an open problem. In fact, Theorem 3.2.3 is appealing and serves as one of the few exceptions, as we notice that such an asymptotic

analysis of the limit distribution is scarce in the literature even for the scalar case. Fortunately, our numerical examples show that the comparison between the orders of R_N and $\hat{\lambda}_1$ under H_1^{ws} or H_1^{ar1} is clear-cut.

We first examine the asymptotic behavior of the test statistic under H_1^{ws} .

Theorem 3.2.4. Under alternative hypothesis H_1^{ws} in (3.3), we have

$$R_N = \mathcal{O}_{\mathbb{P}}(1/N).$$

Theorem 3.2.4 shows that the test statistic R_N tends to zero in probability in the order of $1/N$ under H_1^{ws} .

Under H_1^{ar1} , we need to specify the integral operator Ψ or its kernel ψ . The general ψ is difficult to study due to the intractable asymptotics of the $\hat{c}_N(t, s)$ under H_1^{ar1} . Therefore, we focus on a *constant kernel* function $\psi \equiv a$ for a constant $0 < a < 1$ and the analysis has been very complicated.

Theorem 3.2.5. Under alternative hypothesis H_1^{ar1} in (3.4) with kernel function $\psi(s, t) \equiv a$ for $0 < a < 1$, we have

$$R_N = \mathcal{O}_{\mathbb{P}}(1/N).$$

Theorem 3.2.5 shows that the test statistic R_N also tends to zero in probability in the order of $1/N$ under H_1^{ar1} with a constant kernel. We remark here that Theorem 3.2.5 is likely extendable to a piecewise constant kernel function by repeating the analysis in each piece. However, because it does not bring additional insights, we skip the possible extension such that this study is presented within reasonable limits of space.

Although it is difficult to study the largest eigenvalue $\hat{\lambda}_1$ of the estimated long-run covariance function under these alternative hypotheses, we conjecture that $\hat{\lambda}_1 = \mathcal{O}_{\mathbb{P}}(1/N^{1-\epsilon})$ for some $\epsilon > 0$, which is numerically shown in Section 3.3.3. In this case, again by the result of Dunford and Schwartz (1988, p.1091), we know that $\hat{R}_{\infty}^{(N)} = \mathcal{O}_{\mathbb{P}}(1/N^{1-\epsilon})$ and $\hat{d}_{N,\alpha} = \mathcal{O}_{\mathbb{P}}(1/N^{1-\epsilon})$. Then, we can conclude that under H_1^{ws} or H_1^{ar1} ,

$$\mathbb{P}(R_N < \hat{d}_{N,\alpha}) \rightarrow 1 \quad \text{as } N \rightarrow \infty,$$

which implies that the H_0^{ur}/H_1^{ws} and H_0^{ur}/H_1^{ar1} left-tailed tests are consistent.

3.2.3 Summary of the Asymptotic Tests

Theorems 3.2.2-3.2.5 reveal that the test statistic R_N in (3.6) behaves differently for different hypotheses and thus the decision rule has to be designed accordingly to achieve a desirable power. From the discussions in Section 3.2.2, we summarize our FURT with significance level α as follows:

- for H_1^{ts} , we reject H_0^{ur} if $R_N > \hat{d}_{N,1-\alpha}$;
- for H_1^{ws} or H_1^{ar1} , we reject H_0^{ur} if $R_N < \hat{d}_{N,\alpha}$,

where $\hat{d}_{N,\alpha}$ is the lower-tailed α -quantile of $\hat{R}_\infty^{(N)}$ given in (3.10) and in implementation, it can be determined based on Monte Carlo simulations.

3.3 Simulation Studies

In this section, we conduct simulation studies to examine the finite-sample performance of our proposed FURT. Following the setting in similar FTS tests of Chen and Pun (2019); Horváth et al. (2014); Kokoszka and Young (2016), we choose the kernel function

$$K(t) = \begin{cases} 1, & \text{if } 0 \leq |t| < 0.1, \\ 1.1 - |t|, & \text{if } 0.1 \leq |t| < 1.1, \\ 0, & \text{if } |t| \geq 1.1. \end{cases}$$

and the best smoothing bandwidth $h = N^{2/5}$ in the empirical size and power tests. Moreover, with the same setting, we also conduct a numerical study to examine the rate of changes of the largest eigenvalue $\hat{\lambda}_1$ of estimated long run covariance function $\hat{c}_N(t, s)$ with respect to N .

3.3.1 Details of Simulation and Implementation of FURT

To resemble the real-data analysis, we collect the empirical data of monthly US treasury yield curves from January 1970 to December 2002 in R package “fds,” denoted by

$\{X_n(t)\}_{n=1}^{396}$ (the curves are normalized and defined on $[0, 1]$). We first use these data to calibrate the model parameters under the respective hypothesis. Then, we simulate FTS data of length N , denoted by $\{\widetilde{X}_n(t)\}_{n=1}^N$, under different hypotheses with the calibrated model parameters. We consider different sample sizes: $N = 100$ and 200 to examine the effects of sample size on our FURT. For each hypothesis and sample size N , we repeat the simulation 1000 times to evaluate the testing performance. In all computations, we use 240 equidistant points to estimate every functional data, i.e. the curve values at $t_i = i/240$, $i = 1, \dots, 240$.

To examine the empirical size and power of the FURT, we apply it to the simulated data, which are generated under the hypotheses H_0^{ur} , H_1^{ts} , H_1^{ws} , and H_1^{ar1} in (3.1) and (3.2)-(3.4), respectively. For the model parameter calibration, we simply estimate the mean function $\mu(t)$ by $\hat{\mu}(t) = X_1(t)$ for $t \in [0, 1]$ and estimate the noises with $\{\hat{\eta}_n(t)\}_{n=1}^{395}$ by (3.5) from the real data $\{X_n(t)\}_{n=1}^{396}$. We find the first two principle components $\hat{\phi}_1(t)$ and $\hat{\phi}_2(t)$ of the empirical covariance function of $\{\hat{\eta}_n(t)\}_{n=1}^{395}$, and then compute the standard errors of the first two functional scores of $\{\hat{\eta}_n(t)\}_{n=1}^{395}$, denoted by $\hat{\sigma}_1$ and $\hat{\sigma}_2$. The $\hat{\mu}(t)$ and the $\{\hat{\eta}_n(t)\}_{n=1}^{395}$ are then used in the simulations under different hypotheses.

In the simulations of different models, the $\mu(t)$ is replaced by $\hat{\mu}(t)$ and the noises $\{\tilde{\eta}_n(t)\}_{n=1}^N$ are simulated as follows $\tilde{\eta}_n(t) = a_n \hat{\phi}_1(t) + b_n \hat{\phi}_2(t)$, $t \in [0, 1]$, $n = 1, \dots, N$, where $\{a_n\}_{n=1}^N$ and $\{b_n\}_{n=1}^N$ are i.i.d. or AR(1) random sequences following normal distribution or Student's t -distribution with 5 degrees of freedom (t_5 distribution) with mean equal to 0 and variances equal to $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$, respectively. Specifically,

- to generate t_5 -distributed variables with a specified variance, we multiply the standard t_5 -distributed random variables by a suitable constant ($\sigma \sqrt{3/5}$);
- to generate AR(1) sequences $\{a_n\}_{n=1}^N$, we set $a_1 = \epsilon_1$ and $a_n = \frac{1}{2}a_{n-1} + \frac{\sqrt{3}}{2}\epsilon_n$ for $n = 2, \dots, N$, where $\epsilon_1, \epsilon_2, \dots, \epsilon_N$ follow i.i.d. $N(0, \hat{\sigma}_1^2)$. The same applies to the generation of AR(1) sequences $\{b_n\}_{n=1}^N$.

Then, we are ready to simulate the data $\{\widetilde{X}_n(t)\}_{n=1}^N$ under different hypotheses:

- Under H_0^{ur} in (3.1), the simulated data are constructed as

$$\widetilde{X}_n^{ur}(t) := \hat{\mu}(t) + \sum_{i=1}^n \widetilde{\eta}_n(t), \quad t \in [0, 1], \quad n = 1, 2, \dots, N.$$

- Under H_1^{ts} in (3.2), we simulate a standard Wiener process $\{W(t)\}_{t \in [0, 1]}$ and set the slope $\widetilde{\xi}(t) \equiv 5W(t)$, where $W(t)$ is fixed throughout all simulations. The motivation behind this specification is to ensure that it is significant in the comparison with the noises and there is no specific pattern for the slope. Then, the simulated data under H_1^{ts} are constructed as

$$\widetilde{X}_n^{ts}(t) := \hat{\mu}(t) + n\widetilde{\xi}(t) + \widetilde{\eta}_n(t), \quad t \in [0, 1], \quad n = 1, 2, \dots, N.$$

- Under H_1^{ws} in (3.3), the simulated data under H_1^{ws} are constructed as

$$\widetilde{X}_n^{ws}(t) := \hat{\mu}(t) + \widetilde{\eta}_n(t), \quad t \in [0, 1], \quad n = 1, \dots, N.$$

- Under H_1^{ar1} in (3.4), the simulated data under H_1^{ar1} are constructed as

$$\widetilde{X}_1^{ar1}(t) = \hat{\mu}(t) + \widetilde{\eta}_1(t) \quad \text{and} \quad \widetilde{X}_n^{ar1}(t) := \hat{\mu}(t) + \int \psi(t, s) [\widetilde{X}_{n-1}^{ar1}(s) - \hat{\mu}(s)] ds + \widetilde{\eta}_n(t)$$

for $t \in [0, 1]$, $n = 2, \dots, N$, where we consider two kernel functions:

$$\psi_1(t, s) = 0.9 \quad \text{and} \quad \psi_2(t, s) = 0.3416 * \exp\left(\frac{t^2 + s^2}{2}\right), \quad t, s \in [0, 1],$$

in which ψ_1 is a *constant* kernel and ψ_2 is an *exponential* kernel with the Hilbert–Schmidt norm being approximately 0.5. The setting of the exponential kernel follows that in Horváth et al. (2014).

For each set of the simulated FTS, we construct our FURT statistic in (3.6) or (3.7) with $\{\widetilde{X}_n(t)\}_{n=1}^N$. According to the decision rules in Section 3.2.3, we record the decision of rejection or not. Then for 1000 simulations, the rejection frequency for the simulated FTS under the null hypothesis will be an empirical size and that for the simulated FTS under alternative hypotheses will be an empirical power of FURT.

3.3.2 Simulation Results: Empirical Sizes and Powers

Tables 3.1 and 3.2 present different testing results under different hypotheses. The numbers in the tables are the rejection frequencies (in percentage) among 1000 simulations and they are interpreted as empirical size or power according to null or alternative hypothesis being considered, respectively. For each sample size N , we consider two significance levels $\alpha = 5\%$ and 10% . Since different decision rules are used for H_0^{ur}/H_1^{ts} test and H_0^{ur}/H_1^{ws} (or H_0^{ur}/H_1^{ar1}), the empirical sizes will be examined differently and thus the results for the former are presented in Table 3.1 while the results for the latter are presented in Table 3.2.

Noise type	i.i.d. Normal				AR(1) Normal			
N	100		200		100		200	
α	5%	10%	5%	10%	5%	10%	5%	10%
H_0^{ur}	3.6	5.5	3.7	9.8	3.3	7.4	3.4	8.5
H_1^{ts}	99.4	100.0	100.0	100.0	99.1	100.0	100.0	100.0
Noise type	i.i.d. Student's t_5				AR(1) Student's t_5			
N	100		200		100		200	
α	5%	10%	5%	10%	5%	10%	5%	10%
H_0^{ur}	4.2	10.1	3.0	7.4	4.0	8.8	3.6	7.8
H_1^{ts}	99.6	99.8	99.9	99.9	99.7	99.7	99.7	99.7

Table 3.1 Empirical sizes (%) and powers (%) for FURT's for H_0^{ur} versus H_1^{ts} .

First, we focus on the empirical sizes in the Tables 3.1 and 3.2. The deviation of the estimated asymptotic distribution of R_∞ from the empirical distribution of R_N can be characterized by the difference between the empirical size and the desired significance level. No matter which of the left-tailed and right-tailed significance tests we consider, we can see that all the empirical sizes are close to the corresponding specified significance levels under different stochastic environments. It can also be observed that the empirical size will be generally closer to the significance level when

Noise type	i.i.d. Normal				AR(1) Normal			
N	100		200		100		200	
α	5%	10%	5%	10%	5%	10%	5%	10%
H_0^{ur}	3.5	6.3	4.2	6.7	3.6	6.5	4.0	6.9
H_1^{ws}	//////////		//////////		59.0	75.3	66.8	85.5
H_1^{ar1} (constant kernel)	60.6	85.3	95.4	98.2	62.0	83.1	96.0	98.9
H_1^{ar1} (exponential kernel)	56.4	75.4	60.0	81.3	79.7	90.1	92.0	97.7
Noise type	i.i.d. Student's t_5				AR(1) Student's t_5			
N	100		200		100		200	
α	5%	10%	5%	10%	5%	10%	5%	10%
H_0^{ur}	2.1	6.6	4.1	9.3	2.2	6.9	3.3	9.4
H_1^{ws}	//////////		//////////		66.9	80.8	75.6	85.0
H_1^{ar1} (constant kernel)	62.0	83.1	96.0	98.9	65.4	85.7	98.8	99.6
H_1^{ar1} (exponential kernel)	60.0	76.3	71.4	82.0	79.7	89.2	92.2	95.8

Table 3.2 Empirical sizes (%) and powers (%) for FURT for H_0^{ur} versus H_1^{ws} or H_1^{ar1} .

N increases. It is noteworthy that in the right-tailed test (H_0^{ur}/H_1^{ts}), the empirical size converges faster to the significance level for the normal noises than the t_5 noises but in the left-tailed test (H_0^{ur}/H_1^{ws} (or H_0^{ur}/H_1^{ar1})), the comparison result is opposite, where the convergence is faster for the t_5 noises. It is related to the convergence of the distribution of the test statistic and it happens that different tails would have different convergence rates. With the observations above, the result of Theorem 3.2.1 is validated numerically.

Next, we examine the performance (empirical power) of the tests under different alternative hypotheses. It can be seen from Table 3.1 that the H_0^{ur}/H_1^{ts} test has very high values of power under different stochastic environments. They are mostly very close to 100% while all exceed 99%. As implied by our asymptotic analysis, a significant slope in the trend stationary model is the key distinction from the unit root model. Our FURT works very well for the H_0^{ur}/H_1^{ts} test. For the weakly dependent FTS, we

exclude the studies with the i.i.d. noises since in this case, the H_0^{ur} nests H_1^{ws} and the formulation of the FURT is ill-posed. From Table 3.2, we can observe that the powers of the H_0^{ur}/H_1^{ws} test are lower than that of other tests but we can still obtain about 85% of empirical power when $N = 200$. It can be understood from the matter of fact that H_0^{ur} and H_1^{ws} are similar due to the general definition of the noises. To achieve a desirable power, one would need to have more data, i.e. a larger sample size N . For the AR(1) stationary FTS, two kernel functions are considered, where for the constant kernel, we pick the coefficient of 0.9 that is fairly close to 1, leading to a near unit root test. Again, the empirical powers are high, no matter which kernel function or which type of noises is considered. An interesting observation is that the H_0^{ur}/H_1^{ar1} test with the exponential kernel does not uniformly perform better than the test with the constant kernel, which may be counter-intuitive because the exponential kernel we consider possesses a norm of 0.5, much smaller than 0.9. We thus remark that the norm of the coefficient function of the AR model does not necessarily make a distinction between stationary and non-stationary models as it is an overall measure.

In a brief summary, the results in Tables 3.1 and 3.2 validate the asymptotic theory in Section 3.2 in the sense that the empirical sizes are desirable and the empirical powers are close to 1. Our simulation studies examine different alternative hypotheses, stochastic environments, sample sizes, and significance levels. The plausible results show that our FURT is robust.

3.3.3 Trends of Largest Eigenvalues under H_1^{ws} and H_1^{ar1}

In this subsection, we investigate the growth rate of the largest eigenvalue $\hat{\lambda}_1$ of the estimated long-run covariance function $\hat{c}_N(t, s)$ under H_1^{ws} and H_1^{ar1} with the constant kernel. We adopt the same setting in Section 3.3.1. To examine the growth of $\hat{\lambda}_1$ over a wide range of N values, we consider the simulated data with $N = 800$ and AR(1) normal noises. Then we compute $\hat{\lambda}_1$'s with truncated data of sample size N increasing from 2 to 800. Inspired by Horváth et al. (2014); Kokoszka and Young (2016), we also conjecture that the growth rate will be affected by the bandwidth $h(N)$ and thus we

consider two different bandwidth functions $h_1(N) = N^{\frac{1}{5}}$ and $h_2(N) = N^{\frac{2}{5}}$.

Figure 3.1 Plots of the largest eigenvalues $\hat{\lambda}_1$ of $\hat{c}_N(t, s)$ under H_1^{ws} (left) and H_1^{ar1} (right) with bandwidth $h = N^{\frac{1}{5}}$ (blue), $N^{\frac{2}{5}}$ (red) against the sample size N .

From Figure 3.1, we can see that the $\hat{\lambda}_1$ is generally decreasing in N for both bandwidth functions under both H_1^{ws} and H_1^{ar1} . The convergence rate for the bandwidth $N^{\frac{2}{5}}$ is much faster than that with the bandwidth $N^{\frac{1}{5}}$. Next, we further examine whether the convergence rate of $\hat{\lambda}_1$ is slower than $\mathcal{O}(1/N)$, the convergence rate of R_N under H_1^{ws} and H_1^{ar1} with the constant kernel. To this end, we plot $N\hat{\lambda}_1$ against N in Figure 3.2.

Four panels of Figure 3.2 show that the $N\hat{\lambda}_1$ is generally increasing in N for both bandwidth functions under both H_1^{ws} and H_1^{ar1} , which implies that $N\hat{\lambda}_1$ is dominated by a leading order term of N^ϵ with a positive constant ϵ . Hence, we numerically validate the test consistency we discussed in Section 3.2.2.

3.4 Application of FURT to Intraday Stock Price Curves

A leading example of functional unit root model is the intraday stock price curves. The analysis of the intraday price curves has attracted much attention since 2010. For example, Gabrys et al. (2013, 2010); Kokoszka et al. (2015); Kokoszka and Reimherr (2013); Müller et al. (2011) use appropriate methods to study the shapes of price or

Figure 3.2 Plots of $N\hat{\lambda}_1$ under H_1^{ws} (upper) and H_1^{ar1} (lower) with bandwidth $h = N^{\frac{1}{5}}$ (left, blue), $N^{\frac{2}{5}}$ (right, red) against the sample size N .

volatility curves. For functional FTS testing, Horváth et al. (2013) tested the stationarity of the intraday price curves and Kokoszka and Young (2017) tested the trend stationarity. However, from the results of these two tests, the null hypothesis of stationary or trend stationary is rejected for most of the intraday price curves if we do not make specific transformations on the original curves. Therefore, we apply our FURT on the intraday stock price curves and evaluate its empirical performance.

The data we use are minutely price data of 50 US stocks randomly chosen from the Bats (BYX) Exchange. The period of the data is from 2 Jan 2020 to 29 May 2020,

Figure 3.3 p -values of FURT of 50 intraday stock price curves. The left (resp.) panel presents the p -values of the left-tailed (resp. right-tailed) tests. The horizontal lines of 5% and 10% are added to indicate that the point below the line suggest a rejection of FURT.

resulting in $N = 103$ curves for each stock while some may only have 101 curves due to missing data. The data are accessible from a Russian website *finam.ru* or the corresponding author's website¹. Since not all the intraday price data are available in every minute, we simply adopt B-spline approach to estimate and obtain 390 time points for every day and each stock. Then we apply our FURT on the fitted curves for each stock. Two types of FURTs, according to Section 3.2.3, are considered, namely the left-tailed test (for H_0^{ur}/H_1^{ws} or H_0^{ur}/H_1^{ar1}) and the right-tailed test (for H_0^{ur}/H_1^{ts}). The test results are presented in Figure 3.3 in terms of p -values.

From Figure 3.3, we can first see that the p -values are spread quite uniformly in $[0, 1]$. If we consider H_1^{ws} or H_1^{ar1} as an alternative, we will reject the null hypothesis for 1 stock for the significance level of 5% and 9 stocks for the significance level of 10%, which are marked in blue and light blue, respectively in the left panel of Figure 3.3. If we consider H_1^{ts} as an alternative, we will reject the null hypothesis for 1 stock

¹The FTS data used in our empirical study can be downloaded from <https://personal.ntu.edu.sg/cspun/FURTDData.zip>.

Figure 3.4 Price curves of 8x8 Inc. from 2 Jan 2020 to 28 Feb 2020.

for the significance level of 5% and 4 stocks for the significance level of 10%, which are marked in green and light green, respectively in the right panel of Figure 3.3. Hence, there are 48 stocks (resp. 37 stocks), for which we cannot reject the null hypothesis of unit root model against any alternatives with the significance level of 5% (resp. 10%). The results are expected and aligned with our understanding of the financial market. Figure 3.4 shows the price curves of the first 40 days for "8x8 Inc." as an example, which cannot be rejected in both two tailed tests. The gradually changing colors from yellow to gray and then to blue represent the price curves in 40 days in the time order. From Figure 3.4, we find that the price curves in 40 days does not have a significant trend, but the curves with similar colors are also relatively close. We show the price difference between adjacent days in Figure 3.5. We also use gradually changing colors from yellow to gray and then to blue to represent the 39 differences in the order of time. It is clear that all the curves fluctuate around the horizontal line $y = 0$. Figures 3.4 and 3.5 visualize its unit root non-stationarity, which also validate the result of our unit root test.

Figure 3.5 Difference of price curves of 8x8 Inc. between adjacent days from 2 Jan 2020 to 28 Feb 2020.

3.5 Conclusion

In this chapter, we propose a unit root test for FTS. We derive the limit distribution of the proposed test statistic under the $I(1)$ null hypothesis and investigate the asymptotic behavior of the test statistic under three alternative hypotheses. We show the test consistency theoretically. With the simulated data and real data, we demonstrate the superior FURT performance numerically and empirically.

There are some open problems remaining. It will be interesting to analyze the FURT under more other alternative hypotheses to bring more practical insights. However, the asymptotic theory, especially for the largest eigenvalue $\hat{\lambda}_1$ of the estimated long-run covariance function, is non-trivial. Moreover, the construction of the test statistic also has its critics that limit its applicability and ours is no exception. To derive the asymptotic results, this chapter adopts an innovative approach, which is hopefully inspirational to the studies of nonstationary FTS tests.

This study fills in a research gap on testing a functional unit root model. As we know, the unit root test is a building block of co-integration analysis. We have put it into our research agenda.

3.6 Proof of the theorems in Section 3.2.1

3.6.1 Preliminary Results

Before we present the proof of Theorem 3.2.1, we need some preliminary results. Theorem 3.6.1 is a well-known theorem with the proof in Billingsley (1968). Theorem 3.6.2 resembles Theorem 1.1 in Berkes et al. (2013), it is an important theorem to show the convergence of the sum of the eigenvalues of the long-run covariance function and the sum of the noises.

Theorem 3.6.1. Suppose Z_N, Y_N and Y are random variables taking values in a separable metric space with distance function ρ . If $Y_N \xrightarrow{d} Y$ and $\rho(Z_N, Y_N) \xrightarrow{\mathbb{P}} 0$, then $Z_N \xrightarrow{d} Y$.

Theorem 3.6.2. Under the null hypothesis H_0^{ur} in (3.1),

1. $\sum_{i=1}^{\infty} \lambda_i < \infty$, where $\lambda_1, \lambda_2, \dots$ are eigenvalues of long-run covariance function $c(t, s)$ in (3.8).

2. We can define a sequence of Gaussian processes $\Gamma_N(x, t)$ such that:

$$\{\Gamma_N(x, t), 0 \leq x, t \leq 1\} \stackrel{d}{=} \{\Gamma(x, t), 0 \leq x, t \leq 1\}$$

and

$$\kappa_N = \sup_{0 \leq x \leq 1} \|V_N(x, \cdot) - \Gamma_N(x, \cdot)\| = o_{\mathbb{P}}(1),$$

where $\Gamma(x, t) = \sum_{i=1}^{\infty} \lambda_i^{1/2} W_i(x) \phi_i(t)$ and $V_N(x, t) = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \eta_n(t)$.

3.6.2 Proof of Theorem 3.2.1

Noting that under H_0^{ur} ,

$$Z_N(x, t) = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \hat{\eta}_n(t) = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \{X_{n+1}(t) - X_n(t)\} = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \eta_n(t),$$

we have

$$\sup_{0 \leq x \leq 1} \|Z_N(x, \cdot) - \Gamma_N(x, \cdot)\| = o_{\mathbb{P}}(1)$$

from Theorem 3.6.2. Hence, by Theorem 3.6.1, we have

$$Z_N(x, t) \xrightarrow{d} \Gamma(x, t)$$

and by the continuous mapping theorem, we thus have

$$R_N = \iint Z_N^2(x, t) dx dt \xrightarrow{d} \iint (\Gamma(x, t))^2 dx dt = \sum_{i=1}^{\infty} \lambda_i \int W_i^2(x) dx,$$

which completes the proof of Theorem 3.2.1.

3.7 Proofs of the theorems in Section 3.2.2

3.7.1 Proof of Theorem 3.2.2

Under alternative hypothesis H_1^{ts} ,

$$\begin{aligned} Z_N(x, t) &= \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} [X_{n+1}(t) - X_n(t)] = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} [\xi(t) + \eta_{n+1}(t) - \eta_n(t)] \\ &= \frac{1}{\sqrt{N}} [\lfloor Nx \rfloor \xi(t) + \eta_{\lfloor Nx \rfloor + 1}(t) - \eta_1(t)]. \end{aligned}$$

Hence, we have

$$\begin{aligned} R_N &= \frac{1}{N} \iint [\lfloor Nx \rfloor \xi(t) + \eta_{\lfloor Nx \rfloor + 1}(t) - \eta_1(t)]^2 dx dt \\ &= \frac{1}{N^2} \sum_{n=1}^{N-1} \int [n\xi(t) + \eta_{n+1}(t) - \eta_1(t)]^2 dt \\ &= \frac{(N-1)N(2N-1)}{6N^2} \|\xi\|^2 + \frac{2}{N^2} \sum_{n=1}^{N-1} \int n\xi(t) [\eta_{n+1}(t) - \eta_1(t)] dt \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{N^2} \sum_{n=1}^{N-1} \int [\eta_{n+1}(t) - \eta_1(t)]^2 dt \\
& =: R_1^* + R_2^* + R_3^*.
\end{aligned}$$

By the assumption of non-zero $\|\xi\|^2$, it is easy to see that R_1^* is of order N . Next, we evaluate the orders of R_2^* and R_3^* . For R_2^* , we note that

$$\begin{aligned}
& \left| \sum_{n=1}^{N-1} \int n \xi(t) [\eta_{n+1}(t) - \eta_1(t)] dt \right| \leq \sum_{n=1}^{N-1} n \int |\xi(t) [\eta_{n+1}(t) - \eta_1(t)]| dt \\
& \leq \frac{1}{2} \sum_{n=1}^{N-1} n \int [\xi^2(t) + [\eta_{n+1}(t) - \eta_1(t)]^2] dt \\
& = \frac{N(N-1)}{4} \|\xi\|^2 + \frac{1}{2} \sum_{n=1}^{N-1} n \int [\eta_{n+1}(t) - \eta_1(t)]^2 dt.
\end{aligned}$$

Then a direct calculation yields that

$$\begin{aligned}
\mathbb{E} \left[\sum_{n=1}^{N-1} n \int [\eta_{n+1}(t) - \eta_1(t)]^2 dt \right] & \leq \mathbb{E} \left[\sum_{n=1}^{N-1} 2n \int [\eta_{n+1}^2(t) + \eta_1^2(t)] dt \right] \\
& = 2N(N-1) \mathbb{E} \|\eta_0\|^2 = \mathcal{O}(N^2).
\end{aligned}$$

Hence, we have $R_2^* = \mathcal{O}_{\mathbb{P}}(1)$.

Similarly, we have

$$\mathbb{E}[R_3^*] \leq \mathbb{E} \left[\frac{2}{N^2} \sum_{n=1}^{N-1} \int [\eta_{n+1}^2(t) + \eta_1^2(t)] dt \right] = \frac{4(N-1)}{N^2} \mathbb{E} \|\eta_0\|^2 = \mathcal{O} \left(\frac{1}{N} \right).$$

Therefore, we can conclude that under H_1^{ts} , we have

$$\frac{1}{N} R_N \xrightarrow{\mathbb{P}} \frac{\|\xi\|^2}{3}, \quad \text{as } N \rightarrow \infty.$$

3.7.2 Proof of Theorem 3.2.3

We mainly investigate the order of the estimate of the long-run covariance function $\hat{c}_N(t, s)$. Under the alternative hypothesis H_1^{ts} , we have

$$\hat{c}_N(t, s) = \hat{\gamma}_0(t, s) + \sum_{i=1}^{N-1} K \left(\frac{i}{h(N)} \right) (\hat{\gamma}_i(t, s) + \hat{\gamma}_i(s, t)) \quad (3.11)$$

with

$$\begin{aligned}
\hat{\gamma}_i(t, s) &= \frac{1}{N} \sum_{j=i+1}^N [X_{j+1}(t) - X_j(t)][X_{j-i+1}(s) - X_{j-i}(s)] \\
&= \frac{1}{N} \sum_{j=i+1}^N [\xi(t) + \eta_{j+1}(t) - \eta_j(t)][\xi(s) + \eta_{j-i+1}(s) - \eta_{j-i}(s)], \\
&= \frac{N-i}{N} \xi(t)\xi(s) + \frac{1}{N} \sum_{j=i+1}^N [(\eta_{j+1}(t) - \eta_j(t))(\eta_{j-i+1}(s) - \eta_{j-i}(s)) \\
&\quad + \frac{1}{N} \sum_{j=i+1}^N [\xi(t)(\eta_{j-i+1}(s) - \eta_{j-i}(s)) + (\eta_{j+1}(t) - \eta_j(t))\xi(s)] \\
&=: \hat{\gamma}_i^{(1)}(t, s) + \hat{\gamma}_i^{(2)}(t, s) + \hat{\gamma}_i^{(3)}(t, s)
\end{aligned}$$

for $i = 0, 1, \dots, N-1$. By substituting different parts of $\hat{\gamma}_i^{(j)}$ for $i = 0, 1, \dots, N-1$ into $\hat{c}_N(t, s)$, we denote it by $\hat{c}_N^{(j)}(t, s)$ for $j = 1, 2, 3$ and they compose $\hat{c}_N(t, s)$: $\hat{c}_N(t, s) = \hat{c}_N^{(1)}(t, s) + \hat{c}_N^{(2)}(t, s) + \hat{c}_N^{(3)}(t, s)$.

For the first term, it is easy to obtain that

$$\begin{aligned}
\hat{c}_N^{(1)}(t, s) &= \hat{\gamma}_0^{(1)}(t, s) + \sum_{i=1}^{N-1} K\left(\frac{i}{h(N)}\right) [\hat{\gamma}_i^{(1)}(t, s) + \hat{\gamma}_i^{(1)}(s, t)] \\
&= \xi(t)\xi(s) + \frac{1}{N} \sum_{i=1}^{N-1} 2(N-i)K\left(\frac{i}{h(N)}\right) \xi(t)\xi(s).
\end{aligned}$$

Hence, we have

$$\begin{aligned}
\|\hat{c}_N^{(1)}(t, s)\|^2 &= \left[1 + \frac{1}{N} \sum_{i=0}^{N-1} 2(N-i)K\left(\frac{i}{h(N)}\right)\right]^2 \iint \xi^2(t)\xi^2(s) dt ds \\
&\leq \left[2 \sum_{i=0}^{N-1} \left|K\left(\frac{i}{h(N)}\right)\right|\right]^2 \|\xi\|^4 = \mathcal{O}\left(h^2(N) \left[\int_0^\infty |K(u)| du\right]^2\right) = \mathcal{O}(h^2(N)),
\end{aligned}$$

which implies that

$$\|\hat{c}_N^{(1)}(t, s)\| = \mathcal{O}(h(N)). \quad (3.12)$$

For the second term of $\hat{c}_N^{(2)}(t, s)$, we first expand $\hat{\gamma}_i^{(2)}(t, s)$ into four terms:

$$\begin{aligned}
\hat{\gamma}_i^{(2)}(t, s) &= \frac{1}{N} \sum_{j=i+1}^N [\eta_{j+1}(t)\eta_{j-i+1}(s) + \eta_j(t)\eta_{j-i}(s) - \eta_j(t)\eta_{j-i+1}(s) - \eta_{j+1}(t)\eta_{j-i}(s)] \\
&=: \hat{\gamma}_i^{(2,1)}(t, s) + \hat{\gamma}_i^{(2,2)}(t, s) - \hat{\gamma}_i^{(2,3)}(t, s) - \hat{\gamma}_i^{(2,4)}(t, s).
\end{aligned}$$

Analogously, we write $\hat{c}_N^{(2)}(t, s) := \hat{c}_N^{(2,1)}(t, s) + \hat{c}_N^{(2,2)}(t, s) + \hat{c}_N^{(2,3)}(t, s) + \hat{c}_N^{(2,4)}(t, s)$ where $\hat{c}_N^{(2,k)}(t, s)$ is equal to $\hat{c}_N(t, s)$ in (3.11) with $\hat{\gamma}_i$ replaced by $\hat{\gamma}_i^{(2,k)}$ for $k = 1, 2, 3, 4$. Then, we evaluate $\hat{c}_N^{(2,k)}(t, s)$ sequentially. For the $\hat{c}_N^{(2,1)}(t, s)$, by the ergodic theorem for random variables in the Hilbert space, we have

$$\left\| \hat{\gamma}_0(t, s)^{(2,1)} - \mathbb{E}[\eta_0(t)\eta_0(s)] \right\| = o_{\mathbb{P}}(1).$$

Hence, as in Horváth et al. (2013, Appendix A), we obtain that

$$\left\| \sum_{i=1}^{N-1} K\left(\frac{i}{h(N)}\right) \hat{\gamma}_i^{(2,1)}(t, s) - \tilde{c}(t, s) \right\| = o_{\mathbb{P}}(1), \quad \forall t, s,$$

where $\tilde{c}(t, s) = \sum_{i \geq 1} \mathbb{E}[\eta_0(s)\eta_i(t)]$. Therefore, we have

$$\left\| \hat{c}_N^{(2,1)}(t, s) - c(t, s) \right\| = o_{\mathbb{P}}(1). \quad (3.13)$$

Similarly, we can also obtain that

$$\left\| \hat{c}_N^{(2,2)}(t, s) - c(t, s) \right\| = o_{\mathbb{P}}(1). \quad (3.14)$$

For $\hat{c}_N^{(2,3)}(t, s)$ and $\hat{c}_N^{(2,4)}(t, s)$, we can follow the similar approach in Horváth et al. (2013, Appendix A), where we only need to shift their index notation m to $m - 1$ and $m + 1$, respectively. Then we can obtain that

$$\left\| \hat{c}_N^{(2,3)}(t, s) - [c(t, s) + \mathbb{E}[\eta_0(s)\eta_0(t)] + \mathbb{E}[\eta_0(s)\eta_1(t)]] \right\| = o_{\mathbb{P}}(1) \quad (3.15)$$

and

$$\left\| \hat{c}_N^{(2,4)}(t, s) - [c(t, s) - \mathbb{E}[\eta_0(s)\eta_0(t)] - \mathbb{E}[\eta_0(s)\eta_1(t)]] \right\| = o_{\mathbb{P}}(1). \quad (3.16)$$

By evaluations (3.13) to (3.16), we can conclude that

$$\left\| \hat{c}_N^{(2)}(t, s) \right\| = o_{\mathbb{P}}(1). \quad (3.17)$$

For the third term of $\hat{c}_N^{(3)}(t, s)$, we can rewrite it as

$$\hat{c}_N^{(3)}(t, s)$$

$$\begin{aligned}
&= \frac{1}{N} \xi(t) \left\{ [\eta_{N+1}(s) - \eta_1(s)] + \sum_{i=1}^{N-1} K \left(\frac{i}{h(N)} \right) [\eta_{N+1}(s) + \eta_{N+1-i}(s) - \eta_{i+1}(s) - \eta_1(s)] \right\} \\
&\quad + \frac{1}{N} \xi(s) \left\{ [\eta_{N+1}(t) - \eta_1(t)] + \sum_{i=1}^{N-1} K \left(\frac{i}{h(N)} \right) [\eta_{N+1}(t) + \eta_{N+1-i}(t) - \eta_{i+1}(t) - \eta_1(t)] \right\} \\
&=: \hat{c}_N^{(3,1)}(t, s) + \hat{c}_N^{(3,2)}(t, s).
\end{aligned}$$

It is easy to see that the evaluations on $\hat{c}_N^{(3,1)}(t, s)$ and $\hat{c}_N^{(3,2)}(t, s)$ are the same. For the $\hat{c}_N^{(3,1)}(t, s)$, we note that by the Cauchy–Schwarz inequality,

$$\begin{aligned}
&\mathbb{E} \|\hat{c}_N^{(3,1)}(t, s)\|^2 \\
&\leq \frac{1}{N} \|\xi\|^2 \left[\mathbb{E} \|\eta_{N+1} - \eta_1\|^2 + \sum_{i=1}^{N-1} K^2 \left(\frac{i}{h(N)} \right) \mathbb{E} \|\eta_{N+1} + \eta_{N+1-i} - \eta_{i+1} - \eta_1\|^2 \right] \\
&= \mathcal{O}(1).
\end{aligned}$$

Hence, we can conclude that

$$\|\hat{c}_N^{(3)}(t, s)\| = \mathcal{O}_{\mathbb{P}}(1). \quad (3.18)$$

Finally, by evaluations (3.12), (3.17), and (3.18), we can conclude that under the alternative hypothesis H_1^{ts} , we have

$$\|\hat{c}_N(t, s)\| = \mathcal{O}_{\mathbb{P}}(h(N)),$$

which implies $\hat{\lambda}_1 = \mathcal{O}_{\mathbb{P}}(h(N))$ by the result in Dunford and Schwartz (1988, p.1091).

3.7.3 Proof of Theorem 3.2.4

Under the alternative hypothesis H_1^{ws} , we have

$$Z_N(x, t) = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} [X_{n+1}(t) - X_n(t)] = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} [\eta_{n+1}(t) - \eta_n(t)] = \frac{1}{\sqrt{N}} [\eta_{\lfloor Nx \rfloor + 1}(t) - \eta_1(t)].$$

Hence, under H_1^{ws} ,

$$R_N = \iint Z_N^2(x, t) dx dt = \frac{1}{N} \iint [\eta_{\lfloor Nx \rfloor + 1}(t) - \eta_1(t)]^2 dx dt = \frac{1}{N^2} \sum_{n=1}^{N-1} \int [\eta_{n+1}(t) - \eta_1(t)]^2 dt.$$

Note that in this case,

$$\mathbb{E}[R_N] \leq \mathbb{E} \left[\frac{2}{N^2} \sum_{n=1}^{N-1} \int [\eta_{n+1}^2(t) + \eta_1^2(t)] dt \right] = \frac{4(N-1)}{N^2} \mathbb{E} \|\eta_0\|^2 = \mathcal{O}(1/N).$$

Therefore, we have $R_N = \mathcal{O}_{\mathbb{P}}(1/N)$ under the alternative hypothesis H_1^{ws} .

3.7.4 Proof of Theorem 3.2.5

Under the alternative hypothesis H_1^{ar1} , we have

$$\begin{aligned} R_N &= \frac{1}{N} \iint \left[\sum_{n=1}^{\lfloor Nx \rfloor + 1} \alpha^{\lfloor Nx \rfloor + 1 - n} \eta_n(t) - \eta_1(t) \right]^2 dx dt = \frac{1}{N^2} \int \sum_{i=1}^N \left[\sum_{n=1}^i \alpha^{i-n} \eta_n(t) - \eta_1(t) \right]^2 dt \\ &= \frac{1}{N^2} \int \left\{ \sum_{i=1}^N \left[\sum_{n=1}^i \alpha^{i-n} \eta_n(t) \right]^2 + N \eta_1^2(t) - 2 \eta_1(t) \sum_{i=1}^N \sum_{n=1}^i \alpha^{i-n} \eta_n(t) \right\} dt. \end{aligned}$$

Then we have

$$\begin{aligned} \mathbb{E}[R_N] &= \frac{1}{N} \int \mathbb{E}[\eta_1^2(t)] dt - \frac{2}{N^2} \int \mathbb{E} \left[\eta_1(t) \sum_{i=1}^N \sum_{n=1}^i \alpha^{i-n} \eta_n(t) \right] dt \\ &\quad + \frac{1}{N^2} \int \mathbb{E} \left\{ \sum_{i=1}^N \left[\sum_{n=1}^i \alpha^{i-n} \eta_n(t) \right]^2 \right\} dt. \end{aligned}$$

Next, we evaluate the orders of the three terms above.

First, we have

$$\frac{1}{N} \int \mathbb{E}[\eta_1^2(t)] dt = \frac{1}{N} \mathbb{E} \|\eta_0\|^2 = \mathcal{O} \left(\frac{1}{N} \right).$$

The integral in the second term can be rewritten as

$$\int \mathbb{E} \left[\eta_1(t) \sum_{i=1}^N \sum_{n=1}^i \alpha^{i-n} \eta_n(t) \right] dt = \sum_{i=1}^N \sum_{n=1}^i \alpha^{i-n} \int \mathbb{E}[\eta_n(t) \eta_1(t)] dt.$$

Note that

$$\begin{aligned} \sum_{n=1}^N \int \mathbb{E}[\eta_n(t) \eta_1(t)] dt &= \sum_{n=1}^N \int \mathbb{E}[\eta_{n-1}(t) \eta_0(t)] dt = \int \sum_{n=1}^N \mathbb{E}[\eta_{n-1}(t) \eta_0(t)] dt \\ &\rightarrow \int \frac{1}{2} [C(t, t) + \mathbb{E} \|\eta_0\|^2] dt = \frac{1}{2} \int C(t, t) dt + \frac{1}{2} \mathbb{E} \|\eta_0\|^2 < \infty, \end{aligned} \tag{3.19}$$

when $N \rightarrow \infty$. Since $\sum_{n=1}^{\infty} \int \mathbb{E}[\eta_n(t) \eta_1(t)] dt < \infty$, for any $\varepsilon > 0$, there exists an integer $N_{1,\varepsilon}$ such that when $n > N_{1,\varepsilon}$, we have $|\sum_{k=n}^{\infty} \int \mathbb{E}[\eta_k(t) \eta_1(t)] dt| < \varepsilon$. We set

$$K_\varepsilon \triangleq \sum_{i=1}^{N_{1,\varepsilon}} \sum_{n=1}^i \alpha^{i-n} \int \mathbb{E}[\eta_n(t) \eta_1(t)] dt \quad \text{and} \quad S_1 \triangleq \sum_{n=1}^{\infty} \int \mathbb{E}[\eta_n(t) \eta_1(t)] dt.$$

Then, we have

$$\sum_{i=1}^N \sum_{n=1}^i \alpha^{i-n} \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt = \alpha^{N-N_{1,\varepsilon}} K_\varepsilon + \sum_{i=1}^{N-N_{1,\varepsilon}} \alpha^{i-1} \left[\sum_{n=1}^{N-i+1} \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt \right].$$

Since $S_1 - \varepsilon < \sum_{n=1}^{N-i+1} \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt < S_1 + \varepsilon$ for $i = 1, \dots, N - N_\varepsilon$, we have

$$\sum_{i=1}^{N-N_{1,\varepsilon}} \alpha^{i-1} (S_1 - \varepsilon) < \sum_{i=1}^{N-N_{1,\varepsilon}} \alpha^{i-1} \left\{ \sum_{n=1}^{N-i+1} \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt \right\} < \sum_{i=1}^{N-N_{1,\varepsilon}} \alpha^{i-1} (S_1 + \varepsilon).$$

Subsequently, we have

$$\begin{aligned} & \alpha^{N-N_{1,\varepsilon}} K_\varepsilon + \frac{1 - \alpha^{N-N_{1,\varepsilon}}}{1 - \alpha} (S_1 - \varepsilon) \\ & < \sum_{i=1}^N \sum_{n=1}^i \alpha^{i-n} \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt < \alpha^{N-N_{1,\varepsilon}} K_\varepsilon + \frac{1 - \alpha^{N-N_{1,\varepsilon}}}{1 - \alpha} (S_1 + \varepsilon) \end{aligned}$$

and

$$\left| \sum_{i=1}^N \sum_{n=1}^i \alpha^{i-n} \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt - \frac{1}{1 - \alpha} S_1 \right| < \alpha^{N-N_{1,\varepsilon}} \left| K_\varepsilon - \frac{1}{1 - \alpha} S_1 \right| + \frac{1 - \alpha^{N-N_{1,\varepsilon}}}{1 - \alpha} \varepsilon. \quad (3.20)$$

By (3.20), we can find a number $M_{1,\varepsilon} > \log_\alpha \varepsilon - \log_\alpha \left| K_\varepsilon - \frac{1}{1-\alpha} S_1 \right| + N_{1,\varepsilon}$ such that for any $N' > M_{1,\varepsilon}$, we have

$$\left| \sum_{i=1}^{N'} \sum_{n=1}^i \alpha^{i-n} \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt - \frac{1}{1 - \alpha} S_1 \right| < \frac{2 - \alpha}{1 - \alpha} \varepsilon,$$

which implies that

$$\sum_{i=1}^N \sum_{n=1}^i \alpha^{i-n} \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt \rightarrow \frac{1}{1 - \alpha} \sum_{n=1}^\infty \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt \quad (3.21)$$

as $N \rightarrow \infty$, and conclusively,

$$\frac{2}{N^2} \int \mathbb{E} \left[\eta_1(t) \sum_{i=1}^N \sum_{n=1}^i \alpha^{i-n} \eta_n(t) \right] dt = \mathcal{O} \left(\frac{1}{N^2} \right).$$

For the third term, we first rewrite it as

$$\frac{1}{N^2} \int \mathbb{E} \left\{ \sum_{i=1}^N \left[\sum_{n=1}^i \alpha^{i-n} \eta_n(t) \right]^2 \right\} dt = \frac{2}{N^2} \sum_{i=1}^N \sum_{n=1}^i \left(\sum_{k=0}^{i-n} \alpha^{2k+n-1} \right) \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt$$

$$-\frac{1}{N^2} \sum_{i=1}^N \sum_{k=0}^{i-1} \alpha^{2k+n-1} \int \mathbb{E}[\eta_1(t)\eta_1(t)]dt$$

and note that

$$\frac{1}{N^2} \sum_{i=1}^N \sum_{k=0}^{i-1} \alpha^{2k+n-1} \int \mathbb{E}[\eta_0(t)\eta_0(t)]dt = \frac{1}{N^2} \left[\frac{N}{1-\alpha^2} - \alpha^2 \frac{1-\alpha^{2N}}{(1-\alpha^2)^2} \right] \mathbb{E}\|\eta_0\|^2 = \mathcal{O}\left(\frac{1}{N}\right).$$

We thus focus on the remaining one term. By (3.19) and the Abel Discriminance in series, we have

$$\sum_{n=1}^{\infty} \alpha^{n-1} \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt < \infty.$$

Hence, for any $\varepsilon > 0$, there exists an integer $N_{2,\varepsilon}$ such that when $n > N_{2,\varepsilon}$, we have

$$\left| \sum_{k=n}^{\infty} \alpha^{k-1} \int \mathbb{E}[\eta_k(t)\eta_1(t)]dt \right| < \varepsilon.$$

We set

$$K_{1,\varepsilon} \triangleq \sum_{i=1}^{N_{2,\varepsilon}} \left(\sum_{k=0}^{i-1} \alpha^{2k} \right) \sum_{n=1}^{N_{2,\varepsilon}-i+1} \alpha^{n-1} \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt,$$

$$K_{2,\varepsilon} \triangleq \sum_{i=1}^{N_{2,\varepsilon}} \alpha^{2i-2} \sum_{n=1}^{N_{2,\varepsilon}-i+1} \alpha^{n-1} \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt, \quad \text{and} \quad S_2 \triangleq \sum_{n=1}^{\infty} \alpha^{n-1} \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt.$$

Then we obtain

$$\begin{aligned} & \sum_{i=1}^N \sum_{n=1}^i \left(\sum_{k=0}^{i-n} \alpha^{2k+n-1} \right) \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt = \sum_{i=1}^N \left(\sum_{k=0}^{i-1} \alpha^{2k} \right) \sum_{n=1}^{N-i+1} \alpha^{n-1} \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt \\ & = K_{1,\varepsilon} + \left(\sum_{k=1}^{N-N_{2,\varepsilon}} \alpha^{2k} \right) K_{2,\varepsilon} + \sum_{i=1}^{N-N_{2,\varepsilon}} \left(\sum_{k=0}^{i-1} \alpha^{2k} \right) \sum_{n=1}^{N-i+1} \alpha^{n-1} \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt. \end{aligned}$$

Since $S_2 - \varepsilon < \sum_{n=1}^{N-i+1} \alpha^{n-1} \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt < S_2 + \varepsilon$ for $i = 1, \dots, N - N_{2,\varepsilon}$, we have

$$\begin{aligned} & \sum_{i=1}^{N-N_{2,\varepsilon}} \left(\sum_{k=0}^{i-1} \alpha^{2k} \right) (S_2 - \varepsilon) \\ & < \sum_{i=1}^{N-N_{2,\varepsilon}} \left(\sum_{k=0}^{i-1} \alpha^{2k} \right) \sum_{n=1}^{N-i+1} \alpha^{n-1} \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt < \sum_{i=1}^{N-N_{2,\varepsilon}} \left(\sum_{k=0}^{i-1} \alpha^{2k} \right) (S_2 + \varepsilon). \end{aligned}$$

Simple calculation yields that

$$\left[\frac{N - N_{2,\varepsilon}}{1 - \alpha^2} - \alpha^2 \frac{1 - \alpha^{2N-2N_{2,\varepsilon}}}{(1 - \alpha^2)^2} \right] (S_2 - \varepsilon)$$

$$< \sum_{i=1}^{N-N_{2,\varepsilon}} \left(\sum_{k=0}^{i-1} \alpha^{2k} \right) \sum_{n=1}^{N-i+1} \alpha^{n-1} \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt < \left[\frac{N-N_{2,\varepsilon}}{1-\alpha^2} - \alpha^2 \frac{1-\alpha^{2N-2N_{2,\varepsilon}}}{(1-\alpha^2)^2} \right] (S_2 + \varepsilon).$$

Hence, we obtain that

$$\begin{aligned} & \left| \frac{1}{N} \sum_{i=1}^N \left(\sum_{k=0}^{i-1} \alpha^{2k} \right) \sum_{n=1}^{N-i+1} \alpha^{n-1} \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt - \frac{1}{1-\alpha^2} S_2 \right| \quad (3.22) \\ & < \frac{1}{N-N_{2,\varepsilon}} |K_{1,\varepsilon}| + \frac{\alpha^2 (1-\alpha^{2N-2N_{2,\varepsilon}})}{(N-N_{2,\varepsilon})(1-\alpha^2)} \left| K_{2,\varepsilon} - \frac{1}{1-\alpha^2} S_2 \right| \\ & + \left[\frac{1}{1-\alpha^2} - \frac{\alpha^2 (1-\alpha^{2N-2N_{2,\varepsilon}})}{(N-N_{2,\varepsilon})(1-\alpha^2)^2} \right] \varepsilon. \end{aligned}$$

By (3.22), we can find a number $M_{2,\varepsilon} > \frac{1}{\varepsilon} \left(|K_{1,\varepsilon}| + \frac{\alpha^2}{1-\alpha^2} \left| K_{2,\varepsilon} - \frac{1}{1-\alpha^2} S_2 \right| \right) + N_{2,\varepsilon}$ such that for any $N' > M_{2,\varepsilon}$, we have

$$\left| \frac{1}{N'} \sum_{i=1}^{N'} \left(\sum_{k=0}^{i-1} \alpha^{2k} \right) \sum_{n=1}^{N'-i+1} \alpha^{n-1} \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt - \frac{1}{1-\alpha^2} S_2 \right| < \frac{2-\alpha^2}{1-\alpha^2} \varepsilon,$$

which implies that

$$\frac{1}{N} \sum_{i=1}^N \sum_{n=1}^i \left(\sum_{k=0}^{i-n} \alpha^{2k+n-1} \right) \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt \rightarrow \frac{1}{1-\alpha^2} S_2$$

as $N \rightarrow \infty$. Then it is easy to see that

$$\frac{2}{N^2} \sum_{i=1}^N \sum_{n=1}^i \left(\sum_{k=0}^{i-n} \alpha^{2k+n-1} \right) \int \mathbb{E}[\eta_n(t)\eta_1(t)]dt = \mathcal{O}\left(\frac{1}{N}\right)$$

and

$$\frac{1}{N^2} \int \mathbb{E} \left\{ \sum_{i=1}^N \left[\sum_{n=1}^i \alpha^{i-n} \eta_n(t) \right]^2 \right\} dt = \mathcal{O}\left(\frac{1}{N}\right).$$

To sum up, under the alternative hypothesis H_1^{ar1} , $\mathbb{E}[R_N] = \mathcal{O}(1/N)$, which implies $R_N = \mathcal{O}_{\mathbb{P}}(1/N)$.

Functional Machine Learning

Application on Nanoparticle Size and Size Distributions Prediction

In this chapter, we demonstrate a chemistry-related cooperative study that primarily considers machine learning application on functional data. As mentioned in the Section 1.4, existing methods to study the size of nanoparticles are time-consuming and costly. Therefore, we consider using machine learning methods to analyze the existing LSPR of nanoparticles from chemical experiments, so that the size and size distribution of the corresponding generated nanoparticles can be predicted quickly. If there is a reliable accuracy on the prediction of existing experimental data, it can save a lot of cost for the future related studies. Certainly, we need to solve two problems first.

The first problem is that the existing machine learning methods are basically based on a univariate or multivariate input. We need to make a reasonable transformation of the functional data (LSPR) in the experiment, that is, to extract an effective multivariate informative from the functional data. However, it is difficult to know how to extract the information reasonably before we analyze in depth, so we consider two methods of extraction here. The one is to directly extract the entire functional data, using Fourier expansion method to extract Fourier coefficients. The other is based on some certain

chemical theories. We only extract the information from the peaks of functional data, and some parts that may produce noises are abandoned. Details of feature engineering are in Section 4.3.

Second, because the experiment itself needs to consume much cost, the total amount of experimental data is limited (94 sets). We can not obtain high prediction accuracy by a simple stacking of data, as with some existing machine learning studies. Therefore, a suitable method is particularly important in this study. In this chapter, we consider several common used machine learning methods and make predictions separately, in order to find a method with highest accuracy. We will also analyze the prediction results of different methods individually.

In addition, in this chapter, we omit more of the original analysis of chemical experiments, retain the parts related to statistics. Because this chapter was originally based on the research of chemical experiments, the language of this chapter is relatively simple.

4.1 Introduction of Chemical Experimental Preparation

4.1.1 Preparation of Small Au NS Seeds

Small Au NS seeds were synthesised via seed-mediated method, in which Au^{3+} was reduced to Au using $NaBH_4$ as a reducing agent. Firstly, a 0.25mL of 0.01M $HAuCl_4$ solution was made up to 10mL with 0.1M cetrimonium bromide (CTAB) solution, followed by rapid injection of 0.6mL ice-cold 0.01M $NaBH_4$ solution under stirring. The resultant solution was gently stirred for 3 hours at room temperature. Subsequently, 0.12mL of the seed solution was then injected into a growth solution made of 9.75mL 0.1M CTAB, 190 mL water, 4mL 0.01M $HAuCl_4$, and 15ml 0.1M ascorbic acid. The reaction mixture was shaken gently and left overnight at room temperature. The resultant small Au NS seed sample was washed and concentrated by four times by centrifugation and redispersion in water.

4.1.2 Preparation of Large Gold Nanospheres

Larger Au NS are synthesised by growing the smaller Au NS seeds into larger Au nanopolyhedrons, which are then transformed into rounded Au NS by a mild oxidation process in presence of $HAuCl_4$ and CTAB.

The size of Au nanopolyhedrons depends on the amount of small Au NS seeds, with a strong linear correlation between the cubic power of the Au NS diameter and the reciprocal of the volume of Au NS seed solution added.

To prepare large Au nanopolyhedrons, a varying volume of the seed solution ranging from 0.025mL to 4mL was added into 30mL of 0.025M cetyltrimethylammonium chloride (CTAC) solution. For the seed solution at volumes less than 0.2mL, the seed solution was first diluted by four times with water. 0.75mL of 0.1M ascorbic acid was added to the mixture solution, followed by 1.5mL of 0.01M $HAuCl_4$. The mixture solution was placed in an air-bath shaker at 45C for 3 hours. The obtained Au nanopolyhedrons were centrifuged and redispersed in 30mL of 0.02M CTAB solution.

Finally, to obtain the large rounded Au NSs, the prepared Au nanopolyhedrons were mildly oxidised with 0.2mL of 0.01M $HAuCl_4$ solution. The resultant solution was kept in the air-bath shaker at 45C for another 2 hours. The obtained large Au NSs were centrifuged and redispersed in water.

4.1.3 SEM Characterization of the Synthesised Au NS

The synthesized Au NSs were subject to SEM characterisation. The actual sizes of each Au NS were obtained by measurements, as shown in Figure 4.1. For each sample of Au NSs, the size of 100 particles will be measured to obtain the mean and standard deviation.

At the same time, the corresponding functional data generated by the LSPR was got from the experiment. For example, for the smallest set of Au NS in the experiment, the mean of 100 particle sizes is 21.72nm and standard deviation is 1.59nm. The corresponding curve of LSPR is in Figure 4.2. We have generated 94 sets of Au

Figure 4.1 Measurements of Au NS.

NSs in the experiment, with the mean particle sizes from 21.72nm to 220.64nm. We normalize the curves of LSPR on the interval $[0,1]$ as $\{X_n(t)\}_{n=1}^{94}$ with $t \in [0,1]$, and the corresponding mean particle sizes and standard deviations are $\{Y_{s,n}, Y_{d,n}\}_{n=1}^{94}$. We will apply machine learning methods to estimate mean particle size and standard deviation from the corresponding LSPR.

4.2 Introduction of machine learning methods

4.2.1 Linear regression

Linear regression algorithm is one of the fundamental machine-learning algorithms and plays an important role in the field of artificial intelligence. It contains the properties of understandable concepts and simple procedure. Least-squares estimation is the most well-known estimation method of linear regression. As introduced in Stigler (1986), least-squares estimation was first performed by Legendre (1805) and Gauss (1809) for predicting the planetary movement, and was becoming extensively used in research by the popularization from Quetelet.

Figure 4.2 LSPR of Au NSs with mean particle size = 21.72nm and standard deviation = 1.59nm.

In linear regression, we assume output data is a linear combination of the input data. When the input data are $\{X_{i1}, X_{i2}, \dots, X_{iK}\}_{i=1}^N$ and outputs are $\{Y_i\}_{i=1}^N$, we assume they have the relation

$$Y_i = \sum_{k=1}^K a_k X_{ik}, \quad i = 1, \dots, N.$$

Our aim is to estimate the coefficients $a_1, a_2, \dots, a_K$. We choose the least square estimation method, which is to minimum the sum $\sum_{i=1}^N (Y_i - \sum_{k=1}^K a_k X_{ik})^2$. From the method, the estimate of coefficients are

$$\hat{a} = (X^T X)^{-1} X^T Y,$$

where

$$\hat{a} = (\hat{a}_1, \hat{a}_2, \dots, \hat{a}_K)^T, \quad X = \begin{bmatrix} X_{11} & \cdots & X_{1K} \\ \vdots & \ddots & \vdots \\ X_{N1} & \cdots & X_{NK} \end{bmatrix}, \quad \text{and } Y = (Y_1, Y_2, \dots, Y_N)^T.$$

After that, we estimate the testing data $\hat{Y}$ from $\hat{Y} = \sum_{k=1}^K \hat{a}_k X'_k$, where $X'_1, X'_2, \dots, X'_K$ are the corresponding variables from test data.

4.2.2 Random Forest Regression

The original basic random forest algorithm is an algorithm for classification with a list of decision trees. The algorithm was created by Ho (1995), and was named as random decision forests, Breiman (2001) then developed the algorithm and first named it random forest.

Random forest regression algorithm we use is a combination of decision tree regression and bootstrap. For the given training data set $\{X_{i1}, X_{i2}, \dots, X_{iK}\}_{i=1}^N$ and $\{Y_i\}_{i=1}^N$, we generate M bootstrap samples, and then we generate M decision trees from these bootstrap samples under a fixed classification rule. After that, we substitute the testing data in these M decision trees. For every decision tree, the estimated output is the mean of the output value on the end node, and the final estimated output of testing data is the mean of all estimated outputs.

4.2.3 XGBoost Regression

XGBoost was established by Chen and Guestrin (2016) in the Distributed Machine Learning Community, and became well-known after the victory in Higgs Machine Learning Challenge. After the build of related packages in Python, R and other development languages, XGBoost has been wildly used in recent years.

XGBoost is another algorithm based on the decision tree regression. Different from the random forest algorithm, XGBoost is a not a parallel structure of decision trees. In a simple word, it is a procedure of self-optimization. After the generation of every decision tree, the new one will be improved by a specific rule. Under this situation, the more rounds run in the procedure, the less error will be produced in the regression.

Compared with random forest algorithm, the advantages of XGBoost is that it has a good compatibility for different types of data, it can define the loss function by ourself

and always have a good accuracy rate. The disadvantage is clearly that it will consume more resources of time and space because of a non-parallel structure.

4.2.4 Neural Network

The development of neural network algorithm, which can be said to be one of the most representative machine learning methods, is accompanied by the development of computing ability. The pioneer work of neural network by McCulloch and Pitts (1943) first proposed the neuron calculation model, which can be said to be the basis of neural network theory. The perceptron was proposed in Rosenblatt (1958), and a learning algorithm close to the human learning process was also proposed that time. About 30 years later, Rumelhart et al. (1986) introduced multi-layer perceptron into backpropagation algorithm, which still occupies an important position now, making neural network a hot spot again. Since then, due to the limitation of computing efficiency, the neural network has fallen into silence. Until 2006, Hinton et al. (2006) first proposed the concept of deep belief networks (DBN), which greatly improved the performance of the model and was given the familiar name of “deep learning” for the first time.

Neural network, as the name implies, is an algorithm based on a net-like structure. Simply says, it refines the relation between input and output as a connection of many layers of relationships. We can self-define the number of layers, the number of features on every layer, the type of relationship between the layers and the loss function. After that, the algorithm will improve these relationships in the network round and round to reduce the error. Frequently, the more layers and more features on the layer imply the less error in the final relationship constructed, but also more resources consumed in time and space. In fact, neural network has consumed most of the time in training among all the methods we used in our research.

4.3 Feature Engineering

4.3.1 Introduction of Feature Engineering Methods

Because all the machine learning algorithms need the multi-variate inputs instead of functional inputs. Our first work is to extract the features from the original functional data. We will use two methods as follows.

Our first method is based on the functional data themselves. We assume all functional data lie on $[0, 1]$, then we apply the Fourier expansion on the functional data. The list of Fourier coefficients we get will be the multiple inputs in the machine learning algorithm.

For functional data $X_i(t)$, $t \in [0, 1]$ and $i = 1, \dots, N$. If we need $(2K + 1)$ -dimensional inputs in machine learning algorithms, we apply equations

$$a_{in} = 2 \int_0^1 X_n(t) \cos(2\pi nt) dt, \quad n = 0, \dots, K,$$

and

$$b_{in} = 2 \int_0^1 X_n(t) \sin(2\pi nt) dt, \quad n = 1, \dots, K.$$

The $(2K+1)$ -dimensional data $\{a_{i0}, a_{i1}, b_{i1}, \dots, a_{iK}, b_{iK}\}$ will be the multiple inputs of the machine learning algorithms extracted from original functional data $X_i(t)$.

The advantage of this method is that we can determine the dimension of inputs by ourselves, and when the dimension is big enough, almost all the information of the original functional data can be extracted completely. However, because every input has the same weight in the multiple inputs when be applied into the machine learning algorithms, some useless features will then affect the regression results and consume extra time and space resources in the algorithms. At the same time, they will always be very difficult to be elected artificially before we apply the algorithms in this method. Because of this, we will try another method to extract more useful features from functional data.

From graph 4.3, we can see that there should be a strong relationship between the particle sizes and the position of the first peaks, of the functional data. We can

Figure 4.3 Relation between the particle size (nm) and position of the first peak.

conjecture that position of peaks will be good features in the regression. At the same time, from some related chemical theories, we consider that the full width at half maximum(FWHM) of the peaks may also affect the particle size and size distribution, so we choose the position and FWHM of peaks as the inputs of machine learning algorithms.

From the shapes of functional data, it is easy to find that some of the data have two peaks and others have only one peak. For the data with one peak, we use a Gaussian function to fit the data. For the data with two peaks, we use the sum of two Gaussian functions to fit them. After that, we can use the mean(μ) and standard deviation(σ) of Gaussian function to express the position and FWHM of peaks of the related functional data. For the data with one peak, $\{\mu_1, \sigma_1\}$ will be the extracted multiple input and for the data with two peaks, it will be $\{\mu_1, \sigma_1, \mu_2, \sigma_2\}$. We will apply these features in the machine learning algorithms.

4.3.2 Input Data Preparation

From the previews section, we can see that we have two ways to extract features from the original functional data. For the method of Fourier expansion, we do not need to say too much here, just follow the dimension we set and the equations to get the coefficients. We mainly talk about the way we use to estimate the coefficients of Gaussian function. We know that the general equation of Gaussian function:

$$f(x) = a * e^{-\frac{(x-b)^2}{2c^2}}$$

is determined by three coefficients, height a , mean b and deviation c . The most important coefficients are b and c , which represent the position and deviation (which has a fixed ratio with FWHM), respectively.

From a general observation of all the curves, we find that most of the curves with small particle sizes are showing only one obvious peak, and others showing two peaks. We than separate the curves as two parts from the number of obvious peaks. For some curves, even we can see the second peak, but the too small peak will bring a relatively large error in the estimation of coefficients of Gaussian functions, so we choose to ignore them in this part.

For the curves with only one peak, we use the function “optim” with least square loss function in R and the original Gaussian function to fit the original curves. There is another question is that which part of the curve we choose. Our principle is to priorly fit the top part of the peak. For example, we consider the curve when the particle size=57.83nm. In the Figure 4.4, the hollow points are the original curve, the red, green and blue lines represents the fitted Gaussian function when we choose the 40, 80 and 120 points between the peak, respectively.

From the Figure 4.4, we can see that when we choose 40 points, there is a better fitting on the top points and when we choose 120 points, there is the better fitting on the lower points. Considering that the lower points may be affected by some other unknown factors, we choose to priorly fit the top points. However, too little points chosen will bring larger errors, so the balance needs to be struck. By some experiments, we choose

Figure 4.4 Curve of experimental data (hollow points) for particle size=57.83nm, and fitted Gaussian function when choosing 40 (red), 80 (green) and 120 (blue) points between the peak respectively.

to pick 40 points between the two sides of the top point. Under this setting, we can have a good fitting of the Gaussian function.

For the curves with 2 peaks, the fitting function will become the sum of two Gaussian functions:

$$f(x) = a_1 * e^{-\frac{(x-b_1)^2}{2c_1^2}} + a_2 * e^{-\frac{(x-b_2)^2}{2c_2^2}}.$$

The 6 coefficients should to be fitted and 4 of them (b_1, b_2, c_1, c_2) are what we need. The setting of fitting is almost the same with the curve with single peak, but the range of fitting is a little complex because of the huge difference between the curves. From observation, we can find that every curve has the inflection point of the second peak at a bit right side of 500, so we can choose 500 as the bound of right side. For the left side, it should be related to the curve itself, our principle is to prior fitting the part between top points of two peaks. Here are some examples in Figure 4.5.

Figure 4.5 Curve of experimental data (hollow points) for particle size=128.70nm, 139.71nm, 159.99nm and 184.45nm, and fitted Gaussian functions (red) respectively.

4.4 Testing Procedure and Results Analysis

4.4.1 Testing Procedure

We will use cross validation to test the estimation error of every introduced machine learning method with two kinds of extracted features. In every time's cross validation, we will separate the data as five sets. The data in one of them are chosen as test data, and others as training data. We finish one time's cross validation when all the five sets have been chosen as test data. Because there are 94 data in the data set, it is not a multiple of 5, so in every time's cross validation, we will randomly choose 4 data as the training data in all the five times estimation of one cross validation. For other 90 data, we separate them as five sets with 18 data a set. There is a special case when we choose data with two peaks as the whole data set, there are 46 data in this set. In this case, we will choose one data as the always training data, and other 45 data are separated as five sets with 9 data a set.

In every time's estimation, we will use L_2 error to show the difference between the estimated value and original value. For the estimation of particle sizes, we will use the L_2 rated error to show the difference, that is $(\frac{Y_e - Y_o}{Y_o})^2$, where Y_e is the estimated particle size and Y_o is the original size. For the estimation of deviations, because some of the original deviation are too close to 0, which makes the rated error can be abnormally large. This kind of value may strongly affect the mean error of the estimation, but can not show the difference objectively. For example, the smallest deviation in data set is 1.06nm and the biggest one is 14.63nm. We prefer the estimation errors of 1.06nm and 14.63nm are 0.6nm and 1.4nm than the errors are 0.1nm and 5nm, respectively. However, the results in front has a much larger rated error, which is unreasonable in our experiment. It is for this reason that we choose $(Y_e - Y_o)^2$ instead.

For every machine learning method with one kind of extracted feature, we will do 100 times' cross validations. The final error we show is the square root of mean error of all the cross validations.

Table 4.1

Estimated error of particle size (%) and deviation (nm) of linear regression with Fourier coefficients.

Dimension of Fourier expansion	Rated error of size(%)	Error of deviation(nm)
3	81.98	1.98
5	58.87	1.74
7	40.01	1.52
9	37.72	1.52
11	30.83	1.44
13	32.08	1.48
15	41.64	1.38
17	38.42	1.39
19	44.54	1.43
...	...	...
75	58.85	391.85

4.4.2 Linear Regression

We first talk about the results of linear regression with Fourier expansion. Because there are 76 training data and 18 test data every time, so dimension of Fourier expansion can not exceed 76. We choose all the odd numbers between 3 and 75 as the dimensions of Fourier expansion in our estimation to find which one has the minimum error. The results are represented in Table 4.1. We cut the results of some dimensions here, because the errors have an clear upward tendency when dimensions are bigger than 25. Note that we will use red color to show the minimum errors of size and deviation respectively in every table of Section 4.4.

From the Table 4.1, we can see that the minimum errors of size and deviation exist when dimensions of Fourier expansion are equal to 11 and 15, respectively. The dimensions to find the minimum errors are not very big. The reason may be that in the linear regression, all the features holds equal status. Because of this, when the

dimension becomes big, most of the features, which have little impact on the results, will also affect the regression negatively, which causes over-fitting in the prediction.

Then we consider the linear regression from peaks and FWHMs. Because some curves have one peak and others have two peaks, we first consider all the curves with their first peak. Under this condition, every data only have two features. It is clearly not reasonable to put them into linear regression directly. We consider to use B-spline linear regression here, the basic idea is very simple, we use B-spline basis to extract every feature to more than one feature, and the number of features is related to the degree of freedom of B-spline (often be the degree of freedom minus 3). In most of the time, B-spline linear regression will have a better estimation results than basic linear regression and polynomial linear regression in applications.

By the B-spline linear regression, we can explore the results under different combinations of features and different degrees of freedom of B-spline. We first see results of one to one B-spline linear regression.

From Table 4.2, we can see that when we only choose position of first peak as the feature in B-spline linear regression, the results are very good. The minimum rated error of size (3.85%) has been much better than linear regression with Fourier expansion. The estimation of deviation are both good enough from the peak and FWHM, respectively. It should be noted that, the degree of freedom of best results of deviation estimation are 5, which implies the relation from peaks and FWHMs to deviations are not so close, a high degree of freedom on the estimation will easily cause over-fitting here.

Then we see the estimation results from both peak position and FWHM. We have tried different degrees of freedom (4 to 20) on both the two features. We only show the important part of the results because of the limit of space.

From Tables 4.3 and 4.4, we can see that the results of estimation from both peak and FWHM are not bad, but still worse than the estimation from only peak. This may show that the FWHM of first peak does not have such a strong linear relationship with particle size and deviation. However, we can not receive the conclusion so early. We need to see the results of other different situations and methods first.

Table 4.2

Estimated error of particle size (%) and deviation (nm) of B-spline linear regression with peak position or FWHM individually.

Degree of freedom	Rated error of size from peak(%)	Rated error of size from FWHM(%)	Error of deviation from peak(nm)	Error of deviation from FWHM(nm)
4	5.8155	20.7288	1.085046	1.044668
5	5.033	19.1313	1.015088	1.031579
6	5.1926	20.6385	1.036172	1.058209
7	5.4081	19.7952	1.030635	1.082546
8	4.8412	19.3343	1.038474	1.088715
9	4.6827	19.7966	1.046837	1.072774
10	4.4027	20.5546	1.06532	1.114015
11	3.9655	21.0191	1.12304	1.14026
12	3.851	21.6475	1.331572	1.108857
13	3.9043	22.3019	1.729465	1.089095
14	4.4224	22.3677	2.205546	1.188371
...	...	...	...	...
20	9.6806	26.5783	6.724159	4.454638

Table 4.3

Estimated error of particle size (%) B-spline linear regression with both peak position (first column) and FWHM (first row).

Degree of freedom	4	5	6	7	8
10	4.4776	4.8699	5.0642	4.8249	4.7523
11	4.3314	4.4189	4.7765	5.4014	5.7999
12	4.2715	4.2163	4.8369	6.611	7.6173
13	4.4459	4.4161	4.9179	7.7239	8.4983
14	5.5193	5.3461	6.0622	9.6087	10.5756

Table 4.4

Estimated error of deviation (nm) B-spline linear regression with both peak position (first column) and FWHM (first row).

Degree of freedom	4	5	6	7	8
4	1.663913	1.62835	1.687359	1.742602	1.725226
5	1.264328	1.367106	1.222653	1.318704	1.354107
6	1.256248	1.259441	1.151529	1.324534	1.559567
7	1.2711	1.254946	1.306539	1.197414	1.474128
8	1.325061	1.315125	1.208156	1.18149	1.208615

Table 4.5

Estimated error of particle size (%) B-spline linear regression with both first peak position (first column) and second peak position (first row), for data with two peaks only.

Degree of freedom	4	5	6	7	8
4	4.12449	3.440235	3.511857	4.103009	5.952497
5	4.249787	3.461954	3.016202	3.540414	6.527864
6	4.017791	3.41734	2.72384	3.195863	6.82135
7	2.834473	3.230956	3.290557	4.35887	6.767945
8	9.091974	9.846198	9.657211	11.10754	11.89529

Table 4.6

Estimated error of deviation (nm) B-spline linear regression with both first peak position (first column) and second peak position (first row), for data with two peaks only.

Degree of freedom	4	5	6	7	8
4	1.441054	1.901171	1.740996	1.682099	2.441559
5	1.368532	1.428408	1.481865	1.489281	2.729603
6	1.449892	1.487307	1.801296	1.603756	3.167285
7	1.683162	1.755971	2.306755	2.302142	3.409261
8	4.017013	4.387913	4.290749	4.448702	4.651867

Table 4.7

Estimated error of particle size (%) B-spline linear regression with both first FWHM (first column) and second FWHM (first row), for data with two peaks only.

Degree of freedom	4	5	6	7	8
4	2.6857	2.8964	3.0198	3.7807	7.493
5	2.9609	2.5164	2.8386	3.3749	6.9223
6	3.9831	2.8928	2.9433	3.7523	6.6684
7	4.1362	2.9815	3.2127	5.2364	7.4575
8	4.0094	3.3108	3.6231	4.5539	7.2538

Table 4.8

Estimated error of deviation (nm) B-spline linear regression with both first FWHM (first column) and second FWHM (first row), for data with two peaks only.

Degree of freedom	4	5	6	7	8
4	2.002933	1.863187	3.650194	5.001058	6.139825
5	1.713988	2.295165	3.61604	4.597188	5.663667
6	1.699825	2.073149	4.142663	5.127308	5.756428
7	2.148693	2.576066	4.140247	5.63433	6.172683
8	2.438475	2.862251	3.660585	4.687414	5.60982

Table 4.9

Estimated error of particle size (%) and deviation (nm) of random forest with Fourier coefficients.

Dimension of Fourier(p)	Rated error of size(%)		Dimension of Fourier(p)	Error of deviation(nm)	
	mtry= $\lfloor \sqrt{p} \rfloor$	mtry= $\lfloor p/3 \rfloor$		mtry= $\lfloor \sqrt{p} \rfloor$	Mtry= $\lfloor p/3 \rfloor$
5	33.8132	29.1872	41	1.083572	1.076583
7	32.0404	32.0404	43	1.08734	1.064612
9	25.7102	25.7102	45	1.075793	1.053155
11	24.1367	24.1367	47	1.086323	1.062162
13	28.1545	29.0039	49	1.080856	1.066681
15	27.7034	28.3669	51	1.089248	1.073273

At last of B-spline linear regression, Tables 4.5 - 4.8 represent the estimation results from the data with only two peaks from different choices of features. Because the original data with two peaks are always have relatively large particle size and deviation, the different data sets make the comparison not reasonable. We will make the comparison with other methods from the data with two peaks later.

4.4.3 Random Forest

Similar with the procedure of linear regression. We first talk about the results of random forest with Fourier expansion. The important coefficients of random forest are the number of trees generated (ntree) and number of features sampled at each split (mtry). The larger 'ntree' will bring more stable estimation but more running time of program, so we fix it as 1000 in the estimation of random forest regression. When we choose Fourier coefficients as features, we need to try different dimensions of Fourier expansion, so we just choose some of the commonly used values of 'mtry' here. The important part of results are shown in Table 4.9.

From Table 4.9, we find that the minimum rated errors of size exist when dimension of Fourier expansion equal to 11. The results of two different values of 'mtry' are the

Table 4.10

Estimated error of particle size (%) and deviation (nm) of random forest with peak position and FWHM.

mtry	Rated error of size(%)	Error of deviation(nm)
1	7.561	1.1083
2	5.035	1.1236

same because 'mtry' exactly get the same value when dimension $p=11$ and we have set the same seed here. Compare with the results of linear regression, although the estimation results of particle size with Fourier expansion are not very good in these two methods, we can still find that the best results both exist when dimension $p=11$. The truth may tell us that the best dimension to estimate particle size is around 11, more dimensions have more chance to cause over-fitting.

At the same time, we find that minimum error of deviation is 1.05nm, when $p=45$ and 'mtry'= $[p/3]=15$. The result is much better than linear regression. $p=45$ has been a big dimension, which implies such a complex tree-based construction may better introduce the relation between the Fourier coefficients and deviations.

Then we talk about the estimation of random forest with first peak position and FWHM. In fact, because the limit of features here, we also do not have too much variety in the coefficients.

From Table 4.10, we find that the estimation of size is much better than from Fourier coefficients. It mainly benefit from the close relationship between peak and size. Relatively, the estimation of deviation is not as good as using Fourier expansion. We can find that the random forest with Fourier expansion gets the best estimation of deviation when dimension equal to 45, which implies a complex relationship between the curve and deviation. Because of this, a simple decision tree with two features can not explain the relationship between the curve and deviation very well is reasonable.

Then we see the estimation of random forest with 2 peak positions and FWHMs in Table 4.11. Compared with estimation from linear regressions, it has a much better

Table 4.11

Estimated error of particle size (%) and deviation (nm) of random forest with two peak positions and FWHMs, for data with two peaks only.

mtry	Rated error of size(%)	Error of deviation(nm)
1	2.341	1.259
2	2.362	1.274
3	2.399	1.267
4	2.410	1.264

result on the size and worse result on deviation. We also need to compare with other methods later in Section 4.4.6.

4.4.4 XGBoost

For the XGBoost with Fourier expansion, similar with random forest, it will be very complex to set a lot of different coefficients. However, the coefficients of XGBoost are much more complex than random forest and many of them have strong effect. For some important coefficients, ‘objective’ represents the type of learner, we choose the regression based on linear model because of the good performance in previous linear regression, which is ‘reg: linear’. ‘eval metric’ shows the evaluation metrics for validation data, from the way to show the estimation errors in our research, we choose mean average precision (‘map’) for the estimation of size and root-mean-square deviation (‘rmse’) for the estimation of deviation. ‘nround’ represents the number of trees in one time’s learning, different from the number of trees in random forest, it shows the iterations of the learner, in another word, the number of optimization. In most of the time, larger ‘nround’ gives better result in estimation, but also exist a threshold in the application. Because of this, we choose ‘nround’=25 in the estimation with Fourier expansion and ‘nround’=50 in the estimation with peak position and FWHM, after some simple experiments. ‘max depth’ represents the depth of trees, too small depth will decrease the accuracy but too large one may cause over-fitting. We will try different

Table 4.12

Estimated error of particle size (%) and deviation (nm) of XGBoost with Fourier coefficients.

Dimension of Fourier expansion	Rated error of size(%)		
	max depth=3	max depth=6	max depth=10
5	12.8893	12.9695	12.847
7	10.2354	10.0144	10.0686
9	10.402	10.2395	10.0694
11	8.9343	9.0011	9.0353
13	10.3881	10.0673	9.8719
15	10.1655	9.7277	9.5807

Dimension of Fourier expansion	Error of deviation(nm)		
	max depth=3	max depth=6	max depth=10
43	1.075372	1.10343	1.102783
45	1.07327	1.094175	1.100805
47	1.081096	1.100701	1.114468
49	1.050141	1.07418	1.072922
51	1.059129	1.081532	1.078249
53	1.068727	1.086639	1.085634

depths from 1 to 10 in the training. ‘eta’ represents the step size shrinkage used in update, similarly, larger value of ‘eta’ is easier to cause over-fitting, we will choose different values in the estimation with peak position and FWHM. For the estimation with Fourier expansion, we fixed ‘eta’=0.3 to prevent a too complex experiment, because the value of ‘eta’ has little effect on the accuracy of the estimation when ‘nround’ is big enough.

We first see the results from Fourier expansion. We consider the dimensions of Fourier expansion are the odd numbers from 3 to 75 and ‘max depth’ equals to 3, 6 and 10 respectively.

Table 4.13

Estimated error of particle size (%) and deviation (nm) of XGBoost with peak position and FWHM.

max depth	Rated error of size(%)			max depth	Error of deviation(nm)		
	eta=0.1	eta=0.3	eta=0.5		eta=0.1	eta=0.3	eta=0.5
1	15.8647	5.787	7.0559	1	1.095994	1.10075	1.150423
2	4.7288	4.9219	5.166	2	1.093583	1.195985	1.272854
3	4.2568	4.645	4.9005	3	1.154843	1.255738	1.318141
4	4.2912	4.6339	4.9273	4	1.180191	1.298309	1.342922
5	4.3062	4.6325	4.9584	5	1.210222	1.316131	1.35041
6	4.3185	4.6481	4.9558	6	1.224242	1.32921	1.356862
7	4.3193	4.65	4.9651	7	1.243565	1.341122	1.357734
8	4.3245	4.6468	4.9575	8	1.253122	1.343393	1.361938
9	4.327	4.6513	4.9677	9	1.26349	1.35016	1.360462
10	4.3279	4.6523	4.9586	10	1.267137	1.351568	1.360523

From Table 4.12, we first find that it has a relatively good results in the estimation of particle size, which is different from linear regression and random forest with Fourier expansion. At the same time, the minimum error of deviation is also smaller than previous two methods. Compare with random forest, we find that the dimension of Fourier expansion at the best results of XGBoost are close to which in the random forest, which implies under a similar construction of decision tree, the self-optimizing ability of XGBoost can truly make the estimation better than random forest. We can also find from the table that the maximum depth of the trees has little effect in our estimation, the best estimation results are received when ‘max depth’=3, and when it increases, the over-fitting exists but is not very clear.

For the estimation from first peak position and FWHM, we consider ‘max depth’ from 1 to 10 and ‘eta’ is equal to 0.1, 0.3 and 0.5 respectively. From Table 4.13, the results are better than random forest but worse than B-spline linear regression. It is

Table 4.14

Estimated error of particle size (%) and deviation (nm) of XGBoost with two peak positions and FWHMs, for data with two peaks only.

max	Rated error of size(%)			max	Error of deviation(nm)		
depth	eta=0.1	eta=0.3	eta=0.5	depth	eta=0.1	eta=0.3	eta=0.5
1	3.0467	2.9106	3.2096	1	1.250113	1.271958	1.328052
2	2.8089	2.8338	2.9907	2	1.266544	1.360114	1.42753
3	2.8215	2.8213	2.9356	3	1.298572	1.375809	1.435943
4	2.8273	2.8332	2.9683	4	1.299402	1.373418	1.423063
5	2.8282	2.8256	2.9467	5	1.301521	1.377703	1.421304
6	2.8281	2.8258	2.9598	6	1.297725	1.368457	1.413312
7	2.828	2.828	2.9627	7	1.29827	1.365655	1.415719
8	2.8282	2.8364	2.9782	8	1.294053	1.362448	1.418144
9	2.8281	2.8389	2.9786	9	1.296639	1.363365	1.421776
10	2.8281	2.836	2.9776	10	1.298715	1.361587	1.423217

easy to understand that compared with the tree construction, a simple linear model can explain the relationship between peakFWHM and sizedeviation better. We can also find that we get the minimum error of size and deviation when ‘max depth’ and ‘eta’ are small, which implies the over-fitting of this two coefficients are clear because of the limit of features in XGBoost.

For the case of 2 peak positions and FWHMs in Table 4.14, the setting of coefficients are same with the case of one peak and FWHM. The only interesting thing is that the best estimation of size in XGBoost is worse than random forest, the reason may be that the property of the sample with ‘large value’ are more suitable for random forest regression, which also need further exploration.

4.4.5 Neural Network

Because the structure of neural network is varied and the training of neural network models often takes a lot of time, it is unrealistic for us to try a large number of different neural network structures to find the optimal solution. Therefore, we will explore some typical structures to study the performance of neural network on the premise of ensuring feasibility. In practice, we use the function ‘neuralnet’ in R to realize the prediction by neural network method. On trial, we found that a too simple structure would make it impossible to achieve the partial derivatives stopping criteria of error function (‘threshold’=0.01), within maximum steps of 100,000. Therefore, we always set 3 hidden layers and the number of neurons on each layer is always the power of 2. By this setting, we can maximize the efficiency of prediction on the promise of saving training time. Moreover, because it is the neural network regression algorithm, we always set the activation function as logistic function (‘logistic’) and error function as the sum of squared errors (‘sse’).

We first consider using Fourier coefficients as features. Because of the consideration on time, we do not use all the odd numbers between 3 and 75 as the dimension of Fourier expansion in neural network, but choose some of them (5,11,21,31,41,51,61,71) as a representative. Meanwhile, as previously mentioned, too simple a structure will make training impossible to achieve threshold, and too complex a structure will consume a lot of running time. After trying, we chose the number of neurons on each layer equal to 64, 128 and 256 respectively in the acceptable range.

From Table 4.15, we can see that the prediction of particle size by neural network achieves the highest accuracy among all the methods for the feature of Fourier coefficients. 5.56% has been an acceptable rated error for particle size, which shows that it is truly difficult to explain the relationship between the Fourier coefficients from LSPR and corresponding particle size by a simple linear model or tree based classification model. A complex network model may be more suitable for the elaboration of such a relationship. On the other hand, the prediction of deviation by neural network is not as good as the previous tree based methods, which is related to the distribution of

Table 4.15

Estimated error of particle size (%) and deviation (nm) of neural network with Fourier coefficients, which contains 3 hidden layers with different neurons on each layer.

Dimension of Fourier	Rated error of size(%)			Error of deviation(nm)		
	neuron64	neuron128	neuron256	neuron64	neuron128	neuron256
5	5.769373	5.818739	5.749287	1.667385	1.610102	1.663242
11	6.002732	5.867544	5.647058	1.563551	1.544384	1.622626
21	6.333037	5.609235	5.560895	1.51572	1.564073	1.660251
31	7.226699	6.591176	6.137499	1.321406	1.32599	1.357993
41	11.06406	9.166271	11.67159	1.260149	1.356277	1.346292
51	14.34726	11.24542	14.2279	1.280761	1.364608	1.542192
61	12.27143	11.52972	15.93788	1.386721	1.309067	1.503387
71	12.32462	12.88812	17.73029	1.35776	1.370431	1.426454

deviation in the experimental data and the characteristics of neural network regression. The reason will be discussed in detail in Section 4.4.7.

We consider the situation of the features from first peak position and FWHM. Since there are only two features as input in this case, and it means that the requirement on complexity of neural network structure is higher. After trying, we can only select the number of neurons is equal to 512 in this case. Under this setting, the rated error of particle size is equal to 6.17% and error of deviation is equal to 2.24nm. We can see that the prediction accuracy of neural network is not as good as that of the previous method. One is due to the limitation of parameter selection and the other is that the neural network itself is easier to cause over-fitting than our methods, which will also be explained in detail in Section 4.4.7.

Lastly, we consider the data with only 2 peaks. Because the number of features of the input becomes 4, the choice of parameters becomes looser in this case. We choose the number of neurons is equal to 32, 64, 128 and 256 respectively.

From Table 4.16, which is similar to that of single peak, neural network from 2

Table 4.16

Estimated error of particle size (%) and deviation (nm) of neural network with two peak positions and FWHMs for data with two peaks only, which contains 3 hidden layers with different neurons on each layer.

Neurons on each hidden layer	Rated error of size(%)	Error of deviation(nm)
32	4.0875	1.912498
64	2.9517	2.005164
128	2.8289	2.18467
256	2.97	2.157647

Table 4.17

Best estimated results of different machine learning methods from Fourier coefficients.

Method	Minimum rated error of size(%)	Minimum error of deviation(nm)
Linear regression	30.83	1.38
Random forest	24.14	1.053
XGBoost	8.93	1.050
Neural network	5.56	1.26

peak positions and FWHMs does not show an advantage over other machine learning methods. We can conclude from this experiment that its efficiency of training and prediction are greatly affected by the quantity of input and structure of the data itself. Of course, we have only tried a few parameters in this experiment, and the evaluation of this method in practice needs more and more attempts to draw a rigorous conclusion.

4.4.6 Comparison of the Results between different Methods

After the individual introduction of estimation results of the four machine learning methods, we will make a comparison of the results to find which one is the best estimation method in our experiment. The minimum errors of all the four methods with different features are shown in Tables 4.17 - 4.19.

Table 4.18

Best estimated results of different machine learning methods from first peak position and FWHM.

Method	Minimum rated error of size(%)	Minimum error of deviation(nm)
B-spline linear regression	3.85	1.02
Random forest	5.04	1.11
XGBoost	4.26	1.09
Neural network	6.17	2.24

Table 4.19

Best estimated results of different machine learning methods from two peak positions and FWHMs, for data with two peaks only.

Method	Minimum rated error of size(%)	Minimum error of deviation(nm)
B-spline linear regression	4.22	1.15
Random forest	2.34	1.26
XGBoost	2.81	1.25
Neural network	2.82	1.91

We first see the results from first two tables, because they are from the same data set, the whole 94 data in our experiment. It is very clear that B-spline linear regression from one peak and FWHM have the minimum estimation errors of both the size and deviation. From the previous analysis, we know that such a result is benefited from the strong linear relationship between the peak & FWHM and size & deviation. On another hand, sometimes if we only have the information of the whole curve and it is not convenient for us to measure the peak position and FWHM, we can choose neural network to estimate the particle size and XGBoost to estimate the deviation from the Fourier coefficients of the curve, the result will also be reliable.

Now if the curve of new data have two peaks, we can just choose the experimental data with two peaks as training data to improve the accuracy rate of estimation. Under this situation, B-spline linear regression is still the best choice in the estimation of deviation, but if we want to estimate the particle size, random forest will be the best choice. The rated error 2.34% has been a very good result.

4.4.7 Analysis of the Results

After the comparison of the estimation results between different methods in previous section, we talk about how these different results generated. We will use the 3d fitting graphs of different methods from first peak position and FWHM to size and deviation to explain the different characteristics of these methods.

For every method, we will show the estimation results of size and deviation from first peak position and FWHM. The training data are all the 94 data from experiment and the test data are the dense points at the same intervals in the square area, which covers all the training data.

We use the figure of B-spline linear regression as an example. In Figure 4.6(a), the black points show the particle sizes related with the first peak position and FWHM from experiment. The red surface consists of red points, which shows the estimation from B-spline linear regression. In every method, the parameters are chosen from the best result in previous part, for example, here we only choose peak position as the feature

Figure 4.6 3d graphs between particle size, peak position and FWHM of experimental data (black) and estimated surface from B-spline linear regression (red).

and the degree of freedom of B-spline is chosen as 12 in the estimation of particle size, and we will also choose peak position as the only feature with degree of freedom equal to 5 in the estimation of deviation later. The parameters of other machine learning methods are chosen with the same principle.

In Figure 4.6(a), we can see that the surface from the estimation of B-spline linear regression is very smooth. Because we only use peak position in this estimation, we can see the figure in another direction.

In Figure 4.6(b), we can find that the relationship between particle size and peak position is very close, and the estimation of B-spline linear regression reflects such a relationship in a very good way. Because of this, B-spline linear regression has a good estimation result in our experiment.

Then we see the estimation of deviation from B-spline linear regression. Similar with 4.6, in Figure 4.7(a), the black points show the deviation related with the first peak position and FWHM from experiment. The blue surface consists of blue points, which shows the estimation from B-spline linear regression. Because we still use peak position as the feature only, we can see the graph in the direction as in Figure 4.7(b).

In Figure 4.7(b), it seems that B-spline linear regression does not fit the original

Figure 4.7 3d graphs between deviation, peak position and FWHM of experimental data (black) and estimated surface from B-spline linear regression (blue).

points very well. However, because the estimated surface is very smooth, the situation of over-fitting is hard to occur, which gives a good performance in large number of estimations.

Figures 4.8 and 4.9 show the estimation of size and deviation respectively from random forest. We choose ‘mtry’=2 in the estimation of size and ‘mtry’=1 in the estimation of deviation. We can see them together with the estimation of XGBoost in Figures 4.10 and 4.11, where we choose ‘eta’=0.1 & ‘max depth’=3 in the estimation of size and ‘eta’=0.1 & ‘max depth’=2 in the estimation of deviation. From this four figures, we can find that the estimated results of these tree based methods all have ladder-like distributions. Because we only have two features in the estimation, it makes the shape of ‘staircase’ very clear. However, it is contrary with the continuous experimental results in reality, which gives the reason to have a worse estimation result than the B-spline linear regression with a smooth estimated surface.

At last, we see the result of neural network in Figures 4.12 and 4.13. We choose 3 hidden layers with 512 neurons on each hidden layer both in the estimation of size and deviation. We can clearly find that the two surfaces of estimated results are very rugged

Figure 4.8 3d graph between particle size, peak position and FWHM of experimental data (black) and estimated surface from random forest (red).

Figure 4.9 3d graph between deviation, peak position and FWHM of experimental data (black) and estimated surface from random forest (blue).

Figure 4.10 3d graph between particle size, peak position and FWHM of experimental data (black) and estimated surface from XGBoost (red).

Figure 4.11 3d graph between deviation, peak position and FWHM of experimental data (black) and estimated surface from XGBoost (blue).

Figure 4.12 3d graph between particle size, peak position and FWHM of experimental data (black) and estimated surface from neural network (red).

Figure 4.13 3d graph between deviation, peak position and FWHM of experimental data (black) and estimated surface from neural network (blue).

and messy, which are more obvious in the estimation of deviation. Such estimated results will easily cause the over-fitting because of the limit of features and parameters in the ‘neuralnet’ function we used. Because of this, although the surfaces are fitting well with the training data, it will be hard to give a good estimation result of new test data.

Finally, we simply talk about the estimation results from Fourier expansion. Because the linear relationship between the Fourier coefficients and particle size & deviation is not distinct, linear regression will be hard to give a good estimation in this situation. By comparison, a non-linear and complex black box sized regression in neural network can perform better. However, because of the relative disorder of deviation in experimental results, it is also hard for neural network to solve the problem of over-fitting when estimating the deviation from Fourier coefficients. At the same time, the tree based methods can benefit from the large number of features. Because of this, the tree based methods will give a better result in the estimation of deviation from Fourier expansion.

As a conclusion, if we can measure the peak position and FWHM precisely from the original experimental curve, the B-spline linear regression is the best choice to estimate the particle size and deviation. However, if we sometimes can not measure the features and only have the experimental curve itself. We can also calculate the Fourier coefficients of the curve and then estimate the particle size from neural network and deviation from tree based methods, the estimated error will also be reliable.

4.5 Conclusion

In this study, we apply machine learning methods to modernize the materials characterization workflow, employing algorithms to quantitatively predict nanoparticle size and size distributions based on their optical properties. Using spherical Au nanoparticles with particle sizes ranging from 21.72 to 220.64 nm as our model system, we achieve a low error of 3.85% and 1.02 nm for the prediction of nanoparticle size and size distribution, respectively. We first establish the structure-to-property correlations for

94 distinct samples, using scanning electron microscopy to quantify the nanoparticle size and size distributions as well as extinction spectroscopy to determine the LSPRs of the nanoparticles. To explore the possibility of using LSPR for quantitative size and size distribution prediction, we extract key spectral features such as Fourier coefficients, LSPR positions, and the FWHM of the LSPRs from the extinction spectra. Next, we investigate the performance of various machine learning algorithms in using these spectral features to predict the size and size distributions of the nanoparticles, including regression-based methods, decision tree-based methods, and neural network. In contrast to the common perception that complex network architecture is required for accurate predictions, B-spline analysis generates the best outcomes for both particle size and size distribution using the position of the dipolar resonance. By using the optical properties of nanoparticles to predict size and size distribution, our work can potentially bypass the use of electron microscopy to achieve rapid sample characterization.

Conclusion and Future Research

In this chapter, we will make a brief summary of the whole thesis and discuss the relevant future research.

First, we set up two hypothesis tests of non-stationary functional time series. In Chapter 2, we applied the simple bootstrap method and moving block bootstrap method on the existing functional KPSS test. The newly established bootstrap-based functional KPSS test significantly improved the test power, especially with small sample size. In Chapter 3, we established a functional unit root test, which filled the research gap of unit root test on functional time series. We have solved some difficult problems which are not easily met in other functional time series test, and successfully apply the test to the empirical study of intraday stock price curves.

Second, we completed a cooperative study with chemistry in Chapter 4. By using different machine learning methods, we did the analysis and prediction on the relationship between nanoparticle size & size distribution and corresponding LSPR functional data in chemical experiments, and received a high prediction accuracy.

We finish the thesis with a brief discussion of future research, which is also the work in progress. On the basis of the establishment of functional unit root test, we consider to establish a functional co-integration test. That is, for two functional time series with unit root, to consider whether they exist a linear combination as a stationary functional time series. This is also one of the ultimate goals of our research on non-stationary

functional times series test, which can be very useful in many applications, such as the analysis and prediction of stock price.

Bibliography

- Aue, A., Norinho, D. D., and Hörmann, S. (2015). On the prediction of stationary functional time series. J. Amer. Statist. Assoc., 110(509):378–392.
- Aue, A., Rice, G., and Sönmez, O. (2018). Detecting and dating structural breaks in functional data without dimension reduction. J. R. Stat. Soc. Ser. B. Stat. Methodol., 80(3):509–529.
- Benko, M., Härdle, W., and Kneip, A. (2009). Common functional principal components. Ann. Statist., 37(1):1–34.
- Berkes, I., Horváth, L., and Rice, G. (2013). Weak invariance principles for sums of dependent random functions. Stochastic Process. Appl., 123(2):385–403.
- Bhargava, A. (1986). On the theory of testing for unit roots in observed time series. The Review of Economic Studies, 53(3):369–384.
- Billingsley, P. (1968). Convergence of Probability Measures. John Wiley & Sons, Inc., New York-London-Sydney.
- Bosq, D. (2000). Linear Processes in Function Spaces. Springer-Verlag, New York.
- Breiman, L. (2001). Random forests. Machine learning, 45(1):5–32.

- Bücher, A. and Kojadinovic, I. (2018). A note on conditional versus joint unconditional weak convergence in bootstrap consistency results. J. Theor. Probab., pages 1–21.
- Cavaliere, G. and Georgiev, I. (2017). Bootstrap inference under random distributional limits. Research Institute for Econometrics Discussion Paper, pages 1–17.
- Chang, Y. (2004). Bootstrap unit root tests in panels with cross-sectional dependency. J. Econometrics, 120(2):263–293.
- Chang, Y. (2012). Taking a new contour: A novel approach to panel unit root tests. J. Econometrics, 169(1):15–28.
- Chang, Y. and Park, J. Y. (2003). A sieve bootstrap for the test of a unit root. J. Time Ser. Anal., 24(4):379–400.
- Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pages 785–794.
- Chen, Y. and Pun, C. S. (2019). A bootstrap-based kpss test for functional time series. Journal of Multivariate Analysis, 174:104535.
- Constantinou, P., Kokoszka, P., and Reimherr, M. (2017). Testing separability of space-time functional processes. Biometrika, 104(2):425–437.
- Cuevas, A., Febrero, M., and Fraiman, R. (2006). On the use of the bootstrap for estimating functions with functional data. Comput. Statist. Data Anal., 51(2):1063–1074.
- Dickey, D. A., Bell, W. R., and Miller, R. B. (1986). Unit roots in time series models: Tests and implications. The American Statistician, 40(1):12–26.
- Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. J. Amer. Statist. Assoc., 74(366, part 1):427–431.

- Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica, 49(4):1057–1072.
- Dunford, N. and Schwartz, J. T. (1988). Linear operators, part 1: general theory, volume 10. John Wiley & Sons.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. Ann. Statist., 7(1):1–26.
- Efron, B. and Tibshirani, R. J. (1993). An Introduction to the Bootstrap. Chapman and Hall, New York.
- Erbas, B., Hyndman, R. J., and Gertig, D. M. (2007). Forecasting age-specific breast cancer mortality using functional data models. Stat. Med., 26(2):458–470.
- Gabrys, R., Hörmann, S., and Kokoszka, P. (2013). Monitoring the intraday volatility pattern. Journal of Time Series Econometrics, 5(2):87–116.
- Gabrys, R., Horváth, L., and Kokoszka, P. (2010). Tests for error correlation in the functional linear model. J. Amer. Statist. Assoc., 105(491):1113–1125.
- Gauss, C. F. (1809). *Theoria motus corporum coelestium*. Werke.
- Górecki, T., Hörmann, S., Horváth, L., and Kokoszka, P. (2018). Testing normality of functional time series. J. Time Series Anal., 39(4):471–487.
- Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. Neural computation, 18(7):1527–1554.
- Ho, T. K. (1995). Proceedings of the 3rd international conference on document analysis and recognition. Random Decision Forests, pages 278–282.
- Hörmann, S., Kidziński, L. u., and Hallin, M. (2015). Dynamic functional principal components. J. R. Stat. Soc. Ser. B. Stat. Methodol., 77(2):319–348.

- Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. Ann. Statist., 38(3):1845–1884.
- Hörmann, S. and Kokoszka, P. (2012). Functional time series. In Handbook of Statistics. Elsevier.
- Hörmann, S., Kokoszka, P., and Nisol, G. (2018). Testing for periodicity in functional time series. Ann. Statist., 46(6A):2960–2984.
- Horváth, L., Hušková, M., and Kokoszka, P. (2010). Testing the stability of the functional autoregressive process. J. Multivariate Anal., 101(2):352–367.
- Horváth, L. and Kokoszka, P. (2012). Inference for Functional Data with Applications. Springer, New York.
- Horváth, L., Kokoszka, P., and Reeder, R. (2013). Estimation of the mean of functional time series and a two-sample problem. J. R. Stat. Soc. Ser. B. Stat. Methodol., 75(1):103–122.
- Horváth, L., Kokoszka, P., and Rice, G. (2014). Testing stationarity of functional time series. J. Econometrics, 179(1):66–82.
- Im, K. S., Pesaran, M. H., and Shin, Y. (2003). Testing for unit roots in heterogeneous panels. Journal of econometrics, 115(1):53–74.
- Kargin, V. and Onatski, A. (2008). Curve forecasting by functional autoregression. J. Multivariate Anal., 99(10):2508–2526.
- Kokoszka, P., Miao, H., and Zhang, X. (2015). Functional dynamic factor model for intraday price curves. Journal of Financial Econometrics, 13(2):456–477.
- Kokoszka, P. and Reimherr, M. (2013). Determining the order of the functional autoregressive model. J. Time Series Anal., 34(1):116–129.
- Kokoszka, P. and Young, G. (2016). KPSS test for functional time series. Statistics, 50(5):957–973.

- Kokoszka, P. and Young, G. (2017). Testing trend stationarity of functional time series with application to yield and daily price curves. Statistics and Its Interface, 10(1):81–92.
- Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. Ann. Statist., 17(3):1217–1241.
- Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? J. Econometrics, 54(1-3):159–178.
- Lahiri, S. N. (1999). Theoretical comparisons of block bootstrap methods. Ann. Statist., 27(1):386–404.
- Lee, D. and Schmidt, P. (1996). On the power of the KPSS test of stationarity against fractionally-integrated alternatives. J. Econometrics, 73(1):285–302.
- Legendre, A. M. (1805). Nouvelles méthodes pour la détermination des orbites des comètes. F. Didot.
- Levin, A., Lin, C.-F., and Chu, C.-S. J. (2002). Unit root tests in panel data: asymptotic and finite-sample properties. Journal of econometrics, 108(1):1–24.
- Liebl, D. (2013). Modeling and forecasting electricity spot prices: a functional data perspective. Ann. Appl. Stat., 7(3):1562–1592.
- Maddala, G. S. and Wu, S. (1999). A comparative study of unit root tests with panel data and a new simple test. Oxford Bulletin of Economics and statistics, 61(S1):631–652.
- McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5(4):115–133.
- Moon, H. and Perron, B. (2012). Beyond panel unit root tests: Using multiple testing to determine the nonstationarity properties of individual series in a panel. J. Econometrics, 169(1):29–33.

- Müller, H.-G., Sen, R., and Stadtmüller, U. (2011). Functional data analysis for volatility. Journal of Econometrics, 165(2):233–245.
- Paparoditis, E. (2018). Sieve bootstrap for functional time series. Ann. Statist., 46(6B):3510–3538.
- Paparoditis, E. and Sapatinas, T. (2016). Bootstrap-based testing of equality of mean functions or equality of covariance operators for functional data. Biometrika, 103(3):727–733.
- Park, J. Y. (2002). An invariance principle for sieve bootstrap in time series. Econometric Theory, 18(2):469–490.
- Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regression. Biometrika, 75(2):335–346.
- Pilavakis, D., Paparoditis, E., and Sapatinas, T. (2017). Moving block and tapered block bootstrap for functional time series with an application to the k-sample mean problem. arXiv preprint arXiv:1711.01070.
- Politis, D. N. and Romano, J. P. (1994). Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap. Statist. Sinica, 4(2):461–476.
- Politis, D. N. and White, H. (2004). Automatic block-length selection for the dependent bootstrap. Econometric Rev., 23(1):53–70.
- Pötscher, B. M. and Prucha, I. R. (1997). Dynamic Nonlinear Econometric Models. Springer Berlin Heidelberg, 1 edition.
- Ramsay, J. O. and Silverman, B. W. (2005). Functional Data Analysis. Springer, New York.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review, 65(6):386.

- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. nature, 323(6088):533–536.
- Shang, H. L. (2017a). Forecasting intraday S&P 500 index returns: a functional time series approach. J. Forecast., 36(7):741–755.
- Shang, H. L. (2017b). Functional time series forecasting with dynamic updating: an application to intraday particulate matter concentration. Econom. Stat., 1:184–200.
- Shang, H. L. (2018). Bootstrap methods for stationary functional time series. Stat. Comput., 28(1):1–10.
- Shang, H. L. and Hyndman, R. J. (2013). fds: Functional data sets. R package version 1.7.
- Shorack, G. R. and Wellner, J. A. (1986). Empirical Processes with Applications to Statistics. John Wiley & Sons, Inc., New York.
- Stigler, S. M. (1986). The history of statistics: The measurement of uncertainty before 1900. Harvard University Press.
- Westerlund, J. and Larsson, R. (2015). New tools for understanding the local asymptotic power of panel unit root tests. J. Econometrics, 188(1):59–93.
- Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. Proceedings of the National Academy of Sciences, 102(40):14150–14154.
- Zhang, B., Pan, G., Gao, J., et al. (2018). Clt for largest eigenvalues and unit root testing for high-dimensional nonstationary time series. The Annals of Statistics, 46(5):2186–2215.
- Zhang, R. and Chan, N. H. (2018). Portmanteau-type tests for unit-root and cointegration. J. Econometrics, 207(2):307–324.
- Zvingelis, J. J. (2003). On bootstrap coverage probability with dependent data. In Computer-Aided Econometrics. CRC Press.