UNIVERSITY OF CALIFORNIA, SAN DIEGO

Kernel Methods in Nonparametric Functional Time Series

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Mathematics (with a Specialization in Statistics)

by

Tingyi Zhu

Committee in charge:

Professor Dimitris Politis, Chair
Professor Ery Arias-Castro
Professor Jelena Bradic
Professor Anthony Collins Gamst
Professor Yixiao Sun

2017

The dissertation of Tingyi Zhu is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2017

DEDICATION

To my parents.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First of all, a huge “thank you” to my advisor, Professor Dimitris Politis, for guiding and supporting me over the years. You have set an example of excellence as a researcher, mentor, instructor and role model.

Also, thanks to all my teachers at UCSD, especially Prof. Ery Atias-Castro, Jelena Bradic, Anthony Gamst and Yixiao Sun for serving as members of my dissertation committee. I benefit a lot from your valuable comments and suggestions through the process. To my PhD colleague, Nan Zou, thanks for your help in various aspects of my research. To all my friends at UCSD, thank you for your accompany. Your friendship makes my life a wonderful experience.

Special thanks to Dr. John Gillick, for giving me the opportunity to work on your research project, and your tremendous medical help that relieves me from what I have suffered.

Sincere thanks to my dear friend, Jiayi Wen, for guiding my life journey to San Diego, and also having me meet Fengyi Cai and Michael Teng, who have treated me like their son throughout my time here in San Diego. I will always miss you before the next time we meet.

My foremost thanks to my parents, who created me, raised me, loved and supported me throughout my life.

Lastly, thank you, Lord, for always being there for me.

Chapter 2 and 3 of the dissertation are joint work with Professor Dimitris Politis, published in Electronic Journal of Statistics in 2017. Tingyi Zhu and Dimitris Politis are the primary investigators and authors of this paper.

Chapter 4 is joint work with Professor Dimitris Politis, which is a manuscript in preparation for later submission. Tingyi Zhu and Dimitris Politis are the primary investigators and authors of the manuscript.

VITA

2010	B. S. in Mathematics, Fudan University, Shanghai, China
2012	M. S. in Mathematics, Texas A&M University
2017	Ph. D. in Mathematics, with a Specialization in Statistics, University of California, San Diego

PUBLICATIONS

Tingyi Zhu, Dimitris N. Politis, "Kernel estimates of nonparametric functional autoregression models and their bootstrap approximation", *Electronic Journal of Statistics*, 11(2), 2876-2906, 2017.

Tingyi Zhu, Dimitris N. Politis, "Higher-order accurate spectral density estimation of functional time series", *Manuscript in preparation*, 2017.

ABSTRACT OF THE DISSERTATION

Kernel Methods in Nonparametric Functional Time Series

by

Tingyi Zhu

Doctor of Philosophy in Mathematics (with a Specialization in Statistics)

University of California, San Diego, 2017

Professor Dimitris Politis, Chair

Functional data objects are usually collected sequentially over time exhibiting forms of dependence. Such data structure is known as functional time series. While there is plentiful literature addressing the topics of linear functional processes, relatively few contributions have dealt with nonparametric functional time series, which is the focus of this dissertation. After introducing some background and basics of functional time series in Chapter 1, I address the topics concern-

ing the applications of kernel methods in the analysis of nonparametric functional time series. Specifically, Chapter 2 investigates the kernel estimation of the autoregressive operator in the nonparametric functional autoregression model. A component-wise bootstrap procedure is proposed in Chapter 3 which can be used for estimating the distribution of the kernel estimation and constructing the prediction regions. Chapter 4 tackles the problem of spectral density estimation of functional time series. A class of higher-order accurate spectral density kernel estimators is proposed based on the notion of flat-top kernel.

Chapter 1

Preliminaries of Functional Time Series

1.1 Introduction

Popularized by the pioneering works of Ramsay and Silverman (1997) [58], (2002) [59], Functional Data Analysis (FDA) has emerged as a promising field of statistical research in the past decade. When functional data objects being collected sequentially over time that exhibit forms of dependence, such data are known as functional time series. The typical situation in which functional time series arise is when long continuous records of temporal sequence are segmented into curves over natural consecutive time intervals. One example is the daily price curves of financial transactions. Figure 1.1 shows two consecutive weeks of Mi-

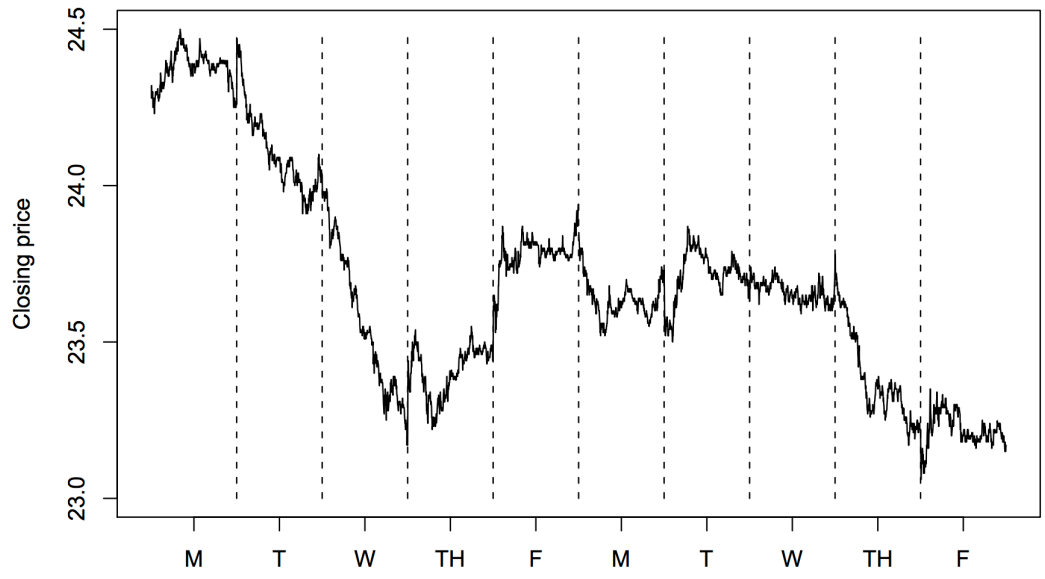

Figure 1.1: Microsoft stock prices in one-minute resolution, May 1-5, 8-12, 2006

crosoft stock prices in May 2006 in one minute resolution. Instead of using the closing daily price of the stock, one can practically think of a price curve that is defined at every trading day, and in such the data shown in Figure 1.1 will be treated as a functional time series with 10 consecutive functional observations. Other examples of data that can be characterized as functional time series include daily curves of electricity consumptions, daily patterns of geophysical, meteorological and environmental data, etc.

Before presenting the main results of our work, we provide in this chapter the general mathematical framework that is required for the statistical research of functional time series, as well as some central concepts and early contributions

in the related field. The chapter follows closely the exposition of Bosq (2000) [6], a monograph of the theory of linear functional time series, and Hörmann and Kokoszka (2012) [26], an in-depth review of the basics of functional dependent data, among the other seminal works in the analysis of functional time series.

1.2 Basics of Functional Time Series

Consider a sequence of functional observations $\{X_t, t \in \mathbb{Z}\}$ that form a functional time series where each X_t is a random function $X_t(\tau), \tau \in [a, b]$. In the context of functional time series analysis, the interval $[a, b]$ is typically normalized to be a unit interval $[0, 1]$ so that the functional observations are modeled as random elements of the separable Hilbert space $L^2([0, 1], \mathbb{R})$ of square integrable real functions defined on $[0, 1]$. The Hilbert space is equipped with a countable basis $\{e_k, k \in \mathbb{Z}^+\}$, the inner product $\langle , \rangle$ and the induced L^2 norm $|| \cdot ||_2$

$$\langle x, y\rangle = \int_0^1 x(\tau)y(\tau)d\tau, \qquad ||u||_2 = \langle y, y\rangle^{1/2}, \qquad x, y \in L^2([0, 1], \mathbb{R}).$$

Each X_t is therefore a square integrable function satisfying $||X_t||_2^2 = \int_0^1 X_t^2(\tau)d\tau < \infty$.

1.2.1 Functional Mean and Covariance Operator

Any curve X in the Hilbert space $L^2([0, 1], \mathbb{R})$ possesses a mean curve

$$\mu = (E[X(\tau)] : \tau \in [0, 1]), \tag{1.1}$$

and a covariance operator C, defines by

$$C(x) = E[\langle X - \mu, x\rangle(X - \mu)]. \tag{1.2}$$

The operator C possesses another form as a kernel operator given by

$$C(x)(\tau) = \int_0^1 c(\tau, \sigma)x(\sigma)d\sigma \tag{1.3}$$

where $c(\tau, \sigma) = \mathrm{Cov}(X(\tau), X(\sigma))$.

Suppose now we have functional observations $X_1, \ldots, X_n$, we can estimate the mean by the sample mean function

$$\hat{\mu}_n(\tau) = \frac{1}{n}\sum_{t=1}^{n} X_t(\tau), \qquad \tau \in [0, 1], \tag{1.4}$$

and the covariance operator by the sample covariance operator

$$\hat{C}_n(x) = \frac{1}{n}\sum_{t=1}^{n}\langle X_t - \hat{\mu}_n, x\rangle(X_t - \hat{\mu}_n). \tag{1.5}$$

For a large class of stationary sequences, both estimators are $\sqrt{n}$-consistent under rather general weak dependence assumptions in the sense that $E||\hat{\mu}_n - \mu||_2^2 = O(n^{-1})$ and $E||\hat{C}_n - C||_{\mathcal{L}}^2 = O(n^{-1})$, where the operator norm $||\cdot||_{\mathcal{L}}$ is, for any operator A, defined as

$$||A||_{\mathcal{L}} = \sup_{||x||_2 \leq 1} ||A(x)||_2. \tag{1.6}$$

As in the multivariate case, C can admits the spectral decomposition

$$C(x) = \sum_{l=1}^{\infty} \lambda_l \langle v_l, x\rangle v_l, \tag{1.7}$$

where $(\lambda_l : l \in \mathbb{Z}^+)$ are the eigenvalues and $(v_l : l \in \mathbb{Z}^+)$ are corresponding normalized eigenfunctions so that $C(v_l) = \lambda_l v_l$ and $||v_l||_2 = 1$. Hence any $X \in L^2([0,1], \mathbb{R})$ allows the Karhunen-Loéve representation $X = \sum_{l=1}^{\infty} \langle X, v_l \rangle v_l$. The coefficients $\langle X, v_l \rangle$ in this expansion are called the functional principal component (FPC) scores of X.

1.2.2 Autocovariance Operator

Statistical analyses of functional data typically focus on the first and second-order characteristics of the law of the functional sequences. When functional data are independent and identically distributed, the entire second-order structure is captured by the covariance operator. However, to obtain a complete description of the second-order structure of sequences of potentially dependent functional data, one needs to consider the autocovariance operators relating different lags of the series, as is the case in multivariate time series.

Before getting to the notion of autocovariance operator, we first introduce the general cross-covariance operator, which is defined for any $X, Y \in L^2([0,1], \mathbb{R})$ by

$$C_{X,Y}(x) = E[\langle X - \mu_X, x \rangle (Y - \mu_Y)], \qquad x \in L^2([0,1], \mathbb{R}) \tag{1.8}$$

where μ_X and μ_Y are the mean curves of X and Y, respectively. For a stationary functional sequence $\{X_t, t \in \mathbb{Z}\}$ with mean μ, we let

$$D(x) = E[\langle X_1 - \mu, x \rangle (X_0 - \mu)], \qquad x \in L^2([0,1], \mathbb{R}) \tag{1.9}$$

be the cross-covariance operator of X_0 and X_1, and

$$D'(x) = E[\langle X_0 - \mu, x\rangle(X_1 - \mu)], \qquad x \in L^2([0,1],\mathbb{R}) \tag{1.10}$$

be the cross-covariance operator of X_1 and X_0. $D(x)$ and $D'(x)$ can be referred to as the autocovariance operator of lag 1 and lag -1, which are two key objects in the estimation of functional autoregressive model that will be introduced in the next section. The autocovariance operator of lag t is defined by

$$\mathscr{R}_t(x) = E[\langle X_0 - \mu, x\rangle(X_t - \mu)], \qquad x \in L^2([0,1],\mathbb{R}). \tag{1.11}$$

Define the autocovariance kernel of lag t as

$$r_t(\tau,\sigma) = \mathrm{cov}(X_{t+s}(\tau), X_s(\sigma)) \tag{1.12}$$

for any $\tau, \sigma \in [0,1]$ and $s \in \mathbb{Z}$. It is easy to see that the autocovariance operator $\mathscr{R}_t : L^2([0,1],\mathbb{R}) \to L^2([0,1],\mathbb{R})$ is induced by the autocovariance kernel r_t through right integration, i.e.,

$$\mathscr{R}_t(x) = \mathrm{cov}[\langle X_0, x\rangle, X_t] = \int_0^1 r_t(\tau,\sigma)x(\sigma)d\sigma \tag{1.13}$$

for $x \in L^2([0,1],\mathbb{R})$.

1.3 Functional Autoregressive Model

The theory of autoregressive and more general linear processes in Hilbert and Banach spaces is developed in the monograph of Bosq (2000) [6]. In this

section, we present a few selected results which provide an introduction to the central ideas.

1.3.1 Existence

The most extensively investigated functional time series model is the linear functional autoregressive model of order one, i.e., FAR(1). We say a zero-mean functional sequence $\{X_t : t \in \mathbb{Z}\}$ in $L^2([0,1],\mathbb{R})$ follows a FAR(1) model if

$$X_t = \Psi(X_{t-1}) + \mathcal{E}_t, \qquad t \in \mathbb{Z} \tag{1.14}$$

where Ψ is a bounded linear operator satisfying $||\Psi||_{\mathcal{L}} < 1$, and $\{\mathcal{E}_t : t \in \mathbb{Z}\}$ is a sequence of independent and identically distributed mean-zero innovations in $x \in L^2([0,1],\mathbb{R})$ satisfying $E||\mathcal{E}_0||_2^2 < \infty$.

Concerning the existence of the sequence $\{X_t : t \in \mathbb{Z}\}$ in (1.14), recall that the scalar AR(1) equations, $X_t = \psi X_{t-1} + \varepsilon_t$, admit the unique causal solution $X_t = \sum_{j=0}^{\infty} \psi^j \varepsilon_{t-j}$ if $|\psi| < 1$. Bosq (2000) [6] proposed a condition analogous to $|\psi| < 1$ for FAR(1) model of Equation (1.14). First we have the following lemma:

Lemma 1.3.1. (Bosq, 2000) *For any bounded linear operator Ψ, the following two conditions are equivalent:*

(c_0) *There exists an integer $j_0 \geq 1$ such that $||\Psi^{j_0}||_{\mathcal{L}} < 1$;*

(c_1) *There exist $a > 0$ and $0 < b < 1$ such that for every $j \geq 0$, $||\Psi^j||_{\mathcal{L}} \leq ab^j$.*

Note that condition (c_0) is weaker than the condition $||\Psi||_{\mathcal{L}} < 1$; in the scalar

case these two conditions are clearly equivalent. Nevertheless, (c_1) is a sufficiently strong condition to ensure the convergence of the series $\sum_j \Psi^j(\mathcal{E}_{t-j})$, and the existence of a stationary causal solution to functional AR(1) equations, as stated in Theorem 1.3.1.

Theorem 1.3.1. (Bosq, 2000) *If condition (c_0) holds, then there is a unique strictly stationary causal solution to 1.14. This solution is given by*

$$X_t = \sum_{j=0}^{\infty} \Psi^j(\mathcal{E}_{t-j}). \tag{1.15}$$

The series converges almost surely in L^2.

1.3.2 The Classical Predictor

Applying $E[\langle \cdot, x\rangle X_{t-1}]$ to (1.14) leads to the functional Yule-Walker equations

$$\begin{aligned} E[\langle X_t, x\rangle X_{t-1}] &= E[\langle \Psi(X_t - 1), x\rangle X_{t-1}] + E[\langle \mathcal{E}_t, x\rangle X_{t-1}] \\ &= E[\langle \Psi(X_{t-1}), x\rangle X_{t-1}]. \end{aligned} \tag{1.16}$$

Recall that $D(x) = E[\langle X_1, x\rangle X_0]$ is the cross-covariance operator of X_0 and X_1. If Ψ' denotes the adjoint operator of Ψ, given by the requirement $\langle \Psi(x), y\rangle = \langle x, \Psi'(y)\rangle$, the operator equation (1.16) yields $D(x) = C(\Psi'(x))$. This formally gives

$$\Psi(x) = D'C^{-1}(x) \tag{1.17}$$

where $D'(x) = E[\langle X_0, x\rangle X_1]$.

The operator D' can be estimated by

$$\hat{D}'_n(x) = \frac{1}{n-1}\sum_{t=2}^{n}\langle X_{t-1}, x\rangle X_t, \tag{1.18}$$

while the operator C^{-1} can be estimated by

$$\hat{C}_n^{-1}(x) = \sum_{l=1}^{d}\hat{\lambda}_l^{-1}\langle \hat{v}_l, x\rangle \hat{v}_l, \tag{1.19}$$

for an appropriated chosen d.

Combining the above results with the approximation $X_t \approx \sum_{l=1}^{d}\langle X_t, \hat{v}_l\rangle\hat{v}_l$ gives the estimator $\hat{\Psi}_n$ of the autoregressive operator Ψ,

$$\begin{aligned}\hat{D}'_n\hat{C}_n^{-1}(x) &\approx \frac{1}{n-1}\sum_{t=2}^{n}\sum_{l=1}^{d}\sum_{l'=1}^{d}\hat{\lambda}_l^{-1}\langle x, \hat{v}_l\rangle\langle X_{t-1}, \hat{v}_l\rangle\langle X_t, \hat{v}_{l'}\rangle\hat{v}_{l'} \\ &= \hat{\Psi}_n(x). \end{aligned}\tag{1.20}$$

This is the estimator of Bosq (2000) [6]. The foregoing gives rise to the one-step ahead functional predictor

$$\hat{X}_{n+1} = \hat{\Psi}_n(X_n). \tag{1.21}$$

Theorem 8.7 of Bosq (2000) [6] provides the strong consistency of $\hat{\Psi}_n$ under certain technical assumptions. It has since been regarded as the classical benchmark of FAR(1) prediction.

1.4 Weak Dependence

What distinguishes time series analysis from other fields of statistics is the attention to temporal dependence of the data. This is also the case as to the analysis of functional time series. In this section, we summarize several frameworks of weak dependence structures that accommodate the temporal dependence of functional time series.

1.4.1 Strong Mixing

The classical approach to weak dependence, developed in the seminal papers of Rosenblatt (1956) [63] and Ibragimov (1962) [31], uses the strong mixing property and its variants like β, ϕ, ρ and ψ mixing.

Suppose $\{X_n\}$ is a sequence of random elements taking values in a measurable space S. Denote by $\mathcal{F}_k^- = \sigma\{\ldots, X_{t-2}, X_{t-1}, X_t\}$ and $\mathcal{F}_k^+ = \sigma\{X_t, X_{t+1}, X_{t+2}, \ldots\}$ the σ-algebras generated by the observations up to time t and after time t, respectively. The general idea of strong mixing property is to measure the maximal dependence between two events lying in the "past" $\mathcal{F}_t^-$ and the future $\mathcal{F}_{t+m}^+$. The fading memory is described by this maximal dependence decaying to zero for m growing to ∞. For example, the well known α-mixing coefficient is given by

$$\alpha(m) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_t^-, B \in \mathcal{F}_{t+m}^+, t \in \mathbb{Z}\}. \qquad (1.22)$$

Definition 1.4.1. *A sequence is said to be α-mixing (or strong mixing), if*

$$\lim_{m\to\infty} \alpha(m) = 0$$

There are two important subclasses of mixing sequences:

Definition 1.4.2. *A sequence is said to be arithmetically α-mixing with a rate $a > 0$ if*

$$\exists\, C > 0 \text{ such that } \alpha(m) \leq Cm^{-a}.$$

It is called geometrically α-mixing if

$$\exists\, C > 0 \text{ and } s \in (0,1), \text{ such that } \alpha(m) \leq Cs^{m}.$$

While mixing assumptions have been widely used in nonparametric statistics involving finite dimensional random variables, it is worth being noted that the general definitions presented above are also applicable for the infinite dimensional variables in functional space. Some general results concerning the strong mixing property of functional sequence can be found in Ferraty and Vieu (2006) [18] (see Proposition 10.3 and 10.4 therein). These results are useful for us to treat mixing sequences in functional space in our kernel nonparametric framework in Chapter 2.

1.4.2 Approximable Functional Sequences

Verifying the general mixing condition of the above type is not easy. Concerning the analysis of functional time series, Hörmann and Kokoszka (2010) [25]

introduced a moment-based notion of dependence which involves m-dependence. The so-called L^p-m-approximable condition for a functional sequence can be immediately verified.

A functional sequence $\{X_n\}$ is said to be m-dependent if for any t, the σ-algebras $\mathcal{F}_t^-$ and $\mathcal{F}_{t+m}^+$ are independent. The notion of m-dependence is the most direct relaxation of independence. However, most time series models are not m-dependent. In practice, we usually see various measures of dependence decay sufficiently fast, as the distance m between the σ-algebras $\mathcal{F}_t^-$ and $\mathcal{F}_{t+m}^+$ increases. This leads to the idea to use m-dependence as a tool to study the properties of sequences with dependence. The general idea is to approximate $\{X_n, n \in \mathbb{Z}\}$ by m-dependent processes $\{X_n^{(m)}, n \in \mathbb{Z}\}, m \geq 1$, and establish that for every n the sequence $\{X_n^{(m)}, m \geq 1\}$ converges in some sense to X_n as $m \to \infty$. The following definition formalizes the idea:

Definition 1.4.3. (Hörmann and Kokoszka, 2010) *A sequence $\{X_n\}$ is called L^p-m-approximable if each X_n admits the representation*

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

where the ε_i are i.i.d elements taking values in a measurable space S, and f is a measurable function $f : S^\infty \to H$. Moreover, if assume that $\{\varepsilon_i'\}$ is an independent copy of $\{\varepsilon_i\}$ defined on the same probability space, then letting

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

we have

$$\sum_{m=1}^{\infty} \nu_p(X_m - X_m^{(m)}) < \infty$$

where $\nu_p = (E||X||^p)^{1/p}$.

Hörmann and Kokoszka (2010) [25] has established the consistency results for the estimation of functional mean, functional principle components and long-run covariance when the condition of m-approximable is applied to the functional sequences; See the details therein.

1.4.3 Cumulant Mixing Condition

Another type of moment-based measures of dependence in time series analysis is the cumulant condition. It is usually referred to as cumulant mixing conditions as it is also related to the mixing conditions. In this section, we present a functional generalization of the cumulant multivariate conditions of Brillinger (2001) [9].

To begin with, we need the notion of cumulant kernel of a functional time series.

Definition 1.4.4. (Panaretos and Tavakoli, 2013) *For a functional time series* $\{X_t(\tau); \tau \in [0,1]\}_{t\in\mathbb{Z}}$ *in the separable Hilbert space* $L^2([0,1],\mathbb{R})$*, the pointwise definition of a kth order cumulant kernel is*

$$cum(X_{t_1}(\tau_1), \ldots, X_{t_k}(\tau_k)) = \sum_{\nu=(\nu_1,\ldots,\nu_p)} (-1)^{p-1}(p-1)! \prod_{l=1}^{p} \mathbb{E}\left[\prod_{j\in\nu_l} X_{t_j}(\tau_j)\right],$$

where the sum extends over all unordered partitions of $\{1, \ldots, k\}$.

With the definition of cumulant kernel in place, we present below the functional cumulant-type mixing conditions. Given a stationary functional sequence $\{X_t\}_{t\in\mathbb{Z}}$, for fixed $l \geq 0$ and $k = 2, 3, \ldots,$

Condition C(l, k). For each $j = 1, \ldots, k-1$,

$$\sum_{t_1,\ldots,t_{k-1}=-\infty}^{\infty} (1 + |t_j|^l)||\text{cum}(X_{t_1}, \ldots, X_{t_{k-1}}, X_0)||_2 < \infty.$$

The notion of cumulant kernels was employed to quantify the weak dependence among functional observations as the foundation of Fourier analysis of functional time series of Panaretos and Tavakoli (2013) [44]. We will make use of the above functional cumulant mixing conditions in our study of higher-order accurate spectral density estimation in Chapter 4.

Chapter 2

Kernel Estimates of Nonparametric Functional Autoregression

2.1 Introduction

The primary goal of functional time series analysis is to provide reliable guesses for the future realizations. In this chapter, we focus our attention on a first-order nonparametric functional autoregression–FAR(1) model which is defined by the recursion:

$$\mathcal{X}_{n+1} = \Psi(\mathcal{X}_n) + \mathcal{E}_{n+1}, \tag{2.1}$$

where the observations $\mathcal{X}_n$ and the error terms $\mathcal{E}_n$ are functions, and distinguished

with the linear FAR(1) model of (1.14), no linearity restrictions are imposed on the functional operator Ψ. Precise definitions and details of the model are stated in Section 2.2. Existing contributions have mostly focused on the functional linear autoregression model while the research addressing the nonlinear model is scarce. We approach this problem by merging the ideas in nonparametric time series and functional regression analysis, extending the theoretical study to the nonparametric model of functional autoregression.

The research pertains to the FAR model can trace back to Bosq (2000) [6], in which the theory of linear processes in functional space was first developed. One of the major contribution of that book was the study of linear autoregressive processes in the Hilbert space. As we have mentioned in Chapter 1.3, under the assumption that the functional operator Ψ in (1.14) is linear, Bosq has derived a one-step ahead predictor $\hat{\Psi}$ based on a functional form of the Yule-Walker equation, which has been regarded as the classical benchmark in FAR(1) prediction. Since then, there has been plentiful literatures on the study of the linear functional processes. We refer the readers Antoniadis and Sapatinas (2003) [3], Antoniadis et al. (2006) [2], Bosq (2007) [8], Kargin and Onatski (2008) [34], Gabrys et al. (2010) [21] and Horváth and Kokoszka (2011) [27], among other contributions. Bosq's predictor in 1.20 has a rather complicated form which makes it unrealistic to implement in practice. Aue et al. (2015) [4] proposed an alternative method of predicting linear FAR(1) process utilizing functional principal component analysis (FPCA). The method

appears to be much more widely applicable under the idea that the dimension reduction with FPCA should lead to a vector-valued time series of FPC scores that can be predicted by existing multivariate methodologies. Hörmann et al. (2015) [24] proposed a dynamic version of functional principal component analysis to address the problem of dimension reduction for functional time series. More recently, Klepsch et al. (2017) [36] used a similar dimension reduction technique to model the FARMA process with an application to traffic data. See also Klepsch and Klüppelberg (2017) [35] in which an innovation algorithm was proposed to obtain the best linear predictor of a functional moving average (FMA) process.

On the other hand, kernel methods have been a powerful tool when dealing with nonparametric models. Numerous early references have investigated its implementations in nonparametric univariate autoregression. To mention some, asymptotic study of the kernel smoother was presented in Robinson (1983) [62], and Masry (1996) [38]. The bootstrap procedures for this model and its validity were provided in Franke et al. (2002) [19]. Pan and Politis (2016) [42] developed a coherent methodology for the construction of bootstrap prediction intervals, which can be successfully applied to the nonlinear univariate autoregression models. Those results can be naturally extended to multivariate time series, but that is not the case for functional time series due to the infinite dimensional nature of functional data.

Nonparametric statistical methods for functional data analysis were estab-

lished in Ferraty and Vieu (2006) [18]. The nonparametric functional regression model, i.e. $\mathcal{Y} = r(\mathcal{X}) + \varepsilon$, has been extensively studied since then. Ferraty et al. (2007) [17] concentrated on the situation where the response variable $\mathcal{Y}$ is scalar and $\mathcal{X}$ takes values in some functional space. Asymptotic properties concerning the kernel estimator $\hat{r}(\cdot)$ of the regressor $r(\cdot)$ have been investigated and the validity of its bootstrap approximation was proved in Ferraty et al. (2010) [15]. The results have been extended to the model with double functional setting (i.e. both $\mathcal{Y}$ and $\mathcal{X}$ are functionals); see Ferraty et al. (2012) [16]. Masry (2005) [39] and Delsol (2009) [13] investigated the same model taking into account dependent functional data.

Motivated by the prior works aforementioned, we investigates the kernel estimator for nonparametric functional autoregression. We show the consistency of the estimator under the assumption of a strong mixing condition on the sample. The proof of its consistency involves a functional central limit theorem for dependent sequence in a triangular array setting.

The rest of this chapter is organized as follows. In Section 2.2, the detailed mathematical background of the model is provided and the functional version kernel estimator of the autoregressive operator is defined. Some notations and necessary assumptions are stated in Section 2.3. Section 2.4 provides the asymptotic results of the proposed estimator. A simulated study is given in Section 2.5 and all proofs are gathered in Section 2.6.

2.2 The FAR(1) model

Let $\{\mathcal{X}_n\}$ be a stationary and α-mixing functional sequence in some separable Hilbert space $\mathbb{H}$ with the usual definition of α-mixing coefficients introduced by Rosenblatt (1956) [63]. $\mathbb{H}$ is endowed with inner product $\langle\cdot,\cdot\rangle$ and corresponding norm $||\cdot||$ (i.e. $||g||^2 = \langle g, g\rangle$), and with orthonormal basis $\{e_k : k = 1, \cdots, \infty\}$. A semi-metric $d(\cdot,\cdot)$ is also defined on $\mathbb{H}$ to measure the proximity between two elements in $\mathbb{H}$. The semi-metric structure d will be the key tool for controlling the good behavior of the estimators whereas some separable Hilbert structure is necessary for studying the operator Ψ component by component. See more details on the two-topology framework in Ferraty et al. (2012) [16].

We consider the following FAR(1) model

$$\mathcal{X}_{i+1} = \Psi(\mathcal{X}_i) + \mathcal{E}_{i+1}, \quad i = 1, 2, \ldots, \tag{2.2}$$

where Ψ is the autoregressive operator mapping functions from $\mathbb{H}$ to $\mathbb{H}$, and the innovations $\mathcal{E}_i$'s are independent and identically distributed (i.i.d.) $\mathbb{H}$-valued random variables satisfy $E(\mathcal{E}_{i+1}|\mathcal{X}_i) = 0$ and $E(||\mathcal{E}_{i+1}||^2|\mathcal{X}_i) = \sigma^2_{\mathcal{E}}(\mathcal{X}_i) < \infty$. Assume here that the model is homoscedastic, that is, $\sigma_{\mathcal{E}}(\mathcal{X}_i) \equiv \sigma_{\mathcal{E}}$. The operator Ψ is not constrained to be linear; this is a Nonparametric Functional Autoregression model.

Remark 2.2.1. Because of the generality of the notation, a higher-order autoregression, say FAR(2), in which $\mathcal{X}_{i+1}$ depends on $\mathcal{X}_i$ and $\mathcal{X}_{i-1}$, can still be written

as FAR(1) by redefining the $\mathcal{X}$ and Ψ; e.g. in the FAR(2) case, one may let $\mathcal{Y}_i = (\mathcal{X}_i, \mathcal{X}_{i-1})$ with an obvious choice for the FAR(1) operator relating $\mathcal{Y}_{i+1}$ to $\mathcal{Y}_i$ only.

Estimation of Ψ is given by the functional version of Nadaraya-Watson estimator of time series

$$\hat{\Psi}_h(\chi) = \frac{\sum_{i=1}^{n-1} \mathcal{X}_{i+1} K(h^{-1} d(\mathcal{X}_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1} d(\mathcal{X}_i, \chi))}, \tag{2.3}$$

where χ is a fixed element in $\mathbb{H}$, K is a kernel function and h is a bandwidth sequence tending to zero as n tends to infinity. The way of choosing a semi-metric $d(\cdot, \cdot)$ was discussed in Ferraty and Vieu (2006) [18]. For a fixed $k \in \mathbb{Z}^+$, applying $\langle \cdot, e_k \rangle$ on both sides of the Eq. (2.2) yields

$$\begin{aligned} \langle \mathcal{X}_{i+1}, e_k \rangle &= \langle \Psi(\mathcal{X}_i) + \mathcal{E}_{i+1}, e_k \rangle \\ &= \langle \Psi(\mathcal{X}_i), e_k \rangle + \langle \mathcal{E}_{i+1}, e_k \rangle \quad i = 1, 2, \ldots \end{aligned}$$

Let $X_{n,k}$, $\varepsilon_{n,k}$ be the jth component of the functional $\mathcal{X}_n$ and $\mathcal{E}_n$ respectively, i.e. $X_{n,k} = \langle \mathcal{X}_n, e_k \rangle$, $\varepsilon_{n,k} = \langle \mathcal{E}_n, e_k \rangle$. Also, define the functional ψ_k from $\mathbb{H}$ to $\mathbb{R}$ such that

$$\psi_k(\cdot) = \langle \Psi(\cdot), e_k \rangle. \tag{2.4}$$

When k is fixed, we will drop the index k for the simplicity of notations, using $\{X_n\}$ and $\{\varepsilon_n\}$ to denote the sequences $\{X_{n,k}\}$ and $\{\varepsilon_{n,k}\}$, respectively. Similarly,

ψ can be used in place of ψ_k. For a fixed k, we obtain

$$X_{i+1} = \psi(\mathcal{X}_i) + \varepsilon_{i+1} \quad i = 1, 2, \ldots \tag{2.5}$$

Eq. (2.5) can be treated as an auxiliary functional autoregressive model with scalar response. The scalar innovations ε_i's are i.i.d. random variables satisfy $E(\varepsilon_{i+1}|\mathcal{X}_i) = 0$ and $E(\varepsilon_{i+1}^2|\mathcal{X}_i) = \sigma_\varepsilon^2 < \infty$. Again, the operator ψ in (2.5) is not constrained to be linear. Accordingly, its kernel estimation is given by

$$\hat{\psi}_h(\chi) = \frac{\sum_{i=1}^{n-1} X_{i+1} K(h^{-1} d(\mathcal{X}_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1} d(\mathcal{X}_i, \chi))}, \tag{2.6}$$

and the two estimators, $\hat{\Psi}_h$ and $\hat{\psi}_h$, are connected in the following way

$$\hat{\psi}_h(\chi) = \langle \hat{\Psi}_h(\chi), e_k \rangle. \tag{2.7}$$

Consistency of both $\hat{\psi}_h(\chi)$ and $\hat{\Psi}_h(\chi)$ will be addressed in Section 2.4. While it is more of the interest to study the estimator $\hat{\Psi}_h(\chi)$, the need for model (2.5) and the estimator $\hat{\psi}_h(\chi)$ will be seen in Section 3.2 where a componentwise bootstrap approximation is proposed.

2.3 Assumptions and notations

In the sequel, χ is a fixed element and $\mathcal{X}$ is a random element of the functional space $\mathbb{H}$. For $k = 1, 2, \ldots$, let $\varphi_{\chi,k}$ be a real-valued function defined as

$$\varphi_{\chi,k}(s) = E[\psi(\mathcal{X}) - \psi(\chi)|d(\mathcal{X}, \chi) = s]$$

$$= E[\langle \Psi(\mathcal{X}) - \Psi(\chi), e_k \rangle | d(\mathcal{X}, \chi) = s].$$

Again, for the simplicity of notations, we drop the index k to use φ_χ in place of $\varphi_{\chi,k}$ when k is fixed. Let F be the distribution of the random variable $d(\mathcal{X}, \chi)$,

$$F_\chi(t) = P(d(\mathcal{X}, \chi) \leq t),$$

which is usually called the small ball probability function in functional data analysis. Also define for $s \in [0, 1]$,

$$\tau_{h\chi}(s) = \frac{F_\chi(hs)}{F_\chi(h)} = P(d(\mathcal{X}, \chi) \leq hs | d(\mathcal{X}, \chi) \leq h).$$

Technical aspects of the functions φ_χ, F_χ and $\tau_{h\chi}$ have been discussed in Ferraty et al.(2007) [17]. The following assumptions are needed:

(A1) ψ is continuous in a neighborhood of χ with respect to the semi-metric d, and $F_\chi(0) = 0$.

(A2) $\varphi_\chi(0) = 0$ and $\varphi'_\chi(0)$ exists.

(A3) $h \to 0$ and $nF_\chi(h) \to \infty$, as $n \to \infty$.

(A4) The kernel function K is supported on $[0, 1]$ and has a continuous derivative with $K'(s) \leq 0$, and $K(1) > 0$.

(A5) For $s \in [0,1]$, $\lim_{h \downarrow 0} \tau_{h\chi}(s) = \tau_{0\chi}(s)$.

(A1)-(A5) are the standard assumptions inherited from those introduced in the independent case in the setting of nonparametric functional regression. For our autoregressive model, additional assumptions (A6)-(A10) below are necessary:

(A6) $\exists\, p > 2, E(|\varepsilon_i|^p|\mathcal{X}) < \infty$.

(A7) $\max(E(|X_{i+1}X_{j+1}||\mathcal{X}_i, \mathcal{X}_j), E(|X_{i+1}||\mathcal{X}_i, \mathcal{X}_j)) < \infty \quad \forall i, j \in \mathbb{Z}$.

(A8) $\{\mathcal{X}_n\}$ is α-mixing process with mixing coefficients $\alpha(n) \leq Cn^{-a}$.

(A9) $\exists\, \nu > 0,$ such that $\Theta(h) = O(F_\chi(h)^{1+\nu}),$ with $a > \dfrac{(1+\nu)p - 2}{\nu(p-2)}$,

where p and a are defined in (A6) and (A8) respectively, and

$$\Theta(s) := \max\{\max_{i \neq j} P(d(\mathcal{X}_i, \chi) \leq s, d(\mathcal{X}_j, \chi) \leq s), F_\chi^2(s)\}.$$

(A10) $\exists\, \gamma > 0$ such that $nF_\chi(h)^{1+\gamma} \to \infty$ and $a > \max\left\{\dfrac{4}{\gamma}, \dfrac{p}{p-2} + \dfrac{2(p-1)}{\gamma(p-2)}\right\}$

where p and a are defined in (A6) and (A8), respectively.

Remark 2.3.1. Delsol (2009) [13] considered a functional regression model with functional variables under dependence. (A6)-(A10) are inherited from the additional assumptions Delsol made to control the dependence between variables.

These assumptions enable us to obtain asymptotic results of the estimator in our autoregressive model, which has a formal resemblance to the analogous regression model. In particular, (A8)-(A10) is a set of conditions on the mixing coefficients, and (A8) is the so-called arithmetically α-mixing condition which is typically satisfied in the case of an autoregressive model. From a theoretical point of view, a more general set of conditions on the mixing coefficients ((H1)-(H2) in Delsol (2009) [13]) is available, see the details therein.

The semi-metric d will act on the asymptotic behavior of the estimator through φ_χ, F_χ and $\tau_{h\chi}$, and the following quantities:

$$M_0 = K(1) - \int_0^1 (sK(s))'\tau_{0\chi}(s)ds,$$
$$M_1 = K(1) - \int_0^1 K'(s)\tau_{0\chi}(s)ds,$$
$$M_2 = K^2(1) - \int_0^1 (K^2)'(s)\tau_{0\chi}(s)ds.$$

2.4 Asymptotic study

2.4.1 Consistency of estimator $\hat{\psi}_h(\chi)$

First, we have the following point-wise asymptotic results for the estimator $\hat{\psi}_h(\chi)$:

Theorem 2.4.1. *Assume (A1)-(A6), then*

$$E[\hat{\psi}_h(\chi)] - \psi(\chi) = \varphi'_\chi(0)\frac{M_0}{M_1}h + O(\frac{1}{nF_\chi(h)}) + o(h), \tag{2.8}$$

Theorem 2.4.2. *Assume (A1)-(A10), then*

$$Var(\hat{\psi}_h(\chi)) = \frac{\sigma_\varepsilon^2}{M_1^2}\frac{M_2}{nF_\chi(h)} + o(\frac{1}{nF_\chi(h)}). \tag{2.9}$$

Theorem 2.4.1 and 2.4.2 can be obtained following Lemma 2.6.1 and 2.6.2, and along the similar lines of the proof of Theorem 1 in Ferraty et al. (2007) [17]. See details of the proof in Section 2.6. Combining 2.4.1 and 2.4.2, we immediately obtain the following corollary:

Corollary 2.4.3. *Assume (A1)-(A10), then*

$$\hat{\psi}_h(\chi) \xrightarrow{p} \psi(\chi). \tag{2.10}$$

The pointwise asymptotic normality is given in Theorem 2.4.4 below:

Theorem 2.4.4. *Assume (A1)-(A10), then*

$$\sqrt{n\hat{F}_\chi(h)}\left(\hat{\psi}_h(\chi) - \psi(\chi) - B_n\right)\frac{M_1}{\sqrt{\sigma_\varepsilon^2 M_2}} \xrightarrow{d} N(0,1), \tag{2.11}$$

where $B_n = h\varphi_\chi{}'(0)M_0/M_1$, *and* $\hat{F}_\chi(h)$ *is the empirical estimation of* $F_\chi(h)$:

$$\hat{F}_\chi(h) = \frac{\#(i : d(\mathcal{X}_i, \chi) \leq h)}{n}.$$

Proof. See Section 2.6. □

The bias term can be ignored under the following additional assumption:

(A11) $\lim_{n\to\infty} h\sqrt{nF_\chi(h)} = 0$.

Corollary 2.4.5. *Assume (A1)-(A11), then*

$$\sqrt{nF_\chi(h)}\left(\hat{\psi}_h(\chi) - \psi(\chi)\right)\frac{M_1}{\sqrt{\sigma_\varepsilon^2 M_2}} \xrightarrow{d} N(0,1). \tag{2.12}$$

2.4.2 Consistency of estimator $\hat{\Psi}_h(\chi)$

Corollary 2.4.3, together with Eq. (2.7), implies the following component-wise consistency of $\hat{\Psi}_h$:

$$\langle \hat{\Psi}_h(\chi) - \Psi(\chi), e_k \rangle \xrightarrow{p} 0, \quad k = 1, \ldots, \infty \tag{2.13}$$

However, (2.13) does not guarantee the consistency of the estimator $\hat{\Psi}_h$ in an infinite-dimensional space. A more general consistency result is desired. We consider the following regularity conditions:

(C1) For each $k \geq 1$, ψ_k is continuous in a neighborhood of χ with respect to semi-metric d, and $F_\chi(0) = 0$.

(C2) For some $\beta > 0$, all $0 \leq s \leq \beta$ and all $k \geq 1$, $\varphi_{\chi,k}(0) = 0$, $\varphi'_{\chi,k}(s)$ exists, and $\varphi'_{\chi,k}(s)$ is Hölder continuous of order $0 < \alpha \leq 1$ at 0, i.e. there exists a $0 < L_k < \infty$ such that $|\varphi'_{\chi,k}(s) - \varphi'_{\chi,k}(0)| \leq L_k s^\alpha$ for all $0 \leq s \leq \beta$. Moreover, $\sum_{k=1}^\infty L_k^2 < \infty$ and $\sum_{k=0}^\infty \varphi'_{\chi,k}(0) < \infty$.

(C3) The bandwidth h satisfies $h \to 0$, $nF_\chi(h) \to \infty$, and $(nF_\chi(h))^{1/2} h^{1+\alpha} = o(1)$.

With additional assumptions on the mixing coefficients and moments, we are able to prove the following limit result for $\hat{\Psi}_h(\chi)$.

Theorem 2.4.6. *For some fixed $\chi \in \mathbb{H}$, assume $\exists\, \delta' > \delta > 0$ such that*

(i) $\frac{2+\delta}{2+\delta'} + \frac{(1-\delta)(2+\delta)}{2} \leq 1$,

(ii) $E||\mathcal{X}_i - \Psi(\chi)||^{2+\delta'} < \infty$,

(iii) $\sum_n \alpha(n)^{\frac{\delta}{2+\delta}} < \infty$,

where $\alpha(\cdot)$ *is the mixing coefficient of the functional sequence* $\{\mathcal{X}_t, t \in \mathbb{N}\}$. *Also assume (A4), (C1)-(C3). Then*

$$\hat{\Psi}_h(\chi) = \Psi(\chi) - \mathcal{B}_n + O_p(\frac{1}{\sqrt{nF_\chi(h)^{1+\delta}}}) \tag{2.14}$$

where

$$\mathcal{B}_n = h\frac{M_0}{M_1}\sum_{k=1}^{\infty} \varphi'_{\chi,k}(0)e_k.$$

Remark 2.4.7. To prove Theorem 2.4.6, we need a functional central limit theorem for dependent sequence in a triangular array setting such as Theorem 2.3 in Politis and Romano (1994) [52]. See Section 2.6 for details of the proof. The assumptions (i)-(iii) show a trade-off between the moment assumptions and the mixing conditions. The conditions on mixing coefficients can be less stringent if higher moments are assumed. The parameter δ' controls the moment while δ controls the mixing condition and they can be chosen under the constraint (i), for example, $\delta = 0.5$ and $\delta' = 5$.

2.5 Simulations

In this section, the theoretical results of the previous sections are illustrated through a simulation study. First we provide the details of the process of simulating

a FAR(1) series. We use a linear functional series here since its stationarity can be guaranteed by existing theories. The performance of the kernel estimation will be shown through the experiments performed on the simulated data.

2.5.1 Data Generating Process

The simulated realization of a linear FAR(1) series has been discussed in Didericksen (2012) [14]. Curves in the series are assumed to be elements of the Hilbert space $L^2[0,1]$. The linear operator $\Psi(\mathcal{X}) = \int_0^1 \psi(s,t)\mathcal{X}(s)ds$ is acted on the functions $\mathcal{X}_i$'s, thereby the series are generated according to the model

$$\mathcal{X}_{n+1}(t) = \int_0^1 \psi(t,s)\mathcal{X}_n(s)ds + \mathcal{E}_{n+1}(t). \tag{2.15}$$

We use the kernel

$$\psi(s,t) = C \cdot s\mathbb{1}\{s \leq t\},$$

such that (2.15) becomes

$$\mathcal{X}_{n+1}(t) = C\int_0^t s\mathcal{X}_n(s)ds + \mathcal{E}_{n+1}(t), \tag{2.16}$$

Here, C is a normalizing constant to be chosen such that $||\Psi|| < 1$, which ensures the existence of a stationary causal solution to FAR(1) model; see Bosq (2000) [6]. We pick $C = 3$, such that $||\Psi|| = 0.5$.

We use the Brownian bridge process as the error process $\mathcal{E}(t)$ (see Didericksen (2012) [14]), which is defined by

$$\mathcal{E}(t) = W(t) - tW(t), \tag{2.17}$$

where $W(\cdot)$ is the standard Wiener process

$$W\left(\frac{k}{K}\right) = \frac{1}{\sqrt{K}} \sum_{j=1}^{k} Z_j, \quad k = 0, 1, \ldots, K,$$

where Z_k's are standard independent normal and $Z_0 = 0$. The interval $[0, 1]$ is equally partitioned such that $0 = t_1 < t_2 < \cdots < t_{99} < t_p = 1$ with $p = 100$. We choose the initial curve $\mathcal{X}_1 = cos(t)$ for $t \in [0, 1]$, and build the series $\mathcal{X}_1, \ldots, \mathcal{X}_{250}$ according to the following scheme for $j = 1, \ldots, 100$:

$$\begin{aligned} \mathcal{X}_1(t_j) &= cos(t_j), \\ \mathcal{X}_i(t_j) &= 3 \int_0^{t_j} s\mathcal{X}_{i-1}(s)ds + \mathcal{E}_i(t_j), \quad i = 2, \ldots, 250. \end{aligned}$$

Figure 2.1 displays the curves $\mathcal{X}_{101}, \mathcal{X}_{102}, ..., \mathcal{X}_{105}$. The last panel is the five curves combined. Out of the 250 curves we generate, the first 200 are used as the learning sample (i.e. $\{\mathcal{X}_i\}_{i=1,...,200}$), and the last 50 make up the testing sample (i.e. $\{\mathcal{X}_i\}_{i=201,...,250}$). The learning sample allows us to compute the kernel estimator while the testing sample will be accessed to assess the performance of the estimator and the behavior of bootstrap approximation.

2.5.2 Computing Kernel Estimator

With the simulated data, we use the learning sample to compute the kernel estimator by Eq. (2.3). The kernel function is chosen to be $K(u) = 1.5(1 - u^2)1_{[0,1]}(u)$, while the bandwidth h is selected through a cross-validation procedure.

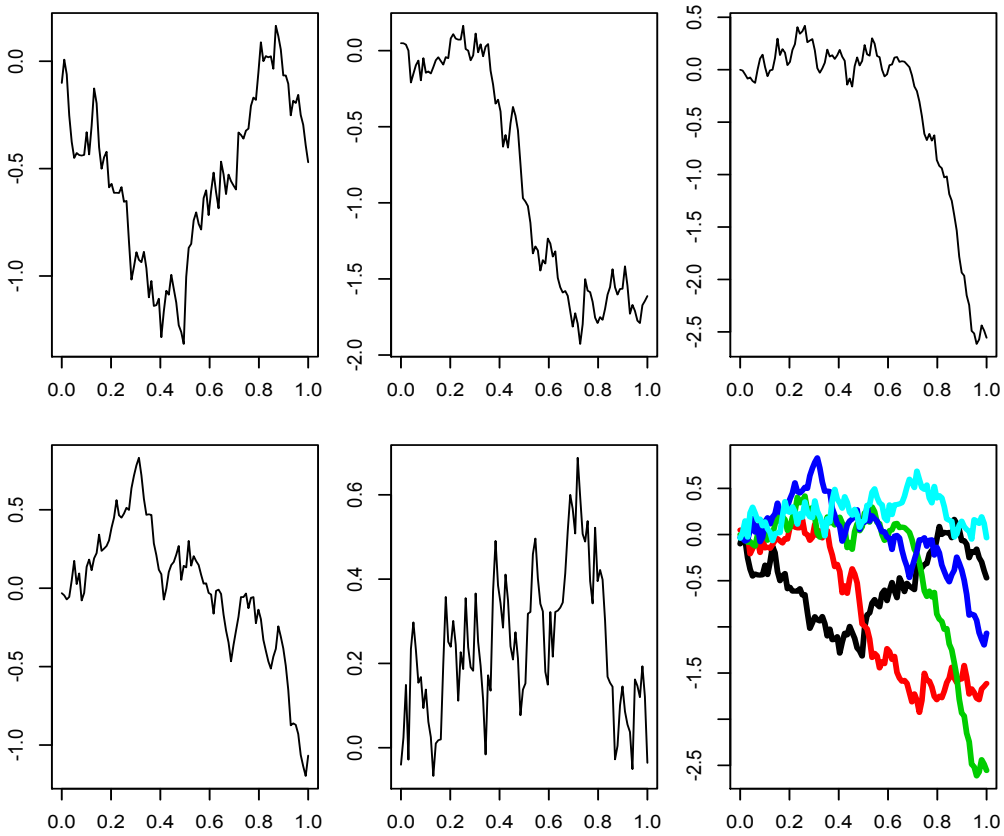

Figure 2.1: 5 Curves $\mathcal{X}_{101}, \mathcal{X}_{102}, ..., \mathcal{X}_{105}$ from the sample.

A projection-based semi-metric is considered:

$$d(\chi_1, \chi_2) = \sqrt{\sum_{k=1}^{J} \langle \chi_1 - \chi_2, v_{k,n} \rangle^2},$$

where $v_{k,n}$, $k = 1, \ldots, J$ are orthonormal eigenfunctions associated with the largest J eigenvalues of the empirical covariance operator of the learning sample:

$$\mathcal{C}(\cdot) = \frac{1}{200} \sum_{i=1}^{200} \langle \mathcal{X}_i, \cdot \rangle \mathcal{X}_i.$$

Figure 2.2 compares the kernel estimation (i.e.$\hat{\Psi}_h(\chi)$) with the true operator (i.e.$\Psi(\chi)$) at $\chi = \mathcal{X}_{201}, \mathcal{X}_{202}, \mathcal{X}_{203}, \mathcal{X}_{204}$, which shows the quality of the kernel estimation (the relatively higher volatility of panel 1 and 4 is due to the scaling difference of y axis).

2.6 Technical Proofs

Throughout this section, given some random variable U, P^U stands for the probability measure induced by U. Since the stationarity of the sequence $\{\mathcal{X}_n\}$, we assume $\mathcal{X}_i's$ are identically distributed as $\mathcal{X}$. Also, the pairs $(X_{i+1}, \mathcal{X}_i)$ have the same joint distribution. So are the functions of them. The kernel estimator $\hat{\psi}_h$ will be decomposed as follows:

$$\hat{\psi}_h(\chi) = \frac{\hat{g}_1(\chi)}{\hat{f}(\chi)},$$

where

$$\hat{g}_1(\chi) = \frac{1}{nF_\chi(h)} \sum_{i=1}^{n} X_{i+1} K\left(\frac{d(\mathcal{X}_i, \chi)}{h}\right)$$

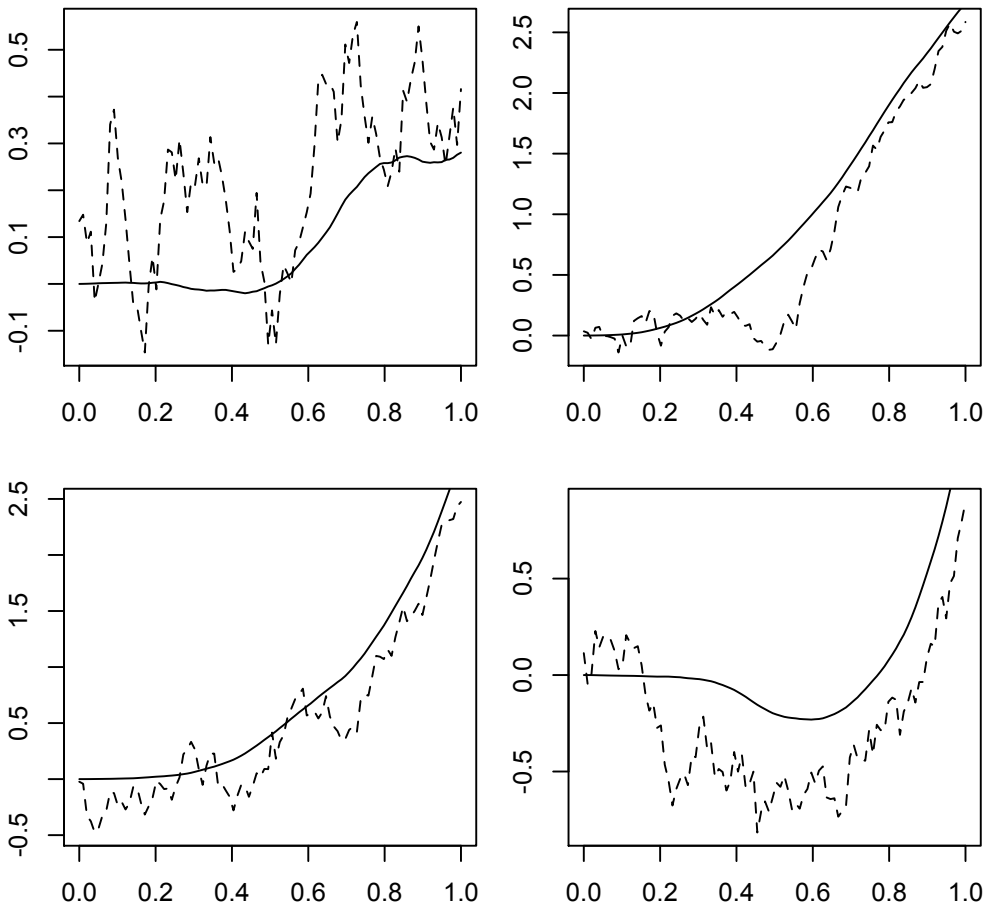

Figure 2.2: Kernel estimations $\hat{\Psi}_h(\chi)$ (dashed lines); true operator $\Psi(\chi)$ (solid lines).

and

$$\hat{f}(\chi) = \frac{1}{nF_\chi(h)} \sum_{i=1}^{n} K\left(\frac{d(\mathcal{X}_i, \chi)}{h}\right).$$

Similarly, the estimator $\hat{\Psi}_h$ can be decomposed as:

$$\hat{\Psi}_h(\chi) = \frac{\hat{g}_2(\chi)}{\hat{f}(\chi)},$$

where

$$\hat{g}_2(\chi) = \frac{1}{nF_\chi(h)} \sum_{i=1}^{n} \mathcal{X}_{i+1} K\left(\frac{d(\mathcal{X}_i, \chi)}{h}\right).$$

2.6.1 Proof of Theorem 2.4.1, 2.4.2 and 2.4.4

To prove these theorems, we need the following lemmas:

Lemma 2.6.1.

$$E[\hat{f}(\chi)] \to K(1) - \int_0^1 K'(s)\tau_{0\chi}(s)ds = M_1, \tag{2.18}$$

$$E[\hat{g}_1(\chi)] \to \psi(\chi)\left(K(1) - \int_0^1 K'(s)\tau_{0\chi}(s)ds\right) = \psi(\chi)M_1. \tag{2.19}$$

Proof. For the first assertion, we have

$$\begin{aligned}
E(\hat{f}(\chi)) &= \frac{1}{F_\chi(h)} E\left(K\left(\frac{d(\mathcal{X}, \chi)}{h}\right)\right) \\
&= \frac{1}{F_\chi(h)} \int K(t) dP^{d(\mathcal{X},\chi)/h}(t) \\
&= \frac{1}{F_\chi(h)} \left(K(1)F_\chi(h) - \int_0^1 K^{'}(s)F_\chi(hs)ds\right) \\
&= K(1) - \int_0^1 K^{'}(s)\frac{F_\chi(hs)}{F_\chi(h)}ds \\
&\to K(1) - \int_0^1 K'(s)\tau_{0\chi}(s)ds = M_1,
\end{aligned}$$

the third line coming from Eq. (14) in Ferraty et al. (2007) [17] and the fifth line follows from the uniform boundedness of the integrand. The second assertion is proved as follows:

$$\begin{aligned}
E(\hat{g}_1(\chi)) &= \frac{1}{F_\chi(h)} E\left(X_{i+1} K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right)\right) \\
&= \frac{1}{F_\chi(h)} E\left(E(X_{i+1}|\mathcal{X}_i) K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right)\right) \\
&= \frac{1}{F_\chi(h)} E\left(\psi(\mathcal{X}_i) K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right)\right) \\
&= \frac{1}{F_\chi(h)} E\left(\psi(\mathcal{X}) K\left(\frac{d(\mathcal{X},\chi)}{h}\right)\right) \\
&= \frac{1}{F_\chi(h)} E\left((\psi(\chi)+o(1)) K\left(\frac{d(\mathcal{X},\chi)}{h}\right)\right) \\
&\to \psi(\chi) M_1
\end{aligned}$$

where the fifth line follows from the continuity of ψ with respect to the semi-metric d. □

Lemma 2.6.2. *Under the assumption (A1)-(A8), we have*

$$Var(\hat{f}(\chi)) = \frac{M_2}{nF_\chi(h)}(1+o(1)), \tag{2.20}$$

$$Var(\hat{g}_1(\chi)) = (\sigma_\varepsilon^2 + \psi^2(\chi))\frac{M_2}{nF_\chi(h)}(1+o(1)), \tag{2.21}$$

$$Cov(\hat{f}(\chi), \hat{g}_1(\chi)) = \psi(\chi)\frac{M_2}{nF_\chi(h)}(1+o(1)). \tag{2.22}$$

Proof. Since the functional sequence $\{\mathcal{X}_n\}$ is stationary and strong mixing, the n pairs $(\mathcal{X}_i, X_{i+1})_{i=1,\cdots,n}$ are identically distributed and the sequence $(\mathcal{X}_i, X_{i+1})_i$ is also strong mixing. Then the results follows directly from Lemma 2.5 in Delsol

(2009) [13]. □

Proof of Theorem 2.4.1. For the proof of Eq. (2.8), we write the following decomposition:

$$E(\hat{\psi}(\chi)) = \frac{E(\hat{g}_1(\chi))}{E(\hat{f}(\chi))} + \frac{A_1}{(E(\hat{f}(\chi)))^2} + \frac{A_2}{(E(\hat{f}(\chi)))^2}, \tag{2.23}$$

with

$$A_1 = E(\hat{g}_1(\chi)(E(\hat{f}(\chi)) - \hat{f}(\chi))) \tag{2.24}$$

and

$$A_2 = E((\hat{f}(\chi) - E(\hat{f}(\chi)))^2 \hat{\psi}_h(\chi)). \tag{2.25}$$

The last two terms on the right-hand side of Eq. (2.23) are negligible since we have $A_1 = O((nF_\chi(h)^{-1}))$ and $A_2 = O((nF_\chi(h)^{-1}))$, both are direct consequences of Lemma 2.6.1 and 2.6.2 (see proof of Lemma 7.5 in Raña et al. (2016) [60]). For the first term on the right-hand side of (2.23), we calculate

$$\frac{E(\hat{g}_1(\chi))}{E(\hat{f}(\chi))} - \psi(\chi) = \frac{E\left((X_{i+1} - \psi(\chi))K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right)\right)}{E\left(K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right)\right)}, \tag{2.26}$$

which follows from the stationarity of the sequence $\{\mathcal{X}_n\}$. Assume $\mathcal{X}_i \sim \mathcal{X}$ for $i = 1, \cdots, \infty$, then we can write the numerator in (2.26) as follows

$$\begin{aligned} E\left((X_{i+1} - \psi(\chi))K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right)\right) &= E\left((\hat{\psi}(\mathcal{X}_i) - \psi(\chi))K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right)\right) \\ &= E\left((\hat{\psi}(\mathcal{X}) - \psi(\chi))K\left(\frac{d(\mathcal{X},\chi)}{h}\right)\right) \end{aligned}$$

$$
\begin{aligned}
&= E\left(\varphi_\chi(d(\mathcal{X},\chi))K\left(\frac{d(\mathcal{X},\chi)}{h}\right)\right)\\
&= \int \varphi_\chi(t)K(\frac{t}{h})dP^{d(\mathcal{X},\chi)}(t)\\
&= \int \varphi_\chi(ht)K(t)dP^{d(\mathcal{X},\chi)/h}(t)\\
&= h{\varphi_\chi}'(0)\int tK(t)dP^{d(\mathcal{X},\chi)/h}(t) + o(h),
\end{aligned}
$$

here P^U stands for the probability measure induced by U, and the last line comes form the first order Taylor's expansion for φ around 0. For the denominator in (2.26), we have

$$
E\left(K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right)\right) = \int K(t)dP^{d(\mathcal{X},\chi)/h}(t).
$$

Consequently,

$$
\frac{E(\hat{g}_1(\chi))}{E(\hat{f}(\chi))} - \psi(\chi) = h{\varphi_\chi}'(0)I + o(h), \tag{2.27}
$$

where

$$
I = \frac{\int tK(t)dP^{d(\mathcal{X},\chi)/h}(t)}{\int K(t)dP^{d(\mathcal{X},\chi)/h}(t)}.
$$

By Lemma 2 in Ferraty et al. (2007) [17], $I \to M_0/M_1$ as $n \to +\infty$. Finally, combining (2.23) and (2.27), we obtain

$$
E[\hat{\psi}_h(\chi)] - \psi(\chi) = \varphi'_\chi(0)\frac{M_0}{M_1}h + O(\frac{1}{nF_\chi(h)}) + o(h),
$$

which completes the proof. □

Proof of Theorem 2.4.2. We use the following decomposition:

$$\begin{aligned} Var(\hat{\psi}_h(\chi)) =& \frac{Var(\hat{g}_1(\chi))}{(E(\hat{f}(\chi)))^2} - 4\frac{E(\hat{g}_1(\chi)Cov(\hat{g}_1(\chi), \hat{f}(\chi)))}{(E(\hat{f}(\chi)))^3} \\ &+ 3Var(\hat{f}(\chi))\frac{(E(\hat{g}_1(\chi)))^2}{(E(\hat{f}(\chi)))^4} + o(\frac{1}{nF_\chi(h)}). \end{aligned} \tag{2.28}$$

Then Theorem 2.4.2 follows from this decomposition together with Lemma 2.6.1 and 2.6.2.

Proof of Theorem 2.4.4. Since the sequence $(\mathcal{X}_i, X_{i+1})_i$ is stationary and strong mixing, this theorem is the same as that obtained in Delsol (2009) [13].

2.6.2 Proof of Theorem 2.4.6

Proof. Consider the expression

$$\sqrt{nF_\chi(h)^{1+\delta}}\left[\hat{\Psi}_h(\chi) - \Psi(\chi) - \mathcal{B}_n\right]. \tag{2.29}$$

Following similar arguments as in the proof of Theorem 4.1 in Ferraty et al. (2012) [16], we have that the above expression has the same asymptotic distribution as

$$\begin{aligned} &\frac{\sqrt{nF_\chi(h)^{1+\delta}}}{M_1}\left[\hat{g}_2(\chi) - E\hat{g}_2(\chi) - (\hat{f}(\chi) - E\hat{f}(\chi))\Psi(\chi)\right] \\ =&\frac{1}{M_1\sqrt{nF_\chi(h)^{1-\delta}}}\sum_{i=1}^{n}\left[\mathcal{X}_{i+1}K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right) - E\left\{\Psi(\mathcal{X})K\left(\frac{d(\mathcal{X},\chi)}{h}\right)\right\}\right. \\ &\left. - \Psi(\chi)K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right) + \Psi(\chi)E\left\{K\left(\frac{d(\mathcal{X},\chi)}{h}\right)\right\}\right] \\ =&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_{n,i} - EZ_{n,i}), \end{aligned}$$

where for $1 \leq i \leq n$,

$$Z_{n,i} = \frac{1}{M_1\sqrt{F_\chi(h)^{1-\delta}}}\left[\mathcal{X}_{i+1}K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right) - \Psi(\chi)K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right)\right].$$

By assumption (i), we can apply Hölder's inequality to obtain

$$\begin{aligned}
&E||Z_{n,i}||^{2+\delta} \\
&= \frac{1}{M_1^{2+\delta}F_\chi(h)^{\frac{(1-\delta)(2+\delta)}{2}}}E\left(||\mathcal{X}_{i+1}-\Psi(\chi)||^{2+\delta}\left\{K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right)\right\}^{2+\delta}\right) \\
&\leq \frac{1}{M_1^{2+\delta}F_\chi(h)^{\frac{(1-\delta)(2+\delta)}{2}}}\left(E||\mathcal{X}_{i+1}-\Psi(\chi)||^{2+\delta'}\right)^{\frac{2+\delta}{2+\delta'}} \\
&\qquad\qquad \times\left\{E\left[K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right)\right]^{\frac{2}{1-\delta}}\right\}^{\frac{(1-\delta)(2+\delta)}{2}}.
\end{aligned}$$

In the above expression, $\left(E||\mathcal{X}_{i+1}-\Psi(\chi)||^{2+\delta'}\right)^{\frac{2+\delta}{2+\delta'}}$ is finite because of assumption (ii). For the last item, we note that

$$K^{\frac{2}{1-\delta}}(t) = K^{\frac{2}{1-\delta}}(1) - \int_t^1 (K^{\frac{2}{1-\delta}}(s))'ds.$$

Applying Fubini's Theorem, we obtain

$$\begin{aligned}
E\left[K\left(\frac{d(\mathcal{X}_i,\chi)}{h}\right)\right]^{\frac{2}{1-\delta}} &= \int_0^1 K^{\frac{2}{1-\delta}}(t)dP^{d(\mathcal{X},\chi)/h}(t) \\
&= K^{\frac{2}{1-\delta}}(1)F_\chi(h) - \int_0^1\left(\int_t^1 (K^{\frac{2}{1-\delta}}(s))'ds\right)dP^{d(\mathcal{X},\chi)/h}(t) \\
&= K^{\frac{2}{1-\delta}}(1)F_\chi(h) - \int_0^1 (K^{\frac{2}{1-\delta}}(s))'F_\chi(hs)ds \\
&= F_\chi(h)\left(K^{\frac{2}{1-\delta}}(1) - \int_0^1 (K^{\frac{2}{1-\delta}}(s))'\tau_{h\chi}(s)ds\right).
\end{aligned}$$

As the factor of $F_\chi(h)$ converges to $M_{\frac{2}{1-\delta}}$ for $h \to 0$ where

$$M_{\frac{2}{1-\delta}} = K^{\frac{2}{1-\delta}}(1) - \int_0^1 (K^{\frac{2}{1-\delta}}(s))'\tau_{0\chi}(s)ds$$

is a constant depending on $K(\cdot)$, the right-hand side is bounded by $CF_\chi(h)$ for a suitable C and all small enough h.

Now we have, $E||Z_{n,i}||^{2+\delta} \leq C < \infty$ for all n. Combining this with assumption (iii), it follows from Theorem 2.3(i) in Politis and Romano (1994) [52] that $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Z_{n,i} - EZ_{n,i})$ converge weakly to a Gaussian measure with mean 0 in $\mathbb{H}$. Hence, (2.29) converges weakly to the same measure and

$$\hat{\Psi}_h(\chi) = \Psi(\chi) - \mathcal{B}_n + O_p(\frac{1}{\sqrt{nF_\chi(h)^{1+\delta}}}). \tag{2.30}$$

Acknowledgement. Chapter 2 is joint work with Professor Dimitris Politis, published in Electronic Journal of Statistics in 2017. Tingyi Zhu and Dimitris Politis are the primary investigators and authors of the paper.

Chapter 3

Bootstrap Approximation of the Kernel Estimators

3.1 Introduction

In this chapter, we propose a bootstrapping procedure in the functional dependent framework that can be used for estimating the distribution of the projection of the kernel estimation introduced in the previous chapter. The bootstrap prediction regions are constructed as a measurement of accuracy for the functional prediction. A regression bootstrap scheme is implemented in the procedure which provides a simplification for the bootstrap method in the autoregression case. Franke et al. (2002) [19] first applied the regression-type bootstrap in univariate nonlinear autoregression for inference of the kernel estimator, and Neumann and

Kreiss (1998) [41] considered to what extent regression-type bootstrap procedures can be successfully applied as long as nonparametric estimators and tests for conditional mean in nonparametric autoregressions are considered. It was mentioned in Kreiss and Lahiri (2012) [37] that the regression-type bootstrap is also valid for the Yule-Walker estimates for coefficients in a parametric AR(p) model, but it might not lead to asymptotically valid results for more general statistics.

The rest of this chapter is organized as follows. A componentwise bootstrap scheme is introduced and its validity is shown in Section 3.2. A simulated study is given in Section 3.3 while Section 3.4 presents the approach of constructing the bootstrap prediction regions. The technical proof of the main theorem in provided in Section 3.5.

3.2 Bootstrap approximation

Ferraty et al. (2010, 2012) [15, 16] have employed both naive and wild bootstrap to approximate the asymptotic distribution of the kernel estimators for nonparametric functional regressions. Their first result showed the validity of bootstrap when the explanatory variable is functional and the response is real. To extend the bootstrap approach to the double functional setting, i.e. when both variables are functional, they introduced the notion of "componentwise bootstrap", in which the idea is to show that the bootstrap approximation has good theoretical

behaviors when functionals are projected to a fixed basis element e_k. Here we take advantage of the auxiliary univariate model in (2.5), extending this componentwise bootstrap idea to the functional autoregression.

First, we propose a bootstrap procedure to approximate the distribution of $\hat{\psi}_h(\chi) - \psi(\chi)$ under the AR model (2.5), which consists of the following steps:

Algorithm 3.2.1.

1. *For $i = 1, \ldots, n$, define $\hat{\varepsilon}_{i,b} = X_{i+1} - \hat{\psi}_b(\mathcal{X}_i)$, where b is a second smoothing parameter.*

2. *Draw n i.i.d. random variables $\varepsilon_1^*, \ldots, \varepsilon_n^*$ from the empirical distribution of $(\hat{\varepsilon}_{1,b} - \bar{\hat{\varepsilon}}_b, \ldots, \hat{\varepsilon}_{n,b} - \bar{\hat{\varepsilon}}_b)$ where $\bar{\hat{\varepsilon}}_b = n^{-1} \sum_{i=1}^n \hat{\varepsilon}_{i,b}$.*

3. *For $i = 1, \ldots, n-1$, let $X_{i+1}^* = \hat{\psi}_b(\mathcal{X}_i) + \varepsilon_{i+1}^*$.*

4. *Define*

$$\hat{\psi}_{hb}^*(\chi) = \frac{\sum_{i=1}^{n-1} X_{i+1}^* K(h^{-1} d(\mathcal{X}_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1} d(\mathcal{X}_i, \chi))}. \tag{3.1}$$

Remark 3.2.1. From a theoretical point of view, the second smoothing parameter b has to be asymptotically larger than h (see condition (D5)), so over-smoothing is needed to make the bootstrap procedure work, as is the case in the functional regression. However, the two bandwidths have to be fixed in practice and a cross-validation procedure is used to determine the bandwidths in the simulation study in Section 2.5.

Remark 3.2.2. It is also worth being noted that instead of generating a real bootstrap sequence, here we are actually generating a scatter plot. Every bootstrap point X^*_{i+1} is generated by the prior true point $\mathcal{X}_i$—see step 3 above—and the desired estimation in the bootstrap world comes from fitting the pairs $(X^*_{i+1}, \mathcal{X}_i)$. This is so called regression bootstrap scheme, introduced in Franke et al. (2002) [19] as one of the three bootstrap schemes proposed for the scalar-valued nonlinear autoregression. The main reason we use the regression bootstrap scheme here is in the fact that when conditioning on the sample $\{\mathcal{X}_1, \ldots, \mathcal{X}_n\}$, it eliminates the random element in the denominator of $\hat{\psi}^*_{hb}(\chi)$—see Eq. (3.1)—which makes the proof of Theorem 3.2.3 proceed (see details in appendix). The regression-type bootstrap is considered as an important simplification for the bootstrap method in autoregression. We refer the readers Neumann and Kreiss (1998) [41] for its applications in nonparametric autoregressions, and Kreiss and Lahiri (2012) [37] for its extensions in parametric time series models.

Theorem 3.2.3. *If conditions of Theorem 2.4.4 hold, as well as assumptions (D1)-(D7) in Appendix, we have*

$$\sup_{y\in\mathbb{R}}\Bigg|P^*\left(\sqrt{nF_\chi(h)}\{\hat{\psi}^*_{hb}(\chi)-\hat{\psi}_b(\chi)\}\le y\right) - P\left(\sqrt{nF_\chi(h)}\{\hat{\psi}_h(\chi)-\psi(\chi)\}\le y\right)\Bigg| \overset{a.s.}{\to} 0, \quad (3.2)$$

where P^* denotes probability conditioned on the sample $\{\mathcal{X}_1, \ldots, \mathcal{X}_n\}$.

Theorem 3.2.3 shows the validity of the bootstrap procedure for $\hat{\psi}_h$. Then,

the bootstrap procedure for $\hat{\Psi}_h$ is proposed as follows:

Algorithm 3.2.2.

1. *For $i = 1, \ldots, n$, define $\hat{\mathcal{E}}_{i,b} = \mathcal{X}_{i+1} - \hat{\Psi}_b(\mathcal{X}_i)$, where b is a second smoothing parameter.*

2. *Draw n i.i.d. random variables $\mathcal{E}_1^*, \ldots, \mathcal{E}_n^*$ from the empirical distribution of $(\hat{\mathcal{E}}_{1,b} - \bar{\hat{\mathcal{E}}}_b, \ldots, \hat{\mathcal{E}}_{n,b} - \bar{\hat{\mathcal{E}}}_b)$ where $\bar{\hat{\mathcal{E}}}_b = n^{-1} \sum_{i=1}^n \hat{\mathcal{E}}_{i,b}$.*

3. *For $i = 1, \ldots, n-1$, let $\mathcal{X}_{i+1}^* = \hat{\Psi}_b(\mathcal{X}_i) + \mathcal{E}_{i+1}^*$.*

4. *Define*

$$\hat{\Psi}_{hb}^*(\chi) = \frac{\sum_{i=1}^{n-1} \mathcal{X}_{i+1}^* K(h^{-1} d(\mathcal{X}_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1} d(\mathcal{X}_i, \chi))}. \tag{3.3}$$

Theorem 3.2.4. *For any $\eta \in \mathbb{H}$ with $||\eta|| = 1$ and any bandwidth h and b, let $\hat{\Psi}_{\eta,h}(\chi) = \langle \hat{\Psi}_h(\chi), \eta \rangle$, $\hat{\Psi}_{\eta,hb}^*(\chi) = \langle \hat{\Psi}_{hb}^*(\chi), \eta \rangle$ and $\Psi_\eta(\chi) = \langle \Psi(\chi), \eta \rangle$. If, in addition to conditions of Theorem 2.4.4, (D1)-(D7) in Appendix hold, we have for every χ*

$$\sup_{y \in \mathbb{R}} \Bigg| P^* \left(\sqrt{n F_\chi(h)} \{ \hat{\Psi}_{\eta,hb}^*(\chi) - \hat{\Psi}_{\eta,b}(\chi) \} \leq y \right) - P \left(\sqrt{n F_\chi(h)} \{ \hat{\Psi}_{\eta,h}(\chi) - \Psi_\eta(\chi) \} \leq y \right) \Bigg| \overset{a.s.}{\to} 0, \tag{3.4}$$

where P^* denotes probability conditioned on the sample $\{\mathcal{X}_1, \ldots, \mathcal{X}_n\}$.

Proof. Choosing a basis with $e_1 = \eta$, this theorem is a direct consequence of Theorem 3.2.3. □

3.3 Simulations

To illustrate Theorem 3.2.4, we compare, for $k = 1, 2, 3, 4$, the density function of the componentwise bootstrapped error $f^*_{k,\chi}$

$$\langle \hat{\Psi}^*_{hb}(\chi) - \hat{\Psi}_b(\chi), v_{k,n} \rangle,$$

with the density function of the component wise true error $f^{true}_{k,\chi}$

$$\langle \hat{\Psi}_h(\chi) - \Psi(\chi), v_{k,n} \rangle.$$

To estimate $f^*_{k,\chi}$, we use the bootstrap procedure as described in Section 3.2:

1. compute $\hat{\Psi}_b(\chi)$ over the learning sample $\mathcal{X}_1, \ldots, \mathcal{X}_{200}$,

2. repeat 200 times the bootstrap algorithm introduced in the previous section to obtain

$$\hat{\Psi}^{*1}_{hb}(\chi), \ldots, \hat{\Psi}^{*200}_{hb}(\chi),$$

3. estimate the density $f^*_{k,\chi}$ over the 200 values

$$\langle \hat{\Psi}^{*1}_{hb}(\chi) - \hat{\Psi}_b(\chi), v_{k,n} \rangle, \ldots, \langle \hat{\Psi}^{*200}_{hb}(\chi) - \hat{\Psi}_b(\chi), v_{k,n} \rangle.$$

The Monte-Carlo scheme is used to estimate $f^{true}_{k,\chi}$:

1. build 200 samples $\{\mathcal{X}^s_1, \ldots, \mathcal{X}^s_{200}\}_{s=1,\ldots,200}$,

2. for the sth sample $\{\mathcal{X}^s_1, \ldots, \mathcal{X}^s_{200}\}$, compute $\hat{\Psi}^s_h$ to obtain

$$\hat{\Psi}^1_h(\chi), \ldots, \hat{\Psi}^{200}_h(\chi),$$

3. estimate the density $f_{k,\chi}^{true}$ over the 200 values

$$\langle \hat{\Psi}_h^1(\chi) - \Psi(\chi), v_{k,n} \rangle, \ldots, \langle \hat{\Psi}_h^{200}(\chi) - \Psi(\chi), v_{k,n} \rangle.$$

The bandwidth $h = h_{CV}$ is selected through a cross-validation procedure and we set $b = h$, which was the same setting used in the context of functional regression in Ferraty et al. (2010) [15]. They have studied the influence of both bandwidths on the behavior of the bootstrap procedure by varying b and h around h_{CV}. Their simulation showed the bootstrap works well for any combination of b and h, which leads to the conclusion that bootstrap results are not sensitive to the bandwidth choice.

Figure 3.1 presents the comparisons between the estimated $f_{k,\chi}^{*}$ and the estimated $f_{k,\chi}^{true}$ for the first four components, at the curves $\chi = \mathcal{X}_{201}, \ldots, \mathcal{X}_{205}$. The density estimation is performed with Gaussian kernel and the bandwidth being chosen three times the rule-of-thumb bandwidth estimator $(1.06\hat{\sigma}n^{-1/5})$ where $\hat{\sigma}$ is the sample standard deviation.

3.4 Bootstrap prediction regions

3.4.1 Construction of bootstrap prediction regions

Politis (2013) [50] constructed bootstrap prediction intervals in regression based on a bootstrap approximation to the distribution of the error in prediction—also called a 'predictive root'. This method was later extended to autoregression by

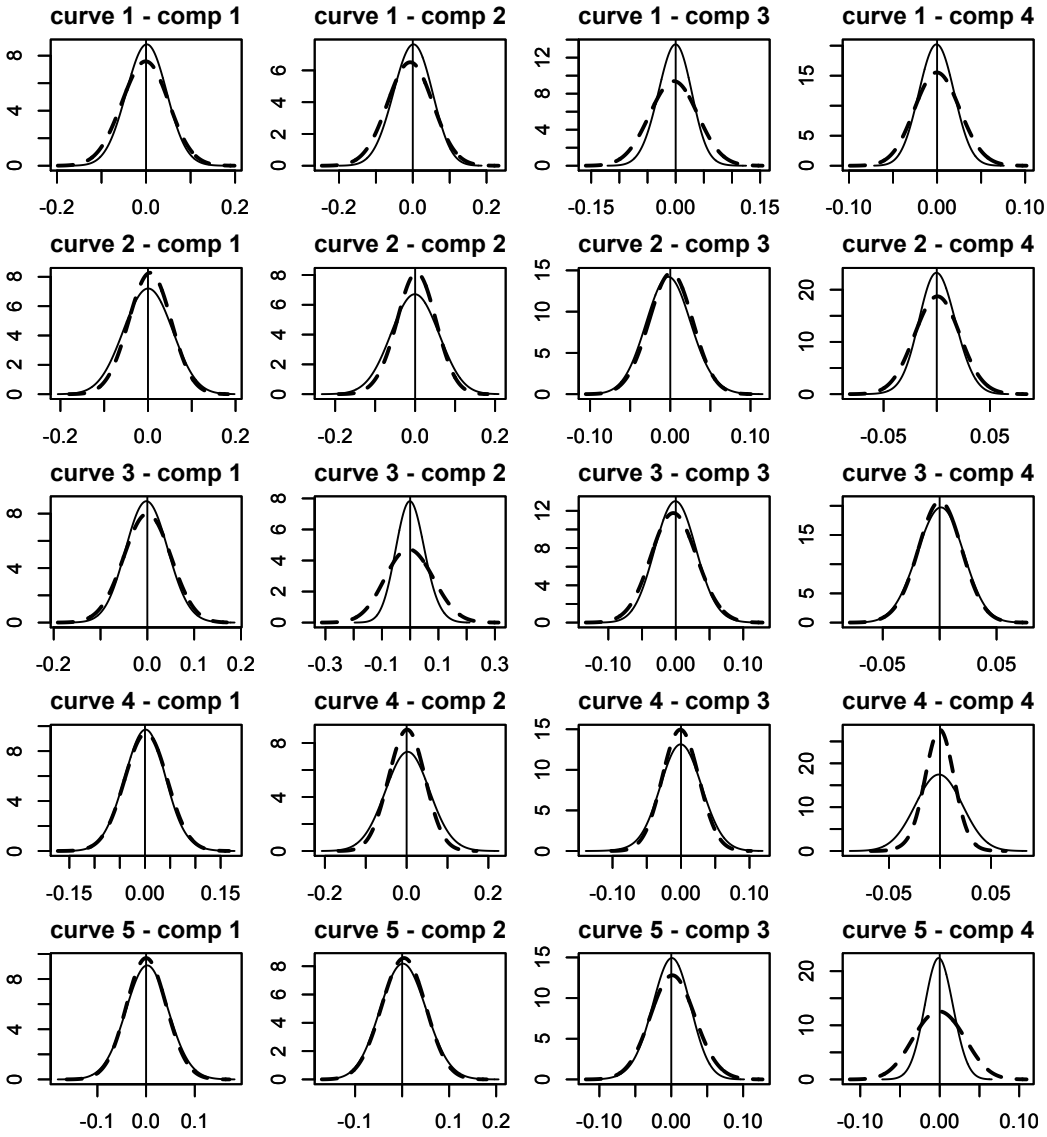

Figure 3.1: Solid line: true error, dashed line: bootstrap error.

Pan and Politis (2016) [42] who constructed bootstrap prediction intervals in the setting of a (univariate) nonparametric autoregression. Given a valid resampling procedure as developed in Section 3.2, we are able to construct bootstrap prediction regions in the functional autoregression model in the spirit of Pan and Politis (2016) [42]. Two bootstrap ideas, forward and backward, were considered in Pan and Politis (2016) [42] for the construction of prediction intervals with conditional validity. It is unrealistic, however, to generate the bootstrap pseudo-data backward in the functional setting. Therefore, we restrict to forward bootstrap scheme in our functional model.

Algorithm 3.4.1. *Bootstrap with fitted residual*

1. *Construct the Nadaraya-Watson kernel estimator* $\hat{\Psi}_h(\cdot)$ *by*

$$\hat{\Psi}_h(\chi) = \frac{\sum_{i=1}^{n-1} \mathcal{X}_{i+1} K(h^{-1} d(\mathcal{X}_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1} d(\mathcal{X}_i, \chi))}. \tag{3.5}$$

2. *Compute the fitted residuals:* $\hat{\mathcal{E}}_{i,b} = \mathcal{X}_i - \hat{\Psi}_b(\mathcal{X}_{i-1})$ *for* $i = 1, \ldots, n-1$*, where* b *is a second smoothing parameter.*

3. *Center the residuals:* $\hat{r}_{i,b} = \hat{\mathcal{E}}_{i,b} - (n-1)^{-1} \sum_{i=1}^{n-1} \hat{\mathcal{E}}_{i,b}$

   (a) *Let the empirical distribution of* $r_{i,b}$ *be denoted by* $\hat{F}_n$*, and draw bootstrap pseudo-residuals* $\mathcal{E}_1^*, \ldots, \mathcal{E}_n^*$ *i.i.d. from* $\hat{F}_n$*.*

(b) Use regression bootstrap scheme to generate the pseudo-data:

$$\mathcal{X}^*_{i+1} = \hat{\Psi}_b(\mathcal{X}_i) + \mathcal{E}^*_{i+1} \quad for\ i = 1, \ldots, n-1$$

(c) Re-estimate Ψ *by*

$$\hat{\Psi}^*_{hb}(\chi) = \frac{\sum_{i=1}^{n-1} \mathcal{X}^*_{i+1} K(h^{-1} d(\mathcal{X}_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1} d(\mathcal{X}_i, \chi))}. \tag{3.6}$$

(d) Calculate the bootstrap predictor

$$\hat{\mathcal{X}}^*_{n+1} = \hat{\Psi}^*_{hb}(\mathcal{X}_n)$$

and the future bootstrap observation

$$\mathcal{X}^*_{n+1} = \hat{\Psi}_b(\mathcal{X}_n) + \mathcal{E}^*_{n+1}$$

where $\mathcal{E}^*_{n+1}$ *is also drawn from* $\hat{F}_n$*.*

(e) Compute $||\mathcal{X}^*_{n+1} - \hat{\mathcal{X}}^*_{n+1}||$*, where* $||\cdot||$ *is a norm of the practitioner's choice; note that difference choices for the norm will lead to prediction regions of different shape.*

4. *Repeat steps (a)-(e) in the above* B *times. The* B *bootstrap replicates* $||\mathcal{X}^*_{n+1} - \hat{\mathcal{X}}^*_{n+1}||$ *are collected in the form of an empirical distribution whose* β*-quantile is denoted* $q(\beta)$*.*

5. *Finally, the* $(1-\beta)100\%$ *bootstrap predictive region for* $\mathcal{X}_{n+1}$ *consists of all* χ *such that*

$$||\chi - \hat{\Psi}_h(\mathcal{X}_n)|| \leq q(\beta).$$

Analogously to Pan and Politis (2016) [42], we also consider using predictive, as opposed to fitted, residuals for the bootstrap. We define the predictive residuals as $\hat{\mathcal{E}}_{t,b}^{(t)} = \mathcal{X}_t - \hat{\mathcal{X}}_t^{(t)}$ where $\hat{\mathcal{X}}_t^{(t)}$ is calculated from the delete-$\mathcal{X}_t$ data set, i.e. the available data for the scatterplot of $\mathcal{X}_k$ vs. $\mathcal{X}_{k-1}$ over which the fitting takes place excludes the single point that corresponds to $k = t$. The forward bootstrap with predictive residuals is similar to Algorithm 3.4.1 except for Step 2.

Algorithm 3.4.2. *Bootstrap with predictive residuals*

1. *Same as step 1 of Algorithm 3.4.1.*

2. *Use the delete-$\mathcal{X}_t$ dataset to compute the delete-one kernel estimator*

$$\hat{\Psi}_b^{(t)}(\chi) = \frac{\sum_{i=1, i\neq t}^{n-1} \mathcal{X}_{i+1} K(b^{-1} d(\mathcal{X}_i, \chi))}{\sum_{i=1, i\neq t}^{n-1} K(b^{-1} d(\mathcal{X}_i, \chi))}, \quad for\ t = 1, \ldots, n-1. \tag{3.7}$$

   Then compute the predictive residuals: $\hat{\mathcal{E}}_{t,b}^{(t)} = \mathcal{X}_t - \hat{\Psi}_b^{(t)}(\mathcal{X}_{t-1})$ for $t = 1, \ldots, n-1$, where b is a second smoothing parameter.

3-5. *Replace $\hat{\mathcal{E}}_{t,b}$ by $\hat{\mathcal{E}}_{t,b}^{(t)}$ in Algorithm 3.4.1; the remaining steps are the same.*

Remark 3.4.1. Recall that prediction intervals are asymptotically valid (in a conditional sense) when the probability of coverage of the future value $\mathcal{X}_{n+1}$ conditional on the observed data $\mathcal{X}_1, \ldots, \mathcal{X}_n$ gets close to the nominal one as n increases. The prediction regions constructed in Algorithm 3.4.1 and 3.4.2 would indeed be asymptotically valid if the predictive root $\mathcal{X}_{n+1} - \hat{\mathcal{X}}_{n+1}$ and the bootstrap predictive

root $\mathcal{X}^*_{n+1} - \hat{\mathcal{X}}^*_{n+1}$ have the same distribution asymptotically. For the predictive root, we have

$$\mathcal{X}_{n+1} - \hat{\mathcal{X}}_{n+1} = \Psi(\mathcal{X}_n) - \hat{\Psi}_h(\mathcal{X}_n) + \mathcal{E}_{n+1} = A + \mathcal{E}_{n+1},$$

where $\mathcal{E}_{n+1}$ is independent of the estimation error A. Similarly, the bootstrap predictive root can be written as

$$\mathcal{X}^*_{n+1} - \hat{\mathcal{X}}^*_{n+1} = \hat{\Psi}_b(\mathcal{X}_n) - \hat{\Psi}^*_{hb}(\mathcal{X}_n) + \mathcal{E}^*_{n+1} = A^* + \mathcal{E}^*_{n+1},$$

where $\mathcal{E}^*_{n+1}$ is independent of A^*. Consequently, the prediction regions of Algorithm 3.4.1 and 3.4.2 would be asymptotically valid provided that the distribution of the true errors $\mathcal{E}_1, \ldots, \mathcal{E}_n$ is captured in the limit by the empirical distribution of the residuals (fitted or predictive).

Corollary 3.4.2. *Assume the conditions of Theorem 2.4.6 and 3.2.4 as well as assumptions strong enough to ensure that* $||\hat{\mathcal{E}}_{1,b}|| \xrightarrow{d} ||\mathcal{E}_1||$ *as* $n \to \infty$ *where* $||\cdot||$ *is the norm appearing in Algorithm 3.4.1 and* $\xrightarrow{d}$ *denotes convergence in distribution. Then, the prediction region of Algorithm 3.4.1 is asymptotically valid. If in addition* $||\hat{\mathcal{E}}^{(1)}_{1,b}|| \xrightarrow{d} ||\mathcal{E}_1||$*, then the prediction region of Algorithm 3.4.2 is asymptotically valid as well.*

Remark 3.4.3. The proof of the corollary is immediate since, under the conditions of Theorem 2.4.6 and 3.2.4, we have both $A \xrightarrow{p} 0$ and $A^* \xrightarrow{p} 0$. Note that the condition $||\hat{\mathcal{E}}_{1,b}|| \xrightarrow{d} ||\mathcal{E}_1||$ would follow if $\hat{F}_n$, the empirical distribution of the

residuals, were shown to converge weakly to the distribution of the true errors as shown e.g. by Franke and Nyarige (2016) [20] in the case of a linear operator Ψ. However, the importance of Corollary 3.4.2 is limited: as discussed in Politis (2015) [51]—asymptotic validity of prediction regions does not tell the whole story since the variability of estimation—which is asymptotically negligible—does not enter the picture. A prediction region will have good finite sample coverage only if the method employed is able to capture the variability of estimation error; that is why Algorithms 3.4.1 and 3.4.2 employ the bootstrap for prediction intervals instead of just relying on the empirical quantiles of $\hat{F}_n$. Indeed, by construction, our model-based bootstrap procedure is capable of approximating the distribution of the re-scaled estimation error $\sqrt{nF_\chi(h)}A$ by that of $\sqrt{nF_\chi(h)}A^*$ due to Theorem 3.2.4. Hence, our bootstrap prediction regions should have good finite sample coverage which is something that can not be captured by the property of asymptotic validity.

As mentioned in the introduction, Aue et al. (2015) [4] proposed a method of predicting FAR(1) using functional principal component analysis (FPCA). In addition, they proposed an algorithm for computing the prediction regions to assess the prediction accuracy. That is the only existing method we could find on the construction of prediction regions for functional time series. In the next subsection, we are going to compare it with the algorithm we proposed.

3.4.2 Monte Carlo studies

In this subsection, we use the Monte Carlo simulations to evaluate the quality of the bootstrap prediction regions with both fitted and predictive residuals. Simulations are performed on the model from the previous section; see Eq. (2.16). 500 'true' datasets each with sample size n are generated and n varies from 100 to 400. For the ith true dataset, we use one of the bootstrap methods to create $B = 1000$ bootstrap sample paths (step 4 of the algorithms), and construct the prediction region.

To assess the corresponding coverage level (CVR) of the constructed region, we also generate 1000 one-step ahead predictions $Y_{n+1,j} = \hat{\Psi}_i(\mathcal{X}_{n,i})$ for $j = 1, 2, ...1000$ where $\hat{\Psi}_i$ is the estimate from the ith 'true' dataset, $\mathcal{X}_{n,i}$ is the ith dataset's last data and $\mathcal{E}_j$ is randomly drawn from the error process (2.17). Then the empirical coverage level from the ith dataset is given by

$$CVR_i = \frac{1}{1000} \sum_{j=1}^{1000} \mathbb{1} \left\{ ||Y_{n+1,j} - \hat{\Psi}_b(\mathcal{X}_n)|| < q_i(\beta) \right\}$$

where $\mathbb{1}(\cdot)$ is the indicator function and β is the nominal coverage level. The coverage level for bootstrap methods is calculated by the average $\{CVR_i\}$ over the 500 'true' datasets, i.e.

$$CVR = \frac{1}{500} \sum_{i=1}^{500} CVR_i.$$

Note that the last observation $\mathcal{X}_{n,i}$ is different for each dataset; hence the coverage CVR represents an unconditional coverage probability.

Prediction regions are constructed with nominal coverage levels of 95% and 90%. Five different norms in the functional space are selected to measure the proximity between functions. The first three are the common l_p norms in the functional space:

- l_1: $||f||_1 = \int_0^1 |f(t)|dt$
- l_2: $||f||_2 = (\int_0^1 |f(t)|^2 dt)^{1/2}$
- l_∞: $||f||_\infty = sup_{t\in[0,1]}|f(t)|$.

We also consider two pointwise measures:

- 1st coordinate: $||x - y||_{coordinate} = |x(0) - y(0)|$
- 1st component: $||x - y||_{component} = |\langle x - y, v_{1,n}\rangle|$

where x and y are functions in $L^2[0, 1]$ and e_1 is the eigenfunction associated with the largest eigenvalue of the sample covariance operator. Table 3.1 presents the empirical coverage rate of the prediction region we construct. The results are promising as the empirical coverage rates pretty well match the nominal coverage rate even with quite small sample size. As expected, when using predictive residuals, the coverage rate is a bit higher compared to fitted residuals.

For the comparison purpose, we apply the algorithm of Aue et al. (2015) [4] on the functional data we generate. Table 3.2 presents the coverage rate of their constructed prediction region. The value of p represents the number of components

Table 3.1: Comparison between empirical coverage rate and nominal coverage rate using (a) fitted residuals and (b) predictive residuals.

fitted residuals	nominal coverage 95%			nominal coverage 90%		
n	100	200	400	100	200	400
l_1	0.956	0.955	0.953	0.914	0.910	0.907
l_2	0.956	0.955	0.953	0.914	0.910	0.908
l_∞	0.959	0.957	0.956	0.918	0.913	0.910
$1st$ coordinate	0.953	0.954	0.953	0.907	0.908	0.904
$1st$ component	0.954	0.955	0.955	0.907	0.909	0.908

(a) Fitted residuals

predictive residuals	nominal coverage 95%			nominal coverage 90%		
n	100	200	400	100	200	400
l_1	0.957	0.955	0.955	0.913	0.910	0.911
l_2	0.957	0.956	0.956	0.914	0.911	0.911
l_∞	0.960	0.959	0.958	0.918	0.915	0.916
$1st$ coordinate	0.956	0.954	0.955	0.911	0.908	0.909
$1st$ component	0.956	0.957	0.956	0.912	0.913	0.910

(b) Predictive residuals

Table 3.2: Empirical coverage rate compared to the nominal coverage rate 95%.

n	100	200	400
$p = 2$	0.859	0.923	0.948
$p = 3$	0.874	0.928	0.945
$p = 4$	0.881	0.931	0.951

to be kept in the prediction algorithm; see Aue et al. (2015) [4] for more details. The experiment is also done on three different sample size, 100, 200 and 400. As is shown in the table, their prediction regions fail to achieve the ideal coverage rate when sample size is small–see the case $n = 100$–and results get better when sample size grows. This phenomenon is not surprising, and explainable. Recall that there are two constituents in the expression of predictive root $\mathcal{X}_{n+1} - \hat{\mathcal{X}}_{n+1}$, the estimation error A and the true error $\mathcal{E}$. The bootstrap predictive root in our algorithm is capable of capturing the distribution of both errors by A^* and $\mathcal{E}^*$. However, the method of Aue et al. (2015) [4] does not attempt to capture the variability of estimation error and thus, as expected, it yields coverages less than nominal (undercoverage). Coverage rate improves under larger sample size since the estimation error diminishes as n grows. As a conclusion, we can say that the prediction regions constructed through our bootstrap procedure can achieve better finite sample coverage.

3.5 Proof of Theorem 3.2.3

Let us first state some additional assumptions which are needed:

(D1) For some $\beta > 0$, all $0 \leq s \leq \beta$ and all χ_1 in a neighborhood of χ with respect to the semi-metric d, $\varphi_{\chi_1}(0) = 0$, $\varphi'_{\chi_1}(s)$ exists, $\varphi'_{\chi_1}(0) \neq 0$, and $\varphi'_{\chi_1}(s)$ is uniformly Hölder continuous of order $0 < \alpha \leq 1$ at 0, i.e. there exists a constant $0 < L < \infty$ such that

$$|\varphi'_{\chi_1}(s) - \varphi'_{\chi_1}(0)| \leq Ls^{\alpha}, \quad \text{for all } 0 \leq s \leq \beta,$$

uniformly in χ_1 in a neighborhood of χ.

(D2) For all $\chi_1 \in \mathbb{H}$, $F_{\chi_1}(0) = 0$ and $F_{\chi_1}(t)/F_{\chi}(t)$ is Hölder continuous of order α in χ_1, uniformly in t in a neighborhood of 0 (with α defined in (D1)), i.e. for some $\beta > 0$, there exists a constant $0 < M < \infty$ such that

$$|(F_{\chi_1}(t) - F_{\chi_2}(t))/F_{\chi}(t)| \leq Md(\chi_1, \chi_2)^{\alpha}, \quad \text{for all } \chi_1, \chi_2 \in \mathbb{H},$$

uniformly for all $0 < t \leq \beta$.

(D3) For all $\chi_1 \in \mathbb{H}$ and all $0 \leq s \leq 1$, $\tau_{0\chi_1}(s)$ exists, $\sup_{\chi_1 \in \mathbb{H},\, 0 \leq s \leq 1} |\tau_{h\chi_1}(s) - \tau_{0\chi_1}(s)| = o(1)$, $M_{0\chi} > 0$, $M_{2\chi} > 0$, $\inf_{d(\chi_1,\chi)<\varepsilon} M_{1\chi_1} > 0$ for some $\varepsilon > 0$, and $M_{k\chi_1}$ is Hölder continuous of order α for $k = 0, 1, 2$, (with α defined in (D1)), i.e. for $k = 0, 1, 2$, there exists a $0 < N_k < \infty$ such that

$$|M_{k\chi_1} - M_{k\chi_2}| \leq N_k d(\chi_1, \chi_2)^{\alpha}, \quad \text{for all } \chi_1, \chi_2 \in \mathbb{H}.$$

(Here the quantities M_k, $k = 0, 1, 2$ are indexed with an additional index χ, which is necessary to the above continuity conditions of M_k as a function of χ.)

(D4) $E(|X_{i+1}||\mathcal{X}_i = \cdot)$ is continuous in a neighborhood of χ with respect to the semi-metric d, and $\sup_{d(\chi_1,\chi)<\epsilon} E(|X_{i+1}|^m|\mathcal{X}_i = \chi_1) < \infty$ for some $\epsilon > 0$ and all $m \geq 1$.

(D5) $b \to 0$, $h/b \to 0$, $h(nF_\chi(h))^{1/2} = O(1)$, $F_\chi(b)^{-1}h/b = O(1)$, $bh^{\alpha-1} = O(1)$, $n^{1/p}F_\chi(h)^{1/2}log(n) = o(1)$, $b^{1+\alpha}(nF_\chi(h))^{1/2} = o(1)$, $F_\chi(b+h)/F_\chi(b) \to 1$, and $[F_\chi(h)/F_\chi(b)]log(n) = o(1)$ (with p and α defined in (A6) and (D1), respectively).

(D6) For each n, there exist $r_n \geq 1$, $l_n > 0$ and curves $t_{1n}, \ldots, t_{r_n n}$ such that $B(\chi, h) \subset \cup_{k=1}^{r_n} B(t_{kn}, l_n)$, with $r_n = O(n^{b/h})$ and $l_n = o(b(nF_\chi(h))^{-1/2})$, where $B(\chi, t) = \{\chi \in \mathbb{H} : d(\chi_1, \chi) \leq t\}$ is the ball in $\mathbb{H}$ with center χ and radius t.

(D7) $a > 4.5$ (with a defined in (A8)).

Note that (D1)-(D3) are regularity conditions related to the smoothness of the functions ψ, φ_χ, F_χ and $\tau_{0\chi}$, which have been used in the context of bootstrap in functional regression. Assumption (D7), along with $n^{1/p}F_\chi(h)^{1/2}log(n) = o(1)$ and $F_\chi(b)^{-1}h/b = O(1)$ in (D5) are additional assumptions made in the dependent case in Raña et al. (2016) [60] which facilitates the proof of Lemma 3.5.2. See more details on these assumptions in Ferraty et al. (2007) [17], (2010) [15].

Proof. The expression between the absolute values of (3.2) can be decomposed as

$$P^*\left(\sqrt{nF_\chi(h)}\{\hat{\psi}^*_{hb}(\chi)-\hat{\psi}_b(\chi)\}\le y\right)- P\left(\sqrt{nF_\chi(h)}\{\hat{\psi}_h(\chi)-\psi(\chi)\}\le y\right)$$
$$= T_1(y)+T_2(y)+T_3(y)$$

where

$$T_1(y)=P^*\left(\sqrt{nF_\chi(h)}\{\hat{\psi}^*_{hb}(\chi)-\hat{\psi}_b(\chi)\}\le y\right) - \Phi\left(\frac{y-\sqrt{nF_\chi(h)}\{E^*\hat{\psi}^*_{hb}(\chi)-\hat{\psi}_b(\chi)\}}{\sqrt{nF_\chi(h)var^*(\hat{\psi}^*_{hb}(\chi))}}\right),$$
$$T_2(y)=\Phi\left(\frac{y-\sqrt{nF_\chi(h)}\{E^*\hat{\psi}^*_{hb}(\chi)-\hat{\psi}_b(\chi)\}}{\sqrt{nF_\chi(h)var^*(\hat{\psi}^*_{hb}(\chi))}}\right) - \Phi\left(\frac{y-\sqrt{nF_\chi(h)}\{E\hat{\psi}_h(\chi)-\psi(\chi)\}}{\sqrt{nF_\chi(h)var(\hat{\psi}_h(\chi))}}\right)$$

and

$$T_3(y)=\Phi\left(\frac{y-\sqrt{nF_\chi(h)}\{E\hat{\psi}_h(\chi)-\psi(\chi)\}}{\sqrt{nF_\chi(h)var(\hat{\psi}_h(\chi))}}\right) - P\left(\sqrt{nF_\chi(h)}\{\hat{\psi}_h(\chi)-\psi(\chi)\}\le y\right).$$

By the asymptotic normality of $\hat{\psi}_h$ given in Theorem 2.4.4, we have $T_3(y)\to 0$. The a.s. convergence to 0 of $T_1(y)$ is given by the asymptotic normality of $\hat{\psi}^*_{hb}$ conditioning on $\mathcal{X}_i, i=1,...,n$, which is proved below.

We decompose $\hat{\psi}^*_{hb}$ as follows

$$\hat{\psi}^*_{hb}(\chi) = \frac{\sum_{i=1}^{n-1} X^*_{i+1} K(h^{-1} d(\mathcal{X}_i, \chi))}{\sum_{i=1}^{n-1} K(h^{-1} d(\mathcal{X}_i, \chi))} = \frac{\hat{g}^*(\chi)}{\hat{f}(\chi)},$$

where

$$\hat{g}^*(\chi) = \frac{1}{nF_\chi(h)} \sum_{i=1}^{n-1} X^*_{i+1} K(h^{-1} d(\mathcal{X}_i, \chi)),$$

$$\hat{f}(\chi) = \frac{1}{nF_\chi(h)} \sum_{i=1}^{n-1} K(h^{-1} d(\mathcal{X}_i, \chi)).$$

Then we have

$$\hat{g}^*(\chi) = \frac{1}{nF_\chi(h)} \sum_{i=1}^{n-1} (\hat{\psi}_b(\mathcal{X}_i) + \varepsilon^*_{i+1}) K(h^{-1} d(\mathcal{X}_i, \chi)),$$

$$E^*(\hat{g}^*(\chi)) = \frac{1}{nF_\chi(h)} \sum_{i=1}^{n-1} (\hat{\psi}_b(\mathcal{X}_i) + E^* \varepsilon^*_{i+1}) K(h^{-1} d(\mathcal{X}_i, \chi)).$$

Therefore,

$$\begin{aligned} \frac{\hat{\psi}^*_{hb}(\chi) - E^*(\hat{\psi}^*_{hb}(\chi))}{\sqrt{var^*(\hat{\psi}^*_{hb}(\chi))}} &= \frac{\frac{\hat{g}^*(\chi)}{\hat{f}(\chi)} - E^*(\frac{\hat{g}^*(\chi)}{\hat{f}(\chi)})}{\sqrt{var^*(\frac{\hat{g}^*(\chi)}{\hat{f}(\chi)})}} \\ &= \frac{\hat{g}^*(\chi) - E^*(\hat{g}^*(\chi))}{\sqrt{var^*(\hat{g}^*(\chi))}} \\ &= \frac{\hat{h}^*(\chi) - E^*(\hat{h}^*(\chi))}{\sqrt{var^*(\hat{h}^*(\chi))}} \end{aligned}$$

where

$$\hat{h}^*(\chi) = \frac{1}{nF_\chi(h)} \sum_{i=1}^{n-1} \varepsilon^*_{i+1} K(h^{-1} d(\mathcal{X}_i, \chi)).$$

$\hat{h}^*(\chi)$ is a sum of a mixing sequence and its asymptotic normality follows from the similar arguments in the proof of Theorem 3 in Delsol (2009) [13].

A special case is when $K(\cdot) = \mathbb{1}_{[0,1]}(\cdot)$, under which

$$\hat{h}^*(\chi) = \frac{1}{\#\{i : d(\mathcal{X}_i, \chi) \leq h\}} \sum_{i:d(\mathcal{X}_i,\chi)\leq h} \varepsilon^*_{i+1},$$

so that $\hat{h}^*(\chi)$ is a sum of independent random variable given $\mathcal{X}_i, i = 1, \ldots, n.$ and the asymptotic normality follows directly.

It remains to consider $T_2(y)$, and its a.s convergence to 0 follows from Lemma 3.5.1 and 3.5.2 that follow:

Lemma 3.5.1. *Assume (A1), (A4) and (D1)-(D5). Then*

$$\frac{var^*[\hat{\psi}^*_{hb}(\chi)]}{var[\hat{\psi}_h(\chi)]} \to 1 \quad a.s.$$

Proof. Define $\hat{\sigma}^2_\varepsilon = n^{-1}\sum_{i=1}^n (\hat{\varepsilon}_{i,b} - \bar{\hat{\varepsilon}}_b)^2$. Then

$$\begin{aligned}
var^*[\hat{\psi}^*_{hb}(\chi)] &= var^*\left[\frac{\sum_{i=1}^{n-1}(\hat{\psi}_b(\mathcal{X}_i) + \varepsilon^*_{i+1})K(h^{-1}d(\mathcal{X}_i,\chi))}{\sum_{i=1}^{n-1} K(h^{-1}d(\mathcal{X}_i,\chi))}\right] \\
&= var^*\left[\frac{\sum_{i=1}^{n-1}\varepsilon^*_{i+1}K(h^{-1}d(\mathcal{X}_i,\chi))}{\sum_{i=1}^{n-1} K(h^{-1}d(\mathcal{X}_i,\chi))}\right] \\
&= \frac{\sum_{i=1}^{n-1} K^2(h^{-1}d(\mathcal{X}_i,\chi))var^*(\varepsilon^*_{i+1})}{\left(\sum_{i=1}^{n-1} K(h^{-1}d(\mathcal{X}_i,\chi))\right)^2} \\
&= \frac{\hat{\sigma}^2_\varepsilon}{\hat{f}(\chi)^2}(nF_\chi(h))^{-2}\sum_{i=1}^{n-1} K^2(h^{-1}d(\mathcal{X}_i,\chi)) \\
&= \frac{\sigma^2_\varepsilon}{E[\hat{f}(\chi)]^2}(nF^2_\chi(h))^{-1}\cdot E[K^2(h^{-1}d(\mathcal{X}_i,\chi))]\cdot(1+o(1))
\end{aligned}$$

$$= \frac{\sigma_\varepsilon^2}{M_1^2}\frac{M_2}{nF_\chi(h)}(1+o(1))$$

$$= var[\hat{\psi}_h(\chi)] + o((nF_\chi(h))^{-1}).$$

Since $var[\hat{\psi}_h(\chi)] = O((nF_\chi(h))^{-1})$ by Theorem 2.4.2, the result follows by dividing $var[\hat{\psi}_h(\chi)]$ on both sides. □

Lemma 3.5.2. *Assume (A1)-(A11) and (D1)-(D7). Then*

$$\sqrt{nF_\chi(h)}\{E^*[\hat{\psi}^*_{hb}(\chi)] - \hat{\psi}_b(\chi) - E[\hat{\psi}_h(\chi)] + \psi(\chi)\} \xrightarrow{a.s.} 0.$$

Proof. Write

$$\begin{aligned}
&E^*[\hat{\psi}^*_{hb}(\chi)] - \hat{\psi}_b(\chi)\\
&= E^*\left[\frac{\sum_{i=1}^{n-1}(\hat{\psi}_b(\mathcal{X}_i)+\varepsilon^*_{i+1})K(h^{-1}d(\mathcal{X}_i,\chi))}{\sum_{i=1}^{n-1}K(h^{-1}d(\mathcal{X}_i,\chi))}\right] - \hat{\psi}_b(\chi)\\
&= \frac{\sum_{i=1}^{n-1}\hat{\psi}_b(\mathcal{X}_i)K(h^{-1}d(\mathcal{X}_i,\chi))}{\sum_{i=1}^{n-1}K(h^{-1}d(\mathcal{X}_i,\chi))} - \hat{\psi}_b(\chi)\\
&= \frac{(nF_\chi(h))^{-1}}{\hat{f}_h(\chi)}\sum_{i=1}^{n-1}\{\hat{\psi}_b(\mathcal{X}_i) - \hat{\psi}_b(\chi)\}K(h^{-1}d(\mathcal{X}_i,\chi))\\
&= \frac{(nF_\chi(h))^{-1}}{\hat{f}_h(\chi)}\sum_{i=1}^{n-1}\{\hat{\psi}_b(\mathcal{X}_i) - \hat{\psi}_b(\chi) - E[\hat{\psi}_b(\mathcal{X}_i)] + E[\hat{\psi}_b(\chi)]\}K(h^{-1}d(\mathcal{X}_i,\chi))\\
&\quad + \frac{(nF_\chi(h))^{-1}}{\hat{f}_h(\chi)}\sum_{i=1}^{n-1}\{E[\hat{\psi}_b(\mathcal{X}_i)] - E[\hat{\psi}_b(\chi)] - \psi(\mathcal{X}_i) + \psi(\chi)\}K(h^{-1}d(\mathcal{X}_i,\chi))\\
&\quad + \frac{(nF_\chi(h))^{-1}}{\hat{f}_h(\chi)}\sum_{i=1}^{n-1}\{\psi(\mathcal{X}_i) - \psi(\chi)\}K(h^{-1}d(\mathcal{X}_i,\chi))
\end{aligned}$$

$$= U_1 + U_2 + U_3$$

Using our Lemma 2.6.2 instead of Theorem 1 in Ferraty et al. (2007) [17], we can follow the lines of the proof of Lemma 5 in Ferraty et al. (2010) [15] to obtain $U_1 = o((nF_\chi(h))^{-1/2})$ a.s. We can also obtain $U_2 = o((nF_\chi(h))^{-1/2})$ a.s. by following the lines of the proof of (8.11) in Raña et al. (2016) [60], which is the extension of the Lemma 6 in Ferraty et al. (2010) [15] to the dependent data case. Finally, it is easy to see from Theorem 2.4.4 that $U_3 = E[\hat{\psi}_h(\chi)] - \psi(\chi) + o((nF_\chi(h))^{-1/2})$, which completes the proof of the lemma. □

Acknowledgement. Chapter 3 is joint work with Professor Dimitris Politis, published in Electronic Journal of Statistics in 2017. Tingyi Zhu and Dimitris Politis are the primary investigators and authors of the paper.

Chapter 4

Higher-order Accurate Spectral Density Estimation of Functional Time Series

4.1 Introduction

Typically, we consider a stationary functional sequence $\{X_t(\tau); \tau \in [0,1]\}_{t\in\mathbb{Z}}$ whose terms are random elements of the separable Hilbert space $L^2([0,1],\mathbb{R})$. The central issue in the analysis of functional time series is to take into account the temporal dependence between the observations. In this respect, a fundamental element is the investigation of second-order characteristics of the functional sequences. A handful of early papers have studied the covariance structure of the

functional sequences with dependence. For example, Bosq (2002) [7], Dehling and Sharipov (2005) [12] consider the estimation of covariance operator for functional autoregressive processes, Horváth and Kokoszka (2010) [25] studied the covariance structure of weakly dependent functional time series under an m-dependence condition; see also Horváth and Kokoszka (2012) [26] for an overview.

Nevertheless, to obtain a complete description of the second-order structure of dependent functional sequences, one needs to consider autocovariance operators, or autocovariance kernels relating different lags of the series, analogous to the autocovariance matrices in the context of multivariate time series analysis. One statistic of interest associated with the autocovariance operator is the long-run covariance kernel (or long-run covariance function) defined as $C(\tau, \sigma) = \sum_{l \in \mathbb{Z}} r_l(\tau, \sigma)$ where $r_t(\tau, \sigma) = \mathrm{cov}(X_{t+s}(\tau), X_s(\sigma))$ for $\tau, \sigma \in [0, 1]$ and $t, s \in \mathbb{Z}$, is the so-called autocovariance kernel. The analysis of the long-run covariance kernel is applicable to general functional dependence sequences without a particular model assumption.

Horváth et. al (2013) [28] proposed the kernel lag-window estimator of $C(\tau, \sigma)$ and showed its consistency under mild conditions. The asymptotic normality of the estimator is established in Berkes et. al. (2016) [5]. The estimation has applications in mean and stationarity testing of functional time series, see Horváth et. al (2015) [29] and Jirak (2013) [32]. Horváth et. al (2016) [30] and Rice and Shang (2016) [61] address the bandwidth selection for the kernel of the lag-window estimator.

Rather than focus on the isolated characteristic like the long-run covariance, Panaretos and Tavakoli (2013) [43] approach the problem of inferring the second-order structure of stationary functional time series via Fourier analysis, formulating a frequency domain framework for weakly dependent functional data. In the frequency domain of functional setting, the entire second-order dynamical properties are encoded in the spectral density kernel which is defined as

$$f_\omega(\tau,\sigma) = \frac{1}{2\pi}\sum_{t\in\mathbb{Z}} \exp(-i\omega t) r_t(\tau,\sigma). \tag{4.1}$$

where the autocovariance kernel $r_t(\tau,\sigma)$ and the spectral density kernel $f_\omega(\tau,\sigma)$ comprise a Fourier pair. The notion of spectral density kernel in the functional setting is a generalization of finite-dimensional notion in the context of spectral density analysis of multivariate time series, which has been extensively studied by prominent statistical researchers; see, e.g. Parzen (1957, 1961) [45, 46], Brillinger and Rosenblatt (1967) [10], Hannan (1970) [23] and Priestley (1981) [57]. A consistent estimate of the spectral density kernel in the form of a weighted average of the periodogram kernel—the functional analogous of periodogram matrix—is also proposed in Panaretos and Tavakoli (2013) [43] under a type of cumulant mixing condition. This weak dependence condition is the functional analog of classical cummulant-type mixing condition of Brillinger (2001) [9].

In this chapter, we propose a new class of spectral density kernel estimators based on the notion of flat-top kernel defined in Politis (2001) [47]; see also Politis and Romano (1995, 1996, 1999) [53, 54, 55]. The new class of estimators employs

the inverse Fourier transform of a flat-top function to construct the weight function smoothing the periodogram. With the choice of a high-order flat-top kernel, it is shown to be able to achieve bias reduction, and hence the higher-order accuracy in terms of optimizing the integrated mean square error (IMSE). It is also nearly equal to the general lag-window type estimators which is a well-know fact in finite-dimensional case; see Brockwell (2013) [11] and Brillinger and Richard (2001) [9].

The higher-order accuracy of flat-top estimation typically comes at the sacrifice of the positive semi-definite property. To address this issue, we show how a flat-top estimator can be modified to become positive semi-definite (even strictly positive definite) while retaining the favorable asymptotic properties. The modification is similar to the one proposed in Politis (2011) [49], for the treatment of flat-top spectral density matrix estimators. In addition, we introduce a data driven bandwidth selection procedure realized by an automatic inspection of the correlation structure.

The rest of this chapter is organized as follows. In the next section, the flat-top estimator of the spectral density kernel is defined after introduction of some basic definitions of the frequency domain framework, and theorems on the asymptotic accuracy are given. Section 4.3 shows the almost equivalence of the proposed estimator in the form of weighted average of periodogram and the flat-top lag-window estimator. A modification of the flat-top spectral density estimator is introduced in Section 4.4 which results into an estimator that is positive

semi-definite while retaining the estimator's higher-order accuracy. Section 4.5 addresses the issue of data-dependent bandwidth choice where an empirical rule of picking bandwidth is proposed. Our favorable asymptotic results are supported by a finite-sample simulation in Section 4.6 where higher-order accuracy of the flat-top estimators are manifested in practice. Finally, the technical proofs are gathered in the in Section 4.7.

4.2 Spectral density kernel estimation

We consider a functional time series $\{X_t(\tau); \tau \in [0,1]\}_{t\in\mathbb{Z}}$ where each $X_t(\cdot)$ belongs to the separable Hilbert space $L^2([0,1],\mathbb{R})$ possessing mean zero, i.e., $\mathbb{E}X_t(\tau) = 0$ for all $\tau \in [0,1]$, and autocovariance kernel $r_t(\tau,\sigma) = \mathbb{E}X_{t+s}(\tau)X_s(\sigma)$ for $\tau,\sigma \in [0,1]$ and $s \in \mathbb{Z}$.

The space is equipped with the inner product $\langle\cdot,\cdot\rangle$ and the induced L^2 norm $||\cdot||_2$,

$$\langle f,g\rangle = \int_0^1 f(\tau)g(\tau)d\tau, \qquad ||g||_2 = \langle g,g\rangle^{1/2}, \qquad f,g \in L^2([0,1],\mathbb{R}).$$

We assume the series $\{X_t\}_{t\in\mathbb{Z}}$ is strictly stationary in the sense that for any finite set of indices $I \subset \mathbb{Z}$ and any $s \subset \mathbb{Z}$, the joint law of $\{X_t, t \in I\}$ is identical to that of $\{X_{t+s}, t \in I\}$.

In addition, the weak dependence structure among the observations $\{X_t\}$ is quantified by employing the notion of *cumulant kernel* of a series which is intro-

duced in Chapter 1. Recall that the pointwise definition of a kth *order cumulant kernel* is

$$\mathrm{cum}(X_{t_1}(\tau_1),\ldots,X_{t_k}(\tau_k)) = \sum_{\nu=(\nu_1,\ldots,\nu_p)} (-1)^{p-1}(p-1)! \prod_{l=1}^{p} \mathbb{E}\left[\prod_{j\in\nu_l} X_{t_j}(\tau_j)\right]$$

where the sum extends over all unordered partitions of $\{1,\ldots,k\}$. We will make use of the following cumulant mixing condition, defined for fixed $l \geq 0$ and $k = 2,3,\ldots$

Condition C(l, k). For each $j = 1,\ldots,k-1$,

$$\sum_{t_1,\ldots,t_{k-1}=-\infty}^{\infty} (1+|t_j|^l)||\mathrm{cum}(X_{t_1},\ldots,X_{t_{k-1}},X_0)||_2 < \infty$$

We inherit the frequency domain framework of functional time series developed in Panaretos and Tavakoli (2013) [43], in which the functional version of discrete Fourier transform is introduced. Given a functional sequence of length T, $\{X_t\}_{t=0}^{T-1}$, the *functional Discrete Fourier Transform* (fDFT) is defined as

$$\tilde{X}_{\omega}^{(T)}(\tau) = (2\pi T)^{1/2} \sum_{t=0}^{T-1} X_t(\tau)\exp(-i\omega t). \tag{4.2}$$

The tensor products of the fDFT leads to the notion of *periodogram kernel*–the function analogue of the periodogram matrix in the multivariate case. The *periodogram kernel* is defined as

$$p_{\omega}^{(T)}(\tau,\sigma) = \tilde{X}_{\omega}^{(T)}(\tau)\tilde{X}_{-\omega}^{(T)}(\sigma). \tag{4.3}$$

The periodogram kernel is asymptotically unbiased under certain cumulant mixing conditions. However, it is not a consistent estimator of the spectral density kernel

f_ω as its asymptotic covariance is not zero. Panaretos and Tavakoli (2013) [43] proposed a consistent estimator $f_\omega^{(T)}$ by convolving the periodogram kernel with a weight function, which has the form

$$f_\omega^{(T)}(\tau,\sigma) = \frac{2\pi}{T}\sum_{s=1}^{T-1} W^{(T)}\left(\omega - \frac{2\pi s}{T}\right) p_{2\pi s/T}^{(T)}(\tau,\sigma). \tag{4.4}$$

The weight function, $W^{(T)}$, is constructed as

$$W^{(T)}(x) = \sum_{j\in\mathbb{Z}} \frac{1}{B_T} W\left(\frac{x+2\pi j}{B_T}\right) \tag{4.5}$$

where $B_T, T = 1, 2, \ldots$ is a sequence of scale parameters with the properties $B_T > 0, B_T \to 0$ and $B_T T \to \infty$ as $T \to \infty$. W is a fixed function satisfying that W is a positive, even function and

$$W(x) = 0 \text{ for } |x| \geq 1; \quad \int_{-\infty}^{\infty} W(x)dx = 1; \quad \int_{-\infty}^{\infty} W(x)^2 dx < \infty.$$

The summation over $j \in \mathbb{Z}$ in (4.5) makes the weight function $W^{(T)}$ periodic with period 2π. The same will be true for the estimator $f_\omega^{(T)}$ by its definition in (4.4). With the above constraints imposed on function W, it has been shown in Panaretos and Tavakoli (2013) [43] that $f_\omega^{(T)}$ is a consistent estimator of the spectral density kernel f_ω in mean square (with respect to Hilbert-Schmidt norm). The bias of $f_\omega^{(T)}$ is partly attributed to the assumption that W is positive and it can potentially be significantly reduced if an appropriate function W is chosen that is not restricted to be positive. To this aim, we propose a class of higher-order accurate estimator by making use of the so-called flat-top kernels in the construction of weight function

$W^{(T)}$. The resulting estimator is shown to achieve bias reduction while retaining the asymptotic covariance structure of $f_\omega^{(T)}$ in (4.4).

To describe our estimator, we need the notion of a "flat-top" kernel. A general flat-top kernel Λ is defined in terms of its Fourier transform λ, which is in turn defined as

$$\lambda(s) = \begin{cases} 1 & \text{if } |s| \leq c, \\ g(s) & \text{otherwise.} \end{cases} \tag{4.6}$$

where $c > 0$ is a parameter and $g : \mathbb{R} \to [-1, 1]$ is a symmetric function, continuous at all but a finite number of points satisfying $g(c) = 1$ and $\int_{\mathbb{R}} g^2(x)dx < \infty$. The flat-top kernel $\Lambda(x)$ is then given by the inverse Fourier transform of $\lambda(s)$

$$\Lambda(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lambda(s) e^{isx} ds. \tag{4.7}$$

Note that in the preceding definition, the function λ, and hence the kernel Λ, depend on the function g and the parameter c, but this dependence will not be explicitly denoted.

The function $\lambda(s)$ is 'flat', i.e., constant, over the region $[-c, c]$, hence the name flat-top for the kernel function $\Lambda(x)$. If a kernel function Λ has finite qth moment, and its moments up to order $q-1$ are equal to zero, i.e. $\int s^q \Lambda(s)ds < \infty$, and $\int s^k \Lambda(s)ds = 0$ for all $k \leq q-1$, then the kernel is said to be of order q. We have the following property concerning the order of the kernel function Λ:

Proposition 4.2.1. *If $\lambda(s)$ is p times differentiable flat-top function and $\lambda^{(p)}$ is Hölder continuous of order $0 < \alpha < 1$, then $\Lambda(x)$ is a kernel of order $p-1$.*

Proof. See Appendix. □

In the following, we will be using $\hat{f}^{(T)}_{\omega,\lambda}$ to denote the flat-top estimator employing the flat-top kernel Λ, which is in turn induced by a flat-top function λ. The estimator is of the same form as (4.4), except the difference in the weight function, i.e.,

$$\hat{f}^{(T)}_{\omega,\lambda}(\tau,\sigma) = \frac{2\pi}{T}\sum_{s=1}^{T-1} W^{(T)}_{\lambda}\left(\omega - \frac{2\pi s}{T}\right) p^{(T)}_{2\pi s/T}(\tau,\sigma), \tag{4.8}$$

where

$$W^{(T)}_{\lambda}(x) = \sum_{j\in\mathbb{Z}} \frac{1}{B_T}\Lambda\left(\frac{x+2\pi j}{B_T}\right), \tag{4.9}$$

with $\Lambda(x)$ being the flat-top kernel induced by a flat-top function $\lambda(s)$ defined in (4.7).

The following theorems investigate the performance of $\hat{f}^{(T)}_{\omega,\lambda}$ employing the general flat-top kernel Λ.

Theorem 4.2.1. *Provided that $B_T \to 0$ and $B_T T \to \infty$ as $T \to \infty$, and assume p is the maximum value that can be attained such that C(p,2) holds; then by choosing an appropriate flat-top kernel Λ of order p, we have*

$$\mathbb{E}[\hat{f}^{(T)}_{\omega,\lambda}(\tau,\sigma)] = f_{\omega}(\tau,\sigma)_{\omega} + O(B_T^p) + O(B_T^{-1}T^{-1}),$$

where the equality holds in L^2, and the error terms are uniform in ω.

Proof. See Appendix. □

Remark 4.2.2. According to Proposition 4.2.1, a sufficient condition for a kernel Λ to be of order p is that λ is a $p+1$ times differentiable flat-top function and $\lambda^{(p+1)}$ is Hölder continuous. On the other hand, the decay rate of bias crucially depends on the cumulant mixing condition satisfied by the functional sequence. A type of moment conditions is provided in Panaretos and Tavakoli (2013) [43], which are sufficient for the cumulant mixing condition to hold for a general linear process of the form $X_t = \sum_{s\in\mathbb{Z}} A_s \varepsilon_{t-s}$; see Proposition 4.1 therein.

The flat-top estimator $\hat{f}^{(T)}_{\omega,\lambda}$ achieves bias improvements while retaining the rate of decay of the covariance structure as stated in the following theorem:

Theorem 4.2.3. *Under C(1,2) and C(1,4),*

$$cov(\hat{f}^{(T)}_{\omega_1,\lambda}(\tau_1,\sigma_1), \hat{f}^{(T)}_{\omega_2,\lambda}(\tau_2,\sigma_2)) = O(B_T^{-2}T^{-1})$$

where the equality holds in L^2, uniformly in the ω's.

Using our Lemma 4.7.2 and 4.7.3, Theorem 4.2.3 can be proved along the same lines of proof of Corollary 3.3 in Panaretos and Tavakoli (2013) [43]. For fixed ω_1, ω_2, the covariance can be shown to have a sharper bound $O(B_T^{-1}T^{-1})$; see Proposition 3.4 in Panaretos and Tavakoli (2013) [43].

Concerning the mean square error, we need the notion of *spectral density operator* of functional time series, which is introduced in Panaretos and Tavakoli (2013) [43]. The spectral density operator $\mathscr{F}_\omega$ is an operator induced by the

spectral density kernel through right integration,

$$\begin{aligned}\mathscr{F}_\omega h(\tau) &= \int_0^1 f_\omega(\tau,\sigma)h(\sigma)d\sigma \\ &= \frac{1}{2\pi}\sum_{t\in\mathbb{Z}} e^{-i\omega t}\int_0^1 r_t(\tau,\sigma)h(\sigma)d\sigma \\ &= \frac{1}{2\pi}\sum_{t\in\mathbb{Z}} e^{-i\omega t}\mathscr{R}_t h(\tau), \end{aligned} \tag{4.10}$$

where $\mathscr{R}_t$ is the *autocovariance operator* induced by the autocovariance kernel through right integration,

$$\mathscr{R}_t h(\tau) = \int_0^1 r_t(\tau,\sigma)h(\sigma)d\sigma = \mathrm{cov}[\langle X_0, h\rangle, X_t(\tau)], \qquad h \in L^2([0,1],\mathbb{R}). \tag{4.11}$$

The spectral density operator $\mathscr{F}_\omega$ is the integral operator with kernel f_ω. Analogously, we denote $\hat{\mathscr{F}}^{(T)}_{\omega,\lambda}$ the operator induced by the the kernel $\hat{f}^{(T)}_{\omega,\lambda}$ through right integration, and thereby the estimator of $\mathscr{F}_\omega$. Combining the results on the asymptotic bias and variance of the spectral density operator, we have the following consistency in *integrated mean square* of the induced estimator $\hat{\mathscr{F}}^{(T)}_{\omega,\lambda}$ for the spectral density operator $\mathscr{F}_\omega$.

Theorem 4.2.4. *Provided assumptions C(p,2) and C(1,4) hold, $B_T \to 0$, $B_T T \to \infty$, then the spectral density operator estimator $\hat{\mathscr{F}}^{(T)}_{\omega,\lambda}$ employing a flat-top kernel Λ of order p is consistent in integrated mean square, that is,*

$$IMSE(\hat{\mathscr{F}}^{(T)}_{\omega,\lambda}) = \int_{-\pi}^{\pi} \mathbb{E}\left|\left|\left|\hat{\mathscr{F}}^{(T)}_{\omega,\lambda} - \mathscr{F}_\omega\right|\right|\right|_2^2 d\omega \to 0, \qquad as\ T\to\infty,$$

where $|||\cdot|||$ is the Hilbert-Schmidt norm. More precisely, $IMSE(\hat{\mathscr{F}}^{(T)}_{\omega,\lambda}) = O(B_T^{2p}) + O(B_T^{-1}T^{-1})$ as $T\to\infty$.

Theorem 4.2.4 gives the rate of convergence of $\hat{\mathscr{F}}_{\omega,\lambda}^{(T)}$ to $\mathscr{F}_\omega$. In the meantime, it also suggests the optimal value of the bandwidth parameter B_T in terms of optimizing the decay rate of integrated mean square error. Apparently, the optimal B_T depends on the cumulant condition a functional sequence possesses, that is, the value of p. For any finite p, the optimal decay rate $O(T^{-2p/(2p+1)})$ can be achieved with $B_T = T^{-1/(2p+1)}$. In the case that $p = \infty$, one can choose $B_T = 1/\log T$ to obtain a favorable rate of $O(\log T/T)$.

4.3 Alternate estimates and flat-top kernel choice

4.3.1 Alternate estimates

The spectral density kernel estimator considered in the previous section has the form of a weighted average of periodogram ordinates. In fact, the weight function $W_\lambda^{(T)}(x)$ in the estimate (4.8) has an alternate form

$$W_\lambda^{(T)}(x) = \frac{1}{2\pi}\sum_{u\in\mathbb{Z}} \lambda(B_T u)e^{-ixu}. \tag{4.12}$$

The equivalence of (4.9) and (4.12) can be easily verified by using Poisson summation formula. Moreover, if the discrete average in (4.8) is replaced by a continuous one, the estimate becomes

$$\begin{aligned}\int_0^{2\pi} W_\lambda^{(T)}(\omega-\alpha)p_\alpha^{(T)}(\tau,\sigma)d\alpha &= \int_0^{2\pi} W_\lambda^{(T)}(\alpha)p_{\omega-\alpha}^{(T)}(\tau,\sigma)d\alpha \\ &= \int_{-\infty}^{\infty} B_T^{-1}\Lambda(B_T^{-1}\alpha)p_{\omega-\alpha}^{(T)}(\tau,\sigma)d\alpha. \end{aligned}\tag{4.13}$$

By Equation (4.2) and (4.3), the periodogram kernel is given by

$$p_{\omega}^{(T)}(\tau,\sigma) = \frac{1}{2\pi}\sum_{|u|<T} \hat{r}_u(\tau,\sigma)e^{-i\omega u} \tag{4.14}$$

where

$$\hat{r}_u(\tau,\sigma) = \frac{1}{T}\sum_{0\le t,t+u\le T-1} X_{t+u}(\tau)X_t(\sigma)$$

is the sample autocovariance kernel. If this is substituted into (4.13), then the estimate takes the form

$$\frac{1}{2\pi}\sum_{|u|<T} \lambda(B_T u)\hat{r}_u(\tau,\sigma)e^{-i\omega u} \tag{4.15}$$

where

$$\lambda(s) = \int_{-\infty}^{\infty} \Lambda(x)e^{-isx}dx.$$

With a flat-top function λ in place, the estimate (4.15) has a formal resemblance to the flat-top estimation for multivariate spectrum explored in Politis (2011) [49] with the bandwidth parameter $m_T = B_T^{-1}$. The estimator has been shown to achieve higher-order accuracy in estimating the spectral density matrix; see Politis (2011) [49] for details. In fact, the estimate (4.15) is the general form of spectral estimation that has been extensively investigated by prominent statistical researchers as early as 1950s and 1960s. See, e.g., Grenander (1951) [22], Parzen (1957) [45] and Priestley (1962) [56].

4.3.2 Flat-top kernel choice

As suggested by Theorem 4.2.1, in order to achieve favorable asymptotic rates, it is desirable to choose a flat-top kernel Λ of higher order, and hence a flat-top function λ as smooth as possible according to Proposition 4.2.1. McMurry and Politis (2004) [40] constructed a member of the flat-top family that is infinitely differentiable, which is defined as

$$\lambda_{ID,b,c}(s) = \begin{cases} 1 & \text{if } |s| \leq c, \\ \exp(-b\exp(-b/(|s|-c)^2)/(|s|-1)^2) & \text{if } c < |s| < 1, \\ 0 & \text{if } |s| \geq 1 \end{cases} \tag{4.16}$$

where $c \in (0, 1]$ determines the region over which λ is identically 1, and $b > 0$ is a shape parameter, making the transition from $\lambda_{ID,b,c}(c) = 1$ to $\lambda_{ID,b,c}(1) = 0$ more or less abrupt.

The function $\exp(-b\exp(-b/(|s|-c)^2)/(|s|-1)^2)$ connects the regions where λ is 0 and the region where λ is 1 in a manner such that $\lambda(s)$ is infinitely differentiable for all s, including where $|s| = c$ and $|s| = 1$. The resulting kernel Λ is of infinite order in the sense that $\Lambda(x)$ decays faster than $|x|^{-m}$, for all positive finite m, as $|x| \to \infty$. Figure 4.1 shows the plots of the infinitely differentiable flat-top function $\lambda_{ID,b,c}(s)$ with $b = 0.25$ and $c = 0.05$, and the resulting kernel $\Lambda(x)$ as well as the corresponding weight function $W_\lambda^{(T)}(x)$. Note that the plot of $W_\lambda^{(T)}(x)$ can be created from either Equation (4.9) or (4.12) as their equivalence stated in Section 4.3.1.

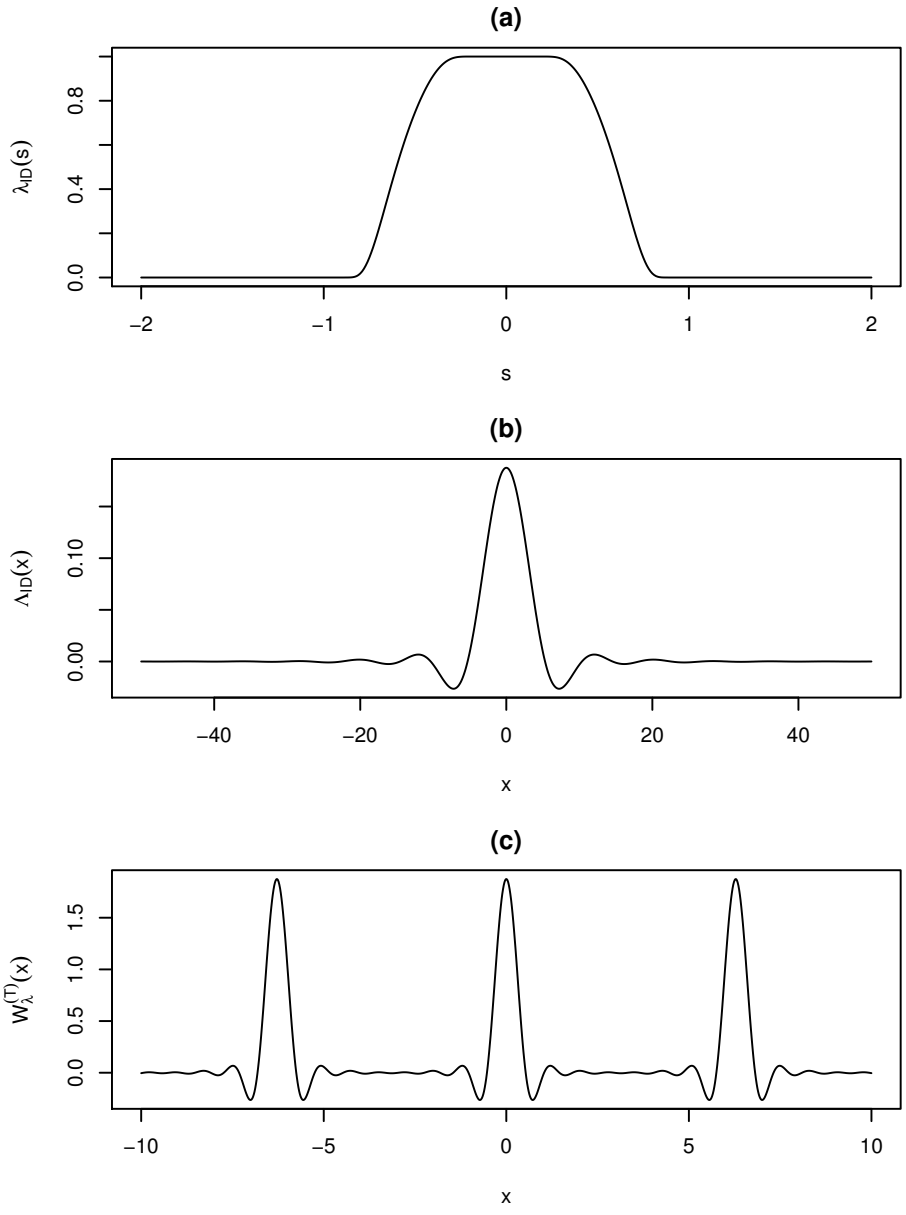

Figure 4.1: (a) Plot of $\lambda_{ID,1/4,0.05}(s)$; (b) Plot of corresponding kernel $\Lambda(x)$ induced by inverse Fourier transform of $\lambda_{ID,1/4,0.05}(s)$; (c) Plot of the corresponding weight function $W_\lambda^{(T)}$ with $B_T = 0.1$.

Nevertheless, while the effectiveness of the flat-top kernels is reflected in Theorem 4.2.1 and 4.2.4, they in fact provide merely theoretical bounds for the decay rate of bias and IMSE. In the meantime, according to Theorem 4.2.1, the reduction of bias of the flat-top estimation could potentially be limited by the order of the cumulant condition, which indicates that an infinite order kernel might not be necessary. That leads us to attempt other choices within the flat-top family. One simple representative flat-top function has the trapezoidal shape defined as

$$\lambda_{TR,c}(s) = \begin{cases} 1 & \text{if } |s| \leq c, \\ \dfrac{|s|-1}{c-1} & \text{if } c < |s| < 1, \\ 0 & \text{if } |s| \geq 1. \end{cases} \tag{4.17}$$

The trapezoidal $\lambda_{TR,c}$ is continuous everywhere and it already exhibits good performance when being implemented for the estimation of spectral density matrix; see Politis (2011) [49]. The infinite differentiable function $\lambda_{ID,b,c}(s)$ looks very much like the trapezoidal, but with smoothed corners.

Another choice to be considered is the flat-top function created by adding a piecewise cubic tail, similar to that of Parzen's (1961) [46] kernel, to the $[-c, c]$

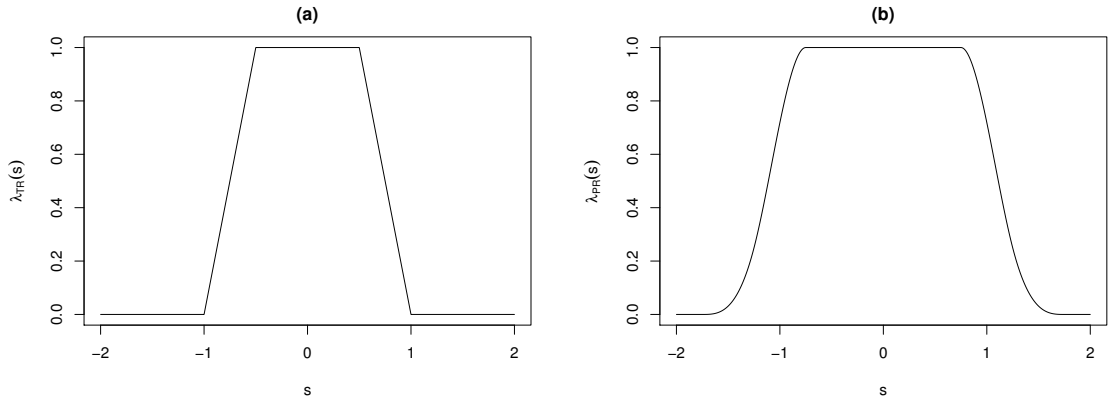

Figure 4.2: (a) Plot of trapezoidal $\lambda_{TR,1/2}(s)$; (b) Plot of flat-top Parzen $\lambda_{PR,3/4}(s)$.

flat-top region. It is defined as

$$\lambda_{PR,c}(s) = \begin{cases} 1 & \text{if } 0 \le s \le c, \\ 1 - 6(s-c)^2 + 6|s-c|^3 & \text{if } c \le s \le c + 1/2, \\ 2(1-|s-c|)^3 & \text{if } c + 1/2 \le s \le c + 1 \\ 0 & \text{if } s \ge c + 1, \\ \lambda_{PR,c}(-s) & \text{if } s < 0. \end{cases} \tag{4.18}$$

Plots of flat-top functions $\lambda_{TR,1/2}$ and $\lambda_{PR,3/4}$ are shown in Figure 4.2. Concerning the choice of parameters of flat-top kernels, i.e., b and c, we refer the readers to Politis (2011) [49] where a detailed discussion is given.

4.4 Positive semi-definite spectral estimation

By employing the infinite-order flat-top kernels, the flat-top estimator $\hat{\mathscr{F}}_{\omega,\lambda}^{(T)}$ is capable of achieving higher-order accuracy with improved estimation bias. The

disadvantage of flat-top kernels, however, is that they are not positive semi-definite. As a result, the operator estimation $\hat{\mathscr{F}}^{(T)}_{\omega,\lambda}$ is not almost surely positive semi-definite for all ω, while it converges to a positive semi-definite operator $\mathscr{F}_\omega$.

The positive semi-definiteness of the estimation is desirable especially in the case of $\omega = 0$ when the object is estimation of a long-run covariance operator. In the context of finite-dimensional time series analysis, the spectral density matrix estimators can be easily adjusted to be positive semi-definite via replacing negative eigenvalues by zeros in the diagonalization of the estimated matrices; see e.g. Politis (2011) [49]. Analogously, we now show how the flat-top operator estimator $\hat{\mathscr{F}}^{(T)}_{\omega,\lambda}$ can be modified to render a positive semi-definite estimator while preserving the asymptotic consistency.

The spectral decomposition of operators in an infinite-dimensional Hilbert space is much more intricate than that in a finite-dimensional context. However, recall that both operators $\mathscr{F}_\omega$ and $\hat{\mathscr{F}}^{(T)}_{\omega,\lambda}$ are induced by kernel functions through right integration, and therefore they are symmetric Hilbert-Schmidt operators that admit the following decompositions

$$\mathscr{F}_\omega(h) = \sum_{j=1}^{\infty} \nu_j \langle h, e_j \rangle e_j, \qquad h \in L^2([0,1], \mathbb{R}) \tag{4.19}$$

$$\hat{\mathscr{F}}^{(T)}_{\omega,\lambda}(h) = \sum_{j=1}^{\infty} \hat{\nu}_j \langle h, \hat{e}_j \rangle \hat{e}_j, \qquad h \in L^2([0,1], \mathbb{R}) \tag{4.20}$$

where (ν_j) and $(\hat{\nu}_j)$ are two sequences of real numbers tending to zero; (e_j) and

$(\hat{e}_j)$ are two orthonormal bases of $L^2([0,1],\mathbb{R})$. We have for $j \geq 1$,

$$\mathscr{F}_\omega(e_j) = \nu_j e_j \quad \text{and} \quad \hat{\mathscr{F}}^{(T)}_{\omega,\lambda}(\hat{e}_j) = \hat{\nu}_j \hat{e}_j;$$

thus (ν_j, e_j) and $(\hat{\nu}_j, \hat{e}_j)$, $j \geq 1$ are complete sequences of eigenelements of $\mathscr{F}_\omega$ and $\hat{\mathscr{F}}^{(T)}_{\omega,\lambda}$ respectively.

Noting that the eigenvalues $\nu_j,\ j \geq 1$ are all non-negative since the operator $\mathscr{F}_\omega$ is positive semi-definite. To fix the possible negativity of $\hat{\mathscr{F}}^{(T)}_{\omega,\lambda}$, let $\tilde{\nu}_j = \max(\hat{\nu}_j, 0)$ for all j, and define the estimator

$$\tilde{\mathscr{F}}^{(T)}_{\omega,\lambda}(h) = \sum_{j=1}^{\infty} \tilde{\nu}_j \langle h, \hat{e}_j \rangle \hat{e}_j, \qquad h \in L^2([0,1],\mathbb{R}). \tag{4.21}$$

We keep nonnegative eigenvalues of $\hat{\mathscr{F}}^{(T)}_{\omega,\lambda}$ and replace negative eigenvalues by zero, which makes the resulting operator $\tilde{\mathscr{F}}^{(T)}_{\omega,\lambda}$ an positive semi-definite estimator. The connection of $\hat{\mathscr{F}}^{(T)}_{\omega,\lambda}$ and $\tilde{\mathscr{F}}^{(T)}_{\omega,\lambda}$ is shown in the following inequality:

Proposition 4.4.1. *Let $\tilde{\mathscr{F}}^{(T)}_{\omega,\lambda}$ be the positive semi-definite operator estimator of $\mathscr{F}_\omega$ defined in (4.10), then for a fixed ω*

$$\left|\!\left|\!\left| \tilde{\mathscr{F}}^{(T)}_{\omega,\lambda} - \mathscr{F}_\omega \right|\!\right|\!\right|_2 \leq \left|\!\left|\!\left| \hat{\mathscr{F}}^{(T)}_{\omega,\lambda} - \mathscr{F}_\omega \right|\!\right|\!\right|_2, \tag{4.22}$$

where $|||\cdot|||_2$ is the Hilbert-Schmidt norm.

Proof. See Appendix. □

A direct consequence of the last result is the following corollary which shows that, in addition to being positive semi-definite, $\tilde{\mathscr{F}}^{(T)}_{\omega,\lambda}$ possesses the same mean square convergence of $\hat{\mathscr{F}}^{(T)}_{\omega,\lambda}$ given in Theorem 4.2.4.

Theorem 4.4.1. *Under the condition of Theorem 4.2.4, the positive semi-definite spectral density operator estimate $\tilde{\mathscr{F}}_{\omega,\lambda}^{(T)}$ employing a flat-top kernel Λ of order p is consistent in integrated mean square with*

$$IMSE(\tilde{\mathscr{F}}_{\omega,\lambda}^{(T)}) = \int_{-\pi}^{\pi} \mathbb{E}\left|\left|\left|\tilde{\mathscr{F}}_{\omega,\lambda}^{(T)} - \mathscr{F}_{\omega}\right|\right|\right|_2^2 d\omega = O(B_T^{2p}) + O(B_T^{-1}T^{-1})$$

where $|||\cdot|||_2$ is the Hilbert-Schmidt norm.

In the case that the estimand $\mathscr{F}_{\omega}$ is not only positives semi-definite but strictly positive definite, it is desirable to have a strictly positive definite estimator of $\mathscr{F}_{\omega}$. A similar modification of $\hat{\mathscr{F}}_{\omega,\lambda}^{(T)}$ can be applied here to make the estimator strictly positive definite. Let $\check{\nu}_j = \max(\check{\nu}_j, \epsilon_T)$ for all j, where $\epsilon_T > 0$ is some chose sequence, and define the estimator

$$\check{\mathscr{F}}_{\omega,\lambda}^{(T)}(h) = \sum_{j=1}^{\infty} \check{\nu}_j \langle h, \hat{e}_j \rangle \hat{e}_j, \qquad h \in L^2([0,1], \mathbb{R}). \tag{4.23}$$

The estimator $\check{\mathscr{F}}_{\omega,\lambda}^{(T)}$ is positive definite and it can be verified that it maintains the high accuracy of the flat-top estimator if $\epsilon_T = O(1/T)$. Thus, $\check{\mathscr{F}}_{\omega,\lambda}^{(T)}$ is a higher-order accurate, strictly positive definite estimator.

4.5 Data-dependent bandwidth choice

As it has been demonstrated in Section 4.3.1 that the lag-window estimate (4.15) with bandwidth m_T is nearly equal to the estimate (4.8) with bandwidths $B_T = m_T^{-1}$, we propose here an empirical rule for choosing the bandwidth B_T in

practice, which resembles the bandwidth choosing rule for the flat-top lag-window introduced in Politis (2011) [49].

Recall that the sample autocovariance kernel

$$\hat{r}_u(\tau,\sigma) = \frac{1}{T}\sum_{0\le t,t+u\le T-1} X_{t+u}(\tau)X_t(\sigma),$$

the proposed bandwidth choice rule is done by a simple inspection of the functional version of correlogram/cross-corregram, i.e. a plot of $\hat{\rho}_m(\tau,\sigma)$ vs. m where

$$\hat{\rho}_m(\tau,\sigma) = \frac{\hat{r}_m(\tau,\sigma)}{\sqrt{\hat{r}_0(\tau,\tau)\hat{r}_0(\sigma,\sigma)}}$$

for all $\tau,\sigma \in [0,1]$.

We look for a point, say $\hat{q}$, after which the correlogram for each pair of (τ,σ) appears negligible, i.e. $\hat{\rho}_m(\tau,\sigma) \simeq 0$ for $|m| > \hat{q}$, and $\hat{\rho}_{\hat{q}}(\tau,\sigma) \neq 0$. Here $\hat{\rho}_m(\tau,\sigma) \simeq 0$ is taken to mean that $\hat{\rho}_m(\tau,\sigma)$ is not taken significantly different from 0. In practice, we determine $\hat{q}$ by considering the correlogram for (τ,σ) over a finite grid of $[0,1]\times[0,1]$. After identifying $\hat{q}$, the recommendation is to take

$$\hat{B}_T = \frac{1}{\max(\lceil \hat{q}/c \rceil, 1)} \tag{4.24}$$

where c is the parameter determines the 'flat-top' region of λ.

From the flat-top lag-window perspective, the intuition behind the above bandwidth choice rule is an effort to extend the 'flat-top' region of λ over the whole of the region where $\hat{\rho}_{\hat{q}}(\tau,\sigma)$ is thought to be significant so as not to downweigh it and introduce bias. As scrutinized in Politis (2011) [49], the 'flat-top' region of λ

can be greater than $[-c, c]$ depending on the choice of function g. The decreasing rate of $g(s)$ near c could be slow enough so that $\lambda(s) \simeq 1$ for an interval much greater than $[-c, c]$; see, for example, (4.16) and Figure 1(a) regarding the infinite differentiable $\lambda_{ID,b,c}(s)$ with $b = 0.25$ and $c = 0.05$. Instead of the interval $[-c, c]$, we consider an 'effective' flat-top region of λ defined as the interval $[-c_{ef}, c_{ef}]$ where c_{ef} is the largest number such that $\lambda(s) \geq 1 - \epsilon$ for all x in $[-c_{ef}, c_{ef}]$; here ϵ is some small number chosen number, e.g. $\epsilon = 0.01$.

Let $\Gamma = \{(i/10, j/10); i, j = 0, \ldots, 9)\}$ be a finite grid of $[0, 1]^2$. Now we can formalize the empirical rule of choosing bandwidth B_T.

EMPIRICAL RULE OF CHOOSING BANDWIDTH B_T.

For $(\tau, \sigma) \in \Gamma$, let $\hat{q}_{\tau,\sigma}$ be the smallest nonnegative integer such that $|\hat{\rho}_{m+\hat{q}_{\tau,\sigma}}(\tau, \sigma)| < C_0\sqrt{\log_{10} T/T}$, for $m = 0, 1, \ldots, K_T$, where $C_0 > 0$ is a fixed constant, and K_T is a positive, nondecreasing integer-valued function of T such that $K_T = o(\log T)$. Then, let $\hat{q} = \max_{(\tau,\sigma)\in\Gamma} \hat{q}_{\tau,\sigma}$, and $B_T = 1/\max(\lceil \hat{q}/c_{ef} \rceil, 1)$.

The constant C_0 and the form of K_T are the practitioner's choice. Politis (2003) [48] makes the concrete recommendations

$$C_0 \simeq 2 \text{ and } K_T = \max(5, \sqrt{\log_{10} T})$$

that have the interpretation of yielding (approximately) 95% simultaneous confidence intervals for $\hat{\rho}_{m+\hat{q}_{\tau,\sigma}}(\tau, \sigma)$ with $m = 1, \ldots, K_T$ by Bonferroni's inequality.

It is also worth noting that by considering the correlogram over the finite grid Γ, we actually generate a matrix of thresholds and, $\hat{q}$ is picked as the maximum among the entries of the matrix, i.e, $\hat{q}_{\tau,\sigma}$ for $(\tau,\sigma) \in \Gamma$. This rule of identifying $\hat{p}$ can be blemished in the situation that certain $\hat{q}_{\tau,\sigma}$ are radically greater compared to the others. Picking $\hat{p}$ to be the average of the matrix entries will be a more reasonable choice when such a special case arises. Nevertheless, if the target is to estimate the spectral kernel f_ω for a particular pair (τ,σ), one can always choose the bandwidth B_T by using the specified $\hat{q}_{\tau,\sigma}$, i.e. $B_T = 1/\max(\lceil \hat{q}_{\tau,\sigma}/c_{ef} \rceil, 1)$.

4.6 Simulations

We now present some numerical simulations to complement our asymptotic results. The main goal of the simulations is to compare the performance of the estimators employing flat-top kernels with that of the non-flat-top estimation, as well as to illustrate the main issues discussed in the paper. The simulations are performed on a simple functional moving average model

$$X_t = A_0\varepsilon_t + A_1\varepsilon_{t-1}. \tag{4.25}$$

The simulations we carry out are analogous to that conducted in Panaretos and Tavakoli (2013) [43]. The innovation functions ε_t's are independent Wiener processes on $[0,1]$, which are represented using a truncated Karhuen-Loève expansion,

$$\varepsilon_t(\tau) = \sum_{k=1}^{1000} \xi_{k,t}\sqrt{\eta_k}e_k(\tau),$$

where $\eta_k = 1/[(k-1/2)^2\pi^2]$, $\xi_{k,t}$ are independent standard Gaussian random variables and $e_k(\tau) = \sqrt{2}\sin[(k-1/2)\pi\tau]$ is orthonormal system in $L^2([0,1],\mathbb{R})$; see Adler (1990) [1]. The operators A_0 and A_1 are constructed so that their image be contained within a 50-dimensional subspace of $L^2([0,1],\mathbb{R})$, spanned by an orthonormal basis $\psi_1,\ldots,\psi_{50}$. Representing ε_t in the e_k basis, and $A_s, s = 1,2$ in the $\psi_m \bigotimes e_k$ basis, we have the matrix representation of the process X_t as $\boldsymbol{X}_t = \boldsymbol{A}_0\boldsymbol{\varepsilon}_t + \boldsymbol{A}_1\boldsymbol{\varepsilon}_{t-1}$, where $\boldsymbol{X}_t$ is a 50×1 matrix, each $\boldsymbol{A}_s$ is a 50×100 matrix, and each $\boldsymbol{\varepsilon}_t$ is a 100×1 matrix.

A stretch of $X_t, t = 0,\ldots,T-1$ is generated for $T = 2^n$ with $n = 6,\ldots,10$. Matrices $\boldsymbol{A}_s, s = 1,2$ are constructed as random Gaussian matrices with independent entries, such that element in jth row are $N(0, j^{-2})$ distributed.

For the simulation, $B = 200$ simulation runs are generated for each T which are used to compute the IMSE by approximating the integral

$$2\int_0^\pi \mathbb{E}\left|\left|\left|\hat{\mathscr{F}}^{(T)}_{\omega,\lambda} - \mathscr{F}_\omega\right|\right|\right|_2^2 d\omega$$

by a weighted sum over the finite grid $\Gamma = \{\pi j/10; j = 0,\ldots,9\}$. We consider the estimators with proposed flat-top kernels and compare them with the *Epanechnikov kernel*, $W(x) = \frac{3}{4}(1-x^2)^+$, which is non-flat-top implemented in the simulations of Panaretos and Tavakoli (2013) [43]. We apply bandwidth $B_T = T^{-1/5}$ for the estimator of each kernel. In addition, the bandwidths of the estimators employing flat-top kernels are also estimated using the empirical rule proposed in Section 4.5.

The simulation results are presented in Table 4.1, entries of which are log-

Table 4.1: Entries represent the logarithm of IMSEs in base 2 of different estimators using (a) bandwidth $B_T = T^{-1/5}$ and (b) Empirical rule of choosing B_T. Sample size ranges from 2^6 to 2^{10}. Minimum IMSE for each T is indicated by boldface.

T	64	128	256	512	1024
Epanechnikov kernel	-6.188	-6.852	-7.588	-8.276	-9.082
$\lambda_{TR,1/2}$ (Trapezoid)	**-6.389**	-7.112	-7.902	-8.719	-9.493
$\lambda_{PR,3/4}$ (Flat-top Parzen)	-6.383	-7.041	-8.018	**-8.846**	-9.710
$\lambda_{ID,1/4,0.05}$ (Flat-top Inf. Diff.)	-6.344	**-7.262**	**-8.074**	-8.832	**-9.719**

(a) Bandwidth $B_T = T^{-1/5}$

T	64	128	256	512	1024
$\lambda_{TR,1/2}$ (Trapezoid)	-6.519	-7.331	-8.260	-9.145	-10.089
$\lambda_{PR,3/4}$ (Flat-top Parzen)	-6.627	-7.592	-8.455	**-9.349**	**-10.313**
$\lambda_{ID,1/4,0.05}$ (Flat-top Inf. Diff.)	**-6.701**	**-7.640**	**-8.538**	-9.327	-10.280

(b) Empirical rule of choosing B_T

arithm of IMSE in base 2. As expected, the estimators employing flat-top kernels show a faster decay rate of IMSE compared to the one with the non-flat-top *Epanechnikov kernel*. The performances of flat-top Parzen and flat-top indefinite differentiable kernels are close, while they slightly outperform the trapezoid as the sample size grows. This might be due to the fact that the smoothness of the flat-top functions is indeed a factor on the decay of IMSE, but over-smoothing might not be necessary as the performance could potentially be limited by the order of cumulant conditions as Theorem 4.2.4 suggests. Also note that implementing the empirical rule of bandwidth choice indeed yields an overall improvement by the comparison of results between Table 4.1(a) and 4.1(b).

4.7 Appendix: Proofs

4.7.1 Proof of Proposition 4.2.1

To prove Proposition 4.2.1, we need the following lemma, which measures the magnitude of Fourier coefficient.

Lemma 4.7.1. *Let f be an integrable function on the interval $[0, 2\pi]$, and $\hat{f}(n)$ be its Fourier coefficients defined by*

$$\hat{f}(n) = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{int} dt. \tag{4.26}$$

If f is p times differentiable and $f^{(p)}$ is Hölder continuous of order $o < \alpha < 1$, then

$$|\hat{f}(n)| = O(|n|^{-p-\alpha}).$$

Proof. By repeated integration by part on Equation (4.26), we have

$$\hat{f}(n) = (in)^{-p} \widehat{f^{(p)}}(n). \tag{4.27}$$

On the other hand, $\widehat{f^{(p)}}(n) = \frac{1}{2\pi} \int_0^{2\pi} f^{(p)}(n) e^{-int} dt = \frac{-1}{2\pi} \int_0^{2\pi} f^{(p)}(n) e^{-in(t+\pi/n)} dt$; by a change of variable, $\widehat{f^{(p)}}(n)$ can be written as

$$\widehat{f^{(p)}}(n) = \frac{1}{4\pi} \int_0^{2\pi} \left(f^{(p)}(t + \frac{\pi}{n}) - f^{(p)}(t) \right) e^{-int} dt.$$

By Hölder continuity of $f^{(p)}$, we have

$$|\widehat{f^{(p)}}(n)| \leq \frac{C}{|n|^{\alpha}}, \tag{4.28}$$

for some constant C. Combining (4.27) and (4.28), we obtain

$$|\hat{f}(n)| = |n|^{-p}|\widehat{f^{(p)}}(n)| = O(|n|^{-p-\alpha}). \tag{4.29}$$

□

Proof of Proposition 4.2.1. For a flat-top function $\lambda(s)$ and its inverse Fourier transform $\Lambda(x)$, we have

$$\Lambda(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \lambda(s)e^{isx}ds, \tag{4.30}$$

$$\lambda(s) = \int_{\mathbb{R}} \Lambda(x)e^{-isx}dx. \tag{4.31}$$

Since the assumption that λ is p times differentiable and $\lambda^{(p)}$ is Hölder continuous of order $0 < \alpha < 1$, by Lemma 4.7.1, we have $|x|^{p+\alpha}|\Lambda(x)| \leq C$ for some constant C, which implies $\Lambda(x)$ has finite moments up to order $p-1$, i.e. $\int_{\mathbb{R}} |x|^k|\Lambda(x)|dx < \infty$ for $0 \leq k \leq p-1$.

By repeated differentiations on both sides of (4.31), we obtain for $k = 1, \ldots, p-2$,

$$\frac{d^k\lambda(s)}{ds^k} = \int_{\mathbb{R}} (-ix)^k\Lambda(x)e^{-isx}dx$$

by dominated convergence theorem. Now that $\lambda(x)$ is flat-top, $\lambda^{(k)}(0)$ is zero for all k, which in turn leads to

$$\int_{\mathbb{R}} x^k\Lambda(x)dx = 0 \text{ for } k = 1, \ldots, p-2 \tag{4.32}$$

if we set $s = 0$ on both sides of (4.32). Therefore, Λ is a kernel of order $p-1$.

4.7.2 Proof of Theorem 4.2.1

To prove Theorem 4.2.1, the following lemmas are necessary.

Lemma 4.7.2. *We have the following properties for the flap-top kernel* $\Lambda(x)$ *and the function* $W_\lambda^{(T)}(x)$*:*

(i) $\int_{\mathbb{R}} \Lambda(x)dx = 1$;

(ii) $\int_{-\pi}^{\pi} W_\lambda^{(T)}(x)dx = 1$;

Let $||f||_\infty = sup_{x\in[a,b]}|f(x)|$*, and denoted by* $V_a^b(h)$ *the total variation of a function* $h : [a, b] \to \mathbb{C}$.

(iii) If $B_T < 1$, $||W_\lambda^{(T)}||_\infty = \frac{1}{B_T}||\Lambda||_\infty + O(B_T)$;

(iv) If $B_T < 1$, $V_{-\pi}^{\pi}(W_\lambda^{(T)}) \leq \frac{1}{B_T} V_{-\pi}^{\pi}(\Lambda)$.

Proof. The statement (i) follows directly from setting $s = 0$ on both sides of Equation (4.31). The statement (ii) is obtained by following the same arguments in the proof of Lemma F.11 in Panaretos and Tavakoli (2013) [43]. For the third statement, recall that

$$W_\lambda^{(T)}(x) = \sum_{j\in\mathbb{Z}} \frac{1}{B_T}\Lambda\left(\frac{x+2\pi j}{B_T}\right).$$

If $B_T < 1$, then for $x \in [-\pi, \pi]$,

$$\begin{aligned}|W_\lambda^{(T)}(x)| &\leq \frac{1}{B_T}\sum_{j\in\mathbb{Z}}\left|\Lambda\left(\frac{x+2\pi j}{B_T}\right)\right| \\ &= \frac{1}{B_T}\left|\Lambda\left(\frac{x}{B_T}\right)\right| + \frac{1}{B_T}\sum_{j\in\mathbb{Z}^+}\left|\Lambda\left(\frac{x+2\pi j}{B_T}\right)\right| + \frac{1}{B_T}\sum_{j\in\mathbb{Z}^-}\left|\Lambda\left(\frac{x+2\pi j}{B_T}\right)\right|.\end{aligned}$$

Consider the second term, by the fact that $|\Lambda(x)| \leq \frac{C}{|x|^{1+\alpha}}$ for large x, we obtain

$$\begin{aligned}\frac{1}{B_T}\sum_{j\in\mathbb{Z}^+}\left|\Lambda\left(\frac{x+2\pi j}{B_T}\right)\right| &\leq \frac{1}{B_T}\sum_{j=1}^{\infty}\frac{C}{\left(\frac{x+2\pi j}{B_T}\right)^{1+\alpha}} \\ &= B_T\sum_{j=1}^{\infty}\frac{C}{(x+2\pi j)^{1+\alpha}} \\ &= O(B_T).\end{aligned}$$

Similarly, the third term above is $O(B_T)$. Therefore, we have

$$|W_\lambda^{(T)}(x)| \leq \frac{1}{B_T}|\Lambda(x/B_T)| + O(B_T),$$

and the statement (iii) then follows from the periodicity of $W_\lambda^{(T)}$. For the last statement, since

$$W_\lambda^{(T)}(x) = \frac{1}{B_T}\Lambda\left(\frac{x}{B_T}\right) + \sum_{j\in\mathbb{Z},j\neq 0}\frac{1}{B_T}\Lambda\left(\frac{x+2\pi j}{B_T}\right),$$

then by the triangle inequality of total variation, we have

$$V_{-\pi}^{\pi}(W_\lambda^{(T)}) \leq V_{-\pi B_T}^{\pi B_T}(\Lambda/B_T) \leq V_{-\pi}^{\pi}(\Lambda/B_T) = \frac{1}{B_T}V_{-\pi}^{\pi}(\Lambda)$$

where the second inequality holds because $B_T < 1$. Here we use several properties of total variation, see Lemma F.6 in Panaretos and Tavakoli (2013) [44]. □

Lemma 4.7.3. *If $B_T \to 0$,*

$$\frac{2\pi}{T}\sum_{s=1}^{T-1}W_\lambda^{(T)}(\omega - 2\pi s/T) = 1 + O(B_T^{-1}T^{-1}).$$

Proof. Let

$$\Delta_n = \int_{-\pi}^{\pi}W_\lambda^{(T)}(\omega-\alpha)d\alpha - \frac{2\pi}{T}\sum_{s=1}^{T-1}W_\lambda^{(T)}(\omega - 2\pi s/T)$$

$$= \int_{-\pi}^{\pi} W'^{(T)}_{\lambda}(\alpha)d\alpha - \frac{2\pi}{T}\sum_{s=1}^{T-1} W'^{(T)}_{\lambda}(2\pi s/T)$$

where $W'^{(T)}_{\lambda}(\alpha) = W^{(T)}_{\lambda}(\omega - \alpha)$. We have

$$\begin{aligned}
|\Delta_n| &\leq \frac{2\pi}{T}\left\{V^{\pi}_{-\pi}(W'^{(T)}_{\lambda}) + ||W'^{(T)}_{\lambda}||_{\infty}\right\} \\
&\leq \frac{2\pi}{T}\left\{\frac{1}{B_T}V^{\pi}_{-\pi}(\Lambda) + \frac{1}{B_T}||\Lambda||_{\infty} + O(B_T)\right\} \\
&= \frac{2\pi}{B_T T}\left\{V^{\pi}_{-\pi}(\Lambda) + ||\Lambda||_{\infty} + O(B_T^2)\right\}.
\end{aligned}$$

where the first inequality above follows from Lemma F.10 in Panaretos and Tavakoli (2013) [44] and the second inequality follows from (iii) and (iv) of Lemma 4.7.2. Hence, $|\Delta_n|$ is of order $O(B_T^{-1}T^{-1})$ as $B_T \to 0$. Then by Lemma 4.7.2(ii), we obtain

$$\frac{2\pi}{T}\sum_{s=1}^{T-1} W^{(T)}_{\lambda}(\omega - 2\pi s/T) = 1 + O(B_T^{-1}T^{-1}).$$

□

Proof of Theorem 4.2.1. By Proposition 2.6 in Panaretos and Tavakoli (2013) [43], under *C(0,2)* the periodogram kernel has expectation

$$\mathbb{E}[p^{(T)}_{2\pi s/T}(\tau,\sigma)] = f_{2\pi s/T}(\tau,\sigma) + O(T^{-1}),$$

then we can write

$$\mathbb{E}[\hat{f}^{(T)}_{\omega,\lambda}(\tau,\sigma)] = \frac{2\pi}{T}\sum_{s=1}^{T-1} W^{(T)}_{\lambda}\left(\omega - \frac{2\pi s}{T}\right)\left\{f_{2\pi s/T}(\tau,\sigma) + O(T^{-1})\right\} = A + B,$$

where

$$A = \frac{2\pi}{T}\sum_{s=1}^{T-1} W^{(T)}_{\lambda}\left(\omega - \frac{2\pi s}{T}\right) f_{2\pi s/T}(\tau,\sigma),$$

$$B = O(T^{-1}) \left\{ \frac{2\pi}{T} \sum_{s=1}^{T-1} W_\lambda^{(T)} \left(\omega - \frac{2\pi s}{T} \right) \right\}.$$

Noting that $||W_\lambda^{(T)}||_\infty = O(B_T^{-1})$ by Lemma 4.7.2, and $||f_\cdot||_\infty = O(1)$, it follows from Lemma F.6(i) and F.10 of Panaretos and Tavakoli (2013) [44] that

$$A = \int_0^{2\pi} W_\lambda^{(T)}(\omega - \alpha) f_\alpha(\tau, \sigma) d\alpha + \varepsilon_T$$

where $\varepsilon_T \sim O(B_T^{-1}T^{-1})$, uniformly in ω. Using Lemma 4.7.3, $B = O(T^{-1})$ if $B_T T \to \infty$. Combining these facts, and with a change of variable $\alpha = \omega - xB_T$ on the integral, we obtain

$$\begin{aligned} \mathbb{E}[\hat{f}_{\omega,\lambda}^{(T)}(\tau, \sigma)] &= \int_0^{2\pi} W_\lambda^{(T)}(\omega - \alpha) f_\alpha(\tau, \sigma) d\alpha + O(B_T^{-1}T^{-1}) + O(T^{-1}) \\ &= \int_{\mathbb{R}} \Lambda(x) f_{\omega - xB_T} dx + O(B_T^{-1}T^{-1}). \end{aligned} \tag{4.33}$$

Following the similar lines in the proof of Lemma F.4 in Panaretos and Tavakoli (2013) [43], we can use the Taylor expansion of $f_{\omega - xB_T}$ to obtain

$$\begin{aligned} \int_{\mathbb{R}} \Lambda(x) f_{\omega - xB_T} dx =& f_\omega + \sum_{k=1}^{p-1} \frac{(-1)^k B_T^k}{k!} \cdot \frac{\partial^k f_\omega}{\partial \omega^k} \cdot \int_{\mathbb{R}} x^k \Lambda(x) dx \\ &+ \frac{B_T^p}{p!} \cdot \sup \left\| \frac{\partial^k f_\omega}{\partial \omega^k} \right\| \cdot \int_{\mathbb{R}} |x|^p \Lambda(x) dx. \end{aligned}$$

With Λ being a kernel of order p, i.e. $\int_{\mathbb{R}} x^k \Lambda(x) dx = 0$ for all $k = 1, \ldots, p-1$, we obtain

$$\int_{\mathbb{R}} \Lambda(x) f_{\omega - xB_T} dx = f_\omega + O(B_T^p). \tag{4.34}$$

Combining (4.33) and (4.34) completes the proof.

4.7.3 Proof of Proposition 4.4.1

Recall the spectral decompositions of the operators $\mathscr{F}_\omega$, $\hat{\mathscr{F}}^{(T)}_{\omega,\lambda}$ and $\tilde{\mathscr{F}}^{(T)}_{\omega,\lambda}$. For $h \in L^2([0,1],\mathbb{R})$,

$$\mathscr{F}_\omega(h) = \sum_{j=1}^{\infty} \nu_j \langle h, e_j \rangle e_j,$$

$$\hat{\mathscr{F}}^{(T)}_{\omega,\lambda}(h) = \sum_{j=1}^{\infty} \hat{\nu}_j \langle h, \hat{e}_j \rangle \hat{e}_j,$$

$$\tilde{\mathscr{F}}^{(T)}_{\omega,\lambda}(h) = \sum_{j=1}^{\infty} \tilde{\nu}_j \langle h, \hat{e}_j \rangle \hat{e}_j.$$

Noting that $\mathscr{F}_\omega$ and $\hat{\mathscr{F}}^{(T)}_{\omega,\lambda}$ are both induced by kernel functions through right integration, the two operators commute, i.e., $\mathscr{F}_\omega \hat{\mathscr{F}}^{(T)}_{\omega,\lambda} = \hat{\mathscr{F}}^{(T)}_{\omega,\lambda} \mathscr{F}_\omega$. Due to the fact that commuting operators can be simultaneously diagonalised, $\mathscr{F}_\omega$ and $\hat{\mathscr{F}}^{(T)}_{\omega,\lambda}$ share the common eigenfunctions, i.e., $e_j = \hat{e}_j$ for all j. Therefore, eigenvalues of $\hat{\mathscr{F}}^{(T)}_{\omega,\lambda} - \mathscr{F}_\omega$ are $\hat{\nu}_j - \nu_j$, $j \geq 1$; and eigenvalues of $\tilde{\mathscr{F}}^{(T)}_{\omega,\lambda} - \mathscr{F}_\omega$ are $\tilde{\nu}_j - \nu_j$, $j \geq 1$. Viewed as an estimator of the nonnegative ν_j, $\tilde{\nu}_j$ is a better estimator than $\hat{\nu}_j$ in the sense that $|\tilde{\nu}_j - \nu_j| \leq |\hat{\nu}_j - \nu_j|$ always holds true. Hence, it follows that

$$\left|\left|\left| \tilde{\mathscr{F}}^{(T)}_{\omega,\lambda} - \mathscr{F}_\omega \right|\right|\right|_2^2 = \sum_{j=1}^{\infty} (\tilde{\nu}_j - \nu_j)^2 \leq \sum_{j=1}^{\infty} (\hat{\nu}_j - \nu_j)^2 = \left|\left|\left| \hat{\mathscr{F}}^{(T)}_{\omega,\lambda} - \mathscr{F}_\omega \right|\right|\right|_2^2,$$

which completes the proof.

Acknowledgement. Chapter 4 is joint work with Professor Dimitris Politis, which is a manuscript in preparation for later submission. Tingyi Zhu and Dimitris Politis are the primary investigators and authors of the manuscript.

Bibliography

[1] Adler, R. J. (1990). An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes. *Institute of Mathematical Statistics Lecture Notes–Monograph Series 12, IMS, Hayward, CA*.

[2] Antoniadis, A., Paparotidis, E. and Sapatinas, T. (2006). A functional wavelet-kernel approach for time series prediction. *Journal of the Royal Statistical Society, Series B*. **68** 837–857.

[3] Antoniadis, A. and Sapatinas, T. (2003). Wavelet methods for continuous time prediction using Hilbert-valued autoregressive processes. *Journal of Multivariate Analysis*. **87** 133–158.

[4] Aue, A., Norinho, D. D. and Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association*. **110** 378–392.

[5] Berkes, I., Horváth, L. and Rice, G. (2016). On the asymptotic normality of kernel estimators of the long run covariance of functional time series. *Journal of Multivariate Analysis*. **144** 150-175.

[6] Bosq, D. (2000). *Linear processes in function space*. Springer, New York.

[7] Bosq, D. (2002). Estimation of mean and covariance operator of autoregressive processes in Banach space. *Stat. Inference Stoch. Process*, **5** 287-306.

[8] Bosq, D. (2007). General linear processes in Hilbert spaces and prediction. *Journal of Statistical Planning and Inference*. **137** 879–894.

[9] Brillinger, D. R. (2001). Time series: data analysis and theory. *Society for Industrial and Applied Mathematics*.

[10] Brillinger, D. R. and Rosenblatt, M. (1967). Asymptotic theory of kth order spectra. *Spectral Analysis of Time Series, (B. Harris, Ed.), Wiley, New York*. 153–188.

[11] Brockwell, P. J. and Richard, A. D. (2013). Time series: theory and methods. *Springer Science & Business Media*.

[12] Dehling, H. and Sharipov, O. S. (2005). Estimation of mean and covariance operator for Banach space valued autoregressive processes with dependent innovations. *Stat. Inference Stoch. Process.* **8** 137–149.

[13] Delsol, L. (2009). Advances on asymptotic normality in non-parametric functional time series analysis. *Statistics: A Journal of Theoretical and Applied Statistics.* **43(1)** 13–33.

[14] Didericksen, D., Kokoszka, P. and Zhang, X. (2012). Empirical properties of forecasts with the functional autoregressive model. *Comput. Stat.* **27(2)** 285–298.

[15] Ferraty, F., Keilegom, I. V. and Vieu, P. (2010). On the validity of the bootstrap in non-parametric functional regression. *Scandinavian Journal of Statistics*. **37** 286–306.

[16] Ferraty, F., Keilegom, I. V. and Vieu, P. (2012). Regression when both response and predictor are functions. *Journal of Multivariate Analysis*. **109** 10–28.

[17] Ferraty, F., Mas, A. and Vieu, P. (2007). Nonparametric regression on functional data: inference and practical aspects. *Aust. N. Z. J. Stat.* **49** 267–286.

[18] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis, Theory and Practice*. Springer, New York.

[19] Franke, J., Kreiss, J.-P. and Mammen, E. (2002). Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli*. **8(1)** 1–37.

[20] Franke, J. and Nyarige, E. (2016). On the residual-based bootstrap for functional autoregression. *working paper, Univ. of Kaiserslautern, Germany*.

[21] Gabrys, R., Horváth, L. and Kokoszka, P. (2010). Tests for error correlation in the functional linear model. *Journal of American Statistical Association*. **105** 1113–1125.

[22] Grenander, U. (1951). On empirical spectral analysis of stochastic processes. *Ark. Mat.*. **1** 503–531.

[23] Hannan, E. J. (1970). Multiple Time Series. *John Wiley, New York*.

[24] Hörmann, S., Kidziński, L. and Hallin, M. (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society, Series B.* **77(2)** 319–348.

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

[26] Hörmann, S. and Kokoszka, P. (2012). Functional Time Series. *Handbook of statistics.* **30** 157–186.

[27] Horváth, L. and Kokoszka, P. (2011). *Inference for functional data with applications.* Springer Series in Statistics, Springer, New York.

[28] Horváth, L., Kokoszka, P. and Reeder, R. (2013). Estimation of the mean of functional time series and a two sample problem. *Journal of the Royal Statistical Society.* **75** 193–122.

[29] Horváth, L. and Rice, G. (2015). Testing for independence between functional time series. *Journal of Econometrics.* **189(2)** 371–382.

[30] Horváth, L., Rice, G. and Whipple, S. (2016). Adaptive bandwidth selection in the long run covariance estimator of functional time series. *Computational Statistics & Data Analysis.* **100** 676–693.

[31] Ibragimov, I. A. (1962). *Some limit theorems for stationary processes.* Teor. Veroyatn. Primen, **7** 361–392.

[32] Jirak, M. (2013). On weak invariance principles for sums of dependent random functionals. *Statistics and Probability Letter.* **83** 2291–2296.

[33] Katznelson, Y. (2004). An introduction to harmonic analysis. *Cambridge University Press.*

[34] Kargin, V. and Onatski, A. (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis.* **99** 2508–2526.

[35] Klepsch, J. and Klüppelberg, C. (2017). An innovations algorithm for the prediction of functional linear processes. *Journal of Multivariate Analysis.* **155** 252–271.

[36] Klepsch, J., Klüppelberg, C. and Wei, T. (2017). Prediction of functional ARMA processes with an application to traffic data. Econometrics and Statistics. **1** 128–149.

[37] Kreiss, J.-P. and Lahiri, S. N. (2012). Bootstrap methods for time series. *Handbook of Statistics: Time Series Analysis: Methods and Applications.* **30(1)**.

[38] Masry, E. (1996). Multivariate regression estimation: Local polynomial fitting for time series. *Stochastic Process. Appl.* **65** 81–101.

[39] Masry, E. (2005). Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Process. Appl.* **115(1)** 155–177.

[40] McMurry, T. and Politis, D. N. (2004). Nonparametric regression with infinite order flat-top kernels. *Journal of Nonparametric Statistics.* **16(3-4)** 549–562.

[41] Neumann, M. H. and Kreiss, J.-P. (1998). Regression-type inference in nonparametric autoregression. *The Annals of Statistics.* **26**, 1570–1613.

[42] Pan, L. and Politis, D. N. (2016). Bootstrap prediction intervals for linear, nonlinear and nonparametric autoregression (with discussion). *Journal of Statistical Planning and Inference.* **177** 1–27.

[43] Panaretos, V. M. and Tavakoli, S. (2013a). Fourier analysis of stationary time series in functional space. *The Annals of Statistics.* **41(2)** 568–603.

[44] Panaretos, V. M. and Tavakoli, S. (2013b). Supplement to "Fourier analysis of stationary time series in functional space.".

[45] Parzen, E. (1957). On the consistent estimation of the spectrum of a stationary time series. *Annal of Mathematical Statistics.* **28(2)** 329–348.

[46] Parzen, E. (1961). Mathematical considerations in the estimation of spectra. *Technometrics.* **3** 167–190.

[47] Politis, D. N. (2001). On nonparametric function estimate with infinite-order flat-top kernels. *Probability and Statistical Models with applications, Ch. Charalambides et al. (Eds.), Chapman and Hall/CRC, Boca Raton.* 469–483.

[48] Politis, D. N. (2003). Adaptive bandwidth choice. *Journal of Nonparametric Statistics.* **15(4-5)** 517–533.

[49] Politis, D. N. (2011). Higher-order accurate, positive semi-definite estimation of large sample covariance and spectral density matrices. *Econometric Theory.* **27(4)** 703–744.

[50] Politis, D. N. (2013). Model-free model fitting and predictive distribution (with discussion). *Test.* **22(2)** 183–250.

[51] Politis, D. N. (2015). *Model-free prediction and regression: A transformation-based approach to inference.* Springer International Publishing.

[52] Politis, D. N. and Romano, J. (1994). Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap. *Statistica Sinica*. **4** 461–476.

[53] Politis, D. N. and Romano, J. (1995). Bias-corrected nonparametric spectral estimation. *Journal of Time Series Analysis*. **16** 67–103.

[54] Politis, D. N. and Romano, J. (1996). On flat-top spectral density estimators for homogeneous random fields. *Journal of Statistical Planning and Inference*. **51** 41–53.

[55] Politis, D. N. and Romano, J. (1996). Multivariate density estimation with general flat-top kernels of infinite order. *Journal of Multivariate Analysis*. **68** 1–25.

[56] Priestley, M. B. (1962). Basic considerations in the estimation of spectra. *Adademic Press, New York*.

[57] Priestley, M. B. (1981). Spectral analysis and time series. *Technometrics*. **4** 551–564.

[58] Ramsay, J. and Silverman, B. W. (1997). *Functional Data Analysis*. Springer-Verlag, New York.

[59] Ramsay, J. and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer-Verlag, New York.

[60] Raña, P., Aneiros, G., Vilar, J. and Vieu, P. (2016). Bootstrap confidence intervals in functional nonparametric regression under dependence. *Electronic Journal of Statistics*. **10** 1973–1999.

[61] Rice, G. and Shang, H. (2016). A plug-in bandwidth selection procedure for long run covariance estimation with stationary functional time series. *Journal of Time Series Analysis*, to appear.

[62] Robinson, P. M. (1983). Nonparametric estimators for time series. *J. Time Ser. Anal.* **4** 185–207.

[63] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci U.S.A.* **42** 43–47.