Higher-order Accurate Spectral Density Estimation of Functional Time Series

Tingyi Zhu and Dimitris N. Politis

Department of Mathematics
University of California San Diego
La Jolla, CA 92093-0112, USA

Abstract

Under the frequency domain framework for weakly dependent functional time series, a key element is the spectral density kernel which encapsulates the second-order dynamics of the process. We propose a class of spectral density kernel estimators based on the notion of a flat-top kernel. The new class of estimators employs the inverse Fourier transform of a flat-top function as the weight function employed to smooth the periodogram. It is shown that using a flat-top kernel yields a bias reduction and results in a higher-order accuracy in terms of optimizing the integrated mean square error (IMSE). Notably, the higher-order accuracy of flat-top estimation comes at the sacrifice of the positive semi-definite property. Nevertheless, we show how a flat-top estimator can be modified to become positive semi-definite (even strictly positive definite) in finite samples while retaining its favorable asymptotic properties. In addition, we introduce a data-driven bandwidth selection procedure realized by an automatic inspection of the estimated correlation structure. Our asymptotic results are complemented by a finite-sample simulation where the higher-order accuracy of flat-top estimators is manifested in practice.

Keywords: Functional time series; spectral density kernel; spectral density estimation flat-top kernel, positive semi-definite estimation

1 Introduction

Functional time series has become a recent focus within the statistical research of functional data analysis due to the fact that functional data are often collected sequentially over time. 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. That amounts to 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) [4], Dehling and Sharipov (2005) [8] consider the estimation of covariance operator for functional autoregressive processes, Horváth and Kokoszka (2010) [11] studied the covariance structure of weakly dependent functional time series under an m-dependence condition; see also Horváth and Kokoszka (2012) [12] 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) [13] 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) [2]. The estimation has applications in mean and stationarity testing of functional time series, see Horváth et. al (2015) [14] and Jirak (2013) [16]. Horváth et. al (2016) [15] and Rice and Shang (2016) [31] 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) [19] 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

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

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) [21, 22], Brillinger and Rosenblatt (1967) [6], Hannan (1970) [10] and Priestley (1981) [30]. 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) [19] under a type of cumulant mixing condition. This weak dependence condition is the functional analog of classical cummulant-type mixing condition of Brillinger (2001) [5].

In this paper, we propose a new class of spectral density kernel estimators based on the notion of flat-top kernel defined in Politis (2001) [23]; see also Politis and Romano (1995, 1996, 1999) [26, 27, 28]. 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 and Richard (2013) [7] and Brillinger (2001) [5].

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) [25], 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 structure of the paper is 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 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 which results into an estimator that is positive semi-definite while retaining the estimator's higher-order accuracy. Section 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 6 where higher-order accuracy of the flat-top estimators are manifested in practice. Finally, the technical proofs are gathered in the Appendix in Section 7.

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. The pointwise definition of a kth *order cumulant kernel* is

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

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 $j = 1, \ldots, k-1$,

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

We inherit the frequency domain framework of functional time series developed in Panaretos and Tavakoli (2013) [19], 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

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

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

$$p_\omega^{(T)}(\tau, \sigma) = \tilde{X}_\omega^{(T)}(\tau)\tilde{X}_{-\omega}^{(T)}(\sigma). \tag{3}$$

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) [19] proposed a consistent estimator $f_\omega^{(T)}$ by convolving the periodogram kernel with a weight function, which has the form

$$f_\omega^{(T)}(\tau, \sigma) = \frac{2\pi}{T} \sum_{s=1}^{T-1} W^{(T)}\left(\omega - \frac{2\pi s}{T}\right) p_{2\pi s/T}^{(T)}(\tau, \sigma). \tag{4}$$

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

$$W^{(T)}(x) = \sum_{j\in\mathbb{Z}} \frac{1}{B_T} W\left(\frac{x + 2\pi j}{B_T}\right) \tag{5}$$

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 (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). With the above constraints imposed on function W, it has been shown in Panaretos and Tavakoli (2013) [19] 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).

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

$$\lambda(s) = \begin{cases} 1 & \text{if } |s| \leq c, \\ g(s) & \text{otherwise.} \end{cases} \tag{6}$$

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 $\lambda(s)$

$$\Lambda(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lambda(s) e^{isx} ds. \tag{7}$$

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 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 $p - 1$.*

In the following, we will be using $\hat{f}_{\omega,\lambda}^{(T)}$ 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), except the difference in the weight function, i.e.,

$$\hat{f}^{(T)}_{\omega,\lambda}(\tau,\sigma) = \frac{2\pi}{T}\sum_{s=1}^{T-1} W^{(T)}_{\lambda}\left(\omega - \frac{2\pi s}{T}\right) p^{(T)}_{2\pi s/T}(\tau,\sigma), \tag{8}$$

where

$$W^{(T)}_{\lambda}(x) = \sum_{j\in\mathbb{Z}} \frac{1}{B_T}\Lambda\left(\frac{x+2\pi j}{B_T}\right), \tag{9}$$

with $\Lambda(x)$ being the flat-top kernel induced by a flat-top function $\lambda(s)$ defined in (7).

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

Theorem 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*

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

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

Remark 2.1. According to Proposition 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 condition is provided in Panaretos and Tavakoli (2013) [19], which is 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.

Remark 2.2. It is worth mentioning that Theorem 2.1-2.3 can hold for a non-flat-top kernel Λ of order p as long as it satisfies properties (i)-(iv) in Lemma 7.2. Nevertheless, in the paper we will be focusing on the flat-top kernels for the simplicity of the proof. In addition, an empirical data-driven bandwidth selection rule is proposed in Section 6, which can be applied exclusively on flat-top kernels.

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 2.2. *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 Lemmas 7.2 and 7.3, Theorem 2.2 can be proved along the same lines of proof of Corollary 3.3 in Panaretos and Tavakoli (2013) [19]. 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) [19].

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)[19]. The spectral density operator $\mathscr{F}_\omega$ is an operator induced by the spectral density kernel through right integration,

$$\mathscr{F}_\omega h(\tau) = \int_0^1 f_\omega(\tau,\sigma)h(\sigma)d\sigma = \frac{1}{2\pi}\sum_{t\in\mathbb{Z}} e^{-i\omega t}\int_0^1 r_t(\tau,\sigma)h(\sigma)d\sigma = \frac{1}{2\pi}\sum_{t\in\mathbb{Z}} e^{-i\omega t}\mathscr{R}_t h(\tau), \tag{10}$$

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

$$\mathscr{R}_t h(\tau) = \int_0^1 r_t(\tau,\sigma)h(\sigma)d\sigma = \mathrm{cov}[\langle X_0, h\rangle, X_t(\tau)], \qquad h \in L^2([0,1],\mathbb{R}). \tag{11}$$

The spectral density operator $\mathscr{F}_\omega$ is the integral operator with kernel f_ω. Analogously, we denote $\hat{\mathscr{F}}_{\omega,\lambda}^{(T)}$ the operator induced by the the kernel $\hat{f}_{\omega,\lambda}^{(T)}$ 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 $\hat{\mathscr{F}}_{\omega,\lambda}^{(T)}$ for the spectral density operator $\mathscr{F}_\omega$.

Theorem 2.3. *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 $\hat{\mathscr{F}}_{\omega,\lambda}^{(T)}$ employing a flat-top kernel Λ of order p is consistent in integrated mean square, that is,*

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

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

Theorem 2.3 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 $O(\log T/T)$.

3 Alternate estimates and flat-top kernel choice

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 (8) has an alternate form

$$W_\lambda^{(T)}(x) = \frac{1}{2\pi}\sum_{u\in\mathbb{Z}} \lambda(B_T u)e^{-ixu}. \tag{12}$$

The equivalence of (9) and (12) can be easily verified by using Poisson summation formula. Moreover, if the discrete average in (8) is replaced by a continuous one, the estimate becomes

$$\begin{aligned}\int_0^{2\pi} W_\lambda^{(T)}(\omega-\alpha)p_\alpha^{(T)}(\tau,\sigma)d\alpha &= \int_0^{2\pi} W_\lambda^{(T)}(\alpha)p_{\omega-\alpha}^{(T)}(\tau,\sigma)d\alpha \\ &= \int_{-\infty}^{\infty} B_T^{-1}\Lambda(B_T^{-1}\alpha)p_{\omega-\alpha}^{(T)}(\tau,\sigma)d\alpha. \end{aligned}\tag{13}$$

By Equation (2) and (3), the periodogram kernel is given by

$$p_\omega^{(T)}(\tau,\sigma) = \frac{1}{2\pi}\sum_{|u|<T} \hat{r}_u(\tau,\sigma)e^{-i\omega u} \tag{14}$$

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 (13), then the estimate takes the form

$$\frac{1}{2\pi}\sum_{|u|<T} \lambda(B_T u)\hat{r}_u(\tau,\sigma)e^{-i\omega u} \tag{15}$$

where

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

With a flat-top function λ in place, the estimate (15) has a formal resemblance to the flat-top estimation for multivariate spectrum explored in Politis (2011) [25] 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) [25] for details. In

fact, the estimate (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) [9], Parzen (1957) [21] and Priestley (1962) [29].

3.2 Flat-top kernel choice

As suggested by Theorem 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 2.1. McMurry and Politis (2004) [18] constructed a member of the flat-top family that is infinitely differentiable, which is defined as

$$\lambda_{ID,b,c}(s) = \begin{cases} 1 & \text{if } |s| \le c, \\ \exp(-b\exp(-b/(|s|-c)^2)/(|s|-1)^2) & \text{if } c < |s| < 1, \\ 0 & \text{if } |s| \ge 1 \end{cases} \tag{16}$$

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 $|x| \to \infty$. Fig.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 (9) or (12) as their equivalence stated in Section 3.1.

Nevertheless, while the effectiveness of the flat-top kernels is reflected in Theorem 2.1 and 2.3, they in fact provide merely theoretical bounds for the decay rate of bias and IMSE. In the meantime, according to Theorem 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

$$\lambda_{TR,c}(s) = \begin{cases} 1 & \text{if } |s| \le c, \\ \dfrac{|s|-1}{c-1} & \text{if } c < |s| < 1, \\ 0 & \text{if } |s| \ge 1. \end{cases} \tag{17}$$

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) [25]. The infinitely differentiable function $\lambda_{ID,b,c}(s)$ looks very much like the trapezoidal $\lambda_{TR,1/2}$ with ultra-smoothed corners.

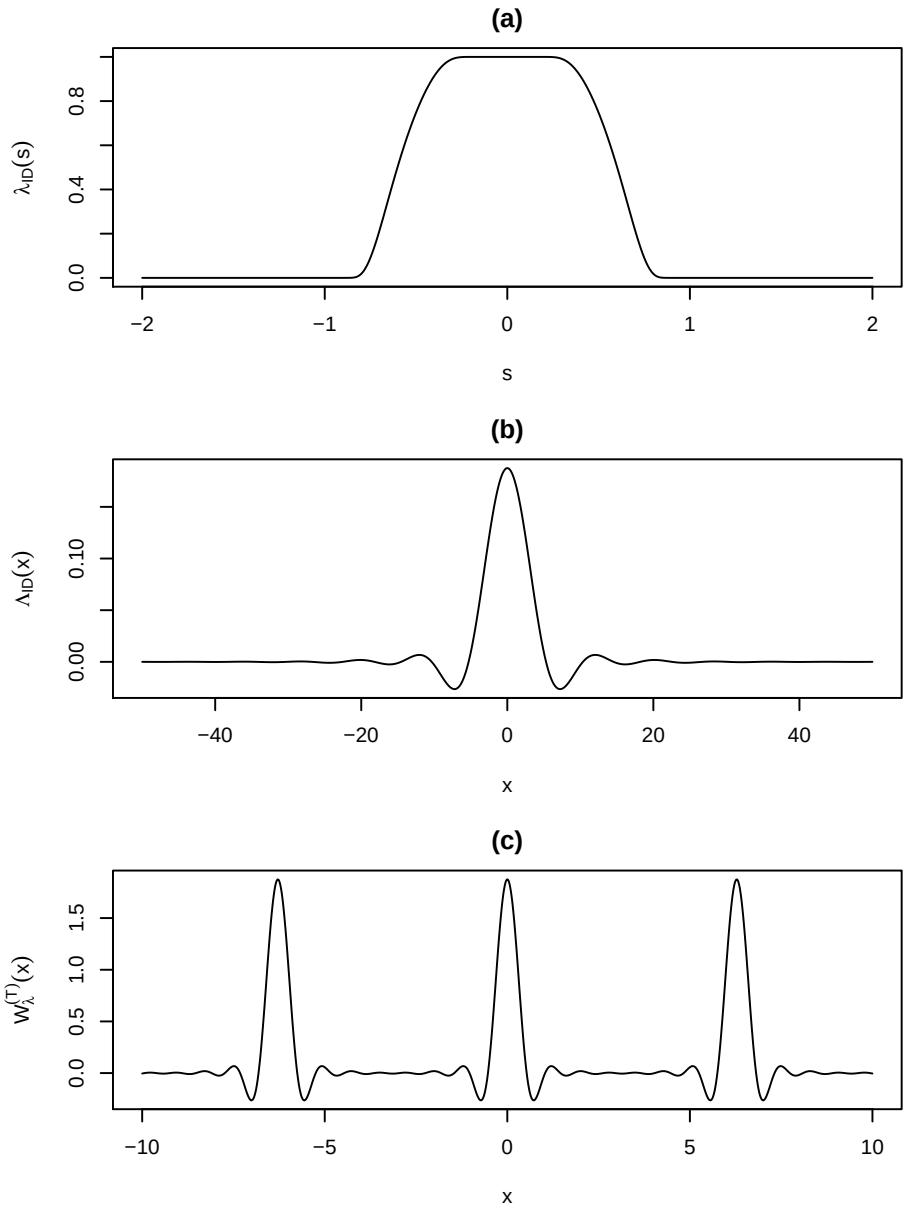


Figure 1: (a) Plot of $\lambda_{ID,1/4,0.05}(s)$; (b) Plot of corresponding kernel $\Lambda(x)$ induced by inverse Fourier transform of $\lambda_{ID,1/4,0.05}(s)$; (c) Plot of the corresponding weight function $W_\lambda^{(T)}$ with $B_T = 0.1$.

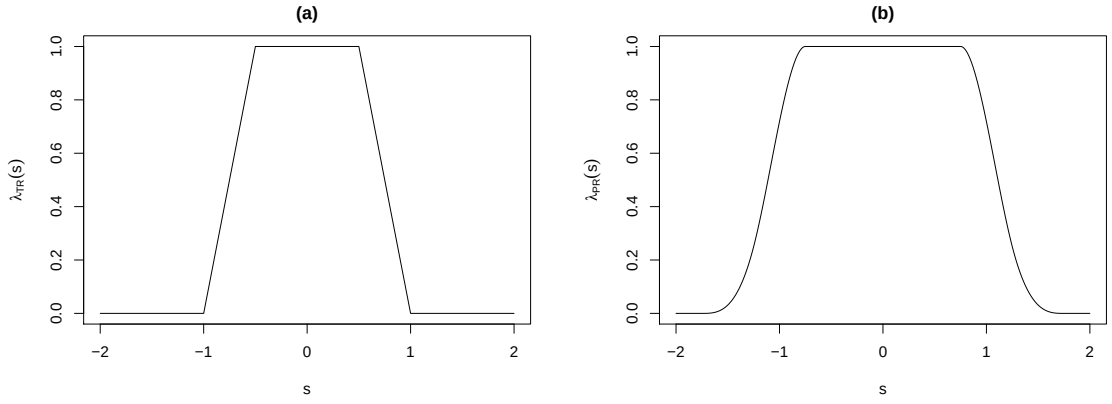


Figure 2: (a) Plot of trapezoidal $\lambda_{TR,1/2}(s)$; (b) Plot of flat-top Parzen $\lambda_{PR,3/4}(s)$.

Another choice to be considered is the flat-top function created by adding a piecewise cubic tail, similar to that of Parzen's (1961) [22] kernel, to the $[-c, c]$ flat-top region. It is defined as

$$\lambda_{PR,c}(s) = \begin{cases} 1 & \text{if } 0 \le s \le c, \\ 1 - 6(s-c)^2 + 6|s-c|^3 & \text{if } c \le s \le c + 1/2, \\ 2(1 - |s-c|)^3 & \text{if } c + 1/2 \le s \le c + 1 \\ 0 & \text{if } s \ge c + 1, \\ \lambda_{PR,c}(-s) & \text{if } s < 0. \end{cases} \tag{18}$$

Plots of flat-top functions $\lambda_{TR,1/2}$ and $\lambda_{PR,3/4}$ are shown in Figure 2. Concerning the choice of parameters of flat-top kernels, i.e., b and c, we refer the readers to Politis (2011) [25] where a detailed discussion is given.

4 Positive semi-definite spectral estimation

By employing the infinite-order flat-top kernels, the flat-top estimator $\hat{\mathscr{F}}^{(T)}_{\omega,\lambda}$ 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) [25]. Analogously, we now show how the flat-top operator estimator $\hat{\mathscr{F}}_{\omega,\lambda}^{(T)}$ 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 $\hat{\mathscr{F}}_{\omega,\lambda}^{(T)}$ are induced by kernel functions through right integration, and therefore they are symmetric Hilbert-Schmidt operators that admit the following decompositions

$$\mathscr{F}_\omega(h) = \sum_{j=1}^{\infty} \nu_j \langle h, e_j \rangle e_j, \qquad h \in L^2([0,1], \mathbb{R}) \tag{19}$$

$$\hat{\mathscr{F}}_{\omega,\lambda}^{(T)}(h) = \sum_{j=1}^{\infty} \hat{\nu}_j \langle h, \hat{e}_j \rangle \hat{e}_j, \qquad h \in L^2([0,1], \mathbb{R}) \tag{20}$$

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 $j \geq 1$,

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

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 $\hat{\mathscr{F}}_{\omega,\lambda}^{(T)}$ 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 $\hat{\mathscr{F}}_{\omega,\lambda}^{(T)}$, let $\tilde{\nu}_j = \max(\hat{\nu}_j, 0)$ for all j, and define the estimator

$$\tilde{\mathscr{F}}_{\omega,\lambda}^{(T)}(h) = \sum_{j=1}^{\infty} \tilde{\nu}_j \langle h, \hat{e}_j \rangle \hat{e}_j, \qquad h \in L^2([0,1], \mathbb{R}). \tag{21}$$

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

Proposition 4.1. *Let $\tilde{\mathscr{F}}_{\omega,\lambda}^{(T)}$ be the positive semi-definite operator estimator of $\mathscr{F}_\omega$ defined in (9), then for a fixed ω*

$$\left|\left|\left| \tilde{\mathscr{F}}_{\omega,\lambda}^{(T)} - \mathscr{F}_\omega \right|\right|\right|_2 \leq \left|\left|\left| \hat{\mathscr{F}}_{\omega,\lambda}^{(T)} - \mathscr{F}_\omega \right|\right|\right|_2, \tag{22}$$

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

A direct consequence of Proposition 4.1 is the following corollary which shows that, in addition to being positive semi-definite, $\tilde{\mathscr{F}}_{\omega,\lambda}^{(T)}$ possesses the same mean square convergence of $\hat{\mathscr{F}}_{\omega,\lambda}^{(T)}$ given in Theorem 2.3.

Theorem 4.1. *Under the condition of Theorem 2.3, 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 case the estimand $\mathscr{F}_{\omega}$ is not only positive semi-definite but strictly positive definite, it may be 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 chosen sequence, and define the estimator

$$\check{\mathscr{F}}_{\omega,\lambda}^{(T)}(h) = \sum_{j=1}^{\infty} \check{\nu}_j \langle h, \hat{e}_j \rangle \hat{e}_j, \qquad h \in L^2([0,1], \mathbb{R}). \tag{23}$$

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.

Remark 4.1. Proposition 4.1 was phrased in terms of the the flat-top spectral estimator $\hat{\mathscr{F}}_{\omega,\lambda}^{(T)}(h)$ and its estimand $\mathscr{F}_{\omega}(h)$. However, an analogous construction, i.e., replacing negative eigenvalues by zero, can be used in different functional estimation situations where the estimator is a (symmetric) Hilbert-Schmidt operator and its estimand is non-negative definite; the proof remains the same. Similarly, if the estimand is strictly positive definite, the construction of eq. (23) may be used.

5 Data-dependent bandwidth choice

As it has been demonstrated in Section 3.1 that the lag-window estimate (15) with bandwidth m_T is nearly equal to the estimate (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 choice rule for flat-top lag-windows introduced in Politis (2003) [24] for the univariate case, and Politis (2011) [25] for the multivariate case.

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-correlogram, 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

$$\hat{B}_T = \frac{1}{\max(\lceil \hat{q}/c \rceil, 1)} \tag{24}$$

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) [25], 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, (16) and Figure 1(a) regarding the infinitely 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) [24] makes the concrete recommendations $C_0 \simeq 2$ 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. $B_T = 1/\max(\lceil \hat{q}_{\tau,\sigma}/c_{ef} \rceil, 1)$.

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 of order one, i.e., the MA(1) model:

$$X_t = A_0\varepsilon_t + A_1\varepsilon_{t-1}. \tag{25}$$

The simulations we carry out are analogous to the ones conducted in Panaretos and Tavakoli (2013) [19]. 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 $n = 6,\ldots,11$. 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) [19]. 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 5.

The simulation results are presented in Table 1, entries of which are logarithm 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 performance of flat-top Parzen's kernel and flat-top infinitely differentiable kernel is close, while each slightly outperforms the trapezoid as the sample size grows. This may be due to the fact that the smoothness of the flat-top functions is indeed a factor on the decay of IMSE. However, a high degree of smoothness is often superfluous as the performance is typically limited by the order of cumulant conditions as Theorem 2.3 suggests.

In case cumulant conditions do not pose a limitation, we might expect that a very smooth flat-top kernel offers an advantage. To verify this assertion, we conducted another simulation experiment using a functional MA(0) series, i.e., letting $A_1 = 0$ in eq. (25). Since the data come from an independent functional time series that satisfies cumulant condition $C(p, 2)$ for $p = \infty$, we may implement the bandwidth $B_T = 1/\log T$. Table 2 is in line with our theory as it shows that the flat-top infinitely differentiable kernel yields better performance than both the trapezoid and flat-top Parzen's kernel as the sample size grows.

7 Appendix: Proofs

7.1 Proof of Proposition 2.1

To prove Proposition 2.1, we need the following lemma, which measures the magnitude of Fourier coefficient.

Lemma 7.1. *Let f be an integrable function on the interval $[0, 2\pi]$, and $\hat{f}(n)$ be its Fourier coefficients defined by*

$$\hat{f}(n) = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{int} dt. \tag{26}$$

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 parts on Equation (26), we have

$$\hat{f}(n) = (in)^{-p} \widehat{f^{(p)}}(n). \tag{27}$$

T	64	128	256	512	1024	2048
Epanechnikov kernel	-6.188	-6.852	-7.588	-8.276	-9.082	-9.816
$\lambda_{TR,1/2}$ (Trapezoid)	**-6.389**	-7.112	-7.902	-8.719	-9.493	-10.146
$\lambda_{PR,3/4}$ (flat-top Parzen)	-6.383	-7.041	-8.018	**-8.846**	-9.710	-10.453
$\lambda_{ID,1/4,0.05}$ (flat-top Inf. Diff.)	-6.344	**-7.262**	**-8.074**	-8.832	**-9.719**	**-10.470**

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

T	64	128	256	512	1024	2048
Epanechnikov kernel	-7.148	-7.668	-8.577	-9.075	-9.950	-10.887
$\lambda_{TR,1/2}$ (Trapezoid)	-7.321	-7.808	-8.789	-9.297	-10.204	-11.193
$\lambda_{PR,3/4}$ (flat-top Parzen)	**-7.343**	-8.028	-8.881	**-9.562**	-10.306	**-11.426**
$\lambda_{ID,1/4,0.05}$ (flat-top Inf. Diff.)	-7.241	**-8.216**	**-8.911**	-9.546	**-10.332**	-11.412

(b) Bandwidth $B_T = 2 \cdot T^{-1/5}$

T	64	128	256	512	1024	2048
$\lambda_{TR,1/2}$ (Trapezoid)	-6.519	-7.331	-8.260	-9.145	-10.089	-11.008
$\lambda_{PR,3/4}$ (flat-top Parzen)	**-6.627**	**-7.592**	**-8.455**	**-9.349**	**-10.313**	**-11.371**
$\lambda_{ID,1/4,0.05}$ (flat-top Inf. Diff.)	-6.118	-6.925	-8.053	-9.056	-10.214	-11.121

(c) Empirical rule of choosing B_T

Table 1: Simulation results of the functional MA(1) model; Entries represent the logarithm of IMSEs in base 2 of different estimators using (a) bandwidth $B_T = T^{-1/5}$, (b) bandwidth $B_T = 2 \cdot T^{-1/5}$ and (c) empirical rule of choosing B_T. Sample size ranges from 2^6 to 2^{11}. Minimum IMSE for each T is indicated by boldface.

T	64	128	256	512	1024	2048
Epanechnikov kernel	-8.628	-9.464	-10.606	-11.267	-11.966	-12.703
$\lambda_{TR,1/2}$ (Trapezoid)	**-9.249**	**-10.011**	-10.754	-11.445	-12.223	-13.021
$\lambda_{PR,3/4}$ (flat-top Parzen)	-8.980	-9.830	-10.626	-11.462	-12.353	-13.112
$\lambda_{ID,1/4,0.05}$ (flat-top Inf. Diff.)	-8.992	-9.971	**-10.845**	**-11.892**	**-12.821**	**-13.787**

Table 2: Simulation results of the functional MA(0) model; Entries represent the logarithm of IMSEs in base 2 of different estimators using bandwidth $B_T = 1/\log T$. Sample size ranges from 2^6 to 2^{11}. Minimum IMSE for each T is indicated by boldface.

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

$$|\widehat{f^{(p)}}(n)| \leq \frac{C}{|n|^{\alpha}}, \tag{28}$$

for some constant C. Combining (27) and (28), we obtain

$$|\hat{f}(n)| = |n|^{-p}|\widehat{f^{(p)}}(n)| = O(|n|^{-p-\alpha}). \tag{29}$$

□

Proof of Proposition 2.1. For a flat-top function $\lambda(s)$ and its inverse Fourier transform $\Lambda(x)$, we have

$$\Lambda(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \lambda(s)e^{isx}ds, \tag{30}$$

$$\lambda(s) = \int_{\mathbb{R}} \Lambda(x)e^{-isx}dx. \tag{31}$$

Since the assumption that λ is p times differentiable and $\lambda^{(p)}$ is Hölder continuous of order $0 < \alpha < 1$, by Lemma 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 (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

$$\int_{\mathbb{R}} x^k\Lambda(x)dx = 0 \text{ for } k = 1, \ldots, p-2 \tag{32}$$

if we set $s = 0$ on both sides of (32). Therefore, Λ is a kernel of order $p-1$.

7.2 Proof of Theorem 2.1

To prove Theorem 2.1, the following lemmas are necessary.

Lemma 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 (31). The statement (ii) is obtained by following the same arguments in the proof of Lemma F.11 in Panaretos and Tavakoli (2013) [19]. 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 $|W_\lambda^{(T)}(x)| \leq \frac{1}{B_T}|\Lambda(x/B_T)| + O(B_T)$. 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) [20]. □

Lemma 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

$$\begin{aligned}\Delta_n &= \int_{-\pi}^{\pi} W_\lambda^{(T)}(\omega-\alpha)d\alpha - \frac{2\pi}{T}\sum_{s=1}^{T-1} W_\lambda^{(T)}(\omega - 2\pi s/T) \\ &= \int_{-\pi}^{\pi} W'^{(T)}_\lambda(\alpha)d\alpha - \frac{2\pi}{T}\sum_{s=1}^{T-1} W'^{(T)}_\lambda(2\pi s/T)\end{aligned}$$

where $W'^{(T)}_\lambda(\alpha) = W_\lambda^{(T)}(\omega-\alpha)$. 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) [20] and the second inequality follows from (iii) and (iv) of Lemma 7.2. Hence, $|\Delta_n|$ is of order $O(B_T^{-1}T^{-1})$ as $B_T \to 0$. Then by Lemma 7.2(ii), we obtain

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

□

Proof of Theorem 2.1. By Proposition 2.6 in Panaretos and Tavakoli (2013) [19], 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

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

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

Noting that $||W^{(T)}_{\lambda}||_{\infty} = O(B_T^{-1})$ by Lemma 7.2, and $||f_{.}||_{\infty} = O(1)$, it follows from Lemma F.6(i) and F.10 of Panaretos and Tavakoli (2013) [20] that

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

where $\varepsilon_T \sim O(B_T^{-1}T^{-1})$, uniformly in ω. Using Lemma 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

$$\begin{aligned}\mathbb{E}[\hat{f}^{(T)}_{\omega,\lambda}(\tau,\sigma)] &= \int_0^{2\pi} W^{(T)}_{\lambda}(\omega - \alpha) f_{\alpha}(\tau,\sigma) d\alpha + O(B_T^{-1}T^{-1}) + O(T^{-1}) \\ &= \int_{\mathbb{R}} \Lambda(x) f_{\omega - xB_T} dx + O(B_T^{-1}T^{-1}). \end{aligned} \tag{33}$$

Following the similar lines in the proof of Lemma F.4 in Panaretos and Tavakoli (2013) [19], 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

$$\int_{\mathbb{R}} \Lambda(x) f_{\omega - xB_T} dx = f_{\omega} + O(B_T^p). \tag{34}$$

Combining (33) and (34) completes the proof.

7.3 Proof of Proposition 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

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

which completes the proof.

Acknowledgement Research was partially supported by NSF grant DMS 16-13026. Many thanks are due to the Co-Editor and two anonymous reviewers for their helpful comments.

References

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

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

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

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

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

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

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

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

[9] U. Grenander, *On empirical spectral analysis of stochastic processes* Ark. Mat., 1, 503-531, 1951.

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

[11] L. Horváth, P. Kokoszka, *Weakly dependent functional data*, Annal of Statistics, 38, 1845-1884, 2010.

[12] L. Horváth, P. Kokoszka, *Inference for functional data with applications*, Springer, New York, 2012.

[13] L. Horváth, P. Kokoszka, R. Reeder, *Estimation of the mean of functional time series and a two sample problem*, Journal of the Royal Statistical Society, 75, 193-122, 2011.

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

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

[16] M. Jirak, *On weak invariance principles for sums of dependent random functionals*, Statistics and Probability Letters, 83, 2291-2296 2013.

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

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

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

[20] V. M. Panaretos, S. Tavakoli, *Supplement to "Fourier analysis of stationary time series in functional space."*, 2013b.

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

[22] E. Parzen, *Mathematical considerations in the estimation of spectra*, Technometrics, 3, 167-190, 1961.

[23] D. N. Politis, *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, 2001.

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

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

[26] D. N. Politis, J. P. Romano, *Bias-corrected nonparametric spectral estimation*, Journal of Time Series Analysis, 16, 67-103, 1995.

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

[28] D. N. Politis, J. P. Romano, *Multivariate density estimation with general flat-top kernels of infinite order*, Journal of Multivariate Analysis, 68, 1-25, 1999.

[29] M. B. Priestley, *Basic considerations in the estimation of spectra*, Academic Press, New York, 1981.

[30] M. B. Priestley, *Spectral analysis and time series*, Technometrics, 4, 551-564, 1962.

[31] G. Rice, H. Shang, *A plug-in bandwidth selection procedure for long run covariance estimation with stationary functional time series*, Journal of Time Series Analysis, 38 (4), 591-609, 2017.