

# Higher-order accurate polyspectral estimation with flat-top lag-windows

Arthur Berg · Dimitris N. Politis

Received: 24 November 2006 / Revised: 4 April 2007 / Published online: 21 September 2007  
© The Institute of Statistical Mathematics, Tokyo 2007

**Abstract** Improved performance in higher-order spectral density estimation is achieved using a general class of infinite-order kernels. These estimates are asymptotically less biased but with the same order of variance as compared to the classical estimators with second-order kernels. A simple, data-dependent algorithm for selecting the bandwidth is introduced and is shown to be consistent with estimating the optimal bandwidth. The combination of the specialized family of kernels with the new bandwidth selection algorithm yields a considerably improved polyspectral estimator surpassing the performances of existing estimators using second-order kernels. Bispectral simulations with several standard models are used to demonstrate the enhanced performance with the proposed methodology.

**Keywords** Bispectrum · Nonparametric estimation · Spectral density · Time series

## 1 Introduction

Lag-window estimation of the high-order spectra under various assumptions is known to be consistent and asymptotically normal (Brillinger and Rosenblatt 1967a,b; Lii and Rosenblatt 1990a,b). However, convergence rates of the estimators depend on the order, or characteristic exponent, of the lag-window used. In general, increasing the order of the lag-window decreases the bias without affecting the order of magnitude

---

A. Berg (✉)  
Department of Statistics, University of Florida, PO Box 110339, Gainesville, FL 32611-0339, USA  
e-mail: berg@ufl.edu

D. N. Politis  
Department of Mathematics, University of California, San Diego, La Jolla, CA 92093-0112, USA  
e-mail: dpolitis@ucsd.edu

of the variance, thus producing an estimator with a faster convergence rate. Although estimators using lag-windows with large orders yield estimates with better mean square error (MSE) rates, they were overlooked and rarely used in practice mainly because of two issues. Firstly, in estimating the second-order spectral density, lag-windows of order larger than two *may* yield negative estimates, despite the fact that the true spectral density is known to be nonnegative. This problem only pertains, if ever, to the second-order spectral density (since higher-order spectra are complex-valued), and is easily remedied by truncating the estimator to zero if it does go negative [thus improving the already optimal convergence rates (Politis and Romano 1995)]. Secondly, when a lag-window has order larger than necessary, the rate of convergence is still optimal, but the multiplicative constant will be suboptimal (Hall and Marron 1988). The second problem is encountered when using a poor choice of large order lag-window like the box-shaped truncated lag-window (Politis and Romano 1995), but there are many other alternatives with descent small-sample performance. Additionally, when the underlying spectral density is sufficiently smooth, this second issue is irrelevant since the lag-window with the largest order performs best. The next section introduces a family of infinite-order lag-windows for estimating the spectral density and higher-order spectra.

The use of infinite-order lag-windows is particularly adept to the estimation of higher-order spectra. Under the typical scenario of exponential decay of the autocovariance function (refer to part (ii) of Theorem 1 within), the MSE rates for estimating the second-order spectral density using a lag-window of order 2 and an infinite-order lag-window are  $N^{-4/5}$  and  $(\log N)/N$  respectively. However, when estimating the third-order spectral density, or bispectrum, the MSE rates become  $N^{-2/3}$  and  $(\log N)/N$  respectively. The disparity grows stronger with yet higher-order spectra.

The problem of choosing the best bandwidth still remains. The optimal bandwidth typically depends on the unknown spectral density leading to a circular problem—estimation of the spectrum requires estimation of the bandwidth which in turn requires estimation of the spectrum. There have been many fixes to this problem; see Jones et al. (1996) for a survey of several methods. Section 3 introduces a new simple, data-dependent method of determining the bandwidth which is shown to converge to the asymptotically ideal bandwidth for flat-top lag-windows. An alternative bandwidth selection algorithm is also included that is designed for use with second-order lag-windows. This algorithm uses the plug-in principle for bandwidth selection but with the flat-top estimators as the plug-in pilots.

Particular attention is given to the bispectrum as it is a key tool in several linearity and Gaussianity tests including Hinich (1982) and Subba Rao and Gabr (1980). The general bandwidth selection algorithm is refined and expanded for the bispectrum. Bispectral simulations compare two different flat-top lag-windows estimators of the bispectrum with accompanying bandwidth selection algorithm to the lag-window estimator using the order two “optimal” lag-window and plug-in bandwidth selection procedure as described in Subba Rao and Gabr (1984).

We define the flat-top lag-window estimate in Sect. 2 and derive its higher-order MSE convergence in Theorem 1 under the ideal bandwidth. In Section 3, a bandwidth selection algorithm tailored to the flat-top estimate is introduced and is shown to automatically adapt to the smoothness of the underlying spectral density and converge

in probability to the ideal bandwidth. The focus is then shifted to the bispectrum in Sect. 4 where the most general function invariant under the symmetries of the bivariate cumulant function is constructed. The bandwidth algorithm is specialized for the bispectrum, and a separate bandwidth algorithm for second-order lag-windows is included that is based on the plug-in method with flat-top estimators as pilots. Simulations of the bispectrum in Sect. 5 exhibit the strength of the flat-top estimators and the bandwidth algorithms.

### 2 Asymptotic performance of a general flat-top window

Let  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$  be a realization of an  $r$ -vector valued  $s^{\text{th}}$ -order stationary (real valued) time series  $X_t = (X_t^{(1)}, \dots, X_t^{(r)})'$  with (unknown) mean  $\boldsymbol{\mu} = (\mu^{(1)}, \dots, \mu^{(r)})'$ . Consider the  $s$ th-order central moment

$$C''_{a_1, \dots, a_s}(\tau_1, \dots, \tau_s) = E \left[ (X_{t+\tau_1}^{(a_1)} - \mu^{(a_1)}) \cdots (X_{t+\tau_s}^{(a_s)} - \mu^{(a_s)}) \right], \tag{1}$$

where the right-hand side is independent of the choice of  $t \in \mathbb{Z}$ . Stationarity allows us to write the above moment as function of  $s - 1$  variables, so we define  $C'_{a_1, \dots, a_s}(\tau_1, \dots, \tau_{s-1}) = C''_{a_1, \dots, a_s}(\tau_1, \dots, \tau_{s-1}, 0)$ . For notational convenience, the sequence  $a_1, \dots, a_s$  will be dropped, so  $C'_{a_1, \dots, a_s}(\tau_1, \dots, \tau_{s-1})$  will be denoted simply by  $C'(\boldsymbol{\tau})$ . Also  $\tau_s$  will occasionally be used, for convenience, with the understanding that  $\tau_s = 0$ .

We express the  $s$ th-order joint cumulant as

$$C_{a_1, \dots, a_s}(\tau_1, \dots, \tau_{s-1}) = \sum_{(v_1, \dots, v_p)} (-1)^{p-1} (p-1)! \mu_{v_1} \cdots \mu_{v_p},$$

where the sum is over all partitions  $(v_1, \dots, v_p)$  of  $\{0, \dots, \tau_{s-1}\}$  and  $\mu_{v_j} = E \left[ \prod_{\tau_i \in v_j} X_{\tau_i}^{(a_i)} \right]$ ; refer to [Jammalamadak \(2006\)](#) for another expression of the joint cumulant. The ( $s$ th-order) spectral density is defined as

$$f(\boldsymbol{\omega}) = \frac{1}{(2\pi)^{s-1}} \sum_{\boldsymbol{\tau} \in \mathbb{Z}^{s-1}} C(\boldsymbol{\tau}) e^{-i\boldsymbol{\tau} \cdot \boldsymbol{\omega}}.$$

We adopt the usual assumption on  $C(\boldsymbol{\tau})$  that it be absolutely summable, thus guaranteeing the existence and continuity of the spectral density. A natural estimator of  $C(\boldsymbol{\tau})$  is given by

$$\widehat{C}(\tau_1, \dots, \tau_{s-1}) = \sum_{(v_1, \dots, v_p)} (-1)^{p-1} (p-1)! \widehat{\mu}_{v_1} \cdots \widehat{\mu}_{v_p}, \tag{2}$$

where

$$\hat{\mu}_{v_j} = \frac{1}{N - \max(v_j) + \min(v_j)} \sum_{k=-\min(v_j)}^{N-\max(v_j)} \prod_{t \in v_j} x_{t+k}^{(a_j)}.$$

It turns out that the second-order and third-order cumulants, those that give rise to the spectrum and bispectrum respectively, are precisely the second-order and third-order central moments in (1). Therefore, in these cases, we can greatly simplify  $\widehat{C}(\boldsymbol{\tau})$  to

$$\widehat{C}(\boldsymbol{\tau}) = \frac{1}{N} \sum_{t=1}^{N-\gamma} \prod_{j=1}^s (x_{t-\alpha+\tau_j}^{(a_j)} - \bar{x}^{(a_j)}), \tag{3}$$

where  $\alpha = \min(0, \tau_1, \dots, \tau_{s-1})$ ,  $\gamma = \max(0, \tau_1, \dots, \tau_{s-1}) - \alpha$ , and  $\bar{x}^{(a_\ell)} = \frac{1}{N} \sum_{j=1}^N x_j^{(a_\ell)}$  for  $\ell = 1, \dots, s-1$ . We extend the domain of  $\widehat{C}$  to  $\mathbb{Z}^s$  by defining  $\widehat{C}(\boldsymbol{\tau}) = 0$  when the sum in (2) or (3) is empty.

Consider a flat-top lag-window function  $\lambda : \mathbb{R}^{s-1} \rightarrow \mathbb{R}$  satisfying the following conditions:

- (i)  $\lambda(\mathbf{x}) \equiv 1$  for all  $\mathbf{x}$  satisfying  $\|\mathbf{x}\| \leq b$ , for some positive number  $b$ .
- (ii)  $|\lambda(\mathbf{x})| \leq 1$  for all  $\mathbf{s}$ .
- (iii) For  $M \rightarrow \infty$  as  $N \rightarrow \infty$ , but with  $M/N \rightarrow 0$ ,

$$\lim_{M \rightarrow \infty} \frac{1}{M^{s-1}} \sum_{\|\mathbf{x}\| \leq N} \lambda\left(\frac{\mathbf{x}}{M}\right) < \infty.$$

- (iv)  $\lambda \in L_2(\mathbb{R}^{s-1})$

The window  $\lambda(\mathbf{x})$  is a ‘‘flat-top’’ because of condition (i); namely, it is constant in a neighborhood of the origin. The constant  $b$  in (i) is used below in constructing the spectral density estimate.

Technically, just requiring  $\lambda$  just to be bounded could replace criterion (ii), but there is no benefit in allowing the window to have values larger than 1. Finally, criteria (iii) and (iv) are satisfied if, for example,  $\lambda$  has compact support.

Define  $\lambda_M(\mathbf{t}) = \lambda(\mathbf{t}/M)$  and consider the smoothed  $s^{\text{th}}$ -order periodogram

$$\hat{f}(\boldsymbol{\omega}) = \frac{1}{(2\pi)^{s-1}} \sum_{\|\boldsymbol{\tau}\| < N} \lambda_M(\boldsymbol{\tau}) \widehat{C}(\boldsymbol{\tau}) e^{-i\boldsymbol{\tau} \cdot \boldsymbol{\omega}}. \tag{4}$$

There is an equivalent expression to this estimator in the frequency domain given by

$$\hat{f}(\boldsymbol{\omega}) = \Lambda_M * I_{a_1, \dots, a_s}(\boldsymbol{\omega}) = \int_{\mathbb{R}^{s-1}} \Lambda_M(\boldsymbol{\omega} - \boldsymbol{\tau}) I_{a_1, \dots, a_s}(\boldsymbol{\tau}) \, d\boldsymbol{\tau},$$

where  $\Lambda_M$  is the Fourier transform of  $\lambda_M$  and  $I_{a_1, \dots, a_s}$  is the  $(s - 1)$ th order periodogram; namely,

$$\Lambda_M(\boldsymbol{\tau}) = \int_{\mathbb{R}^{s-1}} \lambda_M(\boldsymbol{\tau}) e^{-i\boldsymbol{\omega} \cdot \boldsymbol{\tau}} \, d\boldsymbol{\tau}$$

and

$$I_{a_1, \dots, a_s}(\boldsymbol{\omega}) = \frac{1}{(2\pi)^{s-1}} \sum_{\boldsymbol{\tau} \in \mathbb{Z}^{s-1}} \widehat{C}(\boldsymbol{\tau}) e^{-i\boldsymbol{\tau} \cdot \boldsymbol{\omega}}.$$

However, Eq. (4) is computationally simpler, and it is this version that will be used throughout the remainder of this article.

The asymptotic bias convergence rate (and thus the overall MSE convergence rate) of the estimator (4) with a flat-top lag-window  $\lambda$  is superior to traditional estimators using second-order lag-windows. The convergence rates of our estimator improve with the decay rate of the cumulant function  $C(\boldsymbol{\tau})$ —the faster the decay to zero, the faster the convergence. The following theorem outlines convergence rates under three scenarios: when the decay of  $C(\boldsymbol{\tau})$  is polynomial, exponential, and identically zero after some finite time (like an  $MA(q)$  process). Throughout, conditions on the time series are assumed so that

$$\text{var} \left( \widehat{f}(\boldsymbol{\omega}) \right) = O \left( \frac{M^{s-1}}{N} \right). \tag{5}$$

This is a very typical assumption and is satisfied under summability conditions of the cummulants Brillinger and Rosenblatt (1967a) or under certain mixing condition assumptions Lii and Rosenblatt (1990a).

**Theorem 1** *Let  $\{X_t\}$  be an  $r$ -vector valued  $s$ th-order stationary time series with unknown mean  $\boldsymbol{\mu}$ . Let  $\widehat{f}(\boldsymbol{\omega})$  be the estimator as defined in (4) and assume (5) is satisfied.*

- (i) *Assume for some  $k \geq 1$ ,  $\sum_{\boldsymbol{\tau} \in \mathbb{Z}^{s-1}} \|\boldsymbol{\tau}\|^k |C(\boldsymbol{\tau})| < \infty$  and  $M \sim aN^c$  with  $c = (2k + s - 1)^{-1}$ , then*

$$\sup_{\boldsymbol{\omega} \in [-\pi, \pi]^{s-1}} \left| \text{bias} \left\{ \widehat{f}(\boldsymbol{\omega}) \right\} \right| = o \left( N^{\frac{-k}{2k+s-1}} \right) \tag{6}$$

and

$$\text{MSE}(\widehat{f}(\boldsymbol{\omega})) = O \left( N^{\frac{-2k}{2k+s-1}} \right).$$

- (ii) *Assume  $C(\boldsymbol{\tau})$  decreases geometrically fast, i.e.  $|C(\boldsymbol{\tau})| \leq D e^{-d\|\boldsymbol{\tau}\|}$ , for some positive constants  $d$  and  $D$  and  $M \sim A \log N$  where  $A \geq 1/(2db)$ , then*

$$\sup_{\boldsymbol{\omega} \in [-\pi, \pi]^{s-1}} \left| \text{bias} \left\{ \widehat{f}(\boldsymbol{\omega}) \right\} \right| = O \left( \frac{1}{\sqrt{N}} \right) \tag{7}$$

and

$$\text{MSE}(\hat{f}(\omega)) = O\left(\frac{\log N}{N}\right). \tag{8}$$

(iii) Assume  $C(\tau) = 0$  for  $\|\tau\| > q$  and let  $M$  be a constant such that  $bM \geq q$ , then

$$\sup_{\omega \in [-\pi, \pi]^{s-1}} \left| \text{bias} \left\{ \hat{f}(\omega) \right\} \right| = O\left(\frac{1}{N}\right)$$

and

$$\text{MSE}(\hat{f}(\omega)) = O\left(\frac{1}{N}\right).$$

*Remark 1* Equations (6), respectively (7), remain true with the assumptions on  $M$  replaced with  $M^{k+s-1}/N \rightarrow 0$ , respectively  $e^M M^{s-1}/N \rightarrow 0$ .

*Remark 2* Depending on the constant  $A$  in part (ii), the bias in (7) may be as small as  $O((\log N)^{s-1}/N)$ .

*Remark 3* We do not assume the mean  $\mu$  of the time series is known. This adds an extra term of order  $O(M^{s-1}/N)$  to the bias; see the proof of Theorem 1 in the appendix for further details.

*Remark 4* Traditional estimators using second-order lag-windows have bias convergence rates of order  $O(1/M^2)$  regardless of the three scenarios listed in Theorem 1. However when the spectral density is smooth enough, like in the case of an ARMA process (where  $C(\tau)$  decays exponentially), traditional estimators perform considerably worse. For example, estimation of the bispectrum of an ARMA process has an asymptotic MSE rate of  $N^{-2/3}$  in the traditional case, but an asymptotic MSE of  $(\log N)/N$  using flat-top lag-windows. The distinction is even more profound in estimating higher-order spectra where the best rate achieved is  $N^{-4/(3+s)}$  for traditional estimators and again  $(\log N)/N$  using flat-top lag-windows. Even in the worst case of polynomial decay, our proposed estimator still beats, or possibly ties with, traditional estimators in terms of asymptotic MSE rates.

The asymptotic analysis in Theorem 1 relies on having the appropriate bandwidth  $M$  based on the various decay rates  $C(\tau)$ . In the next section we propose an algorithm that, for the most part, automatically detects the correct decay rate of  $C(\tau)$  and supplies the practitioner with an asymptotically consistent estimate of  $M$ .

### 3 A bandwidth selection procedure

For  $\tau \in \mathbb{Z}^{s-1}$ , consider the normalized cumulant function

$$\rho(\tau) = \frac{C(\tau)}{(\prod_{i=1}^s C_{a_i}(0))^{1/2}}$$

with a natural estimator

$$\hat{\rho}(\boldsymbol{\tau}) = \frac{\widehat{C}(\boldsymbol{\tau})}{\left(\prod_{i=1}^s \widehat{C}_{a_i}(0)\right)^{1/2}}.$$

Let  $B_{x,y}$  ( $x, y > 0$ ) denote the set of indices in  $\mathbb{Z}^{s-1}$  contained in the half-open  $s - 1$ -dimensional annulus of inner radius  $x$  and outer radius  $y$ , i.e.

$$B_{x,y} = \{\boldsymbol{\tau} \in \mathbb{Z}^{s-1} : x < \|\boldsymbol{\tau}\| \leq y\}. \tag{9}$$

The following algorithm for estimating the bandwidth of a flat-top estimator is a multivariate extension of an algorithm proposed in Politis (2003).

**Bandwidth selection algorithm**

Let  $k > 0$  be a fixed constant, and  $a_N$  be a nondecreasing sequence of positive integers tending to infinity such that  $a_N = o(\log N)$ . Let  $\hat{m}$  be the smallest number such that

$$|\hat{\rho}(\boldsymbol{\tau})| < k \sqrt{\frac{\log_{10} N}{N}} \quad \text{for all } \boldsymbol{\tau} \in B_{\hat{m}, \hat{m}+a_N}. \tag{10}$$

Then let  $\hat{M} = \hat{m}/b$  (where  $b$  is the ‘‘flat-top radius’’ as defined by condition (i) of a flat-top lag-window).

*Remark 5* A norm was not specified in (9) and any norm may be used. The sup norm, for example, may be preferable to the Euclidean norm in practice since the region in (9) becomes rectangular instead of circular.

*Remark 6* The positive constant  $k$  is irrelevant in the asymptotic theory, but is relevant for finite-sample calculations. In order to determine an appropriate value of  $c$  for computation, we consider the following approximation

$$\sqrt{N} (\hat{\rho}(\boldsymbol{\tau}_0) - \rho(\boldsymbol{\tau}_0)) \sim \mathcal{N}(0, \sigma^2). \tag{11}$$

This approximation holds under general assumptions of the time series and for any fixed  $\boldsymbol{\tau}_0 \in \mathbb{Z}^{s-1}$ . The variance  $\sigma^2$  does not depend on the choice of  $\boldsymbol{\tau}_0$  provided  $\boldsymbol{\tau}_0$  is not a ‘‘boundary point’’; see Brillinger and Rosenblatt (1967b) for more details. Let  $\hat{\sigma}$  be the estimate of  $\sigma$  via a resampling scheme like the block bootstrap. A approximate pointwise 95% confidence bound for  $\rho(\cdot)$  is given by  $\frac{\pm 1.96 \hat{\sigma}}{\sqrt{N}}$ . Therefore if we let  $a_N = 5$ , then  $k = 2\hat{\sigma}$  generates an approximate 95% simultaneous confidence bound by Bonferroni’s inequality by noting that  $\sqrt{\log_{10} N} \approx 1.5$  for moderately sized  $N$ .

The bandwidth selected using the above procedure converges precisely to the ideal bandwidth in each of the three cases of Theorem 1, as is proved in the following theorem under the two natural assumptions in (12) and (13) below.

**Theorem 2** Assume conditions strong enough to ensure that for any fixed  $n$ ,

$$\max_{\tau \in B_{0,n}} |\hat{\rho}(\sigma + \tau) - \rho(\sigma + \tau)| = O_p \left( \frac{1}{\sqrt{N}} \right) \tag{12}$$

uniformly in  $\sigma$ , and for any  $M$ , that may depend on  $N$ , the following holds

$$\max_{\tau \in B_{0,M}} |\hat{\rho}(\sigma + \tau) - \rho(\sigma + \tau)| = O_p \left( \sqrt{\frac{\log M}{N}} \right) \tag{13}$$

uniformly in  $\sigma$ .

- (i) Assume  $C(\tau) \sim A \|\tau\|^{-d}$  for some positive constants  $A$  and  $d \geq 1$ . Then

$$\hat{M} \stackrel{P}{\sim} A_0 \frac{N^{1/2d}}{(\log N)^{1/2d}},$$

where  $A_0 = A^{1/d} / (k^{1/d} b)$ ; here  $A \stackrel{P}{\sim} B$  means  $A/B \rightarrow 1$  in probability.

- (ii) Assume  $C(\tau) \sim A \xi^{\|\tau\|}$  for some positive constant  $A$  and  $|\xi| < 1$ . Then

$$\hat{M} \stackrel{P}{\sim} A_1 \log N,$$

where  $A_1 = -1 / (b \log |\xi|)$ .

- (iii) Suppose  $C(\tau) = 0$  when  $\|\tau\| > q$ , but  $C(\tau) \neq 0$  for some  $\tau$  with norm  $q$ , then  $\hat{M} \stackrel{P}{\sim} q/b$ .

*Remark 7* Under general regularity conditions, (12) holds as does the even stronger assumption of  $\sqrt{N}$  asymptotic normality, and (13) holds from general theory of extremes of dependent sequences; refer to Leadbetter et al. (1983).

### 4 Bispectrum

Now we will focus on estimating the bispectrum using flat-top lag-windows. The third-order cumulant reduces to the third-order central moment with estimator given by (3). It is easily seen that the third-order central moment,  $C(\tau_1, \tau_2)$ , satisfies the following symmetry relations:

$$C(\tau_1, \tau_2) = C(\tau_2, \tau_1) = C(-\tau_1, \tau_2 - \tau_1) = C(\tau_1 - \tau_2, -\tau_2). \tag{14}$$

Naturally, we would expect the lag-window function,  $\lambda(\tau_1, \tau_2)$ , in the estimator (4), to possess the same symmetries. So if a lag-window  $\lambda$  does not a priori have the symmetries as in (14), we can construct a symmetrized version given by

$$\tilde{\lambda} = g(\lambda(x, y), \lambda(y, x), \lambda(-x, y - x), \lambda(y - x, -x), \lambda(x - y, -y), \lambda(-y, x - y)), \tag{15}$$

where  $g$  is any symmetric function (of its six variables); for example  $g$  could be the geometric or arithmetic mean. It is worth noting that the symmetrized version of  $\lambda$  is connected to the theory of group representations of the symmetric group  $S_3$ ; see Berg (2006) for more details.

Several choices of lag-windows are considered in Subba Rao and Terdik (2003) including the so-called “optimal window”,  $\lambda_{\text{opt}}$ , which is in some sense optimal among lag-windows of order 2; refer to Theorem 2 on page 43 of Subba Rao and Gabr (1984). This lag-window is defined as Saito and Tanaka (1985)

$$\lambda_{\text{opt}}(\tau_1, \tau_2) = \frac{8}{\alpha(\tau_1, \tau_2)^2} J_2(\alpha(\tau_1, \tau_2)),$$

where  $J_2$  is the second-order Bessel function of the first kind, and

$$\alpha(x, y) = \frac{2\pi}{\sqrt{3}} \sqrt{x^2 - xy + y^2}.$$

Although  $\lambda_{\text{opt}}$  is optimal among order 2 lag-windows, it is sub-optimal to higher-order lag-windows, such as flat-top lag-windows. Also, since  $\lambda_{\text{opt}}$  is not compactly supported, it has the potential of being computationally taxing.

We detail two simple flat-top lag-windows satisfying the symmetries in (14), but the supply of examples is limitless by (15). (Fig. 1) The first example is a right pyramidal frustum with the hexagonal base  $|x| + |y| + |x - y| = 2$ . We let  $c \in (0, 1)$  be the scaling parameter that dictates when the frustum becomes flat, that is, the flat-top boundary is given by  $|x| + |y| + |x - y| = 2c$ . The equation of this lag-window is given by

$$\lambda_{\text{rpf}}(\tau_1, \tau_2) = \frac{1}{1 - c} \lambda_{\text{rp}}(\tau_1, \tau_2) - \frac{c}{1 - c} \lambda_{\text{rp}}\left(\frac{\tau_1}{c}, \frac{\tau_2}{c}\right),$$

where  $\lambda_{\text{rp}}$  is the equation of the right pyramid with base  $|x| + |y| + |x - y| = 2$ , i.e.,

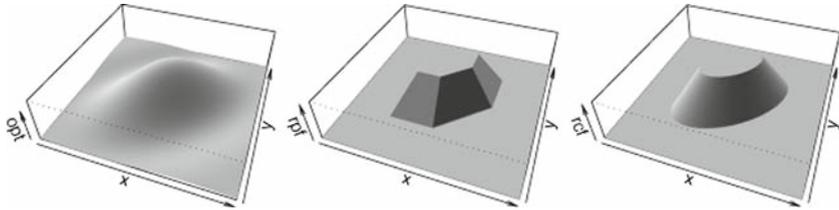
$$\lambda_{\text{rp}}(x, y) = \begin{cases} (1 - \max(|x|, |y|))^+, & -1 \leq x, y \leq 0 \text{ or } 0 \leq x, y \leq 1, \\ (1 - \max(|x + y|, |x - y|))^+, & \text{otherwise.} \end{cases}$$

The second flat-top lag-window that we propose is the right conical frustum with elliptical base  $x^2 - xy + y^2 = 1$ . As in the previous example, there is a scaling parameter  $c \in (0, 1)$ , and the lag-window becomes flat in the ellipse  $x^2 - xy + y^2 = c^2$ . The equation of this lag-window is given by

$$\lambda_{\text{rcf}}(\tau_1, \tau_2) = \frac{1}{1 - c} \lambda_{\text{rc}}(\tau_1, \tau_2) - \frac{c}{1 - c} \lambda_{\text{rc}}\left(\frac{\tau_1}{c}, \frac{\tau_2}{c}\right),$$

where  $\lambda_{\text{rc}}$  is the equation of the right cone with base  $x^2 - xy + y^2 = 1$ , i.e.,

$$\lambda_{\text{rc}}(x, y) = (1 - \sqrt{x^2 - xy + y^2})^+.$$



**Fig. 1** Plots of the three lag-windows,  $\lambda_{\text{opt}}$ ,  $\lambda_{\text{rpf}}$ , and  $\lambda_{\text{ref}}$  (with  $c = 1/2$  in the latter two)

Although in both examples the value for  $b$ , as defined in property (i) of the flat-top lag-window function, is smaller than the parameter  $c$ , the symmetries (14) permit us to only consider the region  $0 \leq y \leq x$  for which a circular arc of radius  $c$  does fit. So in the two examples above, we take the value of  $b$  to be the parameter  $c$ .

The bandwidth selection algorithm can be refined in the context of the bispectrum. The symmetries in (14) allow restriction to the region

$$\{(\tau_1, \tau_2) \in \mathbb{R}^2 \mid 0 \leq \tau_2 \leq \tau_1\}. \tag{16}$$

Here is the modified bandwidth selection algorithm for flat-top kernels that is tailored to the bispectrum:

**Practical bandwidth selection algorithm for the bispectrum**

Let  $\tilde{k} = k_1 > 0$  if  $n = 1$ , otherwise  $\tilde{k} = k_2 > 0$ , and let  $L$  be a positive integer that is  $o(\log N)$ . Order the points  $\{(\tau_1, \tau_2) \in \mathbb{Z}^2 \mid 0 < \tau_2 < \tau_1\} \cup \{(1, 0)\}$  with the usual lexicographical ordering, so  $P_1 = (1, 0)$ ,  $P_2 = (2, 1)$ ,  $P_3 = (3, 1)$ ,  $P_4 = (3, 2)$ , and so forth; in general,  $P_n = (i, j)$  where  $i = \lfloor (\frac{3}{2} + \sqrt{2n - 2}) \rfloor$  and  $j = n - \frac{1}{2}(i^2 - 3i) - 2$ . Let  $\hat{m}$  be the smallest number such that

$$|\hat{\rho}(P_{\hat{m}+\ell})| < \tilde{k} \sqrt{\frac{\log N}{N}} \quad \text{for all } \ell = 1, \dots, L. \tag{17}$$

Then let  $\hat{M} = (\text{first coordinate of } P_{\hat{m}}) / b = \left( \left\lfloor \left( \frac{3}{2} + \sqrt{2\hat{m} - 2} \right) \right\rfloor \right) / b$ .

*Remark 8* Except for the first point,  $(1, 0)$ , this algorithm does not incorporate boundary points since the asymptotic variance is larger on the boundary; the first point is included as there are no interior points with first coordinate equal to 1. The constant  $\tilde{k}$  is adjusted to account for the larger variance in the first point by providing a separate threshold,  $k_1$ , for this point.

*Remark 9* As suggested with the general algorithm, a subsampling procedure should be used to determine the appropriate constants  $k_1$  and  $k_2$ . However, one should be careful when choosing a point  $\tau_0$  for the approximation (11) since high variances at the origin and on the boundary tend to cause high variances near the origin and near the boundary in finite-sample scenarios. Therefore an interior point like  $(6, 3)$  (as opposed to  $(2, 1)$ ) should be used in determining  $k_2$ , and a point like  $(3, 0)$  (as opposed to  $(1, 0)$ ) should be used in determining  $k_1$ .

A modified bandwidth selection procedure is now proposed for use with the sub-optimal lag-windows of order 2. In this case, we propose using a bandwidth selection procedure based on the usual “solve-the-equation plug-in” approach Jones et al. (1996), but with flat-top estimates of the unknown quantities as the plug-in pilots. This will afford faster convergence rates of the bandwidth as compared estimates based on second-order pilots as well as solve the problem of selecting bandwidths for the pilots.

The optimal bandwidth at each point in the region (16), when using differentiable second order kernels, is derived in Subba Rao and Gabr (1984), and is given by

$$M_\lambda(\omega_1, \omega_2) = \left\{ \frac{\pi N}{\|\lambda\|_{L_2} f(\omega_1) f(\omega_2) f(\omega_1 + \omega_2)} \left( \frac{\partial^2 \lambda(\tau_1, \tau_2)}{\partial \tau_1 \partial \tau_1} \Big|_{\tau_1 = \tau_2 = 0} \right)^2 \right. \\ \left. \times \left| \left( \frac{\partial^2}{\partial \omega_1^2} - \frac{\partial^2}{\partial \omega_1 \partial \omega_2} + \frac{\partial^2}{\partial \omega_2^2} \right) f(\omega_1, \omega_2) \right|^2 \right\}^{\frac{1}{6}} \tag{18}$$

Estimates of the spectral density using flat-top lag-windows is discussed above, and estimating the partial derivatives of the bispectrum follow similarly. For instance, the three second order partial derivatives needed in (18) can be estimated by

$$\widehat{f_{\omega_i, \omega_j}}(\omega_1, \omega_2) = \frac{\partial^2}{\partial \omega_i \partial \omega_j} \hat{f}(\omega_1, \omega_2) \\ = \frac{1}{(2\pi)^2} \sum_{\tau_1 = -N}^N \sum_{\tau_2 = -N}^N \tau_i \tau_j \lambda_M(\tau_1, \tau_2) \widehat{C}(\tau_1, \tau_2) e^{-i\tau \cdot \omega} \quad (i, j = 1, 2). \tag{19}$$

By mimicking the proof of Theorem 1, the estimator in (19) has the same asymptotic performance as the estimator  $\hat{f}(\omega)$  in Theorem 1 but under a slightly stronger assumption for part (i) that  $\sum_{\tau \in \mathbb{Z}^2} \|\tau\|^{k+2} |C(\tau)| < \infty$ . We construct the estimator  $\widehat{M}_\lambda$  by replacing the unknown  $f$  and its derivatives in (18) with flat-top estimates. The next theorem provides convergence rates of the plug-in algorithm with flat-top pilots.

**Theorem 3** Assume conditions on  $\hat{\rho}$  such that (12) and (13) of Theorem 2 hold true, and assume conditions strong enough to ensure

$$\text{var} \left( \widehat{f_{\omega_i, \omega_j}} \right) = O \left( \frac{M^{s-1}}{N} \right) \quad (i, j = 1, 2). \tag{20}$$

(i) Assume  $C(\tau) \sim A \|\tau\|^{-d}$  for some positive constants  $A$  and  $d > s + 2$ . Then

$$\widehat{M}_\lambda = M_\lambda \left( 1 + O_p \left( \left( \frac{\log N}{N} \right)^{\frac{[d-s-2]}{2d}} \right) \right).$$

(ii) Assume  $C(\boldsymbol{\tau}) \sim A \xi^{\|\boldsymbol{\tau}\|}$  for some positive constant  $A$  and  $|\xi| < 1$ . Then

$$\widehat{M}_\lambda = M_\lambda \left( 1 + O_p \left( \left( \frac{\log N}{N} \right)^{\frac{1}{2}} \right) \right).$$

(iii) Suppose  $C(\boldsymbol{\tau}) = 0$  when  $\|\boldsymbol{\tau}\| > q$ , but  $C(\boldsymbol{\tau}) \neq 0$  for some  $\boldsymbol{\tau}$  with norm  $q$ , then

$$\widehat{M}_\lambda = M_\lambda \left( 1 + O_p \left( \frac{1}{\sqrt{N}} \right) \right).$$

*Remark 10* Certain mixing condition assumptions can guarantee (20); see Politis (2003) for an example.

In many cases, the convergence is a significant improvement over the traditional plug-in approach with second-order lag-window pilots. For example, the convergence of the bandwidth for data from an ARMA process would be  $M(1 + O_p(N^{-2/9}))$  using second-order pilots and techniques similar to Brockmann et al. (1993); Bühlmann (1996), but by using flat-top pilots, the convergence improves to  $M(1 + O_p(\sqrt{\log N/N}))$ .

### 5 Bispectral simulations

The three lag-windows detailed above— $\lambda_{\text{opt}}$ ,  $\lambda_{\text{rpf}}$ , and  $\lambda_{\text{rcf}}$ —are compared by their mean square error performance in estimating the bispectrum of four standard time series models. Three criteria are used to evaluate the performance of the bispectral estimates. The first two criteria are the estimators performance in estimating the bispectrum at the two points  $(0, 0)$  and  $(2, 1)$ . The bispectrum at the point  $(0, 0)$  is real-valued, and estimates typically have variances significantly larger than estimates at the interior point  $(2, 1)$  (exactly 30-times larger, asymptotically, if the second-order spectrum is flat). The bispectrum at the point  $(2, 1)$  is complex valued and performance is evaluated based on the estimation of the real part, complex part, and absolute value. The third criteria of evaluation is a composite evaluation of performance of the estimators over a rough grid of six points, standardized appropriately (further details below). The simulations are computed with data from the four stationary time series models: iid  $\chi_1^2$ , ARMA(1,1), GARCH(1,1), and bilinear(1,0,1,1). The first two are linear time series models whereas the last two are nonlinear models. Two sample sizes,  $N = 200$  and  $N = 2,000$ , are used throughout. Each simulation is averaged over 500 realizations.

The third criteria of evaluation, the composite evaluation is now described in further detail. The symmetries of  $C$  as given in (14) induce the following symmetries in the spectral density:

$$f(\omega_1, \omega_2) = f(\omega_2, \omega_1) = f(\omega_1, -\omega_1 - \omega_2) = f(-\omega_1 - \omega_2, \omega_2) = f^*(-\omega_1, -\omega_2).$$

The above symmetries in combination with the periodicity of  $f$  imply that  $f$  can be determined over the entire plane just by its values in the closed triangle  $T$  with vertices

$(0, 0)$ ,  $(\pi, 0)$ , and  $(2\pi/3, 2\pi/3)$ . So  $f$  is estimated at  $\binom{n-1}{2} = \frac{(n-1)(n-2)}{2}$  equally spaced points inside  $T$  with coordinates  $\omega_{ij} = \left(\frac{\pi(2i+2j)}{3n}, \frac{2\pi j}{3n}\right)$  where  $i = 1, \dots, n-1$  and  $j = 1, \dots, n-i-1$  (we take  $n = 5$  in the simulations).

The estimates at  $\omega_{ij}$  are standardized to make them comparable. Since, for  $(\omega_1, \omega_2)$  inside  $T$ , [Subba Rao and Gabr \(1984\)](#)

$$\text{var} \left( \hat{f}(\omega_1, \omega_2) \right) \approx \frac{M^2}{N} \frac{\|\lambda\|_{L_2}}{2\pi} f(\omega_1) f(\omega_2) f(\omega_1 + \omega_2),$$

$\hat{f}(\omega_1, \omega_2)$  is standardized by dividing it by  $\sqrt{f(\omega_1) f(\omega_2) f(\omega_1 + \omega_2)}$ . This leads to the composite evaluation of  $\hat{f}$  over a course grid of points by the quantity

$$\text{err}(\lambda) \triangleq \sum_{i=1}^{n-1} \sum_{j=1}^{n-i-1} \left| \frac{\hat{f}(\omega_{ij}) - f(\omega_{ij})}{\sqrt{f(\omega_{ij}^{(1)}) f(\omega_{ij}^{(2)}) f(\omega_{ij}^{(1)} + \omega_{ij}^{(2)})}} \right|$$

and the empirical MSE is calculated by averaging  $\text{err}(\lambda)^2$  over the 500 realizations.

In the tables of MSE estimates below, the first two rows are estimates from the flat-top lag-windows  $\lambda_{\text{rpf}}$  and  $\lambda_{\text{rcf}}$  with the bandwidth derived from the Bandwidth Selection Algorithm for the Bispectrum, as described above, with parameters  $L = 5$ ,  $c = 0.51$ , and  $k$  determined via the block bootstrap (see [Remarks 6 and 9](#)). The third and fourth rows are estimates using  $\lambda_{\text{opt}}$  with bandwidths from the plug-in method with flat-top pilots (f.p.) and second-order pilots (s.p.) respectively. The first column of each table concerns the estimation of the bispectrum at  $(0, 0)$ , taking absolute values if the estimate is complex valued. The next three columns concern the estimation of the real part, complex part, and absolute value of the bispectrum, respectively, at the point  $(2, 1)$ . The last column, labeled  $T_6$ , concerns the composite evaluation over a course grid of 6 points.

Simulations (based on 1,000 realizations) were conducted to determine the optimal finite-sample bandwidth with minimal MSE (checking up to a bandwidth size of 20). In the first three models—IID, ARMA, and GARCH—the optimal bandwidth is one under each evaluation criterion and for each lag-window considered. The estimators with best MSE performance in these models were the estimators with the best bandwidth selection procedure (the choice of lag-window was somewhat secondary). The bilinear model, however, had different optimal bandwidths depending on the particular evaluation criterion and lag-window. The optimal properties of the flat-top lag-window can easily be observed in this model as the optimal bandwidths are all larger than one.

### 5.1 IID data

Identical and independent  $\chi_1^2$  data is generated that has a central third moment  $\mu_3 = 8$ . Therefore the true bispectrum is  $f(\omega_1, \omega_2) \equiv \frac{\mu_3}{(2\pi)^2} \approx 0.202642$ . The following [Tables 1, 2, 3, 4](#) give the empirical MSE calculations of the estimated bispectrum over lengths  $N = 200$  and  $N = 2,000$  based on 500 simulations.

**Table 1** MSE estimates based on iid data for  $N = 200$  and  $N = 2,000$

	$ \hat{f}(0, 0) $	$\text{Re}\hat{f}(2, 1)$	$\text{Im}\hat{f}(2, 1)$	$ \hat{f}(2, 1) $	$T_6$
$N = 200$					
$\lambda_{\text{rpf}}$	0.02796	0.02061	3.131e-04	0.02093	709.4
$\lambda_{\text{rcf}}$	0.02778	0.02060	3.314e-04	0.02094	709.4
$\lambda_{\text{opt}}(\text{f.p.})$	0.02582	0.02086	3.577e-04	0.02122	709.8
$\lambda_{\text{opt}}(\text{s.p.})$	0.02806	0.02116	7.121e-04	0.02187	715.5
$N = 2,000$					
$\lambda_{\text{rpf}}$	2.887e-03	2.063e-03	1.799e-05	2.081e-03	71.19
$\lambda_{\text{rcf}}$	2.865e-03	2.064e-03	1.875e-05	2.083e-03	71.22
$\lambda_{\text{opt}}(\text{f.p.})$	2.616e-03	2.101e-03	2.085e-05	2.121e-03	71.23
$\lambda_{\text{opt}}(\text{s.p.})$	3.294e-03	2.184e-03	1.039e-04	2.288e-03	71.45

**Table 2** MSE estimates based on arma data for  $N = 200$  and  $N = 2,000$

	$ \hat{f}(0, 0) $	$\text{Re}\hat{f}(2, 1)$	$\text{Im}\hat{f}(2, 1)$	$ \hat{f}(2, 1) $	$T_6$
$N = 200$					
$\lambda_{\text{rpf}}$	6.102e-05	2.329e-05	4.468e-06	2.776e-05	313.3
$\lambda_{\text{rcf}}$	6.760e-05	2.435e-05	4.624e-06	2.897e-05	316.5
$\lambda_{\text{opt}}(\text{f.p.})$	4.422e-05	2.172e-05	5.235e-06	2.696e-05	302.8
$\lambda_{\text{opt}}(\text{s.p.})$	1.198e-04	3.088e-05	2.982e-05	6.070e-05	412.0
$N = 2,000$					
$\lambda_{\text{rpf}}$	2.997e-06	2.096e-06	6.896e-08	2.165e-06	24.21
$\lambda_{\text{rcf}}$	3.297e-06	2.137e-06	7.359e-08	2.210e-06	24.59
$\lambda_{\text{opt}}(\text{f.p.})$	3.129e-06	2.132e-06	2.796e-07	2.412e-06	24.74
$\lambda_{\text{opt}}(\text{s.p.})$	2.142e-05	4.222e-06	4.349e-06	8.571e-06	33.53

The flat-top estimators  $\lambda_{\text{rpf}}$ ,  $\lambda_{\text{rcf}}$ , and  $\lambda_{\text{opt}}(\text{f.p.})$  all outperform  $\lambda_{\text{opt}}(\text{s.p.})$  in every criterion considered. For  $N = 2,000$ , the proposed bandwidth procedures perform extremely well (refer to Sect. 5.5) producing the optimal bandwidth one over 95% of the time in each case.

### 5.2 ARMA model

The ARMA(1,1) model

$$X_t = 0.5X_{t-1} - 0.5Z_{t-1} + Z_t$$

is now considered where  $Z_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$ . This time series is Gaussian, so both the bispectrum and normalized bispectrum are identically zero.

**Table 3** MSE estimates based on garch data for  $N = 200$  and  $N = 2,000$

	$ \hat{f}(0, 0) $	$\text{Re} \hat{f}(2, 1)$	$\text{Im} \hat{f}(2, 1)$	$ \hat{f}(2, 1) $	$T_6$
$N = 200$					
$\lambda_{\text{rpf}}$	9.752e-04	5.462e-05	3.92e-05	9.383e-05	113.1
$\lambda_{\text{rcf}}$	1.038e-03	5.800e-05	4.391e-05	1.019e-04	115.1
$\lambda_{\text{opt}}(\text{f.p.})$	6.580e-04	4.345e-05	3.182e-05	7.527e-05	110.1
$\lambda_{\text{opt}}(\text{s.p.})$	3.849e-04	3.488e-05	5.112e-05	8.600e-05	125.1
$N = 2,000$					
$\lambda_{\text{rpf}}$	2.411e-05	2.916e-06	1.555e-06	4.471e-06	7.317
$\lambda_{\text{rcf}}$	2.682e-05	3.050e-06	1.745e-06	4.795e-06	7.401
$\lambda_{\text{opt}}(\text{f.p.})$	1.894e-05	2.528e-06	1.632e-06	4.159e-06	7.026
$\lambda_{\text{opt}}(\text{s.p.})$	5.781e-05	5.577e-06	7.577e-06	1.315e-05	9.021

The flat-top estimators and  $\lambda_{\text{opt}}(\text{f.p.})$  outperform  $\lambda_{\text{opt}}(\text{s.p.})$  even more significantly in this model for every criterion considered. Good performance is again mostly attributed to good bandwidth selection, but true optimal properties of the flat-top lag-windows are present and is addressed for the bilinear model.

### 5.3 GARCH model

We now consider the GARCH(1,1) model

$$\begin{cases} X_t = \sqrt{h_t} Z_t \\ h_t = \alpha_0 + \alpha_1 X_{t-1}^2 + \alpha_2 h_{t-1} \end{cases}$$

where  $\alpha = (0.1, 0.8, 0.1)$  and  $Z_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$ . The theoretical values of the bispectrum are unknown, so they are approximated via simulation over 500 realizations with  $N = 10^5$  and averaging the four estimators.

For  $N = 200$ ,  $\lambda_{\text{opt}}(\text{s.p.})$  performed best at the origin, but considerably worse in the composite criterion. For the larger  $N$ , the flat-top estimators and  $\lambda_{\text{opt}}(\text{f.p.})$  again performed significantly better than  $\lambda_{\text{opt}}(\text{s.p.})$ .

### 5.4 Bilinear model

Finally, we consider the BL(1,0,1,1) bilinear model [Subba Rao and Gabr \(1984\)](#)

$$X_t = aX_{t-1} + bX_{t-1}Z_{t-1} + Z_t$$

where  $a = b = 0.4$  and  $Z_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1)$ . The complete calculations of the bispectrum have been worked out in [Subba Rao and Gabr \(1984\)](#), however the given equation

**Table 4** MSE estimates based on bilinear data for  $N = 200$  and  $N = 2,000$

	$ \hat{f}(0, 0) $	$\text{Re}\hat{f}(2, 1)$	$\text{Im}\hat{f}(2, 1)$	$ \hat{f}(2, 1) $	$T_6$
$N = 200$					
$\lambda_{\text{rpf}}$	5.872 <sub>2,3</sub>	5.421e-04 <sub>4,5</sub>	1.008e-03 <sub>1,2</sub>	1.55e-03 <sub>4,5</sub>	806.8 <sub>2,5</sub>
$\lambda_{\text{rcf}}$	5.956 <sub>2,3</sub>	6.005e-04 <sub>6,5</sub>	1.073e-03 <sub>1,2</sub>	1.673e-03 <sub>6,5</sub>	817.0 <sub>7,6</sub>
$\lambda_{\text{opt}}$ (f.p.)	4.401 <sub>2,1</sub>	4.608e-04 <sub>4,3</sub>	9.654e-04 <sub>1,2</sub>	1.426e-03 <sub>4,3</sub>	807.1 <sub>5,4</sub>
$\lambda_{\text{opt}}$ (s.p.)	2.916 <sub>2,1</sub>	3.926e-04 <sub>4,3</sub>	8.623e-04 <sub>1,2</sub>	1.255e-03 <sub>4,3</sub>	791.4 <sub>5,4</sub>
$N = 2,000$					
$\lambda_{\text{rpf}}$	1.755 <sub>4,3</sub>	7.734e-05 <sub>4,6</sub>	9.867e-05 <sub>1,2</sub>	1.76e-04 <sub>4,6</sub>	71.76 <sub>2,5</sub>
$\lambda_{\text{rcf}}$	1.891 <sub>4,3</sub>	7.792e-05 <sub>6,7</sub>	1.012e-04 <sub>1,2</sub>	1.791e-04 <sub>6,7</sub>	74.69 <sub>6,7</sub>
$\lambda_{\text{opt}}$ (f.p.)	2.119 <sub>4,3</sub>	6.282e-05 <sub>5,6</sub>	9.443e-05 <sub>1,2</sub>	1.572e-04 <sub>5,4</sub>	71.01 <sub>6,7</sub>
$\lambda_{\text{opt}}$ (s.p.)	1.322 <sub>4,3</sub>	5.123e-05 <sub>5,6</sub>	8.064e-05 <sub>1,2</sub>	1.319e-04 <sub>5,4</sub>	72.83 <sub>6,7</sub>

**Table 5** MSE estimates at the origin with bandwidths one through seven and  $N = 200$

$N = 200$	1	2	3	4	5	6	7
$\lambda_{\text{rpf}}$	2.062	<b>1.389</b>	1.71	2.879	4.216	5.849	7.22
$\lambda_{\text{rcf}}$	2.062	<b>1.390</b>	1.864	3.207	4.848	6.502	8.078
$\lambda_{\text{opt}}$	1.823	<b>1.445</b>	2.013	3.13	4.448	5.733	6.852

for the bispectrum does not match-up with simulated estimates. Therefore theoretical values of the bispectrum were computed through simulations as done in the GARCH model. The spectral density equation provided in [Subba Rao and Gabr \(1984\)](#) is correct and was used.

Whereas the previous three models had an optimal bandwidth of one throughout, the optimal bandwidths for the bilinear model is typically much larger and depends on the evaluation criterion considered. The subscripted numbers represent the best and second best bandwidth for each window (as deduced from simulation).

For this model, the flat-top estimators do not perform as well as the other estimators, but the superior asymptotics of the flat-top estimators kick in with larger  $N$  making all the estimators mostly equivalent when  $N = 2,000$ . The particularly good performance of  $\lambda_{\text{opt}}$  (s.p.) at the origin is due to a fortuitous bandwidth selection under sensitive conditions; this is addressed in more detail below.

The bispectrum corresponding to bilinear model resembles a hill peaking at the origin [Subba Rao and Gabr \(1984\)](#). This causes the choice of bandwidth to be particularly delicate when estimating the origin. The following table depicts this delicacy.

The simulations up to this point mostly depict the strength of the bandwidth selection procedure, and not the general asymptotic optimality of the flat-top lag-window. However, if we consider MSE estimates for a fixed set of bandwidths, as in [Table 5](#), the flat-top estimates perform better than  $\lambda_{\text{opt}}$  which improves with  $N$ . [Table 6](#) demonstrates the increased performance at  $N = 2,000$ .

**Table 6** MSE estimates at the origin with bandwidths one through eight and  $N = 2,000$

$N = 2,000$	1	2	3	4	5	6	7	8
$\lambda_{\text{rpf}}$	2.029	0.9465	0.552	<b>0.4687</b>	0.6002	0.8262	1.029	1.237
$\lambda_{\text{rcf}}$	2.029	0.9082	0.5074	<b>0.4917</b>	0.6821	0.9224	1.156	1.346
$\lambda_{\text{opt}}$	1.736	0.8919	0.5444	<b>0.5267</b>	0.6444	0.8001	0.9579	1.099

**Table 7** MSE of  $\hat{M}/M - 1$  for bandwidth selection procedures (a)–(e)

$N$	IID		ARMA		GARCH		Bilinear <sup>d</sup>	
	200	2000	200	2000	200	2000	200	2000
(a)	3.18	0.792	1.54	0.248	6.36	0.968	0.413	0.182
(b)	0.862	0.276	0.232	.050	2.59	0.292	1.63	0.454
(c)	2.71	0.900	0.866	0.142	4.05	0.552	0.633	0.362
(d)	1.45	3.96	1.27	3.22	1.19	3.49	0.185	0.414
(e)	4.66	12.0	4.04	9.36	4.22	9.82	0.0706	0.0394

<sup>a</sup> Bandwidths 5 and 6 were selected as theoretical bandwidths for procedure (a), but this is only approximate as the optimal bandwidth varies. True theoretical bandwidths can be inferred from Table 4

Further illustration of the optimality of the flat-top lag-windows is provided in Politis and Romano (1995) where second-order spectral density estimation with flat-top lag-windows is addressed.

### 5.5 Analysis of bandwidth procedures

Simulations are used to evaluate the various bandwidth selection procedures for the bispectrum. The five bandwidth selection procedures considered are described below.

- (a) Practical bandwidth selection algorithm for the bispectrum of Sect. 4
- (b) Plug-in method at the origin with flat-top pilots
- (c) Plug-in method at the point (2,1) with flat-top pilots
- (d) Plug-in method at the origin with second-order pilots
- (e) Plug-in method at the point (2,1) with second-order pilots.

The pilot estimates for (b) and (c) were derived from the flat-top lag-windows  $\lambda_{\text{rpf}}$  and the trapezoidal flat-top window Politis and Romano (1995). The bandwidths for these pilot estimators are derived from the bandwidth selection algorithm of Sect. 3. The Parzen and optimal lag-windows were used as pilots in (d) and (e) with bandwidths  $\lfloor N^{1/5} \rfloor$  and  $\lfloor N^{1/6} \rfloor$  respectively.

A summary of their performance is tabulated in Table 7.

We see that the simple bandwidth selection algorithm is very effective in producing accurate bandwidths that are consistent. The bandwidth selection procedure (a) can be seen to be quite accurate but tends to produce a few relatively large bandwidths. This error is compounded when squared error loss is used to evaluate the performance. The plug-in method with second-order pilots on the other hand performs very poorly and

does not even appear consistent. Histograms of the five bandwidth selection procedures are available on Politis’ website: <http://www.math.ucsd.edu/~politis/PolyspectraHist.pdf>.

### 6 Conclusions

Flat-top kernels in higher-order spectral density estimation is shown to be asymptotically superior in terms of MSE to any other finite-order kernel estimators. In addition, a very simple bandwidth selection algorithm is included that delivers ideal bandwidths tailored to the flat-top estimators. If one chooses not to adopt the infinite-order flat-top lag-window, then bandwidth selection via the plug-in method with flat-top pilots demonstrates greatly increased performance and should be used. Finite-sample simulations show these flat-top estimators were comparable with, and in many cases outperforming, the popular second-order “optimal” lag-window estimator using the plug-in method with second-order pilots for bandwidth selection. Simulations show the estimation of the bispectrum is quite sensitive to the choice of bandwidth, and this paper delivers the first higher-order accurate bandwidth selection procedures for the bispectrum.

### Appendix A: Technical proofs

*Proof of Theorem 1.* Using property (iii) of the lag-window and the fact that  $E[\widehat{C}(\boldsymbol{\tau})] = C(\boldsymbol{\tau}) + O(1/N)$ , the expectation of  $\widehat{f}(\boldsymbol{\omega})$  can be expressed as

$$\begin{aligned} E[\widehat{f}(\boldsymbol{\omega})] &= \frac{1}{(2\pi)^{s-1}} \sum_{\|\boldsymbol{\tau}\| < N} \left( C(\boldsymbol{\tau}) + O\left(\frac{1}{N}\right) \right) \lambda_M(\boldsymbol{\tau}) e^{-i\boldsymbol{\tau} \cdot \boldsymbol{\omega}} \\ &= \frac{1}{(2\pi)^{s-1}} \sum_{\|\boldsymbol{\tau}\| < N} C(\boldsymbol{\tau}) \lambda_M(\boldsymbol{\tau}) e^{-i\boldsymbol{\tau} \cdot \boldsymbol{\omega}} + O\left(\frac{M^{s-1}}{N}\right). \end{aligned}$$

The bias of  $\widehat{f}(\boldsymbol{\omega})$  is

$$\begin{aligned} E[\widehat{f}(\boldsymbol{\omega})] - f(\boldsymbol{\omega}) &= \underbrace{\frac{1}{(2\pi)^{s-1}} \sum_{\|\boldsymbol{\tau}\| < N} (\lambda_M(\boldsymbol{\tau}) - 1) C(\boldsymbol{\tau}) e^{-i\boldsymbol{\tau} \cdot \boldsymbol{\omega}}}_{A_1} \\ &\quad \times \underbrace{-\frac{1}{(2\pi)^{s-1}} \sum_{\|\boldsymbol{\tau}\| \geq N} C(\boldsymbol{\tau}) e^{-i\boldsymbol{\tau} \cdot \boldsymbol{\omega}}}_{A_2} + O\left(\frac{M^{s-1}}{N}\right). \end{aligned}$$

By the assumption on the summability of  $C(\boldsymbol{\tau})$ ,  $|A_2|$  can be bounded as

$$|A_2| \leq \frac{1}{(2\pi)^{s-1}} \sum_{\|\boldsymbol{\tau}\| \geq N} |C(\boldsymbol{\tau})| \leq \frac{1}{(2\pi)^{s-1} N^k} \sum_{\|\boldsymbol{\tau}\| \geq N} \|\boldsymbol{\tau}\|^k |C(\boldsymbol{\tau})| = o\left(\frac{1}{N^k}\right).$$

Now rewrite  $A_1$  as

$$A_1 = \underbrace{\frac{1}{(2\pi)^{s-1}} \sum_{\|\boldsymbol{\tau}\| \leq bM} (\lambda_M(\boldsymbol{\tau}) - 1) C(\boldsymbol{\tau}) e^{-i\boldsymbol{\tau} \cdot \boldsymbol{\omega}}}_{=0} + \frac{1}{(2\pi)^{s-1}} \sum_{bM < \|\boldsymbol{\tau}\| \leq N} (\lambda_M(\boldsymbol{\tau}) - 1) C(\boldsymbol{\tau}) e^{-i\boldsymbol{\tau} \cdot \boldsymbol{\omega}}.$$

□

*Proof of (i).* Since  $|\lambda(s)| \leq 1$ ,

$$|A_1| \leq \frac{2}{(2\pi)^{s-1}} \sum_{bM < \|\boldsymbol{\tau}\| \leq N} |C(\boldsymbol{\tau})| \leq \frac{2}{(2\pi)^{s-1} (bM)^k} \sum_{bM < \|\boldsymbol{\tau}\| \leq N} \|\boldsymbol{\tau}\|^k |C(\boldsymbol{\tau})| = o\left(\frac{1}{M^k}\right).$$

Equation (6) now follows, and thus  $\text{MSE}(\hat{f}(\boldsymbol{\omega})) \sim o(M^{-2k}) + O(M^{s-1}/N)$ . □

*Proof of (ii).* We have  $\text{bias}(\hat{f}(\boldsymbol{\omega})) = A_1 + O(M^{s-1}/N)$ , where under the assumptions of (ii),

$$|A_1| \leq \frac{2}{(2\pi)^{s-1}} \sum_{bM < \|\boldsymbol{\tau}\| \leq N} |C(\boldsymbol{\tau})| \leq \frac{(2) D}{(2\pi)^{s-1} e^{dbM}} \times \sum_{bM < \|\boldsymbol{\tau}\| \leq N} e^{d(bM - \|\boldsymbol{\tau}\|)} = O\left(e^{-dbM}\right).$$

Therefore  $\text{MSE}(\hat{f}(\boldsymbol{\omega})) \sim O(e^{-2dbM}) + O(M^{s-1}/N)$  is asymptotically minimized when  $M \sim A \log N$  where  $A = 1/(2db)$ , and (8) holds for all  $A \geq 1/(2db)$ . □

*Proof of (iii).* We have  $\text{bias}(\hat{f}(\boldsymbol{\omega})) = A_1 + O(M^{s-1}/N)$ , but under the assumptions of (iii),  $A_1 = 0$ . Hence the bias and variance are  $O(1/N)$ .

*Proof of Theorem 2.* Let  $\boldsymbol{\tau}_{\hat{m}}$  be any element of norm  $\hat{m}$  for which

$$|\hat{\rho}(\boldsymbol{\tau}_{\hat{m}})| > k \sqrt{\frac{\log N}{N}} \tag{21}$$

and let  $\tau'_{\hat{m}} \in B_{\hat{m}, \hat{m}+1}$ , so that  $\hat{m} < \|\tau'_{\hat{m}}\| \leq \hat{m} + 1$ , and

$$|\hat{\rho}(\tau'_{\hat{m}})| < k\sqrt{\frac{\log N}{N}}. \tag{22}$$

Equations (10) and (13) give

$$|\hat{\rho}(\tau_{\hat{m}})| = |\rho(\tau_{\hat{m}})| + o_p\left(\sqrt{\frac{\log \log N}{N}}\right). \tag{23}$$

In part (i),  $\rho(\tau) \sim A\|\tau\|^{-d}$ , so for any  $\epsilon > 0$ , we can find  $\tau_0$  such that

$$A(1 - \epsilon)\|\tau\|^{-d} < \rho(\tau) < A(1 + \epsilon)\|\tau\|^{-d} \tag{24}$$

when  $\|\tau\| > \tau_0$ . Similarly, for any  $\epsilon > 0$ , there exists  $\tau_0$  large enough such that

$$(1 - \epsilon)\hat{m}^{-d} < \|\tau\|^{-d} < (1 + \epsilon)\hat{m}^{-d} \quad \text{for all } \tau \in B_{\hat{m}, \hat{m}+1} \tag{25}$$

when  $\hat{m} > \tau_0$ . Putting Eqs. (21), (22), (23), (24), and (25) together gives, with high probability,

$$A(1 - \epsilon)^2\hat{m}^{-d} < k\sqrt{\frac{\log N}{N}} < A(1 + \epsilon)^2\hat{m}^{-d} \tag{26}$$

up to  $o_p(\sqrt{(\log \log N)/N})$ , which is negligible as  $N$  gets large. Equation (26) is equivalent to

$$\frac{\hat{m}}{(1 + \epsilon)^2} < \frac{A^{1/d}N^{1/2d}}{k^{1/d}(\log N)^{1/2d}} < \frac{\hat{m}}{(1 - \epsilon)^2}$$

with high probability. Therefore

$$\hat{m} \stackrel{P}{\sim} \frac{A^{1/d}N^{1/2d}}{k^{1/d}(\log N)^{1/2d}}.$$

The proof of part (ii) is similar. □

Now we prove part (iii). Note that  $\hat{m} > q$  only if

$$\max_{\tau \in B_{q, q+a_N}} |\hat{\rho}(\tau) - \rho(\tau)| \geq k\sqrt{\frac{\log N}{N}} \tag{27}$$

but since  $C(\tau) = 0$  when  $\|\tau\| > q$ , Eq. (13) then shows

$$\max_{\tau \in B_{q, q+a_N}} |\hat{\rho}(\tau)| = o_p\left(\sqrt{\frac{\log \log N}{N}}\right) \tag{28}$$

since  $a_N = o(\log N)$ . The probability of (27) and (28) happening simultaneously tends to zero, hence  $P(\hat{m} > q) \rightarrow 0$ . Now if  $\hat{m} < q$  then

$$B_{\hat{m}, \hat{m} + a_N} |\hat{\rho}(\boldsymbol{\tau})| = |\rho(\boldsymbol{\tau})| + o_p\left(\sqrt{\frac{\log \log N}{N}}\right)$$

shows that (10) must eventually be violated, hence  $P(\hat{m} < q) \rightarrow 0$  and the result follows.

*Proof of Theorem 3.* Parts (ii) and (iii) follow from Theorems 1 and 2 and the  $\delta$ -method; see Politis (2003) for more details. For part (i), first note that

$$\sum_{\boldsymbol{\tau} \in \mathbb{Z}^{s-1} \setminus \{\mathbf{0}\}} \|\boldsymbol{\tau}\|^\alpha < \infty \text{ if and only if } \alpha > s - 1.$$

In order for  $\sum_{\boldsymbol{\tau} \in \mathbb{Z}^{s-1}} \|\boldsymbol{\tau}\|^{k+2} |C(\boldsymbol{\tau})| < \infty$ , for some  $k \geq 1$ ,  $d$  must satisfy  $d - k - 2 > s - 1$  or  $d > s + k + 1 \geq s + 2$ . Now the results of Theorem 1 hold for  $\widehat{f_{\omega_1, \omega_j}}$  in replace of  $\hat{f}(\omega_1, \omega_2)$  for any positive integer  $k < d - s - 1$ , in particular for  $k = \lceil d - s - 2 \rceil$ . From the proof of Theorem 1, the bias is of order  $o(1/M^k)$ , and since the variance is of smaller order, the result now follows from substituting  $M$  with the rate  $(N/\log N)^{1/2d}$  from Theorem 2 (i). □

### References

Berg, A. (2006). Multivariate lag-windows and group representations. Submitted for publication.

Brillinger, D.R., Rosenblatt, M. (1967a). Asymptotic theory of estimates of  $k$ th order spectra. In B. Harris (Ed.), *Spectral analysis of time series (Proc. advanced sem., madison, Wis., 1966)*, (pp. 153–188). New York: Wiley.

Brillinger, D.R., Rosenblatt, M. (1967b). Computation and interpretation of  $k$ th order spectra. In B. Harris (Ed.), *Spectral analysis of time series (Proc. advanced Sem., Madison, Wis., 1966)*, (pp. 189–232). New York: Wiley.

Brockmann, M., Gasser, T., Herrmann, E. (1993). Locally adaptive bandwidth choice for kernel regression estimators. *Journal of the American Statistical Association*, 88(424), 1302–1309. ISSN 0162-1459.

Bühlmann, P. (1996). Locally adaptive lag-window spectral estimation. *Journal of Time Series Analysis*, 17(3), 247–270. ISSN 0143-9782.

Hall, P., Marron, J.S. (1988). Choice of kernel order in density estimation. *The Annals of Statistics*, 16(1), 161–173. ISSN 0090-5364.

Hinich, M.J. (1982). Testing for Gaussianity and linearity of a stationary time series. *Journal of Time Series Analysis*, 3(3), 169–176. ISSN 0143-9782.

Jammalamadak, S.R., Rao, T.S., Terdik, G. (2006). Higher order cumulants of random vectors and applications to statistical inference and time series. *Sankhyā. The Indian Journal of Statistics*, 68(2), 326–356

Jones, M.C., Marron, J.S., Sheather, S.J. (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91(433), 401–407. ISSN 0162-1459.

Leadbetter, M.R., Lindgren, G., Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. New York: Springer ISBN 0-387-90731-9.

Lii, K.S., Rosenblatt, M. (1990a). Asymptotic normality of cumulant spectral estimates. *Journal of Theoretical Probability*, 3(2), 367–385. ISSN 0894-9840.

- Lii, K.S., Rosenblatt, M. (1990b). Cumulant spectral estimates: bias and covariance. In I. Berkes, E. Csáki, P. Révész (Eds.), *Limit theorems in probability and statistics (Pécs, 1989)*. vol. 57 of *Colloq. Math. Soc. János Bolyai* (pp. 365–405). Amsterdam: North-Holland.
- Politis, D.N. (2003). Adaptive bandwidth choice. *Journal of Nonparametric Statistics*, 15(4–5), 517–533. ISSN 1048-5252.
- Politis, D.N., Romano, J.P. (1995). Bias-corrected nonparametric spectral estimation. *Journal of Time Series Analysis*, 16(1), 67–103. ISSN 0143-9782.
- Saito, K., Tanaka, T. (1985). Exact analytic expression for gabr-rao's optimal bispectral two-dimensional lag window. *Journal of Nuclear Science and Technology*, 22(12), 1033–1035.
- Subba Rao, T., Gabr, M.M. (1980). A test for linearity of stationary time series. *Journal of Time Series Analysis*, 1(2), 145–158 ISSN 0143-9782.
- Subba Rao, T., Gabr, M.M. (1984). An introduction to bispectral analysis and bilinear time series models, *Lecture Notes in Statistics*, vol. 24. New York: Springer. ISBN 0-387-96039-2.
- Subba Rao, T., Terdik, Gy. (2003). On the theory of discrete and continuous bilinear time series models. In D.N. Shanbhag, C.R. Rao, (Eds.), *Stochastic processes: modelling and simulation*, vol. 21. *Handbook of Statistics*. (pp. 827–870). Amsterdam: North-Holland.