Nonlinear spectral density estimation: thresholding the correlogram

Efstathios Paparoditis
Dept. of Mathematics and Statistics
University of Cyprus
P.O.Box 20537
CY 1678 Nicosia, Cyprus

Dimitris N. Politis*
Department of Mathematics
University of California, San Diego
La Jolla, CA 92093-0112, USA
email: dpolitis@ucsd.edu

Original paper: May 2011; this revision: August 2011.

Abstract

Traditional kernel spectral density estimators are linear as a function of the sample autocovariance sequence. The purpose of the present paper is to propose and analyze two new spectral estimation methods that are based on the sample autocovariances in a nonlinear way. The rate of convergence of the new estimators is quantified, and practical issues such as bandwidth and/or threshold choice are addressed. The new estimators are also compared to traditional ones using flat-top lag-windows in a simulation experiment involving sparse time series models.

Key words: Autocovariance matrix, Flat-top lag-windows, Kernel smoothing, Sparsity, Thresholding, Wavelets.

*Research partially supported by NSF grant DMS-10-07513. Many thanks are due to Tim McMurry, Han Xiao, Wei Biao Wu, and Harrison Zhou for helpful suggestions, and to the associate editor and three anonymous referees for their constructive comments.

1 Introduction

Many applications of time series analysis involve the nonparametric estimation of the spectral density function; examples include astronomy, economics, electrical engineering, physics, etc. In fact, the foundation of statistical spectral analysis appears to have been laid by Albert Einstein (1914); see Brillinger (1993) for a historical perspective.

The prevalent spectral estimation method in the literature goes back to Bartlett (1946, 1948) and Daniell (1946) and can be represented in three equivalent ways: (a) kernel smoothing of the periodogram, (b) weighted/tapered Fourier series of the sample autocovariances, and (c) average of short (tapered) periodograms. The distinction between (b) and (c) lies, respectively, in the tapering of the sample autocovariance sequence vs. the actual data found in a short stretch of the series.

By the late 1950s the subject was already well understood as the early books by Grenander and Rosenblatt (1957), and Blackman and Tukey (1959) demonstrate. Influential papers at the time include Hannan (1957, 1958), Parzen (1957, 1961), and Priestley (1962). Many prominent researchers have since contributed to the subject. The book-length treatments in Hannan (1970), Brillinger (1981), Priestley (1981), Brockwell and Davis (1991), Rosenblatt (1985), and Percival and Walden (1993) contain a vast number of additional references.

A relatively recent development is the use of *flat-top* lag-windows defined and analyzed in Politis (2001, 2003, 2011). Flat-top lag-windows represent a generalization of the trapezoidal kernel of Politis and Romano (1995); they have been shown to achieve the optimal rate of convergence in a given smoothness class by automatically *adapting* to the underlying smoothness of the true spectral density. This adaptation has raised parallels with the wavelet literature—see e.g. Donoho and Johnstone (1995). Nevertheless, smoothing with flat-top lag-windows is a *linear* operation on the sample autocovariance sequence whereas typically wavelet methods are nonlinear; cf. Percival and Walden (2000). [Throughout the paper, the term 'linearity' will be understood as 'linearity with respect to the sample autocovariance sequence' not with respect to the data.]

Nonparametric spectral density estimation is an example of function smoothing. Hence,

it is not surprising that wavelet methods have found application here as well. A typical implementation of wavelet smoothing to the spectral estimation problem goes as follows: (a) expand the (logarithm of the) periodogram—or of a different preliminary spectral estimator—in a wavelet series; (b) threshold the wavelet coefficients; and (c) invert the wavelet transform to obtain a smoothed estimate of the (logarithm of the) spectral density; see Moulin (1994), Neumann (1996), Gao (1997), and Walden, Percival, and McCoy (1998).

The downside of the above wavelet method is that the wavelet coefficients are not interpretable, i.e., do not have a significance to the practitioner. By contrast, the autocovariance sequence—which represents the coefficients in a Fourier series expansion of the spectral density—is very meaningful and important in practice. Furthermore, the trigonometric basis is central to the analysis of stationary time series in view of the spectral representation of stationary processes, and the well-known Hilbert space isomorphism; cf. Lamperti (1977).

The purpose of the present paper is to propose and analyze two nonlinear spectral estimation methods that are based on the sample autocovariance. In effect, what we propose is to threshold the sample autocovariances, i.e., a wavelet-like approach using the original Fourier trigonometric orthogonal basis. In Section 2.1, the traditional (linear) spectral estimators are reviewed with an emphasis on flat-top lag windows. The two new nonlinear methods are presented and studied in subsections 2.2 and 2.3 respectively. Section 3 addresses some practical issues (positivity, choice of threshold, etc.), and additionally contains the results of a finite-sample simulation. Appendix A describes the empirical rule of choosing the bandwidth of a flat-top lag window; technical proofs are placed in Appendix B.

Finally, we note that there exist various alternative spectral estimation methods that are also nonlinear as functions of the correlogram. A prime example is estimating the spectral density by that of a fitted autoregressive (AR) model of appropriate degree. A review of nonlinear estimators that are popular in the engineering literature—including the estimators of Capon and Pisarenko—has been given in Haykin (1983). Asymptotic theory for the AR spectral estimators was given by Berk (1974), and for the Capon/Pisarenko estimators by Ioannidis (1994).

2 Spectral density estimation

Suppose $X_1, \ldots, X_n$ are observations from the (strictly) stationary real-valued sequence $\{X_t, t \in \mathbf{Z}\}$ having mean $\mu = EX_t$, and autocovariance sequence $\gamma(s) = E(X_t - \mu)(X_{t+|s|} - \mu)$ where both μ and $\gamma(\cdot)$ are unknown; also define the autocorrelation sequence $\rho(k) = \gamma(k)/\gamma(0)$. Our objective is the nonparametric estimation of the spectral density function $f(w) = (2\pi)^{-1} \sum_{s=-\infty}^{\infty} e^{iws}\gamma(s)$, for w $[-\pi, \pi]$.

2.1 Linear spectral estimation: a brief review

The traditional kernel estimator of $f(w)$ in its lag-window form is defined as:

$$\hat{f}(w) = (2\pi)^{-1} \sum_{s=-\infty}^{\infty} e^{iws} \lambda(\frac{s}{M}) \hat{\gamma}(s), \tag{1}$$

where $\hat{\gamma}(k) = n^{-1} \sum_{i=1}^{n-|k|} (X_i - \bar{X}_n)(X_{i+|k|} - \bar{X}_n)$ is the sample autocovariance at lag k; $\hat{\gamma}(k)$ is defined to be zero for $|k| \geq n$. Well-known choices for the lag-window $\lambda(s)$ have been proposed by Bartlett (1946), Parzen (1961), and Priestley (1962).

The family of 'flat-top' lag-windows was defined in Politis (2001) with typical member

$$\lambda_{g,c}(x) = \begin{cases} 1 & \text{if } |x| \leq c \\ g(x) & \text{else;} \end{cases} \tag{2}$$

here $c > 0$ is a shape parameter, and $g : \mathbb{R} \to [-1, 1]$ is a symmetric function, continuous at all but a finite number of points, and satisfying $g(c) = 1$, and $\int_c^{\infty} g^2(x)dx < \infty$. For better performance, the function $g(x)$ is typically taken nonnegative; see Politis (2011) for an exception based on a modification of the well-known Priestley-Epanechnikov lag-window. The trapezoidal lag-window of Politis and Romano (1995) $\lambda(x) = (\min\{1, 2(1 - |x|)\})^+$. is a prominent example of the 'flat-top' family; here, $(x)^+ = \max(x, 0)$.

It has been found—see e.g. Politis (2001, 2003, 2011)—that estimator (1) using a flat-top lag-window achieves the optimal rate of convergence in a given smoothness class; these optimal rates are delineated in Samarov (1977). Hence, a flat-top lag-window estimator exhibits *adaptivity* to the (unknown) degree of smoothness of the underlying true spectral

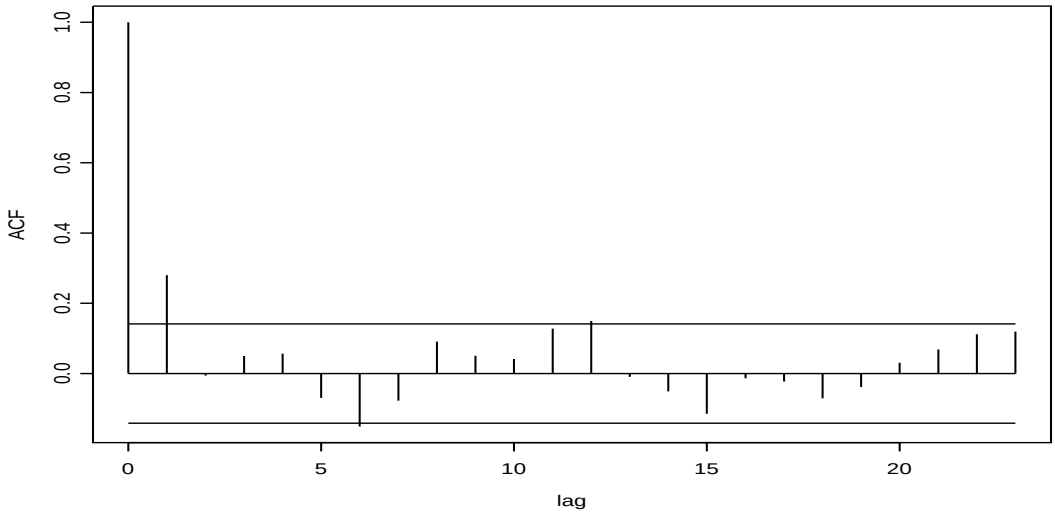

Figure 1: A 'problematic' correlogram from an AR(1) model: $X_t = \rho X_{t-1} + Z_t$ with $\rho = 0.3$ and $n = 500$; superimposed are the empirical rule bands $\pm 2\sqrt{\log_{10} n/n}$.

density; the degree of smoothness can be quantified by the rate of the decay of the autocovariance. Perhaps more importantly, these optimal rates are (almost) achieved even when an empirical data-dependent choice of the bandwidth M is employed—cf. Politis (2003).

The empirical rule of picking M is described in detail in Appendix A. It is based on the correlogram, i.e., a plot of the sample autocorrelation $\hat{\rho}(k) = \hat{\gamma}(k)/\hat{\gamma}(0)$ vs. k. The essence of the empirical rule lies in finding the smallest lag, say $\hat{q}$, after which the correlogram is not significantly different than zero. Although Appendix A gives a precise formula for $\hat{q}$, it should be stressed that the empirical rule of picking M should always be complemented by a visual inspection of the correlogram.

To see why an inspection of the correlogram is important, consider a situation where a practitioner finds that all $\hat{\rho}(k)$'s for $|k| > 1$ (say) are not significantly different than zero with the exception of (say) $\hat{\rho}(12)$ that is just borderline significant. For example, see Figure 1 depicting the 'problematic' correlogram of Politis and White (2004). While a strict application of the empirical rule with $C_0 = 2$ yields $\hat{q} = 12$, note that a slight tweaking of

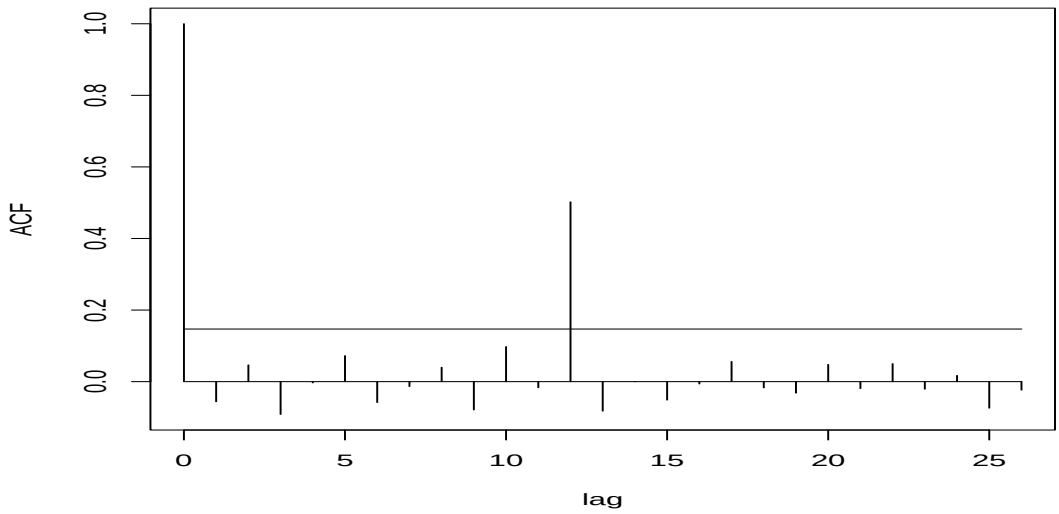

Figure 2: Correlogram from the sparse MA(12) model: $X_t = Z_t + Z_{t-12}$ with $n = 500$; superimposed are the empirical rule bands $\pm 2\sqrt{\log_{10} n/n}$.

the constant C_0 would instead give the answer $\hat{q} = 1$ that is both more reasonable in terms of explaining the autocorrelation structure and significantly better in terms of estimating the spectrum.

Another situation where a correlogram inspection can give crucial information is provided in the simple framework of an MA(q) model, i.e., when the series $\{X_t\}$ satisfies

$$X_t = \sum_{k=0}^{q} \theta_k Z_{t-k} \tag{3}$$

with $\{Z_t\}$ being a stationary white noise sequence. Figure 2 shows the correlogram of a sparse MA(12) model in which all $\hat{\rho}(k)$'s are within the empirical rule band of $\pm 2\sqrt{\log_{10} n/n}$ with the exception of $\hat{\rho}(12)$ whose deviation from zero is undeniably significant. The empirical bandwidth rule estimates $\hat{q} = 12$ which is accurate in view of $q = 12$ being the cut-off point after which all autocorrelations are negligible. However, if indeed it were true—as the correlogram suggests—that many (or all) $\hat{\rho}(k)$'s are negligible for $1 < k < 12$, then it is intuitive that the flat-top estimator $\hat{f}(w)$ would be suboptimal as compared to an esti-

mator that explicitly uses the information on the sparsity of the autocorrelation. Nonlinear estimation—based on thresholding—may give the solution here, and is addressed in the following section.

2.2 Nonlinear spectral estimation: thresholding

A Fourier series where the Fourier coefficient sequence is zero over large sets of consecutive integers is called *lacunary* in Harmonic Analysis; cf. Katznelson (2004). For example, consider the MA(q) model (3) with $\theta_k = 0$ for $0 < k < q$, i.e., a *sparse* MA(q) model. The implication is that $\gamma(k) = 0$ for $0 < |k| < q$, i.e., the only nonzero values of $\gamma(k)$ are $\gamma(0)$ and $\gamma(\pm q)$; see Figure 2 for an empirical example. If $q > 1$, then $\hat{f}(w)$ is seen to unnessecarily include several estimated $\hat{\gamma}(k)$, i.e., the ones with $0 < |k| < \hat{q}$, that ideally should have been set to zero.

So if/when the autocovariance sequence $\gamma(s)$ is lacunary, i.e., it has holes/zeros in-between its non-zero values, then the linear estimator $\hat{f}(w)$ might be suboptimal (in terms of variance) as compared to an 'oracle' estimator that exploits the knowledge regarding which of the $\gamma(s)$ are indeed zero. However, this 'oracle' estimator can be approximated in practice by an estimator that has an implicit test of significance of $\hat{\gamma}(s)$ built-in. With this in mind, we define the *thresholded* spectral density estimator

$$\tilde{f}(w) = (2\pi)^{-1} \sum_{s=-\infty}^{\infty} \tilde{\gamma}(s) e^{iws}; \tag{4}$$

in the above,

$$\tilde{\gamma}(s) = \hat{\gamma}(s) \mathbf{1}\{|\hat{\gamma}(s)| \geq T_n\} \tag{5}$$

where T_n is an appropriate threshold of significance, and $\mathbf{1}\{\cdot\}$ is the indicator function.

To analyze the properties of any spectral estimator it is important to quantify the strength of dependence of the underlying time series $\{X_t\}$. This can be done using cumulant summability conditions Brillinger (1981), moment and mixing conditions as in Rosenblatt (1984, 1985), etc. In the earlier literature, the assumption of linearity of the time series was prevalent; see e.g. Grenander and Rosenblatt (1957) or Hannan (1957, 1958). Recently,

Wu (2005) defined the so-called physical dependence measure conditions that are easy to work with by providing a (nonlinear) structure for $\{X_t\}$, and encompass a large variety of linear and nonlinear time series models; these conditions are summarized in the following assumption.

Assumption W. Assume that the process $\{X_t\}$ has the causal representation $X_t = g(\mathcal{F}^t)$ where g is some measurable function, $\{\epsilon_s,\ s \in \mathbf{Z}\}$ are i.i.d. random variables, and $\mathcal{F}^t$ is the σ-algebra generated by $\{\epsilon_s,\ s \leq t\}$. Let ϵ_0' be an i.i.d. copy of ϵ_0, and let $\mathcal{F}_*^t$ be the σ-algebra generated by $\{\cdots, \epsilon_{-1}, \epsilon_0', \epsilon_1, \cdots, \epsilon_t\}$, i.e., like $\mathcal{F}^t$ but with ϵ_0' in place of ϵ_0. Let $X_t^* = g(\mathcal{F}_*^t)$, and for $p > 1$ define $\delta_p(t) = (E|X_t - X_t^*|^p)^{1/p}$ and $\Theta_p(m) = \sum_{i=m}^{\infty} \delta_p(i)$. Furthermore, let $p' = \min(2, p)$, and define $\Psi_p(m) = \left(\sum_{i=m}^{\infty} \delta_p(i)^{p'}\right)^{1/p'}$, and $\Delta_p(m) = \sum_{i=0}^{\infty} \min\{C_p \Psi_p(m), \delta_p(i)\}$; here $C_p = \sqrt{p-1}$ if $p > 2$, and $C_p = 1/(p-1)$ if $1 < p \leq 2$.

Using Assumption W, Shao and Wu (2007) were able to prove many interesting asymptotic results for traditional (linear) spectral density estimates for linear/nonlinear processes that are sufficiently weakly dependent as measured by the rate of the decay of $\Theta_p(m)$ and $\Delta_p(m)$ as functions of m. We adopt a similar approach in the following theorem.

Theorem 2.1 *For some $p > 4$, assume $E|X_t|^p < \infty$ together with Assumption W with $\Theta_p(m) = O(m^{-\alpha})$ and $\Delta_p(m) = O(m^{-\alpha'})$ where $\alpha \geq \alpha' > 0$ are parameters such that $\alpha > 1/2$ and/or $\alpha' p > 2$. Let ζ be some number such that $\zeta > 1$, and define the threshold*

$$T_n = \zeta c_p' \sqrt{\log n / n} \tag{6}$$

to use in eq. (5) where $c_p' = 6(p+4)e^{p/4}(E|X_t|^4)^{1/4}\Theta_4(0)$. Then,

$$\sup_{w \in [-\pi, \pi]} |\tilde{f}(w) - f(w)| = O_P\left(\frac{\log n}{n}\right)^{\frac{\alpha}{2(1+\alpha)}}.$$

Remark 2.1 The proof of Theorem 2.1 hinges on a bound on the maximal deviation of the sample autocovariance derived by Xiao and Wu (2011); for more details see eq. (10) in Section 3.2. However, in the related problem of estimating the autocovariance matrix

via thresholding, Xiao and Wu (2011) use the threshold $T_n = 2c'_p\sqrt{\log n/n}$, i.e., they use $\zeta = 2$ in eq. (6). As the proof of Theorem 2.1 shows, the more general threshold (6) with $\zeta > 1$ is applicable here, and seems to also be applicable in the matrix setting of Xiao and Wu (2011). The practical applications of choice of threshold are discussed in more detail in Section 3.2.

Remark 2.2 To assess the strength of the assumptions of Theorem 2.1, note that the condition $\Theta_p(m) = O(m^{-\alpha})$ implies that $\sum_{k=m}^{\infty} |\gamma(k)| = O(m^{-\alpha})$. For example, if the series $\{X_t\}$ has a stationary ARMA representation, then condition $\Theta_p(m) = O(m^{-\alpha})$ holds true for *any* $\alpha > 0$. Letting α be an arbitrarily large number, we see that the rate of convergence of the nonlinear estimator $\tilde{f}(w)$ effectively becomes $\sqrt{n/\log n}$, thus attaining the minimax rate in this case; see Samarov (1977). In the case of polynomial decay of autocovariances, i.e., finite α, the result of Theorem 2.1 shows that $\tilde{f}(w)$ comes close to the minimax rate but falls short in claiming that it actually attains it. We conjecture that this is an artifact of our proof only, and that $\tilde{f}(w)$ might attain the minimax rate in this case as well.

2.3 Nonlinear spectral estimation: thresholding combined with a flat-top lag-window

As mentioned in Remark 2.2, it is unclear whether the nonlinear estimator $\tilde{f}(w)$ generally attains the minimax rate of convergence. Since we already have an estimator that does, namely $\hat{f}(w)$ with a flat-top lag-window from Section 2.1, it is interesting to ask if the flat-top estimator $\hat{f}(w)$ can be appropriately modified to efficiently handle lacunary/sparse autocovariance sequences. The answer is yes, and can be traced to the proof of Theorem 2.1. In particular, eq. (17) shows that we can discard large-lag autocovariances, and just focus on thresholding the autocovariances at small lags.

So consider a flat-top lag-window (2) that has finite support. Without loss of generality, we may assume that the support is $[-1, 1]$, i.e., that

$$\lambda_{g,c}(x) = 0 \quad \text{for} \quad |x| > 1 \tag{7}$$

in which case the shape parameter c is restricted to be in the interval $(0, 1]$. We now define the *hybrid* thresholded/flat-top estimator as

$$\check{f}(w) = (2\pi)^{-1} \sum_{s=-M}^{M} e^{iws} \lambda_{g,c}(\frac{s}{M}) \tilde{\gamma}_\delta(s) \tag{8}$$

where, for some $\delta \geq 0$ we let

$$\tilde{\gamma}_\delta(s) = \hat{\gamma}(s) \mathbf{1}\{|\hat{\gamma}(s)| \geq \delta T_n/\zeta\} \tag{9}$$

and $T_n = \zeta c_p' \sqrt{\log n/n}$ as in eq. (6).

For $\delta = \zeta$, there is full thresholding and $\tilde{\gamma}_\zeta(s) = \tilde{\gamma}(s)$ where the latter was defined in (5). The possibility for having $\delta < \zeta$ and even $\delta < 1$ in (9) is that estimator (8) is consistent—and with optimal rate of convergence—even with no thresholding, i.e., with $\delta = 0$. So δ here is a tuning parameter that will hopefully not influence the rate of convergence but may give a better finite-sample behavior in lacunary/sparse settings.

This discussion is made clear in the following theorem.

Theorem 2.2 *Assume the assumptions of Theorem 2.1. Assume eq. (6) and (7), and let* $M = \max(\lceil B_n/c_{ef} \rceil, 1)$ *where* $B_n = c_B(n/\log n)^{1/2(1+\alpha)}$ *for some constant* $c_B > 0$. *Then,*

$$\sup_{w \in [-\pi,\pi]} |\check{f}(w) - f(w)| = O_P\left(\frac{\log n}{n}\right)^{\frac{\alpha}{2(1+\alpha)}}.$$

Note that the choice of c_B is immaterial; only the rate of increase of B_n matters in the above. As with Theorem 2.1, Theorem 2.2 also falls slightly short of showing that $\check{f}(w)$ attains the minimax rate of convergence; this is not surprising since the method of proof is similar to that of Theorem 2.1—see also Remark 2.2.

If one were to use the truncated lag-window $\lambda_{g,c}(x) = \mathbf{1}\{x \in [-1, 1]\}$ together with the choice $\delta = \zeta$ in (9), then Theorem 2.2 would simply reduce to eq. (17) found in the proof of Theorem 2.1. However, the truncated lag-window is arguably the worst representative of the family of flat-top lag-windows as its performance can be quite erratic outside the realm of MA(q) models; see Politis (2011) for more details on choosing a flat-top lag-window. Furthermore, it may be advantageous to choose the tuning parameter δ to be less than ζ; see Section 3.3 for some finite-sample results to this effect.

3 Practical issues and a simulation experiment

3.1 Positivity issues

Neither the flat-top estimator $\hat{f}(w)$, nor its nonlinear counterparts $\tilde{f}(w)$ and $\check{f}(w)$, are almost surely nonnegative. There is an easy fix, however, namely taking the positive part.

Let $\bar{f}(w)$ denote an arbitrary estimator of $f(w)$, and define $\bar{f}^+(w) = \max(\bar{f}(w), 0)$. Noting that

$$|\bar{f}^+(w) - f(w)| \leq |\bar{f}(w) - f(w)|$$

for all $w \in [-\pi, \pi]$, it is apparent that $\bar{f}^+(w)$ is at least as accurate an estimator as $\bar{f}(w)$ is.

If one desires a strictly positive estimator, then—as in Politis (2011)—one can employ the estimator $\bar{f}^+_\epsilon(w) = \max(\bar{f}(w), \epsilon\hat{\gamma}(0)/n)$ where ϵ is some fixed positive number (typically less than one); the presence of $\hat{\gamma}(0)$ in the definition of $\bar{f}^+_\epsilon(w)$ is to make the estimator scale-equivariant.

The following Proposition is now immediate.

Proposition 3.1 *Let τ_n be a sequence such that $\sup_{w\in[-\pi,\pi]} |\bar{f}(w) - f(w)| = O_P(\tau_n)$. Then $\sup_{w\in[-\pi,\pi]} |\bar{f}^+(w) - f(w)| = O_P(\tau_n)$ as well. If $n\tau_n \to \infty$, then we additionally have $\sup_{w\in[-\pi,\pi]} |\bar{f}^+_\epsilon(w) - f(w)| = O_P(\tau_n)$ where $\bar{f}^+_\epsilon(w) = \max(\bar{f}(w), \epsilon\hat{\gamma}(0)/n)$ and ϵ is some fixed positive number.*

In other words, the nonnegative estimator $\bar{f}^+(w)$ and the positive estimator $\bar{f}^+_\epsilon(w)$ inherit the rate of convergence of the original estimator $\bar{f}(w)$.

3.2 Choosing the threshold parameters

In the nonlinear estimators of both Sections 2.2 and 2.3 the threshold $T_n = \zeta c'_p \sqrt{\log n/n}$ is used, where $\zeta > 1$ is some chosen number. Setting aside the choice of ζ, note that the constant c'_p is very difficult to estimate from data because it depends on the quantity $\Theta_4(0)$. Thus, there is a need for a practical, data-dependent choice of the threshold T_n.

Note that the reason the constant c_p' enters here is the fact that $c_p'\sqrt{\log n/n}$ is a bound for the maximum deviation of the sample autocovariances from their mean, i.e.,

$$Prob\{\max_{1\leq k\leq n}|\hat{\gamma}(k)-E\hat{\gamma}(k)|\leq c_p'\sqrt{\log n/n}\}\rightarrow 1 \;\;\text{as}\;\; n\rightarrow\infty; \tag{10}$$

see Lemma 6 of Xiao and Wu (2011). Nevertheless, the empirical bandwidth choice rule of Politis (2003) that is reviewed in Appendix A suggests estimating the aforementioned bound by $C_0\hat{\gamma}(0)\sqrt{\log_{10} n/n}$ with $C_0\simeq 2$. This implies estimating the threshold T_n by

$$\hat{T}_n = 2\zeta\hat{\gamma}(0)\sqrt{\log_{10} n/n}. \tag{11}$$

The sensitivity of the estimator $\tilde{f}(w)$ on the choice of ζ in connection with estimator (11) will be investigated in the finite-sample simulations of Section 3.3; a simple rule-of-thumb is to use $\zeta\simeq 1.5$.

The presence of the tuning parameter δ makes the hybrid estimator $\check{f}(w)$ of Section 2.3 less sensitive to the the choice of threshold T_n. The crucial thing in this case is estimation of the rate of increase of the quantity B_n, i.e., $(n/\log n)^{1/2(1+\alpha)}$, that depends on the unknown value of α. However, the empirical rule of Appendix A can again be called to the rescue. Recall that the optimal bandwith of the flat-top lag-window is proportional to $(n/\log n)^{1/(1+2\alpha)}$ under the assumptions of Theorem 2.2. As shown in Politis (2003), the estimator $\hat{q}$ of the empirical rule comes close to attaining this optimal rate, and this closeness is more pronounced when the degree of smoothness α is large, i.e., when the autocovariances decay reasonably fast. But then, in the case of large α, $\hat{q}$ can also be considered an estimator of the rate by which B_n is assumed to grow in Theorem 2.2.

Since the rate (and not the constant c_B) is important here, the recommendation is to use $\hat{q}$ as a *proxy* for B_n in Theorem 2.2; this recommendation is further re-inforced by the fact that the bandwidth choice there is given by $M=\max(\lceil B_n/c_{ef}\rceil, 1)$ which reduces to the empirical rule of Appendix A by simply replacing B_n with $\hat{q}$. The choice $B_n=\hat{q}$ is employed throughout the finite-sample simulations of Section 3.3 where the effect of choice of δ is also studied.

3.3 A simulation experiment

The finite-sample performance of the aforementioned spectral density estimators was investigated by means of a simulation experiment. The favorable performance of flat-top lag-windows *vis-a-vis* finite order lag-windows such as the Bartlett, Priestley, Parzen, etc. has been demonstrated both asymptotically and in finite-sample simulations—see Politis and Romano (1995), and Politis (2011). Hence, the current simulation focuses on comparing a standard flat-top estimator $\hat{f}$ to the thresholded estimators $\tilde{f}$ and $\check{f}$ proposed in this paper.

For simplicity, the trapezoidal lag-window of Politis and Romano (1995) was used in both $\hat{f}$ and $\check{f}$ together with the empirical bandwidth choice rule of Appendix A. Regarding $\tilde{f}$ and $\check{f}$, the estimated $\hat{T}_n$ from eq. (11) was used throughout, together with the choice $B_n = \hat{q}$ where $\hat{q}$ is defined in the empirical bandwidth choice rule.

Noting that $\hat{f}$ is a special case of $\check{f}$, the tables below include the flat-top estimator $\hat{f}$ in the column of $\check{f}$ with $\delta = 0$. For all estimators, the positive part was used as discussed in Section 3.1, i.e., $\hat{f}^+$, $\tilde{f}^+$ and $\check{f}^+$. R functions to compute the estimators $\hat{f}^+$, $\tilde{f}^+$ and $\check{f}^+$ are provided at: `http://www.math.ucsd.edu/~politis/SOFT/SpecEst.R`.

The simulation was based on 999 replications of time series from the following three models; in all three models, the sequence Z_t is i.i.d. $N(0, 1)$.

- **AR(1) model.** $X_t = \phi X_{t-1} + Z_t$ with $\phi \in \{-0.5, 0, 0.5, 0.9\}$.
- **Sparse MA(q) model.** $X_t = Z_t + Z_{t-q}$ with $q \in \{1, 3, 5, 10\}$.
- **Linear time series with strong dependence and sparsity.** $X_t = \sum_{k=0}^{\infty} \theta_k Z_{t-k}$ in the following four cases; here note that $b_k = 49/(k+7)^2$ for $k = 0, 1, \ldots$.

  A. $\theta_k = b_k$ for all $k = 0, 1, \ldots$.

  B. $\theta_k = b_k \mathbf{1}\{k \text{ is divisible by } 2\}$.

  C. $\theta_k = b_k \mathbf{1}\{k \text{ is divisible by } 3\}$.

  D. $\theta_k = b_k \mathbf{1}\{k \text{ is divisible by } 5\}$.

The AR(1) model was chosen as an easy setting in which thresholding is superfluous. Table 1a shows the empirical estimates of the mean (standardized) L_∞ error, i.e., of $E \sup_{w\in[-\pi,\pi]} \frac{|\hat{f}(w)-f(w)|}{f(w)}$ for the three estimators $\hat{f}^+$, $\tilde{f}^+$ and $\check{f}^+$ for different combinations of ζ and δ, and two different sample sizes, $n = 100$ and 400. Table 1b is similar to Table 1a, using the mean (standardized) integrated squared error (MISE) $E \int_{-\pi}^{\pi} \left(\frac{\hat{f}(w)-f(w)}{f(w)} \right)^2 dw$ to compare the estimators. In both Tables, the division by $f(w)$ is to account for the standard deviation of a general spectral estimator $\hat{f}(w)$ that is asymptotically proportional to $f(w)$. Note also that for the purposes of Tables 1a and 1b—as well as subsequent tables—the L_∞ and MISE errors were approximated by their discrete versions where both estimator and estimand are computed on the Fourier frequencies only.

The results of Tables 1a and 1b are not surprising. In the case $\phi = 0$, where the data are actually white noise, the thresholded estimator $\tilde{f}^+$ is best (with $\zeta = 2 \pm 1/2$) but all other estimators perform quite well here as well. In all other cases, i.e., in all cases where there is some genuine dependence in the data, the flat-top estimator $\hat{f}^+$ is best; the hybrid estimator $\check{f}^+$ with $\delta = 0.33$ comes a close second. Among the thresholded estimators in the non-white cases, the choice $\zeta = 1.5$ seems to be generally preferable. For the tables, an entry of 0.0000 indicates a value that is less than 0.00005.

By contrast to the AR(1) case, the sparse MA(q) model of Tables 2a and 2b presents a situation where thresholding should give improvements. Of course, the MA(1) case is not lacunary; so, as in the AR(1) case, the flat-top estimator $\hat{f}^+$ is best with the hybrid estimator $\check{f}^+$ with $\delta = 0.33$ coming a close second. However, when $q = 3$ or 5, the hybrid estimator $\check{f}^+$ with $\delta = 1$ is the unequivocal winner in the $n = 100$ case, indicating how thresholding together with a flat-top window can manifest a wonderful synergy. In the $n = 400$ case, the hybrid $\check{f}^+$ is a close second to the thresholded estimate $\tilde{f}^+$ with $\zeta \in [1.5, 2]$. The fact that the thresholded estimate $\tilde{f}^+$ is best here is not surprising in view of the fact that the controversial truncated lag-window $\lambda_{g,c}(x) = \mathbf{1}\{x \in [-1, 1]\}$ is actually optimal in the MA(q) case; see the discussion after Theorem 2.2.

The thresholded $\tilde{f}^+$ is also best—and this time by far—in the rather extreme case of

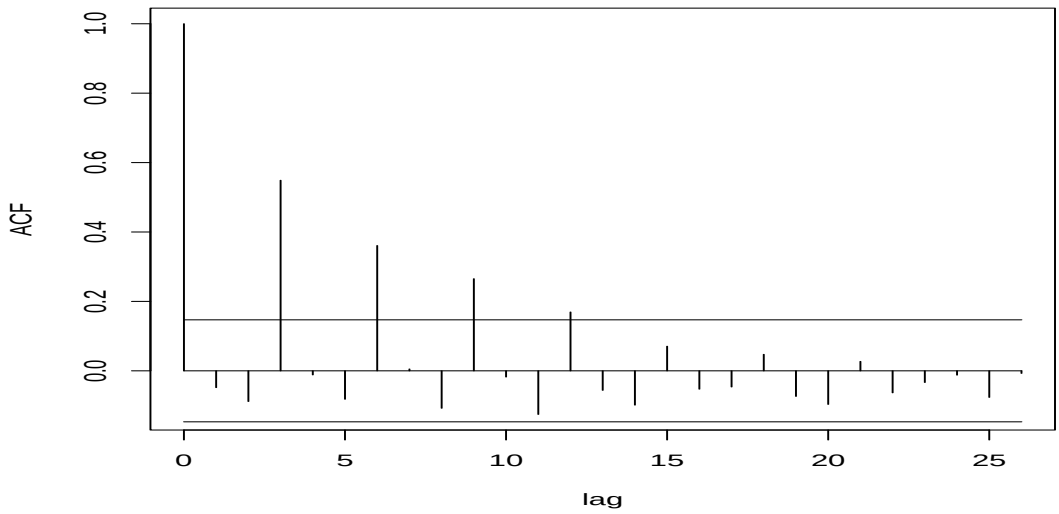

Figure 3: Correlogram from sparse linear time series of Model C with $n = 500$; superimposed are the empirical rule bands $\pm 2\sqrt{\log_{10} n/n}$.

$q = 10$ that was chosen to demonstrate a possible pitfall of the hybrid/flat-top set-up; this pitfall can be pin-pointed to the automatic application of the empirical rule of Appendix A whereby after K_n consecutive non-significant $\hat{\gamma}(k)$'s, the rule stops looking any further. Since $K_n \approx 5$ for sample sizes of the order of a few hundred, the empirical rule treats any sparse MA(q) with $q > 5$ as being white noise resulting in a bandwidth M that is too small. However, as discussed in Section 2.1, the empirical rule should always be complemented by an inspection of the correlogram in which case the significance of $\hat{\gamma}(q)$ should be obvious as a value of $\hat{\rho}(q) \simeq 1/2$ is too large to miss—see Figure 2 for an example.

The set-up of a linear time series with strong dependence is also devised to show-case the potential benefits of thresholding when sparsity is present. Figure 3 depicts a typical correlogram from Model C where the sparsity is apparent. Figure 4 shows, in log-scale, the true spectral density of a sparse linear time series of Model D that is characterized by several prominent peaks. As calculated from a typical data series with $n = 400$, both the thresholded estimator $\tilde{f}^+$ with $\zeta = 1.5$, and the hybrid estimator $\check{f}^+$ with $\delta = 0$, i.e.,

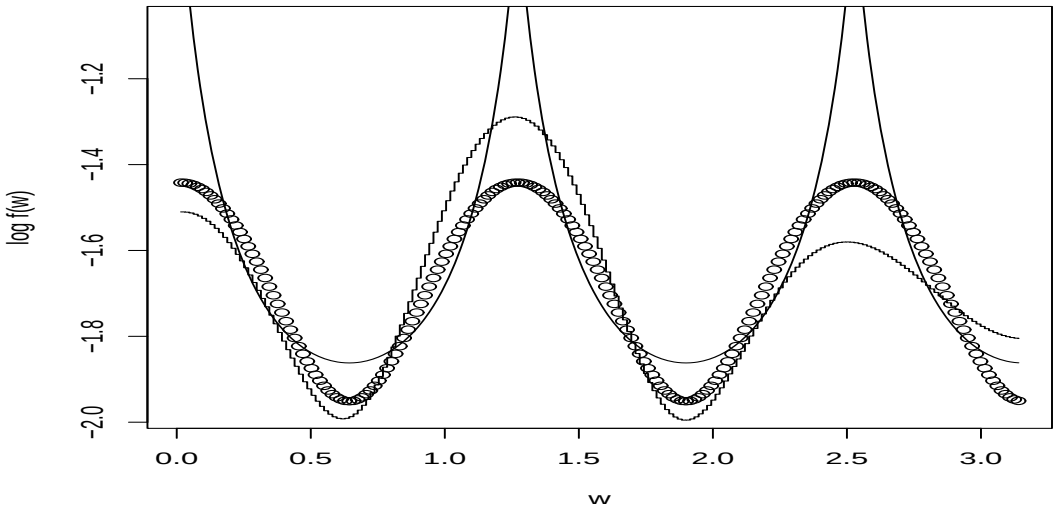

Figure 4: Log-spectral density of sparse linear time series of Model D from a typical data series with $n = 400$. Lines: true spectrum; Points: thresholded estimator $\tilde{f}^+$ with $\zeta = 1.5$; Jagged lines: hybrid estimator $\check{f}^+$ with $\delta = 0$, i.e., the flat-top estimator.

the flat-top estimator, have difficulty approximating these sharp peaks. Interestingly, the hybrid estimator $\check{f}^+$ with $\delta = 1$ is virtually indistinguishable from the thresholded $\tilde{f}^+$ with $\zeta = 1.5$ in the data example associated with Figure 4, and thus is not shown there.

As Tables 3a and 3b show, in case A where there is no sparsity, the flat-top estimator $\hat{f}^+$ is clearly the best. However, in the sparse cases B, C, D, the hybrid estimator $\check{f}^+$ with $\delta = 1$ gains prominence. What is also different here is the slower rate of convergence of all estimators involved; this is manifested by comparing the $n = 100$ to the $n = 400$ lines of Tables 3a and 3b that are within a standard error from each other.

AR(1)	Estimator	$\tilde{f}^+$	with	$\zeta =$	Estimator	$\check{f}^+$	with	$\delta =$
$n = 100$	1.001	1.5	2.0	2.5	0 (i.e., $\hat{f}^+$)	0.33	0.67	1.0
$\phi = -0.5$	0.4010	0.3644	0.4885	0.5051	**0.3027**	**0.3015**	0.3102	0.3253
0	0.0304	**0.0100**	**0.0100**	**0.0100**	0.0179	0.0173	0.0158	0.0150
0.5	0.3901	0.3738	0.4918	0.5050	**0.2971**	**0.3027**	0.3060	0.3234
0.9	1.4254	1.2676	1.3147	1.4030	**0.7350**	0.7455	0.8164	0.9289
$n = 400$								
$\phi = -0.5$	0.3197	0.2389	0.2872	0.2961	**0.1607**	0.1619	0.1714	0.1929
0	0.0130	**0.0025**	**0.0025**	**0.0025**	0.0040	0.0032	0.0031	0.0038
0.5	0.3124	0.2420	0.2866	0.2949	0.1589	**0.1544**	0.1671	0.1924
0.9	1.7118	1.2773	0.9636	0.9834	**0.4785**	0.4886	0.5260	0.6105

Table 1a. AR(1) model. Entries represent empirical estimates of the mean (standardized) L_∞ error, i.e., of $E \sup_{w\in[-\pi,\pi]} \frac{|\hat{f}(w)-f(w)|}{f(w)}$ for different estimators.

AR(1)	Estimator	$\tilde{f}^+$	with	$\zeta =$	Estimator	$\check{f}^+$	with	$\delta =$
$n = 100$	1.001	1.5	2.0	2.5	0 (i.e., $\hat{f}^+$)	0.33	0.67	1.0
$\phi = -0.5$	0.0519	0.0601	0.1161	0.1287	**0.0310**	**0.0308**	0.0337	0.0373
0	0.0032	**0.0001**	**0.0001**	**0.0001**	0.0012	0.0011	0.0009	0.0008
0.5	0.0501	0.0668	0.1187	0.1287	**0.0318**	**0.0321**	0.0335	0.0368
0.9	0.2857	0.2962	0.4040	0.5621	**0.0645**	0.0668	0.0890	0.1423
$n = 400$								
$\phi = -0.5$	0.0257	0.0233	0.0347	0.0381	**0.0075**	0.0080	0.0093	0.0123
0	0.0009	**0.0000**	**0.0000**	**0.0000**	**0.0001**	**0.0000**	**0.0000**	**0.0001**
0.5	0.0252	0.0239	0.0345	0.0379	**0.0076**	**0.0075**	0.0091	0.0124
0.9	0.2259	0.1927	0.1558	0.1893	**0.0165**	0.0180	0.0259	0.0450

Table 1b. AR(1) model. Entries represent empirical estimates of the mean (standardized) L_2 error, i.e., of $E \int_{-\pi}^{\pi} \left(\frac{\hat{f}(w)-f(w)}{f(w)} \right)^2 dw$ for different estimators.

MA(q)	Estimator	$\tilde{f}^+$	with	$\zeta =$	Estimator	$\check{f}^+$	with	$\delta =$
$n = 100$	1.001	1.5	2.0	2.5	0 (i.e., $\hat{f}^+$)	0.33	0.67	1.0
$q = 1$	0.2392	0.2846	0.9025	0.9799	0.1439	**0.1390**	0.1421	0.1445
3	0.2494	0.3156	0.9234	0.9799	0.2847	0.2898	0.2091	**0.1611**
5	0.2333	0.3482	0.9364	0.9800	0.4089	0.4073	0.3148	**0.1772**
10	**0.2663**	0.4339	0.9627	0.9800	1.0054	1.0079	0.9872	0.9793
$n = 400$								
$q = 1$	0.1960	**0.0573**	**0.0564**	0.0635	0.0639	0.0634	0.0609	0.0668
3	0.1856	0.0578	**0.0559**	0.0589	0.1345	0.1352	0.0932	0.0673
5	0.1808	0.0596	**0.0581**	0.0691	0.1945	0.1991	0.1291	0.0748
10	0.1940	**0.0606**	0.0626	0.0708	0.9868	0.9763	0.9792	0.9750

Table 2a. Sparse MA(q) model. Entries represent empirical estimates of the mean (standardized) L_∞ error, i.e., of $E \sup_{w\in[-\pi,\pi]} \frac{|\hat{f}(w)-f(w)|}{f(w)}$ for different estimators.

MA(q)	Estimator	$\tilde{f}^+$	with	$\zeta =$	Estimator	$\check{f}^+$	with	$\delta =$
$n = 100$	1.001	1.5	2.0	2.5	0 (i.e., $\hat{f}^+$)	0.33	0.67	1.0
$q = 1$	0.0240	0.0514	0.2013	0.2226	**0.0090**	**0.0088**	0.0094	0.0104
3	0.0254	0.0595	0.2072	0.2226	0.0238	0.0239	0.0170	**0.0125**
5	0.0232	0.0680	0.2108	0.2226	0.0372	0.0376	0.0293	**0.0144**
10	**0.0283**	0.0896	0.2179	0.2227	0.2204	0.2210	0.2195	0.2192
$n = 400$								
$q = 1$	0.0121	0.0012	**0.0011**	0.0031	0.0017	0.0016	0.0015	0.0021
3	0.0115	0.0013	**0.0011**	0.0018	0.0052	0.0052	0.0035	0.0021
5	0.0111	0.0013	**0.0012**	0.0041	0.0090	0.0090	0.0058	0.0026
10	0.0127	**0.0014**	**0.0014**	0.0041	0.2222	0.2195	0.2217	0.2217

Table 2b. Sparse MA(q) model. Entries represent empirical estimates of the mean (standardized) L_2 error, i.e., of $E \int_{-\pi}^{\pi} \left(\frac{\hat{f}(w)-f(w)}{f(w)} \right)^2 dw$ for different estimators.

LINEAR	Estimator	$\tilde{f}^+$	with	$\zeta =$	Estimator	$\check{f}^+$	with	$\delta =$
$n = 100$	1.001	1.5	2.0	2.5	0 (i.e., $\hat{f}^+$)	0.33	0.67	1.0
A.	1.0105	0.8609	0.9102	0.9467	**0.6163**	0.6203	0.6478	0.7116
B.	0.7769	0.4819	0.4891	0.5383	0.5177	0.4937	0.4939	**0.4707**
C.	0.4921	0.3780	0.4271	0.4776	0.4665	0.4519	0.3838	**0.3522**
D.	0.2985	0.2956	0.3393	0.4461	0.4183	0.3854	0.3017	**0.2576**
$n = 400$								
A.	1.3305	0.8085	0.6843	0.7605	**0.4693**	**0.4690**	0.4862	0.5552
B.	0.7978	0.4779	0.4860	0.5384	0.5346	0.5199	0.4881	**0.4815**
C.	0.4944	0.3714	0.4279	0.4763	0.4652	0.4485	0.3894	**0.3506**
D.	0.3001	0.2948	0.3440	0.4468	0.4279	0.3931	0.3040	**0.2611**

Table 3a. Sparse linear process. Entries represent empirical estimates of the mean (standardized) L_∞ error, i.e., of $E \sup_{w\in[-\pi,\pi]} \frac{|\hat{f}(w)-f(w)|}{f(w)}$ for different estimators.

LINEAR	Estimator	$\tilde{f}^+$	with	$\zeta =$	Estimator	$\check{f}^+$	with	$\delta =$
$n = 100$	1.001	1.5	2.0	2.5	0 (i.e., $\hat{f}^+$)	0.33	0.67	1.0
A.	0.1907	0.1922	0.2667	0.3493	**0.0537**	0.0551	0.0676	0.1088
B.	0.0898	0.0520	0.0657	0.0924	**0.0221**	**0.0227**	0.0298	0.0382
C.	0.0456	0.0345	0.0505	0.0654	**0.0223**	0.0242	0.0241	0.0246
D.	0.0208	0.0214	0.0337	0.0594	0.0237	0.0226	0.0185	**0.0157**
$n = 400$								
A.	0.1736	0.1112	0.1040	0.1417	**0.0166**	0.0171	0.0235	0.0421
B.	0.0930	0.0513	0.0640	0.0919	**0.0226**	0.0238	0.0289	0.0398
C.	0.0462	0.0339	0.0506	0.0648	**0.0224**	0.0234	0.0232	0.0247
D.	0.0209	0.0214	0.0346	0.0596	0.0241	0.0233	0.0187	**0.0162**

Table 3b. Sparse linear process. Entries represent empirical estimates of the mean (standardized) L_2 error, i.e., of $E \int_{-\pi}^{\pi} \left(\frac{\hat{f}(w)-f(w)}{f(w)} \right)^2 dw$ for different estimators.

To elaborate, in the previously discussed AR(1) and MA(q) case, the rate of decay of the quantity $\Theta_p(m)$ of Assumption W is exponential. In other words, we can take $\alpha \to \infty$ in the assumptions of Theorems 2.1 and 2.2, implying that—up to a logarithmic factor—the rate of convergence of all three estimators $\hat{f}^+$, $\tilde{f}^+$ and $\check{f}^+$ is the parametric rate $\sqrt{n}$. However, in the case of a linear time series with $\alpha = 2$, the rate of convergence of the flat-top estimator $\hat{f}^+$ is $n^{2/5}$. Now our Theorems 2.1 and 2.2 imply that $\tilde{f}^+$ and $\check{f}^+$ have a rate of convergence that is at worst $n^{1/3}$, and at best $n^{2/5}$ which is the minimax rate here. In either case, it is apparent that $n = 400$ is still a rather small sample size with a slow polynomial decay of the $\gamma(k)$'s.

All in all, the conclusion of the finite-sample simulation experiment is that the proposed estimators, the thresholded $\tilde{f}^+$ and the hybrid $\check{f}^+$, are both reasonably accurate with the hybrid seeming more versatile and generally preferrable. If one were to use the thresholded estimator $\tilde{f}^+$, then the choice $\zeta \approx 1.5$ seems to give good results overall. For the hybrid estimator $\check{f}^+$, the choice $\delta \approx 0.5$ seems to be the best compromise, i.e., providing some degree of thresholding that is useful in sparse set-ups while at the same time maintaining performance very close to that of the flat-top estimator $\hat{f}^+$ in non-sparse models.

4 Appendix A: Empirical rule of bandwidth choice for flat-top lag-windows

The empirical rule of picking the bandwidth parameter M for estimator (1) using a flat-top lag-window (2) was first proposed in Politis (2003) and further refined in Politis (2011). To describe it, we need to define the *'effective'* flat-top region of $\lambda_{g,c}$ as the interval $[-c_{ef}, c_{ef}]$ where

$$c_{ef} \;\text{ is the largest number such that }\; |\lambda_{g,c}(x) - 1| \leq \epsilon \;\text{ for all }\; x \in [-c_{ef}, c_{ef}]; \qquad (12)$$

here ϵ is some small chosen number, say $\epsilon = 0.01$. Notably, when the flat-top lag-window $\lambda_{g,c}(x)$ is very smooth near the origin, the effective flat-top region $[-c_{ef}, c_{ef}]$ can be quite bigger than the exact flat-top region $[-c, c]$; a prime example is the infinitely differentiable

flat-top lag-window of McMurry and Politis (2004).

EMPIRICAL RULE OF CHOOSING M FOR FLAT-TOP LAG-WINDOW $\lambda_{g,c}$. [Politis (2003, 2011)]

Let $\hat{q}$ be the smallest nonnegative integer such that $|\hat{\rho}(\hat{q}+m)| < C_0\sqrt{\log_{10} n/n}$ for $m = 0, 1, \ldots, K_n$, where $C_0 > 0$ is a fixed constant, and K_n is a positive, nondecreasing integer-valued function of n such that $K_n = o(\log n)$. Then, let $\hat{M} = \max(\lceil \hat{q}/c_{ef} \rceil, 1)$.

The constant C_0 and the form of K_n are the practitioner's choice; indeed, *any* values for $C_0 > 0$ and $1 \leq K_n = o(\log n)$ works equally well in terms of the first-order asymptotic Mean Squared Error of $\hat{f}(w)$. With the usual sample sizes of 100 to 1000, the recommendation $C_0 \simeq 2$ and $K_n \simeq 5$ seems to work well in practice, and is also backed by the interpretation of yielding (approximate) 95% simultaneous confidence intervals for $\rho(\hat{q}+m)$ for $m = 1, \ldots, K_n$ via Bonferroni's inequality. To handle bigger sample sizes, the choice $K_n = \max(5, 3\sqrt{\log_{10} n - 1})$ gives a reasonable rate of increase for K_n while maintaining $K_n \simeq 5$ when $n \leq 1000$.

To appreciate how the above empirical rule of picking M works, consider the simplest case where $\{X_t\}$ follows an MA(q) model as in eq. (3). It then follows that $\gamma(k) = 0$ for $|k| > q$. Then, as shown in Politis (2003), we have $\hat{q} \to q$ in probability as $n \to \infty$. So letting $\hat{M} \approx \hat{q}/c_{ef}$ ensures that the (effective) flat-top region extends exactly over the non-zero portion of the autocovariance. In other words, the autocovariances $\hat{\gamma}(k)$ for $|k| \leq q$ are not distorted at all by the lag-window thus ensuring that $\hat{f}(w)$ has negligible bias. The distortion/tapering starts right after; its purpose is variance reduction since one does not want to include estimates of autocovariances in eq. (1) when it is known that their true value is zero.

As mentioned in the Introduction, the above should be contrasted with the spectral estimation method that is based on averaging short, tapered periodograms. The latter involves tapering of the actual X_t data found in a short stretch of the series—not the

autocovariances; it was proposed by Welch (1967) in order to reduce the bias of the Bartlett (1946) estimator. For more details, see Brillinger (1981), Dahlhaus (1985), Percival and Walden (1993), and Politis (2005).

5 Appendix B: Technical proofs

Proof of Theorem 2.1. The correspondence between the eigenvalues of an autocovariance matrix and the associated spectral density is well known; cf. Brockwell and Davis (1991). Let $\Gamma_n = \left(\gamma(i-j)_{i,j=1,\ldots,n}\right)$ denote the $n \times n$ Toeplitz matrix with i,j element given by $\gamma(i-j)$. Then, as shown by Toeplitz (1911), the spectrum of eigenvalues of matrix Γ_n in the limit (as $n \to \infty$) tends to the set $\{f(w),\ w \in [-\pi, \pi]\}$.

Similarly, define the following $n \times n$ Toeplitz matrices: $\tilde{\Gamma}_{n,T_n} = (\tilde{\gamma}(i-j))$, $\tilde{\Gamma}_{n,T_n,*} = (\tilde{\gamma}_*(i-j))$, and $\hat{\Gamma}_{n,B_n} = (\hat{\gamma}(i-j)\mathbf{1}\{|i-j| \leq B_n\})$ where $B_n = (n/\log n)^{1/2(1+\alpha)}$ and $\tilde{\gamma}_*(i-j) = \hat{\gamma}(i-j)\mathbf{1}\{|\hat{\gamma}(i-j)| \geq T_n \ \& \ |i-j| \leq B_n\}$. With this choice of the bandwidth parameter B_n, the above three matrices are all very close to one another; here 'closeness' is measured in terms of the spectral radius norm λ defined as $\lambda(A) = \max\{|\lambda| \text{ where } \lambda \text{ is an eigenvalue of } A\}$. In particular, in the proof of Theorem 5 of Xiao and Wu (2011) it is established that:

$$\lambda(\tilde{\Gamma}_{n,T_n,*} - \hat{\Gamma}_{n,B_n}) = O_P(B_n^{-\alpha}), \text{ and } \lambda(\tilde{\Gamma}_{n,T_n} - \tilde{\Gamma}_{n,T_n,*}) = O_P(B_n^{-\alpha}). \tag{13}$$

Using a similar estimate as in the proof of Theorem 3 of Meckes (2007), we have:

$$\begin{aligned}(2\pi)^{-1}\lambda(\tilde{\Gamma}_{n,T_n,*} - \Gamma_n) &\geq \sup_{w\in[-\pi,\pi]} (2\pi)^{-1} \Big| \sum_{k=-n}^{n} (1 - \frac{|k|}{n})(\tilde{\gamma}_*(k) - \gamma(k))e^{ikw}\Big| \\ &\geq \Big| \sup_{w\in[-\pi,\pi]} |\tilde{f}_*(w) - f_n(w)| - |A_n| \Big| \end{aligned} \tag{14}$$

by the (inverse) triangle inequality, and letting $\tilde{f}_*(w) = (2\pi)^{-1}\sum_{s=-n}^{n}\tilde{\gamma}_*(s)e^{iws}$, $f_n(w) = (2\pi)^{-1}\sum_{s=-n}^{n}\gamma(s)e^{iws}$, and $A_n = (2\pi)^{-1}\sum_{s=-n}^{n}\frac{|s|}{n}(\tilde{\gamma}_*(s) - \gamma(s))e^{iws}$.

Now we have

$$2\pi|A_n| \leq \frac{1}{n}\sum_{s=-n}^{n}|s||\tilde{\gamma}_*(s) - \gamma(s)| = \frac{1}{n}\sum_{s=-B_n}^{B_n}|s||\tilde{\gamma}_*(s) - \gamma(s)| + \frac{1}{n}\sum_{B_n\leq|s|\leq n}|s||\tilde{\gamma}_*(s) - \gamma(s)|$$

$$= O_P(\frac{B_n^2}{n}) + O_P(B_n^{-\alpha}) = O_P(B_n^{-\alpha}) \tag{15}$$

where the estimate $\sum_{|s|\geq m} |\gamma(s)| = O(m^{-\alpha})$ from Remark 2.2 was used. Thus, putting the above together, we have:

$$\sup_{w\in[-\pi,\pi]} |\tilde{f}_*(w) - f_n(w)| = O_P(B_n^{-\alpha}) + O_P(\lambda(\tilde{\Gamma}_{n,T_n,*} - \Gamma_n)).$$

Notably, the matrix $\hat{\Gamma}_{n,B_n}$ is the banded autocovariance matrix of Wu and Pourahmadi (2009) for which it holds that

$$\lambda(\hat{\Gamma}_{n,B_n} - \Gamma_n) = O_P(B_n^{-\alpha}). \tag{16}$$

But from eq. (13) and (16) it follows that $\lambda(\tilde{\Gamma}_{n,T_n,*} - \Gamma_n) = O_P(B_n^{-\alpha})$ and thus

$$\sup_{w\in[-\pi,\pi]} |\tilde{f}_*(w) - f_n(w)| = O_P(B_n^{-\alpha}).$$

From Remark 2.2 we also have $|f(w) - f_n(w)| \leq \sum_{|s|\geq n} |\gamma(s)| = O(n^{-\alpha}) = o(B_n^{-\alpha})$, and therefore

$$\sup_{w\in[-\pi,\pi]} |\tilde{f}_*(w) - f(w)| = O_P(B_n^{-\alpha}). \tag{17}$$

Finally, look at:

$$2\pi \left(\tilde{f}(w) - \tilde{f}_*(w)\right) = \sum_{B_n\leq|s|\leq n} \hat{\gamma}(s)\mathbf{1}\{|\hat{\gamma}(s)| \geq T_n\}e^{iws} = a_n(w) + b_n(w).$$

In the above,

$$b_n(w) = \sum_{B_n\leq|s|\leq n} E\hat{\gamma}(s)\mathbf{1}\{|\hat{\gamma}(s)| \geq T_n\}e^{iws} = O_P(B_n^{-\alpha}),$$

and

$$a_n(w) = \sum_{B_n\leq|s|\leq n} \left(\hat{\gamma}(s) - E\hat{\gamma}(s)\right)\mathbf{1}\{|\hat{\gamma}(s)| \geq T_n\}e^{iws}.$$

Note that $|a_n(w)| \leq a_n^I + a_n^{II}$ where:

$$a_n^I = \sum_{B_n\leq|s|\leq n} |\hat{\gamma}(s) - E\hat{\gamma}(s)|\,\mathbf{1}\{|\hat{\gamma}(s)| \geq T_n \;\&\; |\gamma(s)| < c_\zeta T_n\}$$

and

$$a_n^{II} = \sum_{B_n \leq |s| \leq n} |\hat{\gamma}(s) - E\hat{\gamma}(s)| \, \mathbf{1}\{|\hat{\gamma}(s)| \geq T_n \ \& \ |\gamma(s)| \geq c_\zeta T_n\};$$

in the above, c_ζ is some constant in the interval $(0, 1 - 1/\zeta)$. Note that

$$Prob\{a_n^I > 0\} \leq Prob\{ \max_{1 \leq |s| \leq n} |\hat{\gamma}(s) - E\hat{\gamma}(s)| \geq (1 - c_\zeta) \ \zeta c_p' \sqrt{\log n/n} \ \}$$

$$\leq Prob\{ \max_{1 \leq |s| \leq n} |\hat{\gamma}(s) - E\hat{\gamma}(s)| \geq c_p' \sqrt{\log n/n} \ \} \to 0 \ \text{ as } \ n \to \infty$$

since $(1 - c_\zeta) \ \zeta > 1$ and invoking eq. (10). In other words, $Prob\{a_n^I = 0\} \to 1$ as $n \to \infty$. Now by an argument analogous to eq. (33) of Xiao and Wu (2011) it follows that $a_n^{II} = O_P(B_n^{-\alpha})$. Therefore, it follows that

$$\sup_{w \in [-\pi, \pi]} |\tilde{f}(w) - \tilde{f}_*(w)| = O_P(B_n^{-\alpha}) \tag{18}$$

Eq. (18) together with (17) prove the theorem.□

Proof of Theorem 2.2. The proof is analogous to the proof of Theorem 2.1 by: (a) replacing the banded autocovariance matrix $\hat{\Gamma}_{n,B_n}$ with the tapered autocovariance matrix proposed by McMurry and Politis (2010) using the lag-window $\lambda_{g,c}(x)$ as taper; (b) replacing the matrix $\tilde{\Gamma}_{n,T_n,*} = (\tilde{\gamma}_*(i-j))$ by the matrix $\tilde{\Gamma}_{n,T_n,\delta} = (\tilde{\gamma}_{\delta,M}(i-j))$ where $\tilde{\gamma}_{\delta,M}(i-j) = \hat{\gamma}(i-j)\mathbf{1}\{|\hat{\gamma}(i-j)| \geq \delta T_n/\zeta \ \& \ |i-j| \leq M\}$; and finally (c) replacing $\tilde{f}_*(w)$ by $\check{f}(w)$.

References

[1] Bartlett, M.S. (1946), On the theoretical specification of sampling properties of autocorrelated time series, *J.Roy.Statist.Soc.Suppl.*, 8, 27-41.

[2] Bartlett, M.S. (1948). Smoothing periodograms from time series with continuous spectra, *Nature*, vol. 161, pp. 686-687.

[3] Berk, K.N. (1974). Consistent autoregressive spectral estimates, *Ann. Statist.*, vol. 2, 489–502.

[4] Blackman, R.B. and Tukey, J.W. (1959). *The Measurement of Power Spectra from the Point of View of Communications Engineering*, Dover, New York.

[5] Brillinger, D.R. (1981), *Time Series: Data Analysis and Theory*, Holden-Day, New York.

[6] Brillinger, D.R. (1993), The Digital Rainbow: Some History and Applications of Numerical Spectrum Analysis, *Canad. J. Statist.*, vol. 21, no. 1, pp. 1-19.

[7] Brockwell, P. J. and Davis, R. A. (1991), *Time Series: Theory and Methods, 2nd ed.*, Springer, New York.

[8] Dahlhaus, R. (1985). On a spectral density estimate obtained by averaging periodograms, *J. Applied Prob.*, vol. 12, pp. 598–610.

[9] Daniell, P.J. (1946). Discussion of paper by M.S. Bartlett. *J. Roy. Statist. Soc. Suppl.*, vol. 8, pp. 88-90.

[10] Donoho, D.L. and Johnstone, I.M. (1995). Adapting to unknown smoothness via wavelet shrinkage, *J. Amer. Statist. Assoc.*, 90, 1200-1224.

[11] Einstein, A. (1914). Méthode pour la détermination des valeurs statistiques d'observations concernant des grandeurs soumises à des fluctuations irrégulières. *Arch. Sci. Phys. et Naturelles Ser. 4*, vol. 37, pp. 254-255.

[12] H.-Y. Gao (1997). Choice of threshold for wavelet shrinkage estimate of the spectrum. *J. Time Ser. Anal.*, vol. 18, pp. 231-251.

[13] Grenander, U. and Rosenblatt, M. (1957), *Statistical Analysis of Stationary Time Series*, Wiley, New York.

[14] Hannan, E.J. (1957). The variance of the mean of a stationary process, *J. Roy. Statist. Soc., Ser. B,* 19, 282-285

[15] Hannan, E.J. (1958). The estimation of the spectral density after trend removal. *J. Roy. Statist. Soc., Ser. B,* 20, 323-333.

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

[17] Haykin, S. (1983), *Nonlinear Methods of Spectral Analysis, 2nd Ed.*, Springer Verlag, Berlin.

[18] Ioannidis, E.E. (1994). On the behavior of a Capon-type spectral density estimator, *Ann. Statist.*, vol. 22, 2089–2114.

[19] Katznelson, Y. (2004). *An Introduction to Harmonic Analysis*, 3rd Ed., Cambridge Univ. Press, Cambridge.

[20] Lamperti, J. (1977). *Stochastic processes.* Applied Mathematical Sciences, vol. 23, Springer-Verlag, New York, Heidelberg, Berlin.

[21] McMurry, T. and Politis, D.N. (2004). Nonparametric regression with infinite order flat-top kernels, *J. Nonparam. Statist.*, vol. 16, no. 3-4, 549–562.

[22] McMurry, T. and Politis, D.N. (2010). Banded and tapered estimates of autocovariance matrices and the linear process bootstrap, *J. Time Ser. Anal.*, vol. 31, pp. 471-482.

[23] Meckes, M.W. (2007). On the spectral norm of a random Toeplitz matrix, *Elec. Comm. Probab.*, vol. 12, pp. 315-325.

[24] Moulin, P. (1994). Wavelet thresholding techniques for power spectrum estimation, *IEEE Trans. Signal Proc.*, vol. 42, pp. 3126-36.

[25] Neumann, M.H. (1996). Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series. *J. Time Ser. Anal.*, 17, no. 6, 601633.

[26] Parzen, E. (1957). On Consistent Estimates of the Spectrum of a Stationary Time Series. *Ann. Math. Statist.*, Vol. 28, No. 2., pp. 329-348.

[27] Parzen, E. (1961), Mathematical Considerations in the Estimation of Spectra, *Technometrics*, vol. 3, 167-190.

[28] Percival, D.B. and Walden, A.T.(1993). *Spectral analysis for physical applications. Multitaper and conventional univariate techniques.* Cambridge University Press, Cambridge.

[29] Percival, D.B. and Walden, A.T.(2000). *Wavelet methods for time series analysis.* Cambridge Series in Statistical and Probabilistic Mathematics, 4. Cambridge University Press, Cambridge.

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

[31] Politis, D.N. (2003). Adaptive bandwidth choice, *J. Nonparam. Statist.,* vol. 15, no. 4-5, 517-533.

[32] Politis, D.N. (2005). Complex-valued tapers, *IEEE Signal Processing Letters*, Volume 12, Issue 7, pp. 512–515.

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

[34] Politis, D.N., and Romano, J.P. (1995), Bias-corrected nonparametric spectral estimation. *J. Time Ser. Anal.*, **16**, 67–103.

[35] Politis, D. N. and White, H. (2004). Automatic block-length selection for the dependent bootstrap, *Econometric Reviews,* vol. 23, no. 1, pp. 53-70, 2004.

[36] Priestley, M.B. (1962), Basic considerations in the estimation of spectra, *Technometrics*, vol. 4, 551-564.

[37] Priestley, M.B. (1981). *Spectral Analysis and Time Series*, Academic Press, New York.

[38] Rosenblatt, M. (1984). Asymptotic normality, strong mixing and spectral density estimates, *Annals Prob.*, 12, 1167-1180.

[39] Rosenblatt, M. (1985), *Stationary Sequences and Random Fields*, Birkhäuser, Boston.

[40] Samarov, A. (1977). Lower bound for the risk of spectral density estimates, *Probl. Inform. Transm.*, 13, pp. 67-72.

[41] Shao, X. and Wu, W.B. (2007). Asymptotic spectral theory for nonlinear time series. *Annals Statist.*, Vol. 35, 1773-1801.

[42] Toeplitz, O. (1911). Zur Theorie der quadratischen und bilinearen Formen von unendlichvielen Veränderlichen, *Math. Annalen*, Vol. 70, pp. 351-376.

[43] Walden, A.T., Percival, D.B., and McCoy, E.J. (1998) Spectrum estimation by wavelet thresholding of multitaper estimators. *IEEE Trans. Signal Process.* 46, no. 12, 31533165.

[44] Welch, P.D. (1967). The use of the Fast Fourier Transform for estimation of spectra: a method based on time averaging over short, modified periodograms, *IEEE Trans. Audio Electroacoustics AU-15*, pp. 70-73.

[45] Wu, W.B. (2005). Nonlinear system theory: another look at dependence, *Proc. Nat. Acad. Sci.*, vol. 102, no. 40, pp. 14150–14154.

[46] Wu, W.B. and Pourahmadi, M. (2009). Banding sample autocovariance matrices of stationary processes, *Statistica Sinica*, vol. 19, pp. 1755-1768.

[47] Xiao, H. and Wu, W.B. (2011). Covariance matrix estimation for stationary time series, Preprint, Dept. of Statistics, Univ. of Chicago.