Complex-valued tapers

Dimitris N. Politis[1]

Abstract: The spectral estimation method based on the average of short, tapered periodograms is re-examined. The bias of such estimators is typically $O(1/b^2)$ where b is the length of the short blocks. Much of the current research on multi-tapering has been focusing on reducing the proportionality constant implicit in the term $O(1/b^2)$. In this note, we show how—with the use of complex-valued tapers—the bias of the spectral estimator can be reduced by orders of magnitude becoming $O(1/b^p)$ for (possibly) high p. Expressions for the estimators' variance and MSE are presented with an aim towards optimal estimation. An automatic method of optimally choosing the block size b is given. Finally, the usage of multiple complex tapers is proposed in an effort to reduce sidelobe size and improve finite-sample performance. **Edics category:** 1.TFSR

Keywords: Bandwidth choice, Bartlett estimator, flat-top lag-windows, multi-tapering, power spectrum.

1 Introduction: Bartlett and Welch

Let $X_1, \ldots, X_n$ be an observed stretch from a real-valued, mean zero, stationary process $\{X_t, t \in \mathbb{Z}\}$ with unknown autocovariance function $\gamma(k) = \text{Cov}(X_t, X_{t+k})$. If $\sum_{s=-\infty}^{\infty} |\gamma(s)| < \infty$, then the spectral density $f(w) = \sum_{s=-\infty}^{\infty} e^{iws}\gamma(s)$ exists and is continuous. In general, for $p \in \mathbb{N}$, consider: **Assumption $\mathbf{A}_p$:** $\sum_{s=-\infty}^{\infty} |s^p\gamma(s)| < \infty$. **Assumption $\mathbf{A}_\infty$:** $|\gamma(k)| \leq c_1 e^{-c_2 k}$ for some $c_1, c_2 > 0$. Assumption A_p implies that the pth derivative of the spectral density, namely $f^{(p)}(w)$, exists and is continuous. Assumption A_∞ holds, e.g., when $\{X_t\}$ follows a stationary ARMA model [3].

Define the sample autocovariance $\hat{\gamma}(k) = \frac{1}{n}\sum_{t=1}^{n-|k|} X_t X_{t+|k|}$. Although $\hat{\gamma}(k)$ is consistent for $\gamma(k)$ for each fixed k, the periodogram $T(w) = \sum_{|s|<n} e^{iws}\hat{\gamma}(s)$ is inconsistent for $f(w)$; see [3]. Under A_1, the bias of $T(w)$ is rather small—approximately $O(1/n)$— but the variance of $T(w)$ tends to $f^2(w)(1 + \mathbf{1}_{\{w/\pi \in \mathbb{Z}\}})$ and not to zero for large n; here $\mathbf{1}_A$ is the indicator of set A: it is equal to 1 or 0 according to whether A is true or not. Using the definition of $\hat{\gamma}(k)$, we also obtain:

$$T(w) = \sum_{|s|<n} e^{iws}\hat{\gamma}(s) = \frac{1}{n}|\sum_{t=1}^{n} e^{iwt}X_t|^2. \tag{1}$$

In 1950, M.S. Bartlett [1] proposed one of the first consistent estimators of $f(w)$. Bartlett's scheme was based on splitting the data into blocks of size b. So let $B_k = (X_k, X_{k+1}, \ldots, X_{k+b-1})$ be the kth block where $k = 1, \ldots, q$ and $q = n-b+1$; alternatively, we may denote $(X_1^{[k]}, X_2^{[k]}, \cdots, X_b^{[k]})$

[1]Professor of Mathematics, Univ. of California at San Diego, La Jolla, CA 92093, USA; e-mail: dpolitis@ucsd.edu. Research partially supported by NSF grants DMS-01-04059 and SES-04-18136. Many thanks are due to Drs. M. Kotsaftis for the inspiration, T. McMurry for a careful proof-reading, and H. Stravinsky for his warmth and kindness.

the entries of B_k. Let $T_k(w) = \frac{1}{b}|\sum_{t=1}^{b} e^{iwt} X_t^{[k]}|^2$ and $\hat{\gamma}^{[k]}(s) = \frac{1}{b}\sum_{t=1}^{b-|s|} X_t^{[k]} X_{t+|s|}^{[k]}$ denote the periodogram and sample autocovariance (respectively) as computed from block B_k. The Bartlett estimator is then defined as $\tilde{T}(w) = \frac{1}{q}\sum_{k=1}^{q} T_k(w)$. The bias of $\tilde{T}(w)$ is equivalent to the bias of each of the $T_k(w)$, i.e., about $O(1/b)$. However, $\mathrm{Var}\,\tilde{T}(w) = O(b/n)$ due to the effect of averaging [2]. Thus, $\tilde{T}(w)$ is consistent for $f(w)$ as long as[2] $b \to \infty$ but $b/n \to 0$ as $n \to \infty$. Letting $\tilde{\gamma}(s) = \frac{1}{q}\sum_{k=1}^{q} \hat{\gamma}^{[k]}(s)$, and noting that $\tilde{\gamma}(s) \approx \hat{\gamma}(s)$ for $|s| < b$, we have

$$\tilde{T}(w) = \sum_{|s|<b} \lambda^B(s/b) e^{iws} \tilde{\gamma}(s) \approx \sum_{|s|<b} \lambda^B(s/b) e^{iws} \hat{\gamma}(s); \tag{2}$$

in the above, $\lambda^B(x) = \max(0, 1 - |x|)$ is the Bartlett lag-window.

Noting that the bias of a periodogram is mostly due to edge effects, Welch [16] proposed tapering each of the short periodograms before taking their average. So let $T_k^\nu(w) = \frac{1}{b}|\sum_{t=1}^{b} e^{iwt} \nu_b(t) X_t^{[k]}|^2$ be the tapered periodogram of block B_k; here $\nu_b(t) = \nu(t/b)$ with $\nu : [0,1] \to \mathbb{R}^+$ being the taper. Welch's spectral estimator is then defined as $\tilde{T}^\nu(w) = \frac{1}{q}\sum_{k=1}^{q} T_k^\nu(w)$. Under standard conditions [3]

$$\mathrm{Var}\ (\tilde{T}^\nu(w)) \approx \Delta_{w,f} \frac{b}{n} \ \text{ where } \ \Delta_{w,f} = f^2(w) \int_{-1}^{1} \frac{\nu_*^2(x)}{\nu_*^2(0)} dx \ (1 + \mathbf{1}_{\{w/\pi \in \mathbb{Z}\}}), \tag{3}$$

where ν_* is the self-convolution of ν defined as $\nu_*(t) = \int \nu(x)\nu(x + |t|)dx$; see [4] [15] [16] [17]. Note that Welch's estimator reduces to Bartlett's when $\nu(x) = 1$; to exclude this possibility consider:

Assumption B. *The taper* $\nu : [0,1] \to \mathbb{R}^+$ *is a continuous, piecewise differentiable function, monotone non-decreasing on* $[0, 1/2]$ *and symmetric about* $1/2$*, that satisfies* $\nu(0) = 0$ *and* $\nu_*(0) = 1$.

Under A_2 and B, it can be shown that $\mathrm{Bias}\,(\tilde{T}^\nu(w)) = O(1/b^2)$, whereas $\mathrm{Var}\,(\tilde{T}^\nu(w)) = O(b/n)$ as before. Therefore, $\tilde{T}^\nu(w)$ outperforms $\tilde{T}(w)$ in terms of Mean Squared Error (MSE). Ever since Welch's 1967 paper [16], much research has been focused on choosing the taper—or combination of tapers—with the goal of reducing the proportionality constant implicit in the term $\mathrm{Bias}\,(\tilde{T}^\nu(w)) = O(1/b^2)$; see [5] [11] [14] [17]. In what follows, we show how—with the use of complex-valued tapers—the bias of Welch's spectral estimator can be reduced by orders of magnitude, becoming $O(1/b^p)$ under Assumption A_p. An automatic method of optimally choosing the block size b is also given, motivated by an analogy with flat-top lag-window spectral estimators [7] [8] [9].

[2]All asymptotic approximations in this paper will be understood to hold as $b \to \infty$ but with $b/n \to 0$.

[3]There is a variety of sufficient conditions that imply equations (3) (5). For example, (cf. [10] p. 455, or [3] p. 351) a sufficient condition is that $X_t = \sum_{i=-\infty}^{\infty} \theta_i Z_{t-i}$, where the Z_t's are i.i.d. with $EZ_t = 0$, and $EZ_t^4 < \infty$, and the θ_i's satisfy $\sum_{i=-\infty}^{\infty} |i|^{1/2}|\theta_i| < \infty$. For different sufficient conditions based on summability of cumulants see [2] p. 26 and p. 144, or [12] p. 134.

2 Tapers, lag-windows and higher-order kernels

The right-hand-side of eq. (2) has the form of a general *lag-window* spectral estimator [3] defined as

$$\hat{f}_\lambda(w) = \sum_{|s|<b} \lambda(s/b) e^{iws} \hat{\gamma}(s) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Lambda_b(w-x) T(x) dx \tag{4}$$

where $\lambda : [-1,1] \to \mathbb{R}^+$. The 2nd equality in (4) is due to the fact $\hat{\gamma}(s) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-iws} T(w) dw$, together with the definition of the kernel $\Lambda_b(w) = \sum_{|s|<b} \lambda(s/b) e^{iws}$, or equivalently, $\Lambda_b(w) = b\Lambda(bw)$ with $\Lambda(w) = \int \lambda(x) e^{iwx} dx$. Under standard conditions, it follows that

$$\mathrm{Var}\,(\hat{f}_\lambda(w)) \approx c_{w,f}\, \frac{b}{n} \quad \text{where} \quad c_{w,f} = f^2(w) \int \lambda^2(x) dx \; (1 + \mathbf{1}_{\{w/\pi \in \mathbb{Z}\}}). \tag{5}$$

The bias of $\hat{f}_\lambda$ depends on the smoothness of λ at the origin. A lag-window λ is said to be of *order* q if its kth derivative $\lambda^{(k)}(0) = 0$ for $k = 1, \ldots, q-1$, but $\lambda^{(q)}(0) \neq 0$. Then, under A_p,

$$\mathrm{Bias}\,(\hat{f}_\lambda(w)) = O(1/b^\zeta) \quad \text{where} \quad \zeta = \min(p, q). \tag{6}$$

The good bias performance comes with a string attached: for λ to have order $q > 2$, its kernel Λ *must* take on some negative values. Since $\lambda^{(k)}(0) = 0$ is equivalent to $\int w^k \Lambda(w) dw = 0$, a nonnegative kernel can have order at most two. Nevertheless, the accuracy of a higher-order $\hat{f}_\lambda$ implies that $\mathrm{Prob}(\hat{f}_\lambda(w) < 0)$ is practically negligible. In addition, there is an easy fix for the nonnegativity: let $\hat{f}_\lambda^+(w) = \max(\hat{f}_\lambda(w), 0)$. Taking the positive-part can only improve the estimator; in fact [9]: $MSE(\hat{f}_\lambda^+(w)) \leq MSE(\hat{f}_\lambda(w))$. In other words, $\hat{f}_\lambda^+$ inherits the higher-order accuracy of $\hat{f}_\lambda$.

Returning to Welch's estimator, note that eq. (2) can be extended to cover the expectation of $\tilde{T}^\nu(w)$; as shown in the Appendix, we have

$$E\tilde{T}^\nu(w) \approx \sum_{|s|<b} \nu_*(s/b) e^{iws} \gamma(s) \approx E\hat{f}_{\nu_*}(w). \tag{7}$$

As $\tilde{T}^\nu(w) \geq 0$, it follows that ν_* is a lag-window of order at most two. As a matter of fact, Assumption B ensures that ν_* has order exactly two, and thus the bias of $\tilde{T}^\nu(w)$ is $O(1/b^2)$.

Intuitively, the Fourier transform (kernel) of ν_* is effectively a squared quantity—and therefore nonnegative—since ν_* is a convolution of ν with itself. In order to open up the possibility of higher-order accuracy of a spectral estimator based on tapers, it is natural to consider the case of complex-valued tapers whose "square" may be negative.

3 Flat-top lag-windows and complex-valued tapers

A family of infinite-order lag-windows is the flat-top family [7] [8] [9] whose typical member is: $\lambda_{g,c}(x) = 1$ if $|x| < c$; $\lambda_{g,c}(x) = g(|x|)$ if $c \leq |x| \leq 1$; and $\lambda_{g,c}(x) = 0$ if $|x| \geq 1$. In the above, c is a fixed number in $(0, 1)$, and g is a continuous function satisfying $g(c) = 1$ and $g(1) = 0$. The simplest such choice is to let g be a straight line; this leads to the trapezoidal flat-top family [9] whose typical member λ_c^T is conveniently described in terms of two Bartlett lag-windows as follows:

$$\lambda_c^T(x) = (h+1)\lambda^B(x) - h\lambda^B(x/c) \quad \text{where} \quad h = \frac{c}{1-c} \tag{8}$$

From eq. (5), we have $\text{Var}\left(\hat{f}_{\lambda_c^T}(w)\right) \approx \frac{3h+1}{h+1}\left(\frac{2b}{3n}\right) f^2(w)(1 + \mathbf{1}_{\{w/\pi \in \mathbb{Z}\}})$. At the same time, $\text{Bias}\left(\hat{f}_{\lambda_c^T}(w)\right) = O(1/b^p)$ under A_p. Thus, letting b proportional to $n^{1/(2p+1)}$, yields $MSE(\hat{f}_{\lambda_c^T}(w)) = O(n^{-2p/(2p+1)})$ which is the best rate possible [13] under A_p. Under A_∞, the bias of $\hat{f}_{\lambda_c^T}(w)$ is negligible as it decays exponentially with b; letting b proportional to $\log n$, we then have $MSE(\hat{f}_{\lambda_c^T}(w)) = O(\log n/n)$ which is very close to the parametric $O(1/n)$ rate. Finally, consider the case where $\gamma(s) = 0$ for all $s >$ some s_0, e.g., when $\{X_t\}$ follows a Moving Average (MA) model of order s_0; in that case, letting $b = s_0/c$ results into $MSE(\hat{f}_{\lambda_c^T}(w)) = O(1/n)$.

It is apparent that to achieve optimal MSEs, the choice of b is crucial. Since the degree of smoothness of f, i.e., the value p in A_p, is not known in advance, the choice of b must be data-dependent. Fortuitously, [8] this choice can be made by a simple inspection of the correlogram. Although it may be overly optimistic to assume that $\{X_t\}$ exactly obeys a finite-order MA model, an MA model may generally serve as an approximation [3]; the aforementioned optimal choice of b under such an approximate MA model motivates the following:

Empirical Rule:[4] *Let* $\hat{b} = \hat{s}_0/c$ *where* $\hat{s}_0$ *is the smallest integer such that* $\hat{\gamma}(s) \simeq 0$ *for all* $s \geq \hat{s}_0$.

It can be shown that the estimator $\hat{b}$ automatically adapts to the (unknown) underlying degree of smoothness p resulting into a near-optimal MSE rate for $\hat{f}_{\lambda_c^T}(w)$. In other words, if we let b_{opt} be the optimal block size b computed under full knowledge of the value p in A_p as well as of $\gamma(k)$, then $\hat{b} \approx b_{opt}$ with high probability; see [8] for details.

Consider a general complex-valued taper $\nu : [0, 1] \to \mathbb{C}$, i.e., $\nu = \nu^R + i\nu^I$ where ν^R, ν^I are real-valued. Let $I_k^\nu(w) = \frac{1}{b}\sum_{t=1}^b e^{iwt}\nu_b(t)X_t^{[k]} \sum_{s=1}^b e^{-iws}\nu_b(s)X_s^{[k]}$ denote the modified periodogram

[4] The condition $\hat{\gamma}(s) \simeq 0$ is really an implied test of significance which is formally described as follows. Let $\hat{\rho}(k) = \hat{\gamma}(k)/\hat{\gamma}(0)$; we will say that $\hat{\gamma}(s) \simeq 0$ and $\hat{\rho}(s) \simeq 0$ for all $s \geq \hat{s}_0$ if $|\hat{\rho}(\hat{s}_0 + k)| < c_0\sqrt{\log_{10} n/n}$, for $k = 1, \ldots, K_n$, where $c_0 > 0$ is a fixed constant, and K_n is a positive, nondecreasing function of n such that $K_n = o(\log n)$. Practically recommended values are $c_0 = 2$ and $K_n = \max(5, \sqrt{\log_{10} n})$; see [8] for more details.

from block B_k, and $\tilde{I}^\nu(w) = \frac{1}{q}\sum_{k=1}^q I_k^\nu(w)$ their average. Because the conjugate of ν is not used in defining the modified periodograms, $I_k^\nu(w)$ and $\tilde{I}^\nu(w)$ are generally complex-valued; however, if ν happens to be real-valued, then they reduce to the real (and nonnegative) $T_k^\nu(w)$ and $\tilde{T}^\nu(w)$ respectively. Our estimator of $f(w)$ will then be the real part of $\tilde{I}^\nu(w)$, namely $\hat{I}^\nu(w) = \mathrm{Re}[\tilde{I}^\nu(w)]$.

From eq. (11) and (12) in the Appendix it follows that the bias of $\hat{I}^\nu(w)$ is tantamount to the bias of a lag-window estimator $\hat{f}_\lambda(w)$ with $\lambda(x) = a(x) = \int \nu^R(s)\nu^R(s+|x|)ds - \int \nu^I(s)\nu^I(s+|x|)ds$. Therefore, we can tailor ν^R and ν^I to give $\hat{I}^\nu(w)$ good bias properties. Note, however, that here the taper ν must satisfy $a(0) = 1$ to avoid 'losing variance' instead of the condition $\nu_*(0) = 1$ of Assumption B; of course, $a(0) = 1$ is equivalent to $\nu_*^R(0) - \nu_*^I(0) = 1$.

Although many choices for ν^R, ν^I are possible, we now focus on a particular simple—but effective—choice. Fix a constant $c \in (0,1)$ and let $\hat{I}_c^T(w)$ be the estimator $\hat{I}^\nu(w)$ using the choices

$$\nu^R(x) = \sqrt{\frac{1}{1-c}}\mathbf{1}_{\{0<x<1\}} \quad \text{and} \quad \nu^I(x) = \sqrt{\frac{1}{1-c}}\mathbf{1}_{\{0<x<c\}}. \tag{9}$$

Eq. (12) implies that $\hat{I}_c^T(w)$ is associated with an implied lag-window $a(k)$ that equals $\lambda_c^T(k)$ as given in eq. (8); thus, the bias of $\hat{I}_c^T(w)$ is tantamount to the bias of the trapezoidal lag-window estimator $\hat{f}_{\lambda_c^T}(w)$, i.e., $\mathrm{Bias}\,(\hat{I}_c^T(w)) = O(1/b^p)$ under A_p. Similarly, eq. (15) shows that $\mathrm{Var}\,(\hat{I}_c^T(w)) \approx \mathrm{Var}\,(\hat{f}_{\lambda_c^T}(w))$, so that $\hat{I}_c^T(w)$ is equivalent to $\hat{f}_{\lambda_c^T}(w)$ in terms of its first two moments. Furthemore, eq. (14) shows that $\hat{I}_c^T(w) = \hat{f}_{\lambda_c^T}(w) + O_P(b/n)$ so the two estimators are asymptotically equivalent. Thus, letting b proportional to $n^{1/(2p+1)}$, yields $MSE(\hat{I}_c^T(w)) = O(n^{-2p/(2p+1)})$ under A_p. Under A_∞, letting b proportional to $\log n$, we obtain $MSE(\hat{I}_c^T(w)) = O(\log n/n)$.

By its construction, $\hat{I}_c^T(w)$ is not necessarily nonnegative, so the practical estimator is $\hat{I}_c^{T,+}(w) = \max(0, \hat{I}_c^T(w))$. However, as previously mentioned, $MSE(\hat{I}_c^{T,+}(w)) \le MSE(\hat{I}_c^T(w))$. Finally, note that by the equivalence of $\hat{I}_c^T(w)$ to $\hat{f}_{\lambda_c^T}(w)$, the "Empirical Rule" for optimally choosing b given above applies *verbatim* to optimally choosing b for the estimator $\hat{I}_c^T(w)$.

4 Multiple complex tapers

The aforementioned optimal MSE properties of $\hat{I}_c^T(w)$ and $\hat{f}_{\lambda_c^T}(w)$ hold for any $c \in (0,1)$. Nevertheless, it is recommended to let c be (around) 0.5 since the extreme values $c = 0$ and $c = 1$ are both to be avoided: the first corresponds to the Bartlett lag-window which is suboptimal—see Figure 1(a,b); on the other hand, $c \to 1$ approximately yields a truncated periodogram whose corresponding kernel (the Dirichlet) has very pronounced sidelobes—see Figure 1 (c,d).

Although more work on the 'optimal' value of c is in order, to fix ideas we now let $\hat{I}^T(w)$ and $\hat{f}_{\lambda^T}(w)$ without a subscript denote $\hat{I}^T_c(w)$ and $\hat{f}_{\lambda^T_c}(w)$ using $c = 1/2$. Similarly, let $\lambda^T(x) = \lambda^T_{1/2}(x)$ be the lag-window that corresponds to $\hat{I}^T(w)$ and $\hat{f}_{\lambda^T}(w)$, and $\Lambda^T(w) = \Lambda^T_{1/2}(w)$ be the corresponding kernel. From eq. (14) it follows that $\Lambda^T(w)$ inherits the tail behavior associated with the Bartlett estimator; this is manifested in Figure 1 (f): except for the first sidelobe—which by necessity must go negative—the rest of the sidelobes are comparable in size to those in Figure 1(b).

The Bartlett-type sidelobes of $\Lambda^T(w)$ are essentially due to the sharp corners in the graph of $\lambda^T(x)$ shown in Figure 1 (e). To effectively smooth-out those sharp corners and reduce the size of the sidelobes, the idea of multiple tapering may be employed; see e.g. [11], [15] and the references therein. For example, consider the lag-window $\bar{\lambda}^T(x) = (1/3)[\lambda^T(x) + \lambda^T(x/0.95) + \lambda^T(x/0.90)]$ with kernel $\bar{\Lambda}^T(w)$. Figure 1 (g) shows that, although the flat-top remains unaffected, the graph of $\bar{\lambda}^T(x)$ has smoother corners. At first glance, the plot of $\bar{\Lambda}^T(w)$ shown in Figure 1 (h) looks much like that of $\Lambda^T(w)$ shown in Figure 1 (f). However, due to the difference in scale of the two figures it is apparent that the kernel $\bar{\Lambda}^T(w)$ has sidelobes of size (at least) 10% less than those of $\Lambda^T(w)$, and this remains true even after adjustment/scaling to ensure the same L_2 norm, i.e., variance.

Finally, note that the lag-window $\bar{\lambda}^T(x)$ corresponds to a simple average of $\hat{I}^T(w)$ with different block sizes. If we let $\hat{I}^{T,b}(w)$ denote the estimator $\hat{I}^T(w)$ using block size b, then $\bar{\lambda}^T(x)$ corresponds to the estimator $(1/3)[\hat{I}^{T,b}(w) + \hat{I}^{T,0.95 \cdot b}(w) + \hat{I}^{T,0.90 \cdot b}(w)]$. Thus, a general multi-taper estimator may be defined as: $\hat{I}^{T,b,mult}(w) = (1/m)\sum_{j=1}^m a_j \hat{I}^{T,d_j b}(w)$ where the a_j are weights such that $\sum_{j=1}^m a_j = 1$, and $d_1 < d_2 < \cdots < d_m$ are constants in the interval $[1-\epsilon, 1+\epsilon]$ for some small $\epsilon > 0$. Note that the "Empirical Rule" for optimally choosing b for $\hat{I}^{T,b,mult}(w)$ should now be modified to read: $\hat{b} = 2\hat{s}_0/d_1$, where $\hat{s}_0$ is the smallest integer such that $\hat{\gamma}(s) \simeq 0$ for all $s \geq \hat{s}_0$. More work is needed, however, to pin-point optimal values for a_j and d_j in connection with $\hat{I}^{T,b,mult}(w)$.

Conclusions

The use of complex-valued tapers is proposed with the goal of reducing the bias in Welch's spectral estimation procedure. It is shown that, with careful choice of the shape of the complex tapers, the bias can be reduced by orders of magnitude provided the underlying true spectral density is smooth enough, i.e., has more than two continuous derivatives. Expressions for the new estimators' variance and MSE are presented, and an automatic method of optimal bandwidth choice is given motivated by an analogy with flat-top lag-window spectral estimators. Finally, the use of multiple complex tapers is proposed in an effort to reduce sidelobe size and improve finite-sample performance.

References

[1] Bartlett, M.S. (1950). Periodogram analysis and continuous spectra, *Biometrika*, 37, 1–16.

[2] Brillinger, D. (1975). *Time Series: Data Analysis and Theory.* Holt, Rinehart and Winston, New York.

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

[4] Dahlhaus, R. (1985). On a spectral density estimate obtained by averaging periodograms. *J. Applied Probability 22*, 598-610.

[5] Papoulis, A. (1973), Minimum Bias Windows for High Resolution Spectral Estimates, *IEEE Trans. Infor. Theory*, Vol. IT-19, pp. 9-12.

[6] Pedrosa, A. and Schmeiser, B. (1993). Asymptotic and finite-sample correlations between obm estimators. *Proceedings of the 25th Winter Simulation Conference*, G.W. Evans et al. (Eds.), Los Angeles, 481–488.

[7] 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.

[8] Politis, D. N. (2003). Adaptive bandwidth choice. *J. Nonparametric Statistics*, vol. 15, no. 4-5, 517-533.

[9] Politis, D. N., and Romano, J.P. (1995). Bias-Corrected Nonparametric Spectral Estimation, *J. Time Ser. Anal.*, 16, 67-104.

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

[11] Riedel, K.S. and Sidorenko, A. (1995). Minimum Bias Multiple Taper Spectral Estimation, *IEEE Trans. Signal Proc.*, Vol. 43, No. 1, pp. 188-195.

[12] Rosenblatt, M. (1985), *Stationary sequences and random fields*, Birkhäuser, Boston.

[13] Samarov, A. M. (1977). A lower bound of the risk in estimates of the spectral density, *Problems of Information Transmission*, 13, no. 1, 67–72.

[14] Thomson, D.J. (1982). Spectrum estimation and harmonic analysis, *Proc. IEEE*, vol. 70, pp. 1055-1096.

[15] Walden, A.T. (2000) A unified view of multitaper multivariate spectral estimation. *Biometrika,* 87, 767-787.

[16] 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 and Electroacoustics AU-15*, 70-73.

[17] Zhurbenko, I.G. (1986). *The Spectral Analysis of Time Series*. Amsterdam: North-Holland.

Appendix: technical remarks and details

Remark. The mean zero assumption regarding $\{X_t\}$ is employed for convenience. A more practical set-up is the following: Let $Y_1, \ldots, Y_n$ be an observed stretch from a real-valued, stationary process $\{Y_t, t \in \mathbb{Z}\}$ with unknown mean $\mu = EY_t$ and autocovariance $\gamma(k) = \text{Cov}(Y_t, Y_{t+k})$. Let $\bar{Y}_n = \frac{1}{n}\sum_{t=1}^n Y_t$, and define the centered version $X_t = Y_t - \bar{Y}_n$ for $t = 1, \ldots, n$. In other words, the data $X_1, \ldots, X_n$ are now a result of centering by a sample mean—as opposed to a true expectation.

The implication of the practical set-up is an error of order $O(b/n)$ in all spectral estimation procedures based on blocks of size b or lag-windows with window-width equal to b. In particular:
(a). The full periodogram $T(w)$ becomes non-informative at the origin, i.e., $T(0) = 0$; this is not true, however, for the short periodograms. In other words, the $T_k(0)$ are informative, and so is their average, i.e., Bartlett's estimator, which is (approximately) unbiased even at $w = 0$. The reason is that although the short periodograms are computed on a block of size b, the data-centering has been performed based on the full-data sample mean, i.e., $X_t = Y_t - \bar{Y}_n$; since $b/n \to 0$, centering at the full-data sample mean is accurate enough for purposes of the short periodograms.
(b). The term $O(b/n)$ must be added to the bias of all spectral estimators considered in the paper. For example, the bias of the classic Welch estimator $\tilde{T}^\nu(w)$ will not be $O(1/b^2)$ but $O(1/b^2)+O(b/n)$. Luckily, this extra bias term can be neglected as it is swamped by the order of magnitude of the spectral estimator's standard deviation which is $O(\sqrt{b/n})$.

The bias of $\hat{I}^\nu(w)$. Consider the complex-valued taper $\nu_b(t) = \nu(t/b)$, and write $\nu_b = \nu_b^R + i\nu_b^I$ where $\nu_b^R(t) = \nu^R(t/b)$ and $\nu_b^I(t) = \nu^I(t/b)$. Let $A(s,k) = \nu_b^R(s)\nu_b^R(s+|k|) - \nu_b^I(s)\nu_b^I(s+|k|)$ and $B(s,k) = \nu_b^R(s)\nu_b^I(s+|k|) + \nu_b^I(s)\nu_b^R(s+|k|)$ so that $\nu_b(s)\nu_b(s+|k|) = A(s,k) + iB(s,k)$.

The modified periodogram from the first block B_1 is analyzed below:

$$\begin{aligned}
I_1^\nu(w) &= \tfrac{1}{b}\textstyle\sum_{t=1}^b e^{iwt}\nu_b(t)X_t \sum_{s=1}^b e^{-iws}\nu_b(s)X_s = \tfrac{1}{b}\sum_{t=1}^b\sum_{s=1}^b e^{iwt}e^{-iws}\nu_b(t)\nu_b(s)X_tX_s \\
&= \tfrac{1}{b}\textstyle\sum_{k=0}^{b-1}\sum_{s=1}^{b-k} e^{iwk}\nu_b(s)\nu_b(s+k)X_sX_{s+k} + \tfrac{1}{b}\sum_{k=-b+1}^{-1}\sum_{s=1-k}^{b} e^{iwk}\nu_b(s)\nu_b(s+k)X_sX_{s+k} \\
&= \tfrac{1}{b}\textstyle\sum_{|k|<b}\sum_{s=1}^{b-|k|} e^{iwk}A(s,k)X_sX_{s+|k|} + \tfrac{i}{b}\sum_{|k|<b}\sum_{s=1}^{b-|k|} e^{iwk}B(s,k)X_sX_{s+|k|}.
\end{aligned}$$

Due to symmetry properties of $B(s,k)$, the term $\sum_{|k|<b}\sum_{s=1}^{b-|k|} e^{iwk}B(s,k)X_sX_{s+|k|}$ is real. Hence,

$$\text{Re}[I_1^\nu(w)] = \frac{1}{b}\sum_{|k|<b}\sum_{s=1}^{b-|k|} e^{iwk}A(s,k)X_sX_{s+|k|}. \tag{10}$$

The expected value of our estimator $\hat{I}^\nu(w) = \text{Re}[\tilde{I}^\nu(w)]$ is identical to the expected value of

$\text{Re}[I_1^\nu(w)]$. It follows that

$$E\hat{I}^\nu(w) = \frac{1}{b}\sum_{|k|<b}\sum_{s=1}^{b-|k|} e^{iwk}A(s,k)\gamma(k) = \sum_{|k|<b} e^{iwk}a_b(k)\gamma(k) \tag{11}$$

where $a_b(k) = \frac{1}{b}\sum_{s=1}^{b-|k|} A(s,k) = \frac{1}{b}\sum_{s=1}^{b-|k|}\left[\nu_b^R(s)\nu_b^R(s+|k|) - \nu_b^I(s)\nu_b^I(s+|k|)\right]$. By a Riemman–sum approximation argument, we have: $a_b(k) \approx a(k/b)$ for large b where

$$a(x) = \int \nu^R(s)\nu^R(s+|x|)ds - \int \nu^I(s)\nu^I(s+|x|)ds. \tag{12}$$

In the case of a real-valued taper, i.e., $\nu^I = 0$ and $\nu = \nu^R$, eq. (7) follows from (11) and (12) above.

Representation of $\hat{I}^\nu(w)$ as a difference of Welch estimators. From eq. (10) it is immediate that $\hat{I}^\nu(w) = \tilde{T}^{\nu^R}(w) - \tilde{T}^{\nu^I}(w)$ where $\tilde{T}^{\nu^R}(w), \tilde{T}^{\nu^I}(w)$ are regular Welch estimators corresponding to the real-valued tapers ν^R, ν^I respectively.

Representation of $\hat{I}_c^T(w)$ as a difference of Bartlett estimators with different block sizes. Eq. (10) with the assignment (9) imply: $\text{Re}[I_k^\nu(w)] = \frac{1}{b}|\sum_{t=1}^b \sqrt{\frac{1}{1-c}}e^{iwt}X_t^{[k]}|^2 - \frac{1}{b}|\sum_{s=1}^{[cb]}\sqrt{\frac{1}{1-c}}e^{iws}X_s^{[k]}|^2$ $= \frac{1}{1-c}\left(\frac{1}{b}|\sum_{t=1}^b e^{iwt}X_t^{[k]}|^2 - \frac{c}{cb}|\sum_{s=1}^{[cb]} e^{iws}X_s^{[k]}|^2\right)$, where $[cb]$ denotes the integer part of cb. We recognize the first term in the parenthesis as the (untapered) periodogram of the block $(X_1^{[k]}, \ldots, X_b^{[k]})$, whereas the second term is c times the (untapered) periodogram of block $(X_1^{[k]}, \ldots, X_{[cb]}^{[k]})$. Taking the average over the $q = n - b + 1$ blocks we obtain:

$$\hat{I}_c^T(w) = \frac{1}{1-c}\frac{1}{q}\sum_{k=1}^q \frac{1}{b}|\sum_{t=1}^b e^{iwt}X_t^{[k]}|^2 - \frac{c}{1-c}\frac{1}{q}\sum_{k=1}^q \frac{1}{cb}|\sum_{s=1}^{[cb]} e^{iws}X_s^{[k]}|^2. \tag{13}$$

Denote the Bartlett estimator using block b as $\hat{f}_b^B(w) = \frac{1}{q}\sum_{k=1}^q \frac{1}{b}|\sum_{t=1}^b e^{iwt}X_t^{[k]}|^2$. Thus, the Bartlett estimator using block $[cb]$ is $\hat{f}_{[cb]}^B(w) = \frac{1}{q'}\sum_{k=1}^{q'} \frac{1}{[cb]}|\sum_{s=1}^{[cb]} e^{iws}X_s^{[k]}|^2$ where $q' = n - [cb] + 1$. Let $\epsilon^2 = \sum_{k=q+1}^{q'} \frac{1}{[cb]}|\sum_{s=1}^{[cb]} e^{iws}X_s^{[k]}|^2 = O_P(q'-q) = O_P(b-cb) = O_P(b)$. Now eq. (13) implies:

$$\hat{I}_c^T(w) = \frac{1}{1-c}\hat{f}_b^B(w) - \frac{c}{1-c}\cdot\frac{1}{q}\left(q'\hat{f}_{[cb]}^B(w) - \epsilon^2\right) = \frac{1}{1-c}\hat{f}_b^B(w) - \frac{c}{1-c}\hat{f}_{[cb]}^B(w) + O_P(\frac{b}{q});$$

using the fact that $b/n \to 0$ and the short-hand notation $h = c/(1-c)$ we finally arrive at:

$$\hat{I}_c^T(w) \approx (h+1)\hat{f}_b^B(w) - h\hat{f}_{[cb]}^B(w). \tag{14}$$

The variance of $\hat{I}_c^T(w)$. Recall that $\text{Var}\,(\hat{f}_b^B(w)) \approx \frac{2}{3}\frac{b}{n}f^2(w)(1 + \mathbf{1}_{\{w/\pi\in\mathbb{Z}\}})$. It can also be shown [6] that the correlation coefficient $\text{Corr}(\hat{f}_b^B(w), f_{[cb]}^B(w)) \approx \frac{3-c}{2}\sqrt{c}$. Thus, from eq. (14), we have

$$\text{Var}\,(\hat{I}_c^T(w)) \approx \frac{3h+1}{h+1}\left(\frac{2b}{3n}\right)f^2(w)(1 + \mathbf{1}_{\{w/\pi\in\mathbb{Z}\}}). \tag{15}$$

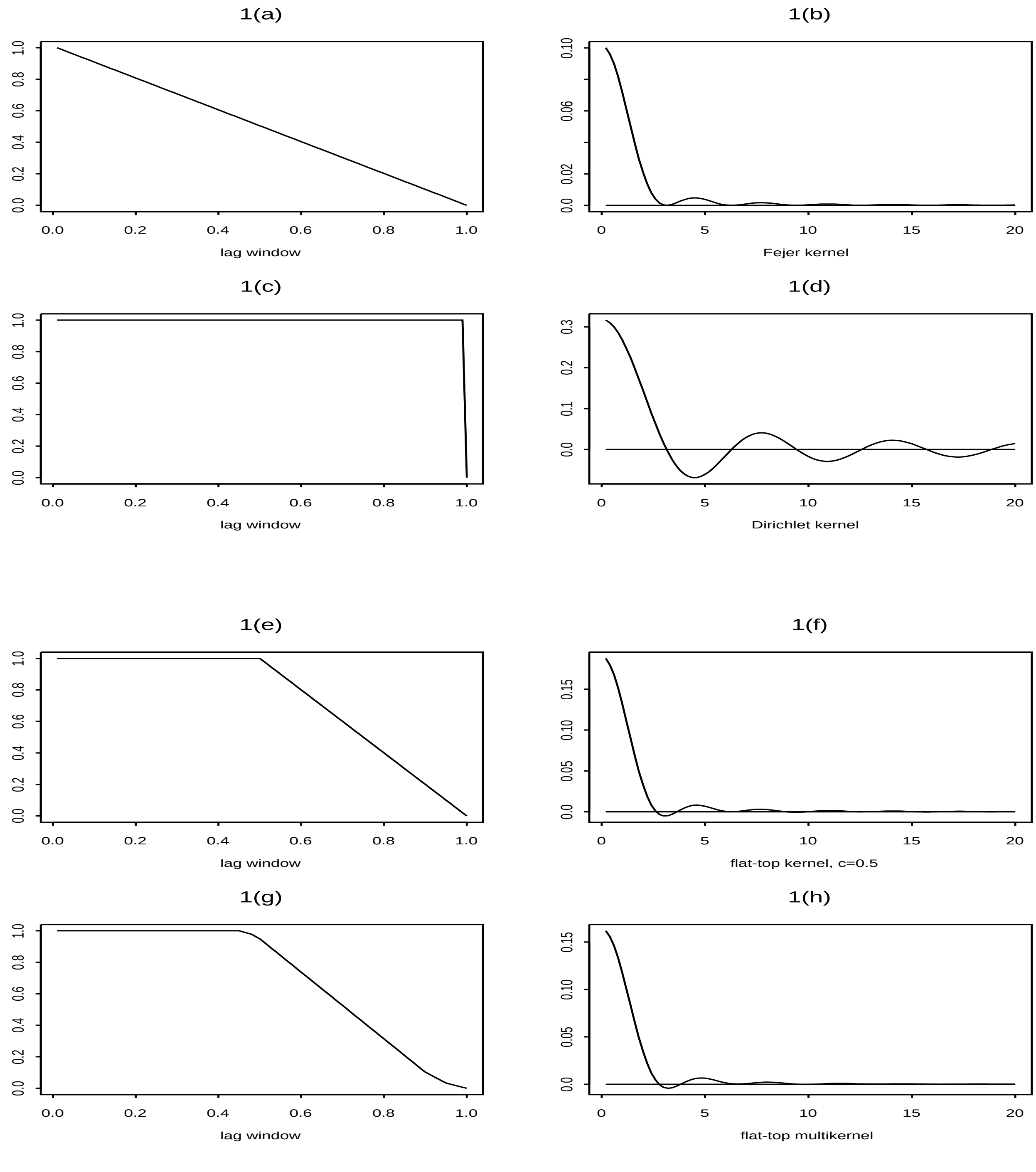

Figure 1: The lag-window $\lambda_c^T(x)$, for $x > 0$, and the kernel $\Lambda_c^T(w)$, for $w > 0$, corresponding to $\hat{I}_c^T(w)$. (a) $\lambda_0^T(x)$; (b) $\Lambda_0^T(w)$; (c) $\lambda_c^T(x)$ with $c \simeq 1$; (d) $\Lambda_c^T(w)$ with $c \simeq 1$; (e) $\lambda^T(x)$, i.e., $\lambda_{1/2}^T(x)$; (f) $\Lambda^T(w)$, i.e., $\Lambda_{1/2}^T(w)$; (g) $\bar{\lambda}^T(x) = (1/3)[\lambda^T(x) + \lambda^T(x/0.95) + \lambda^T(x/0.90)]$; (h) $\bar{\Lambda}^T(w)$.