7

Lag Window Spectral Estimators

7.0 Introduction

In the previous chapter, we considered the periodogram $\hat{S}^{(\mathrm{P})}(\cdot)$ as a seemingly "natural" estimator of a spectral density function (SDF) based upon a time series we regard as a realization of a portion $X_0, X_1, \ldots, X_{N-1}$ of a stationary process. The periodogram unfortunately is of limited use, both because it can be badly biased and because of its unduly large variance, which does not decrease as the sample size N increases. With an appropriate data taper, use of a direct spectral estimator $\hat{S}^{(\mathrm{D})}(\cdot)$ can reduce the bias considerably; however, the variance of $\hat{S}^{(\mathrm{D})}(f)$ has the same properties as that of the periodogram. This is problematic from at least two viewpoints. First, we need to be able to see the basic structure in spectral estimates when plotted versus frequency – the large variability in $\hat{S}^{(\mathrm{D})}(\cdot)$ can mask potentially important features. Second, spectral estimators with better variance properties than $\hat{S}^{(\mathrm{D})}(\cdot)$ can lead to more powerful statistical tests of certain hypotheses of interest.

In this chapter and the next, we look at schemes that use direct spectral estimators as building blocks to form estimators with better variance properties. This chapter is devoted to *lag window spectral estimators*, which reduce variance by smoothing $\hat{S}^{(\mathrm{D})}(\cdot)$ across frequencies (smoothing is actually done by applying a so-called lag window to the estimator $\{\hat{s}_\tau^{(\mathrm{D})}\}$ of the autocovariance sequence (ACVS) corresponding to $\hat{S}^{(\mathrm{D})}(\cdot)$). In Section 7.1 we discuss the rationale for smoothing $\hat{S}^{(\mathrm{D})}(\cdot)$, which leads to the lag window formulation. This estimator is associated with a *smoothing window* (a filter related to the lag window, but operating directly on $\hat{S}^{(\mathrm{D})}(\cdot)$). We discuss the connection between lag and smoothing windows, consider some of their basic properties and introduce the notion of *smoothing window bandwidth*, which characterizes the degree of smoothing to which $\hat{S}^{(\mathrm{D})}(\cdot)$ is subjected. We devote Sections 7.2 and 7.3 to studying the first- and second-moment properties of lag window estimators, after which we consider their large-sample distributions (Section 7.4). Section 7.5 has six examples of lag windows and their associated smoothing windows. We discuss how to choose amongst them in Section 7.6. The common theme of Sections 7.7, 7.8 and 7.9 is setting of the smoothing window bandwidth, which controls the overall quality of the SDF estimator for a particular time series. Section 7.11 discusses computational details. Section 7.12 has five examples of lag window spectral estimation based on actual time series. We close the chapter with a summary (Section 7.13) and exercises for the reader (Section 7.14).

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

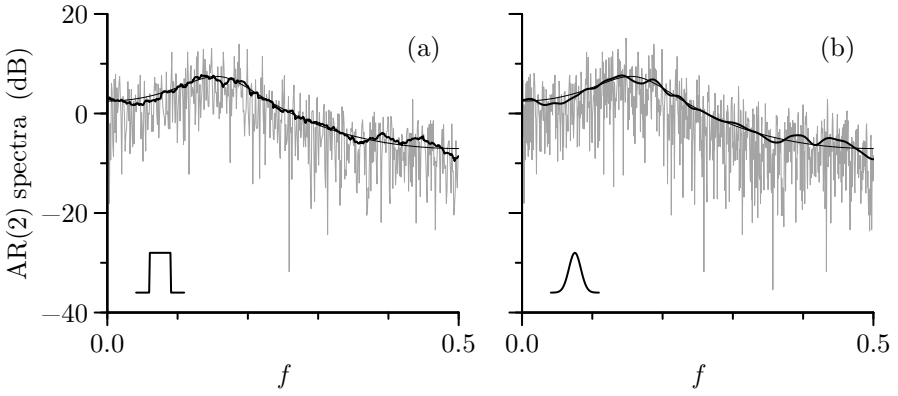

Figure 246 Examples of discretely smoothed direct spectral estimates $\hat{S}^{(\mathrm{DS})}(\cdot)$. The rough-looking light curve in (a) shows the periodogram evaluated over the Fourier frequencies $f_k = k/(N\,\Delta_t)$ for the AR(2) time series shown in Figure 34(a), for which $\Delta_t = 1$. The thick dark curve is $\hat{S}^{(\mathrm{DS})}(\cdot) = \overline{S}(\cdot)$, which is based on a "running average" filter with a span of $2M + 1 = 31$. The impulse response sequence for this filter is depicted in the lower left-hand corner on a linear/linear scale (its vertical axis is not depicted). The thin dark curve is the true SDF for the AR(2) process. Plot (b) shows a second example of $\hat{S}^{(\mathrm{DS})}(\cdot)$. This estimate is based on smoothing the periodogram (now evaluated over the frequencies $f'_k = k/(2N\,\Delta_t)$) using a filter whose impulse response sequence is proportional to samples from a Gaussian probability density function (PDF). Both these examples of $\hat{S}^{(\mathrm{DS})}(\cdot)$ are clearly better overall estimates of the true SDF than the periodogram.

7.1 Smoothing Direct Spectral Estimators

As mentioned in the introduction, lag window estimators are based on smoothing a direct spectral estimator $\hat{S}^{(\mathrm{D})}(\cdot)$ across frequencies. The justification for this approach is the following. Suppose N is large enough so that the periodogram $\hat{S}^{(\mathrm{P})}(\cdot)$ is essentially an unbiased estimator of $S(\cdot)$ and is pairwise uncorrelated at the Fourier frequencies f_k (i.e., the approximations stated by Equation (204a) are valid). If $S(\cdot)$ is slowly varying in the neighborhood of, say, f_k, then

$$S(f_{k-M}) \approx \cdots \approx S(f_k) \approx \cdots \approx S(f_{k+M})$$

for some integer $M > 0$. Thus $\hat{S}^{(\mathrm{P})}(f_{k-M})$, …, $\hat{S}^{(\mathrm{P})}(f_k), \ldots, \hat{S}^{(\mathrm{P})}(f_{k+M})$ are a set of $2M + 1$ approximately unbiased and uncorrelated estimators of the same quantity, namely, $S(f_k)$. Suppose we average them to form the estimator

$$\overline{S}(f_k) \stackrel{\mathrm{def}}{=} \frac{1}{2M+1} \sum_{j=-M}^{M} \hat{S}^{(\mathrm{P})}(f_{k-j}). \tag{246a}$$

(Figure 246(a) shows an example of $\overline{S}(\cdot)$). Under our assumptions we have

$$E\{\overline{S}(f_k)\} \approx S(f_k) \;\text{ and }\; \mathrm{var}\,\{\overline{S}(f_k)\} \approx \frac{S^2(f_k)}{2M+1} \approx \frac{\mathrm{var}\,\{\hat{S}^{(\mathrm{P})}(f_k)\}}{2M+1},$$

where we have made use of Equation (203f) under the simplifying assumption that $f_{k-M} > 0$ and $f_{k+M} < f_{\mathcal{N}}$. If we now consider increasing both the sample size N and the index k in such a way that $k/(N\,\Delta_t) = f_k$ is held constant, we can then let M get large also and claim that $\mathrm{var}\,\{\overline{S}(f_k)\}$ can be made arbitrarily small so that $\overline{S}(f_k)$ is a consistent estimator of $S(f_k)$.

The estimator $\overline{S}(f_k)$ is a special case of a more general spectral estimator of the form

$$\hat{S}^{(\mathrm{DS})}(f'_k) \stackrel{\mathrm{def}}{=} \sum_{j=-M}^{M} g_j \hat{S}^{(\mathrm{D})}(f'_{k-j}) \;\text{ with }\; f'_k = \frac{k}{N'\,\Delta_t}, \tag{246b}$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

where $\{g_j\}$ is a sequence of $2M+1$ smoothing coefficients with $g_{-j} = g_j$; $\hat{S}^{(\mathrm{D})}(\cdot)$ is a direct spectral estimator (Equation (186b)); and N' is a positive integer that controls the spacing of the frequencies over which the smoothing occurs. Typically we choose $N' \geq N$ (the sample size) so that the frequencies f'_k are at least as closely spaced as the Fourier frequencies $f_k = k/(N\,\Delta_\mathrm{t})$. We call $\hat{S}^{(\mathrm{DS})}(f'_k)$ a *discretely smoothed direct spectral estimator* (see Figure 246 for two examples). Note that we can regard $\{g_j\}$ as the coefficients of an LTI digital filter.

Now the estimator $\hat{S}^{(\mathrm{DS})}(f'_k)$ is formed by smoothing the direct spectral estimator $\hat{S}^{(\mathrm{D})}(\cdot)$ with a *discrete* convolution over a *discrete* set of frequencies. Because $\hat{S}^{(\mathrm{D})}(\cdot)$ is defined for all $f \in [-f_\mathcal{N}, f_\mathcal{N}]$, we can also smooth it using a *continuous* convolution over a *continuous* set of frequencies. We thus consider an estimator of the form

$$\hat{S}_m^{(\mathrm{LW})}(f) = \int_{-f_\mathcal{N}}^{f_\mathcal{N}} V_m(f-\phi)\hat{S}^{(\mathrm{D})}(\phi)\,\mathrm{d}\phi, \tag{247a}$$

where the *design window* $V_m(\cdot)$ is a symmetric real-valued $2f_\mathcal{N}$ periodic function that is square integrable over $[-f_\mathcal{N}, f_\mathcal{N}]$ and whose smoothing properties can be controlled by a parameter m (the rationale for putting LW on $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ is given following Equation (248b)).

▷ **Exercise [247]** Recalling from Equations (188a) and (188b) that $\{\hat{s}_\tau^{(\mathrm{D})}\} \longleftrightarrow \hat{S}^{(\mathrm{D})}(\cdot)$, show that we can rewrite Equation (247a) as

$$\hat{S}_m^{(\mathrm{LW})}(f) = \Delta_\mathrm{t} \sum_{\tau=-(N-1)}^{N-1} v_{m,\tau}\hat{s}_\tau^{(\mathrm{D})} \mathrm{e}^{-\mathrm{i}2\pi f\tau\,\Delta_\mathrm{t}}, \tag{247b}$$

where

$$v_{m,\tau} \stackrel{\mathrm{def}}{=} \int_{-f_\mathcal{N}}^{f_\mathcal{N}} V_m(\phi)\mathrm{e}^{\mathrm{i}2\pi\phi\tau\,\Delta_\mathrm{t}}\,\mathrm{d}\phi. \tag{247c}$$ ◁

We note that $\{v_{m,\tau}\}$ and the design window $V_m(\cdot)$ are a Fourier transform pair:

$$\{v_{m,\tau}\} \longleftrightarrow V_m(\cdot).$$

Equation (247b) tells us that we can compute $\hat{S}_m^{(\mathrm{LW})}(f)$ using a finite number of numerical operations even though it is defined in Equation (247a) via a continuous convolution over a continuous set of frequencies. We note that Equation (247b) does not involve $v_{m,\tau}$ for $|\tau| \geq N$ but that $v_{m,\tau}$ need *not* be zero for these values of τ. For theoretical purposes, we will find it convenient to set

$$w_{m,\tau} \stackrel{\mathrm{def}}{=} \begin{cases} v_{m,\tau}, & |\tau| < N; \\ 0, & |\tau| \geq N, \end{cases} \tag{247d}$$

and

$$W_m(f) \stackrel{\mathrm{def}}{=} \Delta_\mathrm{t} \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}\mathrm{e}^{-\mathrm{i}2\pi f\tau\,\Delta_\mathrm{t}} \tag{247e}$$

so that

$$\{w_{m,\tau}\} \longleftrightarrow W_m(\cdot).$$

Note that $W_m(\cdot)$ is identical to the design window $V_m(\cdot)$ if $v_{m,\tau} = 0$ for $|\tau| \geq N$; however, even if this is not true, $V_m(\cdot)$ and $W_m(\cdot)$ are equivalent as far as Equation (247a) is concerned because we have

$$\int_{-f_\mathcal{N}}^{f_\mathcal{N}} W_m(f-\phi)\hat{S}^{(\mathrm{D})}(\phi)\,\mathrm{d}\phi = \int_{-f_\mathcal{N}}^{f_\mathcal{N}} V_m(f-\phi)\hat{S}^{(\mathrm{D})}(\phi)\,\mathrm{d}\phi = \hat{S}_m^{(\mathrm{LW})}(f) \tag{247f}$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

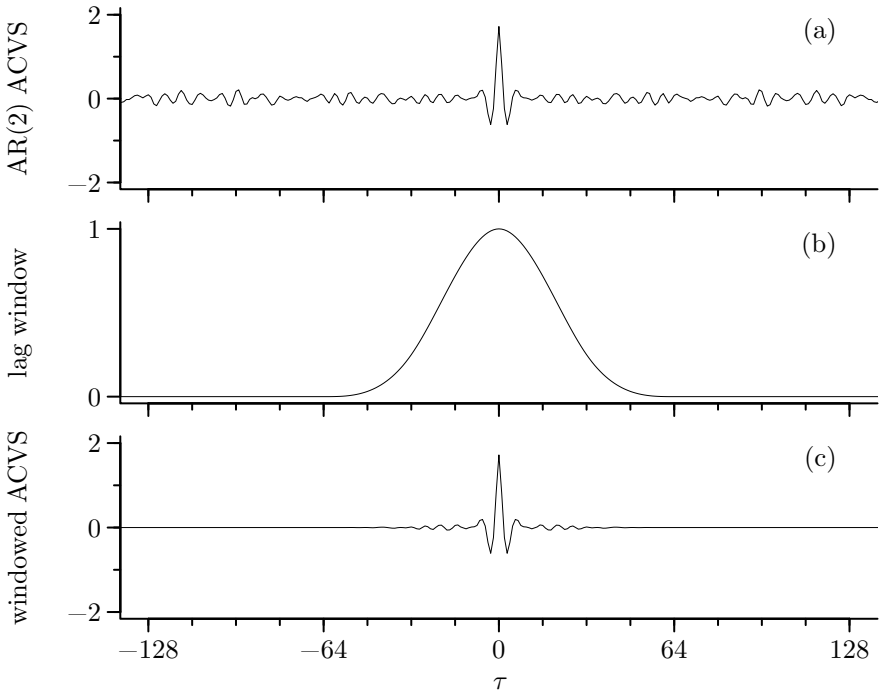

Figure 248 Portion of the biased estimate $\{\hat{s}_\tau^{(\mathrm{P})}\}$ of the ACVS for AR(2) time series shown in Figure 34(a), a Parzen lag window $\{w_{m,\tau}\}$ and their product of $\{w_{m,\tau}\hat{s}_\tau^{(\mathrm{P})}\}$ (the Parzen lag window is given by Equation (275a), and here m = 64). The corresponding spectral estimate $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ is shown in Figure 249.

for all f; i.e., the estimator $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ will be *identically* the same whether we use $V_m(\cdot)$ or $W_m(\cdot)$ (see Exercise [7.1]). The essential distinction between $V_m(\cdot)$ and $W_m(\cdot)$ is that $w_{m,\tau}$ is guaranteed to be 0 for $|\tau| \geq N$ whereas $v_{m,\tau}$ need not be so.

We can thus assume that the estimator $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ can be written as

$$\hat{S}_m^{(\mathrm{LW})}(f) \stackrel{\mathrm{def}}{=} \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} W_m(f-\phi)\hat{S}^{(\mathrm{D})}(\phi)\,\mathrm{d}\phi = \Delta_{\mathrm{t}} \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}\hat{s}_\tau^{(\mathrm{D})} \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_{\mathrm{t}}}, \quad (248a)$$

where $\{w_{m,\tau}\} \longleftrightarrow W_m(\cdot)$ with $w_{m,\tau} = 0$ for $|\tau| \geq N$. If we set

$$\hat{s}_{m,\tau}^{(\mathrm{LW})} \stackrel{\mathrm{def}}{=} \begin{cases} w_{m,\tau}\hat{s}_\tau^{(\mathrm{D})}, & |\tau| \leq N-1; \\ 0, & |\tau| \geq N, \end{cases} \quad (248b)$$

we have $\{\hat{s}_{m,\tau}^{(\mathrm{LW})}\} \longleftrightarrow \hat{S}_m^{(\mathrm{LW})}(\cdot)$; i.e., the sequence $\{\hat{s}_{m,\tau}^{(\mathrm{LW})}\}$ is the estimator of the ACVS corresponding to $\hat{S}_m^{(\mathrm{LW})}(\cdot)$.

We call the function $W_m(\cdot)$ defined in Equation (247e) a *smoothing window* (although most other authors would call it a spectral window, a term we reserve for another concept); its inverse Fourier transform $\{w_{m,\tau}\}$ is called a *lag window* (other names for it are *quadratic window* and *quadratic taper*). We call $\hat{S}_m^{(\mathrm{LW})}(f)$ a *lag window spectral estimator* of $S(f)$ (see item [1] in the Comments and Extensions [C&Es]). In practice, we specify a lag window spectral estimator via either a design window $V_m(\cdot)$ or a lag window $\{w_{m,\tau}\}$ – the smoothing window $W_m(\cdot)$ follows once one of these is given. Figure 248 shows examples of an ACVS estimate, a lag window and their product, while Figure 249 shows the corresponding lag window spectral estimate.

How are the spectral estimators $\{\hat{S}^{(\mathrm{DS})}(f'_k)\}$ and $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ related?

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

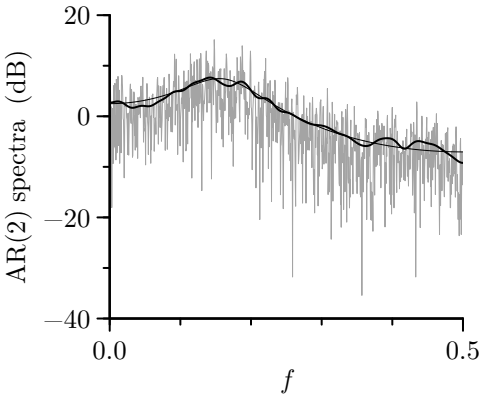

Figure 249 Example of lag window estimate $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ (thick dark curve). This estimate is based on smoothing the periodogram $\hat{S}^{(\mathrm{P})}(\cdot)$ for the AR(2) time series of Figure 34(a) based upon a Parzen lag window $\{w_{m,\tau}\}$ with $m = 64$. The periodogram, evaluated over frequencies $\tilde{f}_k = k/(2N)$, is the rough-looking light curve, while the true SDF for the AR(2) process is the thin dark curve. Note that this lag window estimate is virtually identical to the discretely smoothed direct spectral estimate shown in Figure 246(b). A portion of the ACVS estimate corresponding to $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ is shown in the bottom panel of Figure 248.

▷ **Exercise [249a]** Show that

$$\hat{S}^{(\mathrm{DS})}(f'_k) = \Delta_t \sum_{\tau=-(N-1)}^{N-1} v_{g,\tau} \hat{s}_\tau^{(\mathrm{D})} e^{-i2\pi f'_k \tau \Delta_t}, \tag{249a}$$

where

$$v_{g,\tau} \stackrel{\mathrm{def}}{=} \sum_{j=-M}^{M} g_j e^{i2\pi f'_j \tau \Delta_t} = \sum_{j=-M}^{M} g_j e^{i2\pi j\tau/N'}. \tag{249b}$$ ◁

Hence any discretely smoothed direct spectral estimator can be expressed as a lag window spectral estimator with an implicitly defined lag window given by

$$w_{g,\tau} \stackrel{\mathrm{def}}{=} \begin{cases} v_{g,\tau}, & |\tau| \le N-1; \\ 0, & |\tau| \ge N, \end{cases} \tag{249c}$$

and a corresponding smoothing window given by Equation (247e) with $w_{m,\tau}$ replaced by $w_{g,\tau}$.

Now consider any lag window spectral estimator $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ based upon $\{w_{m,\tau}\} \longleftrightarrow W_m(\cdot)$. Although this estimator is defined for all $f \in \mathbb{R}$, it is completely determined if we know $\{\hat{s}_{m,\tau}^{(\mathrm{LW})}\}$ because $\{\hat{s}_{m,\tau}^{(\mathrm{LW})}\} \longleftrightarrow \hat{S}_m^{(\mathrm{LW})}(\cdot)$. By definition $\hat{s}_{m,\tau}^{(\mathrm{LW})} = 0$ when $|\tau| \ge N$. For $|\tau| < N$, we find, using the same argument that led to Equation (171e), that

$$\{\hat{s}_{m,\tau}^{(\mathrm{LW})} : \tau = -(N-1), \ldots, N-1\} \longleftrightarrow \{\hat{S}_m^{(\mathrm{LW})}(f'_k) : k = -(N-1), \ldots, N-1\},$$

where, as before, $f'_k = k/[(2N-1)\,\Delta_t]$ defines a grid of frequencies almost twice as fine as the Fourier frequencies $f_k \stackrel{\mathrm{def}}{=} k/(N\,\Delta_t)$. Hence $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ is completely determined once we know it over the finite set of frequencies f'_k.

▷ **Exercise [249b]** Show that

$$\hat{S}_m^{(\mathrm{LW})}(f'_k) = \sum_{j=-(N-1)}^{N-1} g_j \hat{S}^{(\mathrm{D})}(f'_{k-j}), \quad \text{where } g_j \stackrel{\mathrm{def}}{=} \frac{W_m(f'_j)}{(2N-1)\,\Delta_t}. \tag{249d}$$ ◁

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

Hence any lag window spectral estimator $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ can be expressed at $f = f'_k$ as a discretely smoothed direct spectral estimator over the set of frequencies $\{f'_j : j = -(N-1), \ldots, N-1\}$ with weights g_j that are proportional to the smoothing window at frequency f'_j.

From these arguments we can conclude that the class of lag window spectral estimators and the class of discretely smoothed direct spectral estimators are equivalent. From a practical point of view, lag window estimators are typically easier to compute because they are based on multiplications whereas discretely smoothed direct spectral estimators are based on convolutions (see Section 7.11 for details). For this reason, we will concentrate mainly on lag window spectral estimators.

Given a direct spectral estimator, a lag window estimator $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ is fully specified by its lag window $\{w_{m,\tau}\}$. So far we have only assumed that $w_{m,\tau} = 0$ for $|\tau| \geq N$, but this condition by itself is not sufficient to get what we want out of $\hat{S}_m^{(\mathrm{LW})}(\cdot)$, namely, an estimator with smaller variance than $\hat{S}^{(\mathrm{D})}(\cdot)$. Indeed, if we let $w_{m,\tau} = 1$ for all $|\tau| \leq N-1$ in Equation (248a), we find that $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ reduces to $\hat{S}^{(\mathrm{D})}(\cdot)$ (compare Equations (188a) and (248a)). Not surprisingly, we must impose some conditions on the lag window $\{w_{m,\tau}\}$ or, equivalently, the smoothing window $W_m(\cdot)$ to ensure that $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ is a reasonable SDF estimator with better statistical properties than $\hat{S}^{(\mathrm{D})}(\cdot)$. We thus make the following three assumptions.

[1] Since $\hat{s}_{-\tau}^{(\mathrm{D})} = \hat{s}_{\tau}^{(\mathrm{D})}$, we require that $w_{m,-\tau} = w_{m,\tau}$ for all τ and all choices of m. This implies that $W_m(\cdot)$ is an even real-valued function that is periodic with a period of $2f_{\mathcal{N}}$.

[2] We require the normalization

$$\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} W_m(f)\,\mathrm{d}f = 1, \quad \text{or, equivalently, } w_{m,0} = 1, \text{ for all } m \tag{250a}$$

(the equivalence follows from $\{w_{m,\tau}\} \longleftrightarrow W_m(\cdot)$). This stipulation is similar to one we made in Section 5.7 to ensure that a low-pass digital filter passes unaltered a locally linear portion of a sequence. In terms of the g_j weights of Equation (249d), the condition $w_{m,0} = 1$ is equivalent to

$$\sum_{j=-(N-1)}^{N-1} g_j = 1. \tag{250b}$$

(Exercise [7.2] is to prove this statement).

[3] We require that $W_m(\cdot)$ acts more and more like a Dirac delta function as $m \to \infty$. In particular, for any $\delta > 0$ and for all $|f| \in [\delta, f_{\mathcal{N}}]$, we require

$$W_m(f) \to 0 \text{ as } m \to \infty. \tag{250c}$$

For a more technical statement of this and other requirements needed to prove various theorems about lag window estimators, see Priestley (1981, pp. 450–1).

In addition to these assumptions, note that, if

$$W_m(f) \geq 0 \text{ for all } m \text{ and } f, \tag{250d}$$

then it follows from Equation (248a) that, since $\hat{S}^{(\mathrm{D})}(f) \geq 0$ necessarily (consider Equation (186b)), we must have $\hat{S}_m^{(\mathrm{LW})}(f) \geq 0$. Interestingly, the condition of Equation (250d) is sufficient, but not necessary, to ensure the nonnegativity of $\hat{S}_m^{(\mathrm{LW})}(f)$; see the discussion on the Daniell smoothing window in Section 7.5. As we shall see in the discussion concerning Figure 282, there are valid reasons for considering smoothing windows for which $W_m(f) < 0$ for some values of f. Thus, whereas we always require the just-listed items [1] to [3] to hold

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

in addition to $w_{m,\tau} = 0$ for $|\tau| \geq N$, the condition of Equation (250d) is desirable but not required. Note that, if this condition were to hold, the assumption of Equation (250a) implies that $W_m(\cdot)$ could then be regarded as a PDF for an RV distributed over the interval $[-f_{\mathcal{N}}, f_{\mathcal{N}}]$.

We also need to define a bandwidth for the smoothing window $W_m(\cdot)$ in order to have some idea about the range of frequencies that influences the value of $\hat{S}_m^{(\mathrm{LW})}(f)$. When the condition of Equation (250d) holds, we can regard $W_m(\cdot)$ as a PDF for an RV that assumes values restricted to the interval $[-f_{\mathcal{N}}, f_{\mathcal{N}}]$. A convenient measure of width is then given by Equation (60c):

$$\beta_W \stackrel{\text{def}}{=} \left(12 \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} f^2 W_m(f)\, \mathrm{d}f \right)^{1/2}. \tag{251a}$$

This definition of window bandwidth is essentially due to Grenander (1951, p. 525) although he omits the factor of 12. As noted in Section 3.4, inclusion of this factor makes the bandwidth of a rectangular smoothing window equal to its natural width (this is related to the Daniell smoothing window – see Section 7.5). A convenient computational formula for β_W is

$$\beta_W = \frac{1}{\Delta_{\mathrm{t}}} \left(1 + \frac{12}{\pi^2} \sum_{\tau=1}^{N-1} \frac{(-1)^\tau}{\tau^2} w_{m,\tau} \right)^{1/2} \tag{251b}$$

(cf. Equation (192b); the extra Δ_{t} in the right-hand side of Equation (192b) can be explained by the fact that $\{w_{m,\tau}\} \longleftrightarrow W_m(\cdot)$, whereas $\{h \star h_\tau\} \longleftrightarrow \Delta_{\mathrm{t}}\, \mathcal{H}(\cdot)$).

There is, however, a potential danger in using β_W to define the smoothing window bandwidth. If the nonnegativity condition of Equation (250d) does not hold, the integral in Equation (251a) can actually be negative, in which case β_W is imaginary (C&E [3] for Section 7.5 has an example of this happening)! Because of this potential problem, we prefer the following measure for the smoothing window bandwidth due to Jenkins (1961):

$$B_W \stackrel{\text{def}}{=} \frac{1}{\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} W_m^2(f)\, \mathrm{d}f}. \tag{251c}$$

Since $W_m(\cdot)$ is real-valued, so must be B_W. This bandwidth measure is just the autocorrelation width of $W_m(\cdot)$ as given by Equation (100f) (the integral in the numerator of Equation (100f) is unity due to Equation (250a)). Since $\{w_{m,\tau}\}$ and $W_m(\cdot)$ are a Fourier transform pair, the Parseval's relationship of Equation (75b) states that

$$\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} W_m^2(f)\, \mathrm{d}f = \Delta_{\mathrm{t}} \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}^2, \tag{251d}$$

so we also have

$$B_W = \frac{1}{\Delta_{\mathrm{t}} \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}^2}. \tag{251e}$$

(See C&E [2] for an additional way of viewing B_W.)

Several other measures of window bandwidth are used widely in the literature (Priestley, 1981, p. 520). For example, Parzen (1957) proposed using the equivalent width of $W_m(\cdot)$:

$$B_{W,P} \stackrel{\text{def}}{=} \mathrm{width_e}\left\{W_m(\cdot)\right\} = \frac{1}{W_m(0)} = \frac{1}{\Delta_{\mathrm{t}} \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}} \tag{251f}$$

(the above follows from Equation (100d) after recalling that $W_m(\cdot)$ integrates to unity and that $w_{m,0} = 1$). Figure 252 shows an example of a smoothing window along with depictions of its bandwidth as measured by β_W, B_W and $B_{W,P}$.

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

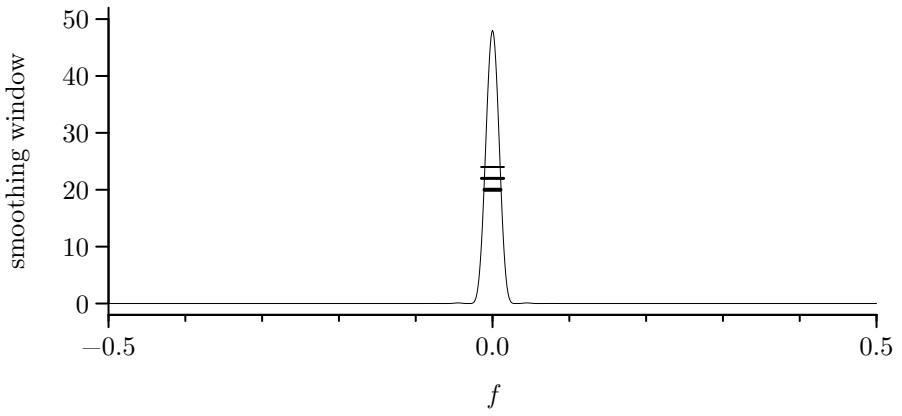

Figure 252 Example of a nonnegative smoothing window $W_m(\cdot)$ and its bandwidth as measured by β_W, B_W and $B_{W,P}$ (given by, from top to bottom, the widths of the three horizontal lines). Here $W_m(\cdot)$ is the Parzen smoothing window of Equation (275b) with $m = 64$. Noting that $W_{64}(0) = 48$, we can visually deduce the half-power width of $W_m(\cdot)$ (another measure of its bandwidth) from the top-most line since it is plotted vertically at 24.

Comments and Extensions to Section 7.1

[1] The lag window spectral estimator $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ is known by many other names in the literature: quadratic window estimator, weighted covariance estimator, indirect nonparametric estimator, windowed periodogram estimator, smoothed periodogram estimator, Blackman–Tukey estimator, spectrograph estimator, spectral estimator of the Grenander–Rosenblatt type and correlogram method power spectral density estimator – and this list is by no means exhaustive! To compound the confusion, some of these names are also used in the literature to refer to estimators that are slightly different from our definition of $\hat{S}_m^{(\mathrm{LW})}(\cdot)$. For example, our definition (Equation (248a)) is essentially a weighting applied to the ACVS estimator $\{\hat{s}^{(\mathrm{D})}\}$ corresponding to *any* direct spectral estimator; other authors define it – at least implicitly – for just the special case of $\{\hat{s}^{(\mathrm{P})}\}$, the ACVS estimator corresponding to the periodogram.

[2] It can be argued from the material in Sections 7.3 and 7.5 that Jenkins's smoothing window bandwidth B_W is the width of the Daniell smoothing window that yields an estimator with the same large-sample variance as one employing $W_m(\cdot)$. Both Priestley (1981) and Bloomfield (2000) argue *against* use of B_W because of this link with variance – they suggest that window bandwidth should be tied instead to the bias introduced by the smoothing window. This property is enjoyed by β_W (see Equation (256d)), but, as noted prior to Equation (251c), we prefer B_W mainly because of its superior practical utility. In most applications, however, as long as the nonnegativity condition of Equation (250d) holds, the measures β_W, B_W and $B_{W,P}$ are interchangeable (see Figure 252).

[3] Considering the defining equality in Equation (248a), we can interpret a lag window estimator $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ as just applying a linear filter to $\hat{S}^{(\mathrm{D})}(\cdot)$, with the intent of extracting a function (the true SDF) buried in rather substantial noise (dictated by the properties of a χ_2^2 distribution). The linear filter in question is specified by the smoothing window $W_m(\cdot)$ and typically acts as a low-pass filter. A complementary interpretation of $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ comes from comparing the Fourier representation for a periodic function $g_p(\cdot)$. with the representation for $\hat{S}^{(\mathrm{D})}(\cdot)$ in terms of $\{\hat{s}_\tau^{(\mathrm{D})}\}$:

$$g_p(t) = \sum_{n=-\infty}^{\infty} G_n e^{i2\pi f_n t} \quad \text{and} \quad \hat{S}^{(\mathrm{D})}(f) = \sum_{\tau=-\infty}^{\infty} \hat{s}_\tau^{(\mathrm{D})} e^{-i2\pi f \tau},$$

where we have set $\Delta_t = 1$ to help with the comparison (see Equations (49c) and (188a), recalling that $\hat{s}_\tau^{(\mathrm{D})} = 0$ when $|\tau| \geq N$). We can attribute any visual choppiness in $g_p(\cdot)$ as being due to the influence of the G_n coefficients with large $|n|$, i.e., those associated with high frequencies f_n. We can alleviate this choppiness by damping down the offending G_n, with the degree of damping ideally increasing as $|n|$ increases (cf. Equation (69b) and the discussion surrounding Figure 70). Similarly, we can attribute

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

the inherently noisy appearance of $\hat{S}^{(\mathrm{D})}(\cdot)$ as being due to the $\hat{s}_\tau^{(\mathrm{D})}$ terms with large $|\tau|$. Hence we can smooth out $\hat{S}^{(\mathrm{D})}(\cdot)$ by damping down the offending $\hat{s}_\tau^{(\mathrm{D})}$. In view of

$$\sum_{\tau=-\infty}^{\infty} w_{m,\tau} \hat{s}_\tau^{(\mathrm{D})} \mathrm{e}^{-\mathrm{i}2\pi f\tau} = \hat{S}_m^{(\mathrm{LW})}(f)$$

we see that a lag window $\{w_{m,\tau}\}$ does exactly that, assuming that it is similar to the one depicted in the middle panel of Figure 248 (the above is essentially the same as Equation (248a)).

[4] We have noted that any discretely smoothed direct spectral estimator $\hat{S}^{(\mathrm{DS})}(\cdot)$ can be reexpressed as a lag window estimator, with the connection between the smoothing coefficients $\{g_j\}$ and the implicit lag window $\{w_{g,\tau}\}$ given by Equations (249b) and (249c). In view of this, let us reconsider the $\hat{S}^{(\mathrm{DS})}(\cdot)$ shown in Figure 246(a), which is based on rectangular coefficients $\{g_j = 1/(2M+1) : j = -M, \ldots, M\}$ applied over the grid of Fourier frequencies $f_k = k/N$ (here $M = 15$). The implicit lag window is given by

$$w_{g,\tau} = \frac{1}{2M+1} \sum_{j=-M}^{M} \mathrm{e}^{\mathrm{i}2\pi j\tau/N} = \mathcal{D}_{2M+1}\left(\tau/N\right), \quad |\tau| \le N-1,$$

where $\mathcal{D}_{2M+1}(\cdot)$ is Dirichlet's kernel (the above follows from Exercise [1.2e]). The function $\mathcal{D}_{2M+1}(\cdot)$ is an even periodic function with a period of unity. Periodicity implies that $w_{g,N-\tau} = w_{g,\tau}$ for $1 \le \tau \le N-1$, as is illustrated by the thick wiggly curve in Figure 254(a). In view of the previous C&E, the fact that this lag window does not damp down the $\hat{s}_\tau^{(\mathrm{P})}$ at the most extreme lags is cause for concern (however, this is mitigated by the fact that $\hat{s}_\tau^{(\mathrm{P})} \to 0$ as $|\tau| \to N-1$). This questionable property is not inherent to rectangular coefficients, but rather holds for all $\{w_{g,\tau}\}$ arising from use of $\{g_j\}$ over the grid of Fourier frequencies: with N' set to N in Equation (249b), we have

$$w_{g,N-\tau} = \sum_{j=-M}^{M} g_j \mathrm{e}^{\mathrm{i}2\pi j(N-\tau)/N} = \sum_{j=-M}^{M} g_j \mathrm{e}^{\mathrm{i}2\pi j} \mathrm{e}^{-\mathrm{i}2\pi j\tau/N} = \sum_{j=-M}^{M} g_j \mathrm{e}^{\mathrm{i}2\pi j\tau/N} = w_{g,\tau},$$

since $g_{-j} = g_j$. As a second example, the thin smooth curve in Figure 254(a) shows $\{w_{g,\tau}\}$ for Gaussian coefficients g_j that are proportional to $\exp(-j^2/150)$ for $j = -35, \ldots, 35$.

A more reasonable implicit lag window arises if we produce $\hat{S}^{(\mathrm{DS})}(\cdot)$ by smoothing the direct SDF estimator over a grid of frequencies that is denser than what is afforded by the Fourier frequencies f_k. In particular, if we use $\tilde{f}_k = k/(2N)$, the implicit lag window corresponding to rectangular coefficients is

$$w_{g,\tau} = \frac{1}{2M+1} \sum_{j=-M}^{M} \mathrm{e}^{\mathrm{i}2\pi j\tau/(2N)} = \mathcal{D}_{2M+1}\left(\tau/(2N)\right), \quad |\tau| \le N-1.$$

The thick wiggly curve in Figure 254(b) shows this lag window when $M = 30$, while the thin smooth curve is for Gaussian coefficients $g_j \propto \exp(-j^2/600)$, $j = -70, \ldots, 70$. In both cases the aberrant increase in $w_{g,\tau}$ as $|\tau| \to N-1$ is gone. The Gaussian coefficients produced the $\hat{S}^{(\mathrm{DS})}(\cdot)$ shown in Figure 246(b), which is markedly smoother than the one in Figure 246(a), despite the fact that both estimates have comparable bandwidths (see the next C&E). While the primary reason for the improved smoothness is the fact that the implicit lag window for Figure 246(b) no longer has ripples, a secondary reason is that the lag window does not ramp up toward unity for $|\tau|$ near $N-1$, as must be the case if we choose to smooth a direct spectral estimator over the Fourier frequencies f_k rather than over $\tilde{f}_k$. Hence it is better to formulate discretely smoothed direct spectral estimators over a grid of frequencies that is finer than the Fourier frequencies.

[5] The measures of smoothing window bandwidth for a lag window estimator we have considered so far (β_W, B_W and $B_{W,P}$) can all be computed once we know the lag window $\{w_{m,\tau}\}$. One idea for defining corresponding bandwidth measures for a discretely smoothed direct spectral estimator $\hat{S}^{(\mathrm{DS})}(\cdot)$ would be to take its defining smoothing coefficients $\{g_j\}$, form the implicit lag window $\{w_{g,\tau}\}$ via

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

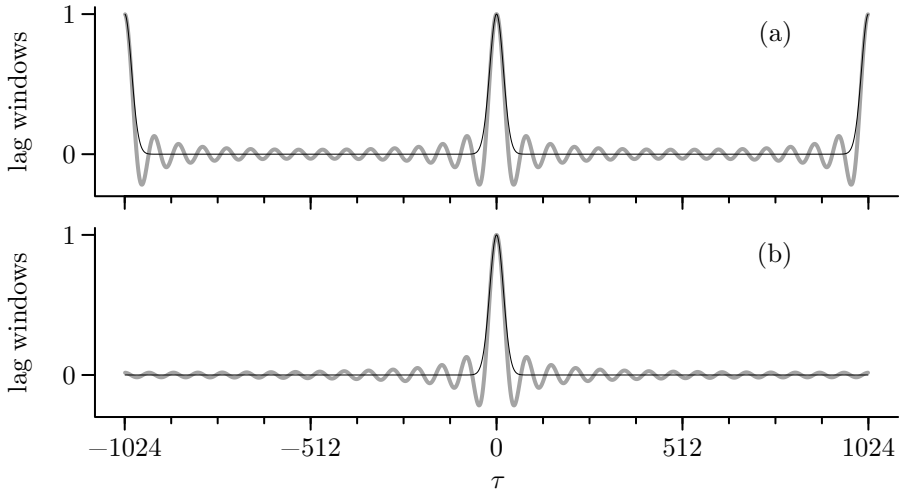

Figure 254 Implicit lag windows $\{w_{g,\tau}\}$ for four discretely smoothed direct spectral estimators $\hat{S}^{(\mathrm{DS})}(\cdot)$. The windows in the top plot correspond to rectangular smoothing coefficients $g_j = 1/31, j = -15, \ldots, 15$, (thick wiggly curve) and Gaussian coefficients $g_j \propto \exp(-j^2/150), j = -35, \ldots, 35$, (thin smooth curve), where the constant of proportionality is set such that $\sum_j g_j = 1$. In both cases, the grid of frequencies over which the direct spectral estimator is smoothed is taken to be the Fourier frequencies $f_k = k/(N\,\Delta_{\mathrm{t}})$, where $N = 1024$. The corresponding curves in the bottom plot are for $g_j = 1/61, j = -30, \ldots, 30$, and $g_j \propto \exp(-j^2/600), j = -70, \ldots, 70$, but now the grid of frequencies is taken to be $\tilde{f}_k = k/(2N\,\Delta_{\mathrm{t}})$ and hence is twice as fine as the grid dictated by the Fourier frequencies.

Equations (249c) and (249b) and then use this as the basis for computing either β_W, B_W or $B_{W,P}$. This approach works in some instances, but can fail when $\{g_j\}$ is used over the grid of Fourier frequencies. As an example, let us reconsider the $\hat{S}^{(\mathrm{DS})}(\cdot)$ shown in Figure 246(a), which is based on the rectangular smoothing coefficients $\{g_j = 1/(2M+1) : j = -M, \ldots, M\}$ with $M = 15$. Since the spacing of the Fourier frequencies is $1/(N\,\Delta_{\mathrm{t}})$, the natural bandwidth is $(2M+1)/(N\,\Delta_{\mathrm{t}})$, which is equal to 0.03027 when $N = 1024$ and $\Delta_{\mathrm{t}} = 1$. Using the implicit lag window (shown by the thick wiggly curve in Figure 254(a)), we obtain the bandwidths listed in the first row of Table 255. While β_W agrees well with the natural bandwidth, the other two measures are off by a factor of two. This disparity can be attributed to how $w_{g,\tau}$ is treated at large $|\tau|$ in Equations (251b), (251e) and (251f). The second row of the table shows that a discrepancy between β_W and the other two measures arises again for the $\hat{S}^{(\mathrm{DS})}(\cdot)$ shown in Figure 246(b), which is based on Gaussian smoothing coefficients $\{g_j\}$ (the implicit lag window is the thin smooth curve in Figure 254(a)). On the other hand, if we use the grid of frequencies $\tilde{f}_k$ with spacing $1/(2N\,\Delta_{\mathrm{t}})$, the natural bandwidth is $(2M+1)/(2N\,\Delta_{\mathrm{t}}) \doteq 0.02979$ for the rectangular smoothing coefficients (M is now set to 30). The third row shows that all three measures now agree well with the natural bandwidth (the corresponding implicit lag window is the thick wiggly curve in Figure 254(b)). The fourth row indicates good agreement between β_W and B_W for the Gaussian coefficients when used with $\tilde{f}_k$, but $B_{W,P}$ is noticeably smaller (the implicit lag window is the thin smooth curve of Figure 254(b)).

Since attaching a bandwidth measure to $\hat{S}^{(\mathrm{DS})}(\cdot)$ via its implicit lag window is problematic, let us consider measures that are based directly on $\{g_j\}$. Let $\Delta_{\mathrm{f}} = 1/(N'\,\Delta_{\mathrm{t}})$ denote the spacing between the frequencies f'_k over which the estimators $\hat{S}^{(\mathrm{D})}(\cdot)$ and $\hat{S}^{(\mathrm{DS})}(\cdot)$ are evaluated in Equation (246b). As suggested by Equation (249d), set $W_m(f'_j) = g_j/\Delta_{\mathrm{f}}$. Use of a Riemann sum to approximate the integrals in Equations (251a) and (251c) yields the following analogs to β_W and B_W:

$$\beta_g \stackrel{\mathrm{def}}{=} \Delta_{\mathrm{f}} \left(24 \sum_{j=1}^{M} j^2 g_j \right)^{1/2} \quad \text{and} \quad B_g \stackrel{\mathrm{def}}{=} \frac{\Delta_{\mathrm{f}}}{\sum_{j=-M}^{M} g_j^2}. \tag{254}$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

g_j	Grid	Bandwidth Measure						
		β_W	B_W	$B_{W,P}$	β_g	B_g	$B_{g,P}$	Natural
Rectangular	$1/N$	0.03025	0.01537	0.01537	0.03026	0.03027	0.03027	0.03027
Gaussian	$1/N$	0.02928	0.01522	0.01071	0.02929	0.02998	0.02120	—
Rectangular	$1/(2N)$	0.02978	0.02979	0.02980	0.02978	0.02979	0.02979	0.02979
Gaussian	$1/(2N)$	0.02928	0.02998	0.02120	0.02928	0.02998	0.02120	—

Table 255 Comparison of window bandwidth measures for discretely smoothed direct spectral estimators $\hat{S}^{(\mathrm{DS})}(\cdot)$. The measures β_W, B_W and $B_{W,P}$ use the implicit lag window associated with $\hat{S}^{(\mathrm{DS})}(\cdot)$, while β_g, B_g and $B_{g,P}$ are based directly on the smoothing coefficients $\{g_j\}$. The smoothing coefficients are either rectangular $g_j = 1/(2M+1)$, $j = -M, \ldots, M$ (with $M = 15$ and 30, respectively, in the first and third rows), or Gaussian $g_j \propto \exp(-j^2/(2\sigma^2))$, $j = -M, \ldots, M$ (with $M = 35$ and $\sigma^2 = 75$ in the second row and $M = 70$ and $\sigma^2 = 300$ in the fourth row). Assuming $\Delta_\mathrm{t} = 1$, the "natural" bandwidth for rectangular smoothing coefficients is $(2M+1)/N$, where $N = 1024$. The grid of frequencies over which $\hat{S}^{(\mathrm{D})}(\cdot)$ is smoothed to produce $\hat{S}^{(\mathrm{DS})}(\cdot)$ corresponds to either the Fourier frequencies $f_k = k/N$ with spacing $1/N$ or $\tilde{f}_k = k/(2N)$ with spacing $1/(2N)$.

The measure B_g is an example of the autocorrelation width for a sequence (see Equation (100e)). The analog of the Parzen measure of bandwidth of Equation (251f) is $B_{g,P} = \Delta_\mathrm{f}/g_0$, which is related to the equivalent width for a sequence (see Equation (100d)). For the rectangular smoothing coefficients $g_j = 1/(2M+1)$, an easy exercise shows that both B_g and $B_{g,P}$ are equal to the natural bandwidth $(2M+1)\,\Delta_\mathrm{f}$, whereas β_g is so to a good approximation since

$$\beta_g = 2\,[M(M+1)]^{1/2}\,\Delta_\mathrm{f} \approx 2(M+\tfrac{1}{2})\Delta_\mathrm{f} = (2M+1)\,\Delta_\mathrm{f}.$$

Table 255 shows that β_g, B_g and $B_{g,P}$ closely agree with, respectively, β_W, B_W and $B_{W,P}$ when the $1/(2N)$ grid is used, but that B_g and $B_{g,P}$ are to be preferred over B_W and $B_{W,P}$ when the $1/N$ grid is used.

7.2 First-Moment Properties of Lag Window Estimators

We now consider the first-moment properties of $\hat{S}_m^{(\mathrm{LW})}(\cdot)$. It follows from Equation (248a) that

$$E\{\hat{S}_m^{(\mathrm{LW})}(f)\} = \int_{-f_\mathcal{N}}^{f_\mathcal{N}} W_m(f-\phi) E\{\hat{S}^{(\mathrm{D})}(\phi)\}\,\mathrm{d}\phi. \tag{255a}$$

▷ **Exercise [255]** Use Equation (186e) to show that

$$E\{\hat{S}_m^{(\mathrm{LW})}(f)\} = \int_{-f_\mathcal{N}}^{f_\mathcal{N}} \mathcal{U}_m(f-f') S(f')\,\mathrm{d}f', \tag{255b}$$

where

$$\mathcal{U}_m(f) \stackrel{\mathrm{def}}{=} \int_{-f_\mathcal{N}}^{f_\mathcal{N}} W_m(f-f'')\mathcal{H}(f'')\,\mathrm{d}f''. \tag{255c}$$ ◁

We call $\mathcal{U}_m(\cdot)$ the *spectral window* for the lag window spectral estimator $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ (just as we called $\mathcal{H}(\cdot)$ the spectral window of $\hat{S}^{(\mathrm{D})}(\cdot)$ – compare Equations (186e) and (255b)). Exercise [7.3] is to show that

$$\mathcal{U}_m(f) = \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}\, h \star h_\tau \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_\mathrm{t}}, \tag{255d}$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

where $\{h \star h_\tau\}$ is the autocorrelation of $\{h_t\}$ (see Equation (192c)). The above is useful for calculating $\mathcal{U}_m(\cdot)$ in practice.

The estimator $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ has several properties that parallel those for $\hat{S}^{(\mathrm{D})}(\cdot)$ described in Section 6.4 – see Exercise [7.4]. In addition, because Equations (186e) and (255b) indicate that $E\{\hat{S}^{(\mathrm{D})}(f)\}$ and $E\{\hat{S}_m^{(\mathrm{LW})}(f)\}$ are both related to $S(f)$ via convolution of a spectral window and the true SDF, we can extend our standard measure of the effective bandwidth of $\hat{S}^{(\mathrm{D})}(\cdot)$ to work with $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ by replacing $\mathcal{H}(f)$ with $\mathcal{U}_m(f)$ in Equation (194).

▷ **Exercise [256]** Taking Equation (255d) as a given, show that

$$B_{\mathcal{U}} \stackrel{\mathrm{def}}{=} \mathrm{width}_{\mathrm{a}}\{\mathcal{U}_m(\cdot)\} = \frac{\left(\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \mathcal{U}_m(f)\,\mathrm{d}f\right)^2}{\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \mathcal{U}_m^2(f)\,\mathrm{d}f} = \frac{\Delta_{\mathrm{t}}}{\sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}^2 (h \star h_\tau)^2}. \quad (256a)$$ ◁

Let us now assume that $\hat{S}^{(\mathrm{D})}(f)$ is an approximately unbiased estimator of $S(f)$, which is a reasonable assumption as long as we have made effective use of tapering (or prewhitening). This allows us to examine the effect of the smoothing window separately on the first-moment properties of $\hat{S}_m^{(\mathrm{LW})}(\cdot)$. From Equation (255a) we then have

$$E\{\hat{S}_m^{(\mathrm{LW})}(f)\} \approx \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} W_m(f - f') S(f')\,\mathrm{d}f'. \quad (256b)$$

The stated requirement involving Equation (250c) says that $W_m(\cdot)$ must act like a Dirac delta function as m gets large. As we shall see in Section 7.5, we cannot make m arbitrarily large if the sample size N is fixed; however, if we let $N \to \infty$, we can also let $m \to \infty$, so we can claim that $\hat{S}_m^{(\mathrm{LW})}(f)$ is an asymptotically unbiased estimator of $S(f)$:

$$E\{\hat{S}_m^{(\mathrm{LW})}(f)\} \to S(f) \text{ as } m, N \to \infty. \quad (256c)$$

For finite sample sizes N and finite values of the smoothing parameter m, however, even if $\hat{S}^{(\mathrm{D})}(\cdot)$ is an approximately unbiased estimator, the smoothing window $W_m(\cdot)$ can introduce significant bias in $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ by inadvertently smoothing together adjacent features in $\hat{S}^{(\mathrm{D})}(\cdot)$ – this happens when the true SDF is not slowly varying. This bias obviously depends upon both the true SDF and the smoothing window. It is informative to derive an expression to quantify this bias (Priestley, 1981, p. 458). Let us take the bias due to the smoothing window alone to be

$$b_W(f) \stackrel{\mathrm{def}}{=} \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} W_m(f - f') S(f')\,\mathrm{d}f' - S(f),$$

a reasonable definition in view of Equation (256b). An argument that exactly parallels the one leading to Equation (192a) yields

$$b_W(f) \approx \frac{S''(f)}{2} \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \phi^2 W_m(\phi)\,\mathrm{d}\phi = \frac{S''(f)}{24} \beta_W^2, \quad (256d)$$

where β_W is the measure of smoothing window bandwidth given in Equation (251a). The bias due to the smoothing window in $\hat{S}_m^{(\mathrm{LW})}(f)$ is thus large if $S(\cdot)$ is varying rapidly at f as measured by its second derivative, and the bias increases as the smoothing window bandwidth increases.

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

Comments and Extensions to Section 7.2

[1] The smoothing window bandwidth B_W measures the width of the central lobe of a smoothing window $W_m(\cdot)$. The measure $B_{\mathcal{U}}$ does the same for a spectral window $\mathcal{U}_m(\cdot)$, which is the convolution of $W_m(\cdot)$ and a spectral window $\mathcal{H}(\cdot)$. Since convolution is a smoothing operation, one interpretation for $\mathcal{U}_m(\cdot)$ is as a smoothed-out version of $W_m(\cdot)$. The central lobe of $\mathcal{U}_m(\cdot)$ should thus be wider than that of $W_m(\cdot)$, so it is reasonable to expect $B_{\mathcal{U}}$ to be bigger than B_W. The following exercise shows this must be true.

▷ **Exercise [257]** Show that $B_{\mathcal{U}} \geq B_W$ (assuming the usual normalization $\sum_t h_t^2 = 1$). Hint: make use of the Cauchy inequality, namely,

$$\left| \sum_{t=0}^{N-1} a_t b_t \right|^2 \leq \sum_{t=0}^{N-1} |a_t|^2 \sum_{t=0}^{N-1} |b_t|^2 , \tag{257a}$$

where the sequences $\{a_t\}$ and $\{b_t\}$ are either real- or complex-valued. (A proof of this inequality can be formulated along the same lines as the one for the Schwarz inequality (Equation (55d)). The proof allows us to note that equality holds if and only if the sequence $\{a_t\}$ is proportional to $\{b_t\}$.) ◁

[2] We have introduced data tapers as a means of compensating for the bias (first moment) of $\hat{S}^{(\mathrm{P})}(\cdot)$ and lag windows as a means of decreasing the variance (second central moment) of direct spectral estimators $\hat{S}^{(\mathrm{D})}(\cdot)$ (including the periodogram). It is sometimes claimed in the literature that a lag window can be used to control both the bias and the variance in $\hat{S}^{(\mathrm{P})}(\cdot)$ (see, for example, Grenander and Rosenblatt, 1984, section 4.2); i.e., we can avoid the use of a nonrectangular data taper. The basis for this claim can be seen by considering the spectral window $\mathcal{U}_m(\cdot)$ in Equation (255d). Equation (255b) says that this spectral window describes the bias properties of a lag window spectral estimator consisting of a lag window $\{w_{m,\tau}\}$ used in combination with a data taper $\{h_t\}$. The following argument shows we can produce the same spectral window by using a different lag window in combination with just the rectangular data taper. First, we note that, in the case of a rectangular data taper, i.e., $h_t = 1/\sqrt{N}$, the spectral window $\mathcal{U}_m(\cdot)$ reduces to

$$\Delta_{\mathrm{t}} \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau} \left(1 - |\tau|/N\right) \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_{\mathrm{t}}}.$$

For any data taper $\{h_t\}$, we can write, using Equation (255d),

$$\begin{aligned} \mathcal{U}_m(f) &= \Delta_{\mathrm{t}} \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau} \frac{\sum_{t=0}^{N-|\tau|-1} h_{t+|\tau|} h_t}{1 - |\tau|/N} \left(1 - |\tau|/N\right) \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_{\mathrm{t}}} \\ &= \Delta_{\mathrm{t}} \sum_{\tau=-(N-1)}^{N-1} w'_{m,\tau} \left(1 - |\tau|/N\right) \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_{\mathrm{t}}}, \end{aligned}$$

where

$$w'_{m,\tau} \stackrel{\mathrm{def}}{=} w_{m,\tau} \frac{\sum_{t=0}^{N-|\tau|-1} h_{t+|\tau|} h_t}{1 - |\tau|/N}. \tag{257b}$$

Thus use of the lag window $\{w'_{m,\tau}\}$ with the rectangular data taper produces the same spectral window $\mathcal{U}_m(\cdot)$ as use of the lag window $\{w_{m,\tau}\}$ with the nonrectangular data taper $\{h_t\}$. Nuttall and Carter (1982) refer to $\{w'_{m,\tau}\}$ as a *reshaped lag window*. As we show by example in C&E [2] for Section 7.5, a potential problem with this scheme is that the smoothing window associated with $\{w'_{m,\tau}\}$ (i.e., its Fourier transform) can have some undesirable properties, including prominent *negative* sidelobes. This means that, for some time series, the associated $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ can be negative at some frequencies.

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

7.3 Second-Moment Properties of Lag Window Estimators

We sketch here a derivation of one approximation to the large-sample variance of $\hat{S}_m^{(\mathrm{LW})}(f)$ – a second, more accurate approximation is considered in C&E [1]. We assume that the direct spectral estimator $\hat{S}^{(\mathrm{D})}(\cdot)$ is approximately uncorrelated on a grid of frequencies defined by $f'_k = k/(N'\,\Delta_\mathrm{t})$, where here $N' \leq N$, and k is an integer such that $0 < f'_k < f_\mathcal{N}$ – see the discussion concerning Equation (207). From Equation (248a) and an obvious change of variable, we have

$$\hat{S}_m^{(\mathrm{LW})}(f) = \int_{-f_\mathcal{N}-f}^{f_\mathcal{N}-f} W_m(-\phi)\hat{S}^{(\mathrm{D})}(\phi+f)\,\mathrm{d}\phi = \int_{-f_\mathcal{N}}^{f_\mathcal{N}} W_m(\phi)\hat{S}^{(\mathrm{D})}(\phi+f)\,\mathrm{d}\phi;$$

here we use the facts that $W_m(\cdot)$ is an even function and that both it and $\hat{S}^{(\mathrm{D})}(\cdot)$ are $2f_\mathcal{N}$ periodic. From Equation (250c) we can assume that, for large N, $W_m(\phi) \approx 0$ for all $|\phi| > J/(N'\,\Delta_\mathrm{t}) = f'_J$ for some positive integer J. Hence

$$\hat{S}_m^{(\mathrm{LW})}(f'_k) \approx \int_{-f'_J}^{f'_J} W_m(\phi)\hat{S}^{(\mathrm{D})}(\phi+f'_k)\,\mathrm{d}\phi \approx \sum_{j=-J}^{J} W_m(f'_j)\hat{S}^{(\mathrm{D})}(f'_j+f'_k)\frac{1}{N'\,\Delta_\mathrm{t}},$$

where we have approximated the Riemann integral by a Riemann sum (see Figure 259). If we note that $f'_j + f'_k = f'_{j+k}$, we now have

$$\begin{aligned}
\mathrm{var}\,\{\hat{S}_m^{(\mathrm{LW})}(f'_k)\} &\approx \frac{1}{(N'\,\Delta_\mathrm{t})^2}\ \mathrm{var}\left\{\sum_{j=-J}^{J} W_m(f'_j)\hat{S}^{(\mathrm{D})}(f'_{j+k})\right\} \\
&\overset{(1)}{\approx} \frac{1}{(N'\,\Delta_\mathrm{t})^2}\sum_{j=-J}^{J} W_m^2(f'_j)\ \mathrm{var}\left\{\hat{S}^{(\mathrm{D})}(f'_{j+k})\right\} \\
&\overset{(2)}{\approx} \frac{1}{(N'\,\Delta_\mathrm{t})^2}\sum_{j=-J}^{J} W_m^2(f'_j)S^2(f'_{j+k}) \\
&\overset{(3)}{\approx} \frac{S^2(f'_k)}{(N'\,\Delta_\mathrm{t})^2}\sum_{j=-J}^{J} W_m^2(f'_j) \\
&\overset{(4)}{\approx} \frac{S^2(f'_k)}{(N'\,\Delta_\mathrm{t})^2}\sum_{j=-\lfloor N'/2\rfloor}^{\lfloor N'/2\rfloor} W_m^2(f'_j) \\
&\overset{(5)}{\approx} \frac{S^2(f'_k)}{N'\,\Delta_\mathrm{t}}\int_{-f_\mathcal{N}}^{f_\mathcal{N}} W_m^2(\phi)\,\mathrm{d}\phi,
\end{aligned} \tag{258}$$

where we have made use of five approximations:

(1) pairwise uncorrelatedness of the components of $\hat{S}^{(\mathrm{D})}(\cdot)$ defined by the grid of frequencies f'_k;
(2) the large-sample variance of $\hat{S}^{(\mathrm{D})}(\cdot)$ as given by Equation (204c), for which we need $0 < f'_{j+k} < f_\mathcal{N}$ for all j;
(3) a smoothness assumption on $S(\cdot)$ asserting that $S(f'_{j+k}) \approx S(f'_k)$ locally;
(4) the previous assumption that $W_m(\phi) \approx 0$ for $|\phi| > f'_J$; and
(5) an approximation of the summation by a Riemann integral.

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

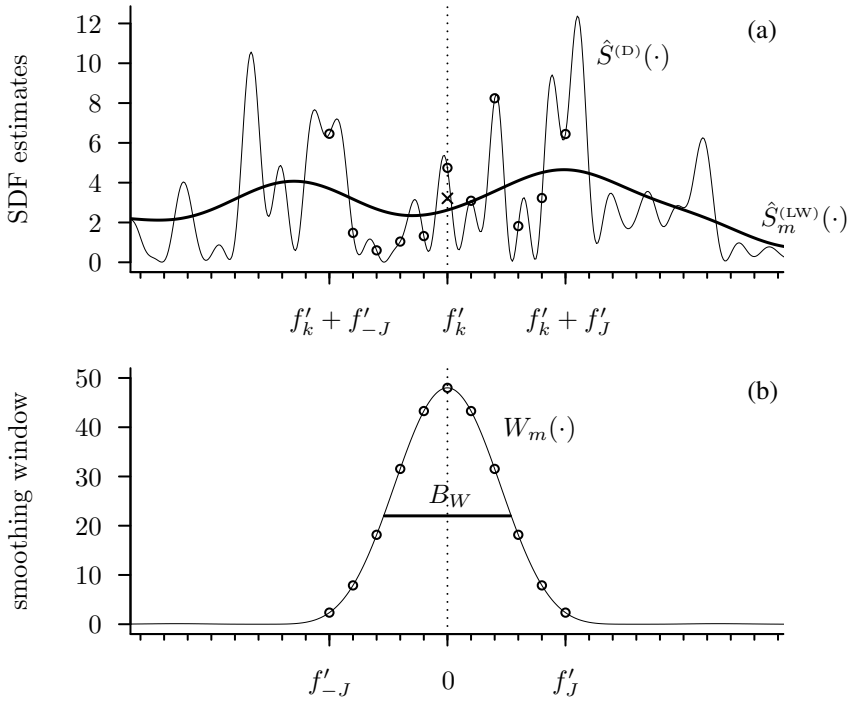

Figure 259 Approximating a lag window estimator via a Riemann sum. The thin curve in plot (a) is a direct spectral estimate $\hat{S}^{(\mathrm{D})}(\cdot)$. The thick curve is a lag window estimate $\hat{S}_m^{(\mathrm{LW})}(\cdot)$, which is the result of convolving $\hat{S}^{(\mathrm{D})}(\cdot)$ with the smoothing window $W_m(\cdot)$ shown in plot (b). The dotted vertical line in (a) indicates the location of the frequency f'_k, with the corresponding estimate $\hat{S}^{(\mathrm{D})}(f'_k)$ marked by a circle. The estimate $\hat{S}_m^{(\mathrm{LW})}(f'_k)$ is formed by multiplying the thin curves in (a) and (b) together and then integrating the product from $-f_{\mathcal{N}}$ to $f_{\mathcal{N}}$. The grid of frequencies over which $\hat{S}^{(\mathrm{D})}(\cdot)$ is approximately both uncorrelated and dictated by a scaled χ^2_2 distribution is given by $f'_j + f'_k$ for integers j such that $0 < f'_j + f'_k < f_{\mathcal{N}}$. The tick marks on the upper horizontal axis show this grid. The smoothing window is essentially zero outside of the interval $[f'_{-J}, f'_J]$ when we set $J = 5$. The eleven values from $\hat{S}^{(\mathrm{D})}(\cdot)$ and $W_m(\cdot)$ used to approximate $\hat{S}_m^{(\mathrm{LW})}(f'_k)$ by a Riemann sum are indicated by the circles. The approximation itself is indicated by the x in plot (a). The width of the thick horizontal line in (b) depicts the smoothing window bandwidth B_W as calculated by Equation (251e). (Here the direct spectral estimator is a periodogram based on the first $N = 256$ values of the AR(2) series of Figure 34(a); the grid is defined by the Fourier frequencies so $N' = N$; $f'_k = k/N' = 38/256$; and $W_m(\cdot)$ is the Parzen smoothing window of Equation (275b) with $m = 64$.)

Finally, we assume that Equation (258) holds for all frequencies (and not just the ones defined by f'_k) and that N' can be related to N by $N' \approx N/C_h$, where C_h is greater than or equal to unity and depends only upon the data taper for the direct spectral estimator (Table 260 gives the value of C_h for some common data tapers, and Equation (262) is a computational formula for it). With these modifications we now have

$$\operatorname{var}\{\hat{S}_m^{(\mathrm{LW})}(f)\} \approx \frac{C_h S^2(f)}{N\,\Delta_{\mathrm{t}}} \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} W_m^2(\phi)\,\mathrm{d}\phi. \tag{259a}$$

Using the definition for the smoothing window bandwidth B_W in Equation (251c), we rewrite the above as

$$\operatorname{var}\{\hat{S}_m^{(\mathrm{LW})}(f)\} \approx \frac{C_h S^2(f)}{B_W N\,\Delta_{\mathrm{t}}}. \tag{259b}$$

Thus, as one would expect, increasing (decreasing) the smoothing window bandwidth causes

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

	$p \times 100\%$ Cosine Taper					Slepian Taper			
	$p =$ 0.0	0.2	0.5	1.0	$NW =$ 1	2	4	8
C_h	1.00	1.12	1.35	1.94		1.34	1.96	2.80	3.98
See Figure 190	(a)	(b)	(c)	(d)		(e)	(f)	(g)	(h)

Table 260 Variance inflation factor C_h for various data tapers. Note that C_h increases as the width of the central lobe of the associated spectral window increases. The values listed here are approximations – valid for large N – to the right-hand side of Equation (262). When $p = 0$, the $p \times 100\%$ cosine taper is the same as the rectangular data taper, which is the taper used implicitly by the periodogram. The choice $p = 1$ yields the Hanning data taper.

the variance to decrease (increase). Because $C_h \geq 1$, we can interpret it as a variance inflation factor due to tapering.

It is also useful to derive a corresponding approximation in terms of the lag window. Because of the Parseval relationship in Equation (251d), the approximation of Equation (259a) can be rewritten as

$$\text{var}\,\{\hat{S}_m^{(\text{LW})}(f)\} \approx \frac{C_h S^2(f)}{N} \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}^2. \tag{260}$$

Since under our assumptions $\hat{S}_m^{(\text{LW})}(f)$ is asymptotically unbiased, it is also a consistent estimator of $S(f)$ if its variance decreases to 0 as $N \to \infty$. From Equations (259a) and (260), the large-sample variance of $\hat{S}_m^{(\text{LW})}(f)$ approaches zero if we make one of the following additional – but equivalent – assumptions:

$$\lim_{N\to\infty} \frac{1}{N\,\Delta_{\text{t}}} \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} W_m^2(\phi)\,\text{d}\phi = 0 \;\text{ or }\; \lim_{N\to\infty} \frac{1}{N} \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}^2 = 0.$$

We can use Equation (251c) to reexpress the above in terms of B_W as

$$\lim_{N\to\infty} \frac{B_W}{\Delta_{\text{f}}} = \infty, \;\text{ where }\; \Delta_{\text{f}} \stackrel{\text{def}}{=} \frac{1}{N\,\Delta_{\text{t}}}.$$

In other words, the smoothing window bandwidth B_W must grow to cover many different Fourier frequency bins. Thus one of these three equivalent assumptions is enough to ensure that $\hat{S}_m^{(\text{LW})}(f)$ is a consistent estimator of $S(f)$ (see also the discussion in Section 7.4).

We now consider the covariance between $\hat{S}_m^{(\text{LW})}(f'_k)$ and $\hat{S}_m^{(\text{LW})}(f'_{k'})$. By the asymptotic unbiasedness of $\hat{S}_m^{(\text{LW})}(f)$, it follows that

$$\text{cov}\,\{\hat{S}_m^{(\text{LW})}(f'_k), \hat{S}_m^{(\text{LW})}(f'_{k'})\} \approx E\{\hat{S}_m^{(\text{LW})}(f'_k)\hat{S}_m^{(\text{LW})}(f'_{k'})\} - S(f'_k)S(f'_{k'}).$$

We can proceed as before to argue that

$$\hat{S}_m^{(\text{LW})}(f'_k) \approx \frac{1}{N'\,\Delta_{\text{t}}} \sum_{j=-J}^{J} W_m(f'_j)\hat{S}^{(\text{D})}(f'_{j+k}).$$

Since $\hat{S}_m^{(\text{LW})}(f'_{k'})$ has a similar expression, it follows that

$$E\{\hat{S}_m^{(\text{LW})}(f'_k)\hat{S}_m^{(\text{LW})}(f'_{k'})\} \approx \frac{1}{(N'\,\Delta_{\text{t}})^2} \sum_{j,j'=-J}^{J} W_m(f'_j)W_m(f'_{j'})E\{\hat{S}^{(\text{D})}(f'_{j+k})\hat{S}^{(\text{D})}(f'_{j'+k'})\}.$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

If f'_k and $f'_{k'}$ are far enough apart so that $|k-k'| > 2J$ (recall that J defines the point at which $W_m(f) \approx 0$ so that $2J$ is proportional to the "width" of $W_m(\cdot)$), we can use the asymptotic uncorrelatedness and unbiasedness of $\hat{S}^{(\mathrm{D})}(\cdot)$ on the grid frequencies defined by f'_k to argue that

$$E\{\hat{S}_m^{(\mathrm{LW})}(f'_k)\hat{S}_m^{(\mathrm{LW})}(f'_{k'})\} \approx \frac{1}{(N'\,\Delta_\mathrm{t})^2} \sum_{j,j'=-J}^{J} W_m(f'_j)W_m(f'_{j'})S(f'_{j+k})S(f'_{j'+k'}).$$

By the same assumption on the smoothness of $S(\cdot)$ as before,

$$\begin{aligned} E\{\hat{S}_m^{(\mathrm{LW})}(f'_k)\hat{S}_m^{(\mathrm{LW})}(f'_{k'})\} &\approx \frac{1}{(N'\,\Delta_\mathrm{t})^2} S(f'_k)S(f'_{k'}) \left(\sum_{j=-J}^{J} W_m(f'_j) \right)^2 \\ &\approx S(f'_k)S(f'_{k'}) \left(\int_{-f_\mathcal{N}}^{f_\mathcal{N}} W_m(\phi)\,\mathrm{d}\phi \right)^2 \\ &= S(f'_k)S(f'_{k'}). \end{aligned}$$

Here we have used the Riemann integral to approximate the summation and the assumption of Equation (250a) to show that the integral is identically 1. It now follows that

$$\mathrm{cov}\,\{\hat{S}_m^{(\mathrm{LW})}(f'_k), \hat{S}_m^{(\mathrm{LW})}(f'_{k'})\} \approx 0, \quad f'_k \neq f'_{k'}, \tag{261a}$$

for N large enough. By an extension of the above arguments, one can show that the same result holds for any two fixed frequencies $f' \neq f''$ for N large enough. Note carefully that uncorrelatedness need not be approximately true when f' and f'' are separated by a distance less than the "width" of $W_m(\cdot)$, which we can conveniently assume to be measured by the smoothing window bandwidth B_W.

Comments and Extensions to Section 7.3

[1] We here derive Equation (260) by a route that makes more explicit the form of the factor C_h. Let $f'_k = k/[(2N-1)\,\Delta_\mathrm{t}]$ be a set of frequencies on a grid almost twice as finely spaced as the Fourier frequencies. From Equation (249d) we have the exact result

$$\hat{S}_m^{(\mathrm{LW})}(f'_k) = \frac{1}{(2N-1)\,\Delta_\mathrm{t}} \sum_{j=-(N-1)}^{N-1} W_m(f'_j)\hat{S}^{(\mathrm{D})}(f'_{k-j}). \tag{261b}$$

This is simply the convolution of $\{W_m(f'_j)\}$ with the "frequency" series $\{\hat{S}^{(\mathrm{D})}(f'_j)\}$. If, for large N, we assume that $W_m(f'_j) \approx 0$ for all $f'_j > J'/[2(N-1)\,\Delta_\mathrm{t}] = f'_{J'}$ for some integer J' and that in the interval $[f'_{k-J'}, f'_{k+J'}]$ the true SDF $S(\cdot)$ is locally constant, then the "frequency" series $\{\hat{S}^{(\mathrm{D})}(f'_j)\}$ is locally stationary. The spectrum of the resultant series $\{\hat{S}_m^{(\mathrm{LW})}(f'_j)\}$ – locally to f'_k – is thus the product of the square of the inverse Fourier transform of $\{W_m(f'_j)\}$ – i.e., $\{w_{m,\tau}^2 : \tau = -(N-1), \ldots, N-1\}$ – with the spectrum of $\hat{S}^{(\mathrm{D})}(f'_j)$, which, under the Gaussian assumption, is $\{S^2(f'_k)r_\tau\}$ (see the discussion following Equation (212b)). Hence the spectrum of $\hat{S}_m^{(\mathrm{LW})}(f'_j)$ – locally to f'_k – is given by $S^2(f'_k)w_{m,\tau}^2 r_\tau$, $\tau = -(N-1), \ldots, N-1$. Since a spectrum decomposes a variance, the variance of $\hat{S}_m^{(\mathrm{LW})}(f'_k)$ is given by

$$\mathrm{var}\,\{\hat{S}_m^{(\mathrm{LW})}(f'_k)\} \approx \Delta_\mathrm{t}\, S^2(f'_k) \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}^2 r_\tau = S^2(f'_k) \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}^2 \sum_{t=0}^{N-|\tau|-1} h_{t+|\tau|}^2 h_t^2, \tag{261c}$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

where we have made use of Equation (213a); see Walden (1990a) for details. For the rectangular taper, $h_t = 1/\sqrt{N}$ for $0 \le t \le N-1$, and hence we have

$$\sum_{t=0}^{N-|\tau|-1} h_{t+|\tau|}^2 h_t^2 = \sum_{t=0}^{N-|\tau|-1} \frac{1}{N^2} = \left(1 - \frac{|\tau|}{N}\right),$$

so that

$$\text{var}\,\{\hat{S}_m^{(\text{LW})}(f_k')\} \approx \frac{S^2(f_k')}{N} \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}^2 \left(1 - \frac{|\tau|}{N}\right),$$

a result of some utility (see, for example, Walden and White, 1990).

If we now simplify Equation (261c) by setting $r_\tau = r_0$ for all τ, we obtain

$$\text{var}\,\{\hat{S}_m^{(\text{LW})}(f_k')\} \approx S^2(f_k') \sum_{t=0}^{N-1} h_t^4 \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}^2 \;\text{ since }\; r_0 = \frac{1}{\Delta_t} \sum_{t=0}^{N-1} h_t^4$$

from Equation (214a). As is discussed in Walden (1990a), this approximation is really only good for relatively "modest" data tapers, i.e., those for which the Fourier transform of $\{r_\tau\}$ – namely, $R(\cdot)$ of Equation (212a) – damps down to zero rapidly (see Figure 212). Comparison of the above with Equation (260) yields

$$C_h = N \sum_{t=0}^{N-1} h_t^4 \tag{262}$$

(see also Hannan, 1970, pp. 271–2, and Brillinger, 1981a, pp. 150–1). The Cauchy inequality of Equation (257a) with $a_t = h_t^2$ and $b_t = 1$ tells us that $C_h \ge 1$ (since we always assume the normalization $\sum_{t=0}^{N-1} h_t^2 = 1$). We also note that $C_h = 1$ only for the rectangular data taper $h_t = 1/\sqrt{N}$. For large N, we can compute C_h approximately for various data tapers by approximating the above summation with an integral. The results for several common data tapers are shown in Table 260.

[2] We give here more details concerning – and an alternative expression for – the approximation to the variance of a lag window spectral estimator given in Equation (261c). First, we note a desirable property that the approximation given by Equation (261c) has, but the less accurate approximation of Equation (260) does not. Now we can regard a direct spectral estimator as a special case of a lag window estimator if we take the lag window to be $w_{m,\tau} = 1$ for all τ (to see this, just compare Equations (188a) and (248a)). With this choice for $w_{m,\tau}$, Amrein and Künsch (2011) point out that the less accurate approximation to $\text{var}\,\{\hat{S}_m^{(\text{LW})}(f)\}$ does not in general match up with the usual approximation to $\text{var}\,\{\hat{S}^{(\text{D})}(f)\}$ given by Equation (204c): making use of Equation (262), the former (Equation (260)) gives

$$\text{var}\,\{\hat{S}_m^{(\text{LW})}(f)\} \approx \frac{C_h S^2(f)}{N} \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}^2 = S^2(f)(2N-1) \sum_{t=0}^{N-1} h_t^4,$$

whereas the latter states that $\text{var}\,\{\hat{S}^{(\text{D})}(f)\} \approx S^2(f)$. In particular, if we consider the rectangular taper $h_t = 1/\sqrt{N}$, we obtain $\text{var}\,\{\hat{S}_m^{(\text{LW})}(f)\} \approx 2S^2(f)$ in comparison to the usual $\text{var}\,\{\hat{S}^{(\text{D})}(f)\} \approx S^2(f)$. By contrast, the more accurate approximation does not exhibit this mismatch because, when $w_{m,\tau} = 1$ (and letting $f = f_k'$), Equation (261c) becomes

$$\text{var}\,\{\hat{S}_m^{(\text{LW})}(f)\} \approx S^2(f) \sum_{\tau=-(N-1)}^{N-1} \sum_{t=0}^{N-|\tau|-1} h_{t+|\tau|}^2 h_t^2 = S^2(f)$$

in light of the following result.

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

▷ **Exercise [263]** Show that, if $\{h_t\}$ is a data taper with the usual normalization $\sum_t h_t^2 = 1$, then

$$\sum_{\tau=-(N-1)}^{N-1} \sum_{t=0}^{N-|\tau|-1} h_{t+|\tau|}^2 h_t^2 = 1.$$

Hint: consider Equations (212a), (213a) and (213b). ◁

Amrein and Künsch (2011) note that an implication of this mismatch is to call into question the use of the approximation given by Equation (260) when the degree of smoothing induced by a lag window is quite small.

Second, we give an alternative expression for the more accurate approximation of Equation (261c). This approximation is based upon Equation (261b), from which it directly follows from Exercise [2.1e] that

$$\operatorname{var}\{\hat{S}_m^{(\mathrm{LW})}(f_k')\} = \frac{1}{(2N-1)^2\Delta_{\mathrm{t}}^2} \sum_{j,j'=-(N-1)}^{N-1} W_m(f_j')W_m(f_{j'}') \operatorname{cov}\{\hat{S}^{(\mathrm{D})}(f_{k-j}'), \hat{S}^{(\mathrm{D})}(f_{k-j'}')\}.$$

From Equation (250c) we can assume that $W_m(f_j') \approx 0$ for all $|j| > J$. In practice we can determine J by setting $f_J' \approx cB_W$, where B_W is the smoothing window bandwidth of Equation (251e), and c is, say, 1 or 2 (J is the integer closest to $(2N-1)B_W\,\Delta_{\mathrm{t}}$ when $c = 1$). We thus have

$$\operatorname{var}\{\hat{S}_m^{(\mathrm{LW})}(f_k')\} \approx \frac{1}{(2N-1)^2\Delta_{\mathrm{t}}^2} \sum_{j,j'=-J}^{J} W_m(f_j')W_m(f_{j'}') \operatorname{cov}\{\hat{S}^{(\mathrm{D})}(f_{k-j}'), \hat{S}^{(\mathrm{D})}(f_{k-j'}')\}.$$

Under the assumptions that we are dealing with a Gaussian stationary process, that $S(\cdot)$ is locally flat from f_{k-J}' to f_{k+J}' and that $0 < f_{k-J}'$ and $f_{k+J}' < f_{\mathcal{N}}$ (with f_{k-J}' not being too near to 0 and f_{k+J}' not being too near to $f_{\mathcal{N}}$), we can use the result of Exercise [212] to simplify the above to

$$\operatorname{var}\{\hat{S}_m^{(\mathrm{LW})}(f_k')\} \approx \frac{S^2(f_k')}{(2N-1)^2\Delta_{\mathrm{t}}^2} \sum_{j=-J}^{J} W_m(f_j') \sum_{j'=-J}^{J} W_m(f_{j'}')R(f_{j-j'}'),$$

where $R(\cdot)$ is defined in Equation (212a) and depends only on the data taper $\{h_t\}$. By a change of variables similar to that used in the solution to Exercise [170], we can rewrite the above as

$$\operatorname{var}\{\hat{S}_m^{(\mathrm{LW})}(f_k')\} \approx \frac{S^2(f_k')}{(2N-1)^2\Delta_{\mathrm{t}}^2} \sum_{l=-2J}^{2J} R(f_l') \sum_{l'=0}^{2J-|l|} W_m(f_{J-l'-|l|}')W_m(f_{J-l'}').$$

The inner summation is essentially an autocorrelation, so it can be efficiently calculated for all lags l using discrete Fourier transforms (see Walden, 1990a, for details); however, as is demonstrated in Figure 212, $R(\cdot)$ typically decreases effectively to 0 rather quickly so that we can shorten the outer summation to, say, its $2L + 1$ innermost terms:

$$\operatorname{var}\{\hat{S}_m^{(\mathrm{LW})}(f_k')\} \approx \frac{S^2(f_k')}{(2N-1)^2\Delta_{\mathrm{t}}^2} \sum_{l=-L}^{L} R(f_l') \sum_{l'=0}^{2J-|l|} W_m(f_{J-l'-|l|}')W_m(f_{J-l'}').$$

For example, if we use a Slepian data taper with $NW \leq 8$, $N = 64$ and $\Delta_{\mathrm{t}} = 1$, we can conveniently let $L = \min(10, 2J)$, where $l = 10$ corresponds to $\eta = l/[(2N-1)\,\Delta_{\mathrm{t}}] = 10/127 \approx 5/64$ in Figure 212.

Finally, we note that, with a trivial adjustment to the limits of the summations, the above expression is a special case of a more general formula given by Amrein and Künsch (2011), which is valid for arbitrary discretely smoothed direct spectral estimators (see their Equation (5)). Their paper gives a rigorous theoretical justification for the more accurate approximation, along with examples demonstrating concretely the improvement it affords.

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

7.4 Asymptotic Distribution of Lag Window Estimators

We consider here the asymptotic (large-sample) distribution of $\hat{S}_m^{(\mathrm{LW})}(f)$. We can write, using Equations (248a) and (249d),

$$\begin{aligned}\hat{S}_m^{(\mathrm{LW})}(f) &= \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} W_m(f-\phi)\hat{S}^{(\mathrm{D})}(\phi)\,\mathrm{d}\phi \\ &\approx \frac{1}{(2N-1)\,\Delta_{\mathrm{t}}} \sum_{j=-(N-1)}^{N-1} W_m(f'_j)\hat{S}^{(\mathrm{D})}(f-f'_{-j}),\end{aligned}$$

with the approximation becoming an equality when $f = f'_k = k/[(2N-1)\,\Delta_{\mathrm{t}}]$ for some integer k. As noted in Section 6.6, under mild conditions and for large N, the $\hat{S}^{(\mathrm{D})}(f-\tilde{f}_{-j})$ terms can be regarded as a set of χ^2 RVs scaled by appropriate multiplicative constants. It follows that $\hat{S}_m^{(\mathrm{LW})}(f)$ is asymptotically a linear combination of χ^2 RVs with weights that depend on the smoothing window $W_m(\cdot)$. The exact form of such a distribution is hard to determine, but there is a well-known approximation we can use (Welch, 1936; Fairfield Smith, 1936; Welch, 1938; Satterthwaite, 1941; Satterthwaite, 1946; Box, 1954; Blackman and Tukey, 1958; and Jenkins, 1961). Assume

$$\hat{S}_m^{(\mathrm{LW})}(f) \stackrel{\mathrm{d}}{=} a\chi^2_\nu;$$

i.e., $\hat{S}_m^{(\mathrm{LW})}(f)$ has the distribution of a chi-square RV scaled by a constant a, where the degrees of freedom ν and a are both unknown. By the properties of the χ^2_ν distribution, we have

$$E\{\hat{S}_m^{(\mathrm{LW})}(f)\} = E\{a\chi^2_\nu\} = a\nu \;\text{ and }\; \mathrm{var}\,\{\hat{S}_m^{(\mathrm{LW})}(f)\} = \mathrm{var}\,\{a\chi^2_\nu\} = 2a^2\nu.$$

We can use these two expressions to derive equations for ν and a in terms of the expected value and variance of $\hat{S}_m^{(\mathrm{LW})}(f)$:

$$\nu = \frac{2\left(E\{\hat{S}_m^{(\mathrm{LW})}(f)\}\right)^2}{\mathrm{var}\,\{\hat{S}_m^{(\mathrm{LW})}(f)\}} \;\text{ and }\; a = \frac{E\{\hat{S}_m^{(\mathrm{LW})}(f)\}}{\nu}. \tag{264a}$$

Under the assumptions of the previous two sections, the large-sample expected value and variance for $\hat{S}_m^{(\mathrm{LW})}(f)$ are, respectively, $S(f)$ and

$$\frac{C_h S^2(f)}{N\,\Delta_{\mathrm{t}}} \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} W_m^2(\phi)\,\mathrm{d}\phi = \frac{C_h S^2(f)}{N} \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}^2$$

(see Equations (256c), (259a) and (260)). If we substitute these values into Equation (264a), we find that, for large samples,

$$\nu = \frac{2N\,\Delta_{\mathrm{t}}}{C_h \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} W_m^2(\phi)\,\mathrm{d}\phi} = \frac{2N}{C_h \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}^2} \;\text{ and }\; a = \frac{S(f)}{\nu}. \tag{264b}$$

Hence we assume

$$\hat{S}_m^{(\mathrm{LW})}(f) \stackrel{\mathrm{d}}{=} S(f)\chi^2_\nu/\nu. \tag{264c}$$

The quantity ν is often called the *equivalent degrees of freedom* (EDOFs) for the estimator $\hat{S}_m^{(\mathrm{LW})}(f)$. From Equation (251e) for the smoothing window bandwidth B_W, we obtain

$$\nu = \frac{2NB_W\,\Delta_{\mathrm{t}}}{C_h}, \tag{264d}$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

so that an increase in B_W (i.e., a greater degree of smoothing) yields an increase in ν. Using $\text{var}\{a\chi^2_\nu\} = 2a^2\nu$, we have (for large samples)

$$\text{var}\{\hat{S}^{(\text{LW})}_m(f)\} \approx \frac{2S^2(f)}{\nu}.$$

As ν increases, this variance decreases, but, for a fixed sample size N, a decrease in the variance of $\hat{S}^{(\text{LW})}_m(f)$ comes at the potential expense of an increase in its bias. Here we must be careful. The above expression for $\text{var}\{\hat{S}^{(\text{LW})}_m(f)\}$ is based on the assumption that $\hat{S}^{(\text{LW})}_m(f)$ is approximately distributed as $a\chi^2_\nu$ with a and ν given above. If, by increasing the bandwidth B_W with the idea of making ν large and hence the variance small, we inadvertently introduce significant bias in $\hat{S}^{(\text{LW})}_m(f)$, then $\hat{S}^{(\text{LW})}_m(f)$ will deviate substantially from its assumed distribution, and the above expression for $\text{var}\{\hat{S}^{(\text{LW})}_m(f)\}$ can be misleading.

The $a\chi^2_\nu$ approximation to the distribution of $\hat{S}^{(\text{LW})}_m(f)$ allows us to construct an approximate confidence interval (CI) for $S(f)$ at a fixed f. Let $Q_\nu(p)$ represent the $p \times 100\%$ percentage point of the χ^2_ν distribution; i.e., $\mathbf{P}\left[\chi^2_\nu \le Q_\nu(p)\right] = p$. We thus have, for $0 \le p \le 1/2$,

$$\mathbf{P}\left[Q_\nu(p) \le \chi^2_\nu \le Q_\nu(1-p)\right] = 1 - 2p.$$

Since the distribution of $\hat{S}^{(\text{LW})}_m(f)$ is approximately that of the RV $a\chi^2_\nu$ with $a = S(f)/\nu$, it follows that $\nu\hat{S}^{(\text{LW})}_m(f)/S(f)$ is approximately distributed as χ^2_ν. Hence

$$\mathbf{P}\left[Q_\nu(p) \le \frac{\nu\hat{S}^{(\text{LW})}_m(f)}{S(f)} \le Q_\nu(1-p)\right] = \mathbf{P}\left[\frac{\nu\hat{S}^{(\text{LW})}_m(f)}{Q_\nu(1-p)} \le S(f) \le \frac{\nu\hat{S}^{(\text{LW})}_m(f)}{Q_\nu(p)}\right] = 1-2p,$$

from which it follows that

$$\left[\frac{\nu\hat{S}^{(\text{LW})}_m(f)}{Q_\nu(1-p)}, \frac{\nu\hat{S}^{(\text{LW})}_m(f)}{Q_\nu(p)}\right] \tag{265}$$

is a $100(1-2p)\%$ CI for $S(f)$. Note that this CI applies only to $S(f)$ at one particular value of f. We cannot get a confidence band for the entire function $S(\cdot)$ by this method.

Since $\hat{S}^{(\text{LW})}_m(f)$ is approximately distributed as an $a\chi^2_\nu$ RV, we can employ – when ν is large (say greater than 30) – the usual scheme of approximating a χ^2_ν distribution by a Gaussian distribution with the same mean and variance; i.e., the RV

$$\frac{\hat{S}^{(\text{LW})}_m(f) - S(f)}{S(f)(2/\nu)^{1/2}} \stackrel{\text{d}}{=} \frac{a\chi^2_\nu - a\nu}{(2a^2\nu)^{1/2}}$$

is approximately Gaussian distributed with zero mean and unit variance. Under these circumstances, an approximate $100(1-2p)\%$ CI for $S(f)$ has the form

$$\left[\frac{\hat{S}^{(\text{LW})}_m(f)}{1 + \Phi^{-1}(1-p)(2/\nu)^{1/2}}, \frac{\hat{S}^{(\text{LW})}_m(f)}{1 + \Phi^{-1}(p)(2/\nu)^{1/2}}\right],$$

where $\Phi^{-1}(p)$ is the $p \times 100\%$ percentage point of the standard Gaussian distribution.

The CI for $S(f)$ given in Equation (265) has a width of

$$\hat{S}^{(\text{LW})}_m(f)\left[\frac{\nu}{Q_\nu(p)} - \frac{\nu}{Q_\nu(1-p)}\right],$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

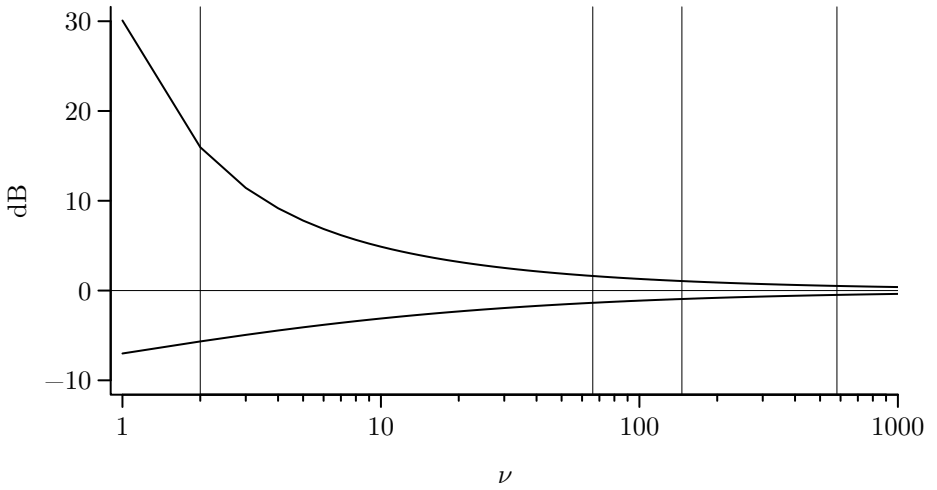

Figure 266 Upper and lower additive factors needed to create a 95% confidence interval (CI) for $10 \log_{10}(S(f))$ based upon a spectral estimator with ν equivalent degrees of freedom (EDOFs). Here ν ranges from 1 up to 1000 in increments of 1. The thin vertical lines indicate $\nu = 2$, 66, 146 and 581. The latter three values are the smallest integer-valued EDOFs needed to achieve a 95% CI for $10 \log_{10}(S(f))$ whose width is less than, respectively, 3 dB, 2 dB and 1 dB. The case $\nu = 2$ is associated with direct SDF estimators $\hat{S}^{(\mathrm{D})}(\cdot)$ at frequencies $0 < f < f_{\mathcal{N}}$ (see Equations (204b) and (205b)), while $\nu = 1$ is associated with $\hat{S}^{(\mathrm{D})}(0)$, $\hat{S}^{(\mathrm{D})}(f_{\mathcal{N}})$ and DCT-based periodograms (see Equations (204b) and (219a)).

which unfortunately depends on $\hat{S}_m^{(\mathrm{LW})}(f)$. This fact makes plots of $\hat{S}_m^{(\mathrm{LW})}(f)$ versus f difficult to interpret: the width of a CI for $S(f)$ is proportional to $\hat{S}_m^{(\mathrm{LW})}(f)$ and thus varies from frequency to frequency. If we assume that $S(f) > 0$ and $\hat{S}_m^{(\mathrm{LW})}(f) > 0$, we can write

$$\mathbf{P}\left[\frac{\nu \hat{S}_m^{(\mathrm{LW})}(f)}{Q_\nu(1-p)} \le S(f) \le \frac{\nu \hat{S}_m^{(\mathrm{LW})}(f)}{Q_\nu(p)}\right]$$
$$= \mathbf{P}\left[10 \log_{10}\left(\frac{\nu \hat{S}_m^{(\mathrm{LW})}(f)}{Q_\nu(1-p)}\right) \le 10 \log_{10}(S(f)) \le 10 \log_{10}\left(\frac{\nu \hat{S}_m^{(\mathrm{LW})}(f)}{Q_\nu(p)}\right)\right].$$

It follows that

$$\left[10 \log_{10}(\hat{S}_m^{(\mathrm{LW})}(f)) + 10 \log_{10}\left(\frac{\nu}{Q_\nu(1-p)}\right), 10 \log_{10}(\hat{S}_m^{(\mathrm{LW})}(f)) + 10 \log_{10}\left(\frac{\nu}{Q_\nu(p)}\right)\right] \tag{266a}$$

is a $100(1-2p)\%$ CI for $10 \log_{10}(S(f))$. Note that the width of this CI, namely

$$10 \log_{10}\left(\frac{\nu}{Q_\nu(p)}\right) - 10 \log_{10}\left(\frac{\nu}{Q_\nu(1-p)}\right) = 10 \log_{10}\left(\frac{Q_\nu(1-p)}{Q_\nu(p)}\right), \tag{266b}$$

is now independent of $10 \log_{10}(\hat{S}_m^{(\mathrm{LW})}(f))$. This is a rationale for plotting SDF estimates on a decibel scale. Figure 266 shows $10 \log_{10}(\nu/Q_\nu(p))$ and $10 \log_{10}(\nu/Q_\nu(1-p))$ versus ν (upper and lower thick curves, respectively) for $p = 0.025$ and integer-valued ν from 1 to 1000. (Note that, when $\nu = 2$, the CI width given by Equation (266b) reduces to what is given in Equation (205a), which is appropriate for a CI based upon a direct spectral estimator.)

As a concrete example, let us use the Parzen lag window estimate $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ shown in Figure 249 to create 95% CIs for both $S(f)$ and $10 \log_{10}(S(f))$ (in contrast to what happens

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

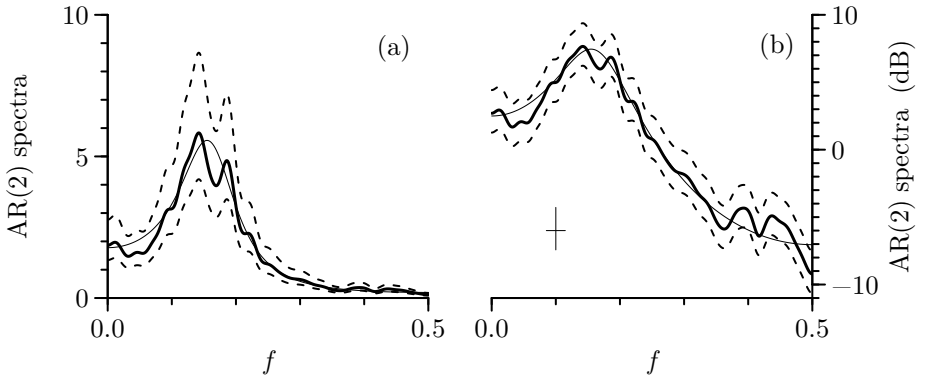

Figure 267 95% confidence intervals (CIs) for an SDF based upon a lag window estimate. The thick solid curves in both plots are the Parzen lag window estimate $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ shown in Figure 249 (this estimate makes use of the AR(2) time series of Figure 34(a)). The dashed curves show the upper and lower limits of 95% CIs for the true AR(2) SDF based upon $\hat{S}_m^{(\mathrm{LW})}(\cdot)$. The true AR(2) SDF is shown as the thin solid curve. The two plots are the same except for the vertical axes: in (a), this axis is on a linear scale, while it is on a decibel scale in (b). The true SDF is trapped by the CIs at most frequencies, but it does fall barely below them before and after $f = 0.4$ and above them around the Nyquist frequency $f_{\mathcal{N}} = 0.5$ (while readily apparent in plot (b), these patterns are almost impossible to see in (a)). The height and width of the crisscross in the lower left-hand portion of plot (b) are equal to, respectively, the width of each 95% CI and the bandwidth measure $B_{\mathcal{U}}$ of Equation (256a).

in practical applications, here we know $S(f)$ because $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ is based upon the AR(2) time series of Figure 34(a), with an SDF dictated by Equations (34), (33) and (145a)). We first need to determine the EDOFs ν associated with the lag window estimate, which, according to Equation (264b), depends upon the sample size N, the variance inflation factor C_h and the lag window $\{w_{m,\tau}\}$. Here $N = 1024$, $C_h = 1$ because $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ is based upon the periodogram (see Table 260), and $w_{m,\tau}$ is given by the forthcoming Equation (275a) with m set to 64, yielding

$$\sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}^2 \doteq 34.51.$$

Hence $\nu \doteq 59.34$. From Equation (265) we obtain

$$\left[\frac{\nu \hat{S}_m^{(\mathrm{LW})}(f)}{Q_\nu(0.975)}, \frac{\nu \hat{S}_m^{(\mathrm{LW})}(f)}{Q_\nu(0.025)}\right] \doteq \left[0.72\, \hat{S}_m^{(\mathrm{LW})}(f), 1.49\, \hat{S}_m^{(\mathrm{LW})}(f)\right]$$

as a 95% CI for $S(f)$, where $Q_\nu(0.025) \doteq 39.94$ and $Q_\nu(0.975) \doteq 82.52$ are the lower and upper 2.5% percentage points of a chi-square distribution with $\nu \doteq 59.34$ degrees of freedom. Figure 267(a) shows $\hat{S}_m^{(\mathrm{LW})}(f)$ versus f (thick solid curve) along with the corresponding CIs (upper and lower dashed curves) and the true SDF (thin solid curve). Note that the width of the CI at f is proportional to $\hat{S}_m^{(\mathrm{LW})}(f)$. Simple adjustments to Equation (266a) give the lower and upper parts of the CI for $10 \log_{10}(S(f))$:

$$10 \log_{10}\left(\frac{\nu}{Q_\nu(0.975)}\right) + 10 \log_{10}(\hat{S}_m^{(\mathrm{LW})}(f)) \doteq -1.43 + 10 \log_{10}(\hat{S}_m^{(\mathrm{LW})}(f)),$$
$$10 \log_{10}\left(\frac{\nu}{Q_\nu(0.025)}\right) + 10 \log_{10}(\hat{S}_m^{(\mathrm{LW})}(f)) \doteq 1.72 + 10 \log_{10}(\hat{S}_m^{(\mathrm{LW})}(f)).$$

The width of each CI (3.15 dB) no longer depends upon the particular value of $\hat{S}_m^{(\mathrm{LW})}(f)$. Figure 267(b) is similar to Figure 267(a), but now shows $10 \log_{10}(\hat{S}_m^{(\mathrm{LW})}(f))$, its related CI

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

and the true $10 \log_{10}(S(f))$ versus f. The CI based on $10 \log_{10}(\hat{S}_m^{(\mathrm{LW})}(f))$ generally traps the true $10 \log_{10}(S(f))$, but fails to do so at certain high frequencies (just before and after $f = 0.4$, and also around $f = 0.5$). Because a log transform preserves order, this statement is also true for $\hat{S}_m^{(\mathrm{LW})}(f)$ and $S(f)$, but this is not easy to see in Figure 267(a).

Since the upper (lower) limit of the CI for the unknown $S(f)$ is the same distance above (below) $\hat{S}_m^{(\mathrm{LW})}(f)$ for any given f when plotted on a decibel scale, it is useful to depict these distances as the vertical part of a crisscross. An example is shown in the lower left-hand part of Figure 267(b), for which the lengths of the crisscross above and below the horizontal line are, respectively, 1.72 and 1.43 dB. The horizontal part of the crisscross indicates the bandwidth measure $B_{\mathcal{U}}$ of Equation (256a) (here $B_{\mathcal{U}} \doteq 0.0296$). If $|f - f'| < B_{\mathcal{U}}$, estimates $\hat{S}_m^{(\mathrm{LW})}(f)$ and $\hat{S}_m^{(\mathrm{LW})}(f')$ tend to be positively correlated, with the strength of correlation increasing as $|f - f'|$ decreases; on the other hand, if $|f - f'| \geq B_{\mathcal{U}}$, the two estimates should be approximately uncorrelated. We cannot expect $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ to be able to resolve features in $S(\cdot)$ whose width is smaller than $B_{\mathcal{U}}$; on the other hand, we can expect to see bumps and valleys in $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ whose widths are characterized by $B_{\mathcal{U}}$, as the example in Figure 267(b) illustrates. We shall show similar crisscrosses on most forthcoming linear/decibel plots of $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ and other nonparametric SDF estimates.

7.5 Examples of Lag Windows

We consider here six examples of lag and smoothing windows that can be used in practice to form a specific lag window spectral estimator $\hat{S}_m^{(\mathrm{LW})}(\cdot)$. Five of our examples (the Bartlett, Daniell, Bartlett–Priestley, Parzen and Papoulis lag windows) have been used extensively in practical applications. As we document by example in Section 7.12, the remaining example (the Gaussian lag window) has some attractive properties that can recommend its use over the five standard windows even though it has not seen much use in practice. Two other lag windows – mainly of historical interest – are the subjects of Exercises [7.7] and [7.8], and we discuss some other nonstandard windows in C&E [5].

For each of the specific lag windows, we show a figure with four plots (Figures 270, 272b, 274b, 276, 278 and 279). Plot (a) in each figure shows an example of the lag window $w_{m,\tau}$ versus lag τ for a sample size $N = 64$ and a particular setting of the window parameter m. Plot (b) shows the corresponding smoothing window $W_m(\cdot)$ on a decibel scale versus frequency (the sampling interval Δ_{t} is taken to be 1, so that the Nyquist frequency $f_{\mathcal{N}}$ is 1/2). Because $w_{m,-\tau} = w_{m,\tau}$ and $W_m(-f) = W_m(f)$ by assumption, we need only plot $w_{m,\tau}$ for $\tau \geq 0$ and $W_m(f)$ for $f \geq 0$. We indicate the smoothing window bandwidths β_W (Equation (251b)) and B_W (Equation (251e)) by horizontal lines plotted, respectively, 3 and 6 dB down from $W_m(0)$ (the widths of the displayed lines are equal to $\beta_W/2$ and $B_W/2$). Plots (c) and (d) show two different spectral windows $\mathcal{U}_m(\cdot)$ for the same m and N. From Equation (255c) we see that $\mathcal{U}_m(\cdot)$ depends upon both the smoothing window $W_m(\cdot)$ and the data taper $\{h_t\}$ through its associated spectral window $\mathcal{H}(\cdot)$. Plots (c) and (d) show $\mathcal{U}_m(\cdot)$ corresponding to, respectively, the rectangular data taper (see Figures 190(a) and 191(a)) and a Slepian data taper with $NW = 4$ (Figures 190(g) and 191(g)). We have included plots (c) and (d) to emphasize the point that the expected value of a lag window spectral estimator $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ depends upon both the smoothing window $W_m(\cdot)$ and the data taper used in the corresponding direct spectral estimator $\hat{S}^{(\mathrm{D})}(\cdot)$. Plots (c) and (d) also depict our standard measure of the effective bandwidth of $\hat{S}_m^{(\mathrm{LW})}(\cdot)$, namely, $B_{\mathcal{U}}$ (Equation (256a)). This measure is shown as a horizontal line plotted 3 dB down from $\mathcal{U}_m(0)$ – the width of the line is $B_{\mathcal{U}}/2$.

In all six examples, the smoothing window $W_m(\cdot)$ consists of a central lobe (i.e., the one centered about zero frequency) and several sidelobes. As we note in the next section, these sidelobes ideally should be as small as possible (to minimize what is defined there as smoothing window leakage). The magnitude of the peak in the first sidelobe relative to the

magnitude of the peak in the central lobe is thus of interest, as is the rate of decay of the envelope formed by the peaks of the sidelobes. This envelope typically decays approximately as f^{α} for some $\alpha < 0$. For example, if $\alpha = -1$, a doubling of frequency corresponds to a doubling in the decay of the magnitude of the peaks in the sidelobes of $W_m(\cdot)$. On a decibel (i.e., $10 \log_{10}$) scale such a doubling corresponds to a $10 \log_{10}(2) \approx 3$ dB drop in a frequency octave (doubling of frequency). The magnitude of the envelope thus decays at 3 dB per octave. We quote this decay rate for each of the lag windows because it is a good guide to how fast the sidelobes decay with increasing frequency (see, however, the discussions about the Daniell, Bartlett–Priestley and Gaussian windows shown in Figures 272b(b), 274b(b) and 278(b)).

Table 279 at the end of this section lists approximations for the asymptotic variance of $\hat{S}_m^{(\mathrm{LW})}(f)$, the EDOFs ν, the smoothing window bandwidth B_W and the alternative bandwidth measure β_W for the six lag windows. This table shows a number of interesting relationships. For example, for fixed m, we have $\mathrm{var}\,\{\hat{S}_m^{(\mathrm{LW})}(f)\} \to 0$ as $N \to \infty$ for lag window estimators based on these six windows. Also we can deduce that $\mathrm{var}\,\{\hat{S}_m^{(\mathrm{LW})}(f)\}$ is inversely proportional either to β_W (for the Daniell, Bartlett–Priestly, Parzen, Gaussian and Papoulis windows) or to β_W^2 (for the Bartlett window). Since the bias due to the smoothing window alone is proportional to β_W^2 (see Equation (256d)), there is a trade-off between the bias and variance in lag window spectral estimators, or, to state it more picturesquely, "bias and variance are antagonistic" (Priestley, 1981, p. 528).

Bartlett Window (Figure 270)
The lag window for Bartlett's spectral estimator (Bartlett, 1950) is given by

$$w_{m,\tau} = \begin{cases} 1 - |\tau|/m, & |\tau| < m; \\ 0, & |\tau| \geq m, \end{cases} \tag{269a}$$

where m is an integer-valued window parameter that can assume values from 1 up to N. Bartlett's window applies linearly decreasing weights to the sample ACVS up to lag m and zero weights thereafter. The parameter m can be interpreted as a truncation point beyond which the ACVS is considered to be zero. The corresponding smoothing window is

$$W_m(f) = \Delta_{\mathrm{t}} \sum_{\tau=-m}^{m} \left(1 - \frac{|\tau|}{m}\right) \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_{\mathrm{t}}} = \mathcal{F}(f), \tag{269b}$$

where the latter quantity is Fejér's kernel (see Equation (174c)). Note that the smoothing window is of the same form as the two-sided Cesàro partial sum we met in Equation (79). From the plots of $\mathcal{F}(\cdot)$ in Figure 176, we see that the width of the central lobe decreases as m increases, so the amount of smoothing decreases as m increases (a convention for m we maintain for the other lag windows we discuss). Since Fejér's kernel is nonnegative everywhere, it follows that Bartlett's estimator of an SDF is also nonnegative everywhere. The magnitude of the peak of the first sidelobe of Fejér's kernel is down about 13 dB from its value at $f = 0$. The envelope of Bartlett's smoothing window decays as approximately f^{-2}, which corresponds to a decay of 6 dB per frequency octave.

Note from plots (c) and (d) in Figure 270 that the two spectral windows $\mathcal{U}_m(\cdot)$ corresponding to the default data taper and the Slepian data taper both have a decay rate similar to that of the smoothing window shown in plot (b). This is due to the fact that $\mathcal{U}_m(\cdot)$ is the convolution of the smoothing window $W_m(\cdot)$ and the spectral window $\mathcal{H}(\cdot)$ associated with the data taper. In the case of the default data taper (plot (c)), both $W_m(\cdot)$ and $\mathcal{H}(\cdot)$ are in fact Fejér's kernel, so the decay rate of their convolution must be the same. In the case of the Slepian data taper, a glance at its spectral window $\mathcal{H}(\cdot)$ in Figure 191(g) shows that $\mathcal{U}_m(\cdot)$ is

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

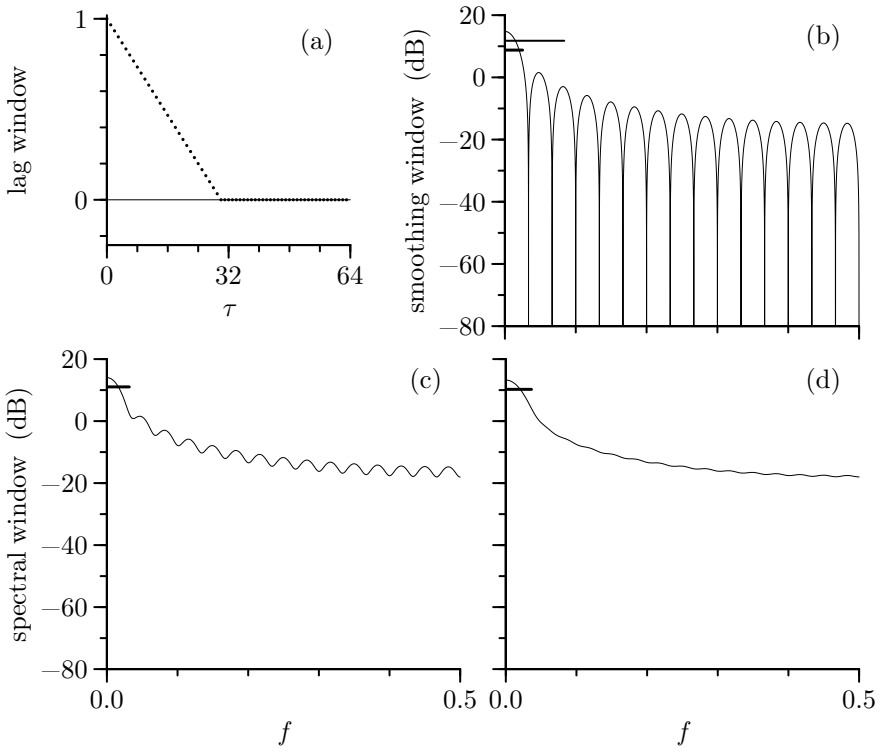

Figure 270 Bartlett lag, smoothing and spectral windows for $m = 30$ and $N = 64$ (the smoothing window bandwidth is $B_W \doteq 0.05$). For a description of the contents of each of the four plots, see the second paragraph from the beginning of this section (Section 7.5).

essentially a smoothed version of Fejér's kernel and hence has the same decay rate as Fejér's kernel. Now let us make a practical interpretation. The user who selected a Slepian data taper with $NW = 4$ in order to achieve low bias in a direct spectral estimate $\hat{S}^{(\mathrm{D})}(\cdot)$ of an SDF with a suspected large dynamic range might be shocked to find that smoothing $\hat{S}^{(\mathrm{D})}(\cdot)$ with Bartlett's window yields an *overall* spectral window $\mathcal{U}_m(\cdot)$ that decays at the same rate as the spectral window for the periodogram!

The rationale Bartlett (1950) used to come up with the lag window of Equation (269a) has an interesting connection to Welch's overlapped segment averaging spectral estimator (see Section 8.8). Bartlett did not apply any tapering (other than the default rectangular taper) so that the direct spectral estimator of Equation (188a) to be smoothed was just the periodogram. Bartlett argued that one could reduce the sampling fluctuations of the periodogram by

[1] splitting the original sample of N observations into N/m contiguous blocks, each block containing m observations (we assume that N/m is an integer);
[2] forming the periodogram for each block; and
[3] averaging the N/m periodograms together.

Let $\hat{S}^{(\mathrm{P})}_{k,m}(\cdot)$ be the periodogram for the kth block. Bartlett reasoned that $\hat{S}^{(\mathrm{P})}_{j,m}(f)$ and $\hat{S}^{(\mathrm{P})}_{k,m}(f)$ for $j \neq k$ should be approximately uncorrelated (with the approximation improving as m gets larger – Exercise [7.9] investigates this claim). The reduction in variance should thus be inversely proportional to the number of blocks N/m. Bartlett worked with the unbiased

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

ACVS estimator

$$\hat{s}_\tau^{(\mathrm{U})} = \frac{N}{N-|\tau|}\hat{s}_\tau^{(\mathrm{P})},$$

in terms of which we can write

$$\hat{S}_{k,m}^{(\mathrm{P})}(f) = \Delta_\mathrm{t} \sum_{\tau=-m}^{m} \left(1 - \frac{|\tau|}{m}\right) \hat{s}_{k,\tau}^{(\mathrm{U})} \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_\mathrm{t}},$$

where $\{\hat{s}_{k,\tau}^{(\mathrm{U})}\}$ denotes the unbiased ACVS estimator based upon just the data in the kth block. If we average the N/m periodogram estimates together, we get

$$\begin{aligned}\frac{1}{N/m}\sum_{k=1}^{N/m} \hat{S}_{k,m}^{(\mathrm{P})}(f) &= \Delta_\mathrm{t} \sum_{\tau=-m}^{m} \left(1 - \frac{|\tau|}{m}\right) \left[\frac{1}{N/m}\sum_{k=1}^{N/m} \hat{s}_{k,\tau}^{(\mathrm{U})}\right] \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_\mathrm{t}} \\ &= \Delta_\mathrm{t} \sum_{\tau=-m}^{m} \left(1 - \frac{|\tau|}{m}\right) \bar{s}_\tau^{(\mathrm{U})} \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_\mathrm{t}},\end{aligned}$$

where $\bar{s}_\tau^{(\mathrm{U})}$ is the average of the $\hat{s}_{k,\tau}^{(\mathrm{U})}$ terms. Bartlett then argued that $\bar{s}_\tau^{(\mathrm{U})}$ ignores information concerning s_τ that is contained in pairs of data values in adjacent blocks. He suggested replacing $\bar{s}_\tau^{(\mathrm{U})}$ with $\hat{s}_\tau^{(\mathrm{U})}$, the unbiased estimator of s_τ obtained by using all N observations. This substitution yields the estimator of $S(f)$ given by

$$\Delta_\mathrm{t} \sum_{\tau=-m}^{m} \left(1 - \frac{|\tau|}{m}\right) \hat{s}_\tau^{(\mathrm{U})} \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_\mathrm{t}} = \Delta_\mathrm{t} \sum_{\tau=-m}^{m} \frac{1-|\tau|/m}{1-|\tau|/N} \hat{s}_\tau^{(\mathrm{P})} \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_\mathrm{t}}.$$

The common form for Bartlett's lag window spectral estimator (in the case of a default rectangular taper), namely,

$$\hat{S}_m^{(\mathrm{LW})}(f) = \Delta_\mathrm{t} \sum_{\tau=-m}^{m} \left(1 - \frac{|\tau|}{m}\right) \hat{s}_\tau^{(\mathrm{P})} \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_\mathrm{t}},$$

can be regarded as an approximation to the above under the assumption that N is much larger than m since then

$$\frac{1-|\tau|/m}{1-|\tau|/N} = \left(1 - \frac{|\tau|}{m}\right)\left(1 + \frac{|\tau|}{N} + \left[\frac{|\tau|}{N}\right]^2 + \cdots\right) \approx 1 - \frac{|\tau|}{m}.$$

Bartlett noted that the choice of m is a compromise between reducing the variance of $\hat{S}_m^{(\mathrm{LW})}(f)$ and reducing its resolution (i.e., smearing out local features).

Daniell (or Rectangular) Window (Figures 272a and 272b)
As we noted after Equation (71), the quantity

$$\frac{1}{b-a}\int_a^b f(x)\,\mathrm{d}x$$

is often called the average value of the function $f(\cdot)$ in the interval $[a, b]$. Daniell's lag window spectral estimator (Daniell, 1946) is simply the average value of the direct spectral estimator $\hat{S}^{(\mathrm{D})}(\cdot)$ in an interval of length $1/(m\,\Delta_\mathrm{t})$ around each value of f:

$$\hat{S}_m^{(\mathrm{LW})}(f) = m\,\Delta_\mathrm{t} \int_{f-1/(2m\,\Delta_\mathrm{t})}^{f+1/(2m\,\Delta_\mathrm{t})} \hat{S}^{(\mathrm{D})}(f)\,\mathrm{d}f,$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

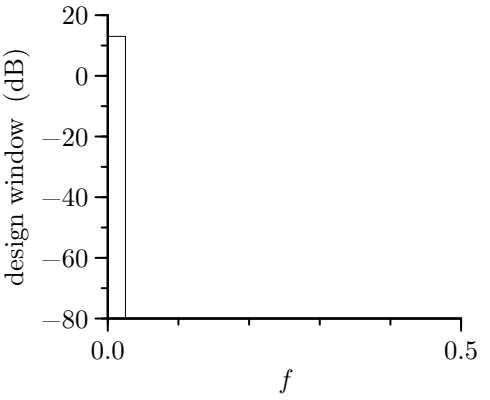

Figure 272a Daniell design window for $m = 20$.

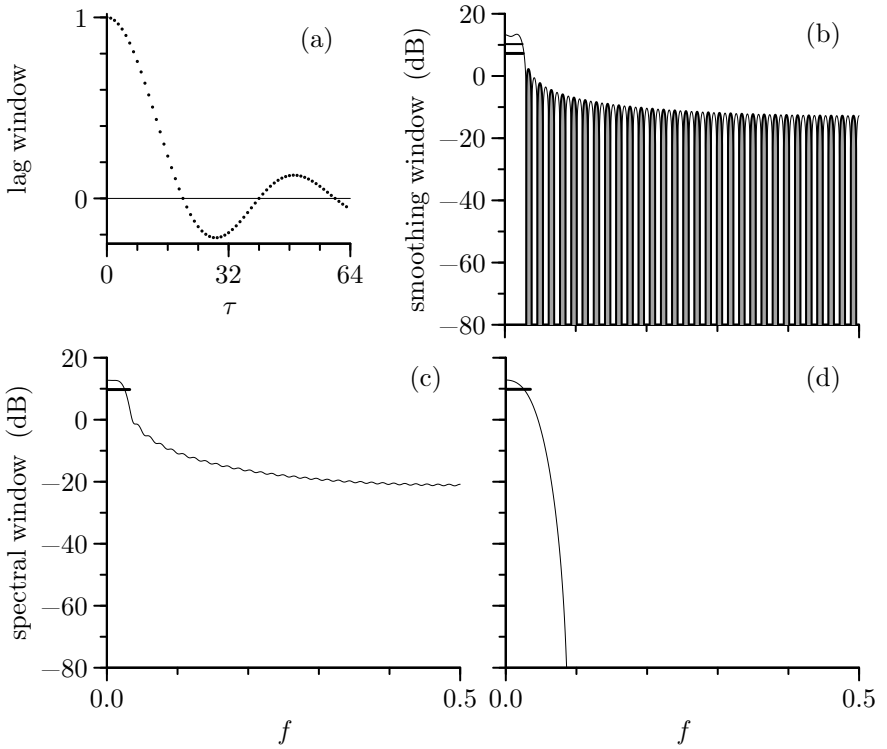

Figure 272b Daniell windows for $m = 20$ and $N = 64$ ($B_W \doteq 0.05$).

where, for f near $\pm f_{\mathcal{N}}$, we use the fact that $\hat{S}^{(\mathrm{D})}(\cdot)$ is $2f_{\mathcal{N}}$ periodic. Here the parameter $m \geq 1$ does not correspond to a truncation point – and need not be an integer – as in the case of the Bartlett lag window; however, it does control the degree of averaging to which $\hat{S}^{(\mathrm{D})}(\cdot)$ is subjected: the smaller m is, the greater the amount of smoothing. Comparison with Equation (247a) shows that the design window for Daniell's estimator is

$$V_m(f) = \begin{cases} m\,\Delta_{\mathrm{t}}, & |f| \leq 1/(2m\,\Delta_{\mathrm{t}}); \\ 0, & 1/(2m\,\Delta_{\mathrm{t}}) < |f| \leq f_{\mathcal{N}} \end{cases} \tag{272}$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

(the above also defines the design window for $|f| > f_{\mathcal{N}}$ since it is periodic with a period of $2f_{\mathcal{N}}$). The corresponding lag window is given by Equation (247d) in conjunction with Equation (247c):

$$w_{m,\tau} = \begin{cases} m\,\Delta_{\mathrm{t}} \int_{-1/(2m\,\Delta_{\mathrm{t}})}^{1/(2m\,\Delta_{\mathrm{t}})} \mathrm{e}^{\mathrm{i}2\pi f \tau\,\Delta_{\mathrm{t}}}\,\mathrm{d}f = \dfrac{\sin(\pi\tau/m)}{\pi\tau/m}, & |\tau| < N; \\ 0, & |\tau| \geq N \end{cases} \tag{273a}$$

(we interpret the ratio above as unity when $\tau = 0$). The smoothing window for Daniell's spectral estimator is thus

$$W_m(f) = \Delta_{\mathrm{t}} \sum_{\tau=-(N-1)}^{N-1} \frac{\sin(\pi\tau/m)}{\pi\tau/m} \mathrm{e}^{-\mathrm{i}2\pi f \tau\,\Delta_{\mathrm{t}}}. \tag{273b}$$

Figure 272a shows the Daniell design window $V_m(\cdot)$ for $m = 20$. Because the sidelobes of the corresponding smoothing window $W_m(\cdot)$ alternate between positive and negative values, we have plotted the quantity $|W_m(\cdot)|$ in Figure 272b(b) rather than $W_m(\cdot)$. The negative sidelobes are shaded – unfortunately, there are so many sidelobes that this shading is barely visible. The magnitude of the peak of the first sidelobe in $|W_m(\cdot)|$ is down about 11 dB from its value at $f = 0$. The envelope of the sidelobes decays as approximately f^{-1}, which corresponds to a decay of 3 dB per frequency octave. However, since Equation (247f) implies that we can just as well regard $V_m(\cdot)$ in Equation (272) as Daniell's smoothing window, we can also argue that the latter has effectively *no* sidelobes! This example shows that the smoothing window $W_m(\cdot)$ has to be interpreted with some care. Because $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ is the cyclic convolution of two periodic functions, one of them (namely, $\hat{S}^{(\mathrm{D})}(\cdot)$) with Fourier coefficients $\{\hat{s}_\tau^{(\mathrm{D})}\}$ that are identically 0 for lags $|\tau| \geq N$, there is an inherent ambiguity in $W_m(\cdot)$, which we have sidestepped by arbitrarily setting the Fourier coefficients of $W_m(\cdot)$ (namely, $\{w_{m,\tau}\}$) to 0 for lags $|\tau| \geq N$. Thus, the fact that a smoothing window has negative sidelobes does *not* automatically imply that the corresponding lag window spectral estimator can sometimes take on negative values!

Note that, in contrast to what happened with Bartlett's window (Figure 270), the two spectral windows $\mathcal{U}_m(\cdot)$ in Figure 272b now have a markedly different appearance. The first of these is again dominated by Fejér's kernel (the spectral window for the rectangular data taper shown in Figure 191(a)), but the second reflects the convolution of $V_m(\cdot)$ in Equation (272) and the spectral window for the Slepian data taper (shown in Figure 191(g)). We can make another practical interpretation. The user who chose, say, the Daniell smoothing window because of its lack of sidelobes and used the default rectangular taper might be surprised to find that because of the effect of Fejér's kernel – basically reflecting the finiteness of the data – the *overall* spectral window $\mathcal{U}_m(\cdot)$ again decays as just f^{-2}, i.e., the same rate as the sidelobes of the periodogram!

Bartlett–Priestley (Quadratic or Epanechnikov) Window (Figures 274a and 274b)
The Bartlett–Priestley window is specified via the following design window:

$$V_m(f) = \begin{cases} 3m\,\Delta_{\mathrm{t}} \left[1 - (2fm\,\Delta_{\mathrm{t}})^2\right]/2, & |f| \leq 1/(2m\,\Delta_{\mathrm{t}}); \\ 0, & 1/(2m\,\Delta_{\mathrm{t}}) < |f| \leq f_{\mathcal{N}}, \end{cases} \tag{273c}$$

where the parameter m can assume any value greater than or equal to 1 (the window is defined for $|f| > f_{\mathcal{N}}$ by periodic extension). The burden of Exercise [7.10a] is to show that the

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

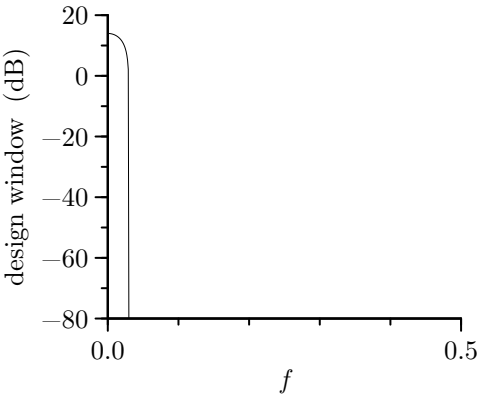

Figure 274a Bartlett–Priestley design window for m = 16.67.

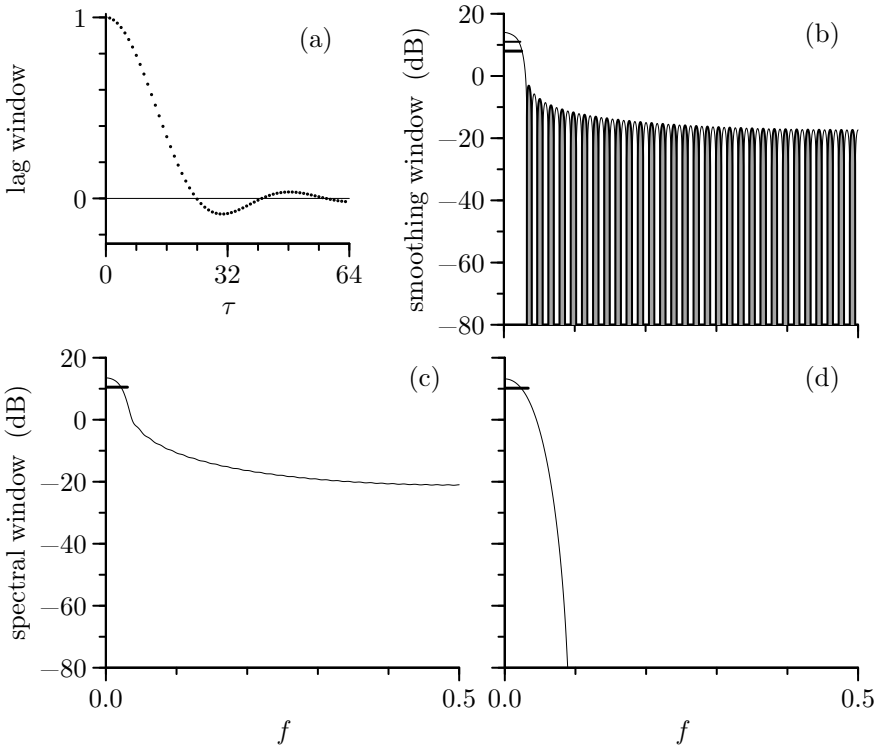

Figure 274b Bartlett–Priestley windows for m = 16.67 and N = 64 ($B_W \doteq 0.05$).

corresponding lag window is given by

$$w_{m,\tau} = \begin{cases} 1, & \tau = 0; \\ \dfrac{3m^2}{\pi^2\tau^2}\left[\dfrac{\sin(\pi\tau/m)}{\pi\tau/m} - \cos(\pi\tau/m)\right], & 1 \le |\tau| < N; \\ 0, & |\tau| \ge N. \end{cases} \tag{274}$$

Figure 274a shows the Bartlett–Priestley design window for $m = 16.67$, while Figure 274b depicts the corresponding lag, smoothing and spectral windows for this same value

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

of m and for $N = 64$. The design window visually resembles the Daniell design window (Figure 272a), but, while the latter transitions to zero abruptly, the Bartlett–Priestley window does so in a continuous manner. The Daniell and Bartlett–Priestley smoothing windows are also similar (plot (b) of Figures 272b and 274b), but the central lobe of the latter is smoother than the former. As is the case for the Daniell smoothing window, the Bartlett–Priestley smoothing window has negative sidelobes. The magnitudes of the sidelobes in the Daniell case are approximately 5 dB higher than in the Bartlett–Priestley case, but the decay rates are essentially the same; however, in view of Equation (247f), we can argue that, like the Daniell window, the Bartlett–Priestley smoothing window is equivalent to one with no sidelobes. The Daniell and Bartlett–Priestley spectral windows that are associated with the rectangular data taper (plot (c) of Figures 272b and 274b) are quite similar, with the exception of small ripples in the Daniell window not present in the Bartlett–Priestley. The two windows that are associated with the Slepian taper (plot (d) of Figures 272b and 274b) are virtually identical. As demonstrated in Figure 319 in Section 7.12, the discontinuous nature of the Daniell design window can result in a lag window estimate that is choppier looking than the comparable Bartlett–Priestley estimate, even though overall the two estimates track each other well because the first-moment properties of the Daniell and Bartlett–Priestley windows are nearly identical.

As discussed in Priestley (1962), the Bartlett–Priestley window arises as the solution to an optimality problem that involves minimizing an approximation to the normalized mean square error $E\{[\hat{S}_m^{(\mathrm{LW})}(f) - S(f)]^2/S^2(f)\}$ at one fixed frequency f with respect to a restricted class of design windows proposed in Parzen (1957). Priestley (1981, pp. 444–5) notes that Bartlett (1963) independently introduced this window but that Epanechnikov (1969) established its optimality properties in a more general context.

Parzen Window (Figure 276)
Parzen (1961) suggested the following lag window:

$$w_{m,\tau} = \begin{cases} 1 - 6\,(\tau/m)^2 + 6\,(|\tau|/m)^3\,, & |\tau| \le m/2; \\ 2\,(1 - |\tau|/m)^3\,, & m/2 < |\tau| < m; \\ 0, & |\tau| \ge m, \end{cases} \tag{275a}$$

where the window parameter m is an integer that can range from 1 up to N. The corresponding smoothing window is given by

$$W_m(f) = \frac{A_m(f)\,\Delta_{\mathrm{t}}}{m^3 \sin^4(\pi f\,\Delta_{\mathrm{t}})}, \tag{275b}$$

where, for m even,

$$A_m(f) \stackrel{\mathrm{def}}{=} 4\left[3 - 2\,\sin^2(\pi f\,\Delta_{\mathrm{t}})\right] \sin^4(m\pi f\,\Delta_{\mathrm{t}}/2)$$

(Priestley, 1981, p. 444), while, for m odd,

$$\begin{aligned} A_m(f) \stackrel{\mathrm{def}}{=}\ & [3 - 2\sin^2(\pi f\,\Delta_{\mathrm{t}})][2 - \sin^2(m\pi f\,\Delta_{\mathrm{t}})] \\ & - \cos(\pi f\,\Delta_{\mathrm{t}})\cos(m\pi f\,\Delta_{\mathrm{t}})[6 - \sin^2(\pi f\,\Delta_{\mathrm{t}})] \end{aligned}$$

(Exercise [7.11] gives equations for $W_m(0)$, which needs special care since the above equations are indeterminate at $f = 0$). The Parzen lag window can be derived by taking the Bartlett lag window (treated as a continuous function of τ) with parameter $m/2$, convolving it with itself, and then rescaling and sampling the resulting function (see the lower right-hand plot

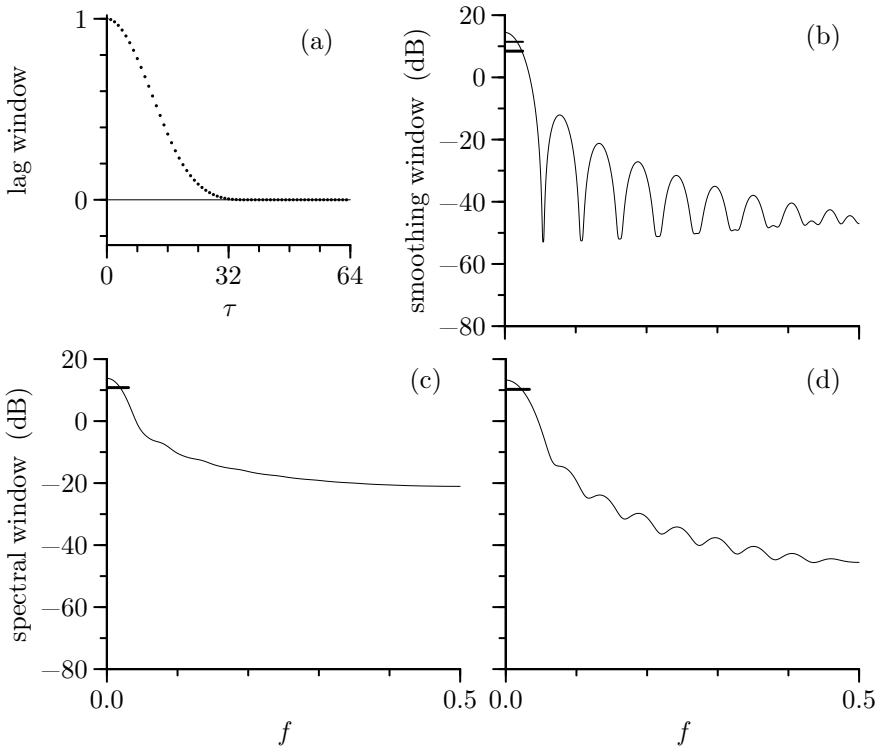

Figure 276 Parzen windows for $m = 37$ and $N = 64$ ($B_W \doteq 0.05$).

of Figure 69 and Equation (104)). In fact, for continuous τ, the Bartlett lag window and the Parzen lag window are related to the PDFs of the sum of, respectively, two and four uniformly distributed RVs (the functions shown in the right-hand column of Figure 69 are proportional to the PDFs for the sum of two, three and four such RVs). Because of the central limit theorem, we can regard the Parzen lag window – and hence its smoothing window – as having approximately the shape of a Gaussian PDF (cf. the lower right-hand plot of Figure 69).

Parzen SDF estimates, like the Bartlett, Daniell and Bartlett–Priestley estimates, are always nonnegative. As f increases, the envelope of Parzen's smoothing window decreases as approximately f^{-4}, i.e., 12 dB per octave. The magnitude of the peak of the first sidelobe is down about 28 dB from the magnitude of the central lobe. Thus the sidelobes decay much more rapidly than those of Bartlett's smoothing window (6 dB per octave), and the first sidelobe is also much smaller (28 dB down as compared to 13 dB).

Note that the two spectral windows $\mathcal{U}_m(\cdot)$ in Figure 276 are quite different. As for the Bartlett, Daniell and Bartlett–Priestley windows, the first of these is again dominated by Fejér's kernel, but the second is the convolution of Parzen's $W_m(\cdot)$ with the spectral window for the Slepian data taper and has sidelobes about 25 dB below that of the first at $f = 1/2$.

Gaussian Window (Figure 278)

The fact that the Parzen lag window was designed to have approximately the shape of a Gaussian PDF leads us to consider the following *Gaussian lag window*:

$$w_{m,\tau} = \begin{cases} \mathrm{e}^{-\tau^2/m^2}, & |\tau| < N; \\ 0, & |\tau| \geq N, \end{cases} \tag{276}$$

where the parameter m can assume *any* positive value (the smoothing window bandwidth

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

B_W decreases as m increases, as is true for all the other lag windows we have considered so far). Figure 278(a) shows an example of $\{w_{m,\tau}\}$ for the case $N = 64$ and $m = 16$. As expected, this window is visually quite similar to the Parzen lag window of Figure 276(a). The smoothing window for $\{w_{m,\tau}\}$ is

$$W_m(f) = \Delta_t \sum_{\tau=-(N-1)}^{N-1} e^{-\tau^2/m^2} e^{-i2\pi f \tau \Delta_t}. \quad (277)$$

As is also true for the Daniell and Bartlett–Priestley smoothing windows, $W_m(\cdot)$ has some negative sidelobes. This fact is demonstrated for our example in Figure 278(b), which plots $|W_m(f)|$ versus f with shading used to indicate negative sidelobes. The presence of negative sidelobes opens up the possibility that a lag window estimator based on $\{w_{m,\tau}\}$ might potentially be negative at some frequencies; however, as is also true in the Daniell and Bartlett–Priestley cases, this cannot happen, as the following exercise indicates.

▷ **Exercise [277]** Given any direct spectral estimator $\hat{S}^{(D)}(\cdot)$ based upon a time series of length N, show that, for all f,

$$\hat{S}_m^{(LW)}(f) = \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} W_m(f-\phi)\hat{S}^{(D)}(\phi)\,d\phi \geq 0,$$

where $W_m(\cdot)$ is given by Equation (277). ◁

Note that the magnitudes of the sidelobes of $W_m(\cdot)$ in Figure 278(b) are substantially smaller than those shown in Figure 276(b) for the Parzen smoothing window. The magnitude of the peak of the first sidelobe in $|W_m(\cdot)|$ is down about 79 dB from its value at $f = 0$, as compared to 28 dB for the Parzen window. As a result, the spectral window shown in Figure 278(d) more closely resembles the ones associated with the sidelobe-free Daniell and and Bartlett–Priestley windows (plot (d) of Figures 272b and 274b) than that associated with the Parzen window (Figure 276(d)). The envelope of the sidelobes for the Gaussian smoothing window decays as approximately $f^{-0.7}$, which corresponds to a decay of 2 dB per frequency octave. This poor decay rate (the slowest we have seen so far) is not particular important because the sidelobes are so small in magnitude.

Although, as we shall see in Section 7.12, the Gaussian lag window is competitive with the Daniell, Bartlett–Priestley and Parzen windows, it has not seen much use in practical applications. Neave (1972) and Harris (1978) include a Gaussian window $\exp(-\tau^2/m^2)$, $\tau \in \mathbb{R}$, in their investigation of windows in an abstract context (i.e., not necessarily for use as lag windows). One disadvantage they point out is the discontinuity that arises when this window is truncated to a finite interval ($[-(N-1), N-1]$ in our case). Depending upon the settings for m and N, this discontinuity can be large enough to cause a prominent ringing in the Fourier transform of either the truncated window or samples thereof. This ringing might adversely impact the performance of the Gaussian window in comparison to windows exhibiting continuity (an example being the Parzen and window). This discontinuity is, however, a moot point for a Gaussian *lag* window due to the nature of $\{\hat{s}_\tau^{(D)}\}$. By construction, $\hat{s}_\tau^{(D)} = 0$ for all $|\tau| \geq N$, which means that eliminating the discontinuity in Equation (276) by redefining $w_{m,\tau}$ to be $e^{-\tau^2/m^2}$ for all τ yields the same lag window estimator $\hat{S}_m^{(LW)}(\cdot)$ as with the existing definition. Thus, while the Gaussian window might not be appealing when viewed abstractly, it is well-suited for use as a lag window (the forthcoming Figures 319 and 320 document an example where the Gaussian lag window outperforms the Daniell, Bartlett–Priestley and Parzen lag windows on an actual time series).

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

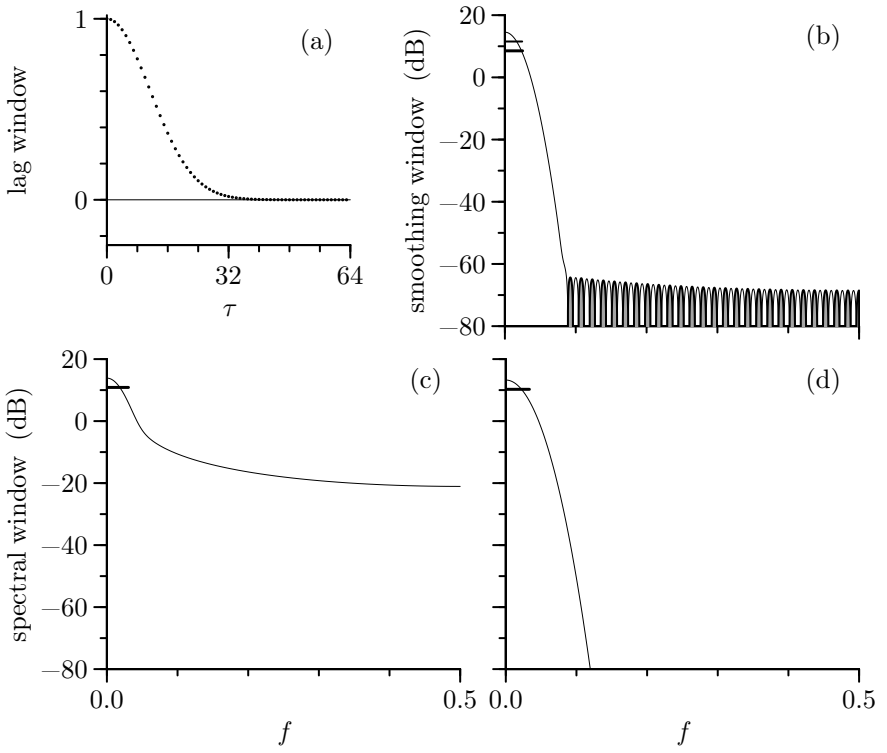

Figure 278 Gaussian windows for $m = 16$ and $N = 64$ ($B_W \doteq 0.05$).

Papoulis Window (Figure 279)
Papoulis (1973) found that the continuous τ analog of the following lag window produces a window with a certain minimum bias property:

$$w_{m,\tau} = \begin{cases} \dfrac{1}{\pi}|\sin(\pi\tau/m)| + (1 - |\tau|/m)\cos(\pi\tau/m), & |\tau| < m; \\ 0, & |\tau| \geq m, \end{cases} \tag{278}$$

where m is an integer ranging from 1 up to N (Bohman, 1961, derived this window earlier in the context of characteristic functions). The rationale behind this lag window is as follows. Equation (256d) tells us that the bias in a lag window spectral estimator due to the smoothing window alone is proportional to β_W^2, where β_W is defined in Equation (251a). The Papoulis window is the solution to the continuous τ analog of the following problem: for fixed $m > 0$, amongst all lag windows $\{w_{m,\tau}\}$ with $w_{m,\tau} = 0$ for $|\tau| \geq m$ and with a corresponding smoothing window $W_m(\cdot)$ such that $W_m(f) \geq 0$ for all f, find the window such that $\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} f^2 W_m(f)\,\mathrm{d}f \propto \beta_W^2$ is minimized.

The derivation of the Papoulis smoothing window is left as an exercise. Comparison of Figures 276 and 279 shows that the Parzen and the Papoulis windows have quite similar characteristics.

Comments and Extensions to Section 7.5

[1] We motivated our discussion of lag window estimators by first considering discretely smoothed direct spectral estimators $\{\hat{S}^{(\mathrm{DS})}(f'_k)\}$. As shown by Equation (246b), the latter are based on a set of smoothing coefficients $\{g_j\}$, which implicitly define a lag window $\{w_{g,\tau}\}$ (see Equations (249a),

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

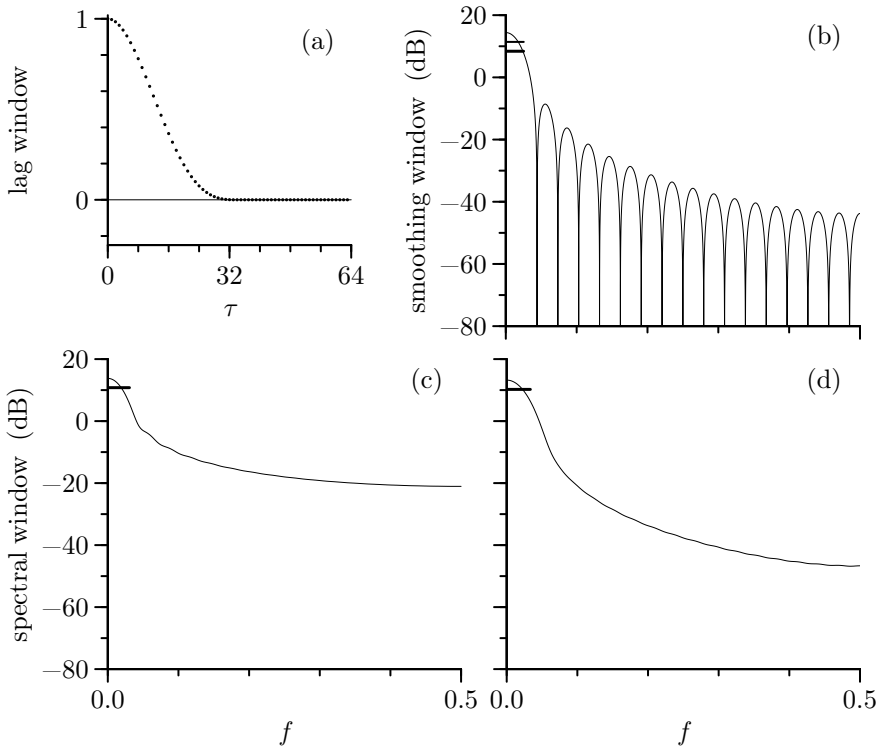

Figure 279 Papoulis windows for $m = 34$ and $N = 64$ ($B_W \doteq 0.05$).

Estimator	Asymptotic Variance	ν	B_W	β_W
Bartlett	$\frac{0.67 m C_h S^2(f)}{N}$	$\frac{3N}{mC_h}$	$\frac{1.5}{m\,\Delta_t}$	$\frac{0.92}{\sqrt{m}\,\Delta_t}$
Daniell	$\frac{mC_h S^2(f)}{N}$	$\frac{2N}{mC_h}$	$\frac{1}{m\,\Delta_t}$	$\frac{1}{m\,\Delta_t}$
Bartlett–Priestley	$\frac{1.2 mC_h S^2(f)}{N}$	$\frac{1.67N}{mC_h}$	$\frac{0.83}{m\,\Delta_t}$	$\frac{0.77}{m\,\Delta_t}$
Parzen	$\frac{0.54 mC_h S^2(f)}{N}$	$\frac{3.71N}{mC_h}$	$\frac{1.85}{m\,\Delta_t}$	$\frac{1.91}{m\,\Delta_t}$
Gaussian	$\frac{1.25 mC_h S^2(f)}{N}$	$\frac{1.60N}{mC_h}$	$\frac{0.80}{m\,\Delta_t}$	$\frac{0.78}{m\,\Delta_t}$
Papoulis	$\frac{0.59 mC_h S^2(f)}{N}$	$\frac{3.41N}{mC_h}$	$\frac{1.70}{m\,\Delta_t}$	$\frac{1.73}{m\,\Delta_t}$

Table 279 Asymptotic variance, equivalent degrees of freedom ν, smoothing window bandwidth B_W and alternative bandwidth measure β_W (used in Equation (256d) for the bias due to the smoothing window alone) for six lag window spectral density estimators. The tabulated quantities are approximations to the formulae given in Equations (260), (264b), (251e) and (251b). The quantity C_h – a variance inflation factor – depends upon the data taper used in the corresponding direct spectral estimator $\hat{S}^{(\mathrm{D})}(\cdot)$ (see Table 260 and Equation (262)).

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

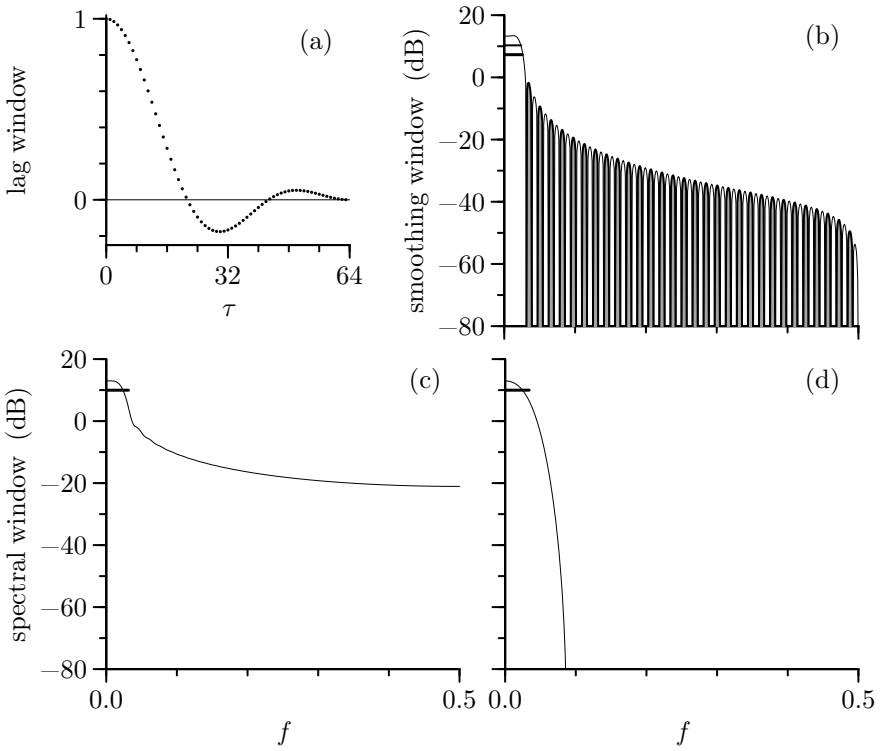

Figure 280 Modified Daniell windows for $M = 3$ and $N = 64$ ($B_W \doteq 0.05$).

(249b) and (249c)). The relationship between $\{w_{g,\tau}\}$ and $\{g_j\}$ is similar to that between a lag window and its smoothing window, suggesting that we can sample a smoothing window to obtain smoothing coefficients.

As an example of obtaining $\{g_j\}$ in this manner, let us consider smoothing coefficients of the form

$$g_j = \begin{cases} 1/(2M), & |j| < M; \\ 1/(4M), & |j| = M; \\ 0, & \text{otherwise.} \end{cases}$$

We can consider these weights as being generated by sampling from a Daniell smoothing window $V_m(\cdot)$ (see Equation (272)) with an "end point" adjustment. Bloomfield (2000) calls the resulting estimator $\{\hat{S}^{(\mathrm{DS})}(f'_k)\}$ a *modified Daniell* spectral estimator. Note that $\{g_j\}$ can be regarded as a low-pass LTI filter. Since $\sum_{j=-M}^{M} g_j = 1$, this filter has the normalization that we argued in Section 5.7 is appropriate for a smoother. Assume the frequencies involved in $\{\hat{S}^{(\mathrm{DS})}(f'_k)\}$ are given by $f'_k = \tilde{f}_k = k/(2N\,\Delta_t)$. We can use Equations (249b) and (249c) to reexpress $\hat{S}^{(\mathrm{DS})}(\cdot)$ as a lag window spectral estimator $\hat{S}_m^{(\mathrm{LW})}(\cdot)$. The corresponding lag and smoothing windows are shown, respectively, in Figures 280(a) and (b) for the case $M = 3$ and $N = 64$. As for the usual Daniell window in Figure 272b, the smoothing window has negative sidelobes (shaded in plot (b)), but, because the estimate is necessarily nonnegative, these sidelobes are again an artifact caused by truncation. The spectral windows for the modified Daniell window (plots (c) and (d)) closely resemble those for the usual Daniell window ((c) and (d) of Figure 272b).

The device of sampling from a smoothing window to produce the coefficients for a discretely smoothed direct spectral estimator can obviously be applied to other smoothing windows besides the Daniell window. A variation on this idea is to generate the coefficients from, say, the Parzen smoothing window (with a renormalization to ensure that the weights sum to unity). This form of spectral estimator is discussed in Cleveland and Parzen (1975) and Walden (1990a).

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

[2] We introduced the idea of a reshaped lag window in C&E [2] for Section 7.2 (see Equation (257b)). Figure 282 gives an example for such a window formed from the Parzen lag window with $m = 37$ and the Slepian data taper with $NW = 4$ (a combination appearing in Figure 276). When used in conjunction with the rectangular data taper, the lag window in Figure 282(a) thus yields exactly the same spectral window as shown in Figure 276(d). The corresponding smoothing window is shown in plot (b), where now we plot $|W_m(\cdot)|$ rather than $W_m(\cdot)$. It has a prominent single *negative* sidelobe (indicated by the shaded area). That this lag window can lead to negative spectral estimates is demonstrated in plot (c). This plot shows two lag window spectral estimates for the first 64 values of the realization of the AR(4) process $\{X_t\}$ depicted in Figure 34(e). The thin curve is the lag window estimate formed using the Parzen lag window in combination with the Slepian data taper; the thick curve uses the reshaped lag window with the rectangular data taper. The shaded areas under the latter curve indicate the frequencies where $\hat{S}_m^{(\mathrm{LW})}(f)$ is negative so that $|\hat{S}_m^{(\mathrm{LW})}(f)|$ is plotted rather than $\hat{S}_m^{(\mathrm{LW})}(f)$. The true SDF for the AR(4) process is shown as the thick curve in plot (d); the thin curve there is $E\{\hat{S}_m^{(\mathrm{LW})}(\cdot)\}$, which is the result of convolving the spectral window in Figure 276(d) with the true SDF and hence is the same for both spectral estimates depicted in 282(c). (In view of Equation (255a), there are two other ways of constructing $E\{\hat{S}_m^{(\mathrm{LW})}(\cdot)\}$. First, it is the convolution of $E\{\hat{S}^{(\mathrm{D})}(\cdot)\}$ given by the thin curve in the $NW = 4$ plot of Figure 193 with the smoothing window of Figure 276(b). Second, it is the convolution of $E\{\hat{S}^{(\mathrm{P})}(\cdot)\}$ shown by the thin curve in Figure 178(b) with the smoothing window whose absolute value is shown in Figure 282(b). Note in the latter case that, even though the smoothing window has a prominent negative sidelobe, the resulting $E\{\hat{S}_m^{(\mathrm{LW})}(\cdot)\}$ is still totally positive.)

While it is possible to create a lag window estimator using a rectangular data taper whose *expected value* is the same as another lag window estimator using a non-rectangular taper, this example shows that, for a particular time series, the *actual* SDF estimates can be qualitatively quite different: the reshaped lag window estimator can assume undesirable negative values, whereas, for the Bartlett, Daniell, Bartlett–Priestley, Parzen, Gaussian and Papoulis lag windows in combination with an arbitrary data taper, the lag window estimator is guaranteed to be nonnegative, thus matching a basic property of true SDFs.

[3] We noted following our description of Grenander's measure β_W of smoothing window bandwidth (Equation (251a)) that β_W could assume an imaginary value if the smoothing window is negative at some frequencies. The smoothing window shown in Figure 282(b) is negative for $f \in [0.05, 0.5]$. A computation indicates that $\beta_W \doteq \sqrt{(-0.0087)} \doteq 0.093\mathrm{i}$, hence illustrating this potential problem. By contrast, Jenkin's measure B_W of smoothing window bandwidth (Equation (251c)) is 0.057 and, as depicted by the short horizontal line in Figure 282(b), is intuitively sensible.

[4] As pointed out in our discussion of the Daniell window, smoothing a direct spectral estimator using its rectangular design window (see Figure 272a) is entirely equivalent to smoothing it using its corresponding smoothing window (see Figure 272b(b)), even though the latter has sidelobes that rapidly oscillate between positive and negative values. Because of this equivalence we can regard the Daniell smoothing window as having no sidelobes. Taking this point of view, there is an interesting pattern in plots (c) and (d) in the figures for the Bartlett, Daniell, Bartlett–Priestley, Parzen, Gaussian and Papoulis lag windows (Figures 270, 272b, 274b, 276, 278 and 279). These plots show the spectral window $\mathcal{U}_m(\cdot)$, which, from Equation (255c), is the convolution of the smoothing window $W_m(\cdot)$ and the spectral window $\mathcal{H}(\cdot)$ determined by the data taper of the underlying direct spectral estimator. The rolloff in each plot (c) reflects the slower of the rolloffs of $W_m(\cdot)$ and $\mathcal{H}(\cdot)$ for the rectangular data taper, i.e., Fejér's kernel $\mathcal{F}(\cdot)$. In all six cases, Fejér's kernel dominates the rolloff, so all of the (c) plots look similar. The rolloff in each plot (d) reflects the slower of the rolloffs of $W_m(\cdot)$ and $\mathcal{H}(\cdot)$ for a Slepian data taper. Since $\mathcal{H}(\cdot)$ now damps down so fast, the rolloff is dominated by the smoothing window $W_m(\cdot)$ for the Bartlett, Parzen and Papoulis windows, but *not* for the Daniell and Bartlett–Priestley windows. In these two cases, the rolloff is dominated by $\mathcal{H}(\cdot)$ since $W_m(\cdot)$ can be regarded as either a rectangular or a quadratic window with an "infinitely fast" rolloff. The Gaussian case requires special attention. A comparison of Figures 191(g) and 278(b) shows that $\mathcal{H}(\cdot)$ and $W_m(\cdot)$ are similar in that each has a prominent central lobe surrounded by small sidelobes. Their convolution $\mathcal{U}(\cdot)$ has a central lobe that is necessarily wider than that of $W_m(\cdot)$, with an accompanying suppression of its sidelobes.

The important point to keep in mind is that the central lobe and the decay rate of the sidelobes of the spectral window $\mathcal{U}_m(\cdot)$ depend on both the spectral window corresponding to the data taper and the smoothing window. Thus, for example, use of the Parzen smoothing window with a sidelobe decay rate

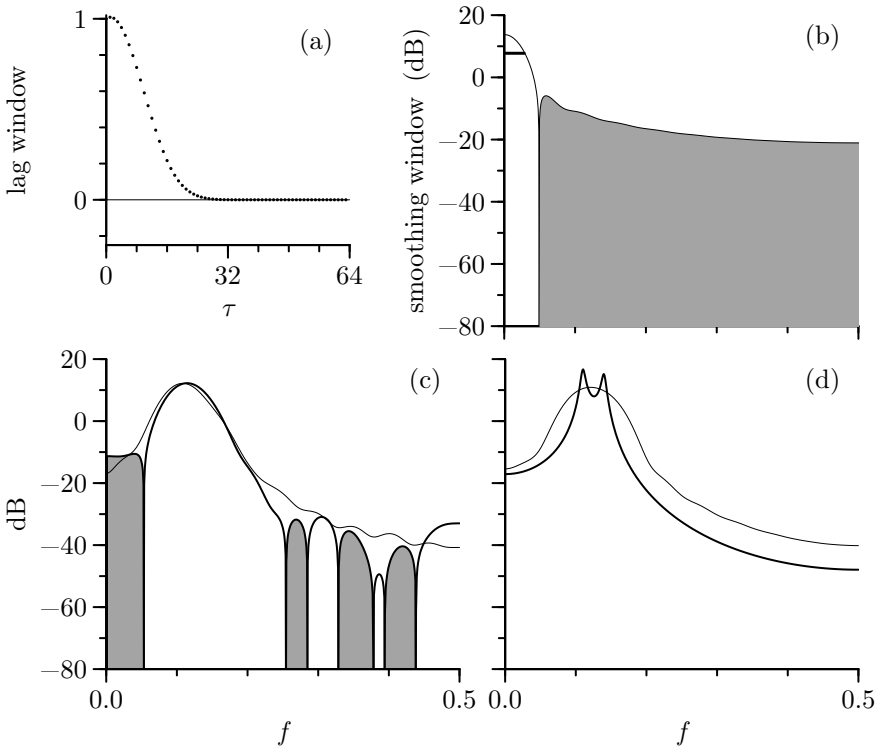

Figure 282 Reshaped lag window formed from a Parzen lag window with $m = 37$ and a Slepian data taper with $NW = 4$ for sample size $N = 64$ (cf. Figure 276). Plot (a) shows the reshaped lag window, while (b) shows the absolute value of the corresponding smoothing window $W_m(\cdot)$, with shaded portions indicating where $W_m(f) < 0$ (the width of the short horizontal line is equal to half of $B_W \doteq 0.057$). The thin curve in plot (c) shows an ordinary SDF estimate based upon a Parzen lag window with $m = 37$ and a Slepian data taper with $NW = 4$ for the first 64 values of the AR(4) time series shown in Figure 34(e); the thick curve shows the absolute value of the corresponding reshaped lag window estimate, with shaded portions indicating where this estimate is negative. The thin curve in (d) shows the theoretical expected value corresponding to the two lag window estimates in (c), while the thick curve is the true AR(4) SDF.

of 12 dB per octave does not imply that the sidelobes of $\mathcal{U}_m(\cdot)$ also have this decay rate *unless* the data taper for the direct spectral estimator is suitably chosen.

[5] We have presented data tapers as a means of reducing the bias in direct spectral estimators, and lag windows as a way of decreasing their variance. Despite differences in their intended purposes, a glance at Figures 185 and 248 shows that tapering and lag windowing have strong similarities in how they are implemented. In both cases, there is a sequence (either a time series $\{X_t : t = 0, 1, \ldots, N-1\}$ or an ACVS estimator $\{\hat{s}_\tau^{(\mathrm{D})} : \tau = -(N-1), -(N-2), \ldots, N-1\}$) whose beginning and end we would like to damp down to zero. We accomplish this by a point-by-point multiplication of the sequence by another sequence (the data taper or the lag window). There are, of course, obvious differences between data tapers and lag windows. First, the indices t for $\{h_t\}$ and τ for $\{w_{m,\tau}\}$ have different ranges. Second, the usual normalization for a data taper is $\sum_t h_t^2 = 1$, whereas the requirement $w_{m,0} = 1$ sets the normalization for a lag window. If we ignore these two differences, certain data tapers and lag windows are remarkably similar. Figure 283 shows curves for a Slepian data taper and a Parzen lag window that give the appearance of being almost identical.

The similarity in appearance between certain data tapers and lag windows suggests that, with adjustments in indexing and normalization, a good data taper can be transformed into a good lag window and *vice versa*. To some extent, this suggestion is correct; however, because data tapers and lag windows are designed with different goals in mind (lessening leakage and smoothing bumpy direct spectral estimators, respectively), there are subtle differences between what constitutes a good data taper and a

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

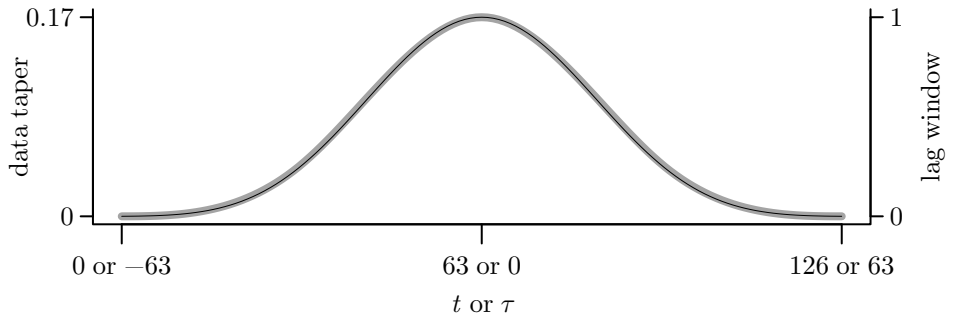

Figure 283 Comparison of Parzen lag window $\{w_{m,\tau}\}$ with $m = 64$ (thick gray curve) and Slepian data taper $\{h_t\}$ for $N = 127$ and $NW = 3.5$ (thin black curve). When regarded as lag windows, both displayed curves have associated smoothing window bandwidth $B_W \doteq 0.029$.

good lag window. For example, a lag window whose smoothing window is nonnegative would seem to be advantageous because this constraint rules out the possibility of a negative SDF estimate (however, as demonstrated by the Daniell, Bartlett–Priestley and Gaussian smoothing windows shown in parts (b) of Figures 272b, 274b and 278, this constraint is sufficient – but not necessary – to guarantee $\hat{S}_m^{(\mathrm{LW})}(f) \geq 0$). Since the smoothing window is the DFT of the lag window, this constraint takes the form of requiring the DFT to be nonnegative. The DFTs for all eight data tapers shown in Figure 190 have sidelobes that are negative. Transforming a data taper into a lag window can thus potentially lead to a negative SDF estimate at some frequencies. On the other hand, we typically do not judge the quality of a data taper based directly on its DFT, but rather on its spectral window (the squared magnitude of its DFT).

Let us first consider converting data tapers into lag windows, starting with the $p \times 100\%$ cosine tapers of Equation (189a). Adaptation of the defining equation yields the lag window

$$w_{m,\tau} = \begin{cases} 1, & 0 \leq |\tau| \leq m - \lfloor pM/2 \rfloor - 1; \\ \dfrac{1}{2}\left[1 - \cos\left(\dfrac{\pi(m - |\tau|)}{\lfloor pM/2 \rfloor + 1}\right)\right], & m - \lfloor pM/2 \rfloor \leq |\tau| \leq m - 1; \\ 0, & \tau \geq m, \end{cases} \tag{283}$$

where m can be any positive integer, and $M \stackrel{\mathrm{def}}{=} 2m - 1$. Figure 284 shows (a) the lag window and (b) the smoothing window for the case $p = 1$ and $m = 27$ with $N = 64$, in addition to the spectral windows corresponding to using (c) the rectangular data taper and (d) the $NW = 4$ Slepian data taper. The smoothing window has some negative sidelobes, so Figure 284(b) is a plot of $|W_m(f)|$ in decibels versus f, with shading indicating the sidelobes for which $W_m(f) < 0$. To ascertain the effect of these negative sidelobes on $\hat{S}_m^{(\mathrm{LW})}(\cdot)$, we generated 100,000 independent realizations of length $N = 64$ from the AR(4) process of Equation (35a). Figure 193 indicates the Hanning data taper yields a direct spectral estimator $\hat{S}^{(\mathrm{D})}(\cdot)$ with overall bias properties that are good in comparison to the other seven tapers considered in that figure, so we computed this $\hat{S}^{(\mathrm{D})}(\cdot)$ for each realization and then formed $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ using the 100% cosine lag window shown in Figure 284(a). We found 99,859 of the lag window SDF estimates to be negative at some frequencies (Figure 291 shows an additional case). Thus, in contrast to the negative sidelobes of the Daniell, Bartlett–Priestley and Gaussian smoothing windows, those of the 100% cosine smoothing window really can lead to negative SDF estimates. (Similar results are found for $p \times 100\%$ cosine lag windows with $0 \leq p < 1$.)

The fact that the spectral window for a data taper is necessarily nonnegative suggests another approach for generating a lag window from a data taper, with the advantage that the resulting smoothing window is nonnegative. Recall that, if $\{h_t\} \longleftrightarrow H(\cdot)$, then $\{h \star h_t\} \longleftrightarrow |H(\cdot)|^2 = \mathcal{H}(\cdot)$ (the spectral window). Additionally, if $\{h_t : t = 0, 1, \ldots, N-1\}$ has the usual normalization $\sum_t h_t^2 = 1$, then its autocorrelation $\{h \star h_\tau : \tau = -(N-1), -(N-2), \ldots, N-1\}$ has the proper length, indexing and normalization (since $h \star h_0 = \sum_t h_t^2$) to be a lag window for a time series of length N. The

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

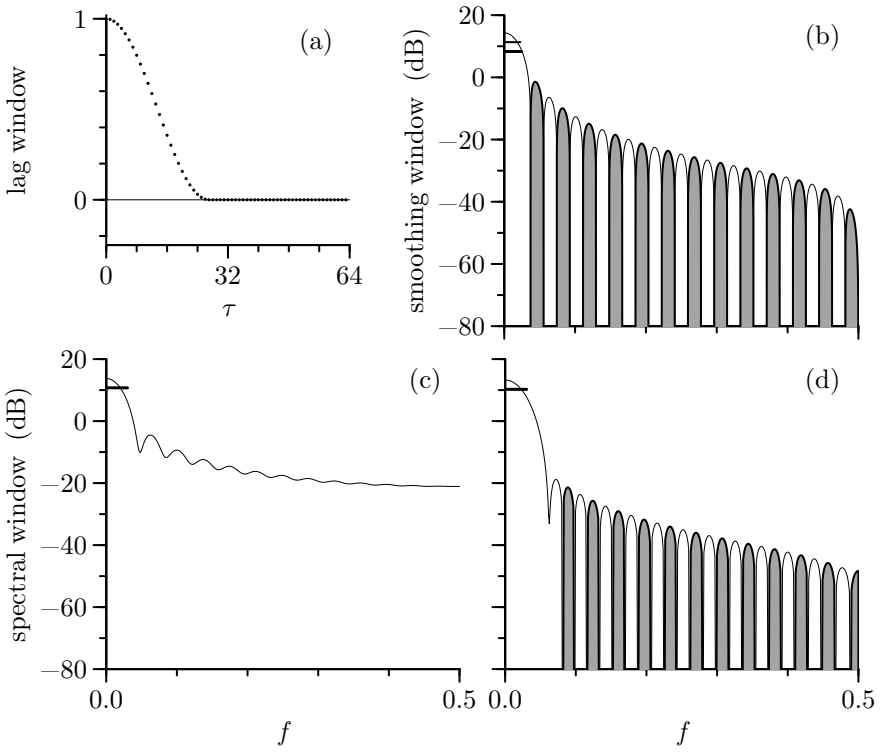

Figure 284 100% cosine windows for $N = 64$ and $m = 27$ ($B_W \doteq 0.05$).

smoothing window for the lag window defined by $w_{m,\tau} = h \star h_\tau$ is the same as the spectral window $\mathcal{H}(\cdot)$ and hence is nonnegative. If we use the $p \times 100\%$ cosine tapers with this scheme, we can create a different lag window for each choice of $p \in [0, 1]$ for a given N; however, the smoothing window bandwidths B_W for these windows are limited in range. For example, when $N = 64$, we find that $0.023 < B_W < 0.033$, so we cannot, for example, generate a lag window whose smoothing window bandwidth is approximately 0.05 (the nominal value used in our examples of other lag windows).

A variation on the autocorrelation scheme that yields a wider range for B_W is to generate a taper of length $m \leq N$ for a given p and then pad this taper with $N - m$ zeros. This padded taper of length N is then autocorrelated to create the desired lag window. With $N = 64$ and $p = 1$, we now find $0.032 < B_W \leq 1$ as we vary m from 1 to 64, while the choice $p = 0.2$ yields $0.024 < B_W \leq 1$ (for comparison, note that the spacing between Fourier frequencies is $1/N \doteq 0.016$). The parameter m is a truncation point since $w_{m,m} = 0$ whereas $w_{m,\tau} \neq 0$ for $|\tau| < m$. Figure 285 shows an example of a lag window created by autocorrelating a 100% cosine data taper of length $m = 41$ (zero padded to length $N = 64$). This lag window compares quite favorably with Parzen and Papoulis lag windows with similar smoothing window bandwidths of $B_W = 0.05$ (see Figures 276 and 279).

The greater flexibility of the Slepian tapers facilitates the creation of suitable lag windows either directly or via an autocorrelation. Figure 191 makes it clear that, by increasing NW, we can increase the width of the central lobe of the Fourier transform for the Slepian tapers well beyond what we can achieve by varying p within the $p \times 100\%$ cosine tapers. We can thus use a Slepian taper of length $2N - 1$ to directly create a lag window appropriate for a time series of length N, with the smoothing window bandwidth B_W being controlled by adjusting $m = (2N - 1)W$ (note that B_W increases as m increases, which is the opposite of what happens for the lag windows listed in Table 279). Figure 286 shows an example of a lag window created in this manner with $m = 10.2$. The Slepian lag window is quite competitive with other lag windows we have examined. Its smoothing window does have some negative sidelobes, but, over these sidelobes, $|W(f)|$ is always less than -110 dB and hence are not

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

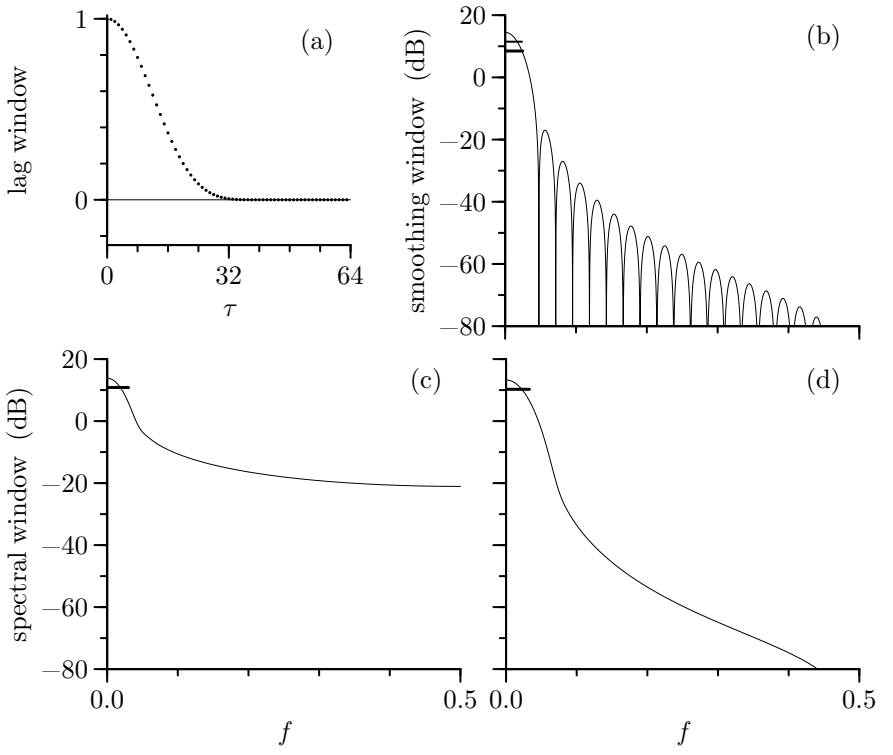

Figure 285 Autocorrelated 100% cosine windows for $N = 64$ and $m = 41$ ($B_W \doteq 0.05$).

visible in Figure 286(b). There is thus the possibility of negative SDF estimates, but a repetition of our experiment involving 100,000 realizations of the AR(4) process of Equation (35a) failed to find *any* cases for which $\hat{S}_m^{(\mathrm{LW})}(f) < 0$. (Exercise [7.13] invites the reader to form a Slepian-based lag window with a nonnegative smoothing window by autocorrelating a Slepian taper of length N.)

Two comments are in order. First, it is desirable in principle for the upper limit of the smoothing window bandwidth B_W to be twice the Nyquist frequency (this width would be appropriate for smoothing a periodogram of a white noise process since it would recover $\hat{S}_m^{(\mathrm{LW})}(f) = \Delta_\mathrm{t}\, \hat{s}_0^{(\mathrm{P})}$ as the SDF estimator). Thus, when $\Delta_\mathrm{t} = 1$, we would ideally have B_W range all the way up to unity. By making $m = (2N - 1)W$ sufficiently large for a Slepian lag window, we can get B_W close to unity, but we cannot get $B_W = 1$ exactly, as can be done with, for example, the Parzen and Papoulis lag windows by setting $m = 1$. Second, while the scheme of creating a lag window by reindexing and renormalizing a data taper works well with the Slepian taper, it is not so successful with other tapers. Suppose we set m equal to N in Equation (283) to create a lag window based upon a $p \times 100\%$ cosine taper. Our only option for controlling the smoothing window bandwidth is to vary p. In our $N = 64$ example, B_W can only vary from 0.008 to 0.021, which is too small for this scheme to be practical.

Now let's consider converting lag windows into data tapers. A lag window always has an odd length, whereas a data taper needs to be of the same length as the time series under study. Unfortunately time series with even lengths are more often than not the case these days because measurement systems preferentially collect time series with lengths set to a power of two. When N is even, an *ad hoc* solution to this mismatch is to base the data taper on a lag window of odd length $N - 1$ after padding the latter with a single zero. This approach unfortunately has the effect of shortening the time series by one value, but we note that a commonly used alternative to our definition of the Hanning data taper does this also (see Exercise [6.18a]).

Another approach for extracting a data taper is possible for the Bartlett, Parzen and Papoulis lag windows, for which the lag window parameter m represents a truncation point. We can regard these

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

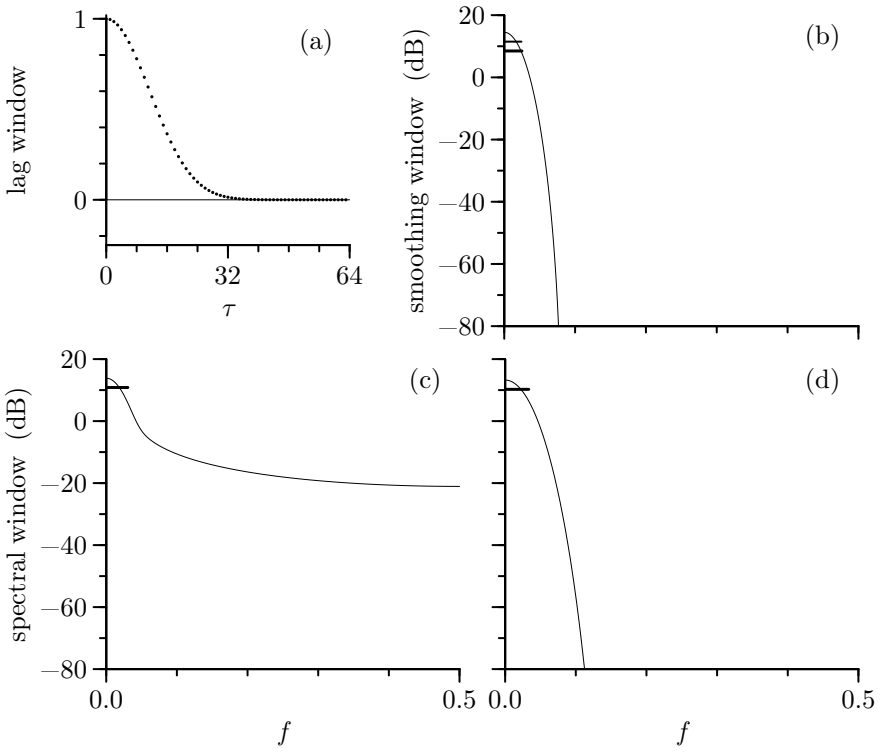

Figure 286 Slepian windows for $N = 64$ and $m = 10.2$ ($B_W \doteq 0.05$).

windows as arising from samples of a continuous function, say $h(\cdot)$, whose support is over the finite interval $(-1, 1)$; i.e., $h(t) = 0$ when $|t| \geq 1$. We can create a data taper appropriate for any sample size by setting

$$\tilde{h}_t = h\left(\frac{2(t+1)}{N+1} - 1\right), \quad t = 0, 1, \ldots, N-1, \tag{286}$$

and then defining $h_t = \tilde{h}_t / \sqrt{\sum_t \tilde{h}_t^2}$ to obtain a properly normalized taper. We can also regard the Daniell, Bartlett–Priestley and Gaussian lag windows as arising from samples of continuous functions, but not ones with finite support; however, here we can use the parameter m to define a suitable $h(\cdot)$ so that we can obtain $\tilde{h}_t$ as per Equation (286).

Figures 287(a) to (e) show examples of data tapers based upon the (a) Daniell, (b) Bartlett–Priestley, (c) Parzen, (d) Gaussian and (e) Papoulis lag windows, with corresponding plots in Figure 288 showing their spectral windows. For comparison, Figures 287(f) and 288(f) show a conventional Slepian taper with $NW = 3.14$. The Parzen and Papoulis tapers have a similar autocorrelation width of $B_{\mathcal{H}} \doteq 0.04$, so we set the smoothing window parameters m in the Daniell, Bartlett–Priestley and Gaussian cases (and NW in the Slepian case) to achieve this same spectral window bandwidth. We can judge the relative merits of these tapers by studying the sidelobes in their spectral windows. Since smaller sidelobes imply better protection against leakage, the Daniell and Bartlett–Priestley tapers are clearly inferior choices. The spectral window for the Slepian taper has sidelobes uniformly smaller than those for the Gaussian taper, but the same cannot be said for the Parzen and Papoulis cases. While the Slepian taper should yield a $\hat{S}^{(\mathrm{D})}(\cdot)$ with less bias because the sidelobes of its spectral window closest to the central lobe are quite a bit smaller than in the Parzen and Papoulis cases, conceivably there are SDFs for which use of the latter two tapers would result in less bias because their more distant sidelobes are smaller.

In conclusion, with sufficient care, it is possible to convert data tapers into decent lag windows, and lag windows into acceptable data tapers. C&E [1] in the next section discusses the relative merits

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

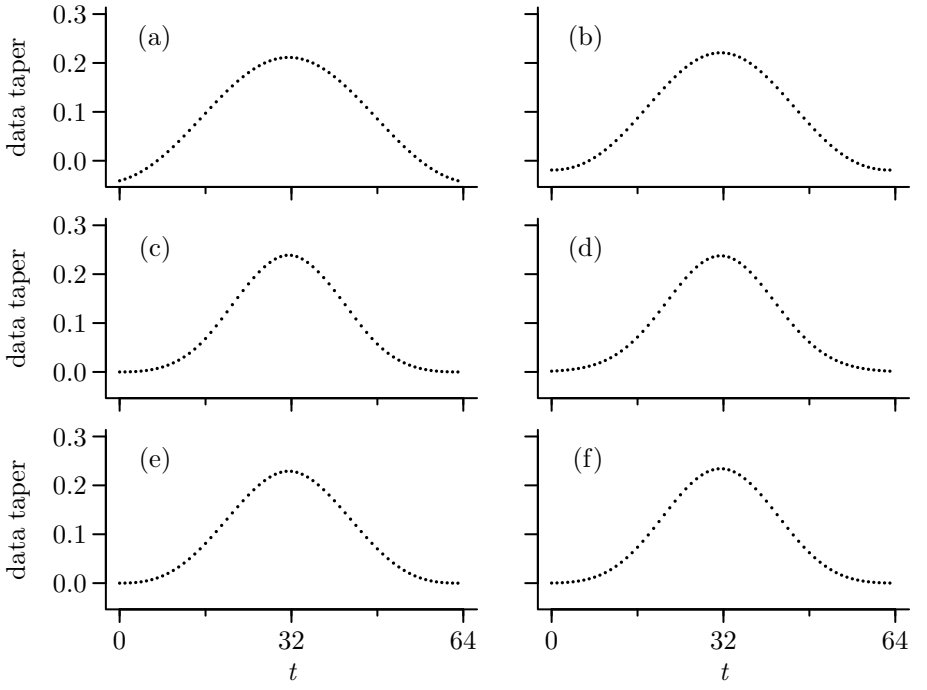

Figure 287 Six data tapers for use with a time series of length $N = 64$ (cf. Figure 190). The first five tapers are extracted from the (a) Daniell, (b) Bartlett–Priestley, (c) Parzen, (d) Gaussian and (e) Papoulis lag windows, while the final taper (f) is a zeroth-order Slepian data taper with $NW = 3.14$. The lag window parameter m is set to, respectively, 0.754, 0.528 and 0.435 for the Daniell, Bartlett–Priestley and Gaussian windows.

of the specific lag windows we generated from data tapers, pointing out that the Slepian lag window and the autocorrelated 100% cosine lag window are quite competitive with conventional lag windows. The Parzen and Papoulis data tapers we generated from lag windows might also have some appeal over conventional data tapers.

7.6 Choice of Lag Window

In the previous section we considered the Bartlett, Daniell, Bartlett–Priestley, Parzen, Gaussian and Papoulis lag window spectral estimators, each of which has a window parameter m controlling its lag window and smoothing window. For a particular SDF estimation problem, which estimator should we use, and how should we set m? We address the first of these questions in this section, and the second in the next section.

Several different criteria have been proposed in the literature for evaluating different lag window spectral estimators. One of the more useful is based on the concept of *smoothing window leakage*. Since

$$\hat{S}_m^{(\mathrm{LW})}(f) = \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} W_m(f-\phi)\hat{S}^{(\mathrm{D})}(\phi)\,\mathrm{d}\phi$$

(Equation (248a)), a lag window estimator of $S(f)$ is the result of integrating the product of the direct spectral estimator $\hat{S}^{(\mathrm{D})}(\cdot)$ and the smoothing window $W_m(\cdot)$ after the latter has been shifted so that its central lobe is centered at frequency f. Under our operational assumption that $S(\cdot)$ is slowly varying, we want $\hat{S}_m^{(\mathrm{LW})}(f)$ to be influenced mainly by values in $\hat{S}^{(\mathrm{D})}(\cdot)$ with frequencies "close" to f. We define "close" here to mean those frequencies lying within

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

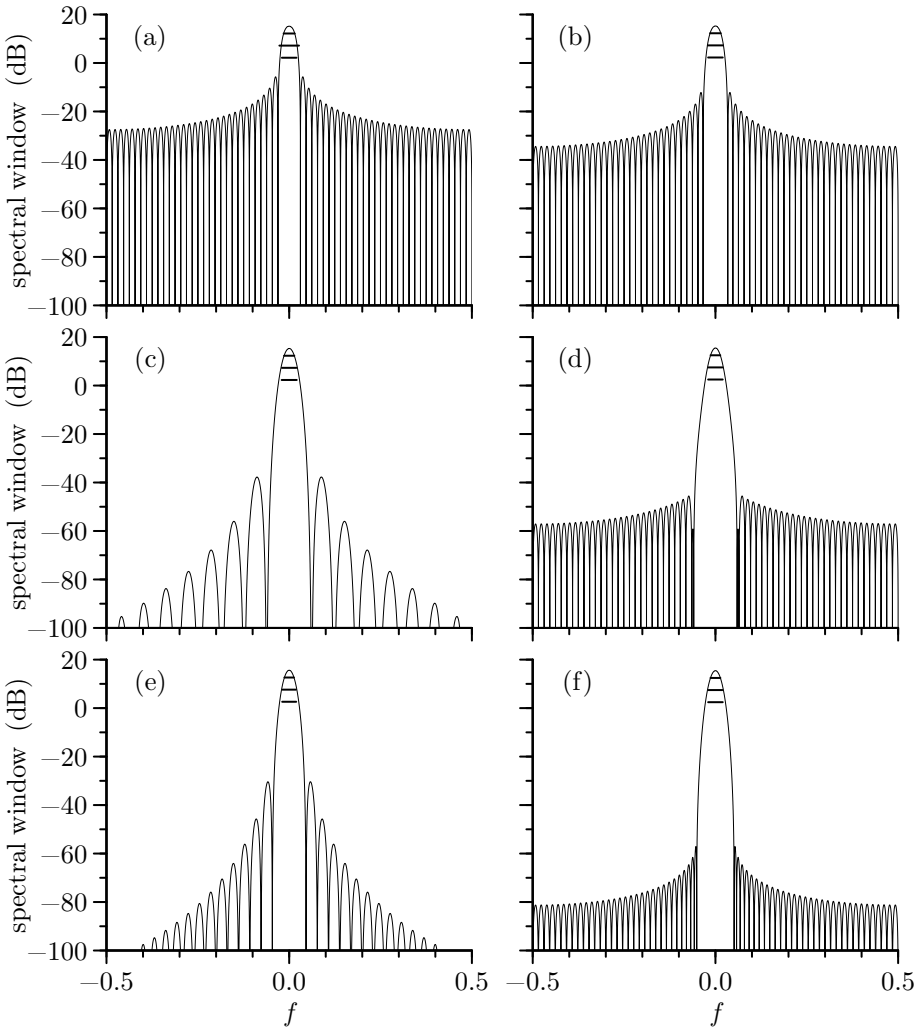

Figure 288 Spectral windows for the six data tapers shown on Figure 287 (cf. Figure 191). The three horizontal solid lines below the central lobes are bandwidth measures based upon, from the top on down, the half-power, variance and autocorrelation widths. The autocorrelation width is $B_{\mathcal{H}} \doteq 0.04$ for all six spectral windows.

the central lobe of the shifted smoothing window $W_m(f - \cdot)$. If this smoothing window has significant sidelobes and if the dynamic range of $\hat{S}^{(\mathrm{D})}(\cdot)$ is large, $\hat{S}_m^{(\mathrm{LW})}(f)$ can be unduly influenced by values in $\hat{S}^{(\mathrm{D})}(\cdot)$ lying under one or more of the sidelobes of the smoothing window. If this in fact happens, we say that the estimate $\hat{S}_m^{(\mathrm{LW})}(f)$ suffers from smoothing window leakage.

One criterion for window selection is thus to insist that the smoothing window leakage be small. If we have two different lag window estimators whose smoothing windows have the same bandwidth B_W (a measure of the width of the central lobe of $W_m(\cdot)$), this criterion would dictate picking the window whose sidelobes are in some sense smaller. Considering the six smoothing windows discussed in the previous section, the first sidelobe for the Bartlett window is 13 dB down from its central lobe, and the decay rate for its sidelobes is 6 dB per octave; the first sidelobes of the Parzen and Papoulis windows are 28 and 23 dB down, and both have decay rates of 12 dB per octave; for the Gaussian window, the comparable measures are

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

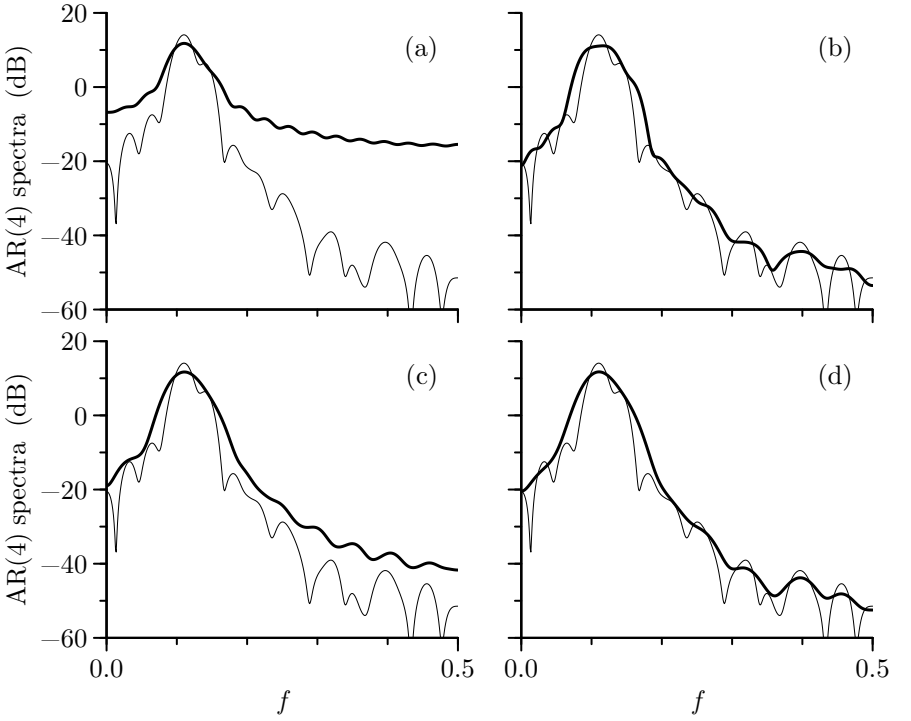

Figure 289 Direct spectral estimate for the first 64 values of the AR(4) time series of Figure 34(e) (thin curves in each plot), along with lag window estimates based upon (a) Bartlett, (b) Daniell, (c) Parzen and (d) Gaussian lag windows (thick curves). The direct spectral estimate is based upon the Hanning data taper (Equation (189a) with $p = 1$ and $N = 64$). The Bartlett lag window is defined by Equation (269a) with $m = 30$ (Figure 270(a) shows this window); the Daniell, by Equation (273a) with $m = 20$ and $N = 64$ (Figure 272b(a)); the Parzen, by Equation (275a) with $m = 37$ (Figure 276(a)); and the Gaussian, by Equation (276) with $m = 16$ (Figure 278(a)). The lag window parameter m is set in each case so that the smoothing window bandwidth is the same ($B_W \doteq 0.05$).

79 dB down and 2 dB per octave; and arguably the Daniell and Bartlett–Priestley smoothing windows have no sidelobes, so our first choice would be one of these. Based upon how far down their first sidelobes are from their central lobes, our choices amongst the four remaining windows would be, from first to last, the Gaussian, Parzen, Papoulis and Bartlett smoothing windows.

Figure 289 shows two examples of lag window estimates where smoothing window leakage is evident, and two for which it is not. The thin curve in each plot is a direct spectral estimate $\hat{S}^{(\mathrm{D})}(\cdot)$ based upon the Hanning data taper for the first 64 values of the AR(4) time series displayed in Figure 34(e). The thick curves show corresponding lag window estimates based upon (a) Bartlett, (b) Daniell, (c) Parzen and (d) Gaussian lag windows. The windows themselves are exactly the ones shown in, respectively, Figures 270(a), 272b(a), 276(a) and 278(a), all of which have a smoothing window bandwidth of $B_W \doteq 0.05$. The lag window estimates $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ should ideally be a smoothed version of $\hat{S}^{(\mathrm{D})}(\cdot)$, but $\hat{S}_m^{(\mathrm{LW})}(f)$ is systematically above $\hat{S}^{(\mathrm{D})}(f)$ at high frequencies for the Bartlett and Parzen lag windows. This deviation is due to smoothing window leakage and is more prominent in the Bartlett estimate than in the Parzen estimate, a pattern consistent with the sidelobes of their smoothing windows. By contrast, the Daniell and Gaussian estimates track $\hat{S}^{(\mathrm{D})}(\cdot)$ nicely at high frequencies, with the Gaussian estimate being a little smoother in appearance. (Figure 319(b) later in this chapter has an additional illustration of smoothing window leakage.)

There are valid reasons for *not* making smoothing window leakage our only criterion.

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

First, the degree of distortion due to this leakage can be controlled somewhat by the smoothing parameter m (see Figure 292). Second, smoothing window leakage is relevant largely because it affects the bias in $\hat{S}_m^{(\mathrm{LW})}(f)$ due to the sidelobes of the smoothing window. Note that this source of bias is *not* the same as the smoothing window bias of Equation (256d), which reflects the bias introduced by the central lobe of the smoothing window (so-called "local" bias, which results in a loss of resolution). The very rectangular Daniell window is particularly susceptible to local bias (Walden and White, 1984). Smoothing window leakage does not take into account either the variance of $\hat{S}_m^{(\mathrm{LW})}(f)$ or local bias. Third, this leakage is a significant problem only when $\hat{S}^{(\mathrm{D})}(\cdot)$ has a large dynamic range. If this is not the case, the sidelobes of $W_m(\cdot)$ have little influence on $\hat{S}_m^{(\mathrm{LW})}(\cdot)$, and hence smoothing window leakage is not relevant.

A second consideration is that $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ should be a smoothed version of $\hat{S}^{(\mathrm{D})}(\cdot)$. In Section 3.6 we noted that, when smoothing a function, a Gaussian kernel is preferable to a rectangular one from the point of view of smoothness because the former has a monotone attenuation property that the latter lacks. This desirable property translates here into a requirement that the transfer function for a smoothing window decrease monotonically in magnitude. Now the "transfer function" here is in fact the inverse Fourier transform of the smoothing window, which is just the sequence $\{w_{m,\tau}\}$ (see Exercise [5.17]). By this smoothness requirement, we would prefer smoothing windows whose lag windows decay monotonically to 0. This is true for the Bartlett, Parzen, Papoulis and Gaussian lag windows (see plot (a) of Figures 270, 276, 279 and 278), but *not* for the Daniell and Bartlett–Priestley lag windows (plot (a) of Figures 272b and 274b) – this is not surprising in the Daniell case because in fact the Daniell smoothing window is rectangular. A practical interpretation is that use of the Daniell or Bartlett–Priestley windows can lead to undesirable ripples in $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ that are not present in $\hat{S}^{(\mathrm{D})}(\cdot)$. A close examination of the Daniell lag window estimate in Figure 289(b) – aided by a comparison with the ripple-free Gaussian estimate below it – reveals a hint of these ripples (the forthcoming Figure 319 has a better example).

If smoothing window leakage does not come into play for the direct spectral estimate to be smoothed, our preference is for either the Parzen or Papoulis smoothing window from amongst the five standard windows we have discussed; if leakage proves significant for these windows, our choice from amongst these five is then either the Daniell or the Bartlett–Priestley smoothing window, but the Gaussian window can outperform both of these windows by providing a ripple-free estimate with comparable protection against leakage. However, except in cases where the dynamic range of $\hat{S}^{(\mathrm{D})}(\cdot)$ is large enough to make smoothing window leakage an issue, the following quote from Jenkins and Watts (1968, p. 273) is germane: "... the important question in empirical spectral analysis is the choice of [smoothing window] bandwidth and *not* the choice of [smoothing] window."

Comments and Extensions to Section 7.6

[1] In addition to the five standard lag windows and the Gaussian window, we have also considered three additional lag windows in previous C&Es, namely, the 100% cosine, autocorrelated 100% cosine and Slepian lag windows (see Figures 284, 285 and 286). When applied to the Hanning-based direct spectral estimate shown in Figure 289, autocorrelated 100% cosine and Slepian lag window estimates with parameters chosen such that $B_W \doteq 0.05$ are visually identical to the Gaussian estimate shown in Figure 289(d). Figure 291 shows the corresponding 100% cosine lag window estimate, which suffers from smoothing window leakage and also is negative over certain frequency intervals, rendering this lag window the least attractive of the three additional windows. (Exercise [7.14] invites the reader to compute the expected value of the five lag window estimators shown in Figures 289 and 291, with the take-home message being that the smoothing window leakage evident in these figures is quite likely to be repeated for other AR(4) realizations.)

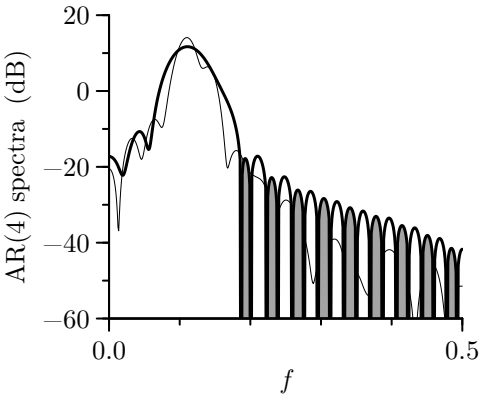

Figure 291 As in Figure 289, but now showing a lag window estimate based upon the 100% cosine lag window as defined by Equation (283) with $p = 1$, $m = 27$ and $M = 53$, yielding a smoothing window bandwidth of $B_W \doteq 0.05$ (Figure 284(a) shows this window). This estimate is negative at some frequency intervals, over which $10 \log_{10}(|\hat{S}_m^{(\mathrm{LW})}(f)|)$ is shown with shading underneath.

7.7 Choice of Lag Window Parameter

In the previous section, we argued that any one of several specific lag windows would be a reasonable choice. Unfortunately, the choice of the window parameter m is *not* so easy: a sensible choice of m depends in part upon the shape of the unknown SDF we want to estimate.

To illustrate the importance in selecting m properly, Figure 292 plots

$$\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} W_m(f - f') S(f') \, \mathrm{d}f'$$

versus f for the case where $S(\cdot)$ is the SDF of the AR(4) process $\{X_t\}$ of Equation (35a); $W_m(\cdot)$ is Parzen's smoothing window; and $m = 32$, 64, 128 and 256, with corresponding smoothing window bandwidths $B_W \doteq 0.058$, 0.029, 0.014 and 0.007. The true SDF for this AR(4) process (indicated by the thick curves in Figure 292) has two sharp peaks close to each other. For both $m = 32$ and 64, the amount of smoothing done by Parzen's lag window spectral estimator is such that the expected value of the estimator (given by the thin curves) does not accurately represent the true SDF in the region of the two peaks. The amount of smoothing is related to the width of the central lobe of Parzen's smoothing window (shown – centered about $f = 1/10$ – by the thin curves in the lower left-hand part of each plot). As we increase m to 128 or 256, the width of this central lobe decreases enough so that the expected value of the estimator more accurately represents the true SDF – in fact, the agreement is so good in the $m = 256$ case that the thick and thin curves are indistinguishable in the plot. (Note that this example considers the bias due *only* to the smoothing window. As we have seen in Figure 193, the bias in the direct spectral estimator itself can be a significant factor for small N even if we are careful to use a good data taper.)

To make a choice of m that yields an approximately unbiased estimator, we must have some idea of the "size" of the important features in $S(\cdot)$ – the very SDF we are trying to estimate! In many situations this is simply not known a priori. For example, spectral analysis is often used as an exploratory data analysis tool, in which case little is known about $S(\cdot)$ in advance. There are cases, however, where it is possible to make an intelligent guess about the typical shape of $S(\cdot)$ either from physical arguments or from analyses of previous data of a similar nature. As an example, let us assume that the typical SDF to be expected is the same as the SDF for the AR(2) process of Equation (34). This SDF is shown in all four plots of

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

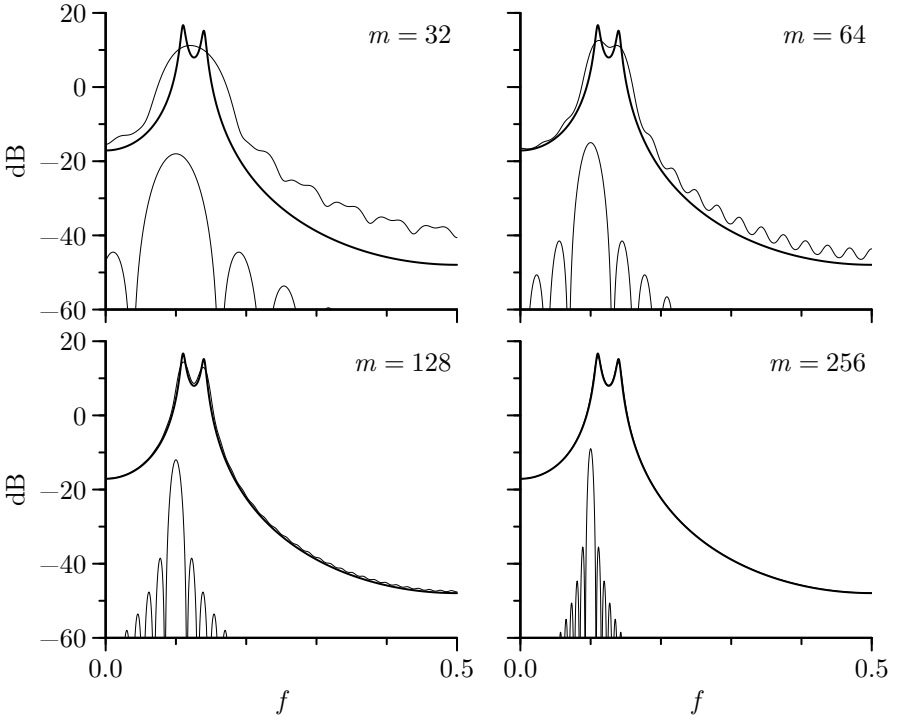

Figure 292 Effect of window parameter m on the expected value of an unbiased direct spectral estimator that has been smoothed using Parzen's lag window. The thick curve in each plot is the SDF for the AR(4) process of Equation (35a); the thin curves that follow the thick curves to varying degrees show the component of $E\{\hat{S}_m^{(\mathrm{LW})}(\cdot)\}$ solely due to Parzen's smoothing window for m = 32, 64, 128 and 256; and the thin curves in the lower left-hand corners show the shape of the smoothing window itself.

the forthcoming Figure 294 as a thin dark curve. We can define the *spectral bandwidth* B_S as roughly the width of the *smallest* feature of interest in the projected SDF $S(\cdot)$. Here we take this feature to be the broad peak, the maximum for which is at $f_0 \doteq 0.155$. We can measure the peak's width in terms of half-power points; i.e., those points $f_1 < f_0$ and $f_2 > f_0$ such that

$$S(f_1) = S(f_2) = \frac{S(f_0)}{2} \text{ and hence } B_S = f_2 - f_1.$$

Here $S(f_0) \doteq 7.5\,\mathrm{dB}$, and, since the frequencies $f_1 \doteq 0.085$ and $f_2 \doteq 0.208$ yield the required conditions, we have $B_S \doteq 0.123$. As a second example, suppose that the expected SDF is the AR(4) SDF shown in Figure 292, which has two prominent peaks of comparable width. If we take the one centered about $f_0 \doteq 0.110$ to be the smallest feature of interest, we find $f_1 \doteq 0.107$ and $f_2 \doteq 0.114$, yielding $B_S \doteq 0.007$. For either example, we can then use either Table 279 or Equation (251e) to find a value of m such that the smoothing window bandwidth B_W is acceptable (note that the bandwidths B_S and B_W have the same units as f, namely, cycles per unit time). Priestley (1981) recommends setting m such that $B_W = B_S/2$ so that the smoothing window does not smooth out any important features in the SDF. Assuming use of the Parzen window for our two examples, Table 279 says that

$$m = \frac{1.85}{B_W} = \frac{3.6}{B_S} \doteq 30 \text{ for the AR(2) case, and } m = 565 \text{ for the AR(4) case.}$$

This advice, however, must be used with caution, since the amount of data N that is available can be such that the variance of $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ is unacceptably large if we were to use the value

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

of m chosen by this bandwidth matching criterion. (Priestley has an extensive discussion on selecting m such that a trade-off is made between the bias – due to smoothing window leakage – and the variance of a lag window spectral estimator.)

In the more common case where there is little prior knowledge of the true SDF, we can entertain choosing the window parameter m using the objective criteria discussed in Sections 7.8, 7.9 and 7.10. The following three subjective methods also merit discussion (Priestley, 1981, pp. 539–42).

The first method is known as *window closing*. The idea is to compute a sequence of different SDF estimates for the same time series using different smoothing window bandwidths that range from large to small. For large values of B_W, the estimates will look smooth, but, as B_W decreases, the estimates will progressively exhibit more detail until a point is reached where the estimates are more "erratic" in form. Based upon a subjective evaluation of these estimates, we can hopefully pick a value of m that is appropriate in the sense that the resulting estimate is neither too smooth nor too erratic. Since $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ is supposed to be a smoothed version of $\hat{S}^{(\mathrm{D})}(\cdot)$, our reference point in these comparisons should be how well the former captures – and points out – the important features of the latter. These visual comparisons can reject some estimates as obviously being "too smooth" or "too erratic."

As an example, let us consider four Parzen lag window estimates for the AR(2) time series shown in Figure 34(a). There is no need for tapering since the SDF is simple enough that the periodogram is essentially bias free when $N = 1024$ (cf. Figure 177). Figure 294 shows $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ with $m = 4$, 16, 29 and 64 (thick dark curves) along with the periodogram (rough-looking light curves) and the true SDF (thin dark curves). The width of the crisscross in each plot depicts the bandwidth measure $B_{\mathcal{U}}$, while its height shows the length of a 95% CI for $10 \log_{10}(S(f))$ based upon $10 \log_{10}(\hat{S}_m^{(\mathrm{LW})}(f))$. A comparison of the four $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ with the true SDF shows the importance of properly choosing the smoothing parameter m: whereas the $m = 4$ estimate is oversmoothed and the $m = 64$ estimate is undersmoothed, the $m = 16$ and $m = 29$ estimates agree with the true SDF much better. Note that a 95% CI for $10 \log_{10}(S(f))$ based upon the oversmoothed $m = 4$ estimate would in fact *fail* to include the true value of the SDF over the majority of the frequencies between 0 and $1/2$, whereas the opposite is true for CIs based upon the other three estimates. We can quantify how well a particular $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ agrees with the true SDF $S(\cdot)$ by computing the mean square error

$$\mathrm{MSE}(m) = \frac{1}{N+1} \sum_{k=0}^{N} \left(\hat{S}_m^{(\mathrm{LW})}(\tilde{f}_k) - S(\tilde{f}_k) \right)^2, \quad \text{where} \quad \tilde{f}_k = k/(2N). \tag{293}$$

The above gives 1.309, 0.105, 0.049 and 0.110 for $m = 4$, 16, 29 and 64. In fact, the choice with the smallest MSE(m) amongst all possible m for this particular AR(2) time series is the $m = 29$ lag window estimate, which fortuitously is quite close to the setting $m = 30$ dictated by the $B_W = B_S/2$ criterion. Here the subjective criterion of window closing would likely rule out the $m = 4$ choice because $\hat{S}_m^{(\mathrm{LW})}(f)$ is systematically above $\hat{S}^{(\mathrm{D})}(f)$ at high frequencies; however, it would be difficult to choose visually between the $m = 16$ and $m = 29$ estimates even though MSE(16) is approximately twice as large as MSE(29). We might also rule out the $m = 64$ choice as being too erratic even though the visually more pleasing $m = 16$ choice has approximately the same MSE(m). On the other hand, the maximum difference between any two of the $m = 16$, 29 and 64 estimates is less than 2.2 dB in magnitude (a factor of 1.7), so these estimates all agree to well within the same order of magnitude.

As a second example, let us consider four Parzen lag window estimates for the AR(4) time series of length $N = 1024$ shown in Figure 34(e). Figures 182(e) and 187(e) show the periodogram and a direct spectral estimate based upon the Hanning data taper for this series. It is evident that the periodogram suffers from leakage, so we form lag window estimates

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

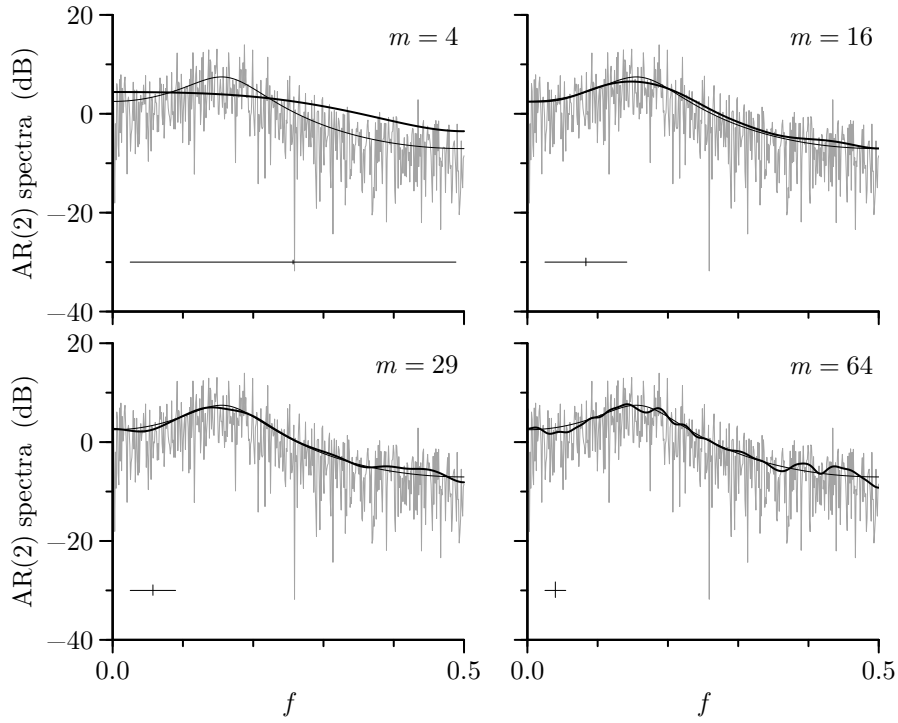

Figure 294 Lag window spectral estimates $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ for AR(2) series of Figure 34(a) using a Parzen smoothing window with m = 4, 16, 29 and 64 (thick dark curves). The thin dark curve on each plot is the true AR(2) SDF, while the rough-looking light curve is the periodogram at the Fourier frequencies. The width of the crisscross emanating from the lower left-hand corner of each plot gives the bandwidth measure $B_{\mathcal{U}}$ as per Equation (256a), while its height gives the length of a 95% CI for 10 $\log_{10}(S(f))$ as per Equation (266a).

using the Hanning-based $\hat{S}^{(\mathrm{D})}(\cdot)$. Four such estimates with m set to 64, 128, 179 and 565 are shown in Figure 295, along with the underlying direct spectral estimate and the true SDF. Here we can rule out the $m = 64$ estimate because it oversmooths the twin peaks and is systematically above the direct spectral estimate at high frequencies, which is indicative of smoothing window leakage (the ripples in $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ at those frequencies are another indication – cf. the $m = 64$ case in Figure 292). The $m = 565$ estimate is much more erratic looking than the remaining two estimates, of which we might prefer the $m = 128$ estimate over $m = 179$ because the latter is a bit bumpier looking. Calculations indicate that

$$\mathrm{MSE}(m) = \begin{cases} 10.241, & m = 64; \\ 4.886, & m = 128; \\ 4.228, & m = 179; \\ 10.329, & m = 565. \end{cases}$$

Visibly $m = 128$ might seem to be a better choice than $m = 179$, but it actually isn't. The subjectively chosen $m = 128$ estimate is, however, not too far off the mark since MSE(128) is only about 15% larger than MSE(179), which is the minimizer of MSE(m) over all possible m for this particular AR(4) time series. (It is of interest to note that the $m = 64$ and $m = 565$ estimates have approximately the same MSE, with the former being bias dominated, and the latter, variance dominated. The setting $m = 565$ comes from the $B_W = B_S/2$ criterion, which is geared toward eliminating bias in $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ and does not explicitly take into account its variance.)

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

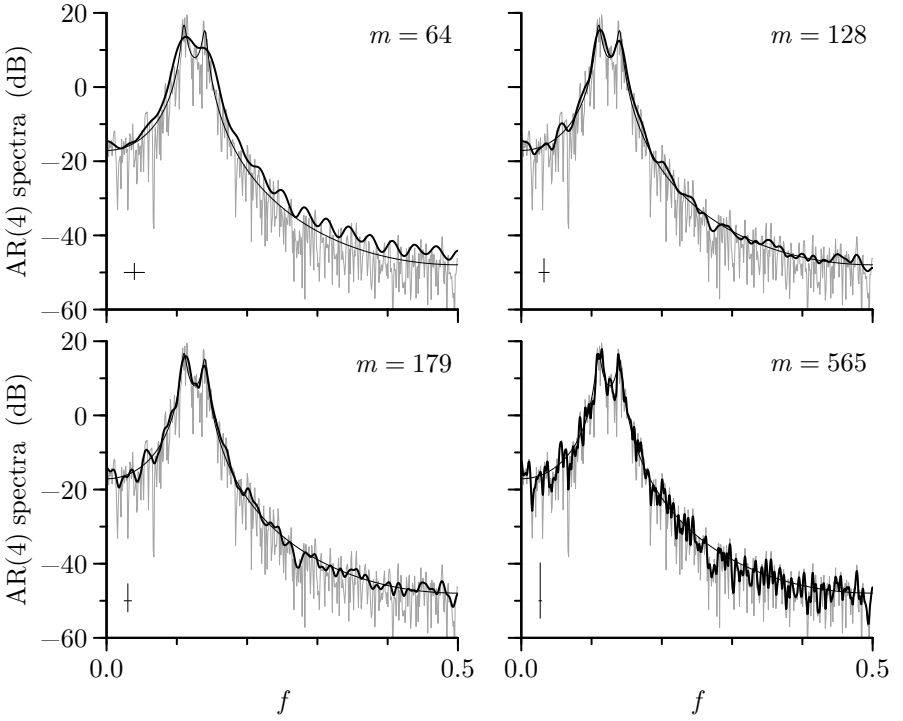

Figure 295 As in Figure 294, but now showing lag window spectral estimates $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ for AR(4) series of Figure 34(e) using a Parzen smoothing window with m = 64, 128, 179 and 565 (thick dark curves). The thin dark curve in each plot is now the true AR(4) SDF, while the rough-looking light curve is the Hanning-based direct spectral estimate.

The second subjective method is based upon the sample ACVS. If our spectral estimator is of the truncation type (such as the Bartlett, Parzen or Papoulis estimators), we can argue that we should choose m so that $s_\tau \approx 0$ for all $|\tau| \geq m$. Thus we could plot $\hat{s}_\tau^{(\mathrm{P})}$ versus τ and pick m as that point starting at which the sample ACVS shows only small fluctuations around 0. Although crude, this method has evidently been quite effective in a number of problems and is arguably preferable to choosing a completely arbitrary value for m. There are two dangers with this approach. First, sampling fluctuations can cause the sample ACVS to deviate substantially from the true ACVS. Second, if $S(\cdot)$ contains a large peak with a wide bandwidth and a small peak with a narrow bandwidth, the effect of the small peak will not be apparent at all in the sample ACVS (see Figure 125 supports this assertion). The value of m selected by this method will reflect only the bandwidth of the large peak. We can thus miss some important features in the SDF by using this method.

As an example, let us reconsider Figure 248(a), which shows $\{\hat{s}_\tau^{(\mathrm{P})}\}$ for the AR(2) time series behind the four lag window estimates shown in Figure 248. Here $\hat{s}_\tau^{(\mathrm{P})}$ over the range $\tau = 5, \ldots, 128$ fluctuates about zero with no particularly dominant value. We might be tempted to pick $m = 5$ by the subjective sample ACVS method. An $m = 5$ Parzen lag window estimate is only a marginal improvement over the $m = 4$ estimate shown in Figure 248. This subjective method does not serve us as well as window closing for this particular example.

The final subjective method is simply to let m be some fixed proportion of N, say 20% or 30%, or to let $m = \sqrt{N}$. Such recommendations are fundamentally unsound since they do not take into account the underlying process that generated the time series. Although we might have no prior knowledge of the spectral bandwidth, it is desirable to choose m based

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

upon some properties of the process. This is precisely the aim of the previous two subjective techniques.

Comments and Extensions to Section 7.7

[1] While B_W is a measure of the width of the central lobe of $W_m(\cdot)$, the effective bandwidth $B_{\mathcal{U}}$ measures the same for the spectral window associated with a lag window estimator. Hence, while B_W takes into account just the lag window, $B_{\mathcal{U}}$ takes into account both the lag window and the data taper. In situations where we know the spectral bandwidth B_S, a possible refinement to selecting the lag window parameter m via the relationship $B_W = B_S/2$ is to use $B_{\mathcal{U}} = B_S/2$ instead. If anything, this refinement will call for *less* smoothing since $B_{\mathcal{U}} \geq B_W$ always (see Exercise [257]). A disadvantage of this new approach is the absence of simple formulae that link $B_{\mathcal{U}}$ to m for various combinations of lag windows and data tapers, i.e., formulae analogous to $B_W = 1.85/(m\,\Delta_t)$ for the Parzen lag window (see Table 279). Nonetheless, for a given data taper and lag window, we can use Equation (256a) to compute $B_{\mathcal{U}}$ for various choices of m and thus to set m such that $B_{\mathcal{U}}$ is as close as possible to $B_S/2$.

As examples, let us reconsider the AR(2) and AR(4) processes. Suppose we want to form a lag window estimate for the former based upon a sample of size $N = 1024$, the default (rectangular) data taper and the Parzen lag window. The $B_{\mathcal{U}} = B_S/2$ criterion leads us to select $m = 30$, i.e., the same value we got using $B_W = B_S/2$. Considering now the AR(4) process, suppose we again set $N = 1024$ and select the Parzen lag window, but now we use the Hanning data taper. Here the $B_{\mathcal{U}} = B_S/2$ criterion sets $m = 714$, whereas $B_W = B_S/2$ leads to $m = 565$. When the choice $m = 714$ is used on the AR(4) time series behind the four Parzen lag window estimates displayed in Figure 295, the resulting estimate (not shown) is even bumpier looking than the $m = 565$ case. Here MSE(714) $\doteq$ 13.156, in comparison to MSE(565) $\doteq$ 10.329 and MSE(179) $\doteq$ 4.228 (the minimum over all possible choices of m). The criteria $B_W = B_S/2$ and $B_{\mathcal{U}} = B_S/2$ are both geared toward eliminating bias, but do not take into consideration variance. Hence, if we use either criteria for setting m, there is no guarantee of getting an estimate with a sensible trade-off between bias and variance.

[2] The MSE of Equation (293) is only one of many possibilities for quantifying how well a particular SDF estimate matches up with the true SDF. Two variations on the MSE appearing in the literature are the normalized mean square error

$$\text{NMSE}(m) = \frac{1}{N+1} \sum_{k=0}^{N} \left(\frac{\hat{S}_m^{(\text{LW})}(\tilde{f}_k) - S(\tilde{f}_k)}{S(\tilde{f}_k)} \right)^2 \tag{296a}$$

and the mean square log error

$$\text{MSLE}(m) = \frac{1}{N+1} \sum_{k=0}^{N} \left(\log\left(\hat{S}_m^{(\text{LW})}(\tilde{f}_k)\right) - \log\left(S(\tilde{f}_k)\right) \right)^2 \tag{296b}$$

(see, e.g., Lee, 1997, who points out that both are problematic when $S(\tilde{f}_k) = 0$). If we momentarily assume that $\hat{S}_m^{(\text{LW})}(\tilde{f}_k) \stackrel{\text{d}}{=} S(\tilde{f}_k)\chi^2_\nu/\nu$ for all k (a dicey assumption for $\tilde{f}_k$ close to 0 or $f_{\mathcal{N}}$), then

$$E\{\text{NMSE}(m)\} = \frac{2}{\nu} \quad \text{and} \quad E\{\text{MSLE}(m)\} = \left[\psi\left(\frac{\nu}{2}\right) - \log\left(\frac{\nu}{2}\right)\right]^2 + \psi'\left(\frac{\nu}{2}\right), \tag{296c}$$

where $\psi(\cdot)$ and $\psi'(\cdot)$ are the digamma and trigamma functions. Neither of these expected values involves the true SDF, whereas

$$E\{\text{MSE}(m)\} = \frac{2}{(N+1)\nu} \sum_{k=0}^{N} S^2(\tilde{f}_k) \tag{296d}$$

does (Exercise [7.15] is to verify Equations (296c) and (296d)). While NMSE and MSLE do not depend preferentially on large values of the SDF, MSE does. Returning to the AR(2) example, for which $m = 29$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

minimized MSE(m), we find that NMSE(m) and MSLE(m) are both minimized when $m = 22$ (the estimate $\hat{S}_{22}^{(\mathrm{LW})}(\cdot)$ is shown in the right-hand plot of the forthcoming Figure 310 and is a compromise between $\hat{S}_{16}^{(\mathrm{LW})}(\cdot)$ and $\hat{S}_{29}^{(\mathrm{LW})}(\cdot)$ shown in Figure 294; its maximum deviations from these are, respectively, 0.6 dB and 0.5 dB). For the AR(4) example, the minimum values are MSE(179), NMSE(154) and MSLE(130) (the estimate $\hat{S}_{154}^{(\mathrm{LW})}(\cdot)$ is in appearance halfway between $\hat{S}_{128}^{(\mathrm{LW})}(\cdot)$ and $\hat{S}_{179}^{(\mathrm{LW})}(\cdot)$ in Figure 295, with maximum deviations less than 1.7 dB and 1.2 dB; the $\hat{S}_{130}^{(\mathrm{LW})}(\cdot)$ estimate is always within 0.4 dB of $\hat{S}_{128}^{(\mathrm{LW})}(\cdot)$, which visual inspection of Figure 295 arguably suggests to be the most appropriate amongst the four estimates displayed there).

The *Kullback–Leibler (KL) discrepancy* between two PDFs motivates another measure of interest (see, e.g., Ombao et al., 2001). If $p(\cdot)$ and $q(\cdot)$ are the PDFs for two nonnegative RVs, the KL discrepancy takes the form

$$d(p(\cdot), q(\cdot)) \stackrel{\text{def}}{=} \int_0^\infty p(x) \log\left(\frac{p(x)}{q(x)}\right) \mathrm{d}x \tag{297a}$$

(Kullback and Leibler, 1951). The above is minimized when $p(\cdot)$ and $q(\cdot)$ are the same. We note that $d(p(\cdot), q(\cdot)) \neq d(q(\cdot), p(\cdot))$ in general, so the discrepancy depends on the ordering. The PDF corresponding to the assumption $\hat{S}_m^{(\mathrm{LW})}(\tilde{f}_k) \stackrel{\mathrm{d}}{=} S(f)\chi_\nu^2/\nu$ is

$$f(x) = \left(\frac{\nu}{2S(\tilde{f}_k)}\right)^{\nu/2} \frac{x^{\nu/2-1}\mathrm{e}^{-\nu x/(2S(\tilde{f}_k))}}{\Gamma\left(\nu/2\right)}, \quad 0 \le x < \infty, \tag{297b}$$

which is the same as the PDF for a gamma RV with shape parameter $\nu/2$ and rate parameter $\nu/(2S(\tilde{f}_k))$. Setting $p(\cdot)$ and $q(\cdot)$ in Equation (297a) to two gamma PDFs – with ν taken to be the same but allowing the SDFs to be distinct – inspires the following measure:

$$\mathrm{KL}(m) = \frac{1}{N+1} \sum_{k=0}^{N} \left[\frac{S(\tilde{f}_k)}{\hat{S}_m^{(\mathrm{LW})}(\tilde{f}_k)} - \log\left(\frac{S(\tilde{f}_k)}{\hat{S}_m^{(\mathrm{LW})}(\tilde{f}_k)}\right) - 1 \right] \tag{297c}$$

(see Exercise [7.16a] for details). The fact that $x - \log(x) - 1 \ge 0$ for all $x > 0$ tells us that the above must be nonnegative (here we ignore the special cases $S(\tilde{f}_k) = 0$ and $\hat{S}_m^{(\mathrm{LW})}(\tilde{f}_k) = 0$); in addition, it attains its minimum value of zero when each $\hat{S}_m^{(\mathrm{LW})}(\tilde{f}_k)$ is identical to $S(f_k)$. Exercise [7.16b] is to show that

$$E\{\mathrm{KL}(m)\} = \frac{\nu}{\nu-2} - \log\left(\frac{\nu}{2}\right) + \psi\left(\frac{\nu}{2}\right) - 1 \tag{297d}$$

under the previous assumption $\hat{S}_m^{(\mathrm{LW})}(\tilde{f}_k) \stackrel{\mathrm{d}}{=} S(\tilde{f}_k)\chi_\nu^2/\nu$, but now with the additional constraint that $\nu > 2$. The above is similar to $E\{\mathrm{NMSE}(m)\}$ and $E\{\mathrm{MSLE}(m)\}$ in its lack of dependence on the true SDF. For the AR(2) example, KL(m) is minimized when $m = 22$, in agreement with the minimizers for both NMSE(m) and MSLE(m). For the AR(4) example, the minimizer is $m = 129$, which is in close agreement with $m = 130$, the minimizer for MSLE(m) (the $\hat{S}_{129}^{(\mathrm{LW})}(\cdot)$ estimate is always within 0.2 dB of $\hat{S}_{128}^{(\mathrm{LW})}(\cdot)$ displayed in Figure 295).

7.8 Estimation of Spectral Bandwidth

We have looked at the bandwidth B_W of a smoothing window and how to compute it using Equation (251e). To compute the spectral bandwidth B_S, we must have good knowledge of the SDF we are trying to estimate. Can we estimate the spectral bandwidth directly from a time series? If this were possible, it would clearly be beneficial since we could use the estimated spectral bandwidth to guide the setting of B_W. The following argument shows that, if we assume the time series is drawn from a Gaussian stationary process with a dominantly unimodal SDF, it is possible to produce a satisfactory estimator for the spectral bandwidth (for details, see Walden and White, 1990).

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

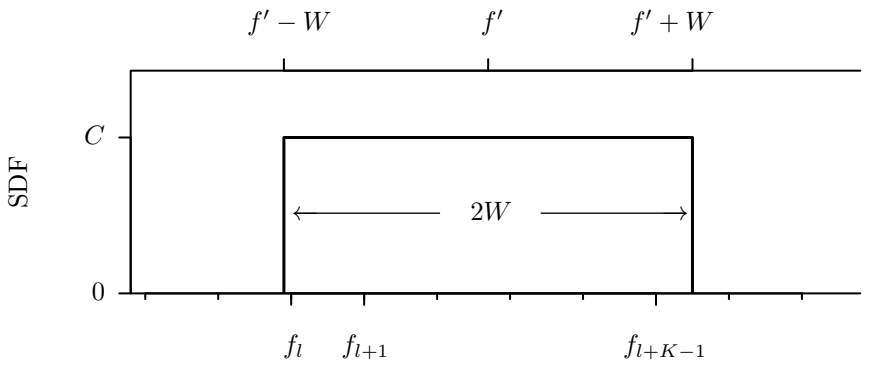

Figure 298 SDF for ideal band-pass process with bandwidth of $2W$ centered about frequency f'. The tick marks on the bottom horizontal axis show the locations of the Fourier frequencies, three of which are labelled.

The standard estimator of the variance s_0 of a time series drawn from a stationary process with unknown mean is

$$\hat{s}_0^{(\mathrm{P})} = \frac{1}{N} \sum_{t=0}^{N-1} \left(X_t - \overline{X} \right)^2 .$$

Assume for convenience that the process mean is known to be 0 so that we can take

$$\hat{s}_0^{(\mathrm{P})} = \frac{1}{N} \sum_{t=0}^{N-1} X_t^2$$

(this allows us to simplify the calculations in the rest of this section considerably – the resulting estimator of the spectral bandwidth is still useful even if the process mean must be estimated). The expected value of $\hat{s}_0^{(\mathrm{P})}$ is

$$E\{\hat{s}_0^{(\mathrm{P})}\} = \frac{1}{N} \sum_{t=0}^{N-1} E\{X_t^2\} = s_0.$$

▷ **Exercise [298]** By evoking the Isserlis theorem (Equation (30)), show that

$$\operatorname{var}\{\hat{s}_0^{(\mathrm{P})}\} = \frac{2}{N} \sum_{\tau=-(N-1)}^{N-1} \left(1 - \frac{|\tau|}{N} \right) s_\tau^2 . \qquad (298a)$$ ◁

In Section 7.4 the distribution of the spectral estimator $\hat{S}_m^{(\mathrm{LW})}(f)$ was approximated by that of a χ_ν^2 RV times a constant a. If we make the same approximation here, namely,

$$\hat{s}_0^{(\mathrm{P})} \stackrel{\mathrm{d}}{=} b \chi_\eta^2 \qquad (298b)$$

(Rice, 1945), we obtain

$$\eta = \frac{2 \left(E\{\hat{s}_0^{(\mathrm{P})}\} \right)^2}{\operatorname{var}\{\hat{s}_0^{(\mathrm{P})}\}} = \frac{N s_0^2}{\sum_{\tau=-(N-1)}^{N-1} (1 - |\tau|/N)\, s_\tau^2} \qquad (298c)$$

as an expression for the *degrees of freedom in a time series* of length N (analogous to the degrees of freedom in a spectral estimator – see Equation (264a)). Note that both concepts involve quadratic forms (see Chapter 8), and some researchers make this point explicit by calling η the number of degrees of freedom of order 2 (Kikkawa and Ishida, 1988). Exercise [9.2e] uses a first-order autoregressive process to explore how η depends on $\{s_\tau\}$.

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

We now need to relate η somehow to the width of the dominant hump of a unimodal SDF. Suppose for the moment that $\{X_t\}$ is an ideal band-pass process with an SDF defined over $f \in [-f_{\mathcal{N}}, f_{\mathcal{N}}]$ by

$$S(f) = \begin{cases} C, & f' - W \le |f| \le f' + W; \\ 0, & \text{otherwise}, \end{cases} \tag{299}$$

where $C > 0$; $f' > 0$ is the center frequency of the pass-band; and $2W > 0$ is the corresponding bandwidth (see Figure 298). We assume that $f' - W > 0$ and $f' + W < f_{\mathcal{N}}$. For this simple example we can relate the bandwidth $2W$ to the degrees of freedom in a sample of length N from $\{X_t\}$ by the following argument. From a special case of Equation (237b) and from the fact that $\hat{S}^{(\mathrm{P})}(\cdot)$ is $2f_{\mathcal{N}}$ periodic, we can write, with $f_k = k/(N\,\Delta_{\mathrm{t}})$ (i.e., the Fourier frequencies),

$$\begin{aligned} \hat{s}_0^{(\mathrm{P})} &= \frac{1}{N\,\Delta_{\mathrm{t}}} \sum_{k=0}^{N-1} \hat{S}^{(\mathrm{P})}(f_k) \\ &= \frac{1}{N\,\Delta_{\mathrm{t}}} \sum_{k=-\lfloor (N-1)/2 \rfloor}^{\lfloor N/2 \rfloor} \hat{S}^{(\mathrm{P})}(f_k) \approx \frac{1}{N\,\Delta_{\mathrm{t}}} \sum_{f'-W \le |f_k| \le f'+W} \hat{S}^{(\mathrm{P})}(f_k) \end{aligned}$$

since $\hat{S}^{(\mathrm{P})}(f_k)$ should be small for those f_k such that $S(f_k) = 0$. Because $\hat{S}^{(\mathrm{P})}(\cdot)$ is symmetric about zero, we also have

$$\hat{s}_0^{(\mathrm{P})} \approx \frac{2}{N\,\Delta_{\mathrm{t}}} \sum_{f'-W \le f_k \le f'+W} \hat{S}^{(\mathrm{P})}(f_k) = \frac{2}{N\,\Delta_{\mathrm{t}}} \sum_{k=0}^{K-1} \hat{S}^{(\mathrm{P})}(f_{l+k}),$$

where f_l and K are such that $f_{l-1} < f' - W$ and $f_{l+K} > f' + W$ but $f' - W \le f_l < \cdots < f_{l+K-1} \le f' + W$. Since the spacing between adjacent Fourier frequencies is $\Delta_{\mathrm{f}} = 1/(N\,\Delta_{\mathrm{t}})$, we have $K \approx 2W/\Delta_{\mathrm{f}} = 2WN\,\Delta_{\mathrm{t}}$. From Section 6.6 we know that, for those f_{l+k} in the pass-band,

$$\hat{S}^{(\mathrm{P})}(f_{l+k}) \stackrel{\mathrm{d}}{=} \frac{S(f_{l+k})}{2} \chi_2^2 = \frac{C}{2} \chi_2^2$$

to a good approximation and that the $\hat{S}^{(\mathrm{P})}(f_{l+k})$ RVs at distinct Fourier frequencies are approximately independent of each other. Since the sum of K independent χ_2^2 RVs has a χ_{2K}^2 distribution, we can conclude that, to a good approximation,

$$\hat{s}_0^{(\mathrm{P})} \stackrel{\mathrm{d}}{=} \frac{C}{N\,\Delta_{\mathrm{t}}} \chi_{2K}^2.$$

Matching the above with Equation (298b) yields $\eta = 2K \approx 4WN\,\Delta_{\mathrm{t}}$, so the width of the pass-band $2W$ is approximately $\eta/(2N\,\Delta_{\mathrm{t}})$. For an ideal band-pass process, we can thus take the spectral bandwidth B_S to be $\eta/(2N\,\Delta_{\mathrm{t}})$.

Suppose now that $\{X_t\}$ is either an ideal low-pass or high-pass process; i.e., its SDF is given by Equation (299) with either $f' = 0$ (low-pass case) or $f' = f_{\mathcal{N}}$ (high-pass), and we assume that $W < f_{\mathcal{N}}$. The width of the passband for both processes is $2W$, the same as for an ideal band-pass process (to support this claim, recall that the SDF is an even periodic function with a period of $2f_{\mathcal{N}}$); however, the burden of Exercise [7.19] is to show that $2W$ is now approximately $\eta/(N\,\Delta_{\mathrm{t}})$. For ideal low-pass or high-pass processes, we can thus take the spectral bandwidth B_S to be $\eta/(N\,\Delta_{\mathrm{t}})$.

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

For a time series of length N drawn from a process with an SDF whose shape is dominantly unimodal and resembles that of a band-pass, low-pass or high-pass process, we can use the arguments we have just made to *define* the bandwidth of the series as

$$B_T = \frac{\eta}{(1+\delta_{\mathrm{BP}})N\,\Delta_{\mathrm{t}}} = \frac{s_0^2}{(1+\delta_{\mathrm{BP}})\,\Delta_{\mathrm{t}}\sum_{\tau=-(N-1)}^{N-1}(1-|\tau|/N)\,s_\tau^2}, \tag{300a}$$

where $\delta_{\mathrm{BP}} = 1$ for the band-pass case and $= 0$ otherwise. An obvious estimator of B_T is

$$\hat{B}_T \stackrel{\mathrm{def}}{=} \frac{(\hat{s}_0^{(\mathrm{P})})^2}{(1+\delta_{\mathrm{BP}})\,\Delta_{\mathrm{t}}\sum_{\tau=-(N-1)}^{N-1}(1-|\tau|/N)(\hat{s}_\tau^{(\mathrm{P})})^2}. \tag{300b}$$

Walden and White (1990) find that, for the band-pass case, $\hat{B}_T$ is a biased estimator of B_T, but that $5\hat{B}_T/3 - 1/(N\,\Delta_{\mathrm{t}})$ is approximately unbiased.

▷ **Exercise [300]** Using their result, show that

$$\tilde{B}_T \stackrel{\mathrm{def}}{=} \frac{5}{6}\left(\frac{(\hat{s}_0^{(\mathrm{P})})^2}{\Delta_{\mathrm{t}}\sum_{\tau=-(N-1)}^{N-1}(1-|\tau|/N)(\hat{s}_\tau^{(\mathrm{P})})^2}\right) - \frac{1}{N\,\Delta_{\mathrm{t}}} \tag{300c}$$

is an approximately unbiased estimator of B_T in the band-pass, low-pass and high-pass cases. ◁

Thus the unbiased estimator of B_T does not depend on the nature of the unimodality.

We hope it is clear that, just as the smoothing window should be unimodal for its bandwidth measure B_W to be useful, so should the SDF for B_T to be useful. Kikkawa and Ishida (1988) say similarly that second-order parameters (such as η) "... are good measures for both low-pass and band-pass processes." Distinct bimodality in a SDF will lead to B_S and B_T being very different since B_S will be a measure of the width of the smaller mode, while B_T will be an average over the entire SDF (see Exercise [7.20]). If B_T is meaningful and if we replace B_S by $\tilde{B}_T$ and adopt the recommendation $B_W = \tilde{B}_T/2$ as before, the practical results are generally quite satisfactory; however, as we emphasized in the previous section, the $B_W = B_S/2$ criterion for setting the smoothing window bandwidth is geared toward eliminating bias, but does not pay attention to the resulting variance of the lag window estimator – the same comment holds for the $B_W = \tilde{B}_T/2$ criterion.

As a first example, consider the AR(2) process of Equation (34), which has an SDF that is dominantly unimodal and has a shape reminiscent of a band-pass process (this SDF is shown as a thin dark curve in each of the plots in Figure 294). Using the AR(2) series of Figure 34(a), we obtain $\tilde{B}_T \doteq 0.248$. Setting $B_W = 1.85/m = \tilde{B}_T/2$ yields $m = 15$ for the Parzen lag window. This value produces a lag window estimate that is almost identical to the $m = 16$ case shown in Figure 294 and hence is quite acceptable. The three AR(2) series in Figures 34(b), (c) and (d) lead to comparable results, namely, $\tilde{B}_T \doteq 0.247$, 0.267 and 0.265, resulting in $m = 15$, 14 and 14. For comparison, Exercise [7.20] says that $B_T \to 0.265$ as $N \to \infty$, yielding $m = 14$, while the $B_W = B_S/2$ criterion recommends $m = 30$ based on $B_S \doteq 0.123$.

For a second example, let us consider the AR(4) process of Equation (35a). The thin dark curve in Figure 301 shows that its SDF is *not* dominantly unimodal because of the prominent twin peaks, but let's consider using $\tilde{B}_T$ anyway. For the AR(4) series of Figure 34(e), we get $\tilde{B}_T \doteq 0.036$, so equating $B_W = 1.85/m$ to $\tilde{B}_T/2$ yields $m = 102$. The corresponding Parzen spectral estimate is shown in Figure 301 and – except for the fact that the twin peaks

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

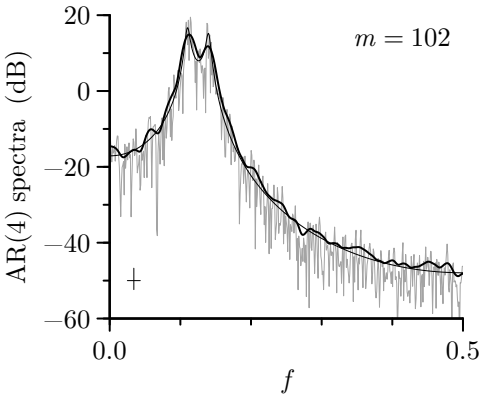

Figure 301 As in Figure 295, but with lag window spectral estimate for AR(4) series of Figure 34(e) using a Parzen smoothing window with m = 102. This choice comes from setting the smoothing window bandwidth B_W to be half of the estimated time series bandwidth measure $\tilde{B}_T$.

are slightly blurred out in comparison to the $m = 128$ and $m = 179$ estimates shown in Figure 295 – is quite acceptable. The AR(4) series in Figures 34(f), (g) and (h) give $\tilde{B}_T \doteq 0.030$, 0.043 and 0.034, resulting in $m = 121$, 85 and 107. Exercise [7.20] says that here $B_T \to 0.033$ as $N \to \infty$, which corresponds to $m = 111$. Recall that the $B_W = B_S/2$ criterion (based on the higher of the twin peaks) yields $m = 565$ from $B_S \doteq 0.007$, which is more than four times smaller than the large sample B_T. We can attribute this discrepancy to lack of unimodality in this example.

7.9 Automatic Smoothing of Log Spectral Estimators

In the previous sections we have considered smoothing a direct spectral estimator by convolving it with a smoothing window. We present here an alternative approach in which we smooth $\log(\hat{S}^{(\mathrm{D})}(\cdot))$ instead of $\hat{S}^{(\mathrm{D})}(\cdot)$. This approach leads to an objective procedure for determining the amount of smoothing that is optimal in a certain mean square sense. For further details, see Wahba (1980), who originally proposed this technique for smoothing the log of the periodogram, and Sjoholm (1989), who extended it to the case of all direct spectral estimators.

For a given SDF $S(\cdot)$, let us define the *log spectral density function* as

$$C(f) = \lambda \log(S(f)), \tag{301a}$$

where λ is a constant facilitating use of different logarithmic scales (for example, if we set $\lambda = 10 \log_{10}(\mathrm{e})$, then $C(\cdot)$ is expressed in decibels). We assume that $C(\cdot)$ has a Fourier series representation

$$C(f) = \Delta_{\mathrm{t}} \sum_{\tau=-\infty}^{\infty} c_\tau \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_{\mathrm{t}}}, \quad \text{where } c_\tau \stackrel{\mathrm{def}}{=} \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} C(f) \mathrm{e}^{\mathrm{i}2\pi f \tau \Delta_{\mathrm{t}}} \, \mathrm{d}f,$$

so that $\{c_\tau\} \longleftrightarrow C(\cdot)$. If $\lambda = 1$, the sequence $\{c_\tau\}$ is sometimes referred to as the *cepstrum* (Bogert et al., 1963) and is in many ways analogous to the ACVS.

Given observed values of $X_0, \ldots, X_{N-1}$ from a stationary process with SDF $S(\cdot)$, a natural estimator of $C(f)$ is

$$\hat{C}^{(\mathrm{D})}(f) \stackrel{\mathrm{def}}{=} \lambda \log(\hat{S}^{(\mathrm{D})}(f)), \tag{301b}$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

where $\hat{S}^{(\mathrm{D})}(\cdot)$ is a direct spectral estimator. If sufficient care is taken in selecting a data taper when needed, then $\hat{S}^{(\mathrm{D})}(f) \stackrel{\mathrm{d}}{=} S(f)\chi^2_\nu/\nu$ to a good approximation, where $\nu = 2$ when $0 < f < f_{\mathcal{N}}$, while $\nu = 1$ when $f = 0$ or $f = f_{\mathcal{N}}$ (see Equation (204b)). It follows that

$$\lambda \log\left(\hat{S}^{(\mathrm{D})}(f)\right) \stackrel{\mathrm{d}}{=} \lambda \log\left(S(f)\chi^2_\nu/\nu\right) \text{ and hence } \hat{C}^{(\mathrm{D})}(f) \stackrel{\mathrm{d}}{=} C(f) - \lambda \log(\nu) + \lambda \log\left(\chi^2_\nu\right)$$

approximately, from which we get

$$E\{\hat{C}^{(\mathrm{D})}(f)\} \approx C(f) - \lambda \log(\nu) + \lambda E\{\log\left(\chi^2_\nu\right)\} \text{ and } \operatorname{var}\{\hat{C}^{(\mathrm{D})}(f)\} \approx \lambda^2 \operatorname{var}\{\log\left(\chi^2_\nu\right)\}.$$

For positive integers ν, Bartlett and Kendall (1946) show that

$$E\{\log\left(\chi^2_\nu\right)\} = \log(2) + \psi(\nu/2) \text{ and } \operatorname{var}\{\log\left(\chi^2_\nu\right)\} = \psi'(\nu/2), \tag{302a}$$

where $\psi(\cdot)$ and $\psi'(\cdot)$ are the digamma and trigamma functions. In particular, we have

$$E\{\log\left(\chi^2_\nu\right)\} = \begin{cases} -\log(2) - \gamma, & \nu = 1; \\ \log(2) - \gamma, & \nu = 2; \end{cases} \text{ and } \operatorname{var}\{\log\left(\chi^2_\nu\right)\} = \begin{cases} \pi^2/2, & \nu = 1; \\ \pi^2/6, & \nu = 2, \end{cases} \tag{302b}$$

where $\gamma \doteq 0.5772$ is Euler's constant. These results yield

$$E\{\hat{C}^{(\mathrm{D})}(f)\} \approx \begin{cases} C(f) - \lambda \log(2) - \lambda\gamma, & f = 0 \text{ or } f_{\mathcal{N}}; \\ C(f) - \lambda\gamma, & 0 < f < f_{\mathcal{N}}; \end{cases} \tag{302c}$$

and

$$\operatorname{var}\{\hat{C}^{(\mathrm{D})}(f)\} \approx \begin{cases} \lambda^2\pi^2/2, & f = 0 \text{ or } f_{\mathcal{N}}; \\ \lambda^2\pi^2/6, & 0 < f < f_{\mathcal{N}}. \end{cases} \tag{302d}$$

This variance involves *known* constants – in particular it does *not* depend on $C(f)$.

As noted in the discussion surrounding Equation (207), there exists a grid of frequencies defined by $f'_k = k/(N'\Delta_{\mathrm{t}})$ with $N' \leq N$ such that the $\hat{S}^{(\mathrm{D})}(f'_k)$ RVs are approximately pairwise uncorrelated for $k = 0, \ldots, \lfloor N'/2 \rfloor$. With the additional assumption that $\{X_t\}$ is a Gaussian process, we can argue that the $\hat{C}^{(\mathrm{D})}(f'_k)$ RVs are also approximately pairwise uncorrelated. Let us now define the following estimator of c_τ:

$$\hat{c}'^{(\mathrm{D})}_\tau = \frac{1}{N'\Delta_{\mathrm{t}}} \sum_{k=-(N_L-1)}^{N_U-1} \hat{C}^{(\mathrm{D})}(f'_k) \mathrm{e}^{\mathrm{i}2\pi f'_k \tau \Delta_{\mathrm{t}}}, \quad \tau = -(N_L-1), \ldots, N_U - 1, \tag{302e}$$

where $N_L = \lfloor (N'-1)/2 \rfloor + 1$ and $N_U = \lfloor N'/2 \rfloor + 1$ so that $N_U + N_L - 1 = N'$ (recall that $\lfloor x \rfloor$ is the integer part of x). Note that, if N' is odd so that $N_L = N_U$, then Equation (302e) mimics the expression for $\hat{s}^{(\mathrm{P})}_\tau$ in Equation (171f). In view of Equations (100g), (101a) and (101b), we can write

$$\hat{C}^{(\mathrm{D})}(f'_k) = \Delta_{\mathrm{t}} \sum_{\tau=-(N_L-1)}^{N_U-1} \hat{c}'^{(\mathrm{D})}_\tau \mathrm{e}^{-\mathrm{i}2\pi f'_k \tau \Delta_{\mathrm{t}}}, \qquad k = -(N_L-1), \ldots, N_U - 1,$$

and we have $\{\hat{c}'^{(\mathrm{D})}_\tau\} \longleftrightarrow \{\hat{C}^{(\mathrm{D})}(f'_k)\}$. In view of Equations (74a) and (75a), by defining $\hat{c}'^{(\mathrm{D})}_\tau = 0$ for $\tau \geq N_U$ and $\tau \leq -N_L$, we also have $\{\hat{c}'^{(\mathrm{D})}_\tau\} \longleftrightarrow \hat{C}^{(\mathrm{D})}(\cdot)$, and hence

$$\hat{C}^{(\mathrm{D})}(f) = \Delta_{\mathrm{t}} \sum_{\tau=-(N_L-1)}^{N_U-1} \hat{c}'^{(\mathrm{D})}_\tau \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_{\mathrm{t}}}.$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

For a given lag window $\{w_{m,\tau}\}$, let us now define the following analog to the lag window spectral estimator $\hat{S}_m^{(\mathrm{LW})}(f)$:

$$\hat{C}_m^{(\mathrm{LW})}(f) = \Delta_\mathrm{t} \sum_{\tau=-(N_L-1)}^{N_U-1} w_{m,\tau} \hat{c}_\tau'^{(\mathrm{D})} \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_\mathrm{t}}. \tag{303a}$$

Just as $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ is a smoothed version of $\hat{S}^{(\mathrm{D})}(\cdot)$, so is $\hat{C}_m^{(\mathrm{LW})}(\cdot)$ a smoothed version of $\hat{C}^{(\mathrm{D})}(\cdot)$. As a measure of how well $\hat{C}_m^{(\mathrm{LW})}(\cdot)$ estimates $C(\cdot)$, let us consider the mean integrated square error (MISE):

$$E\left\{ \int_{-f_\mathcal{N}}^{f_\mathcal{N}} \left(\hat{C}_m^{(\mathrm{LW})}(f) - C(f) \right)^2 \mathrm{d}f \right\} = I_m + \Delta_\mathrm{t} \sum_{\substack{\tau \geq N_U \\ \tau \leq -N_L}} c_\tau^2, \tag{303b}$$

where

$$I_m \stackrel{\mathrm{def}}{=} \Delta_\mathrm{t} \sum_{\tau=-(N_L-1)}^{N_U-1} E\left\{ \left(w_{m,\tau} \hat{c}_\tau'^{(\mathrm{D})} - c_\tau \right)^2 \right\}$$

(this is an application of Parseval's theorem as stated by Equation (75b); see also the discussion surrounding Equation (52)). For a given lag window, the value of I_m depends on the window parameter m. The idea is to estimate I_m by, say, $\hat{I}_m$ and then to pick m such that $\hat{I}_m$ is minimized. Note that, if the right-most term in Equation (303b) is small relative to I_m, then I_m is a good approximation to the MISE. In any case, I_m is the only part of this error depending on the lag window and hence under our control. Now

$$I_m = \Delta_\mathrm{t} \sum_{\tau=-(N_L-1)}^{N_U-1} \left[w_{m,\tau}^2 E\{ (\hat{c}_\tau'^{(\mathrm{D})})^2 \} - 2 w_{m,\tau} c_\tau E\{ \hat{c}_\tau'^{(\mathrm{D})} \} + c_\tau^2 \right]. \tag{303c}$$

The burden of Exercise [7.21] is to show that, for large N',

$$E\{ \hat{c}_\tau'^{(\mathrm{D})} \} \approx c_\tau - \frac{\lambda \gamma \delta_\tau}{\Delta_\mathrm{t}} \quad \text{and} \quad E\{ (\hat{c}_\tau'^{(\mathrm{D})})^2 \} \approx \left(c_\tau - \frac{\lambda \gamma \delta_\tau}{\Delta_\mathrm{t}} \right)^2 + \frac{\lambda^2 \pi^2 \left(1 + \delta_\tau + \delta_{\tau - \frac{N'}{2}} \right)}{6 N' \Delta_\mathrm{t}^2}, \tag{303d}$$

where δ_τ is Kronecker's delta function (thus $\delta_\tau = 1$ if $\tau = 0$ and $= 0$ otherwise).

▷ **Exercise [303]** Show that Equations (303c) and (303d) lead to the approximation

$$I_m \approx \frac{\lambda^2 \gamma^2}{\Delta_\mathrm{t}} + \Delta_\mathrm{t} \sum_{\tau=-(N_L-1)}^{N_U-1} c_\tau^2 \left(1 - w_{m,\tau} \right)^2 + \frac{\lambda^2 \pi^2}{6 N' \Delta_\mathrm{t}} \sum_{\tau=-(N_L-1)}^{N_U-1} w_{m,\tau}^2 \left(1 + \delta_\tau + \delta_{\tau - \frac{N'}{2}} \right).$$

◁

Since $w_{m,0} = 1$ always, c_0^2 does not influence the above approximation; in view of Equation (303d), an approximately unbiased estimator of c_τ^2 for $0 < |\tau| < N'/2$ is given by $(\hat{c}_\tau'^{(\mathrm{D})})^2 - (\lambda^2 \pi^2 / (6 N' \Delta_\mathrm{t}^2))$. This suggests the following estimator for I_m:

$$\begin{aligned} \hat{I}_m \stackrel{\mathrm{def}}{=} & \frac{\lambda^2 \gamma^2}{\Delta_\mathrm{t}} + \Delta_\mathrm{t} \sum_{\tau=-(N_L-1)}^{N_U-1} \left((\hat{c}_\tau'^{(\mathrm{D})})^2 - \frac{\lambda^2 \pi^2}{6 N' \Delta_\mathrm{t}^2} \right) \left(1 - w_{m,\tau} \right)^2 \\ & + \frac{\lambda^2 \pi^2}{6 N' \Delta_\mathrm{t}} \sum_{\tau=-(N_L-1)}^{N_U-1} w_{m,\tau}^2 \left(1 + \delta_\tau + \delta_{\tau - \frac{N'}{2}} \right). \end{aligned} \tag{303e}$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

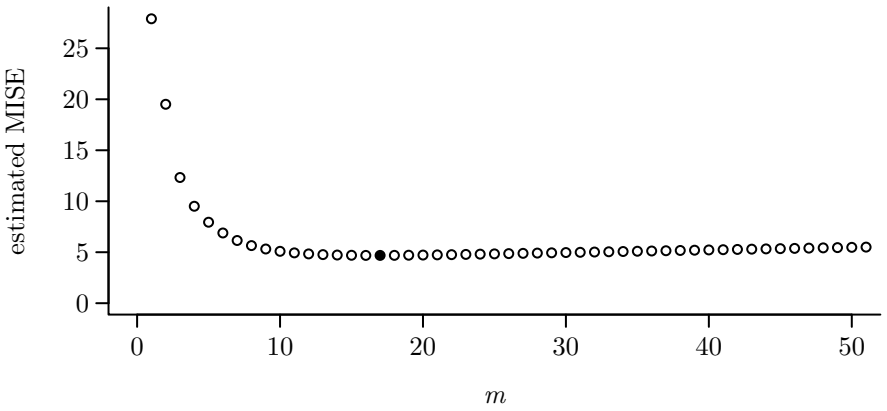

Figure 304 $\hat{I}_m$ of Equation (303e) versus m for the AR(2) time series of Figure 34(a) based upon the log periodogram and the Parzen lag window. The minimum value is at m = 17 (solid circle).

We can then objectively select m by picking the value for which $\hat{I}_m$ is smallest.

Sjoholm (1989) points out that an estimate of $S(\cdot)$ based on smoothing $\log(\hat{S}^{(\mathrm{D})}(\cdot))$ might have poor quantitative properties in that power is not necessarily preserved: since smoothing is done in log space, a high-power region can be disproportionately modified if it is adjacent to a low-power region. If $\hat{C}_m^{(\mathrm{LW})}(\cdot)$ is to be used for other than qualitative purposes, the just-described procedure for selecting m can be regarded merely as a way of establishing the appropriate smoothing window bandwidth, so that, once m is known, we can use it in computing $\hat{S}_m^{(\mathrm{LW})}(\cdot)$. To see the justification for using m as determined by the $\hat{I}_m$ criterion to form $\hat{S}_m^{(\mathrm{LW})}(\cdot)$, consider, for example, the Parzen lag window for which $w_{m,\tau} = 0$ for $|\tau| \geq m$ as per Equation (275a). Assuming $m \leq N_L$ (always the case in practice) and recalling that $\{\hat{c}_\tau'^{(\mathrm{D})}\} \longleftrightarrow \hat{C}^{(\mathrm{D})}(\cdot)$ and $\{\hat{s}_\tau^{(\mathrm{D})}\} \longleftrightarrow \hat{S}^{(\mathrm{D})}(\cdot)$, a study of Equations (303a) and (248a) indicates that

$$\hat{C}_m^{(\mathrm{LW})}(f) = \Delta_\mathrm{t} \sum_{\tau=-(m-1)}^{m-1} w_{m,\tau} \hat{c}_\tau'^{(\mathrm{D})} \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_\mathrm{t}} = \int_{-f_\mathcal{N}}^{f_\mathcal{N}} W_m(f-\phi) \hat{C}^{(\mathrm{D})}(\phi)\, \mathrm{d}\phi$$

and

$$\hat{S}_m^{(\mathrm{LW})}(f) = \Delta_\mathrm{t} \sum_{\tau=-(m-1)}^{m-1} w_{m,\tau} \hat{s}_\tau^{(\mathrm{D})} \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_\mathrm{t}} = \int_{-f_\mathcal{N}}^{f_\mathcal{N}} W_m(f-\phi) \hat{S}^{(\mathrm{D})}(\phi)\, \mathrm{d}\phi.$$

Use of the same smoothing window $W_m(\cdot)$ to create both $\hat{C}_m^{(\mathrm{LW})}(\cdot)$ and $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ makes sense since our choice of the smoothing window bandwidth should take into consideration both the bandwidths of the important features in $C(\cdot)$ and $S(\cdot)$ and how smooth we want the estimators to be. The bandwidths for $C(\cdot)$ and $S(\cdot)$ are arguably comparable since $C(\cdot)$ is just $S(\cdot)$ after a log transformation, and the degrees of smoothness are also relatively the same for $\hat{C}_m^{(\mathrm{LW})}(\cdot)$ and $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ when the latter is not subject to smoothing window leakage.

As a first example, let us consider the AR(2) time series of length $N = 1024$ shown in Figure 34(a). Figure 304 shows a plot of $\hat{I}_m$ versus m based upon the periodogram at the Fourier frequencies k/N and the Parzen lag window (here we can let $N' = N$ since a coarser grid is not needed because tapering is not needed). The minimizing value is $m = 17$ (the other three AR(2) time series shown in Figure 34 give comparable values of 16, 17 and

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

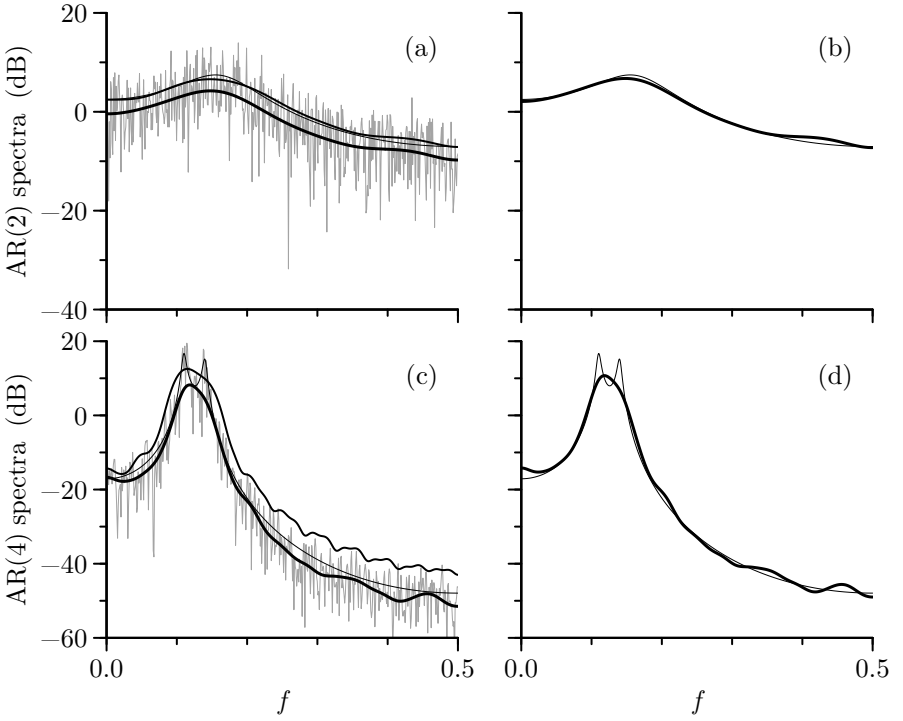

Figure 305 Comparison of estimates $\hat{C}_m^{(\mathrm{LW})}(\cdot)$ of log SDFs $C(\cdot) = 10 \log_{10}(S(\cdot))$ with estimates $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ of SDFs $S(\cdot)$. The rough-looking light curve in plot (a) is the periodogram $\hat{S}^{(\mathrm{P})}(\cdot)$ for the AR(2) series of Figure 34(a), upon which thin, medium and thick dark curves are superimposed. The thin and medium curves track each other closely and depict, respectively, the true AR(2) SDF and an $m = 17$ Parzen lag window SDF estimate $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ (this setting for m is dictated by the $\hat{I}_m$ criterion – see Figure 304). The thick dark curve below these two curves shows the log SDF estimate $\hat{C}_m^{(\mathrm{LW})}(\cdot)$, which is also based on the $m = 17$ Parzen lag window. Plot (b) compares the true AR(2) SDF with $\hat{C}_m^{(\mathrm{LW})}(\cdot)$ after correcting the latter for bias by adding $10 \log_{10}(\mathrm{e})\,\gamma$. The bottom row of plots has the same structure as the top row, but now involve the AR(4) series of Figure 34(e). The rough-looking light curve in (c) is a direct spectral estimate $\hat{S}^{(\mathrm{D})}(\cdot)$ based on the Hanning data taper rather than the periodogram, and the estimates $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ and $\hat{C}_m^{(\mathrm{LW})}(\cdot)$ are based on an $m = 45$ Parzen lag window (again the $\hat{I}_m$ criterion dictates this choice).

18). The erratic-looking light curve in Figure 305(a) shows the periodogram $\hat{S}^{(\mathrm{P})}(\cdot)$ for this AR(2) time series on a decibel scale, which is the same as displaying – on a linear scale – the estimate $\hat{C}^{(\mathrm{D})}(\cdot)$ of the log SDF defined by $\hat{C}^{(\mathrm{D})}(f) = 10 \log_{10}(\hat{S}^{(\mathrm{P})}(f))$. The true AR(2) SDF $S(\cdot)$ is the thin black curve, which, because we are displaying it on a decibel scale, is the same as the true log SDF $C(\cdot)$. This SDF is intertwined with a medium black curve showing an $m = 17$ Parzen lag window estimate $\hat{S}_m^{(\mathrm{LW})}(\cdot)$, i.e., $\hat{S}^{(\mathrm{P})}(\cdot)$ convolved with a Parzen smoothing window. The thick black curve below the thin and medium curves is the smoothed log periodogram $\hat{C}_m^{(\mathrm{LW})}(\cdot)$, i.e., $\hat{C}^{(\mathrm{D})}(\cdot)$ convolved with an $m = 17$ Parzen smoothing window. The estimate $\hat{C}_m^{(\mathrm{LW})}(\cdot)$ is systematically below the true log SDF. This bias arises because $\hat{C}_m^{(\mathrm{LW})}(\cdot)$ is a smoothed version of $\hat{C}^{(\mathrm{D})}(\cdot)$, which is a biased estimator of $C(\cdot)$; i.e., $E\{\hat{C}^{(\mathrm{D})}(f)\} \approx C(f) - \lambda\gamma$ for $0 < f < f_{\mathcal{N}}$, where here $\lambda = 10 \log_{10}(\mathrm{e})$. Figure 305(b) shows the true AR(2) log SDF (again as a thin black curve) with $\hat{C}_m^{(\mathrm{LW})}(\cdot)$ after correcting it for bias by the addition of $10 \log_{10}(\mathrm{e})\,\gamma$. The agreement between this bias-corrected $\hat{C}_m^{(\mathrm{LW})}(\cdot)$ and the true log SDF $C(\cdot)$ is as good as the agreement between $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ and the true SDF $S(\cdot)$.

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

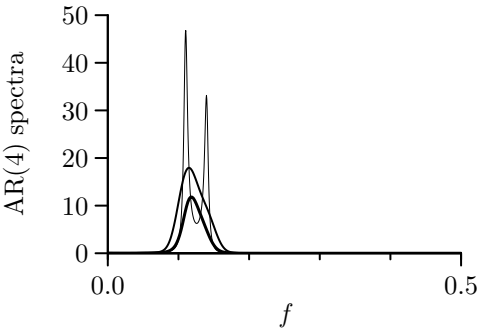

Figure 306 Comparison of true AR(4) SDF with a lag window estimate $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ and a log SDF estimate $\hat{C}_m^{(\mathrm{LW})}(\cdot)$ (corrected for bias by addition of $10\ \log_{10}(\mathrm{e})\ \gamma$) for AR(4) time series of Figure 34(e) (thin, medium and thick curves, respectively). Figures 305(c) and (d) show these spectra on linear/dB scales, whereas here we use a linear/linear scale.

A more challenging example is provided by the AR(4) time series of Figure 34(e), the results for which are shown in the bottom row of Figure 305. Because of bias in the periodogram, we define the basic log SDF estimator $\hat{C}^{(\mathrm{D})}(f) = 10\ \log_{10}(\hat{S}^{(\mathrm{D})}(f))$ in terms of a direct spectral estimator $\hat{S}^{(\mathrm{D})}(\cdot)$ based upon the Hanning data taper. This estimate is shown as the rough-looking light curve in Figure 305(c). We take the grid of frequencies over which the RVs $\hat{S}^{(\mathrm{D})}(f'_k)$ can be considered as approximately pairwise uncorrelated to be $f'_k = 2k/N$, i.e., a grid twice as coarse as the one defined by the Fourier frequencies. Using the Parzen lag window again, the $\hat{I}_m$ now selects $m = 45$ (the other three AR(4) time series in Figure 34 give $m = 86$, 46 and 50). The medium dark curve in Figure 305(c) shows the $m = 45$ Parzen lag window estimate $\hat{S}_m^{(\mathrm{LW})}(\cdot)$, which, when compared to the true AR(4) SDF (thin dark curve), is seen to suffer from smoothing window leakage at high frequencies and to smear out the high-power twin-peaks portion of the SDF. In comparison, the $m = 45$ Parzen estimate $\hat{C}_m^{(\mathrm{LW})}(\cdot)$ (thick dark curve) does not suffer from smoothing window leakage, but still blurs the twin peaks. In addition, $\hat{C}_m^{(\mathrm{LW})}(\cdot)$ is systematically below the true log SDF, which again can be attributed to the bias in $\hat{C}^{(\mathrm{D})}(\cdot)$. Figure 305(d) compares $\hat{C}_m^{(\mathrm{LW})}(\cdot)$ with the true SDF after correcting for bias by addition of $10\ \log_{10}(\mathrm{e})\ \gamma$. The agreement between the bias-corrected $\hat{C}_m^{(\mathrm{LW})}(\cdot)$ and the log SDF $C(\cdot)$ is now better than that between $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ and $S(\cdot)$ except in the twin peaks region, where the bias-corrected $\hat{C}_m^{(\mathrm{LW})}(\cdot)$ undershoots $C(\cdot)$. To focus on this discrepancy in more detail, Figure 306 shows plots of $S(f)$, $\hat{S}_m^{(\mathrm{LW})}(f)$ and $10^{(\hat{C}_m^{(\mathrm{LW})}(f)+10\ \log_{10}(\mathrm{e})\ \gamma)/10}$ versus f; i.e., all spectra are now shown on a linear scale rather on the decibel scales of Figures 305(c) and (d). While both estimates smear out the twin peaks badly, the integral of $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ is very close to the sample variance of the AR(4) time series (which in turn is reasonably close to the AR(4) process variance, i.e., the integral of the true SDF), whereas the integral of the estimate based on $\hat{C}_m^{(\mathrm{LW})}(\cdot)$ is about 40% smaller than the sample variance. This illustrates that smoothing $\log(\hat{S}^{(\mathrm{D})}(\cdot))$ can lead to an SDF estimate that does *not* constitute an analysis of the sample variance of a time series, which is one of the key goals of a proper spectral analysis.

Comments and Extensions to Section 7.9

[1] The starting point in estimating the log spectrum is the computation of $\hat{C}^{(\mathrm{D})}(f'_k) = \lambda \log(\hat{S}^{(\mathrm{D})}(f'_k))$, which is problematic if $\hat{S}^{(\mathrm{D})}(f'_k) = 0$. Although generally $\hat{S}^{(\mathrm{D})}(f) > 0$ for $f \neq 0$, we can have $\hat{S}^{(\mathrm{D})}(0) = 0$ due to centering the data (see Exercise [6.15b] and the discussion concerning Equation (195b)). This would force $\hat{C}^{(\mathrm{D})}(0)$ to be $-\infty$, in which case we need to redefine $\hat{C}^{(\mathrm{D})}(0)$ somehow. A simple solution is to set $\hat{C}^{(\mathrm{D})}(0) = \hat{C}^{(\mathrm{D})}(f'_1)$, but the merits of this approach are currently unknown.

[2] In addition to Wahba (1980), the log periodogram is used in a number of papers as the starting point

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

for SDF estimation. Pawitan and O'Sullivan (1994) use Wahba's smoothed log-periodogram estimator as the first iteration in a scheme leading to a penalized quasi-likelihood estimator. This estimator is defined as the maximizer of a roughness penalty combined with the Whittle likelihood function (Whittle, 1953; Dzhaparidze and Yaglom, 1983). This likelihood function is based on the assumption that the periodogram over the Fourier frequencies $f_k = k/(N\,\Delta_t)$, $k = 1, 2, \ldots, \lfloor (N-1)/2 \rfloor$, consists of independent $S(f)\chi^2_2/2$ RVs (C&E [5] for Section 6.6 has a discussion related to this assumption). Kooperberg et al. (1995) also make use of the Whittle likelihood, but now in combination both with cubic splines to express the log of the SDF and with indicator functions to handle the line components in processes with mixed spectra (see Section 4.4). Other papers that make the same assumption about the periodogram are Stoica and Sandgren (2006), who advocate smoothing the log periodogram by thresholding the corresponding estimated cepstrum, and Moulin (1994) and Gao (1993, 1997), who develop wavelet-based techniques for smoothing the log periodogram (see also section 10.6 of Percival and Walden, 2000). A common problem with these and other periodogram-based estimators is that, as noted in Section 6.3, the periodogram can be severely biased even for sample sizes that might seem to be large, and smoothing its log does nothing to compensate for this bias.

7.10 Bandwidth Selection for Periodogram Smoothing

In Section 7.1 we looked briefly at the periodogram to motivate discretely smoothed direct spectral estimators $\hat{S}^{(\mathrm{DS})}(\cdot)$ (see Equations (246a) and (246b)). Here we consider the periodogram as the basis for the SDF estimator

$$\hat{S}^{(\mathrm{DSP})}_m(f_k) \stackrel{\mathrm{def}}{=} \sum_{j=-\lfloor N/2 \rfloor}^{N-\lfloor N/2 \rfloor - 1} g_{m,j}\, \hat{S}^{(\mathrm{P})}(f_{k-j}), \tag{307}$$

where $f_k = k/(N\,\Delta_t)$ is a Fourier frequency, and $\{g_{m,j}\}$ is a set of N nonnegative smoothing coefficients such that $g_{m,-j} = g_{m,j}$ when the index j above is positive, $g_{m,-N/2} = 0$ when N is even, and $\sum_j g_{m,j} = 1$. The parameter m controls the degree of smoothing, and, once m is set, the smoothing coefficients are fully determined.

▷ **Exercise [307]** Show that $\hat{S}^{(\mathrm{DSP})}_m(\cdot)$ is a special case of $\hat{S}^{(\mathrm{DS})}(\cdot)$ of Equation (246b). ◁

We refer to $\hat{S}^{(\mathrm{DSP})}_m(f_k)$ as a *discretely smoothed periodogram.* Note that Equation (307) makes use of the periodogram over all possible Fourier frequencies f_k while recognizing that the periodogram is an even periodic function of frequency with a period of $2f_{\mathcal{N}}$ (see Equation (170c)). The particular form this equation takes is computationally appealing (this is the subject of Exercise [7.22a]). The smoothing coefficients are usually based on samples from a symmetric PDF. For example, a Gaussian PDF yields coefficients proportional to $\exp(-(mj)^2)$, where, in keeping with earlier use of m in this chapter (see, e.g., Table 279), the bandwidth associated with $\hat{S}^{(\mathrm{DSP})}_m(\cdot)$ decreases as m increases. Exercises [7.22b] to [7.22d] consider the EDOFs, the standard measure of effective bandwidth and the smoothing window bandwidth for $\hat{S}^{(\mathrm{DSP})}_m(\cdot)$. (In addition to the first part of Section 7.1, the following items in C&Es for previous sections deal with material of relevance to $\hat{S}^{(\mathrm{DSP})}_m(\cdot)$: items [4] and [5] of Section 7.1; [1] and [2] of Section 7.3; and [1] of Section 7.5.)

Our task is to objectively select an estimator from amongst the set of estimators $\hat{S}^{(\mathrm{DSP})}_m(\cdot)$ indexed by the smoothing parameter m. A number of papers have considered this problem under the rubric of bandwidth or span selection. These include Hurvich (1985), Beltrão and Bloomfield (1987), Hurvich and Beltrão (1990), Lee (1997, 2001), Fan and Kreutzberger (1998), Ombao et al. (2001), Hannig and Lee (2004) and Mombeni et al. (2017). All of these define estimators that are essentially special cases of $\hat{S}^{(\mathrm{DSP})}_m(\cdot)$. Here we present the method of Ombao et al. (2001). As is commonly done in the literature advocating smoothing of the periodogram, they assume that the periodogram $\hat{S}^{(\mathrm{P})}(f_k)$ over the $\lfloor (N-1)/2 \rfloor$ Fourier

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

frequencies satisfying $0 < f_k < f_{\mathcal{N}}$ constitute a set of independent $S(f_k)\chi^2_2/2$ RVs; in addition $\hat{S}^{(\mathrm{P})}(0)$ and $\hat{S}^{(\mathrm{P})}(f_{N/2})$ are assumed to be $S(f_k)\chi^2_1$ RVs that are independent both of one another and with the RVs associated with $0 < f_k < f_{\mathcal{N}}$ (note that $\hat{S}^{(\mathrm{P})}(f_{N/2})$ disappears when N is odd). The periodogram for Gaussian white noise agrees with this assumption (see the discussion surrounding Equation (203b)); for other processes satisfying certain stringent conditions, it is asymptotically reasonable (see C&E [5] for Section 6.6); but it can be dicey for time series encountered in practical applications (in particular, the assumption implies that the periodogram is unbiased, and there is ample evidence of its bias at certain frequencies for the ocean wave and chaotic beam data of Figures 225 and 228).

To select m, Ombao et al. (2001) propose minimizing a generalized cross-validated criterion that arises from the following considerations. McCullagh and Nelder (1989) discuss *deviance* as a measure of the goodness of fit between observations and a model for the observations (see their sections 2.3 and 8.3). Deviance is defined as the log of the ratio of two likelihood functions, one associated with a fully saturated model (i.e., the number of parameters in the model is equal to the number of observations), and the other, with a non-saturated model. In the current context, the observations consist of the periodogram at the Fourier frequencies, and the parameters in the fully saturated model are the values of the SDF at these frequencies. The non-saturated model is associated with the discretely smoothed periodogram. For $0 < f_k < f_{\mathcal{N}}$, the distribution of $\hat{S}^{(\mathrm{P})}(f_k)$ is taken to be that of the RV $S(f_k)\chi^2_2/2$, which has a PDF given by $f(u; S(f_k)) = \exp(-u/S(f_k))/S(f_k)$ (cf. Equation (209a)). The log likelihood function for a single observation $\hat{S}^{(\mathrm{P})}(f_k)$ is

$$\ell(S(f_k) \,|\, \hat{S}^{(\mathrm{P})}(f_k)) \stackrel{\mathrm{def}}{=} \log\left(f(\hat{S}^{(\mathrm{P})}(f_k); S(f_k))\right) = -\frac{\hat{S}^{(\mathrm{P})}(f_k)}{S(f_k)} - \log\left(S(f_k)\right).$$

The maximum likelihood estimator of $S(f_k)$, i.e., the value $S(f_k)$ that maximizes the log likelihood, is just $\hat{S}^{(\mathrm{P})}(f_k)$. The corresponding maximized log likelihood is

$$\ell(\hat{S}^{(\mathrm{P})}(f_k) \,|\, \hat{S}^{(\mathrm{P})}(f_k)) = -1 - \log\left(\hat{S}^{(\mathrm{P})}(f_k)\right).$$

Deviance is defined to be the difference between the above and the same log likelihood, but now evaluated at $\hat{S}^{(\mathrm{DSP})}_m(f_k)$:

$$\begin{aligned} D(\hat{S}^{(\mathrm{P})}(f_k), \hat{S}^{(\mathrm{DSP})}_m(f_k)) &\stackrel{\mathrm{def}}{=} \ell(\hat{S}^{(\mathrm{P})}(f_k) \,|\, \hat{S}^{(\mathrm{P})}(f_k)) - \ell(\hat{S}^{(\mathrm{DSP})}_m(f_k) \,|\, \hat{S}^{(\mathrm{P})}(f_k)) \\ &= \frac{\hat{S}^{(\mathrm{P})}(f_k)}{\hat{S}^{(\mathrm{DSP})}_m(f_k)} - \log\left(\frac{\hat{S}^{(\mathrm{P})}(f_k)}{\hat{S}^{(\mathrm{DSP})}_m(f_k)}\right) - 1. \end{aligned}$$

Given the assumption that the periodogram across the Fourier frequencies constitutes a set of independent RVs, Ombao et al. (2001) argue that an appropriate overall measure of deviance is

$$\frac{1}{M}\sum_{k=0}^{M-1} q_k D(\hat{S}^{(\mathrm{P})}(f_k), \hat{S}^{(\mathrm{DSP})}_m(f_k)), \tag{308}$$

where $M = \lfloor (N+1)/2 \rfloor$, and $q_k = 1$ with two exceptions, namely, $q_0 = 1/2$ and, if N is even, $q_{M-1} = 1/2$ also (these exceptions handle the periodogram at zero and Nyquist frequencies, where its distribution is assumed to be that of a $S(f_k)\chi^2_1$ RV rather than a $S(f_k)\chi^2_2/2$ RV; however, if the time series has been centered by subtracting off its sample mean (usually done in practical applications), then Exercise [6.15b] says that $\hat{S}^{(\mathrm{P})}(0) = 0$, in which case

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

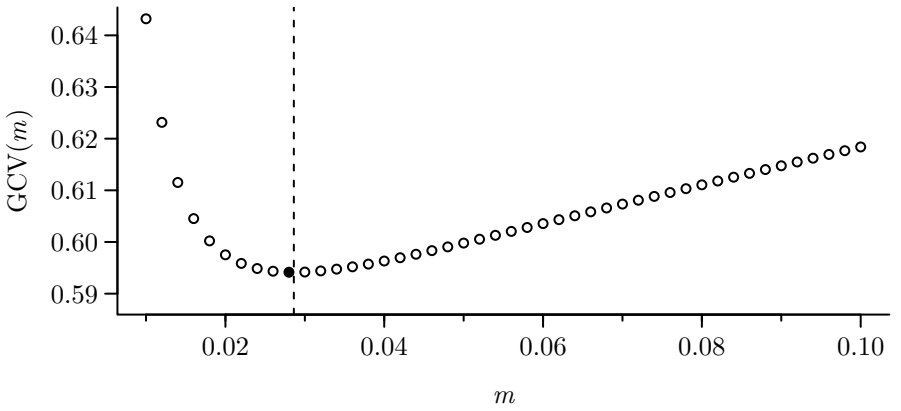

Figure 309 GCV(m) of Equation (309a) (with $\hat{S}_m^{(\mathrm{DSP})}(f_k)$ formed using weights $g_{m,j} \propto \exp(-(mj)^2)$) versus m = 0.01, 0.012, . . . , 0.1 for the AR(2) time series of Figure 34(a). The minimum value over the displayed grid is at m = 0.028 (solid circle). The vertical dashed line indicates the minimizer of GCV(m) picked out by the function `optimize` in R, namely, $m \doteq 0.0286$.

$k = 0$ in Equation (308) must be changed to $k = 1$). Their generalized cross-validated deviance method for selecting m is to minimize Equation (308) after division by $(1 - g_{m,0})^2$ (the so-called model degrees of freedom), i.e., to minimize

$$\mathrm{GCV}(m) \stackrel{\mathrm{def}}{=} \frac{1}{M(1-g_{m,0})^2} \sum_{k=0}^{M-1} q_k \left[\frac{\hat{S}^{(\mathrm{P})}(f_k)}{\hat{S}_m^{(\mathrm{DSP})}(f_k)} - \log\left(\frac{\hat{S}^{(\mathrm{P})}(f_k)}{\hat{S}_m^{(\mathrm{DSP})}(f_k)} \right) - 1 \right] \tag{309a}$$

(if, for example, $g_{m,j}$ is proportional to $\exp(-(mj)^2)$, then $g_{m,0} = 1/\sum_j \exp(-(mj)^2)$; if $k = 0$ in Equation (308) is changed to $k = 1$, then the same must be done in Equation (309a)). The role of $(1 - g_{m,0})^2$ is vital: without it, the fact that the minimum value of $x - \log(x) - 1$ over all positive x is zero and occurs at $x = 1$ says that the criterion would always set $\hat{S}_m^{(\mathrm{DSP})}(\cdot)$ to $\hat{S}^{(\mathrm{P})}(\cdot)$, i.e., no smoothing at all. If $\hat{S}_m^{(\mathrm{DSP})}(\cdot)$ is the same as $\hat{S}^{(\mathrm{P})}(\cdot)$, we must have $g_{m,0} = 1$, and $1/M(1 - g_{m,0})^2$ is infinite. As $\hat{S}_m^{(\mathrm{DSP})}(\cdot)$ moves away from the periodogram, the weight $g_{m,0}$ decreases from unity, and $1/M(1 - g_{m,0})^2$ gets smaller. By contrast the summation in Equation (309a) gets larger, so the two components of that equation play off against each other. In effect $1/(1 - g_{m,0})^2$ manages a trade-off between bias and variance in the discretely smoothed periodogram: as $\hat{S}_m^{(\mathrm{DSP})}(\cdot)$ gets smoother, the magnitude of its bias in $\hat{S}_m^{(\mathrm{DSP})}(\cdot)$ increases, while its variance decreases. Ombao et al. (2001) recommend section 6.9 of Hastie and Tibshirani (1990) for a discussion that illuminates the connection between GCV(m) and cross validation.

As a first example, let us consider again the AR(2) time series shown in Figure 34(a). The circles in Figure 309 show GCV(m) versus $m = 0.01, 0.012, \ldots, 0.1$ for a discretely smoothed periodogram using weights g_j proportional to $\exp(-(mj)^2)$. The value $m = 0.028$ minimizes GCV(m) over this grid (shown by the solid circle). The best value for m need not be on this grid, and indeed a nonlinear optimization routine picks the slightly higher off-grid value of $m \doteq 0.0286$ (indicated in the figure by the vertical dashed line). The left-hand plot of Figure 310 shows $\hat{S}_m^{(\mathrm{DSP})}(\cdot)$ for this setting of m (thick dark curve), along with the true AR(2) SDF (thin dark curve) and the periodogram (rough-looking light curve). The mean square error for $\hat{S}_m^{(\mathrm{DSP})}(\cdot)$, namely,

$$\mathrm{MSE}(m) = \frac{1}{N/2+1} \sum_{k=0}^{N/2} \left(\hat{S}_m^{(\mathrm{DSP})}(f_k) - S(f_k) \right)^2, \tag{309b}$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

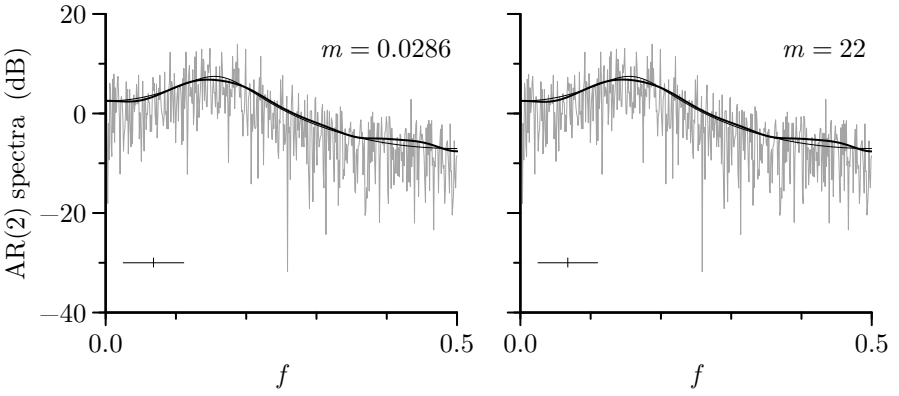

Figure 310 Discretely smoothed periodogram $\hat{S}_m^{(\mathrm{DSP})}(\cdot)$ using weights $g_{m,j} \propto \exp(-(mj)^2)$ with $m \doteq 0.0286$ for AR(2) series of Figure 34(a) (thick dark curve in left-hand plot). The thin dark curve (barely distinguishable from the thick curve) is the true AR(2) SDF, while the rough-looking light curve is the periodogram at the Fourier frequencies. The right-hand plot is similar, but now shows an $m = 22$ Parzen lag window estimate $\hat{S}_m^{(\mathrm{LW})}(\cdot)$. The widths of the crisscrosses indicate the bandwidth measure $B_{\mathcal{U}}$ (via Equation (347b) for $\hat{S}_m^{(\mathrm{DSP})}(\cdot)$ and via (256a) for $\hat{S}_m^{(\mathrm{LW})}(\cdot)$), while their heights display a 95% CI for 10 $\log_{10}(S(f))$ along the lines of Equation (266a) (with ν set by Equation (347a) for $\hat{S}_m^{(\mathrm{DSP})}(\cdot)$ and by Equation (264b) for $\hat{S}_m^{(\mathrm{LW})}(\cdot)$).

is 0.0562 for the m picked out by the GCV criterion (recall that $N = 1024$ for the AR(2) time series). The minimizer of MSE(m) is $m \doteq 0.0381$, for which MSE$(m) \doteq 0.0453$. The ratio of these two MSEs is 1.240, so $\hat{S}_m^{(\mathrm{DSP})}(\cdot)$ with m chosen by the GCV criterion has an MSE that is 25% larger than the best possible m (the three AR(2) time series in Figures 34(b), (c) and (d) lead to ratios of 2.385, 1.321 and 1.004).

The discretely smoothed periodogram selected by GCV(m) for the AR(2) time series has associated EDOFs $\nu \doteq 175.2$ and bandwidth measure $B_{\mathcal{U}} \doteq 0.0862$ (see Equations (347a) and (347b)). These quantities dictate the height and width of the crisscross in the left-hand plot of Figure 310. The right-hand plot illustrates the close relationship discussed in Section 7.1 between discretely smoothed periodograms and lag windows estimates. The thick dark curve shows an $m = 22$ Parzen lag window estimate $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ based on the periodogram (the two other curves are the same as in the left-hand plot). The setting $m = 22$ yields a bandwidth measure of $B_{\mathcal{U}} \doteq 0.0849$, which is as close as we can get to the measure associated with the discretely smoothed periodogram $\hat{S}_m^{(\mathrm{DSP})}(\cdot)$ The maximum difference between the two estimates is minuscule (0.07 dB). The crisscrosses in the two plots are also virtually identical (the EDOFs for the lag window estimate are $\nu \doteq 172.6$).

As a second example, we turn again to the AR(4) time series of Figure 34(e). Figure 173(e) shows its periodogram, which is badly biased over a substantial interval of frequencies. Since smoothing this periodogram cannot lead to an acceptable SDF estimate, let us instead consider a Hanning-based direct spectral estimate $\hat{S}^{(\mathrm{D})}(\cdot)$. Figure 187(e) indicates the bias in this estimate is substantially less than the periodogram's, and hence $\hat{S}^{(\mathrm{D})}(\cdot)$ is a more suitable candidate for smoothing. To do so, we need only replace $\hat{S}^{(\mathrm{P})}(f_k)$ with $\hat{S}^{(\mathrm{D})}(f_k)$ in Equation (307) and then use the resulting estimator along with $\hat{S}^{(\mathrm{D})}(f_k)$ in Equation (309a); however, minimizing this redefined GCV(m) to set m lacks theoretical justification because the $\hat{S}^{(\mathrm{D})}(f_k)$ RVs at the Fourier frequencies are not independent. Ignoring this fact and applying the GCV criterion anyway, we obtain the estimate shown in the left-hand side of Figure 311, for which $m \doteq 0.2313$ (the minimizer of GCV(m)). As determined by Equation (309b) after replacing $\hat{S}_m^{(\mathrm{DSP})}(\cdot)$ with the discretely smoothed direct spectral estimate, the corresponding MSE is 4.258. The best possible MSE is 4.256, which corresponds

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

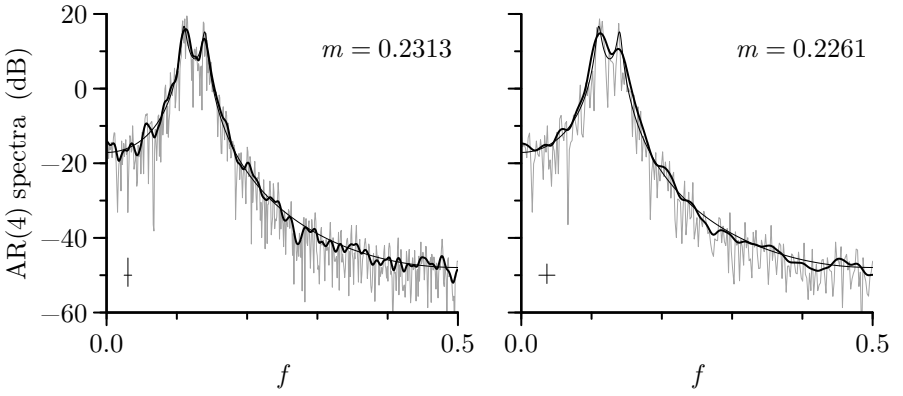

Figure 311 Discretely smoothed Hanning-based direct spectral estimate using weights $g_{m,j} \propto \exp(-(mj)^2)$ with $m \doteq 0.2313$ for AR(4) series of Figure 34(e) (thick dark curve in left-hand plot). The thin dark curve is the true AR(4) SDF, while the rough-looking light curve is $\hat{S}^{(\mathrm{D})}(\cdot)$ at the Fourier frequencies f_k, $k = 0, 1, \ldots, N/2$. The right-hand plot is similar, but now the light curve is $\hat{S}^{(\mathrm{D})}(\cdot)$ at f_{2k}, $k = 0, 1, \ldots, N/4$ (i.e., a grid twice as coarse as the Fourier frequencies), while the thick dark curve is $\hat{S}^{(\mathrm{D})}(\cdot)$ after smoothing over this coarser grid. The widths of the crisscrosses indicate the bandwidth measure $B_{\mathcal{U}}$, and their heights display a 95% CI for 10 $\log_{10}(S(f))$ along the lines of Equation (266a) with ν set appropriately. Exercise [7.23] addresses creation of the left-hand crisscross; for the right-hand plot, $B_{\mathcal{U}}$ and ν are determined using Equations (347b) and (347a) with N replaced by $N/2$.

to setting m to 0.2348. The ratio between these MSEs is 1.0003 (the three AR(4) time series in Figures 34(f), (g) and (h) lead to ratios of 2.810, 1.001 and 1.013).

Despite lacking firm theoretical backing, the GCV criterion appears to select a reasonable bandwidth for a Hanning-based direct spectral estimator $\hat{S}^{(\mathrm{D})}(\cdot)$, at least for the AR(4) process of Equation (35a). Nonetheless it is of interest to consider the following modification to the GCV scheme because it has some theoretical appeal. When a Hanning data taper is used, it is reasonable to take $\hat{S}^{(\mathrm{D})}(f_{2k})$, $k = 0, 1, \ldots, \lfloor N/4 \rfloor$, to be approximately independent RVs obeying rescaled chi-square distributions with either one or two degrees of freedom; i.e., we entertain the same assumption that Ombao et al. (2001) make about the periodogram, but now over a grid of frequencies twice as coarse as the one defined by the Fourier frequencies f_k. If we regard the direct spectral estimates $\hat{S}^{(\mathrm{D})}(f_{2k})$ as surrogates for a periodogram for a time series of length $\lfloor N/2 \rfloor$ and if we plug them into straightforward modifications of Equations (307) and (309a), we are led to the discretely smoothed direct spectral estimate displayed in the right-hand plot of Figure 311. The minimizer of GCV(m) is $m \doteq 0.2261$, and the associated MSE over the grid of frequencies f_{2k} is 7.916. The minimizer of MSE(m) is $m \doteq 0.4148$. Its associate MSE is 5.752, so the ratio of these MSEs is 1.376 (the ratios are 1.560, 1.418 and 2.836 for the AR(4) series in Figures 34(f), (g) and (h)). Superficially it seems that the modified GCV scheme is not working as well as the one that just ignores theoretical considerations! (Exercise [7.24] asks the reader to verify parts of Figures 310 and 311 and then to generate other AR(2) and AR(4) time series and analyze these for comparison.)

A critique of the modified GCV scheme is that it smooths over only half as many values of $\hat{S}^{(\mathrm{D})}(\cdot)$ as the scheme that just ignores correlations between $\hat{S}^{(\mathrm{D})}(f_k)$ and $\hat{S}^{(\mathrm{D})}(f_{k\pm1})$. A possible refinement is to use the modified scheme to pick m and then smooth $\hat{S}^{(\mathrm{D})}(\cdot)$ over the Fourier frequencies with a newly determined m set such that the resulting bandwidth $B_{\mathcal{U}}$ matches the one associated with the modified scheme. The bandwidth associated with the estimate shown in the right-hand plot of Figure 311 is $B_{\mathcal{U}} \doteq 0.02291$, which is gotten from Equation (347b) by using $N/2 = 512$ in place of N and by using $m \doteq 0.2261$ to set the $N/2$ weights $g_{m,j}$. Regarding the right-hand side of Equation (347b) as a function of m with N

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

set to 1024, we can find the value of m yielding N weights such that the corresponding $B_{\mathcal{U}}$ is also 0.02291. This leads to using $m \doteq 0.1099$ and to an associated MSE of 7.567. The ratio of this to be best possible MSE is 1.316, which is worse than the value of close to unity for the first scheme we considered (the ratios are 1.464, 1.341 and 2.943 for the AR(4) series in Figures 34(f), (g) and (h)).

Comments and Extensions to Section 7.10

[1] The terms between the brackets in Equation (309a) defining GCV(m) are quite similar to the ones between the brackets in Equation (297c), which defines a discrepancy measure inspired by the Kullback–Leibler (KL) discrepancy between two PDFs. Hannig and Lee (2004) propose a method for selecting m that is based directly on the KL discrepancy. Their method selects m as the minimizer of

$$\frac{1}{M-1}\sum_{k=1}^{M-1}\left[\sum_{j=-J}^{J} g_{m,j} + \frac{\hat{S}_m^{(\mathrm{DSP})}(f_k) - \sum_{j=-J}^{J} g_{m,j}\hat{S}^{(\mathrm{P})}(f_{k-j})}{\frac{1}{2J+1}\sum_{j=-J}^{J}\hat{S}^{(\mathrm{P})}(f_{k-j})} - \log\left(\frac{\hat{S}_m^{(\mathrm{DSP})}(f_k)}{\hat{S}^{(\mathrm{P})}(f_k)}\right)\right],$$

where J is a small integer that controls local averaging of the periodogram (both with weights $g_{m,j}$ and with weights $1/(2J+1)$). If we set $J = 0$ so that local averaging is eliminated, the above is proportional to

$$\sum_{k=1}^{M-1}\left[\frac{\hat{S}_m^{(\mathrm{DSP})}(f_k)}{\hat{S}^{(\mathrm{P})}(f_k)} - \log\left(\frac{\hat{S}_m^{(\mathrm{DSP})}(f_k)}{\hat{S}^{(\mathrm{P})}(f_k)}\right)\right]. \tag{312a}$$

If we eliminate the terms corresponding to zero and Nyquist frequency in GCV(m), the result is proportional to

$$\frac{1}{(1-g_{m,0})^2}\sum_{k=1}^{M-1}\left[\frac{\hat{S}^{(\mathrm{P})}(f_k)}{\hat{S}_m^{(\mathrm{DSP})}(f_k)} - \log\left(\frac{\hat{S}^{(\mathrm{P})}(f_k)}{\hat{S}_m^{(\mathrm{DSP})}(f_k)}\right)\right]. \tag{312b}$$

The two methods swap the roles of $\hat{S}^{(\mathrm{P})}(f_k)$ and $\hat{S}_m^{(\mathrm{DSP})}(f_k)$; moreover, since $x - \log(x)$ is minimized for positive x when $x = 1$, it follows that Equation (312a) is minimized when $\hat{S}_m^{(\mathrm{DSP})}(f_k)$ is set to $\hat{S}^{(\mathrm{P})}(f_k)$, i.e., the discretely smoothed periodogram degenerates to the periodogram. Local averaging evidently plays a role comparable to that governed by $g_{m,0}$ in Equation (312b) (without $1/(1-g_{m,0})^2$, Equation (312b) is also minimized when $\hat{S}_m^{(\mathrm{DSP})}(f_k)$ is set to $\hat{S}^{(\mathrm{P})}(f_k)$). Hannig and Lee (2004) note that J manages the trade-off between bias and variance in $\hat{S}_m^{(\mathrm{DSP})}(f_k)$, but its use introduces an element of subjectivity to their procedure that does not exist in the Ombao et al. (2001) procedure. The Hannig and Lee (2004) procedure has some asymptotic appeal since the authors were able to show that their discrepancy estimator is consistent.

[2] The fact that all lag window estimators have a corresponding well-defined ACVS is put to good use in Chapter 11 when we discuss generating realizations of time series with an SDF specified by a lag window estimator (an important topic if, e.g., bootstrapping is of interest). It would be nice to do the same with discretely smoothed periodograms, but defining an ACVS for these is not straightforward. To understand why this is so, let us first review some facts about the periodogram.

Since $\{\hat{s}_\tau^{(\mathrm{P})}\} \longleftrightarrow \hat{S}^{(\mathrm{P})}(\cdot)$, the ACVS estimator corresponding to the periodogram is the biased estimator of the ACVS. Given the periodogram, we can retrieve its corresponding ACVS estimator using the inverse Fourier transform:

$$\hat{s}_\tau^{(\mathrm{P})} = \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \hat{S}^{(\mathrm{P})}(f)\mathrm{e}^{\mathrm{i}2\pi f\tau\,\Delta_{\mathrm{t}}}\,\mathrm{d}f, \quad \tau \in \mathbb{Z}$$

(this is Equation (171c)). The above might suggest that we need to know the periodogram for all $f \in [-f_{\mathcal{N}}, f_{\mathcal{N}}]$ to get $\{\hat{s}_\tau^{(\mathrm{P})}\}$; however, by definition $\hat{s}_\tau^{(\mathrm{P})} = 0$ for $|\tau| \geq N$, and we also have

$$\{\hat{s}_\tau^{(\mathrm{P})} : \tau = -N, \ldots, N-1\} \longleftrightarrow \{\hat{S}^{(\mathrm{P})}(\tilde{f}_k) : k = -N, \ldots, N-1\} \tag{312c}$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

where $\tilde{f}_k = k/(2N\,\Delta_t)$ (the above is a slight variation on Equation (171e) – it follows from an argument analogous to the one leading to that equation). We can thus retrieve the nonzero portion of $\{\hat{s}_\tau^{(P)}\}$ using the inverse DFT:

$$\hat{s}_\tau^{(P)} = \frac{1}{2N\,\Delta_t} \sum_{k=-N}^{N-1} \hat{S}^{(P)}(\tilde{f}_k) e^{i2\pi \tilde{f}_k \tau\,\Delta_t} \tag{313a}$$

(cf. Equation (171f)). Thus, given the periodogram over the grid of frequencies $\tilde{f}_k$ twice as fine as that of the Fourier frequencies f_k, we can get $\{\hat{s}_\tau^{(P)} : \tau \in \mathbb{Z}\}$. Exercise [6.12] demonstrates that it is not possible to recover $\{\hat{s}_\tau^{(P)}\}$ given the periodogram over just the Fourier frequencies.

Defining an ACVS estimator, say $\{\hat{s}_\tau^{(DSP)}\}$, to go along with a discretely smoothed periodogram is complicated by the fact that $\hat{S}_m^{(DSP)}(\cdot)$ is specified only at the Fourier frequencies. Just as having access to $\hat{S}^{(P)}(\cdot)$ only at these frequencies is not enough to uniquely specify an ACVS estimator, the same is true for $\hat{S}_m^{(DSP)}(\cdot)$. If, however, we were to define $\hat{S}_m^{(DSP)}(\tilde{f}_k)$ for odd indices k, we could then obtain $\{\hat{s}_\tau^{(DSP)} : \tau = 0, \ldots, N\}$ by replacing $\hat{S}^{(P)}(\tilde{f}_k)$ with $\hat{S}_m^{(DSP)}(\tilde{f}_k)$ on the right-hand side of Equation (312c) (we need only worry about odd k because $\hat{S}_m^{(DSP)}(\tilde{f}_k) = \hat{S}_m^{(DSP)}(f_{k/2})$ for even k). Linear interpolation is one option: given the estimates at the Fourier frequencies provided by Equation (307), define estimates between these frequencies (i.e., at frequencies $\tilde{f}_k$ indexed by odd k) using

$$\hat{S}_m^{(DSP)}(\tilde{f}_k) = \frac{\hat{S}_m^{(DSP)}(\tilde{f}_{k-1}) + \hat{S}_m^{(DSP)}(\tilde{f}_{k+1})}{2} = \frac{\hat{S}_m^{(DSP)}(f_{(k-1)/2}) + \hat{S}_m^{(DSP)}(f_{(k+1)/2})}{2}, \tag{313b}$$

and, if N is odd, additionally define $\hat{S}_m^{(DSP)}(\tilde{f}_{-N}) = \hat{S}_m^{(DSP)}(f_{-(N-1)/2})$ (note that $\tilde{f}_{-N} = -f_{\mathcal{N}}$). A second option is a modification of Equation (307) that makes use of the periodogram over frequencies $\tilde{f}_{k-j}$ rather than just the Fourier frequencies:

$$\hat{S}_m^{(DSP)}(\tilde{f}_k) = \sum_{j=-N}^{N-1} \tilde{g}_{m,j} \hat{S}^{(P)}(\tilde{f}_{k-j}), \quad \text{where } \tilde{g}_{m,j} = \begin{cases} g_{m,j/2}, & j \text{ even;} \\ 0, & j \text{ odd.} \end{cases} \tag{313c}$$

When k is even so that $\tilde{f}_k$ is a Fourier frequency, the above reduces to Equation (307); when k is odd, the above defines a discretely smoothed periodogram at the non-Fourier frequencies. This definition combines values of the periodogram at non-Fourier frequencies using the same weighting scheme employed to create $\hat{S}_m^{(DSP)}(f_k)$ at the Fourier frequencies. Exercise [7.25] explores these and other options for defining $\hat{s}_\tau^{(DSP)}$.

[3] A lag window estimator $\hat{S}_m^{(LW)}(\cdot)$ that is based on the periodogram is an alternative to a discretely smoothed periodogram $\hat{S}_m^{(DSP)}(\cdot)$; however, even when confined to just the Fourier frequencies, $\hat{S}_m^{(LW)}(\cdot)$ is not a special case of $\hat{S}_m^{(DSP)}(\cdot)$. We thus cannot use the GCV procedure as described by Equation (309a) to set the parameter m for the lag window $\{w_{m,\tau}\}$, but the procedure can be adapted to do so, as follows. We start by reworking Exercise [249b] so that Equation (249d) uses the periodogram over the grid of frequencies $\tilde{f}_k = k/(2N\,\Delta_t)$ (the Fourier frequencies $f_k = k/(N\,\Delta_t)$ are a subset of these):

$$\hat{S}_m^{(LW)}(\tilde{f}_k) = \sum_{j=-(N-1)}^{N-1} g_j \hat{S}^{(P)}(\tilde{f}_{k-j}), \quad \text{where } g_j = \frac{1}{2N-1} \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau} e^{-i2\pi \tilde{f}_j \tau\,\Delta_t}$$

(the expression for g_j above involves an appeal to Equation (247e)). We then make two adjustments to the GCV criterion. On the right-hand side of Equation (309a), we replace $\hat{S}_m^{(DSP)}(f_k)$ with $\hat{S}_m^{(LW)}(f_k)$. We also change $g_{m,0}$ to $2g_0$, where g_0 is dictated by the above, and the factor of 2 is introduced to correct for the $\tilde{f}_k$ grid having twice as many frequencies as the f_k grid. These adjustments yield the adaptation

$$\mathrm{GCV}(m) = \frac{1}{M} \sum_{k=0}^{M-1} \frac{q_k}{(1-2g_0)^2} \left[\frac{\hat{S}^{(P)}(f_k)}{\hat{S}_m^{(LW)}(f_k)} - \log\left(\frac{\hat{S}^{(P)}(f_k)}{\hat{S}_m^{(LW)}(f_k)} \right) - 1 \right]. \tag{313d}$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

Application of this criterion to the AR(2) time series of Figure 34(a) picks out an $m = 22$ Parzen lag window estimate, which is shown in the right-hand plot of Figure 310 and is virtually identical to the discretely smoothed periodogram picked out by the GCV criterion of Equation (309a) (left-hand plot). (Exercise [7.26] addresses the extent to which other realizations of the AR(2) process yields similar results.)

[4] We noted in C&E [2] for Section 7.7 that the MSE is but one way to measure the discrepancy between estimated and true SDFs. The normalized mean square error NMSE(m) of Equation (296a) is a viable alternative if the true SDF is positive at all frequencies, as is the case for the AR(2) and AR(4) SDFs we've used as examples. For the estimate $\hat{S}^{(\mathrm{DSP})}(\cdot)$ of the AR(2) SDF shown in the left-hand plot of Figure 310, the theoretical minimizer of NMSE(m) is $m \doteq 0.0283$ for which NMSE(m) $\doteq 0.014939$. The NMSE is 0.014944 for the m picked out by the GCV criterion (0.0286). The ratio of these two NMSEs is 1.0003, and the corresponding ratios for the other three AR(2) time series are 1.012, 1.133 and 1.002; by contrast, the MSE-based ratios for the four series are 1.240, 2.385, 1.321 and 1.004. The NMSE measure thus views the GCV bandwidth selection procedure more favorably than does the MSE. Turning now to the AR(4) SDF estimate in the left-hand plot of Figure 311, the theoretical minimizer of NMSE(m) is $m \doteq 0.1858$ for which NMSE(m) $\doteq 0.1433$. The NMSE is 0.1537 for the m picked out by the GCV criterion (0.2313). The ratio of these NMSEs is 1.072, and the ratios for the other AR(4) series are 1.647, 1.364 and 1.323. The MSE-based ratios are 1.0003, 2.810, 1.001 and 1.013. With the exception of the second AR(4) series, the NMSE measure views the GCV bandwidth selection procedure less favorably than does the MSE – the opposite of what happened in the AR(2) case! Exercise [7.27] invites the reader to look at many other realizations of the AR(2) and AR(4) processes in part to explore more thoroughly the evaluation of the GCV procedure by different discrepancy measures.

7.11 Computational Details

Here we give details about computation of a lag window spectral estimate $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ given a time series we regard as a realization of a portion $X_0, X_1, \ldots, X_{N-1}$ of a stationary process with sampling interval Δ_t and unknown mean μ, SDF $S(\cdot)$ and ACVS $\{s_\tau\}$. Pathways for computing $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ from $\{X_t\}$ are shown in Figure 315 and involve four Fourier transform pairs. Because a lag window estimator is built upon a direct spectral estimator $\hat{S}^{(\mathrm{D})}(\cdot)$, the first three pairs are the same as the ones shown in Figure 221 for computing $\hat{S}^{(\mathrm{D})}(\cdot)$. The left-hand sides of these three pairs are

(1) $\{\tilde{X}_t : t = 0, 1, \ldots, M-1\}$, the time series after centering by subtracting off its sample mean $\overline{X}$ and then padding with at least $N-1$ zeros to form a sequence of length $M \geq 2N-1$ (see Equation (220e));
(2) $\{\tilde{h}_t \tilde{X}_t\}$, the product of $\tilde{X}_t$ and the zero-padded data taper $\tilde{h}_t$ (Equation (220e)); and
(3) $\{\tilde{h}\tilde{X} \star \tilde{h}\tilde{X}_\tau\} \stackrel{\mathrm{def}}{=} \{\tilde{s}_\tau^{(\mathrm{D})}\}$, the cyclic autocorrelation of $\{\tilde{h}_t \tilde{X}_t\}$ as per Equation (220b).

The right-hand sides are

(1) $\{\tilde{\mathcal{X}}_k : k = 0, 1, \ldots, M-1\}$, the DFT of $\{\tilde{X}_t\}$;
(2) $\{\tilde{H} * \tilde{\mathcal{X}}_k\}$, the frequency domain cyclic convolution of the DFT $\{\tilde{H}_k\}$ of $\{\tilde{h}_t\}$ and the DFT $\{\tilde{\mathcal{X}}_k\}$ as per Equation (220d); and
(3) $\{|\tilde{H} * \tilde{\mathcal{X}}_k|^2\} \stackrel{\mathrm{def}}{=} \{\tilde{S}^{(\mathrm{D})}(\tilde{f}'_k)\}$, the square modulus of $\{\tilde{H} * \tilde{\mathcal{X}}_k\}$, where $\tilde{f}'_k = k/M$.

The lag window $\{w_{m,\tau}\}$ enters the picture as a contributor to the fourth Fourier transform pair in Figure 315. Using this window, we define the sequence $\{\tilde{w}_{m,\tau}\}$ of length M by

$$\tilde{w}_{m,\tau} = \begin{cases} w_{m,\tau}, & 0 \leq \tau \leq N-1; \\ 0, & N \leq \tau \leq M-N; \\ w_{m,M-\tau}, & M-N+1 \leq \tau \leq M-1. \end{cases} \tag{314}$$

We denote its DFT by $\{\tilde{W}(\tilde{f}'_k)\}$. The left-hand side of (4) is obtained by multiplying the sequences $\{\tilde{w}_{m,\tau}\}$ and $\{\tilde{s}_\tau^{(\mathrm{D})}\}$ term by term to produce $\{\tilde{w}_{m,\tau}\tilde{s}_\tau^{(\mathrm{D})}\} \stackrel{\mathrm{def}}{=} \{\tilde{s}_{m,\tau}^{(\mathrm{LW})}\}$. Its DFT –

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

$$
\begin{array}{ccccc}
\{\tilde{X}_t\} & \overset{(1)}{\longleftrightarrow} & \{\tilde{\mathcal{X}}_k\} & & \\
\downarrow \text{mult} & & \downarrow \text{conv} & & \\
\{\tilde{h}_t \tilde{X}_t\} & \overset{(2)}{\longleftrightarrow} & \{\tilde{H} * \tilde{\mathcal{X}}_k\} & & \\
\downarrow \text{auto} & & \downarrow \text{mod sq} & & \\
\{\tilde{s}_\tau^{(\mathrm{D})}\} = \{\tilde{h}\tilde{X} \star \tilde{h}\tilde{X}_\tau\} & \overset{(3)}{\longleftrightarrow} & \{|\tilde{H} * \tilde{\mathcal{X}}_k|^2\} & = & \{\tilde{S}^{(\mathrm{D})}(\tilde{f}'_k)\} \\
\downarrow \text{mult} & & \downarrow \text{conv} & & \\
\{\tilde{s}_{m,\tau}^{(\mathrm{LW})}\} = \{\tilde{w}_{m,\tau}\tilde{s}_\tau^{(\mathrm{D})}\} & \overset{(4)}{\longleftrightarrow} & \{\tilde{W} * \tilde{S}^{(\mathrm{D})}(\tilde{f}'_k)\} & = & \{\tilde{S}_m^{(\mathrm{LW})}(\tilde{f}'_k)\}
\end{array}
$$

Figure 315 Pathways for computing $\tilde{S}_m^{(\mathrm{LW})}(\cdot)$, from which, upon multiplication by Δ_t, we can obtain the lag window estimate $\hat{S}_m^{(\mathrm{LW})}(f'_k) = \Delta_\mathrm{t}\, \tilde{S}_m^{(\mathrm{LW})}(\tilde{f}'_k)$, where $f'_k = k/(M\,\Delta_\mathrm{t})$ and $\tilde{f}'_k = k/M$ (adapted from figure 1 of Van Schooneveld and Frijling, 1981). This figure incorporates Figure 221, which shows computational pathways for direct spectral estimates.

denoted as $\{\tilde{W} * \tilde{S}^{(\mathrm{D})}(\tilde{f}'_k)\} = \{\tilde{S}_m^{(\mathrm{LW})}(\tilde{f}'_k)\}$ – can be obtained directly or indirectly via the frequency domain cyclic convolution of $\{\tilde{W}(\tilde{f}'_k)\}$ and $\{\tilde{S}^{(\mathrm{D})}(\tilde{f}'_k)\}$ using Equation (220d). From the left-hand side of (4), we obtain the lag window-based ACVS estimate:

$$
\hat{s}_{m,\tau}^{(\mathrm{LW})} = \begin{cases} \tilde{s}_{|\tau|}^{(\mathrm{LW})}, & |\tau| \le N-1; \\ 0, & \text{otherwise.} \end{cases}
$$

From the right-hand side, we obtain the desired lag window spectral estimate:

$$
\hat{S}_m^{(\mathrm{LW})}(f'_k) = \Delta_\mathrm{t}\, \tilde{S}_m^{(\mathrm{LW})}(\tilde{f}'_k), \qquad k = 0, \ldots, \lfloor M/2 \rfloor.
$$

As shown in Figure 315, we can use any one of eight computational pathways to obtain $\{\hat{S}_m^{(\mathrm{LW})}(f'_k)\}$ via $\{\tilde{S}_m^{(\mathrm{LW})}(\tilde{f}'_k)\}$ from $\{X_t\}$, each of which we might prefer under certain conditions. For example, the pathway

$$
\begin{array}{ccccc}
\{\tilde{X}_t\} & & & & \\
\downarrow \text{mult} & & & & \\
\{\tilde{h}_t \tilde{X}_t\} & \overset{(2)}{\longrightarrow} & \{\tilde{H} * \tilde{\mathcal{X}}_k\} & & \\
 & & \downarrow \text{mod sq} & & \\
\{\tilde{h}\tilde{X} \star \tilde{h}\tilde{X}_\tau\} & \overset{(3)}{\longleftarrow} & \{|\tilde{H} * \tilde{\mathcal{X}}_k|^2\} & & \\
\downarrow \text{mult} & & & & \\
\{\tilde{w}_{m,\tau}\tilde{s}_\tau^{(\mathrm{D})}\} & \overset{(4)}{\longrightarrow} & \{\tilde{W} * \tilde{S}^{(\mathrm{D})}(\tilde{f}'_k)\} & = & \{\tilde{S}_m^{(\mathrm{LW})}(\tilde{f}'_k)\}
\end{array}
$$

requires two DFTs and one inverse DFT and allows us to easily obtain $\{\hat{S}^{(\mathrm{D})}(f'_k)\}$, $\{\hat{s}_\tau^{(\mathrm{D})}\}$, $\{\hat{s}_{m,\tau}^{(\mathrm{LW})}\}$ and $\{\hat{S}_m^{(\mathrm{LW})}(f'_k)\}$ (in that order). If we are not interested in examining the ACVS

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

estimates $\{\hat{s}_\tau^{(\mathrm{D})}\}$ and $\{\hat{s}_{m,\tau}^{(\mathrm{LW})}\}$, then we could use the pathway

$$\begin{array}{ccc}
\{\tilde{X}_t\} & & \\
\Big\downarrow \text{mult} & & \\
\{\tilde{h}_t \tilde{X}_t\} & \xrightarrow{(2)} & \{\tilde{H} * \tilde{\mathcal{X}}_k\} \\
& & \Big\downarrow \text{mod sq} \\
& & \{|\tilde{H} * \tilde{\mathcal{X}}_k|^2\} \\
& & \Big\downarrow \text{conv} \\
& & \{\tilde{W} * \tilde{S}^{(\mathrm{D})}(\tilde{f}'_k)\}\} \;=\; \{\tilde{S}_m^{(\mathrm{LW})}(\tilde{f}'_k)\},
\end{array}$$

which involves one DFT and one convolution. This second pathway might require fewer numerical operations than the first if the sequence $\{\tilde{W}(\tilde{f}'_k)\}$ is sufficiently short so that the convolution can be computed efficiently with Equation (220d) (in fact, we now only need $M \geq N$ instead of $M \geq 2N-1$ because padding with $N-1$ or more zeros is only required to correctly compute $\{\hat{s}_\tau^{(\mathrm{D})}\}$ and $\{\hat{s}_{m,\tau}^{(\mathrm{LW})}\}$). Note also that this pathway is exactly how in practice we would compute the discretely smoothed direct spectral estimator $\hat{S}^{(\mathrm{DS})}(\cdot)$ of Equation (246b).

Figure 317 gives an example of computing a lag window spectral estimate by extending the example of computing a direct spectral estimate shown in Figure 223. The four top rows in both figures are the same and are described in Section 6.7, to which we refer the reader for a description. Turning to the final two rows of Figure 317, the left-hand plot on the fifth row is $\{\tilde{w}_{m,\tau}\}$, formed here using a Parzen lag window with $m = 10$; the right-hand plot is the real-valued sequence $\{\tilde{W}(\tilde{f}'_k)\}$, the DFT of $\{\tilde{w}_{m,\tau}\}$. The left-hand plot on the last row is $\{\tilde{s}_{m,\tau}^{(\mathrm{LW})}\}$, from which we can form $\{\hat{s}_{m,\tau}^{(\mathrm{LW})}\}$; the right-hand plot is $\{\tilde{S}_m^{(\mathrm{LW})}(\tilde{f}'_k)\}$, from which we can compute $\{\hat{S}_m^{(\mathrm{LW})}(f'_k)\}$, a smoothed version of $\{\hat{S}^{(\mathrm{D})}(f'_k)\}$.

7.12 Examples of Lag Window Spectral Estimators

Ocean Wave Data

Here we continue the analysis of the ocean wave time series that we started in Section 6.8. Figure 225 shows the time series, while Figure 226 displays four direct spectral estimates for this series, namely, the periodogram and three estimates based upon Slepian tapers with NW set to $1/\Delta_\mathrm{t}$, $2/\Delta_\mathrm{t}$ and $4/\Delta_\mathrm{t}$. There is evidence of bias due to leakage in the periodogram and – to a lesser extent – in the estimate based on the $NW = 1/\Delta_\mathrm{t}$ Slepian taper. As discussed in Section 6.8, part of the rationale for collecting this time series was to investigate the rate at which the SDF is decreasing over the frequency range 0.2 to 1.0 Hz. While the periodogram exhibits leakage in this frequency range, none of the Slepian-based direct spectral estimates do; however, to illustrate a point concerning smoothing window leakage (see the discussion surrounding Figure 319(a)), we will concentrate on the $NW = 2/\Delta_\mathrm{t}$ direct spectral estimate (Figure 226(c)) even though the $NW = 1/\Delta_\mathrm{t}$ estimate (Figure 226(b)) is adequate for this range of frequencies. We want to smooth this direct spectral estimate to get a better visual representation of the rate at which the power in the SDF decreases over this frequency range. We do so using a lag window approach.

Figure 318 shows a Parzen lag window estimate with parameter $m = 150$ (dark curve). This SDF estimate is a smoothed version of the $NW = 2/\Delta_\mathrm{t}$ direct spectral estimate (the rough-looking light curve). We set $m = 150$ to achieve a smoothing window bandwidth B_W comparable to a visual determination of the width of the broad peak in $\hat{S}^{(\mathrm{D})}(\cdot)$ at $f =$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

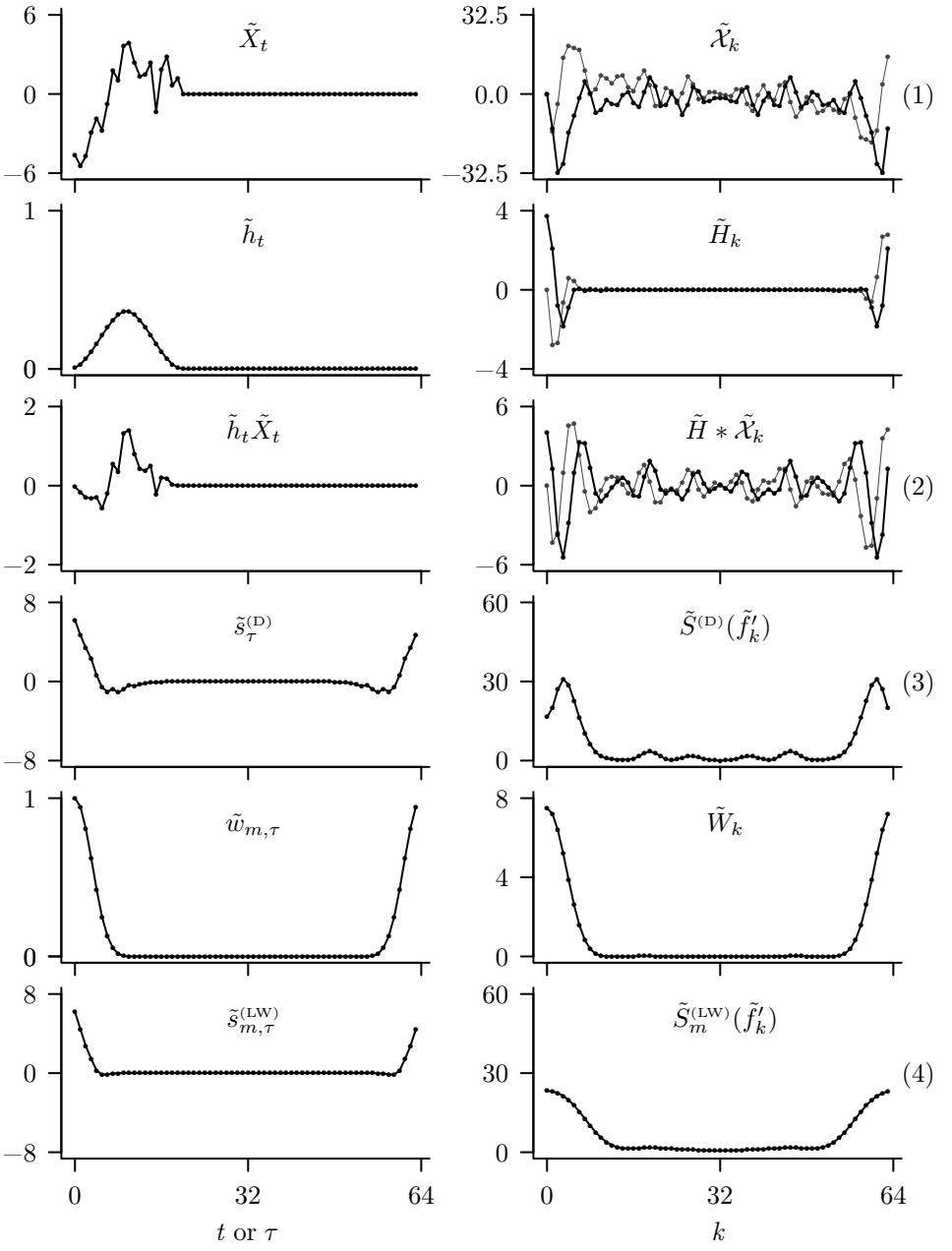

Figure 317 Illustration of computational pathways for lag window spectral estimates (cf. Figure 315). The upper four rows are identical to Figure 223, which illustrates computational pathways for direct spectral estimates.

0.160 Hz. This choice produces a smoothed version of $\hat{S}^{(\mathrm{D})}(\cdot)$ that smears out this peak somewhat (and the one at $f = 1.469$ Hz). To help assess this lag window estimate, we display a crisscross with an interpretation similar to the crisscrosses for the direct spectral estimates in Figure 226; however, the horizontal width of the crisscross here now depicts $B_{\mathcal{U}} \doteq 0.050$ Hz, the measure of effective bandwidth appropriate for a lag window estimator (Equation (256a)). This bandwidth can again be interpreted as a rough measure of the distance in frequency

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

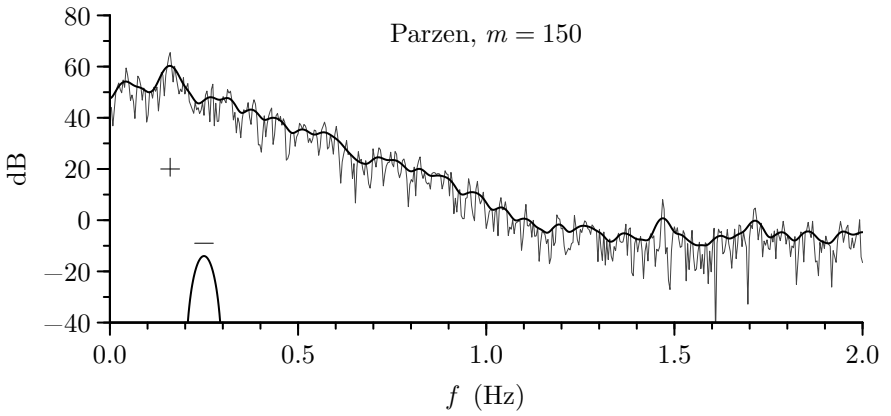

Figure 318 Parzen lag window spectral estimate $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ with $m = 150$ for ocean wave data (dark curve). The light curve shows the underlying $\hat{S}^{(\mathrm{D})}(\cdot)$ (also shown in Figure 226(c)). The width and height of the crisscross gives the bandwidth measure $B_{\mathcal{U}}$ (Equation (256a)) and the length of a 95% CI for 10 $\log_{10} S(f)$) (Equation (266a)). The central lobe for the smoothing window $W_m(\cdot)$ is shown in the lower left-hand corner with a horizontal line depicting its bandwidth B_W.

between adjacent uncorrelated spectral estimates. The vertical height of this crisscross represents the length of a 95% CI for $10\,\log_{10}(S(f))$ based upon $10\,\log_{10}(\hat{S}_m^{(\mathrm{LW})}(f))$. This length is computed as described at the end of Section 7.4 (the associated EDOFs for this estimate are $\nu \doteq 12.6$). The central lobe of the corresponding smoothing window $W_m(\cdot)$ (shifted so that its peak value is plotted at $1/4$ Hz and -14 dB) is shown in the lower left-hand corner. There is a horizontal line above this smoothing window, which depicts the smoothing window bandwidth B_W. This bandwidth can be computed either via its definition in Equation (251e) or via the simple approximating formula $1.85/(m\,\Delta_{\mathrm{t}})$ stated in Table 279, both of which yield 0.049 Hz (note that this is slightly less than $B_{\mathcal{U}} \doteq 0.050$ Hz, as is reasonable given that the latter takes into account the data taper in addition to the lag window).

While the $m = 150$ Parzen estimate tracks the $NW = 2/\Delta_{\mathrm{t}}$ direct spectral estimate nicely and is considerably smoother, it is arguably too bumpy to adequately represent the rolloff pattern over the range of frequencies of main interest ($0.2\,\mathrm{Hz} \le f \le 1.0\,\mathrm{Hz}$). We thus consider a second Parzen estimate, but now one that is associated with a larger smoothing window bandwidth B_W. If we let $m = 55$, we obtain the estimate shown in Figure 319(a) (this figure has a layout similar to that of Figure 318). For this estimate we have $B_W \doteq 0.135$ Hz, which is almost three times larger than the one associated with the $m = 150$ estimate. The $m = 55$ Parzen estimate is much smoother looking (a comparison of the crisscrosses in Figures 318 and 319(a) shows the inherent trade-off between variability – as measured by the length of a 95% CI – and the effective bandwidth as measured by $B_{\mathcal{U}}$). While this new estimate arguably offers a better representation of the rolloff pattern for $0.2\,\mathrm{Hz} \le f \le 1.0\,\mathrm{Hz}$, it does not do well at other frequencies. The peaks at 0.160 and 1.469 Hz have been smeared out almost to the point of being unrecognizable. There is also evidence of smoothing window leakage between frequencies 1.0 Hz and 2.0 Hz: the dark curve in Figure 319(a) should be a smoothed version of the $NW = 2$ direct spectral estimate (the light curve), but note that the dark curve is consistently *higher* than the light curve (by contrast, this is not the case in the high-power portion of the spectral estimate, which in this example corresponds to low-frequency values).

To confirm that the $m = 55$ Parzen estimate does indeed suffer from smoothing window leakage at high frequencies, we can compare it to Daniell and Bartlett–Priestley lag window

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

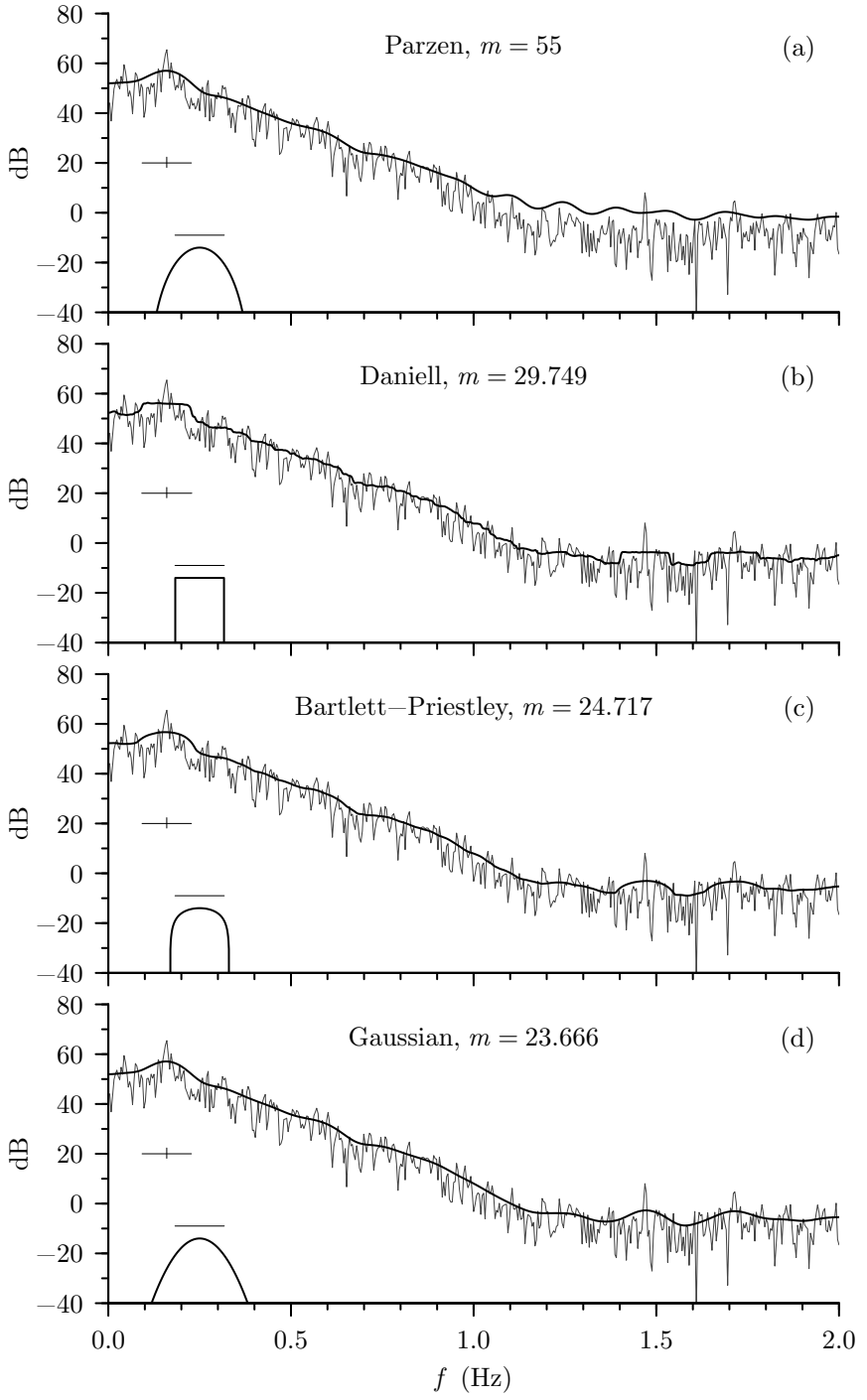

Figure 319 Four lag window spectral estimates $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ for ocean wave data (dark curves). The layout is the same as for Figure 318, with the exception that the central lobe for the design window $V_m(\cdot)$ – rather than the smoothing window $W_m(\cdot)$ – is shown in the lower left-hand corners for the Daniell and Bartlett–Priestley estimates (plots (b) and (c)).

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

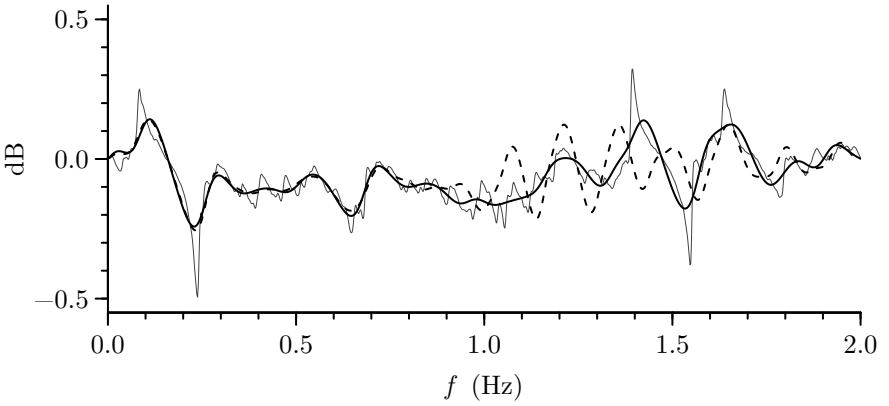

Figure 320 First difference $10\ \log_{10}(\hat{S}_m^{(\mathrm{LW})}(\tilde{f}_k)) - 10\ \log_{10}(\hat{S}_m^{(\mathrm{LW})}(\tilde{f}_{k-1}))$ of lag window SDF estimates versus $\tilde{f}_k$. The dashed, light jagged and dark smooth curves are for, respectively, the Parzen, Bartlett–Priestley and Gaussian estimates shown in parts (a), (c) and (d) of Figure 319.

estimates, both of which are necessarily leakage free. For these two estimates we set m such that the resulting smoothing window bandwidths are the same as that of the $m = 55$ Parzen smoothing window (i.e., $B_W \doteq 0.135\,\mathrm{Hz}$). The appropriate choices for the Daniell and Bartlett–Priestley smoothing windows are, respectively, $m = 29.749$ and $m = 24.717$. The estimates are shown in Figures 319(b) and (c). Neither of these estimates is systematically above the underlying direct spectral estimate in the 1.0 to 2.0 Hz region, which confirms that the Parzen estimate suffers from leakage. The Daniell estimate, however, is not nearly as smooth as the Parzen estimate. The Bartlett–Priestley estimate is superior to the Daniell estimate in terms of smoothness, but a Gaussian lag window estimate offers further improvement even though its associated smoothing window is not entirely free of sidelobes (see Figure 278(b)). This estimate is shown in Figure 319(d), for which we have set $m = 23.666$ to achieve a smoothing window bandwidth of $B_W \doteq 0.135\,\mathrm{Hz}$ (the same as for the other three estimates in Figure 319; for the record, we note that the EDOFs for all four estimates are $\nu \doteq 34.4$). All the estimates we have considered are evaluated over the grid of frequencies $\tilde{f}_k = k/(2N\,\Delta_t)$. To qualitatively assess smoothness, Figure 320 shows the first difference $10\ \log_{10}(\hat{S}_m^{(\mathrm{LW})}(\tilde{f}_k)) - 10\ \log_{10}(\hat{S}_m^{(\mathrm{LW})}(\tilde{f}_{k-1}))$ versus $\tilde{f}_k$ for the Parzen (dashed curve), Bartlett–Priestley (light jagged) and Gaussian (dark solid) estimates. The Bartlett–Priestley estimate has the most rapidly changing first difference, and the Gaussian, the least, with the Parzen being similar to the Gaussian for $f < 0.9\,\mathrm{Hz}$ and having a ringing pattern over the frequencies for which it suffers from leakage. Since rapid changes in the first difference are an indication of lack of smoothness, we see that the Gaussian estimate combines the desirable properties of the $m = 55$ Parzen estimate (smoothness) with those of the Bartlett–Priestley estimate (freedom from smoothing window leakage), thus giving us the estimate of choice amongst the four in Figure 319.

To summarize, we have considered five different lag window estimates, all of which are smoothed versions of an erratic $NW = 2$ direct spectral estimate. The $m = 150$ Parzen lag window estimate (Figure 318) shows the SDF rolling off over $0.2\,\mathrm{Hz} \le f \le 1.0\,\mathrm{Hz}$ in a complicated non-monotonic manner, which does not make it easy to compare with physical theories. The $m = 55$ Parzen estimate (Figure 319(a)) is a more useful estimate over the frequency range of interest, but a comparison of this estimate with its corresponding $\hat{S}^{(\mathrm{D})}(\cdot)$ indicates that it suffers from smoothing window leakage over $1.0\,\mathrm{Hz} \le f \le 2.0\,\mathrm{Hz}$ (an important point to note here is that we should routinely compare any $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ we are entertaining

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

with its corresponding $\hat{S}^{(\mathrm{D})}(\cdot)$ to make sure that the former is a reasonably smoothed version of the latter). The Daniell, Bartlett–Priestley and Gaussian estimates (Figures 319(b), (c) and (d)) show no evidence of leakage, with the Gaussian being the estimate of choice because of its superior smoothness properties. Note that the crisscrosses in all four plots in Figures 319 are virtually identical, which says that the four estimates shown there have the same bandwidths and EDOFs. Thus, while all four estimates share some statistical properties, there are important differences amongst them, and these differences cannot be deduced based on bandwidths and EDOFs alone. (We return to an analysis of these data in Sections 8.9, 9.12, 10.15 and 11.6)

Ice Profile Data

Let us now consider two arctic sea-ice profiles (this analysis expands upon an unpublished 1991 report by W. Fox, Applied Physics Laboratory, University of Washington). Ice morphology was measured by a profiling device moving at a constant speed in a straight line roughly parallel to the surface of the ice. The distance between the profiler and the ice surface was recorded at a constant rate in time to obtain measurements spaced every $\Delta_{\mathrm{t}} = 1.7712$ m apart. We can regard measurements along this transect as a "time" series related to variations in the surface height of the ice over distance, i.e., a sea-ice profile. The two top plots of Figure 322 employ similar vertical scales to show two profiles. The first profile (upper plot) is fairly rough and has $N_R = 1121$ data points, while the second (middle) is relatively smooth and has $N_S = 288$ points. The bottom plot gives an expanded view of the smooth profile. In what follows, we center the profiles by subtracting off their sample means – these are $\overline{X}_R \doteq -19.92$ for the rough profile and $\overline{X}_S \doteq -8.93$ for the smooth.

One reason for collecting these profiles was to develop a way of generating representative artificial profiles for use in simulations. It is clear from the top two plots of Figure 322 that it is unrealistic to regard both profiles as portions of two realizations from a single stationary process – the mean values and the variances of the two profiles are quite different. Superficially, however, the profiles in the bottom and top plots appear to have similar "spikiness," so one simple model worth investigating is that the spectral content of the two profiles is the same except for a scaling factor; i.e., if we let $S_R(\cdot)$ and $S_S(\cdot)$ represent the SDFs for the rough and smooth profiles, we have $S_R(f) = cS_S(f)$ over all f for some constant $c > 0$. Here we informally test this hypothesis by comparing estimated spectra for the two series.

The light choppy curve in Figure 323(a) – replicated in plots (c) and (e) – is the periodogram $\hat{S}_R^{(\mathrm{P})}(\cdot)$ for the rough profile; similarly, the periodogram $\hat{S}_S^{(\mathrm{P})}(\cdot)$ for the smooth profile appears in Figures 323(b), (d) and (f). An examination of direct spectral estimates using Slepian data tapers with $NW = 1/\Delta_{\mathrm{t}}, 2/\Delta_{\mathrm{t}}$ and $4/\Delta_{\mathrm{t}}$ shows that, for both profiles, these estimates do not differ in any substantial way from the periodograms. Thus, in contrast to the ocean wave series, the periodograms $\hat{S}_R^{(\mathrm{P})}(\cdot)$ and $\hat{S}_S^{(\mathrm{P})}(\cdot)$ appear to be bias free. Figure 323 displays both periodograms over positive Fourier frequencies less than or equal to the Nyquist frequency $f_{\mathcal{N}} \doteq 0.2823$. The spacing between these frequencies is $1/(N_R\,\Delta_{\mathrm{t}}) \doteq 0.0005$ and $1/(N_S\,\Delta_{\mathrm{t}}) \doteq 0.0020$ cycles/meter for the rough and smooth profiles. The fact that $\hat{S}_S^{(\mathrm{P})}(\cdot)$ looks smoother than $\hat{S}_R^{(\mathrm{P})}(\cdot)$ is merely because the smooth profile has a smaller sample size (cf. Figure 207). Note that, since the units for the profiles are in feet and since the "time" variable is measured in meters, the SDFs has units of squared feet per cycle per meter.

We next smooth the two periodograms to better compare them. Since the dynamic ranges of the periodograms are not large (less than about 40 dB), smoothing window leakage should not be an issue here, so we can safely use the Parzen lag window. The two periodograms suggest that the underlying SDFs are dominantly unimodal with the mode at zero frequency, so we can use the estimator $\tilde{B}_T$ of Equation (300c) to compute the bandwidth for the two profiles. For the rough profile we obtain $\tilde{B}_T = 0.0239$, and for the smooth profile, $\tilde{B}_T =$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

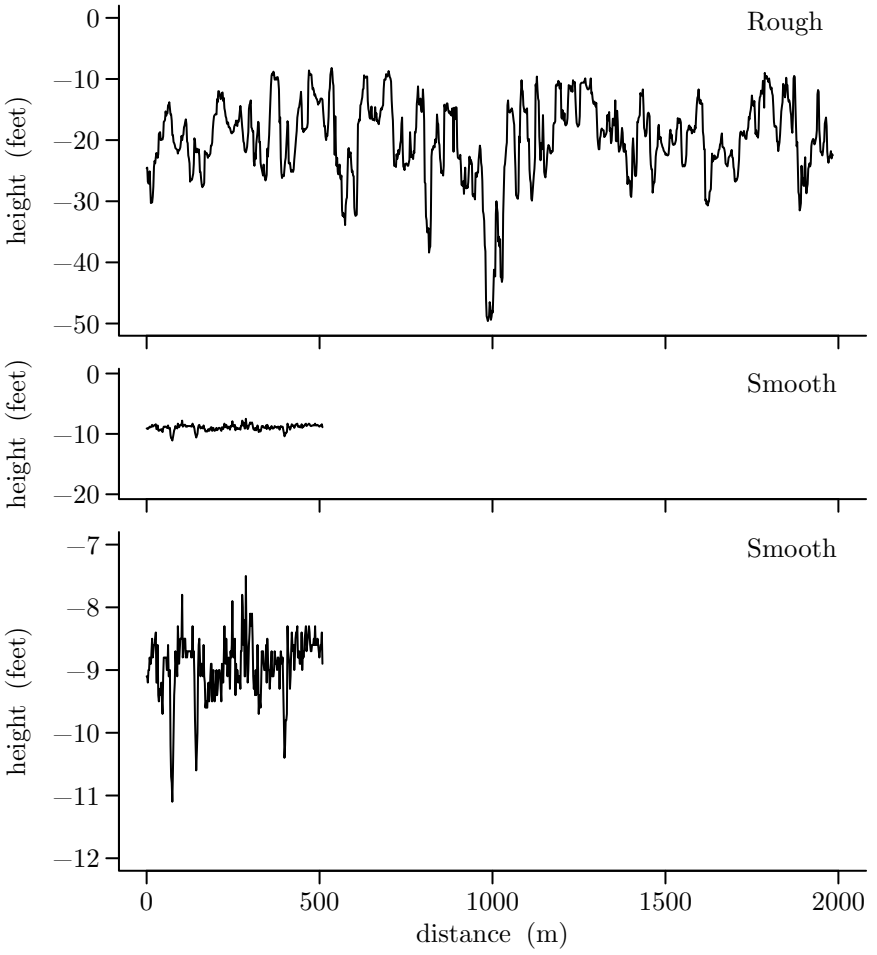

Figure 322 Two segments of ice profile data (courtesy of W. Fox, Applied Physics Laboratory, University of Washington). The top plot shows a relatively rough profile, while the middle, a shorter but relatively smooth one (also shown on the bottom plot with an expanded scale). The "time" values are really distances measured in meters, while the time series values are in feet. The distance between observations is $\Delta_t = 1.7712$ m. The rough profile has 1121 data points, while the smooth one has 288 points.

0.1069. Following the advice of Section 7.8, we select the lag window parameter m for each profile such that the resulting smoothing window bandwidth B_W is equal to $\tilde{B}_T/2$. Using the approximation $B_W = 1.85/(m\,\Delta_t)$ from Table 279, we obtain

$$m = \left[\frac{3.7}{\tilde{B}_T\,\Delta_t}\right] = \begin{cases} 88, & \text{for the rough profile;} \\ 20, & \text{for the smooth profile,} \end{cases} \tag{322}$$

where $[x]$ here denotes x rounded to the nearest integer. The resulting lag window estimators $\hat{S}^{(\mathrm{LW})}_{R,88}(\cdot)$ and $\hat{S}^{(\mathrm{LW})}_{S,20}(\cdot)$ are shown as the dark smoother curves in Figures 323(a) and (b).

Rather than objectively smoothing the periodograms through estimation of the spectral bandwidth, we can also entertain the objective methods described in Sections 7.9 (automatic smoothing after a log transformation) and 7.10 (bandwidth selection via generalized cross-validation). Using the first of these two methods with the Parzen lag window, we set m to the

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

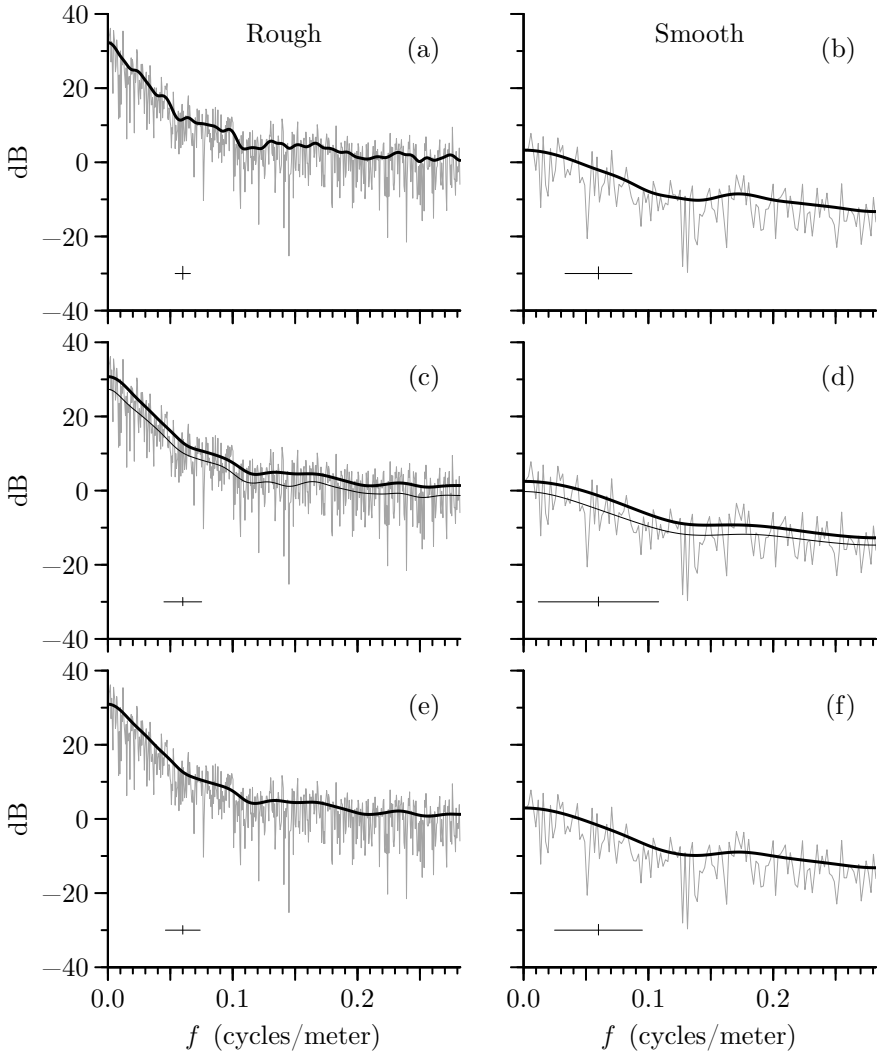

Figure 323 Periodogram for rough ice profile data (light choppy curves replicated in three left-hand plots) and for smooth profile (right-hand plots). The darkest curve in each plot is a smoothed version of the corresponding periodogram. Plots (a) and (b) show Parzen lag window estimates with the lag window parameter m set by equating the smoothing window bandwidth B_W to $\tilde{B}_T/2$, where $\tilde{B}_T$ is the estimator of the bandwidth of the time series given by Equation (300c); the darkest curves in plots (c) and (d) are Parzen lag window estimates with m set to the minimizer of $\hat{I}_m$ of Equation (303e), while the smooth curves just below these are smoothed versions of the log periodogram, again based on the Parzen lag window; and, finally, plots (e) and (f) are discretely smoothed periodograms with the smoothing parameters m set to the minimizer of GCV(m) of Equation (309a). The crisscross in the lower left-hand corner of each plot shows the bandwidth measure $B_{\mathcal{U}}$ and the length of 95% CIs.

minimizer of $\hat{I}_m$ of Equation (303e), yielding

$$m = \begin{cases} 35, & \text{for the rough profile;} \\ 11, & \text{for the smooth profile} \end{cases} \tag{323}$$

(due to centering of the time series, the periodograms at $f = 0$ are zero, so their logs are problematic – C&E [1] for Section 7.9 discusses how to handle this difficulty). We can use the selected m in conjunction with the Parzen lag window to smooth either the periodogram or

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

the log periodogram. The thicker of the two smooth curves in Figures 323(c) and (d) show the smoothed periodograms $\hat{S}^{(\mathrm{LW})}_{R,35}(\cdot)$ and $\hat{S}^{(\mathrm{LW})}_{S,11}(\cdot)$ (the crisscrosses refer to these estimates); the thinner ones show the smoothed log periodograms. The smoothed log periodograms are systematically below the corresponding smoothed periodograms. Since spectral analysis should provide a frequency-based decomposition of the sample variance of a time series, the integral of an SDF estimate over $f \in [-f_{\mathcal{N}}, f_{\mathcal{N}}]$ should ideally equal the sample variance. The sample variance for the rough ice profile is 44.95. Numerical integrations of the periodogram and the smoothed periodogram in Figure 323(c) agree with this value; integration of the SDF corresponding to the smoothed log periodogram yields 19.41, i.e., 43% of the desired value. If we attempt to correct for bias in the smoothed log periodogram by displacing it upwards by $10 \log_{10}(\mathrm{e})\,\gamma \doteq 2.5$ dB, (see the discussion surrounding Figure 305(b)), the smoothed periodogram and the bias-corrected smoothed log periodogram agree better visually (not shown in the plots), but numerical integration of the SDF corresponding to the latter yields 34.57, which is 77% of the desired value. Similar results hold for the smooth ice profile. The sample variance is 0.2345, which is in agreement with integration of the periodogram and smoothed periodogram in Figure 323(d); by contrast, integration of the SDFs corresponding to the smoothed log periodogram without and with bias correction yield 0.1150 and 0.2048, i.e., 49% and 87% of the desired value. These examples illustrate the fact that smoothing the log periodogram can yield a corresponding SDF estimate that does not constitute a proper frequency-based analysis of variance.

Figures 323(e) and (f) show discretely smoothed periodograms objectively chosen using the method described in Section 7.10. For a given m, the smoothed periodograms $\hat{S}^{(\mathrm{DSP})}_{R,m}(\cdot)$ and $\hat{S}^{(\mathrm{DSP})}_{S,m}(\cdot)$ are defined in accordance with Equation (307), where the smoothing coefficients $g_{m,j}$ are proportional to $\exp(-(mj)^2)$ and obey the constraint $\sum_j g_{m,j} = 1$. We set m to the minimizer of GCV(m) of Equation (309a) (however, we adjust the lower limit of the summation there to be $k = 1$ so as to avoid taking the log of $\hat{S}^{(\mathrm{P})}_R(0)$ and $\hat{S}^{(\mathrm{P})}_S(0)$, both of which are zero due to centering the time series). This gives

$$m \doteq \begin{cases} 0.0459, & \text{for the rough profile;} \\ 0.0712, & \text{for the smooth profile,} \end{cases}$$

which yield the dark smooth curves $\hat{S}^{(\mathrm{DSP})}_{R,0.0459}(\cdot)$ and $\hat{S}^{(\mathrm{DSP})}_{S,0.0712}(\cdot)$ shown in plots (e) and (f). For comparison with the Parzen lag window estimates shown in plots (a) to (d), we compute the bandwidth measures for the discretely smoothed periodograms, obtaining $B_{\mathcal{U}} \doteq 0.0278$ for the rough profile and $B_{\mathcal{U}} \doteq 0.0703$ for the smooth (see Equation (347b)). We then determine Parzen lag window estimates $\hat{S}^{(\mathrm{LW})}_{R,m}(\cdot)$ and $\hat{S}^{(\mathrm{LW})}_{S,m}(\cdot)$ with m chosen so that the associated bandwidth measures for these estimates are as close as possible to the measures for the discretely smoothed periodograms. This gives

$$m = \begin{cases} 38, & \text{for the rough profile;} \\ 15, & \text{for the smooth profile,} \end{cases} \tag{324}$$

which are comparable to the selections stated in Equation (323) that yield the Parzen lag window estimates shown in plots (c) and (d). The estimates $\hat{S}^{(\mathrm{DSP})}_{R,0.0459}(\cdot)$ and $\hat{S}^{(\mathrm{LW})}_{R,38}(\cdot)$ are almost the same, differing by less than 0.24 dB across the Fourier frequencies; likewise, $\hat{S}^{(\mathrm{DSP})}_{S,0.0712}(\cdot)$ and $\hat{S}^{(\mathrm{LW})}_{S,15}(\cdot)$ differ overall by less than 0.13 dB.

The crisscrosses in the lower left-hand corners of Figures 323(a) to (f) show, as usual, the bandwidth measure $B_{\mathcal{U}}$ (Equation (256a) for plots (a) to (d) and Equation (347b) for (e) and (f)) and the length in decibels of 95% CIs for the true SDFs at a given frequency (these

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

	Rough Profile			Smooth Profile		
	(a)	(c)	(e)	(b)	(d)	(f)
$B_{\mathcal{U}}$ (cycles/meter)	0.0122	0.0302	0.0278	0.0536	0.0964	0.0703
95% CI Length (dB)	3.54	2.22	2.31	3.33	2.46	2.89

Table 325 Lengths of horizontal and vertical components of crisscrosses in Figure 323.

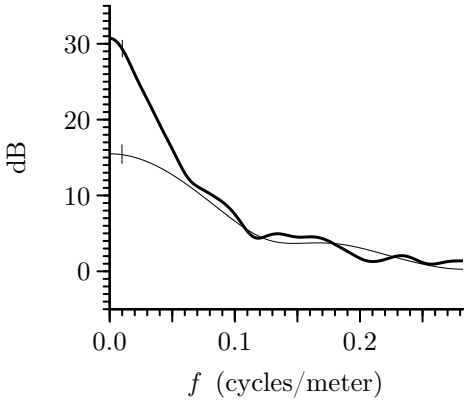

Figure 325 Parzen lag window estimates $\hat{S}^{(\mathrm{LW})}_{R,35}(\cdot)$ (dark curve, which is the same as the one in Figure 323(c)) and $\hat{S}^{(\mathrm{LW})}_{S,11}(\cdot)$ (light curve, which reproduces the one in Figure 323(d), but with an upward displacement of 13 dB).

length are dictated by Equation (266b) with EDOFs η computed using Equation (264b) with $C_h = 1$ for plots (a) to (d) and using Equation (347a) for (e) and (f)). Table 325 lists $B_{\mathcal{U}}$ and the 95% CI lengths for the six smoothed periodograms. The CI lengths range from 2.22 to 3.54 dB, which are nearly two orders of magnitude smaller than the length associated with an unsmoothed periodogram (21.6 dB). The smoothed periodograms for the smooth series look similar, but, for the rough series, the smoothed periodogram in Figure 323(a) looks noticeably bumpier than the ones in (c) and (e). All six SDF smoothed periodograms capture the germane patterns in the corresponding raw periodograms.

Figure 325 compares the Parzen lag window estimates $\hat{S}^{(\mathrm{LW})}_{R,35}(\cdot)$ and $\hat{S}^{(\mathrm{LW})}_{S,11}(\cdot)$ of Figures 323(c) and (d). The thick solid curve shows $\hat{S}^{(\mathrm{LW})}_{R,35}(\cdot)$ while the thin curve is $\hat{S}^{(\mathrm{LW})}_{S,11}(\cdot)$ after it has been moved up by 13 dB – this corresponds to multiplying $\hat{S}^{(\mathrm{LW})}_{S,11}(\cdot)$ by a factor of 20. With this adjustment, there is reasonably good agreement between the two SDF estimates from about 0.05 cycles/meter up to the Nyquist frequency; however, for frequencies less than 0.05 cycles/meter, $\hat{S}^{(\mathrm{LW})}_{R,35}(\cdot)$ and $20\,\hat{S}^{(\mathrm{LW})}_{S,11}(\cdot)$ diverge. Vertical lines emanating from the lag window estimates at $f = 0.01$ cycles/meter indicate what 95% CIs would be when based upon the unadjusted and adjusted lag window estimates. These lines show that the difference between the two spectral estimates is much larger than can be reasonably explained by their inherent variability. Based upon these qualitative results, we can conclude that the simple model $S_R(f) = cS_S(f)$ is not viable.

Atomic Clock Data

Our third example involves the differences in times as kept by two atomic clocks. Figure 326(a) shows these time differences after adjustment for some isolated so-called phase

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

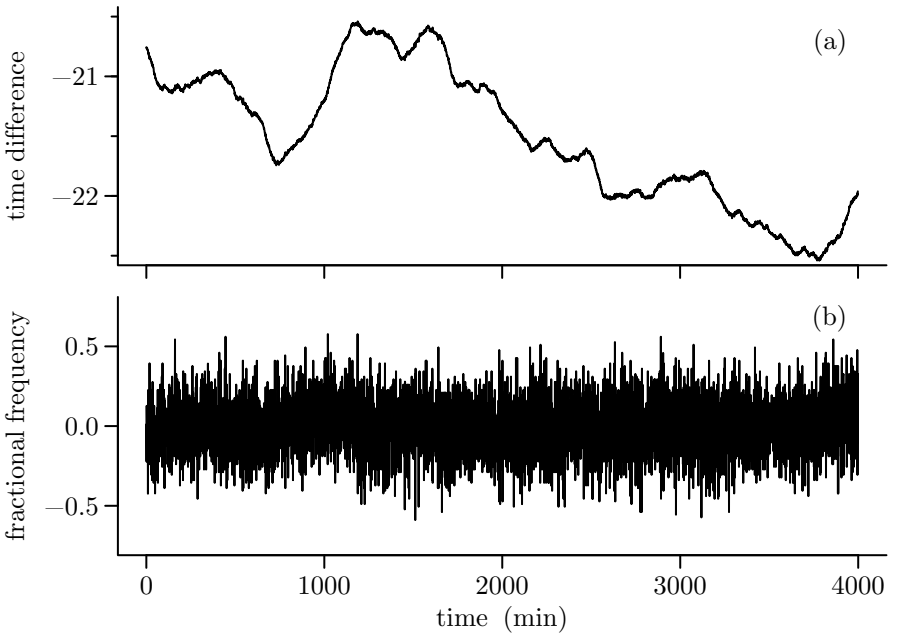

Figure 326 Atomic clock data (courtesy of L. Schmidt and D. Matsakis, US Naval Observatory). Plot (a) shows the time differences $\{X_t\}$ between two atomic clocks after removing a small number of phase glitches, while plot (b) shows the corresponding fractional frequency deviates $\{Y_t\}$ after multiplication by 10^{12}. The time differences are measured in nanoseconds, while the fractional frequency deviates are unitless.

glitches, which are small anomalous steps in time that are easy to detect and that do not reflect the inherent timekeeping ability of these clocks. The time differences are measured in nanoseconds (i.e., 10^{-9} of a second) at a sampling interval of $\Delta_{\mathrm{t}} = 1$ min. Since there are $N = 4000$ measurements in all, the total span of the data is close to 2.8 days. An atomic clock keeps time by counting the number of oscillations from an embedded frequency standard, which is a hydrogen maser for both clocks under study. This standard oscillates at a nominal frequency. If the standard were to generate oscillations that never deviated from its nominal frequency, the clock could keep time perfectly. The joint imperfection in the frequency standards embedded in the two atomic clocks is reflected by their fractional frequency deviates. If we consider the time differences to be a portion of a realization of a stochastic process $\{X_t\}$, the fractional frequency deviates are a realization of the first difference of this process divided by Δ_{t}, but only after we reexpress the latter in the same units as X_t (i.e., $\Delta_{\mathrm{t}} = 60 \times 10^9$ nanoseconds). These deviates are thus given by

$$Y_t \stackrel{\text{def}}{=} \frac{X_t - X_{t-1}}{\Delta_{\mathrm{t}}}$$

and are shown in Figure 326(b) after multiplication by 10^{12} to simplify labeling the vertical axis. We can regard $\{Y_t\}$ as a filtered version of $\{X_t\}$.

▷ **Exercise [326]** Appeal to Equations (143d), (143e) and (143b) to claim that, if $\{X_t\}$ possesses an SDF $S_X(\cdot)$, then $\{Y_t\}$ has an SDF given by

$$S_Y(f) = \frac{4 \sin^2(\pi f \, \Delta_{\mathrm{t}})}{\Delta_{\mathrm{t}}^2} S_X(f). \qquad \triangleleft$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

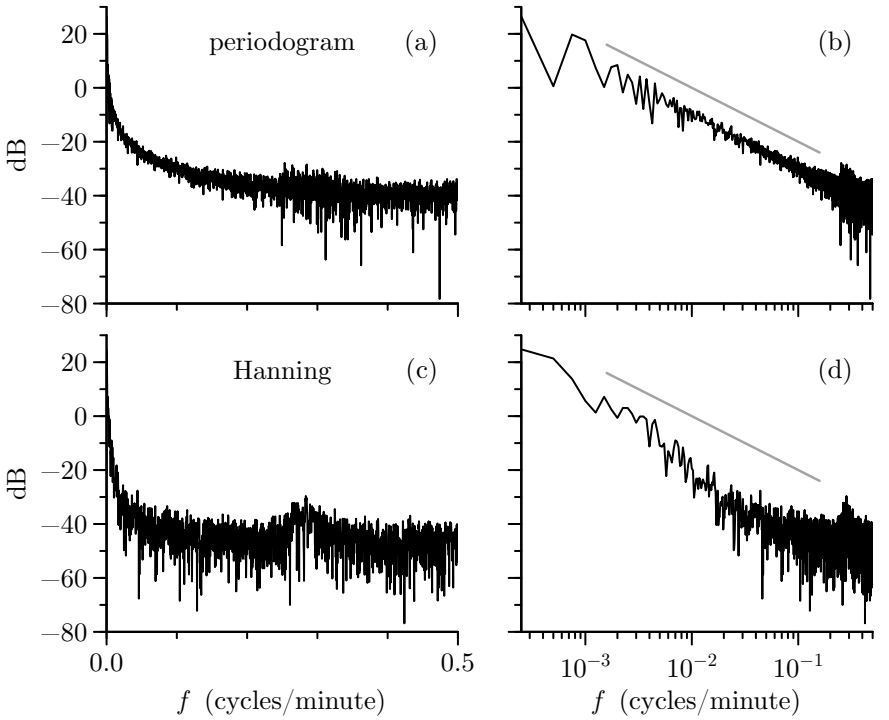

Figure 327 Periodogram (top row of plots) and direct spectral estimate based upon a Hanning data taper (bottom row) for atomic clock time differences. The left-hand column shows the SDF estimates in decibels versus a linear frequency scale, whereas the right-hand column show the same estimates, but versus a logarithmic scale. The gray lines in the right-hand column indicate the rolloff of an SDF proportional to f^{-2}.

Let us first consider estimation of the SDF $S_X(\cdot)$ for the time differences $\{X_t\}$. The upper row of plots in Figure 327 shows the periodogram $\hat{S}_X^{(\mathrm{P})}(\cdot)$ versus the positive Fourier frequencies on both a decibel versus linear scale (plot (a)) and a decibel versus log scale (b). The bottom row shows corresponding plots for a direct SDF estimate $\hat{S}_X^{(\mathrm{D})}(\cdot)$ based upon the Hanning data taper $\{h_t\}$ (Equation (189a) with $p = 1$; experimentation with other tapers indicates that the Hanning taper is a good choice here). To form this estimate, we centered $\{X_t\}$ by subtracting off its sample mean $\overline{X} \doteq -21.5$ and normalized $\{h_t\}$ such that $\sum h_t^2 = 1$. A comparison of $\hat{S}_X^{(\mathrm{P})}(\cdot)$ with $\hat{S}_X^{(\mathrm{D})}(\cdot)$ shows that there is substantial leakage in the periodogram. The leakage here is potentially misleading. A common practice amongst analysts of atomic clocks is to model the SDF $S_X(\cdot)$ for $\{X_t\}$ over selected ranges of frequencies in terms of canonical power-law processes. A power-law process dictates that $S_X(f) = C|f|^\alpha$. Restricting our attention to $f > 0$, this model implies that $\log_{10}(S_X(f)) = \log_{10}(C) + \alpha \log_{10}(f)$ and hence that, when $\log_{10}(S_X(f))$ is plotted versus $\log_{10}(f)$, we should see a line with a slope dictated by the power-law exponent α. One canonical choice for the power-law exponent is $\alpha = -2$, which corresponds to so-called random-walk phase noise (other canonical choices are 0, -1, -3 and -4; see Barnes et al., 1971). Since a decibel scale is merely a relabeling of a log scale, the format of Figure 327(b) allows placement of a gray line to depict the pattern for the SDF of a power-law process with $\alpha = -2$. The rolloff rate corresponding to this process is in good agreement with that exhibited by the periodogram over low frequencies ($f \leq 0.1$ cycles/minute); however, as Figure 327(d) shows, this agreement vanishes when the periodogram is replaced by a leakage-reduced direct SDF estimate. The fact that the

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

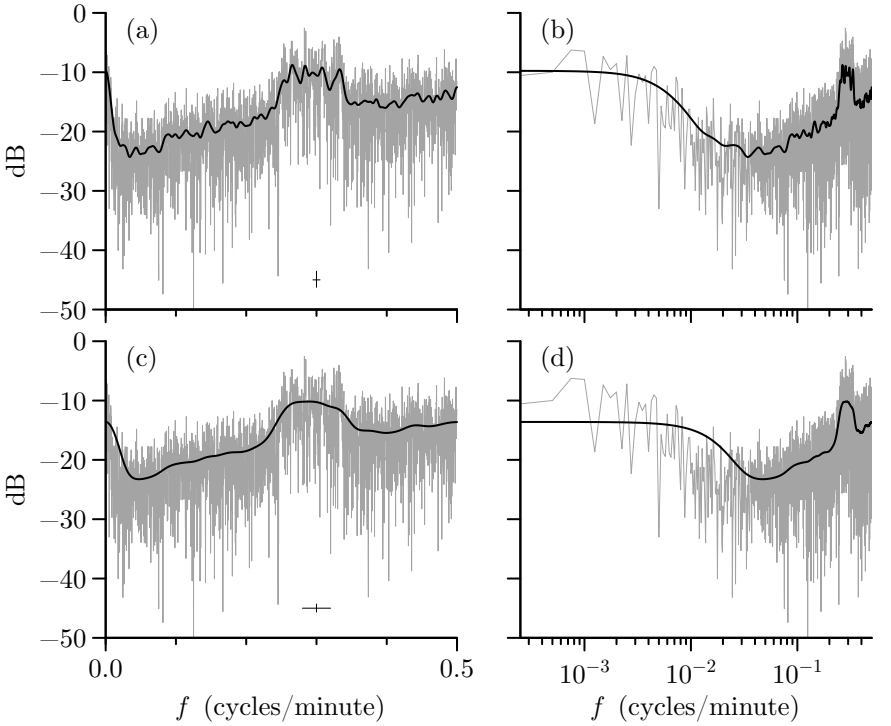

Figure 328 Periodogram for atomic clock fractional frequency deviates (light choppy curves) along with corresponding Gaussian lag window estimates (dark curves) with $m = 80$ (top row) and 20 (bottom). The horizontal (frequency) axes are linear in the left-hand column and logarithmic in the right. The crisscrosses in plots (a) and (c) show the bandwidth measure $B_{\mathcal{U}}$ (Equation (256a)) and the length of 95% CIs based upon the lag window estimates.

periodogram falsely suggests that a power-law process with $\alpha = -2$ might be an appropriate model for $\{X_t\}$ over low frequencies is attributable to its spectral window, which is Fejér's kernel. This window has sidelobes whose envelope decays at a rate of approximately f^{-2} (see the discussion surrounding Figure 270, where Fejér's kernel turns up as the smoothing window associated with the Bartlett lag window). Figure 327(d) suggests that, at frequencies lower than about 0.05 cycles/minute, the true SDF for $\{X_t\}$ decays at a faster rate than f^{-2}. Hence the overall pattern exhibited by the periodogram is really due to Fejér's kernel rather than the true SDF.

Let us now look at the estimation of the SDF $S_Y(\cdot)$ for the fractional frequency deviates $\{Y_t\}$. Figure 328 shows the periodogram $\hat{S}_Y^{(\mathrm{P})}(\cdot)$ in decibels versus the positive Fourier frequencies (light choppy curves), both on linear (left-hand column) and log (right-hand) scales. In contrast to $\hat{S}_X^{(\mathrm{P})}(\cdot)$, here the periodogram shows no evidence of leakage, an assessment based on comparing it to a Hanning-based direct SDF estimate (not shown). Because the periodogram is so erratic, it is helpful to compute associated Gaussian lag window estimates (Equation (276)) to better see the overall pattern of the SDF. Two such estimates are shown in Figure 328 as dark curves, the one in the top row with m set to 80, and in the bottom row, to 20. In contrast to Figure 319, here Parzen lag window estimates (not shown) are virtually identical to the displayed Gaussian lag window estimates once the lag window parameter has been set to achieve similar smoothing window bandwidths B_W. There are two prominent peaks evident in all the estimates, a narrow one centered at $f = 0$ cycles/minute, and a broad one at $f = 0.28$ cycles/minute. The bandwidth of the lag window estimate in the top row is narrow enough to preserve the narrow low-frequency peak, but at the cost of introducing

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

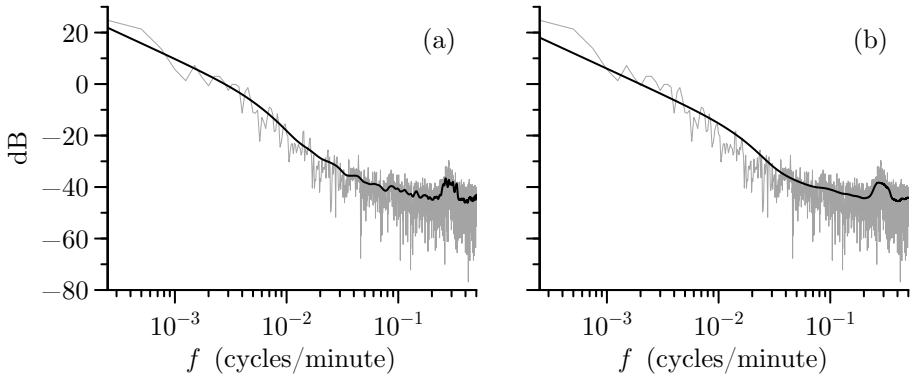

Figure 329 Hanning-based direct spectral estimate $\hat{S}_X^{(\mathrm{D})}(\cdot)$ for atomic clock time differences (light choppy curves, a reproduction of the dark choppy curve in Figure 327(d)), along with postcolored estimates $\hat{S}_X^{(\mathrm{PC})}(\cdot)$ based upon Gaussian lag window estimates $\hat{S}_{Y,m}^{(\mathrm{LW})}(\cdot)$ (thick curves). For plots (a) and (b), the underlying lag window estimates are the dark curves shown in, respectively, Figures 328(b) and (d).

unappealing bumpiness elsewhere; by contrast, the wider bandwidth in use in the bottom row yields a much smoother looking SDF estimate overall, but blurs out the low-frequency peak. This example illustrates that, because its bandwidth is constant across frequencies, a single lag window estimator cannot be easily adjusted to uniformly handle an SDF with peaks having quite different bandwidths.

Since the fractional frequency deviates $\{Y_t\}$ are the result of filtering the time differences $\{X_t\}$ with a rescaled first difference filter, and since the dynamic range of $S_Y(\cdot)$ is evidently markedly smaller than that of $S_X(\cdot)$, we can regard the first difference filter as a prewhitening filter. This suggests that we can estimate $S_X(\cdot)$ based on a lag window estimator $\hat{S}_{Y,m}^{(\mathrm{LW})}(\cdot)$ using the postcoloring procedure described in Section 6.5, thus obtaining the estimator

$$\hat{S}_X^{(\mathrm{PC})}(f) \stackrel{\mathrm{def}}{=} \frac{\Delta_{\mathrm{t}}^2}{4\sin^2(\pi f \Delta_{\mathrm{t}})} \hat{S}_{Y,m}^{(\mathrm{LW})}(f). \tag{329}$$

Figures 329(a) and (b) show postcolored estimates (dark curves) based upon, respectively, the $m = 80$ and 20 Gaussian lag window estimates shown in Figures 328(b) and (d). Both postcolored estimates are overlaid on the Hanning-based direct SDF estimate $\hat{S}_X^{(\mathrm{D})}(\cdot)$ (light choppy curves) shown in Figure 328(d). The postcolored estimates appear to be smoother versions of $\hat{S}_X^{(\mathrm{D})}(\cdot)$, with the differences between the $\hat{S}_X^{(\mathrm{PC})}(\cdot)$'s being due to the bandwidth properties of the underlying lag window estimates. The close agreement illustrates the fact that tapering and prewhitening are alternative ways of creating SDF estimates that do not suffer from severe bias as does the periodogram. (C&E [3] discusses a subtle issue that arises in interpreting the postcolored estimator $\hat{S}_X^{(\mathrm{PC})}(\cdot)$, which calls into question the assumption that $\{X_t\}$ is a stationary process.)

As a concluding comment, we note that Barnes et al. (1971) and numerous subsequent papers about the analysis of clock data advocate the use of power-law processes to model both time differences and fractional frequency deviates. Such processes are reasonable models if SDF estimates, when plotted on a log/log scale, appear to obey a piecewise linear pattern, with each line spanning at least a decade of frequencies; i.e., letting $[f_{\mathrm{L}}, f_{\mathrm{H}}]$ denote an interval over which linearity hold, we desire $f_{\mathrm{H}}/f_{\mathrm{L}} \geq 10$. While there is compelling evidence that atomic clocks prevalent in the 1970s tended to have SDFs with this pattern, Figures 327(d), 328(b), 328(d) and 329 call into question piecewise linearity as a useful characterization for atomic

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

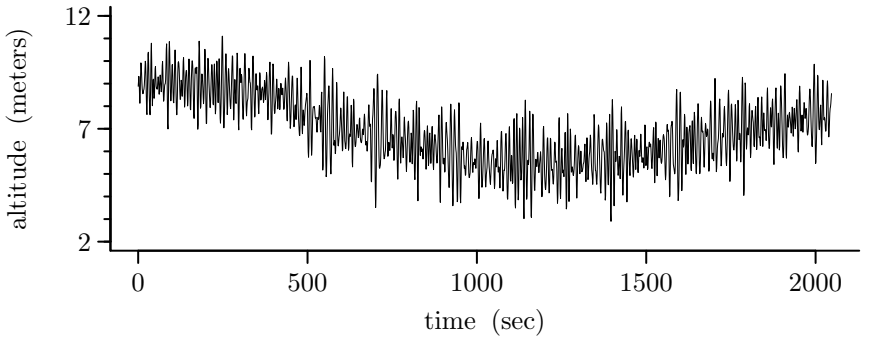

Figure 330 Ship altitude data (courtesy of J. Mercer, Applied Physics Laboratory, University of Washington). There are N = 2048 samples taken Δ_t = 1 sec apart.

clocks of recent vintage. This example shows the need to carefully assess through spectral analysis the appropriateness of power-law processes as models for particular time series.

Ship Altitude Data

As our fifth example, we consider a time series collected on an oceanographic cruise in open waters of the Pacific Ocean (Mercer et al., 2009). To carry out certain experiments, the ship was maneuvered to stay at a fixed latitude and longitude (lat/long), but moderate winds and seas hampered this goal. The ship's lat/long position was monitored using the Global Positioning System, which also recorded the ship's altitude as a function of time. Figure 330 shows $N = 2048$ measurements of altitude collected at $\Delta_t = 1$ sec intervals (the series thus spans a little over 34 min). Prominent low- and high-frequency variations are visible – the former are due in part to tides and currents, and the latter, to swells. Understanding the nature of this time series is of interest to help interpret various oceanographic measurements collected aboard the ship.

We begin our analysis by considering the periodogram. Figure 331(a) shows $\hat{S}^{(\mathrm{P})}(\cdot)$ on a decibel scale over the grid of positive Fourier frequencies. In contrast to the periodograms shown in Figures 226(a), 327(a) and 229(a), here we do not see obvious indications of inhomogeneous variability across frequencies, which is a tip-off that the periodogram might suffer from leakage. Figure 331(b) shows a corresponding direct spectral estimate $\hat{S}^{(\mathrm{D})}(\cdot)$ based upon the Hanning data taper $\{h_t\}$ (Equation (189a), with $p = 1$ and C set such that $\sum_t h_t^2 = 1$). As usual, we centered the time series by subtracting off its sample mean prior to applying the Hanning taper. Here $\hat{S}^{(\mathrm{P})}(\cdot)$ and $\hat{S}^{(\mathrm{D})}(\cdot)$ have the same general pattern across frequencies, but there are some subtle differences that become more apparent once we reduce the inherent variability by applying a lag window. The light and dark curves in Figure 331(c) show lag window estimates based upon, respectively, $\hat{S}^{(\mathrm{P})}(\cdot)$ and $\hat{S}^{(\mathrm{D})}(\cdot)$. For both estimates, we used the Gaussian lag window of Equation (276) with the parameter m set to 80 (we chose this setting subjectively using the window-closing method described in Section 7.7). There are two intervals of frequencies over which we see systematic differences between the two lag window estimates: one at low frequencies, and the other at high. In both intervals, the estimate based upon the periodogram is higher than the one employing the Hanning-based direct spectral estimate. Both intervals trap low-power regions that are adjacent to high-power regions, so we can chalk up the discrepancy between the two estimates as being due to leakage in the periodogram. Comparison of the Hanning-based direct spectral estimate with ones using Slepian data tapers with $NW = 2$, 3, 4 and 5 (not shown) indicates that the Hanning-based estimate adequately ameliorates bias in the periodogram. The bias in the low-frequency region is close to 10 dB in places (i.e., a potentially serious discrepancy of an order of magnitude), whereas

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

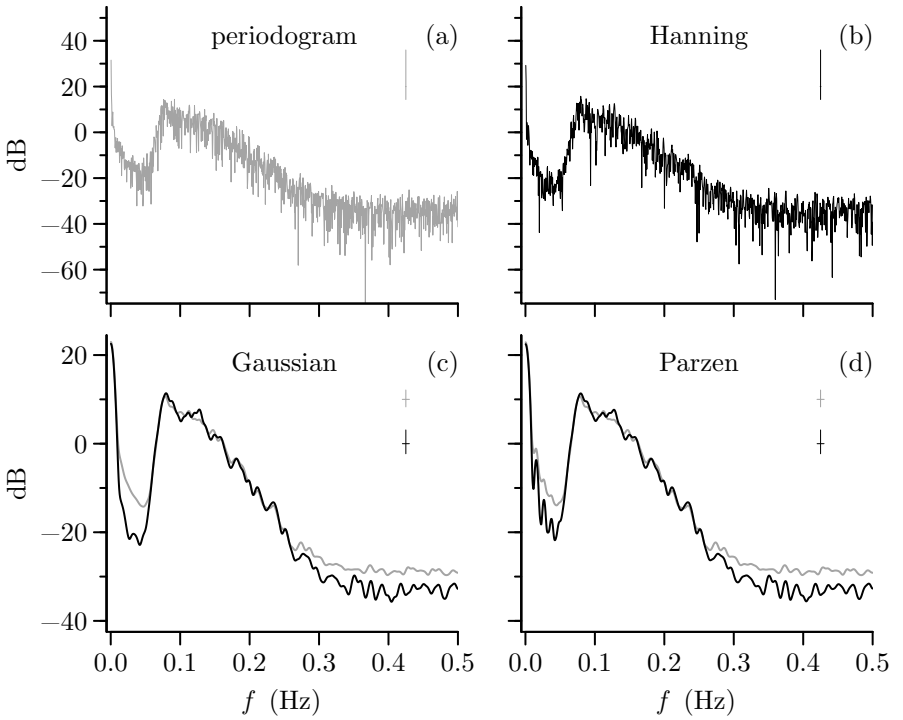

Figure 331 Periodogram (a) and a direct spectral estimate based upon a Hanning data taper (b) for the ship altitude data. Plot (c) shows two lag window estimates using a Gaussian lag window with parameter $m = 80$. The light and dark curves are smoothed versions of, respectively, the periodogram and the Hanning-based direct spectral estimate. Plot (d) is similar to (c), but now uses a Parzen lag window with parameter $m = 185$. The widths of the crisscrosses show either the bandwidth measure $B_{\mathcal{H}}$ of Equation (194) (plots (a) and (b)) or $B_{\mathcal{U}}$ of Equation (256a) ((c) and (d)); their heights show the length of a 95% CI for 10 $\log_{10}(S(f))$ (Equation (266a), with ν set to 2 in plots (a) and (b) and with ν set via Equations (264b) and (262) in (c) and (d)). In plots (c) and (d), the top and bottom crisscrosses are associated with, respectively, the periodogram-based and Hanning-based lag window estimates.

it is smaller in the high-frequency region (around 5 dB overall). This example of leakage contrasts in two ways with what we found for the ocean wave, atomic clock and chaotic beam data. First, leakage here is not confined to a single interval whose upper limit is the Nyquist frequency, but rather two distinct intervals. An SDF can theoretically have low power regions anywhere within the frequency range $[0, f_{\mathcal{N}}]$, opening up the possibility of leakage in disjoint intervals. Second, the variability in the periodogram for the ship altitude data appears to be consistent with what statistical theory dictates, whereas this was not the case with the other three examples. Apparent homogeneity of variance in the periodogram (when displayed on a log-based scale) is thus not a reliable indicator that the periodogram does not suffer from significant leakage.

Finally let us comment upon our choice of a Gaussian lag window. Figure 331(d) is similar to Figure 331(c) but is based on the Parzen lag window (Equation (275a)). Here we set the smoothing parameter m to 185, which forces the Parzen smoothing window bandwidth B_W to be the same as that of the Gaussian window (0.01 Hz – see Table 279). The Parzen lag window estimates are in good agreement with the Gaussian estimates over frequencies higher than about 0.05 Hz, but there are noticeable differences over the low-frequency interval $[0.01, 0.05]$. In particular, there is a pronounced ringing in the Parzen estimate associated with the Hanning data taper that is absent in the corresponding Gaussian estimate. This ringing is attributable to smoothing window leakage, which is caused by the sidelobes of the

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

Parzen smoothing window. The Gaussian estimate is free of such leakage, a fact that can be confirmed by comparing it with leakage-free Daniell and Bartlett–Priestley lag window estimates (not shown) with the smoothing parameter set such that $B_W = 0.01\,\text{Hz}$ also. Thus, whereas the Hanning data taper successfully reduces leakage in the periodogram, an inappropriate choice of a lag window can inadvertently introduce another form of leakage. As was also true for the ocean wave data, the Gaussian lag window estimate amalgamates the attractive properties of a Daniell or Bartlett–Priestley lag window estimate (leakage-free) and of a Parzen estimate (smoothness outside of intervals where smoothing window leakage is a concern). (The ship altitude data we have examined consists of the first 2048 values in a longer series of length $N = 14{,}697$ – Exercise [7.29] invites the reader to analyze the entire series. See also Section 11.6, which considers simulating series with statistical properties similar to those of the ship altitude series).

Atmospheric CO_2 Data

Our final example is a time series of monthly mean atmospheric CO_2 values (in parts per million by volume (ppmv)) from air samples collected at Mauna Loa, Hawaii, from March 1958 to February 2017. Ignoring the facts that the number of days in a month varies from 28 to 31 and that data for 7 of the 708 months are interpolated rather than directly measured, we regard this series as being collected at equally spaced times with a sampling interval of $\Delta_t = 1/12$ year. This time series has generated considerable interest. Keeling et al. (2009) note that it is the longest available continuous record of measured atmospheric CO_2 concentrations and argue that it is "... a precise record and a reliable indicator of the regional trend in the concentrations of atmospheric CO_2 in the middle layers of the troposphere" because of "... the favorable site location, continuous monitoring, and careful selection and scrutiny of the data." The light wiggly curve in Figure 333(a) shows a plot of the series and indicates an increasing trend upon which are superimposed seasonal variations. The dramatic upward trend is part of the scientific evidence that the concentration in atmospheric CO_2 is on the increase over the 59-year period depicted.

Because of the importance of this time series, it is of interest to consider statistical models for it as an aid in understanding its nature. Here we explore how far we can get with a simple model in which we assume the series to be a realization of a nonstationary process $\{X_t\}$ that is the sum of three components, namely, a low-order polynomial trend, seasonal variations and a stochastic noise process:

$$X_t = T_t + S_t + Y_t,$$

where $\{T_t\}$ is the trend, $\{S_t\}$ is a deterministic seasonal component (i.e., $S_{t+12} = S_t$ for all t) and $\{Y_t\}$ is the noise, which we assume to be stationary with zero mean and with SDF $S_Y(\cdot)$. The dark smooth curve in Figure 333(a) shows the result of fitting a quadratic trend, i.e.,

$$T_t = \alpha + \beta t + \gamma t^2,$$

to $\{X_t\}$, with the three parameters estimated by least squares. The estimated trend $\widehat{T}_t$ agrees reasonably well with the backbone of the data. With the trend defined in this manner, we could proceed to investigate the residuals $R_t \stackrel{\text{def}}{=} X_t - \widehat{T}_t$, a pathway we return to in Section 10.15. Here we take an alternative approach in which we employ linear time-invariant (LTI) filters to eliminate both the trend (assumed to be well-approximated by a quadratic polynomial) and the seasonal component. This approach is appealing because we can easily keep track of the effect of LTI filtering operations on the noise process, thus allowing qualitative evaluation of simple hypotheses about its SDF. Because subtracting off a linear least squares fit is not an LTI filter operation, it is harder to deduce the spectral properties of $\{Y_t\}$ from $\{R_t\}$ (in addition the residuals do not encapsulate just the noise process, but the seasonal component as well).

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

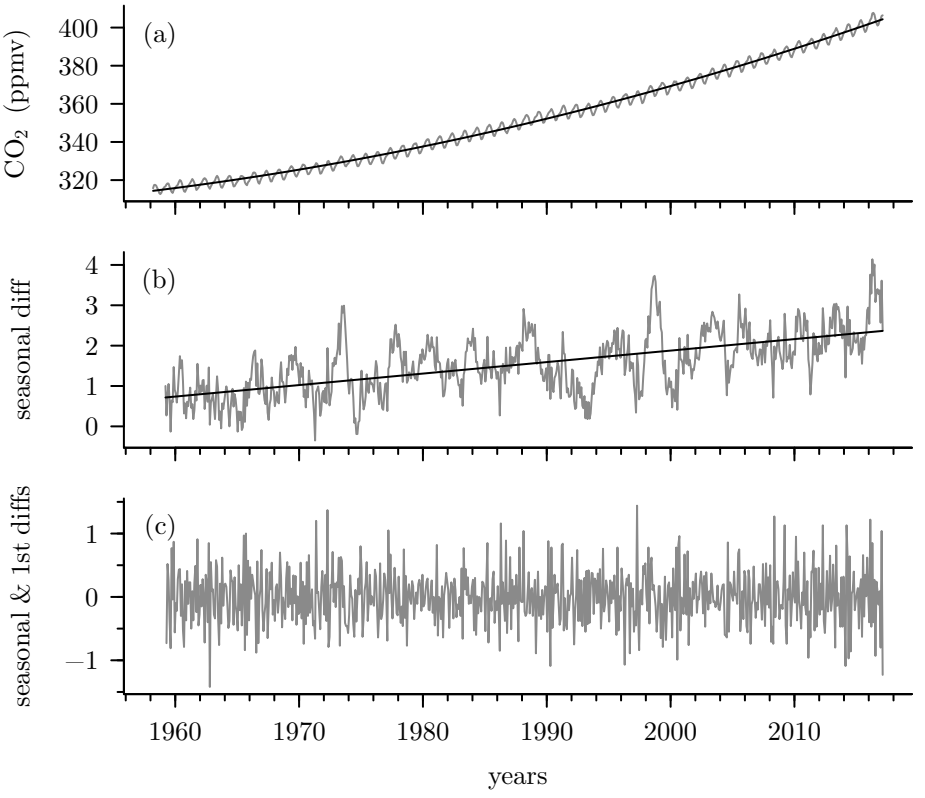

Figure 333 Atmospheric CO_2 data (P. Tans, Earth System Research Laboratory, National Oceanic & Atmospheric Administration, and R. Keeling, Scripps Institution of Oceanography – see the "Data" part of the website for the book for details on how to obtain these data). The light wiggly line in plot (a) depicts monthly values $\{X_t\}$ of atmospheric CO_2 concentration at Mauna Loa, Hawaii, from March 1958 to February 2017, while the dark smooth curve is a quadratic polynomial fit using least squares. The similarly displayed parts of plot (b) are the result $\{X_t^{(1)}\}$ of applying a lag $p = 12$ seasonal differencing filter to $\{X_t\}$ and a line fit via least squares. Plot (c) shows the result $\{X_t^{(2)}\}$ of taking $\{X_t^{(1)}\}$ and subjecting it to a first difference filter.

The task at hand is thus to find LTI filters that eliminate a quadratic trend and the seasonal component. As discussed in Section 2.8, we can remove seasonality in data having a period of $p = 12$ using a lag p seasonal differencing operation (Box et al., 2015; Brockwell and Davis, 2016). This procedure yields

$$X_t^{(1)} \stackrel{\text{def}}{=} X_t - X_{t-p} = T_t^{(1)} + Y_t^{(1)},$$

where $T_t^{(1)} \stackrel{\text{def}}{=} T_t - T_{t-p}$ and $Y_t^{(1)} \stackrel{\text{def}}{=} Y_t - Y_{t-p}$. For a quadratic trend we have

$$T_t^{(1)} = \alpha + \beta t + \gamma t^2 - \big(\alpha + \beta(t-p) + \gamma(t-p)^2\big) = (\beta - \gamma p)p + 2\gamma p t \stackrel{\text{def}}{=} \alpha' + \beta' t;$$

i.e., a lag p seasonal differencing filter reduces a quadratic trend to a linear trend. As demonstrated by Equations (39) and (40), we can reduce a linear trend to a constant by applying a first difference filter, leading us to consider

$$X_t^{(2)} \stackrel{\text{def}}{=} X_t^{(1)} - X_{t-1}^{(1)} = T_t^{(2)} + Y_t^{(2)} = \beta' + Y_t^{(2)},$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

where $T_t^{(2)} \stackrel{\text{def}}{=} T_t^{(1)} - T_{t-1}^{(1)}$ and $Y_t^{(2)} \stackrel{\text{def}}{=} Y_t^{(1)} - Y_{t-1}^{(1)}$. The light rough curve in Figure 333(b) shows $\{X_t^{(1)}\}$. Under our assumed model, this series should be the sum of a linear trend and a realization of a stationary process that has been subjected to a lag p seasonal differencing filter. The dark line in Figure 333(b) is what we get by fitting a line to $\{X_t^{(1)}\}$ via least squares. This line appears to be a reasonable summary of the upward pattern in the series. Figure 333(c) shows $\{X_t^{(2)}\}$, the result of subjecting $\{X_t^{(1)}\}$ to a first difference filter. According to our model, this series is a realization of a stationary process with mean β' and with an SDF given by

$$S_{X^{(2)}}(f) = |G_1(f)|^2 \cdot |G_{12}(f)|^2 S_Y(f), \tag{334a}$$

where $G_1(\cdot)$ and $G_{12}(\cdot)$ are the transfer functions defined as per Equation (143e) and associated with, respectively, the first difference and lag $p = 12$ seasonal differencing filters. Their associated squared gain functions are both special cases of the one arising in Equation (158), which, upon adjusting for the fact that $\Delta_t \neq 1$ here, gives

$$|G_1(f)|^2 = 4\sin^2(\pi f\, \Delta_t) \;\text{ and }\; |G_{12}(f)|^2 = 4\sin^2(12\pi f\, \Delta_t).$$

Figure 335(a) shows $|G_1(f)|^2$ (dashed curve), $|G_{12}(f)|^2$ (dark solid curve) and their product (light curve) versus f on a linear/decibel scale. The squared gain function $|G_1(\cdot)|^2$ is equal to zero at zero frequency, whereas $|G_{12}(\cdot)|^2$ is zero for $f = k$ cycles per year, $k = 0, 1, \ldots, 6$ – hence their product is also equal to zero at integer-valued frequencies.

We now consider a spectral analysis of $\{X_t^{(2)}\}$, with the idea of using Equation (334a) to deduce the nature of $S_Y(\cdot)$, the unknown SDF for the noise process. Figure 335(b) shows the periodogram $\hat{S}^{(\mathrm{P})}_{X^{(2)}}(f)$ versus f (light curve) and a corresponding Parzen lag window estimate $\hat{S}^{(\mathrm{LW})}_{X^{(2)},m}(f)$ with lag window parameter $m = 185$ (dark curve), which reduces the variability inherent in the periodogram (comparisons with direct spectral estimates using Hanning and Slepian data tapers indicate that the periodogram suffices for our purposes here). Equation (334a) suggests estimating $S_Y(f)$ via

$$\hat{S}_Y(f) \stackrel{\text{def}}{=} \frac{\hat{S}^{(\mathrm{LW})}_{X^{(2)},m}(f)}{|G_1(f)|^2 \cdot |G_{12}(f)|^2};$$

however, this estimate is problematic because it is infinite at integer-valued frequencies. Instead we take an approach whereby we hypothesize an SDF for $\{Y_t\}$, use Equation (334a) to determine the corresponding SDF for $\{X_t^{(2)}\}$ and then compare this theoretical SDF with the estimate $\hat{S}^{(\mathrm{LW})}_{X^{(2)},m}(\cdot)$. If the comparison is favorable, we have informal evidence in favor of the hypothesized SDF.

Let us start with a very simple hypothesis, namely, that $\{Y_t\}$ is a white noise process with unknown variance σ_Y^2, which says that its SDF is given by $S_Y(f) = \sigma_Y^2\, \Delta_t$. Under this hypothesis, the variances of $\{X_t^{(2)}\}$ and $\{Y_t\}$ are related by $\sigma^2_{X^{(2)}} = 4\sigma_Y^2$ (an easy exercise, but this result also follows from Equation (42) within Exercise [2.8]). We can use the sample variance $\hat{\sigma}^2_{X^{(2)}}$ for $\{X_t^{(2)}\}$ to estimate the unknown variance σ_Y^2 via $\hat{\sigma}_Y^2 = \hat{\sigma}^2_{X^{(2)}}/4$, from which we get the SDF estimate $\hat{S}_Y(f) = \hat{\sigma}_Y^2\, \Delta_t$. Plugging this estimate into Equation (334a) yields

$$\hat{S}_{X^{(2)}}(f) \stackrel{\text{def}}{=} |G_1(f)|^2 \cdot |G_{12}(f)|^2 \hat{S}_Y(f). \tag{334b}$$

This estimate is shown as the light curve in Figure 335(c) along with the Parzen lag window estimate (dark curve). If we use the crisscross to assess the sampling variability in $\hat{S}^{(\mathrm{LW})}_{X^{(2)},m}(\cdot)$, we see significant mismatches between this estimate and $\hat{S}_{X^{(2)}}(\cdot)$ over a considerable range of

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

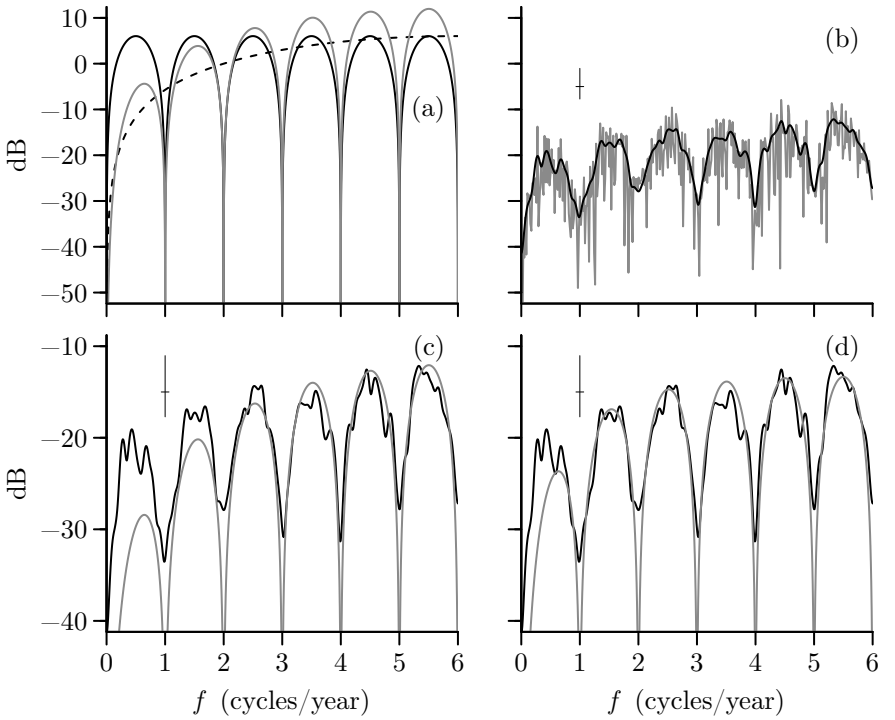

Figure 335 Squared gain functions, spectral estimates and hypothesized SDFs for CO_2 data. Plot (a) shows the squared gain functions for the lag $p = 12$ seasonal differencing filter (dark solid curve) and the first difference filter (dashed curve), along with their product (light curve). Plot (b) shows the periodogram (light bumpy curve) for $\{X_t^{(2)}\}$, i.e., the CO_2 data after subjection to both seasonal and first differencing. Superimposed upon the periodogram is a corresponding Parzen lag window estimate (dark curve) with smoothing window parameter $m = 185$. The width of the crisscross shows the bandwidth measure $B_{\mathcal{U}}$ (Equation (256a)) for the Parzen estimate, while its height is the length of a 95% CI for 10 $\log_{10}(S_{X^{(2)}}(f))$ (Equation (266a), with $\nu \doteq 14$). The curve for the Parzen estimate and the crisscross are reproduced in plots (c) and (d). The light curve in (c) shows a theoretical SDF for $\{X_t^{(2)}\}$ under the hypothesis that the underlying noise process $\{Y_t\}$ for the CO_2 data is white noise, while the one in (d) assumes noise dictated by a first-order autoregressive process.

low frequencies. This informal test suggests that the hypothesis that $\{Y_t\}$ is white noise is not tenable – we need to consider some form of colored noise. (There are also mismatches at the integer-valued frequencies, but these are of less concern: the bandwidth of the Parzen estimate – indicated by the width of the crisscross in Figure 335(c) – is wider than the narrow dips in $\hat{S}_{X^{(2)}}(\cdot)$, so we cannot expect the Parzen estimate to faithfully capture these narrow-band features.)

Accordingly, let us now entertain the hypothesis that $\{Y_t\}$ is a first-order autoregressive (AR(1)) process, which is a popular model for colored noise in climate research (von Storch and Zwiers, 1999). In keeping with Equation (144c), we presume $Y_t = \phi Y_{t-1} + \epsilon_t$, where $|\phi| < 1$ is an unknown parameter, and $\{\epsilon_t\}$ is a white noise process with mean zero and unknown variance σ_ϵ^2. We need to estimate ϕ and σ_ϵ^2 somehow based upon the observed filtered series $\{X_t^{(2)}\}$. A simple approach is to obtain expressions for ϕ and σ_ϵ^2 in terms of the lag $\tau = 0$ and $\tau = 1$ elements of the ACVS $\{s_{X^{(2)},\tau}\}$ for $\{X_t^{(2)}\}$. We can then use these expressions in conjunction with the standard estimators $\hat{s}^{(\mathrm{P})}_{X^{(2)},0}$ and $\hat{s}^{(\mathrm{P})}_{X^{(2)},1}$ to estimate ϕ and σ_ϵ^2. Two applications of Equation (42) yield

$$s_{X^{(2)},0} = 4s_{Y,0} - 4s_{Y,1} + 2s_{Y,11} - 4s_{Y,12} + 2s_{Y,13}$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

and

$$s_{X^{(2)},1} = -2s_{Y,0} + 4s_{Y,1} - 2s_{Y,2} + s_{Y,10} - 2s_{Y,11} + s_{Y,12} - 2s_{Y,13} + s_{Y,14}.$$

Since the ACVS for $\{Y_t\}$ is given by $s_{Y,\tau} = \phi^{|\tau|}\sigma_\epsilon^2/(1-\phi^2)$ (see Equation (44c)), we have

$$s_{X^{(2)},0} \approx \frac{4\sigma_\epsilon^2}{1+\phi} \text{ and } s_{X^{(2)},1} \approx -\frac{2\sigma_\epsilon^2(1-\phi)}{1+\phi},$$

where the approximations involve dropping terms ϕ^τ such that $\tau \geq 10$ and hence are good as long as $|\phi|$ is not too close to unity. Solving for ϕ and σ_ϵ^2 leads to the estimators

$$\hat{\phi} \stackrel{\text{def}}{=} \frac{2\hat{s}^{(\mathrm{P})}_{X^{(2)},1}}{\hat{s}^{(\mathrm{P})}_{X^{(2)},0}} + 1 \text{ and } \hat{\sigma}_\epsilon^2 \stackrel{\text{def}}{=} \frac{\hat{s}^{(\mathrm{P})}_{X^{(2)},0}(1+\hat{\phi})}{4}.$$

Setting $p = 1$ and $\phi_{1,1} = \phi$ in Equation (145a) gives us the SDF for an AR(1) process, which, when adjusted for Δ_t not being unity, yields

$$S_Y(f) = \frac{\sigma_\epsilon^2\,\Delta_\mathrm{t}}{|1-\phi\,\mathrm{e}^{-\mathrm{i}2\pi f\,\Delta_\mathrm{t}}|^2} \text{ and the estimator } \hat{S}_Y(f) = \frac{\hat{\sigma}_\epsilon^2\,\Delta_\mathrm{t}}{|1-\hat{\phi}\,\mathrm{e}^{-\mathrm{i}2\pi f\,\Delta_\mathrm{t}}|^2}.$$

Applying this procedure to the CO_2 data, we obtain the estimates $\hat{\phi} \doteq 0.36$ and $\hat{\sigma}_\epsilon^2 \doteq 0.064$, from which we can then form $\hat{S}_Y(\cdot)$. Using this SDF estimate in Equation (334b) gives us the imputed SDF $\hat{S}_{X^{(2)}}(\cdot)$ for the filtered process $\{X_t^{(2)}\}$ under the hypothesis that the noise process $\{Y_t\}$ is an AR(1) process. The light curve in Figure 335(d) shows this imputed SDF, which we need to compare with the lag window estimate $\hat{S}^{(\mathrm{LW})}_{X^{(2)},m}(\cdot)$ (dark curve). While there is markedly better agreement between $\hat{S}^{(\mathrm{LW})}_{X^{(2)},m}(\cdot)$ and the imputed SDF here than under the white noise hypothesis (Figure 335(c)), there are still significant mismatches at low frequencies (i.e., less than about 0.5 cycles/year) as assessed by the crisscross. We can conclude that a potentially important part of the covariance structure of the noise process is not being captured by an AR(1) model.

To conclude, this example demonstrates that spectral analysis is useful in assessing the effect of LTI filtering operations on a time series. If our operating model for the series has additive trend, seasonal and noise components, removal of the first two by operations other than filtering (e.g., least squares) can make it difficult to determine the SDF for the noise component. For the CO_2 data, we found an AR(1) model to be adequate except at capturing certain low-frequency fluctuations, where the model underestimates the contribution due to these frequencies. To compensate for this underestimation, our spectral analysis suggests that an alternative model worth exploring is one exhibiting a power law over low frequencies. We introduced the notion of power-law processes in our discussion of the atomic clock data. They come into play here because processes obeying a power law at low frequencies can attach more importance to those frequencies than an AR(1) process is capable of doing.

Comments and Extensions to Section 7.12

[1] The usual way we have plotted SDFs and their estimates is on a decibel scale versus frequency on a linear scale, i.e., 10 $\log_{10}(S(f))$ versus f (as examples, see Figures 319, 323 and 327(c)). Occasionally we have utilized a linear/linear scale (Figures 172 and 267(a)) and a decibel/log scale (Figures 327(b) and (d) and Figure 329). Figure 338a illustrates some strengths and weaknesses of these three ways of displaying an SDF and, for completeness, shows a fourth possibility we have not considered so far,

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

namely, a linear/log scale (plot (b)). All four plots in this figure depict the same SDF $S(\cdot)$, which is the sum of the SDF shown in Figure 125(c) and of the SDF for an AR(1) process dictated by the model $X_t = 0.95X_{t-1} + \epsilon_t$, where $\{\epsilon_t\}$ is a white noise process with zero mean and variance 0.25 (for convenience, we take the sampling interval Δ_t to be unity so that the Nyquist frequency is $f_{\mathcal{N}} = 0.5$). Plot (a) is on a linear/linear scale and clearly shows that the SDF has a prominent low-frequency component in addition to twin peaks located near 0.11 and 0.14 cycles per unit time. Since the integral of $S(\cdot)$ over $[-f_{\mathcal{N}}, f_{\mathcal{N}}]$ is equal to the variance of the underlying stationary process, the area under the curve in plot (a) is equal to half the process variance. This plot clearly shows that low-frequency fluctuations contribute more to the overall variance than do fluctuations attributable to either one of the twin peaks individually. By contrast, the usual decibel/linear plot (c) shows an additional peak near $f = 0.35$, which is insignificant on the linear/linear plot. The ability to showcase interesting features in an SDF that are overshadowed by more dominant components is a strength of a decibel/linear plot, but a downside is that we must exercise discipline when using this plot to interpret the relative importance of various portions of the SDF: visually it appears that the peak at $f = 0.35$ is of greater importance as a contributor to the overall variance than actually is the case. (When plotting a nonparametric SDF estimate rather than a theoretical SDF, an additional advantage to a decibel/linear scale is the ease with which we can display pointwise confidence intervals for the true unknown SDF using a crisscross, as is demonstrated in Figure 267.) The decibel/log scale of plot (d) allows us to study how the SDF varies as f decreases to zero while still pulling out the small peak at $f = 0.35$ in addition to the twin peaks. This particular SDF flattens out with decreasing frequency, an aspect of $S(\cdot)$ not readily apparent in plots (a) and (c). In addition, we noted in our discussion of the atomic clock data that a linear feature in an SDF displayed on decibel/log scale indicates that a power-law process might be a useful characterization for the SDF over a certain range of frequencies. In plot (d) there is a linear feature at low frequencies. Its slope is zero, thus indicating that $S(\cdot)$ varies as a power-law process with exponent zero (i.e., white noise) as f decreases to zero. Depending on what we wish to emphasize, some combination of plots (a), (c) and (d) might be appropriate to use.

Figure 338a(b) shows the SDF on a linear/log scale, which we have not found occasion to use previously. Such a plot might be of interest if there were fine structure in the SDF at low frequencies that is jammed together on a linear/linear plot. This is not the case in Figure 338a(b), so there is no clear rationale for advocating this plot over (a), (c) or (d). In particular, if we want to point out which frequencies are the dominant contributors to the overall variance, it would be better to use Figure 338a(a) rather than Figure 338a(b) because the latter is open to the misinterpretation that low-frequency fluctuations are much more important than those coming from the twin peaks. There is, however, an interesting variation on a linear/log plot called a *variance-preserving SDF* that seeks to prevent such a misinterpretation and that has found use in practical applications (see, for example, Emery and Thomson, 2001). For a stationary process $\{X_t\}$ possessing an SDF $S(\cdot)$ and a sampling interval of unity, the fundamental relationship

$$\text{var}\,\{X_t\} = \int_{-1/2}^{1/2} S(f)\,\mathrm{d}f = 2\int_0^{1/2} S(f)\,\mathrm{d}f$$

motivates plotting $S(f)$ versus $f \in [0, 1/2]$ with the interpretation that $S(f)\,\mathrm{d}f$ is the contribution to the process variance due to frequencies in a small interval about f. Letting λ denote an arbitrary nonzero constant, use of

$$\frac{\mathrm{d}(\lambda\,\log_{\mathrm{e}}(f))}{\mathrm{d}f} = \frac{\lambda}{f}$$

allows us to reexpress the fundamental relationship as

$$\text{var}\,\{X_t\} = \frac{2}{\lambda}\int_0^{1/2} fS(f)\,\mathrm{d}(\lambda\,\log_{\mathrm{e}}(f)).$$

The above motivates plotting the so-called variance-preserving SDF $fS(f)/\lambda$ versus $\lambda\,\log_{\mathrm{e}}(f)$ with the interpretation that $\frac{fS(f)}{\lambda}\,\mathrm{d}(\lambda\,\log_{\mathrm{e}}(f))$ also represents the contribution to the process variance due to frequencies in a small interval about f. Figure 338b shows the same SDF depicted in Figure 338a, but now in a variance-preserving manner (here we set $\lambda = \log_{10}(\mathrm{e})$, leading to $\lambda\,\log_{\mathrm{e}}(f) = \log_{10}(f)$). Since

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

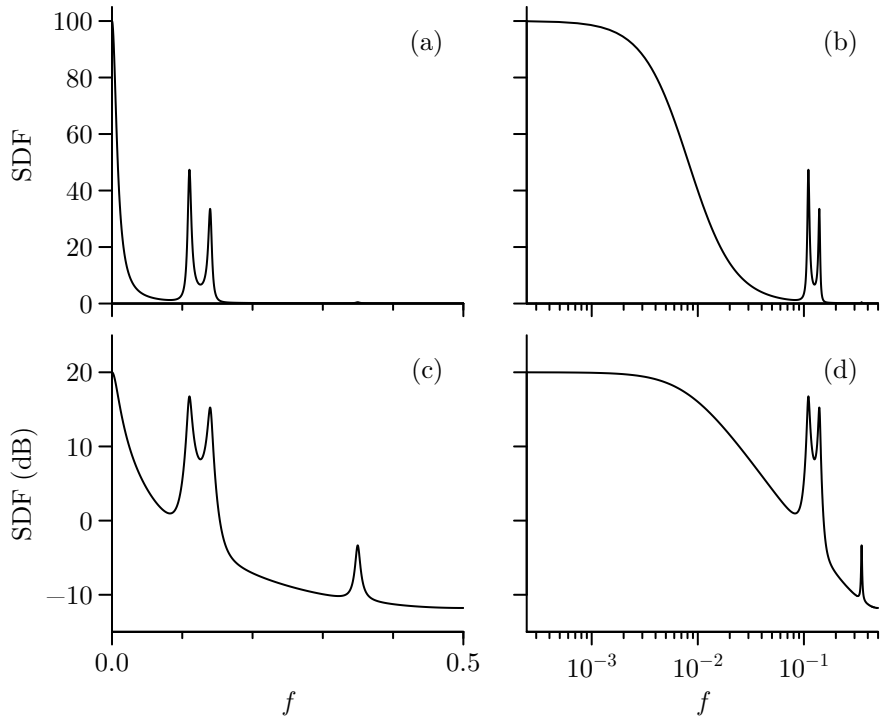

Figure 338a Four ways of plotting the same spectral density function $S(f)$. Plot (a) shows the SDF on a linear scale versus frequency, also on a linear scale; (b) uses a linear/log scale; (c), a decibel/linear scale; and (d), a decibel/log scale. The displayed SDF is the sum of three autoregressive SDFs, namely, ones corresponding to the AR(4) process of Equation (35a), the AR(2) process of Equation (125) and the AR(1) process defined by $X_t = 0.95X_{t-1} + \epsilon_t$, where $\{\epsilon_t\}$ is zero mean white noise process with variance 0.25.

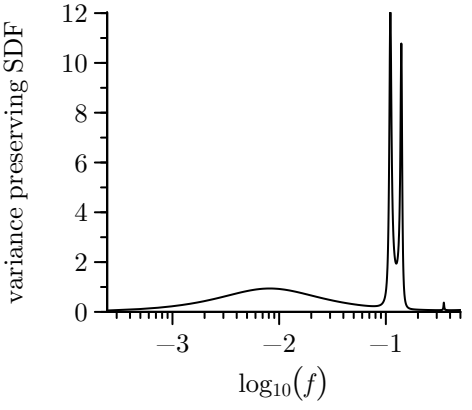

Figure 338b As in Figure 338a, but with the spectral density function displayed as a variance-preserving spectrum; i.e., $fS(f)/\log_{10}(e)$ is plotted versus $\log_{10}(f)$.

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

$fS(f)$ is virtually zero for values of $\log_{10}(f)$ smaller than what are shown in Figure 338b, the area under the curve depicted there is a tiny amount less than the area under the curve in Figure 338a(a), i.e., half of the process variance. Advocates of this type of plot might argue that it is to be preferred over the plot shown in Figure 338a(b) because, while both offer depictions against frequency on a log scale, the variance-preserving SDF allows us to see the relative contributions to the overall variance more clearly; however, the example we have put together illustrates an important caveat, namely, that distortions can arise, leading to an improper interpretation of the properties of the underlying SDF. Specifically, the variance-preserving SDF in Figure 338b distorts the low-frequency content of the SDF such that it appears as a broad peak centered near $\log_{10}(f) = -2.1$ (i.e., $f \doteq 0.008$). A practitioner might be tempted to regard this peak as evidence for a weak tendency in the process to have oscillations centered about this frequency, an interpretation that is not consistent with the actual SDF for the process. Because of the danger of similar misinterpretations (which, unfortunately, have occurred in the literature), we do not advocate use of the variance-preserving SDF when unaccompanied by one or more of the standard plots illustrated in Figure 338a (for example, the false impression created by Figure 338b could be corrected by considering it alongside Figure 338a(b), which shows that the SDF does not have an actual peak around $f \doteq 0.008$).

[2] In C&E [3] for Section 7.10 we noted an adaptation of the GCV criterion for use in determining the smoothing parameter m for a lag window estimator. Using this adaptation with the Parzen lag window and with the rough and smooth sea-ice profiles yields $m = 39$ for the former and 15 for the latter. These choices are almost identical to the choices $m = 38$ and 15 displayed in Equation (324) – these yield the Parzen lag window estimates that are the closest in bandwidth to those of the discretely smoothed periodograms of Figures 323(e) and (f). For these two time series the GCV criteria for discretely smoothed periodograms and for lag window estimators yield smoothed periodograms that are almost identical.

[3] Here we reconsider the postcolored SDF estimator $\hat{S}_X^{(\mathrm{PC})}(\cdot)$ for the atomic clock time differences $\{X_t\}$ (the thick curves in Figure 329 show two such estimates). As indicated by its definition in Equation (329), this estimator is the product of a lag window estimator and the squared gain function for a rescaled first difference filter. We can write the lag window estimator as

$$\hat{S}_{Y,m}^{(\mathrm{LW})}(f) = \Delta_{\mathrm{t}} \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau} \hat{s}_\tau^{(\mathrm{D})} \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_{\mathrm{t}}} = \Delta_{\mathrm{t}} \left(\hat{s}_0^{(\mathrm{D})} + 2 \sum_{\tau=1}^{N-1} w_{m,\tau} \hat{s}_\tau^{(\mathrm{D})} \cos(2\pi f \tau \Delta_{\mathrm{t}}) \right)$$

(see Equation (248a)). Since the cosine function is bounded and continuous with derivatives of all orders, this estimator must also be such since it is just a linear combination of cosines with different frequencies. In addition, its derivative at $f = 0$ must be zero (an easy exercise), which implies that we can approximate $\hat{S}_{Y,m}^{(\mathrm{LW})}(f)$ by a constant, say C, over a small interval of frequencies centered about zero. We can use the small angle approximation $\sin(x) \approx x$ to approximate the squared gain function over this same interval, yielding

$$\hat{S}_X^{(\mathrm{PC})}(f) \approx \frac{C}{4\pi^2 f^2}$$

for f close to zero. Thus $\hat{S}_X^{(\mathrm{PC})}(f) \to \infty$ as $|f| \to 0$ in a manner such that the integral of $\hat{S}_X^{(\mathrm{PC})}(f)$ over $[-f_{\mathcal{N}}, f_{\mathcal{N}}]$ necessarily diverges to infinity. A fundamental property of an SDF is that it integrates to the variance of the associated stationary process, which must be finite. Evidently the SDF estimator $\hat{S}_X^{(\mathrm{PC})}(\cdot)$ does not correspond to the SDF for a stationary process, which suggests that $\{X_t\}$ is not a stationary process. Note that we would *not* reach this same conclusion based upon the Hanning-based direct spectral estimate $\hat{S}_X^{(\mathrm{D})}(\cdot)$ (the light choppy curves in Figure 329), which necessarily integrates to a finite value. Even though Figure 329 shows $\hat{S}_X^{(\mathrm{PC})}(\cdot)$ and $\hat{S}_X^{(\mathrm{D})}(\cdot)$ to be similar over the displayed frequencies, there is a key difference between them: whereas $\hat{S}_X^{(\mathrm{PC})}(f)$ explodes as $|f| \to 0$, $\hat{S}_X^{(\mathrm{D})}(f)$ converges to a finite value (this follows from an argument similar to the one we used for $\hat{S}_{Y,m}^{(\mathrm{LW})}(\cdot)$, which says that $\hat{S}_X^{(\mathrm{D})}(f)$ is well approximated by a constant over a sufficiently small interval about zero).

Because $\hat{S}_X^{(\mathrm{PC})}(\cdot)$ does not integrate to a finite value, it is not a valid SDF; on the other hand, the lag window estimate $\hat{S}_{Y,m}^{(\mathrm{LW})}(\cdot)$ for its rescaled first difference does integrate finitely and is a valid

SDF. This suggests that, even if we cannot regard $\{X_t\}$ as a stationary process, we can take its first difference (or a rescaling thereof) to be such. Yaglom (1958) expands the definition of an SDF to include stochastic processes $\{X_t\}$ that are nonstationary, but whose first- or higher-order differences are stationary. Suppose that $\{X_t\}$ is a stochastic process whose dth-order difference

$$X_t^{(d)} \stackrel{\text{def}}{=} \sum_{k=0}^{d} \binom{d}{k} (-1)^k X_{t-k}$$

is a stationary process with SDF $S_{X^{(d)}}(\cdot)$, where d is a nonnegative integer (for convenience, we take the sampling interval Δ_t to be unity for $\{X_t\}$ and hence also for $\{X_t^{(d)}\}$). Thus $X^{(0)} = X_t$, $X^{(1)} = X_t - X_{t-1}$, $X^{(2)} = X_t - 2X_{t-1} + X_{t-2}$ and so forth. Note that $X^{(d)} = X_t^{(d-1)} - X_{t-1}^{(d-1)}$ for all $d \geq 1$. Now, if $\{X_t\}$ were a stationary process with SDF $S_X(\cdot)$, then Exercise [5.4a] and the notion of a cascaded filter (see Section 5.5) tell us that the SDFs for $\{X_t^{(d)}\}$ and $S_X(\cdot)$ would be related by

$$S_{X^{(d)}}(f) = 4^d \sin^{2d}(\pi f) S_X(f).$$

If $\{X_t\}$ is instead nonstationary, Yaglom (1958) uses the above relationship to *define* an SDF for it:

$$S_X(f) \stackrel{\text{def}}{=} \frac{S_{X^{(d)}}(f)}{4^d \sin^{2d}(\pi f)}.$$

Here the nonstationary process $\{X_t\}$ is said to have *stationary differences of order d*. In general a process with dth-order stationary differences can be either stationary or nonstationary. A related concept is an *intrinsically stationary process of order d*: $\{X_t\}$ is said to be such if $\{X_t\}, \{X_t^{(1)}\}, \ldots, \{X_t^{(d-1)}\}$ are all nonstationary, but $\{X_t^{(d)}\}$ is stationary (see Chilès and Delfiner, 2012, or Stein, 1999). While $S_X(\cdot)$ need not possess all of the properties of a proper SDF (in particular, integration to a finite value), it can – with care – be dealt with in the same manner as the SDF for a stationary process. This theory allows us to interpret the postcolored SDF estimator of Equation (329) as an SDF for a first-order intrinsically stationary process.

7.13 Summary of Lag Window Spectral Estimators

Suppose we have a time series that is a portion $X_0, X_1, \ldots, X_{N-1}$ of a real-valued zero mean stationary process $\{X_t\}$ with a sampling interval of Δ_t (and hence a Nyquist frequency of $f_{\mathcal{N}} = 1/(2\,\Delta_t)$) and with an autocovariance sequence (ACVS) $\{s_\tau\}$ and spectral density function (SDF) $S(\cdot)$ related by $\{s_\tau\} \longleftrightarrow S(\cdot)$. (If $\{X_t\}$ has an unknown mean, we must replace X_t with $X_t' = X_t - \overline{X}$ in all computational formulae, where $\overline{X} = \sum_{t=0}^{N-1} X_t/N$ is the sample mean.) Using a data taper $\{h_t\} \longleftrightarrow H(\cdot)$ with normalization $\sum_t h_t^2 = 1$, we start by forming a direct spectral estimator for $S(\cdot)$:

$$\hat{S}^{(\mathrm{D})}(f) \stackrel{\text{def}}{=} \Delta_t \left| \sum_{t=0}^{N-1} h_t X_t \mathrm{e}^{-\mathrm{i}2\pi f t\,\Delta_t} \right|^2 = \Delta_t \sum_{\tau=-(N-1)}^{N-1} \hat{s}_\tau^{(\mathrm{D})} \mathrm{e}^{-\mathrm{i}2\pi f \tau\,\Delta_t}, \quad \text{(see (186b) and (188a))}$$

where the ACVS estimator $\{\hat{s}_\tau^{(\mathrm{D})}\}$ is given by

$$\hat{s}_\tau^{(\mathrm{D})} \stackrel{\text{def}}{=} \begin{cases} \sum_{t=0}^{N-|\tau|-1} h_{t+|\tau|} X_{t+|\tau|} h_t X_t, & |\tau| \leq N-1; \\ 0, & |\tau| \geq N \end{cases} \qquad \text{(see (188b))}$$

(note that $\{\hat{s}_\tau^{(\mathrm{D})}\} \longleftrightarrow \hat{S}^{(\mathrm{D})}(\cdot)$). A lag window estimator $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ is the result of smoothing this direct spectral estimator across frequencies:

$$\hat{S}_m^{(\mathrm{LW})}(f) \stackrel{\text{def}}{=} \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} W_m(f-\phi) \hat{S}^{(\mathrm{D})}(\phi)\,\mathrm{d}\phi, \qquad \text{(see (248a))}$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

where $W_m(\cdot)$ is a smoothing window with a parameter m that controls the degree of smoothing to be imposed upon $\hat{S}^{(\mathrm{D})}(\cdot)$. The rationale for smoothing $\hat{S}^{(\mathrm{D})}(\cdot)$ is to obtain an estimator with reduced variance, which is desirable because direct spectral estimators are highly variable (see, e.g., the choppy curves in Figure 187). While the above equation emphasizes that $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ is a filtered version of $\hat{S}^{(\mathrm{D})}(\phi)$, actual computation of a lag window estimator makes use of the inverse Fourier transform of $W_m(\cdot)$, namely, $\{w_{m,\tau}\} \longleftrightarrow W_m(\cdot)$:

$$\hat{S}_m^{(\mathrm{LW})}(f) = \Delta_{\mathrm{t}} \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau} \hat{s}_\tau^{(\mathrm{D})} \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_{\mathrm{t}}} \qquad \text{(see (248a))}$$

(note that $\{w_{m,\tau}\hat{s}_\tau^{(\mathrm{D})}\} \longleftrightarrow \hat{S}_m^{(\mathrm{LW})}(\cdot)$). The sequence $\{w_{m,\tau}\}$ is known as a lag window, from which $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ gets its name. We assume that $w_{m,0} = 1$, that $w_{m,\tau} = 0$ for $|\tau| \geq N$ and that $w_{m,-\tau} = w_{m,\tau}$ for all τ. These conditions imply that the smoothing window $W_m(\cdot)$ is an even real-valued periodic function with a period of $2f_{\mathcal{N}}$ whose integral over $[-f_{\mathcal{N}}, f_{\mathcal{N}}]$ is unity. We can characterize the degree of smoothing that $W_m(\cdot)$ imparts to $\hat{S}^{(\mathrm{D})}(\cdot)$ using the smoothing window bandwidth measure given by

$$B_W \stackrel{\mathrm{def}}{=} \frac{1}{\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} W_m^2(f)\,\mathrm{d}f} = \frac{1}{\Delta_{\mathrm{t}} \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}^2}. \qquad \text{(see (251c) and (251e))}$$

A typical smoothing window resembles a bell-shaped curve (see Figure 259(b)), and B_W is a measure of the width of the curve.

Turning now to an examination of the statistical properties of $\hat{S}_m^{(\mathrm{LW})}(\cdot)$, we have

$$E\{\hat{S}_m^{(\mathrm{LW})}(f)\} = \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \mathcal{U}_m(f - f')S(f')\,\mathrm{d}f', \qquad \text{(see (255b))}$$

where

$$\mathcal{U}_m(f) \stackrel{\mathrm{def}}{=} \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} W_m(f - f'')\mathcal{H}(f'')\,\mathrm{d}f'' = \sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}\, h \star h_\tau \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_{\mathrm{t}}},$$

(see (255c) and (255d))

is the spectral window associated with $\hat{S}_m^{(\mathrm{LW})}(\cdot)$; in the above, $\mathcal{H}(\cdot)$ is the spectral window for the underlying direct spectral estimator $\hat{S}^{(\mathrm{D})}(\cdot)$ (by definition, this window is the square modulus of the Fourier transform $H(\cdot)$ of the data taper $\{h_t\}$ used to create $\hat{S}^{(\mathrm{D})}(\cdot)$), and $\{h \star h_\tau\}$ is the autocorrelation of $\{h_t\}$ (see Equation (192c)). Under appropriate conditions, by letting the smoothing window bandwidth B_W decrease towards zero as the sample size N increases to infinity, the spectral window $\mathcal{U}_m(\cdot)$ acts as a Dirac delta function, allowing us to claim that $\hat{S}_m^{(\mathrm{LW})}(f)$ is asymptotically an unbiased estimator of $S(f)$. With proper care in selecting the data taper and lag window, it is often possible for $E\{\hat{S}_m^{(\mathrm{LW})}(f)\}$ to be approximately equal to $S(f)$ for finite N, but the quality of the approximation for a particular N depends highly upon the complexity of the true SDF. We take the effective bandwidth $B_{\mathcal{U}}$ of $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ to be the autocorrelation width of its spectral window:

$$B_{\mathcal{U}} \stackrel{\mathrm{def}}{=} \mathrm{width_a}\,\{\mathcal{U}_m(\cdot)\} = \frac{\left(\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \mathcal{U}_m(f)\,\mathrm{d}f\right)^2}{\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \mathcal{U}_m^2(f)\,\mathrm{d}f} = \frac{\Delta_{\mathrm{t}}}{\sum_{\tau=-(N-1)}^{N-1} w_{m,\tau}^2 (h \star h_\tau)^2}. \qquad \text{(see (256a))}$$

Note that $B_{\mathcal{U}}$ depends upon both the lag window and the data taper.

As to the second-moment properties of $\hat{S}_m^{(\mathrm{LW})}(\cdot)$, under suitable conditions,

$$\mathrm{var}\,\{\hat{S}_m^{(\mathrm{LW})}(f)\} \approx \frac{C_h S^2(f)}{B_W N \Delta_{\mathrm{t}}}, \qquad \text{(see (259b))}$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

where $C_h \geq 1$ is a variance inflation factor due solely to tapering:

$$C_h = N \sum_{t=0}^{N-1} h_t^4. \qquad \text{(see (262))}$$

This factor is unity if and only if $h_t = 1/\sqrt{N}$ (the default data taper, which turns $\hat{S}^{(\mathrm{D})}(\cdot)$ into the periodogram). Table 260 gives large sample approximations to C_h for some commonly used data tapers. If $B_W \to 0$ and $N \to \infty$ in such a way that $B_W N \to \infty$, then var $\{\hat{S}_m^{(\mathrm{LW})}(f)\}$ decreases to 0, which, when coupled with asymptotic unbiasedness, says that $\hat{S}_m^{(\mathrm{LW})}(f)$ is a consistent estimator of $S(f)$. Also, for $f_j' = j/(N' \Delta_t)$, cov $\{\hat{S}_m^{(\mathrm{LW})}(f_j'), \hat{S}_m^{(\mathrm{LW})}(f_k')\} \approx 0$, where j and k are distinct integers such that $0 \leq j, k \leq \lfloor N'/2 \rfloor$; here $N' = N/c$ where $c \geq 1$ is a factor that depends primarily upon the bandwidth B_W of the smoothing window $W_m(\cdot)$ (see Section 7.3).

A large-sample approximation to the distribution of $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ is that of a scaled chi-square random variable:

$$\hat{S}_m^{(\mathrm{LW})}(f) \stackrel{\mathrm{d}}{=} \frac{S(f)}{\nu} \chi_\nu^2,$$

where ν is the equivalent degrees of freedom for $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ given by

$$\nu = \frac{2NB_W \Delta_t}{C_h} \qquad \text{(see (264d))}$$

(as ν gets large, $\hat{S}_m^{(\mathrm{LW})}(f)$ becomes asymptotically Gaussian distributed with mean $S(f)$ and variance $2S^2(f)/\nu$). The above leads to approximate $100(1-2p)\%$ pointwise confidence intervals (CIs) of the form

$$\left[\frac{\nu \hat{S}_m^{(\mathrm{LW})}(f)}{Q_\nu(1-p)}, \frac{\nu \hat{S}_m^{(\mathrm{LW})}(f)}{Q_\nu(p)} \right] \qquad \text{(see (265))}$$

for $S(f)$ and of the form

$$\left[10 \log_{10}(\hat{S}_m^{(\mathrm{LW})}(f)) + 10 \log_{10}\left(\frac{\nu}{Q_\nu(1-p)}\right), 10 \log_{10}(\hat{S}_m^{(\mathrm{LW})}(f)) + 10 \log_{10}\left(\frac{\nu}{Q_\nu(p)}\right) \right] \qquad \text{(see (266a))}$$

for $10 \log_{10}(S(f))$ (the width of the latter CI is independent of $\hat{S}_m^{(\mathrm{LW})}(f)$).

Specific lag windows we've considered in detail are the Bartlett (Equation (269a)), Daniell (Equation (273a)), Bartlett–Priestley (Equation (274)), Parzen (Equation (275a)), Gaussian (Equation (276)) and Papoulis (Equation (278)) windows, with the Gaussian having particularly attractive properties. The smoothing window bandwidth B_W for each of these windows is controlled by a smoothing window parameter m whose setting is critical for obtaining a $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ that is appropriate for a particular time series. A subjective method for setting m is window closing, which consists of visually comparing lag window estimates with different choices for m to the underlying $\hat{S}^{(\mathrm{D})}(\cdot)$ and picking m such that the resulting $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ best captures the salient features in $\hat{S}^{(\mathrm{D})}(\cdot)$. One objective method for setting m is to estimate the bandwidth B_S of the unknown SDF and to pick m such that $B_W = B_S/2$. The theory behind this method assumes that the unknown SDF is dominantly unimodal, which is often – but certainly not always – the case in practical applications. A second method is based upon smoothing an estimator of the log of the SDF rather than SDF itself, which allows for m to be set by minimizing an appealing mean integrated square error criterion. While this method yields acceptable results for certain SDFs, it is problematic in other cases. A third method uses generalized cross-validation to objectively pick the appropriate degree of smoothing for a discretely smoothed periodogram.

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

7.14 Exercises

[7.1] Show that Equation (247f) is true.

[7.2] Show that Equation (250b) is equivalent to $w_{m,0} = 1$.

[7.3] Show that Equation (255d) is true. Does this equation still hold if we drop our assumption that $w_{m,\tau} = 0$ for $|\tau| \geq N$? If we replace $W_m(\cdot)$ with $V_m(\cdot)$ in Equation (255c), why does this not change the definition of $\mathcal{U}_m(\cdot)$? (Section 7.1 describes the distinction between $W_m(\cdot)$ and $V_m(\cdot)$.)

[7.4] (a) As an extension to Equation (186g), show that

$$E\{\hat{S}_m^{(\mathrm{LW})}(f)\} = \Delta_\mathrm{t} \sum_{\tau=-(N-1)}^{N-1} \left(w_{m,\tau} \sum_{t=0}^{N-|\tau|-1} h_{t+|\tau|} h_t \right) s_\tau \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_\mathrm{t}}. \tag{343a}$$

(b) As extensions to Equations (188d) and (188e), show that

$$\int_{-f_\mathcal{N}}^{f_\mathcal{N}} E\{\hat{S}_m^{(\mathrm{LW})}(f)\}\,\mathrm{d}f = s_0 \quad \text{and} \quad \int_{-f_\mathcal{N}}^{f_\mathcal{N}} \mathcal{U}_m(f)\,\mathrm{d}f = 1$$

under the usual assumptions $\sum_t h_t^2 = 1$ and $w_{m,0} = 1$.

[7.5] Use the approach of Section 7.4 to show that the equivalent number of degrees of freedom for the spectral estimator defined by Equation (246b) is

$$\nu = \frac{2N}{C_h \sum_{\tau=-(N-1)}^{N-1} \left| \sum_{j=-M}^{M} g_j \mathrm{e}^{\mathrm{i}2\pi j \tau / N'} \right|^2}.$$

[7.6] Suppose $\hat{S}(f)$ is a nonparametric SDF estimator with a distribution that is well approximated by that of the RV $a\chi_\nu^2$, where, as usual, χ_ν^2 is a chi-square RV with ν degrees of freedom, and a is a constant. If $\hat{S}(f)$ is an unbiased estimator of the true SDF $S(f)$ at frequency f, what must a be equal to? Using this value for a, provide a brief argument that, if $S(f)$ is finite, then $\mathrm{var}\,\{\hat{S}(f)\} \to 0$ as $\nu \to \infty$.

[7.7] An early attempt to produce a spectral estimator with better variance properties than the periodogram was to truncate the summation in Equation (170d) at $m < N$ to produce

$$\hat{S}^{(\mathrm{TP})}(f) \stackrel{\mathrm{def}}{=} \Delta_\mathrm{t} \sum_{\tau=-m}^{m} \hat{s}_\tau^{(\mathrm{P})} \mathrm{e}^{-\mathrm{i}2\pi f \tau \Delta_\mathrm{t}}. \tag{343b}$$

The idea is to discard estimates of the ACVS for which there are relatively little data. This estimator is known as the *truncated periodogram*. Show that $\hat{S}^{(\mathrm{TP})}(\cdot)$ can be regarded as a special case of a lag window spectral estimator, and determine its lag window and its smoothing window. Argue that the truncated periodogram need *not* be nonnegative at all frequencies.

[7.8] Blackman and Tukey (1958, section B.5) discuss the following class of lag windows:

$$w_{m,\tau} = \begin{cases} 1 - 2a + 2a\cos\left(\pi\tau/m\right), & |\tau| \leq m; \\ 0, & |\tau| > m, \end{cases} \tag{343c}$$

where $m < N$ and a is a parameter in the range $0 < a \leq 1/4$. If we let $a = 0.23$, then, among all smoothing windows corresponding to lag windows of the above form, we obtain the smoothing window that minimizes the magnitude in the first sidelobe relative to the magnitude of the peak in the central lobe. This particular case is called the *Hamming lag window*. A slight variation is to let $a = 1/4$. The resulting lag window is called the *Hanning lag window*.

(a) Determine the smoothing window for the lag window of Equation (343c).

(b) Show that, if we use the rectangular data taper, then $\hat{S}_m^{(\mathrm{LW})}(f)$ can be expressed as a linear combination of the truncated periodogram of Equation (343b) at three frequencies.

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

[7.9] This exercise examines an assumption of approximate uncorrelatedness made in Bartlett (1950) as part of formulating the Bartlett lag window of Equation (269a) (some aspects of Exercise [6.27f] parallel this exercise). Let N_S denote a block size, and consider a time series $X_0, X_1, \ldots, X_{N-1}$ of sample size $N = 2N_S$ from which we construct two periodograms, one making use of the first block of N_S values, and the other, the second block (we assume $\Delta_t = 1$ for convenience):

$$\hat{S}_0^{(\mathrm{P})}(f) \stackrel{\mathrm{def}}{=} \frac{1}{N_S}\left|\sum_{t=0}^{N_S-1} X_t e^{-i2\pi ft}\right|^2 \quad \text{and} \quad \hat{S}_{N_S}^{(\mathrm{P})}(f) \stackrel{\mathrm{def}}{=} \frac{1}{N_S}\left|\sum_{t=N_S}^{2N_S-1} X_t e^{-i2\pi ft}\right|^2.$$

Suppose the series is a portion of a Gaussian AR(1) process $X_t = \phi X_{t-1} + \epsilon_t$ with an ACVS given by $s_\tau = \phi^{|\tau|}/(1-\phi^2)$, where $\phi = 0.8$, and $\{\epsilon_t\}$ is a zero mean Gaussian white noise process with variance $\sigma_\epsilon^2 = 1$ (see Exercise [2.17a]). With N_S set to 16, use Equation (243b) to obtain the covariances and variances needed to compute the correlation ρ_k between $\hat{S}_0^{(\mathrm{P})}(\tilde{f}_k)$ and $\hat{S}_{N_S}^{(\mathrm{P})}(\tilde{f}_k)$ at frequencies $\tilde{f}_k = k/(2N_S)$, $k = 0, 1, \ldots, N_S$ (i.e., a grid twice as fine as that of the Fourier frequencies). Plot ρ_k versus $\tilde{f}_k$. Create similar plots with N_S set to 64 and then to 256. Finally repeat this exercise with ϕ set to 0.9, then 0.99 and finally -0.99. Comment on how well $\hat{S}_0^{(\mathrm{P})}(\tilde{f}_k)$ and $\hat{S}_{N_S}^{(\mathrm{P})}(\tilde{f}_k)$ are approximately uncorrelated for the twelve pairs of ϕ and N_S under consideration.

[7.10] (a) Verify that the lag window $\{w_{m,\tau}\}$ of Equation (274) corresponds to the Bartlett–Priestley design window of Equation (273c).

(b) Using $V_m(\cdot)$ as a proxy for $W_m(\cdot)$ in Equation (251c), show that B_W is approximately $5/(6m\,\Delta_t) \doteq 0.83/(m\,\Delta_t)$ for the Bartlett–Priestley window (this approximation appears in Table 279). Explore the accuracy of this approximation by making use of Equation (251e).

(c) Using $V_m(\cdot)$ as a proxy for $W_m(\cdot)$ in Equation (251a), show that β_W is approximately $\sqrt{3/5}/(m\,\Delta_t) \doteq 0.77/(m\,\Delta_t)$ for the Bartlett–Priestley window (this approximation appears in Table 279). Explore the accuracy of this approximation by making use of Equation (251b).

[7.11] Show that, for the Parzen smoothing window of Equation (275b), $W_m(0) = 3m\,\Delta_t/4$ for even m and $W_m(0) = (1+3m^4)\,\Delta_t/(4m^3)$ for odd m.

[7.12] The Parzen lag window $\{w_{m,\tau}\}$ of Equation (275a) is defined such that $w_{m,\tau} = 0$ for $|\tau| \geq m$, where m is a positive integer we restricted to be such that $m \leq N$. With this restriction, the corresponding smoothing window $W_m(\cdot)$ of Equation (275b) does not depend on N. To emphasize this fact, denote this window henceforth as $W_{m,N\text{-free}}(\cdot)$; i.e., assuming $m \leq N$ and $w_{m,\tau} = 0$ for all $|\tau| \geq m$, we have $\{w_{m,\tau} : \tau \in \mathbb{Z}\} \longleftrightarrow W_{m,N\text{-free}}(\cdot)$.

Now consider the case $m > N$. With the smoothing window defined as per Equation (247e), show that

$$W_m(f) = \Delta_t\,(2N-1)\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} \mathcal{D}_{2N-1}(f'\Delta_t) W_{m,N\text{-free}}(f-f')\,\mathrm{d}f', \tag{344}$$

where $\mathcal{D}_{2N-1}(\cdot)$ is Dirichlet's kernel (Equation (17c)). Argue that the corresponding lag window estimator $\hat{S}_m^{(\mathrm{LW})}(f)$ can be expressed not only in the usual way per Equation (247f) as

$$\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} W_m(f-f')\hat{S}^{(\mathrm{D})}(f')\,\mathrm{d}f', \quad \text{but also as} \quad \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} W_{m,N\text{-free}}(f-f')\hat{S}^{(\mathrm{D})}(f')\,\mathrm{d}f'.$$

Create plots similar to those in Figure 276 with the only difference being that m is set to 100 rather than 37 (recall that $N = 64$, and hence we are in the $m > N$ case). For comparison with $W_m(\cdot)$, also plot $W_{m,N\text{-free}}(\cdot)$ in a manner similar to Figure 276(b). Comment upon your findings. (While we have focused on the Parzen lag window here, the key results of this exercise also hold for other windows defined such that $w_{m,\tau} = 0$ for all $|\tau| \geq m$. Two examples are the Bartlett and Papoulis windows (Equations (269a) and (278)).)

[7.13] Create lag windows appropriate for a time series of length $N = 64$ by autocorrelating a Slepian data taper with NW set to each of the following 161 values: 1, 1.05, ..., 8.95, 9 (C&E [5] for

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

Section 7.5 has a discussion on formulating a lag window by autocorrelating a data taper). For each of the resulting 161 lag windows, compute the smoothing window bandwidth B_W using Equation (251e). Using the setting for NW with B_W closest to 0.05, create plots similar to those comprising Figure 286. Comment upon your findings.

[7.14] (a) Compute and plot $E\{\hat{S}_m^{(\mathrm{LW})}(f)\}$ versus f for the five lag window estimators involved in Figures 289 and 291. For comparison, display also in each plot a reproduction of the thin curve in the lower left-hand plot of Figure 193, which depicts the expected value of the underlying direct spectral estimator $\hat{S}^{(\mathrm{D})}(\cdot)$ that involves the Hanning data taper and the AR(4) process of Equation (35a). Hints: use Equations (343a) and (186g) along with the procedure given in Exercise [6.31d] for forming the ACVS for the AR(4) process.

(b) Repeat part (a) for the Papoulis, autocorrelated 100% cosine and Slepian lag windows.

[7.15] Verify Equations (296c) and (296d). Hint: make use of Equation (302a).

[7.16] (a) Show that, if we take the PDFs $p(\cdot)$ and $q(\cdot)$ in the Kullback–Leibler (KL) discrepancy of Equation (297a) to correspond to those of the RVs $S_p(f)\chi^2_{\nu_p}/\nu_p$ and $S_q(f)\chi^2_{\nu_q}/\nu_q$, where $S_p(f) > 0$, $S_q(f) > 0$, $\nu_p > 0$ and $\nu_q > 0$, then

$$d(p(\cdot), q(\cdot)) = \frac{\nu_q}{2}\left[\frac{S_p(f)}{S_q(f)} - \log\left(\frac{S_p(f)}{S_q(f)}\right)\right] + \eta(\nu_p, \nu_q), \tag{345a}$$

where $\eta(\nu_p, \nu_q)$ is to be determined as part of the exercise (the forms for $p(\cdot)$ and $q(\cdot)$ can be deduced from Equation (297b)). Show that, if ν_p and ν_q are both equal to, say, ν, then the above becomes

$$d(p(\cdot), q(\cdot)) = \frac{\nu}{2}\left(\left[\frac{S_p(f)}{S_q(f)} - \log\left(\frac{S_p(f)}{S_q(f)}\right)\right] - 1\right).$$

Hint: Gradshteyn and Ryzhik (1980) state the following definite integrals:

$$\int_0^\infty x^a \mathrm{e}^{-bx}\,\mathrm{d}x = \frac{\Gamma(a+1)}{b^{a+1}}, \quad \text{when } a > -1 \text{ and } b > 0; \text{ and}$$

$$\int_0^\infty x^{c-1}\mathrm{e}^{-dx}\log(x)\,\mathrm{d}x = \frac{\Gamma(c)}{d^c}\left[\psi(c) - \log(d)\right], \quad \text{when } c > 0 \text{ and } d > 0,$$

where $\psi(\cdot)$ is the digamma function. (Note that, conditional on $\nu_p = \nu_q = \nu$, the KL discrepancy is the same to within a constant of proportionality $\nu/2$. The choice $\nu = 2$ conveniently sets this constant to be unity and leads to the measure given by Equation (297c).)

(b) Under the assumptions that $\hat{S}_m^{(\mathrm{LW})}(\tilde{f}_k) \stackrel{\mathrm{d}}{=} S(\tilde{f}_k)\chi^2_\nu/\nu$ for all k and that $\nu > 2$, verify Equation (297d). Hint: make use of Equation (302a).

(c) In addition to $d(p(\cdot), q(\cdot))$, Kullback and Leibler (1951) explore $d(q(\cdot), p(\cdot))$ as a measure of discrepancy. In our context, the measure inspires

$$\mathrm{LK}(m) \stackrel{\mathrm{def}}{=} \frac{1}{N+1}\sum_{k=0}^{N}\left[\frac{\hat{S}_m^{(\mathrm{LW})}(\tilde{f}_k)}{S(\tilde{f}_k)} - \log\left(\frac{\hat{S}_m^{(\mathrm{LW})}(\tilde{f}_k)}{S(\tilde{f}_k)}\right) - 1\right] \tag{345b}$$

as an alternative to $\mathrm{KL}(m)$; i.e., we swap $S(\tilde{f}_k)$ and $\hat{S}_m^{(\mathrm{LW})}(\tilde{f}_k)$ in Equation (297c). Assuming that $\hat{S}_m^{(\mathrm{LW})}(\tilde{f}_k) \stackrel{\mathrm{d}}{=} S(\tilde{f}_k)\chi^2_\nu/\nu$ for all $\tilde{f}_k$, determine $E\{\mathrm{LK}(m)\}$, and comment briefly on how it compares to $E\{\mathrm{KL}(m)\}$ (Equation (297d)).

(d) Kullback and Leibler (1951) also explore $d(p(\cdot), q(\cdot)) + d(q(\cdot), p(\cdot))$ as a symmetric measure of discrepancy that does not depend on the ordering of $p(\cdot)$ and $q(\cdot)$. This measure inspires consideration of

$$\mathrm{KLLK}(m) \stackrel{\mathrm{def}}{=} \mathrm{KL}(m) + \mathrm{LK}(m). \tag{345c}$$

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

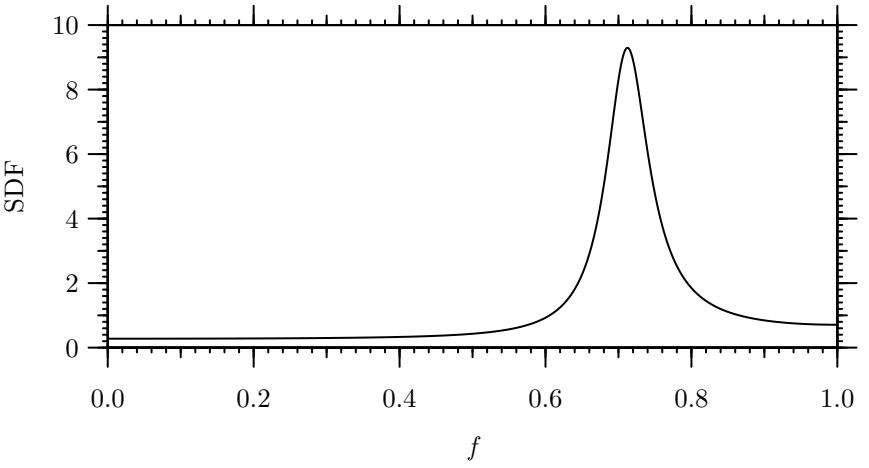

Figure 346 Hypothesized SDF. Note that $S(f)$ versus f is plotted on a *linear/linear* scale, not on the usual linear/decibel scale.

Formulate a succinct expression for KL(m) + LK(m), and determine its expected value by appealing to previous parts of this exercise.

[7.17] (a) For each of the four AR(2) time series displayed in plots (a) to (d) of Figure 34 (downloadable from the "Data" part of the website for the book), compute MSE(m), NMSE(m), MSLE(m), KL(m), LK(m) and KLLK(m) as per Equations (293), (296a), (296b), (297c), (345b) and (345c) for Parzen lag window estimates with m = 16, 32 and finally 64. Compare these measures to their corresponding expected values (see Equations (296d), (296c) and (297d) and parts (c) and (d) of Exercise [7.16]). Comment briefly on your findings.

(b) Generate a large number, say N_{R}, of simulated time series of length N = 1024 from the AR(2) process of Equation (34) (take "large" to be at least 1000 – see Exercise [597] for a description of how to generate realizations from this process). For each simulated series, compute MSE(m) for Parzen lag window estimates with m = 16, 32 and finally 64. Do the same for NMSE(m), MSLE(m), KL(m), LK(m) and KLLK(m). Create a scatterplot of the N_{R} values for NMSE(16) versus those for MSE(16) and then also for m = 32 and m = 64. Do the same for the remaining 14 pairings of the six measures. For a given m, compute the sample correlation coefficients corresponding to the 15 scatterplots. The sample correlation coefficients for one of the 15 pairings should be almost indistinguishable from unity for all three settings of m – offer an explanation as to why this happens. For each m, average all N_{R} values of MSE(m). Compare this average with its expected value. Do the same for the remaining five measures. Comment briefly on your findings.

[7.18] Suppose an investigator plans to collect a time series that can be modeled by a stationary process with SDF $S(\cdot)$. The sampling interval Δ_{t} is set at $1/2$ sec so the Nyquist frequency $f_{\mathcal{N}}$ is 1 Hz. Suppose that the rough shape of $S(\cdot)$ is known from either physical theory or previous experiments and is given by Figure 346. The investigator plans to use the Parzen lag window spectral estimator.

(a) What data taper would you recommend if the sample size N of the collected time series is 128? Would you change your recommendation if the sample size were increased to 1024? State the reasons for your recommendations.

(b) Would prewhitening be useful here? State the reasons for your answer.

(c) Determine the spectral bandwidth of $S(\cdot)$ by examining Figure 346. What does this imply about the size of the window bandwidth of the Parzen smoothing window? For the data taper(s) you selected in part (a) for sample sizes of N = 128 and 1024, determine what values of m (if any) achieve the desired window bandwidth. Determine the corresponding EDOFs ν.

[7.19] Assuming that f' = 0 and $W < f_{\mathcal{N}}$ in Equation (299) so that the SDF is that of an ideal low-pass rather than an ideal band-pass process, suitably modify the argument following that equation to

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

conclude that $2W$ is now approximately equal to $\eta/(N\,\Delta_t)$ rather than the value derived for the band-pass case, namely, $\eta/(2N\,\Delta_t)$. In doing so, make the assumption that $\hat{S}^{(P)}(0)$ is negligible (as would be the case for a centered time series – see Exercise [6.15b]). Repeat this exercise, but now assuming that $f' = f_{\mathcal{N}}$ so that the SDF is that of an ideal high-pass process (assume that $\hat{S}^{(P)}(f_{\mathcal{N}})$ is negligible when N is even).

[7.20] Show that, as $N \to \infty$,

$$B_T \to \frac{s_0^2}{2\int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} S^2(f)\,\mathrm{d}f} \stackrel{\text{def}}{=} B_T^{(\infty)},$$

where B_T is the time series bandwidth measure of Equation (300a) with δ_{BP} set to unity. What is $B_T^{(\infty)}$ when $S(\cdot)$ is the SDF for the ideal band-pass process of Equation (299)? Show via computations that $B_T^{(\infty)} \doteq 0.265$ for the AR(2) process of Equation (34) and $\doteq 0.033$ for the AR(4) process of Equation (35a). (The forthcoming Equations (508a) and (508b) state the ACVSs for these two processes.)

[7.21] Verify the two approximations stated in Equation (303d) under the assumption that N' is large.

[7.22] This exercise considers computational details and three measures associated with the discretely smoothed periodogram $\hat{S}_m^{(DSP)}(\cdot)$ of Equation (307).

(a) Sections 6.7 and 7.11 give details on how to efficiently compute direct spectral estimates (including the periodogram) and lag window estimates using discrete Fourier transforms. Provide similar details for computing $\hat{S}_m^{(DSP)}(\cdot)$ over the grid of Fourier frequencies.

(b) Starting with the obvious analog of Equation (264a), use an approach similar to what was used in Section 7.4 to argue that, by making certain assumptions (to be stated as carefully as possible as part of the exercise), we can take the EDOFs for $\hat{S}_m^{(DSP)}(f_k)$ to be

$$\nu = \frac{2}{\sum_{j=-\lfloor N/2 \rfloor}^{N-\lfloor N/2 \rfloor - 1} g_{m,j}^2}. \tag{347a}$$

How reasonable are the assumptions behind this expression for ν?

(c) Use an approach similar to the one leading to Equation (256a) to show that the standard measure for the effective bandwidth of $\hat{S}_m^{(DSP)}(\cdot)$ is

$$B_{\mathcal{U}} = \frac{1}{\Delta_t \sum_{\tau=-(N-1)}^{N-1} \left| \sum_{j=0}^{N-1} g_{m,j-\lfloor N/2 \rfloor} e^{i2\pi f_j \tau \Delta_t} \right|^2 \left(1 - \frac{|\tau|}{N}\right)^2}. \tag{347b}$$

Comment briefly upon how the above compares to Equation (256a).

(d) Use Equations (251c) and (258) to motivate the following as a definition for the smoothing window bandwidth associated with $\{g_{m,j}\}$:

$$\frac{1}{N\,\Delta_t \sum_{j=-\lfloor N/2 \rfloor}^{N-\lfloor N/2 \rfloor - 1} g_{m,j}^2}.$$

[7.23] Adapt Equations (347a) and (347b) to handle discrete smoothing of a Hanning-based direct SDF estimator $\hat{S}^{(D)}(\cdot)$ across the Fourier frequencies, i.e., an estimator formed as per the right-hand side of Equation (307), but with $\hat{S}^{(P)}(f_{k-j})$ replaced by $\hat{S}^{(D)}(f_{k-j})$. In adapting Equation (347a), assume that the correlation between $\hat{S}^{(D)}(f_k)$ and $\hat{S}^{(D)}(f_{k\pm l})$ is $1/2$ when $l = 1$ and is 0 for all other positive integers l of interest (Equation (241) suggests this approximation if we simplify $4/9$ and $1/36$ to $1/2$ and 0). Using weights $g_{m,j} \propto \exp(-(mj)^2)$ with $m = 0.2313$, verify the left-hand plot of Figure 311, including the crisscross (the required AR(4) time series is available via the website for the book). The crisscross depends in part on the value of ν from your adaptation of Equation (347a). By appealing to Table 279, use this value to determine a setting m for a Parzen lag window estimator such that the resulting EDOF agrees with ν as closely as possible. Using this setting for m, compute a Parzen lag window estimate for the AR(4) time series that also utilizes

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

the Hanning-based $\hat{S}^{(\mathrm{D})}(\cdot)$. How well does this estimate match up with the estimate shown in the left-hand plot of Figure 311?

[7.24] Verify the contents of the left-hand plot of Figure 310 and both plots in Figure 311 (the required AR(2) and AR(4) time series are available on the book's website). Verify that the GCV-based MSEs associated with the SDFs estimates depicted in these three plots are correctly stated in the discussion surrounding these plots (namely, 0.0562 for the left-hand plot of Figure 310, and 4.258 and 7.916 for the left- and right-hand plots in Figure 311). Verify that the ratios of the GCV-based MSEs to the associated best possible MSEs are also stated correctly (1.240, 1.0003 and 1.376, respectively). Generate a large number N_{R} of additional realizations of length $N = 1024$ of the AR(2) and AR(4) processes of Equations (34) and (35a), and compute the ratio of the GCV-based MSE to the best possible MSE for each of these realizations. Are the ratios associated with Figures 310 and 311 typical of what you obtained from the additional realizations? (For the above, take N_{R} to be at least 1000, and see Exercises [597] and [11.1] for descriptions of how to generate realizations of the AR(2) and AR(4) processes.)

[7.25] Suppose that, given a time series of length N, we have calculated its periodogram $\hat{S}^{(\mathrm{P})}(f_k)$ over the N Fourier frequencies $f_k = k/(N\,\Delta_{\mathrm{t}})$ satisfying $-f_{\mathcal{N}} \le f_k < f_{\mathcal{N}}$. We then use these periodogram values to form a discretely smoothed periodogram $\hat{S}_m^{(\mathrm{DSP})}(f_k)$ over these same frequencies using a set of weights $\{g_{m,j}\}$ (see Equation (307)). Let $\tilde{f}_k \stackrel{\mathrm{def}}{=} k/(2N\,\Delta_{\mathrm{t}})$, and note that $\tilde{f}_{2k} = f_k$. As discussed in C&E [2] for Section 7.10, we can take

$$\hat{s}_\tau^{(\mathrm{DSP})} \stackrel{\mathrm{def}}{=} \frac{1}{2N\,\Delta_{\mathrm{t}}} \sum_{k=-N}^{N-1} \hat{S}_m^{(\mathrm{DSP})}(\tilde{f}_k) \mathrm{e}^{\mathrm{i}2\pi\tilde{f}_k\tau\,\Delta_{\mathrm{t}}} \tag{348}$$

to be an ACVS estimator corresponding to the discretely smoothed periodogram once we define $\hat{S}_m^{(\mathrm{DSP})}(\tilde{f}_k)$ when k is odd (we already know $\hat{S}_m^{(\mathrm{DSP})}(\tilde{f}_k)$ for even k since $\hat{S}_m^{(\mathrm{DSP})}(\tilde{f}_k) = \hat{S}_m^{(\mathrm{DSP})}(f_{k/2})$). Here we explore options for doing so.

(a) Show that, if we set $\hat{S}_m^{(\mathrm{DSP})}(\tilde{f}_k)$ for positive odd indices k to *any* nonnegative values so desired and if we set $\hat{S}_m^{(\mathrm{DSP})}(\tilde{f}_{-k}) = \hat{S}_m^{(\mathrm{DSP})}(\tilde{f}_k)$, then $\{\hat{s}_\tau^{(\mathrm{DSP})} : \tau \in \mathbb{Z}\}$ corresponds to the ACVS for a harmonic process.

(b) Compute ACVS estimates $\{\hat{s}_\tau^{(\mathrm{DSP})}(\cdot) : \tau = 0, 1, \ldots, 1024\}$ corresponding to the discretely smoothed periodogram of Figure 310(a) by setting $\hat{S}_m^{(\mathrm{DSP})}(\tilde{f}_k)$ at odd indices k in three different ways: using (i) zeros, (ii) the linear interpolation scheme of Equation (313b) and (iii) the convolution scheme of Equation (313c). Comment briefly on how these three estimates compare. Given that each of these ACVS estimates is associated with a harmonic process, how would you go about estimating s_τ for $\tau > 1024$? (Figure 34(a) shows the AR(2) series used to form the discretely smoothed periodogram of interest here – this series is downloadable from the book's website).

(c) A harmonic process does not possess a proper SDF, so the ACVS given in part (a) is disconcerting since the intent of a discretely smoothed periodogram is to serve as an SDF estimator. Construct a proper SDF, say $S_{\mathrm{PC}}(\cdot)$, that is piecewise constant over intervals of the form $(\tilde{f}_k - 1/(4N), \tilde{f}_k + 1/(4N))$ such that $S_{\mathrm{PC}}(\tilde{f}_k) = \hat{S}_m^{(\mathrm{DSP})}(\tilde{f}_k)$ for all k, and show that its ACVS is given by

$$s_{\mathrm{PC},\tau} = \frac{\sin\left(\pi\tau/(2N)\right)}{\pi\tau/(2N)} \hat{s}_\tau^{(\mathrm{DSP})}, \quad \tau \in \mathbb{Z}$$

(when $\tau = 0$, interpret the above ratio to be unity).

(d) For each of the three ACVS estimates computed in part (b), compute corresponding $s_{\mathrm{PC},\tau}$ sequences for $\tau = 0, 1, \ldots, 4096$. Comment briefly on how $s_{\mathrm{PC},\tau}$ compares to $\hat{s}_\tau^{(\mathrm{DSP})}$ in each case and how the three $s_{\mathrm{PC},\tau}$ sequences compare to one another. Comment on how these $s_{\mathrm{PC},\tau}$ sequences compare to the usual biased estimator $\hat{s}_\tau^{(\mathrm{P})}$ at lags $\tau \ge 1024$.

[7.26] Figure 310 shows the periodogram for the AR(2) time series of Figure 34(a) along with two smoothed versions thereof. The left-hand plot depicts a discretely smoothed periodogram $\hat{S}_m^{(\mathrm{DSP})}(\cdot)$ with the smoothing parameter m chosen by the GCV criterion of Equation (309a); the right-hand,

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

a Parzen lag window estimate $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ with m chosen via Equation (313d). Create similar figures for the AR(2) time series shown in Figures 34(b), (c) and (d) (downloadable from the book's website). For a given series, how well do $\hat{S}_m^{(\mathrm{DSP})}(\cdot)$ and $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ agree?

[7.27] We have discussed three objective methods for determining the amount of smoothing to be applied to a direct spectral estimator: estimation of the spectral bandwidth (Section 7.8), estimation of a mean integrated square error in log space (Section 7.9) and a generalized cross-validation criterion (Section 7.10). This exercise seeks to evaluate these three methods using different measures of how well a particular SDF estimate matches up with the true SDF.

(a) Generate a large number N_{R} of realizations of length $N = 1024$ from the AR(2) process of Equation (34) (take N_{R} to be at least 1000, and see Exercise [597] for a description of how to generate the realizations). For each realization, form periodogram-based Parzen lag window estimates $\hat{S}_m^{(\mathrm{LW})}(f_k)$ over the Fourier frequencies $f_k = k/N$, $k = 0, 1, \ldots, N/2$, with m set, first, to $[3.7/\tilde{B}_T]$, where $\tilde{B}_T$ is the estimator of the time series bandwidth given by Equation (300c), and $[x]$ is x rounded to the nearest integer (cf. the discussion surrounding Equation (322)); and, second, to the minimizer of $\hat{I}_m$ of Equation (303e). For each realization, also form a discretely smoothed periodogram $\hat{S}_m^{(\mathrm{DSP})}(f_k)$ as per Equation (307), with weights $g_{m,j} \propto \exp(-(mj)^2)$ such that $\sum_j g_{m,j} = 1$, and with m set to be the minimizer of Equation (309a). Evaluate the quality of the two lag window estimates and the discretely smoothed periodogram using at least two of the following seven measures $M_1, \ldots, M_7$:

$$
\begin{aligned}
M_1 &= \frac{1}{N/2+1} \sum_{k=0}^{N/2} \left(\hat{S}_m(f_k) - S(f_k) \right)^2 \\
M_2 &= \frac{1}{N/2+1} \sum_{k=0}^{N/2} \left(\log\left(\hat{S}_m(f_k) \right) - \log\left(S(f_k) \right) \right)^2 \\
M_3 &= \frac{1}{N/2+1} \sum_{k=0}^{N/2} \left(\frac{\hat{S}_m(f_k) - S(f_k)}{S(f_k)} \right)^2 \\
M_4 &= \frac{1}{N/2+1} \sum_{k=0}^{N/2} \left[\frac{S(f_k)}{\hat{S}_m(f_k)} - \log\left(\frac{S(f_k)}{\hat{S}_m(f_k)} \right) - 1 \right] \\
M_5 &= \frac{1}{N/2+1} \sum_{k=0}^{N/2} \left[\frac{\hat{S}_m(f_k)}{S(f_k)} - \log\left(\frac{\hat{S}_m(f_k)}{S(f_k)} \right) - 1 \right] \\
M_6 &= M_4 + M_5 \\
M_7 &= \max_k |10 \log_{10}(\hat{S}_m(f_k)/S(f_k))|,
\end{aligned}
$$

where $\hat{S}_m(f_k)$ is either $\hat{S}_m^{(\mathrm{LW})}(f_k)$ or $\hat{S}_m^{(\mathrm{DSP})}(f_k)$, and $S(\cdot)$ is the true AR(2) SDF (measures M_1 to M_6 are adaptations of MSE of Equation (293), MSLE of Equation (296b), NMSE of Equation (296a), KL of Equation (297c), LK of Equation (345b) and KLLK of Equation (345c)). Average each of your chosen measures over all N_{R} realizations. Comment on your findings.

(b) Repeat part (a), but now using the AR(4) process of Equation (35a) (Exercise [11.1] describes generation of the required realizations). Replace the periodogram with a Hanning-based direct SDF estimator throughout, blindly using exactly the same methods for selecting m and the same measures for evaluating the resulting SDF estimates as before. Comment on your findings.

[7.28] Figure 319 shows four lag window estimates $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ for the ocean wave time series, all based upon a direct spectral estimate $\hat{S}^{(\mathrm{D})}(\cdot)$ using a $NW = 2/\Delta_{\mathrm{t}}$ Slepian data taper. Conceptually each $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ is the convolution of $\hat{S}^{(\mathrm{D})}(\cdot)$ with a smoothing window $W_m(\cdot)$, i.e., the Fourier transform of the chosen lag window $\{w_{m,\tau}\}$. Suppose we subject $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ to an additional smoothing op-

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press

eration identical to the first; i.e., we form $\hat{S}^{(\mathrm{RLW})}(\cdot)$ by convolving $\hat{S}_m^{(\mathrm{LW})}(\cdot)$ with $W_m(\cdot)$, where "RLW" stands for "repeated lag window."

(a) Show that the repeated lag window estimator is an ordinary lag window estimator (i.e., one based directly on $\hat{S}^{(\mathrm{D})}(\cdot)$) by a suitable definition of a new lag window. In terms of $W_m(\cdot)$, what is the smoothing window that is used to produce $\hat{S}^{(\mathrm{RLW})}(\cdot)$ from $\hat{S}^{(\mathrm{D})}(\cdot)$?

(b) For the lag windows used in Figure 319, show that the smoothing window bandwidth for $\hat{S}^{(\mathrm{RLW})}(\cdot)$ is greater than that for $\hat{S}_m^{(\mathrm{LW})}(\cdot)$. What conditions do we need to impose on other lag windows for this relationship to hold?

(c) Recreate the contents of Figure 319 but using the appropriate repeated lag windows in place of the original lag windows and showing the central lobes of the smoothing windows for all four lag windows (rather than the lobes of the design windows for the Daniell and Bartlett–Priestley lag windows). Compare the smoothing window bandwidths (Equation (251e)), the standard measure of effective bandwidths (Equation (256a)) and EDOFs (Equation (264b)) associated with the original and repeated lag window estimates. Comment upon your findings. (The book's website has the required data.)

[7.29] Figure 330 shows the ship altitude data we used in the analysis presented in Section 7.12. This series has 2048 values, but it is the first part of a longer series of length $N = 14{,}697$ that is available on the book's website. Repeat the analysis of the ship altitude data, but now using this longer series. Do any of the conclusions we made based on the shorter series need to be changed?

https://doi.org/10.1017/9781139235723.008 Published online by Cambridge University Press