10

Harmonic Analysis

10.0 Introduction

If a stationary process has a purely continuous spectrum, it is natural to estimate its spectral density function (SDF) since this function is easier to interpret than the integrated spectrum. Estimation of the SDF has occupied our attention in the previous four chapters. However, if we are given a sample of a time series drawn from a process with a purely discrete spectrum (i.e., a "line" spectrum for which the integrated spectrum is a step function), our estimation problem is quite different: we must estimate the location and magnitude of the jumps in the integrated spectrum, which requires estimation techniques that differ from what we have already studied. It is more common, however, to come across processes whose spectra are a mixture of lines and an SDF stemming from a so-called "background continuum." In Section 4.4 we distinguished two cases. If the SDF for the continuum is that of white noise, we said that the process has a discrete spectrum – as opposed to a *purely* discrete spectrum, which has only a line component; on the other hand, if the SDF for the continuum differs from that of white noise (sometimes called "colored" noise), we said that the process has a mixed spectrum (see Figure 121).

In this chapter we begin with discrete parameter harmonic processes with random phases (these processes have a purely discrete spectrum). In the next section we first recall how we defined these processes in Section 2.6. We then use some standard concepts from tidal analysis to motivate and illustrate use of these processes as models.

10.1 Harmonic Processes – Purely Discrete Spectra

A real-valued discrete parameter harmonic process with random phases can be written as

$$X_t = \mu + \sum_{l=1}^{L} D_l \cos\left(2\pi f_l t\, \Delta_{\mathrm{t}} + \phi_l\right), \tag{511}$$

where μ, $D_l \geq 0$ and $f_l > 0$ are in general unknown real-valued constants; $L \geq 1$ is the number of harmonic components; and the ϕ_l terms are independent real-valued random variables (RVs) – representing random phase angles – with a rectangular (uniform) distribution on $(-\pi, \pi]$ (see Exercise [37]; Equation (511) is the same as Equation (35d) except for the insertion of the sampling interval Δ_{t}, which makes the frequencies f_l have physically meaningful units of cycles per unit time). Since $\{X_t\}$ is a discrete parameter process, we can consider

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

each f_l to lie in the interval $(0, f_{\mathcal{N}}]$ due to the aliasing effect, where $f_{\mathcal{N}} \stackrel{\text{def}}{=} 1/(2\,\Delta_{\text{t}})$ is the Nyquist frequency (see Sections 3.9 and 4.5).

This model is in one sense not really a statistical model: for each realization of the process the RVs ϕ_l are fixed quantities, so each realization is essentially deterministic and free from what is commonly thought of as random variation. For this reason the process is often dismissed as of no interest; however, its simple properties can prove useful in practice as we will demonstrate by looking at ocean tide prediction.

The theory of gravitational tidal potential predicts that at a fixed location the hourly height of the ocean tide at the hour with index t can be written as

$$X_t = \mu + \sum_{l=1}^{L'} \alpha_{t,l} H_l \cos\left(2\pi f_l t\,\Delta_{\text{t}} + v_{t,l} + u_{t,l} - g_l\right) \quad \text{with } \Delta_{\text{t}} = 1 \text{ hour}, \tag{512a}$$

where the sum is over L' primary tidal constituent frequencies, and the harmonic constants H_l and g_l depend only upon the location of interest and must be estimated. (These primary constituent frequencies derive from consideration of a frictionless and inertia-free ocean covering the whole globe; this crude model is inadequate for the real ocean, various hydrographic effects such as the locations of land and rivers making it necessary to introduce the location-dependent amplitudes H_l and phases g_l.) The $\alpha_{t,l}$ and $u_{t,l}$ terms are the so-called nodal parameters (these are essentially corrections for the amplitudes H_l and phases g_l), and $v_{t,l}$ is the phase of the lth constituent of the potential over Greenwich at $t = 0$ (these three quantities can be found from, e.g., Doodson and Warburg, 1941). For locations such as Honolulu, where the tides are virtually "linear" (Munk and Cartwright, 1966), this theory is sufficient, but at many British ports and elsewhere the distortion of tides by shallow-water effects often cannot be ignored. Nonlinear interactions are incorporated into harmonic tidal prediction by including constituents with frequencies that are sums and differences of the L' primary constituents. The total number of constituents required, say L, can exceed 100 in complicated shallow water areas such as Anchorage, Alaska, or Southampton, England.

In practice the terms of known form, namely the nodal parameters $\alpha_{t,l}$ and $u_{t,l}$ – which are continuously but slowly varying – are updated at fixed periods when computing the right-hand side of Equation (512a). Sufficient accuracy is often obtained by regarding them as fixed over a span of a year and updating them annually, although every 60 days is preferred (whichever span is used, the $v_{t,l}$ terms are also updated by "resetting" the index t to 0 at the start of each span). This implies that in Equation (512a) both the amplitudes and phases of the cosines vary slowly and slightly with time. This is a good example of what Thomson (1982) called the "convenient fiction" of pure line components. To illustrate the complicated forms that can occur, predicted heights of the tide at Southampton, England, are given in Figure 513 for two different segments of time in 1977. This uses $L = 102$ frequencies and 60 day updating of $\alpha_{t,l}$, $u_{t,l}$ and $v_{t,l}$ (see Walden, 1982). The two segments differ substantially even though they arise for the same model, which illustrates the point that lots of sinusoidal terms – when summed – can interact to give quite complicated patterns.

10.2 Harmonic Processes with Additive White Noise – Discrete Spectra

As we have noted, the model given by Equation (511) is not truly statistical in the sense that, once the random phases ϕ_l are set, there are no remaining stochastic components (the time series is just a linear combination of cosines with known frequencies, amplitudes and phases). A more realistic and useful model is given by

$$X_t = \mu + \sum_{l=1}^{L} D_l \cos\left(2\pi f_l t\,\Delta_{\text{t}} + \phi_l\right) + \epsilon_t, \tag{512b}$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

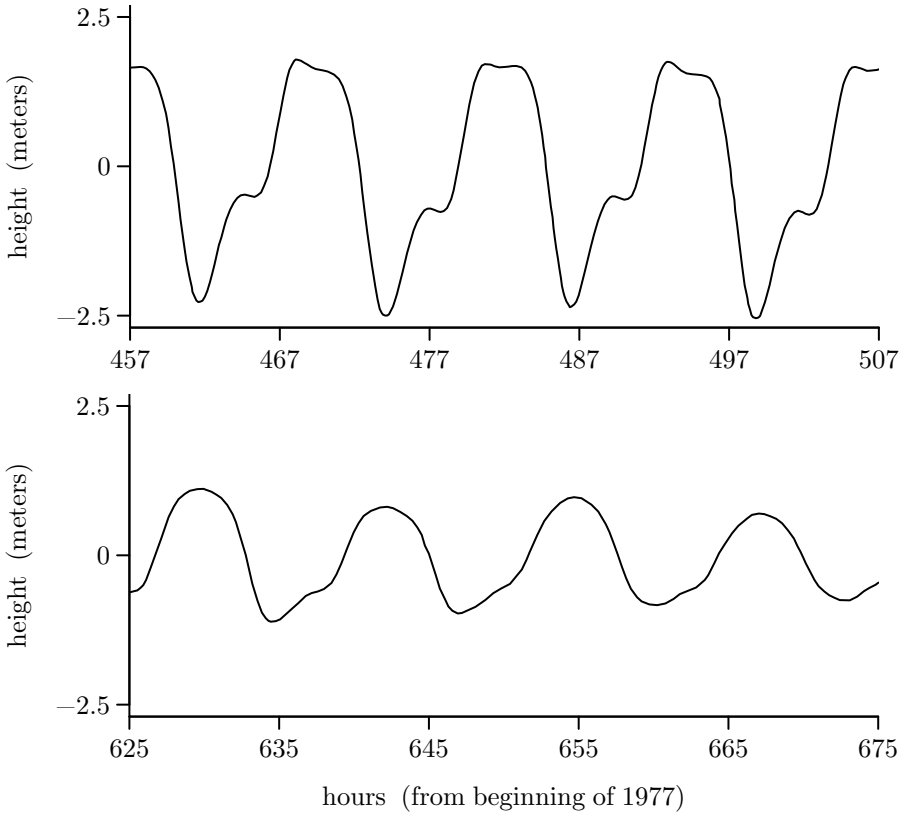

Figure 513 Predicted spring and neap tidal curves at Southampton (January, 1977). Tidal elevations relative to the mean sea level (in meters) are plotted versus an hour count starting from the beginning of 1977. The upper and lower plots show, respectively, spring and neap tides.

which differs from Equation (511) only in that ϵ_t has been added to represent observational error (always present in any sort of physical measurements). The background continuum $\{\epsilon_t\}$ is assumed to be a real-valued white noise process with zero mean and variance σ_ϵ^2. Each ϵ_t is also assumed to be independent of each ϕ_l. Whereas the model of Equation (511) has a purely discrete spectrum, that of Equation (512b) has a discrete spectrum.

Equation (512b) arises more from analytical convenience than realism – many measuring instruments are band-limited, so that $\{\epsilon_t\}$ is not truly white. We looked at predicted tidal heights in the previous section. Observed tidal heights – i.e., predictions plus noise – certainly have a mixed spectrum with colored rather than white noise (see Section 10.4). However, we shall concentrate in what follows on the model of Equation (512b) since it is commonly assumed and – if nothing else – can serve an approximation to the mixed process case.

Let us assume initially that $L = 1$ and that the frequency $f_1 = f$ is known. The unknown parameters in Equation (512b) are now just $D_1 = D$ and σ_ϵ^2. The model can then be rewritten as

$$X_t = \mu + A\cos(2\pi f t\,\Delta_t) + B\sin(2\pi f t\,\Delta_t) + \epsilon_t, \tag{513}$$

where $A \stackrel{\text{def}}{=} D\cos(\phi_1)$ and $B \stackrel{\text{def}}{=} -D\sin(\phi_1)$. Note that, for a given realization, A and B are just constants so that Equation (513) expresses a linear regression model. For a time series of length N that is assumed to be a realization of $X_0, X_1, \ldots, X_{N-1}$, we can estimate μ, A and B by least squares (see, e.g., Bloomfield, 2000, or Weisberg, 2014). For real sinusoids,

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

this amounts to minimizing the residual sum of squares

$$\mathrm{SS}(\mu, A, B) \stackrel{\text{def}}{=} \|\boldsymbol{X} - \boldsymbol{H\beta}\|^2 = \sum_{t=0}^{N-1} \left[X_t - \mu - A\cos\left(2\pi f t\,\Delta_{\mathrm{t}}\right) - B\sin\left(2\pi f t\,\Delta_{\mathrm{t}}\right)\right]^2, \tag{514a}$$

where $\boldsymbol{X}^T \stackrel{\text{def}}{=} [X_0, X_1, \ldots, X_{N-1}]$; $\boldsymbol{\beta}^T \stackrel{\text{def}}{=} [\mu, A, B]$;

$$\boldsymbol{H}^T \stackrel{\text{def}}{=} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \cos\left(2\pi f\,\Delta_{\mathrm{t}}\right) & \cos\left(4\pi f\,\Delta_{\mathrm{t}}\right) & \cdots & \cos\left([N-1]2\pi f\,\Delta_{\mathrm{t}}\right) \\ 0 & \sin\left(2\pi f\,\Delta_{\mathrm{t}}\right) & \sin\left(4\pi f\,\Delta_{\mathrm{t}}\right) & \cdots & \sin\left([N-1]2\pi f\,\Delta_{\mathrm{t}}\right) \end{bmatrix};$$

and $\|\boldsymbol{V}\|$ refers to the Euclidean norm of the vector $\boldsymbol{V}$. Since

$$\frac{\mathrm{dSS}}{\mathrm{d}\boldsymbol{\beta}} = -2\boldsymbol{H}^T\left(\boldsymbol{X} - \boldsymbol{H\beta}\right)$$

(see, e.g., Rao, 1973, p. 71), we can set the above equal to the zero vector to obtain the following normal equations (here $\hat{\mu}$, $\hat{A}$ and $\hat{B}$ represent their solutions):

$$\sum_{t=0}^{N-1} X_t = N\hat{\mu} + \hat{A}\sum_{t=0}^{N-1}\cos\left(2\pi f t\,\Delta_{\mathrm{t}}\right) + \hat{B}\sum_{t=0}^{N-1}\sin\left(2\pi f t\,\Delta_{\mathrm{t}}\right);$$

$$\begin{aligned}\sum_{t=0}^{N-1} X_t\cos\left(2\pi f t\,\Delta_{\mathrm{t}}\right) &= \hat{\mu}\sum_{t=0}^{N-1}\cos\left(2\pi f t\,\Delta_{\mathrm{t}}\right) + \hat{A}\sum_{t=0}^{N-1}\cos^2(2\pi f t\,\Delta_{\mathrm{t}}) \\ &\quad + \hat{B}\sum_{t=0}^{N-1}\cos\left(2\pi f t\,\Delta_{\mathrm{t}}\right)\sin\left(2\pi f t\,\Delta_{\mathrm{t}}\right); \end{aligned} \tag{514b}$$

$$\begin{aligned}\sum_{t=0}^{N-1} X_t\sin\left(2\pi f t\,\Delta_{\mathrm{t}}\right) &= \hat{\mu}\sum_{t=0}^{N-1}\sin\left(2\pi f t\,\Delta_{\mathrm{t}}\right) + \hat{B}\sum_{t=0}^{N-1}\sin^2(2\pi f t\,\Delta_{\mathrm{t}}) \\ &\quad + \hat{A}\sum_{t=0}^{N-1}\cos\left(2\pi f t\,\Delta_{\mathrm{t}}\right)\sin\left(2\pi f t\,\Delta_{\mathrm{t}}\right). \end{aligned} \tag{514c}$$

We can now use Exercises [1.2c] and [1.3] to reduce these three equations somewhat. For example, Equation (514b) becomes

$$\begin{aligned}\sum_{t=0}^{N-1} X_t\cos\left(2\pi f t\,\Delta_{\mathrm{t}}\right) &= \hat{\mu}\frac{\cos\left([N-1]\pi f\,\Delta_{\mathrm{t}}\right)\sin\left(N\pi f\,\Delta_{\mathrm{t}}\right)}{\sin\left(\pi f\,\Delta_{\mathrm{t}}\right)} \\ &\quad + \hat{A}\left[\frac{N}{2} + \frac{\sin\left(N2\pi f\,\Delta_{\mathrm{t}}\right)}{2\sin\left(2\pi f\,\Delta_{\mathrm{t}}\right)}\cos\left([N-1]2\pi f\,\Delta_{\mathrm{t}}\right)\right] \\ &\quad + \hat{B}\left[\frac{\sin\left(N2\pi f\,\Delta_{\mathrm{t}}\right)}{2\sin\left(2\pi f\,\Delta_{\mathrm{t}}\right)}\sin\left([N-1]2\pi f\,\Delta_{\mathrm{t}}\right)\right], \end{aligned}$$

with similar expressions for the other two normal equations. Using these expressions, it can be argued (see Exercise [10.1]) that

$$\left|\hat{\mu} - \overline{X}\right|, \quad \left|\hat{A} - \frac{2}{N}\sum_{t=0}^{N-1} X_t\cos\left(2\pi f t\,\Delta_{\mathrm{t}}\right)\right| \text{ and } \left|\hat{B} - \frac{2}{N}\sum_{t=0}^{N-1} X_t\sin\left(2\pi f t\,\Delta_{\mathrm{t}}\right)\right|$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

are all close to zero (at least for large N), where, as usual, $\overline{X} \stackrel{\text{def}}{=} \sum_{t=0}^{N-1} X_t/N$, the sample mean. Thus we have, to a good approximation,

$$\hat{\mu} \approx \overline{X}, \;\; \hat{A} \approx \frac{2}{N} \sum_{t=0}^{N-1} X_t \cos(2\pi f t \, \Delta_t) \;\text{ and }\; \hat{B} \approx \frac{2}{N} \sum_{t=0}^{N-1} X_t \sin(2\pi f t \, \Delta_t). \tag{515a}$$

If in fact $f = k/(N\,\Delta_t)$, where k is an integer such that $1 \le k < N/2$, it follows from Exercise [1.3] that the approximations in Equation (515a) are in fact equalities. As we have noted in Chapters 3 and 6, this special set of frequencies – along with 0 and, if N is even, $f_{\mathcal{N}}$ – is known as the *Fourier frequencies* or *standard frequencies*. Since any frequency f in the interval $(0, f_{\mathcal{N}}]$ is at most a distance of $1/(N\,\Delta_t)$ away from a Fourier frequency not equal to 0 or $f_{\mathcal{N}}$, it is intuitively reasonable that the approximations in Equation (515a) are good.

Reverting back to the general case $L \ge 1$, we can rewrite the model of Equation (512b) as

$$X_t = \mu + \sum_{l=1}^{L} [A_l \cos(2\pi f_l t \, \Delta_t) + B_l \sin(2\pi f_l t \, \Delta_t)] + \epsilon_t, \tag{515b}$$

where $A_l \stackrel{\text{def}}{=} D_l \cos(\phi_l)$ and $B_l \stackrel{\text{def}}{=} -D_l \sin(\phi_l)$. For the remainder of this section, let us assume that each f_l is a Fourier frequency not equal to either 0 or $f_{\mathcal{N}}$. The deductions already made are equally valid here, and so $\hat{\mu} = \overline{X}$,

$$\hat{A}_l = \frac{2}{N} \sum_{t=0}^{N-1} X_t \cos(2\pi f_l t \, \Delta_t) \;\text{ and }\; \hat{B}_l = \frac{2}{N} \sum_{t=0}^{N-1} X_t \sin(2\pi f_l t \, \Delta_t) \tag{515c}$$

are the *exact* least squares estimators of μ, A_l and B_l, respectively. Let us consider some of the statistical properties of $\hat{A}_l$ and $\hat{B}_l$ *under the assumption that each ϕ_l is a constant* – this implies that A_l and B_l are constants. Equation (515b) is thus a standard multiple linear regression model, and the fact that the RVs ϵ_t are uncorrelated means that the RVs X_t are also uncorrelated. Note also that

$$\begin{aligned} E\{\hat{A}_l\} &= \frac{2}{N} \sum_{t=0}^{N-1} E\{X_t\} \cos(2\pi f_l t \, \Delta_t) \\ &= \frac{2}{N} \sum_{t=0}^{N-1} \sum_{k=1}^{L} \Big[A_k \cos(2\pi f_k t \, \Delta_t) \cos(2\pi f_l t \, \Delta_t) \\ &\qquad + B_k \sin(2\pi f_k t \, \Delta_t) \cos(2\pi f_l t \, \Delta_t) \Big] \end{aligned}$$

since $\sum_t \mu \cos(2\pi f_l t \, \Delta_t) = 0$ (see Exercise [1.2d]) and $E\{\epsilon_t\} = 0$. By interchanging the summations and using the orthogonality relationships

$$\sum_{t=0}^{N-1} \cos(2\pi f_k t \, \Delta_t) \cos(2\pi f_l t \, \Delta_t) = \sum_{t=0}^{N-1} \sin(2\pi f_k t \, \Delta_t) \cos(2\pi f_l t \, \Delta_t) = 0,$$

$$\sum_{t=0}^{N-1} \cos^2(2\pi f_l t \, \Delta_t) = N/2,$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

$k \neq l$ (see Exercise [1.3c]), we find that $E\{\hat{A}_l\} = A_l$ and, likewise, $E\{\hat{B}_l\} = B_l$. From the definitions of $\hat{A}_l$ and $\hat{B}_l$ and the fact that the RVs X_t are uncorrelated, we have

$$\operatorname{var}\{\hat{A}_l\} = \frac{4}{N^2} \sum_{t=0}^{N-1} \operatorname{var}\{X_t\} \cos^2(2\pi f_l t\, \Delta_t) = \frac{2\sigma_\epsilon^2}{N} = \operatorname{var}\{\hat{B}_l\} \tag{516a}$$

when each ϕ_l is fixed. It also follows from the orthogonality relationships that, for $k \neq l$,

$$\operatorname{cov}\{\hat{A}_k, \hat{B}_l\} = \operatorname{cov}\{\hat{A}_l, \hat{B}_l\} = \operatorname{cov}\{\hat{A}_k, \hat{A}_l\} = \operatorname{cov}\{\hat{B}_k, \hat{B}_l\} = 0.$$

We also note that $E\{\hat{\mu}\} = \mu$, $\operatorname{var}\{\hat{\mu}\} = \sigma_\epsilon^2/N$ and $\operatorname{cov}\{\hat{\mu}, \hat{A}_l\} = \operatorname{cov}\{\hat{\mu}, \hat{B}_l\} = 0$ for all l.

We can estimate σ_ϵ^2 by the usual formula used in multiple linear regression (Weisberg, 2014), namely,

$$\begin{aligned} \hat{\sigma}_\epsilon^2 &= \frac{1}{N-2L-1} \sum_{t=0}^{N-1} \left(X_t - \overline{X} - \sum_{l=1}^{L} \left[\hat{A}_l \cos(2\pi f_l t\, \Delta_t) + \hat{B}_l \sin(2\pi f_l t\, \Delta_t) \right] \right)^2 \\ &= \frac{1}{N-2L-1} \left\| \boldsymbol{X} - \boldsymbol{H}\hat{\boldsymbol{\beta}} \right\|^2, \end{aligned} \tag{516b}$$

where now

$$\boldsymbol{H}^T \stackrel{\text{def}}{=} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \cos(2\pi f_1 \Delta_t) & \cos(4\pi f_1 \Delta_t) & \cdots & \cos([N-1]2\pi f_1 \Delta_t) \\ 0 & \sin(2\pi f_1 \Delta_t) & \sin(4\pi f_1 \Delta_t) & \cdots & \sin([N-1]2\pi f_1 \Delta_t) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \cos(2\pi f_L \Delta_t) & \cos(4\pi f_L \Delta_t) & \cdots & \cos([N-1]2\pi f_L \Delta_t) \\ 0 & \sin(2\pi f_L \Delta_t) & \sin(4\pi f_L \Delta_t) & \cdots & \sin([N-1]2\pi f_L \Delta_t) \end{bmatrix}$$

and $\hat{\boldsymbol{\beta}}^T \stackrel{\text{def}}{=} [\hat{\mu}, \hat{A}_1, \hat{B}_1, \ldots, \hat{A}_L, \hat{B}_L]$. The divisor $N - 2L - 1$ is due to the fact that we have estimated $2L + 1$ parameters from the data.

If the frequencies f_l are not all of the form $k/(N\,\Delta_t)$, the estimators $\hat{\mu}$, $\hat{A}_l$ and $\hat{B}_l$ can be regarded as approximate least squares estimates of μ, A_l and B_l. Priestley (1981, p. 394) argues that in this case

$$E\{\hat{A}_l\} = A_l + O\left(\frac{1}{N}\right) \quad \text{and} \quad E\{\hat{B}_l\} = B_l + O\left(\frac{1}{N}\right), \tag{516c}$$

which means that, for example, there exists a constant c (independent of N) such that

$$\left| E\{\hat{A}_l\} - A_l \right| < \frac{c}{N}$$

for all N (note, however, that c need not be a small number!).

To summarize, in order to estimate A_l and B_l, we have regarded each ϕ_l as fixed and have derived some statistical properties of the corresponding estimators ($\hat{A}_l$ and $\hat{B}_l$) conditional on ϕ_l being fixed. With this conditioning, the only stochastic part of $\{X_t\}$ is $\{\epsilon_t\}$, and the process $\{X_t\}$ is uncorrelated, which makes it easy to deduce the statistical properties of $\hat{A}_l$ and $\hat{B}_l$. While this approach is reasonable for amplitude estimation, it is problematic for spectral estimation. Fixing each ϕ_l causes $\{X_t\}$ to be nonstationary since its mean value, namely,

$$E\{X_t\} = \mu + \sum_{l=1}^{L} A_l \cos(2\pi f_l t\, \Delta_t) + B_l \sin(2\pi f_l t\, \Delta_t),$$

varies with time. Hence $\{X_t\}$ is outside of the framework of the spectral representation theorem, which provides the basis for spectral estimation. Proper theoretical treatment of spectral estimation would require reintroducing the ϕ_l terms as RVs.

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

Comments and Extensions to Section 10.2

[1] Here we look at the complex-valued counterpart of Equation (512b), namely,

$$Z_t = \mu + \sum_{l=1}^{L} D_l e^{i(2\pi f_l t \Delta_t + \phi_l)} + \epsilon_t, \tag{517a}$$

where, while the amplitude D_l is real-valued, now $\{Z_t\}$ and μ are both complex-valued, $|f_l| \in (0, f_{\mathcal{N}}]$ – hence $f_l = 0$ is prohibited – and $\{\epsilon_t\}$ is proper complex-valued white noise with zero mean and variance $\sigma_\epsilon^2 = E\{|\epsilon_t|^2\}$ (see Exercise [32] and the discussion preceding it). Although Equation (512b) is much more useful for physical applications, the complex-valued form has three advantages:

[1] it leads to more compact mathematical expressions;
[2] it brings out more clearly the connection with the periodogram; and
[3] it is the form most often used in the electrical engineering literature.

The equivalent of Equation (513) is

$$Z_t = \mu + Ce^{i2\pi f t \Delta_t} + \epsilon_t, \tag{517b}$$

where $C \stackrel{\text{def}}{=} D\exp(i\phi)$ is the complex-valued amplitude of the complex exponential. Again, if we regard ϕ as a constant for a given realization, we can estimate the parameters μ and C by least squares:

$$\mathrm{SS}(\mu, C) = \|\boldsymbol{Z} - \boldsymbol{H}\boldsymbol{\beta}\|^2 = \sum_{t=0}^{N-1} \left| Z_t - \mu - Ce^{i2\pi f t \Delta_t} \right|^2,$$

where $\boldsymbol{Z}^T \stackrel{\text{def}}{=} \left[Z_0, Z_1, \ldots, Z_{N-1}\right]$, $\boldsymbol{\beta}^T \stackrel{\text{def}}{=} [\mu, C]$ and

$$\boldsymbol{H}^T \stackrel{\text{def}}{=} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & e^{i2\pi f \Delta_t} & e^{i4\pi f \Delta_t} & \cdots & e^{i(N-1)2\pi f \Delta_t} \end{bmatrix}.$$

If we set

$$\frac{d\mathrm{SS}}{d\boldsymbol{\beta}} = -\boldsymbol{H}^T \left(\boldsymbol{Z}^* - \boldsymbol{H}^*\boldsymbol{\beta}^*\right)$$

to zero and take the complex conjugate of the resulting expressions (the asterisk denotes this operation), we obtain the following normal equations:

$$\sum_{t=0}^{N-1} Z_t = N\hat{\mu} + \hat{C}\sum_{t=0}^{N-1} e^{i2\pi f t \Delta_t} \quad \text{and} \quad \sum_{t=0}^{N-1} Z_t e^{-i2\pi f t \Delta_t} = \hat{\mu}\sum_{t=0}^{N-1} e^{-i2\pi f t \Delta_t} + N\hat{C}.$$

From Exercises [1.2c] and [1.2d], we know that

$$\sum_{t=0}^{N-1} e^{i2\pi f_k t \Delta_t} = 0 \text{ for } f_k \stackrel{\text{def}}{=} \frac{k}{N\Delta_t}, \quad 1 \le k < N.$$

If we restrict f to this set of frequencies, the following are exact least squares estimators:

$$\hat{\mu} = \frac{1}{N}\sum_{t=0}^{N-1} Z_t \text{ and } \hat{C} = \frac{1}{N}\sum_{t=0}^{N-1} Z_t e^{-i2\pi f t \Delta_t}. \tag{517c}$$

Note that $\hat{C}$ would reduce to $\hat{\mu}$ if we were to allow f to be zero (recall the assumption $0 < |f_l| \le f_{\mathcal{N}}$); i.e., C would be the amplitude associated with zero frequency and would be confounded with the mean

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

value μ (the mean is sometimes called the DC – *direct current* – component in the electrical engineering literature).

If we introduce several frequencies, our complex-valued model now becomes

$$Z_t = \mu + \sum_{l=1}^{L} C_l e^{i2\pi f_l t \Delta_t} + \epsilon_t, \tag{518a}$$

where we assume that each f_l is one of the Fourier frequencies not equal to 0 or $\pm f_{\mathcal{N}}$. Under the same assumptions and restrictions as for their real sinusoidal counterpart, the following results hold:

$$E\{\hat{C}_l\} = C_l, \quad \text{var}\{\hat{C}_l\} = \frac{\sigma_\epsilon^2}{N} \quad \text{and} \quad \text{cov}\{\hat{C}_k, \hat{C}_l\} = 0, \quad k \neq l,$$

where $\hat{C}_l$ is the least squares estimate of C_l (see Exercise [10.2]). In addition we have

$$E\{\hat{\mu}\} = \mu, \quad \text{var}\{\hat{\mu}\} = \frac{\sigma_\epsilon^2}{N} \quad \text{and} \quad \text{cov}\{\hat{\mu}, \hat{C}_l\} = 0.$$

In order to estimate σ_ϵ^2, we use the complex analog of the usual formula in linear regression:

$$\hat{\sigma}_\epsilon^2 = \frac{1}{N - L - 1} \left\| \boldsymbol{Z} - \boldsymbol{H}\hat{\boldsymbol{\beta}} \right\|^2, \tag{518b}$$

where

$$\boldsymbol{H}^T \stackrel{\text{def}}{=} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & e^{i2\pi f_1 \Delta_t} & e^{i4\pi f_1 \Delta_t} & \cdots & e^{i(N-1)2\pi f_1 \Delta_t} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{i2\pi f_L \Delta_t} & e^{i4\pi f_L \Delta_t} & \cdots & e^{i(N-1)2\pi f_L \Delta_t} \end{bmatrix}$$

and $\hat{\boldsymbol{\beta}}^T \stackrel{\text{def}}{=} \left[\hat{\mu}, \hat{C}_1, \ldots, \hat{C}_L\right]$. The divisor $N - L - 1$ is due to the fact that we have estimated $L + 1$ complex-valued parameters from our complex-valued data. Under a Gaussian assumption, $\hat{\sigma}_\epsilon^2$ is an unbiased estimator of σ_ϵ^2 (Miller, 1973, theorem 7.3); however, this result depends upon $\{\epsilon_t\}$ being proper complex-valued white noise (to the best of our knowledge, the properties of the expectation of $\hat{\sigma}_\epsilon^2$ in the improper case are an open question).

10.3 Spectral Representation of Discrete and Mixed Spectra

The Lebesgue decomposition theorem for integrated spectra (Section 4.4) says that a combination of spectral lines plus white noise (a "discrete" spectrum) or a combination of spectral lines plus colored noise (a "mixed" spectrum) has an integrated spectrum $S^{(\text{I})}(\cdot)$ given by

$$S^{(\text{I})}(f) = S_1^{(\text{I})}(f) + S_2^{(\text{I})}(f),$$

where $S_1^{(\text{I})}(\cdot)$ is absolutely continuous and $S_2^{(\text{I})}(\cdot)$ is a step function. Consider the real-valued process

$$X_t = \sum_{l=1}^{L} D_l \cos\left(2\pi f_l t \Delta_t + \phi_l\right) + \eta_t, \tag{518c}$$

which is the same as Equation (512b) with two substitutions: (1) the process mean μ is set to 0, and (2) the process $\{\eta_t\}$ is not necessarily white noise, but it is independent of each ϕ_l. The spectral representation theorem in combination with a decomposition of the orthogonal process $\{Z(f)\}$ states that

$$X_t = \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} e^{i2\pi f t \Delta_t} \, dZ(f) = \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} e^{i2\pi f t \Delta_t} \, dZ_1(f) + \int_{-f_{\mathcal{N}}}^{f_{\mathcal{N}}} e^{i2\pi f t \Delta_t} \, dZ_2(f),$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

where $\{Z_1(f)\}$ and $\{Z_2(f)\}$ are each orthogonal processes;

$$E\{|\mathrm{d}Z_1(f)|^2\} = S_\eta(f)\,\mathrm{d}f \text{ and } E\{|\mathrm{d}Z_2(f)|^2\} = \mathrm{d}S_2^{(\mathrm{I})}(f); \tag{519a}$$

$S_\eta(\cdot)$ is the SDF of $\{\eta_t\}$; and $E\{\mathrm{d}Z_1^*(f)\,\mathrm{d}Z_2(f')\} = 0$ for all f and f' (Priestley, 1981, pp. 252–3). Recall from Section 4.1 that, by putting

$$C_l = D_l \mathrm{e}^{\mathrm{i}\phi_l}/2 \text{ and } C_{-l} = C_l^*, \quad l = 1, \ldots, L,$$

we can write

$$X_t = \sum_{l=-L}^{L} C_l \mathrm{e}^{\mathrm{i}2\pi f_l t\,\Delta_\mathrm{t}} + \eta_t, \tag{519b}$$

where $C_0 \stackrel{\mathrm{def}}{=} 0$, $f_0 \stackrel{\mathrm{def}}{=} 0$ and $f_{-l} = -f_l$. Furthermore, $E\{X_t\} = 0$ and

$$\mathrm{var}\,\{X_t\} = \sum_{l=-L}^{L} E\{|C_l|^2\} + \mathrm{var}\,\{\eta_t\} = \sum_{l=-L}^{L} D_l^2/4 + \sigma_\eta^2. \tag{519c}$$

Corresponding to this result we have

$$E\{|\mathrm{d}Z_2(f)|^2\} = \begin{cases} E\{|C_l|^2\} = D_l^2/4, & f = \pm f_l; \\ 0, & \text{otherwise.} \end{cases} \tag{519d}$$

It follows from the argument leading to Equation (37b) that

$$\mathrm{cov}\,\{X_{t+\tau}, X_t\} = \sum_{l=1}^{L} D_l^2 \cos(2\pi f_l \tau\,\Delta_\mathrm{t})/2 + \mathrm{cov}\,\{\eta_{t+\tau}, \eta_t\}. \tag{519e}$$

Comments and Extensions to Section 10.3

[1] For complex exponentials, we have

$$Z_t = \sum_{l=1}^{L} D_l \mathrm{e}^{\mathrm{i}(2\pi f_l t\,\Delta_\mathrm{t} + \phi_l)} + \eta_t \tag{519f}$$

(this complex-valued process $\{Z_t\}$ should not be confused with the complex-valued process $\{Z(f)\}$ with orthogonal increments – the notation is unfortunately similar). This is the same as Equation (517a) except that (1) the process mean $\mu = 0$ and (2) the proper complex-valued process $\{\eta_t\}$ need not be a white noise process. If we set $C_l = D_l \exp(\mathrm{i}\phi_l)$ so that Equation (519f) becomes

$$Z_t = \sum_{l=1}^{L} C_l \mathrm{e}^{\mathrm{i}2\pi f_l t\,\Delta_\mathrm{t}} + \eta_t, \tag{519g}$$

we see that Equation (519g) is a one-sided version of Equation (519b); i.e., l is only positive in Equation (519g), but note that f_l can be either positive or negative. Thus $E\{Z_t\} = 0$ and

$$\mathrm{var}\,\{Z_t\} = \sum_{l=1}^{L} E\{|C_l|^2\} + \mathrm{var}\,\{\eta_t\} = \sum_{l=1}^{L} D_l^2 + \sigma_\eta^2. \tag{519h}$$

For this complex-valued model we also have

$$E\{|\mathrm{d}Z_2(f)|^2\} = \begin{cases} E\{|C_l|^2\} = D_l^2, & f = f_l; \\ 0, & \text{otherwise,} \end{cases} \tag{519i}$$

and

$$\mathrm{cov}\,\{Z_{t+\tau}, Z_t\} = \sum_{l=1}^{L} D_l^2 \mathrm{e}^{\mathrm{i}2\pi f_l \tau\,\Delta_\mathrm{t}} + \mathrm{cov}\,\{\eta_{t+\tau}, \eta_t\}. \tag{519j}$$

Proof of this final result is the subject of Exercise [10.3].

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

10.4 An Example from Tidal Analysis

To illustrate some of the ideas introduced so far in this chapter, we will look at the observed heights of sea levels as recorded by tide gauges. It should be noted that sea level as defined here excludes the effect of individual waves – tidal gauges do *not* measure the instantaneous height of the sea at any particular time but rather the average level about which the waves are oscillating. We can thus write

$$Y_t = X_t + \eta_t,$$

where Y_t is the observed height of the sea level at time t; X_t is the predicted height as given by Equation (512a) – with L' increased to L – under the assumption that $\alpha_l \stackrel{\text{def}}{=} \alpha_{t,l}$, $v_l \stackrel{\text{def}}{=} v_{t,l}$ and $u_l \stackrel{\text{def}}{=} u_{t,l}$ are varying slowly enough that they can be assumed to be constant over the span of data under analysis; and η_t is the unexplained component. Thus (Murray, 1964, 1965)

$$\begin{aligned} Y_t &= \mu + \sum_{l=1}^{L} \alpha_l H_l \cos\left(2\pi f_l t\, \Delta_{\mathrm{t}} + v_l + u_l - g_l\right) + \eta_t \\ &= \mu + \sum_{l=1}^{L} \alpha_l H_l \big[\cos\left(g_l\right) \cos\left(2\pi f_l t\, \Delta_{\mathrm{t}} + v_l + u_l\right) \\ &\qquad + \sin\left(g_l\right) \sin\left(2\pi f_l t\, \Delta_{\mathrm{t}} + v_l + u_l\right)\big] + \eta_t \\ &= \mu + \sum_{l=1}^{L} A_l \left[\alpha_l \cos\left(2\pi f_l t\, \Delta_{\mathrm{t}} + v_l + u_l\right)\right] \\ &\qquad + B_l \left[\alpha_l \sin\left(2\pi f_l t\, \Delta_{\mathrm{t}} + v_l + u_l\right)\right] + \eta_t, \end{aligned}$$

where $A_l \stackrel{\text{def}}{=} H_l \cos\left(g_l\right)$ and $B_l \stackrel{\text{def}}{=} H_l \sin\left(g_l\right)$. This is a slightly more complicated version of Equation (515b), and the process $\{\eta_t\}$ need not be white noise. The unknowns H_l and g_l can be found from A_l and B_l. The frequencies f_l are not in general related to the Fourier frequencies. The vertical lines in the top plot of Figure 521 indicate the locations of the $L = 102$ frequencies used in the model – the frequency $M_1 \doteq 0.0402557$ cycles/hour and its multiples $M_k = k \cdot M_1$ for $k = 2$, 3, 4, 5, 6 and 8 are indicated by thicker and longer lines than the other 95 frequencies.

It is convenient to write

$$Y_t = \sum_{l=0}^{2L} \beta_l \theta_{l,t} + \eta_t,$$

where, by definition, $\beta_0 = \mu$; $\beta_{2l-1} = H_l \cos\left(g_l\right)$; $\beta_{2l} = H_l \sin\left(g_l\right)$; $\theta_{0,t} = 1$;

$$\theta_{2l-1,t} = \alpha_l \cos\left(2\pi f_l t\, \Delta_{\mathrm{t}} + v_l + u_l\right) \quad \text{and} \quad \theta_{2l,t} = \alpha_l \sin\left(2\pi f_l t\, \Delta_{\mathrm{t}} + v_l + u_l\right).$$

Least squares then gives

$$\sum_t Y_t \theta_{m,t} = \sum_{l=0}^{2L} \beta_l \sum_t \theta_{l,t} \theta_{m,t}$$

for each of the $2L + 1$ parameters. These equations can be written in matrix form as

$$\boldsymbol{\Theta\beta} = \boldsymbol{c},$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

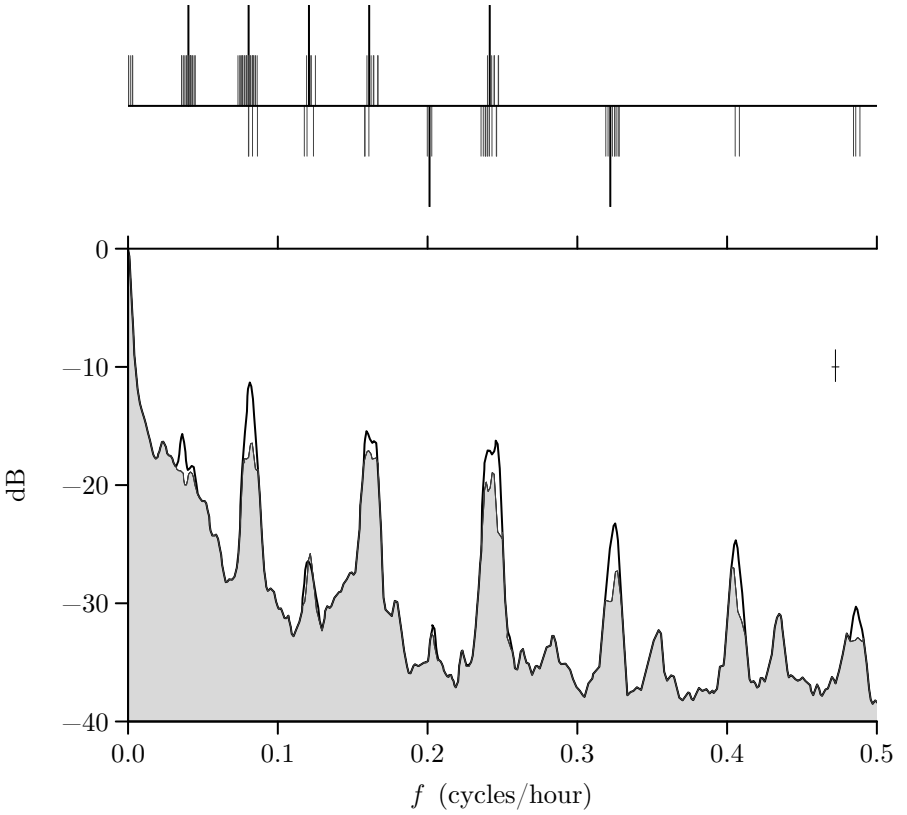

Figure 521 Tidal frequencies (upper plot) and estimated power spectra for Portsmouth residuals for 1971 (lower plot). In the upper plot, each frequency in the $L = 102$ model is indicated by a vertical line originating up or down from the thin horizontal line – there are thus 102 vertical lines in all, but their widths are such that most of them are not distinctly visible. The vertical lines going up indicate those frequencies that are in both the $L = 102$ and $L = 60$ models; the lines going down correspond to frequencies in the $L = 102$ model only. There are also seven lines that are thicker and twice as long as the others – five going up and two going down. These represent, from left to right, the tidal frequencies M_1, M_2, M_3, M_4, M_5, M_6 and M_8. In the lower plot, the thick curve is the spectrum for the 60 constituent model, while the thin curve at the top of the shaded area gives the spectrum for the 102 constituent model. A Parzen lag window with $m = 400$ was used. The height and width of the crisscross in the upper right-hand of this plot indicate, respectively, the width of a 95% confidence interval and the bandwidth measure $B_{\mathcal{U}}$ of Equation (256a).

where the matrix $\boldsymbol{\Theta}$ has $\sum_t \theta_{j-1,t}\theta_{i-1,t}$ for its (i,j)th element and

$$\boldsymbol{c} \stackrel{\text{def}}{=} \begin{bmatrix} \mu \\ \sum_t Y_t \alpha_1 \cos\left(2\pi f_1 t\, \Delta_{\mathrm{t}} + v_l + u_1\right) \\ \vdots \\ \sum_t Y_t \alpha_1 \sin\left(2\pi f_L t\, \Delta_{\mathrm{t}} + v_L + u_L\right) \end{bmatrix}.$$

With $\boldsymbol{\beta}$ estimated by $\hat{\boldsymbol{\beta}}$, H_l and g_l are found from

$$\hat{H}_l = \left(\hat{\beta}_{2l}^2 + \hat{\beta}_{2l-1}^2\right)^{1/2} \quad \text{and} \quad \hat{g}_l = \tan^{-1}\left(\hat{\beta}_{2l}/\hat{\beta}_{2l-1}\right), \quad l = 1, \ldots, L.$$

In standard linear regression, the noise process is required to be uncorrelated. In our current application, even before examining estimates of η_t (formed from $Y_t - \hat{Y}_t$, where $\hat{Y}_t$ is

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

the least squares estimate), we would expect the η_t terms to be correlated. For example, seasonal groupings of meteorological effects and, on a shorter scale, surge-generated groupings of residuals will be present. In general, for correlated noise, the maximum likelihood solution to the parameter estimation corresponds to *generalized* rather than ordinary least squares. Remarkably, ordinary least squares estimates are asymptotically efficient for a polynomial regression model such as is under consideration here (see, e.g., section 7.7 of Priestley, 1981, and also the Comments and Extensions [C&Es] for this section).

Walden (1982) calculated the harmonic constants H_l and g_l for Portsmouth, England, using three years of carefully edited consecutive hourly height data (1967–9) – over 26,000 points in all. The harmonic constants were estimated using least squares. These estimates were used to "predict" tidal elevations for each of the years 1962 and 1964–74. These predictions were in turn subtracted from the corresponding observed sea levels to produce residual series of hourly heights. This process was carried out for the so-called orthodox constituents ($L = 60$ harmonic terms comprising the primary frequencies plus some shallow-water frequencies) and for orthodox plus shallow-water terms ($L = 102$ harmonic terms) – see Figure 521. The lower plot of Figure 521 shows the power spectra of the 1971 Portsmouth residuals (note that, although the harmonic constants were computed using three years of data, each residual sequence is one year in length; i.e., $N = 365 \times 24 = 8760$). The thick curve and the thin curve tracing the top of the shaded area show, respectively, spectral estimates based upon the residuals for the $L = 60$ and 102 constituent models. A Parzen lag window was used for both estimates. After some testing, the smoothing parameter m for this window was chosen as 400. From Table 279 the smoothing window bandwidth B_W is $1.85/(m\,\Delta_t) \doteq 0.0046$ cycles/hour, and the bandwidth measure $B_{\mathcal{U}}$ of Equation (256a) is 0.0047. While these bandwidths are not sufficient to separate individual constituents, increasing m – i.e., decreasing the bandwidths – led to no visual improvement in the form of the spectral estimate. With $L = 102$, there is a diminution – as compared to the $L = 60$ analysis – of the contribution in the frequency bands close to integer multiples of 0.0805 cycles/hour. Since this frequency is equivalent to 1 cycle/12.42 hours (a semi-diurnal term), multiples of 2, 3 and so forth correspond to quarter-diurnal, sixth-diurnal and so forth terms. These are approximately the positions of the majority of the additional frequencies in the $L = 102$ model – see Figure 521. Note that, even for the $L = 102$ model, there is still considerable low-frequency power in the residual process so that it cannot be reasonably approximated as white noise.

How can we explain the residual power at low frequencies and in the tidal bands (i.e., the frequency ranges encompassing semi-, quarter-, eighth-diurnal, etc., tides)? The prediction model of Equation (512a) does not include nongravitational effects. For example, pressure fluctuations due to weather systems – which can induce a sea level response – have a continuous noise spectrum. Storm surges will show primarily in the band 0.03 to 0.08 cycles/hour. Land and sea breezes will cause discernible sea level changes and have associated spectral *lines* at 1 and 2 cycles/day as well as a continuous noise component (Munk and Cartwright, 1966). The overall spectrum is of course a mixed spectrum. At higher frequencies, residual power can arise through the omission of shallow-water constituents or through interaction between the tide and local mean sea level. Aliasing could also be involved. Suppose there is nonnegligible power in the *continuous time* residual series at harmonics of the main semidiurnal M_2 constituent at frequencies corresponding to M_{14} and M_{16}, i.e., approximately 0.5636 and 0.6441 cycles/hour. Sampling every hour gives a Nyquist frequency of $f_{\mathcal{N}} = 1/2$ cycle/hour. Since power from frequency $2f_{\mathcal{N}} - f$ would be aliased with f under this sampling scheme (see Equation (122b), recalling that spectra are symmetric about zero frequency), we note that 0.5636 cycles/hour aliases to 0.4364 cycles/hour, while 0.6441 aliases to 0.3559. Thus the residual power at M_{14} and M_{16} in the continuous series would appear at 0.436 and

0.356 cycles/hour in the sample spectrum. For Portsmouth there is a peak close to both of these frequencies (for other data – such as Southampton for 1955 – there is a noticeable peak about the first frequency only).

Comments and Extensions to Section 10.4

[1] If the SDF of the background continuum $\{\eta_t\}$ is colored, there is no efficiency to be gained *asymptotically* by including information on the covariance structure of the continuum. Hannan (1973), in examining a single sinusoid with frequency f_1, explained that the regression component is by nature a "signal" sent at a single frequency so that only the SDF of the background continuum at that frequency matters. Ultimately, as the sample size increases, we can gain nothing from knowledge of the covariance structure of the background continuum. He also pointed out that, for finite sample sizes and a continuum with a very irregular SDF near f_1, the influence of the covariance structure of the continuum might be appreciable even though its influence vanishes as the sample size increases.

10.5 A Special Case of Unknown Frequencies

Up to this point we have assumed that the frequencies $f_l \in (0, f_{\mathcal{N}}]$ in Equation (512b) are known in advance (as in the example of Section 10.4). If they are not known, then the models are no longer multiple linear regressions. The usual procedure is to somehow estimate the f_l terms and then to use these estimates in finding the other unknowns.

The classic procedure for estimating f_l – which dates from the 1890s – is based upon examining the periodogram, which we have already met in Chapter 6. The basic idea behind periodogram analysis in the current context is as follows. From Equation (512b), note that D_l^2 represents the squared amplitude of the sinusoid with frequency f_l. We also have $D_l^2 = A_l^2 + B_l^2$ from reexpressing Equation (512b) as Equation (515b). If $f_l \neq f_{\mathcal{N}}$, we can estimate D_l^2 by using the estimators for A_l and B_l in Equation (515c) to obtain

$$\begin{aligned}\hat{D}_l^2 &= \left[\frac{2}{N}\sum_{t=0}^{N-1} X_t \cos\left(2\pi f_l t\,\Delta_{\mathrm{t}}\right)\right]^2 + \left[\frac{2}{N}\sum_{t=0}^{N-1} X_t \sin\left(2\pi f_l t\,\Delta_{\mathrm{t}}\right)\right]^2 \\ &= \frac{4}{N^2}\left|\sum_{t=0}^{N-1} X_t \mathrm{e}^{-\mathrm{i}2\pi f_l t\,\Delta_{\mathrm{t}}}\right|^2 \end{aligned} \tag{523a}$$

(a practical note: if the process mean μ is unknown, we typically need to replace X_t by $X_t - \overline{X}$ – see the discussion surrounding Equation (184)). It follows from Equation (519d) that the contribution to the power spectrum at $f = \pm f_l$ is $D_l^2/4$, which is estimated by $\hat{D}_l^2/4$. Recalling Equation (170d), the periodogram at f_l is

$$\hat{S}^{(\mathrm{P})}(f_l) = \frac{\Delta_{\mathrm{t}}}{N}\left|\sum_{t=0}^{N-1} X_t \mathrm{e}^{-\mathrm{i}2\pi f_l t\,\Delta_{\mathrm{t}}}\right|^2 = \frac{N\,\Delta_{\mathrm{t}}}{4}\hat{D}_l^2, \tag{523b}$$

so the scheme is to plot $\hat{S}^{(\mathrm{P})}(f)$ versus f and look for "large" values of $\hat{S}^{(\mathrm{P})}(f)$. Note that, for any $f \in (0, f_{\mathcal{N}})$ not equal to an f_l in Equation (512b), we can imagine adding an $(L+1)$st term to the model with $f_{L+1} = f$ and $D_{L+1} = 0$. The estimator $\hat{D}_{L+1}^2$ should thus be an estimator of zero and consequently should be small.

Now let us assume that each f_l in Equation (512b) is a positive Fourier frequency not equal to $f_{\mathcal{N}}$ (hence each f_l can be written in the form $k_l/(N\,\Delta_{\mathrm{t}})$, where k_l is an integer – unique for each different f_l – such that $0 < k_l < N/2$).

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

▷ **Exercise [524]** Show that

$$E\{\hat{S}^{(\mathrm{P})}(f_l)\} = \frac{N\,\Delta_\mathrm{t}}{4} D_l^2 + \sigma_\epsilon^2\,\Delta_\mathrm{t}. \tag{524a}$$ ◁

The first term on the right-hand side above is $N\,\Delta_\mathrm{t}$ times the contribution to the power spectrum from the line at $f = \pm f_l$, while the second is the value of the white noise contribution to the spectrum. If $f_l \in (0, f_\mathcal{N})$ is one of the Fourier frequencies *not* in Equation (512b), then

$$E\{\hat{S}^{(\mathrm{P})}(f_l)\} = \sigma_\epsilon^2\,\Delta_\mathrm{t}$$

by the artifact of adding a fictitious term to the model with zero amplitude (we cannot use this same trick to handle the case $f_l = f_\mathcal{N}$ – the above equation still holds then, but a special argument is needed and is left as an exercise for the reader).

Thus, in the case where each f_l is of a special nature (i.e., a Fourier frequency), plotting $\hat{S}^{(\mathrm{P})}(f_l)$ versus f_l on the grid of standard frequencies and searching this plot for large values is a reasonable way of finding which frequencies f_l belong in the model (Section 6.6 discusses the sampling properties for $\hat{S}^{(\mathrm{P})}(f_l)$).

Comments and Extensions to Section 10.5

[1] For the complex exponential model of Equation (517a), the equivalent of Equation (524a) is

$$E\{\hat{S}^{(\mathrm{P})}(f_l)\} = \left(N D_l^2 + \sigma_\epsilon^2\right)\Delta_\mathrm{t},$$

where now $f_l \in (-f_\mathcal{N}, 0) \cup (0, f_\mathcal{N})$.

10.6 General Case of Unknown Frequencies

Assuming the general case of Equation (512b) for which f_l is not necessarily one of the Fourier frequencies, we have the following result, valid for all f and with μ known to be zero:

$$E\{\hat{S}^{(\mathrm{P})}(f)\} = \sigma_\epsilon^2\,\Delta_\mathrm{t} + \sum_{l=1}^{L} \frac{D_l^2}{4}\left[\mathcal{F}(f + f_l) + \mathcal{F}(f - f_l)\right], \tag{524b}$$

where $\mathcal{F}(\cdot)$ is Fejér's kernel – see Equation (174c) for its definition and also Figure 176. The proof of Equation (524b) is Exercise [10.4], which can be most easily seen by expressing the increments of $S_2^{(\mathrm{I})}(\cdot)$ in Section 10.3 in terms of the Dirac delta function (see Equation (519a) along with Equation (519d)):

$$\mathrm{d}S_2^{(\mathrm{I})}(f) = \sum_{l=1}^{L} \frac{D_l^2}{4}\left[\delta(f - f_l) + \delta(f + f_l)\right]\mathrm{d}f.$$

What does Equation (524b) tell us about the usefulness of the periodogram for identifying f_l in Equation (512b)? Consider the case where there is only one frequency, f_1 (i.e., $L = 1$ in Equation (512b)). Equation (524b) reduces to a constant term, $\sigma_\epsilon^2\,\Delta_\mathrm{t}$, plus the sum of $D_1^2/4$ times two Fejér kernels centered at frequencies $\pm f_1$. If $f_1 = k_1/(N\,\Delta_\mathrm{t})$ for some integer k_1 and if we evaluate $E\{\hat{S}^{(\mathrm{P})}(f)\}$ at just the Fourier frequencies, Equation (524b) is in agreement with our earlier analysis: a plot of $E\{\hat{S}^{(\mathrm{P})}(f)\}$ versus those frequencies has a single large value at f_1 against a "background" of level $\sigma_\epsilon^2\,\Delta_\mathrm{t}$. If, however, f_1 is not of the form $k_1/(N\,\Delta_\mathrm{t})$, the plot is more complicated since we are now sampling Fejér's kernel at points other than the "null points." For example, if f_1 falls exactly halfway between two of the

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

Fourier frequencies and if we only plot $E\{\hat{S}^{(\mathrm{P})}(\cdot)\}$ at the Fourier frequencies, we would find the two largest values at $f_1 \pm 1/(2N\,\Delta_\mathrm{t})$; moreover, these values would have approximately the same amplitude due to the symmetry of the Fejér kernel (the qualifier "approximately" is needed due to the contribution from the second Fejér kernel centered at $-f_1$). For large enough N and for f_1 not too close to zero or Nyquist frequency, the size of these two values is about 40% of the value of $E\{\hat{S}^{(\mathrm{P})}(f_1)\}$ (see Exercise [10.5]). Thus, we need to plot $\hat{S}^{(\mathrm{P})}(\cdot)$ at more than just the $\lfloor N/2 \rfloor + 1$ Fourier frequencies. In practice a grid twice as fine as that of the Fourier frequencies suffices – at least initially – to aid us in searching for periodic components. An additional rationale for this advice is that $\hat{S}^{(\mathrm{P})}(\cdot)$ is a trigonometric polynomial of degree $N-1$ and hence is uniquely determined by its values at N points. (There is, however, a potential penalty for sampling twice as finely as the Fourier frequencies – the loss of independence between the periodogram ordinates. This presents no problem when we are merely examining plots of the periodogram ordinates to assess appropriate terms for Equation (512b), but it is a problem in constructing valid statistical tests based upon periodogram ordinates.)

Let us consider the relationship between the periodogram and least squares estimation for the single-frequency model

$$X_t = D\cos(2\pi f_1 t\,\Delta_\mathrm{t} + \phi) + \epsilon_t = A\cos(2\pi f_1 t\,\Delta_\mathrm{t}) + B\sin(2\pi f_1 t\,\Delta_\mathrm{t}) + \epsilon_t. \quad (525a)$$

This is the model described by Equations (512b) and (513) with μ again known to be zero. We can find the exact least squares estimators of A and B, say $\hat{A}$ and $\hat{B}$, from the normal equations

$$\sum_{t=0}^{N-1} X_t \cos(2\pi f_1 t\,\Delta_\mathrm{t}) = \hat{A}\sum_{t=0}^{N-1} \cos^2(2\pi f_1 t\,\Delta_\mathrm{t}) + \hat{B}\sum_{t=0}^{N-1} \cos(2\pi f_1 t\,\Delta_\mathrm{t})\sin(2\pi f_1 t\,\Delta_\mathrm{t})$$

$$\sum_{t=0}^{N-1} X_t \sin(2\pi f_1 t\,\Delta_\mathrm{t}) = \hat{A}\sum_{t=0}^{N-1} \cos(2\pi f_1 t\,\Delta_\mathrm{t})\sin(2\pi f_1 t\,\Delta_\mathrm{t}) + \hat{B}\sum_{t=0}^{N-1} \sin^2(2\pi f_1 t\,\Delta_\mathrm{t})$$

(cf. Equations (514b) and (514c)). The explicit expressions for $\hat{A}$ and $\hat{B}$ are somewhat unwieldy but simplify considerably if we restrict f_1 to be of the form $f_k' = k/(2N\,\Delta_\mathrm{t})$, i.e., a frequency on the grid of frequencies twice as fine as the Fourier frequencies. If we rule out the zero and Nyquist frequencies so that $0 < f_k' < f_{\mathcal{N}}$, we have

$$\sum_{t=0}^{N-1} \cos^2(2\pi f_k' t\,\Delta_\mathrm{t}) = \sum_{t=0}^{N-1} \sin^2(2\pi f_k' t\,\Delta_\mathrm{t}) = \frac{N}{2}$$

and

$$\sum_{t=0}^{N-1} \cos(2\pi f_k' t\,\Delta_\mathrm{t})\sin(2\pi f_k' t\,\Delta_\mathrm{t}) = 0$$

(see Exercise [1.3a]). The simplified normal equations lead to

$$\hat{A} = \frac{2}{N}\sum_{t=0}^{N-1} X_t \cos(2\pi f_k' t\,\Delta_\mathrm{t}) \;\text{ and }\; \hat{B} = \frac{2}{N}\sum_{t=0}^{N-1} X_t \sin(2\pi f_k' t\,\Delta_\mathrm{t})$$

as the exact least squares estimators of A and B when $f_1 = f_k'$. The residual sum of squares is $\mathrm{SS}(f_k')$, where

$$\mathrm{SS}(f) \stackrel{\mathrm{def}}{=} \sum_{t=0}^{N-1} \left(X_t - \left[\hat{A}\cos(2\pi f t\,\Delta_\mathrm{t}) + \hat{B}\sin(2\pi f t\,\Delta_\mathrm{t})\right]\right)^2. \quad (525b)$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

If we substitute for $\hat{A}$ and $\hat{B}$ and simplify somewhat (part of the burden of Exercise [10.6a]), we obtain

$$\text{SS}(f'_k) = \sum_{t=0}^{N-1} X_t^2 - \frac{2}{\Delta_t} \hat{S}^{(\text{P})}(f'_k), \tag{526a}$$

thus establishing an interesting connection between the periodogram and the residual sum of squares from a single-frequency regression model. This connection, however, banks on the single frequency f_1 being of the special form $f'_k = k/(2N\,\Delta_t)$. Exercise [10.7] demonstrates that this connection breaks down when f_1 is not of this special form; however, as long as $f_1 \in (0, f_{\mathcal{N}})$ but is not too close to the zero or Nyquist frequencies, we can expect $\hat{S}^{(\text{P})}(f_1)$ to be closely approximated by

$$\widetilde{S}(f_1) \stackrel{\text{def}}{=} \frac{\Delta_t}{2} \left[\sum_{t=0}^{N-1} X_t^2 - \text{SS}(f_1) \right]. \tag{526b}$$

We assumed f_1 to be known, but in general it will not be. As N gets large, any f_1 satisfying $0 < f_1 < f_{\mathcal{N}}$ will become closer and closer to a frequency of the form $f'_k = k/(2N\,\Delta_t)$, and so Equation (526a) says that any large-sample results based upon *maximizing* the periodogram will be the same as those based on *minimizing* the residual sum of squares for a single-frequency regression model.

In the case where

[1] the error terms $\{\epsilon_t\}$ in the single-frequency model are an independent and identically distributed sequence of RVs with zero mean and finite variance and
[2] $\hat{f}_1$ represents the frequency of the maximum value of the periodogram,

Whittle (1952) and Walker (1971) showed that

$$E\{\hat{f}_1\} = f_1 + O\left(N^{-1}\right) \quad \text{and} \quad \text{var}\,\{\hat{f}_1\} \approx \frac{3}{N^3 R\,\pi^2 \Delta_t^2}, \tag{526c}$$

where

$$R \stackrel{\text{def}}{=} \frac{A^2 + B^2}{2\sigma_\epsilon^2} = \frac{D^2}{2\sigma_\epsilon^2} \tag{526d}$$

is the signal-to-noise ratio; i.e., since

$$\text{var}\,\{X_t\} = \frac{A^2 + B^2}{2} + \sigma_\epsilon^2,$$

R is just the ratio of the two terms on the right-hand side, the first due to the sinusoid (the "signal"), and the second to the white noise process (the "noise"). Since the mean square error (MSE) is the variance plus the bias squared, the contribution of the variance to the MSE is $O\left(N^{-3}\right)$, whereas that of the bias is $O\left(N^{-2}\right)$. Hence the MSE of $\hat{f}_1$ is dominated by bias rather than variance.

The $O\left(N^{-3}\right)$ rate at which $\text{var}\,\{\hat{f}_1\}$ decreases is somewhat surprising (the rate $O\left(N^{-1}\right)$ is more common, as in Equation (164b) for example), but Bloomfield (2000, p. 32) gives the following argument to lend credence to it (see also Quinn, 2012, section 3). Suppose R is reasonably large so that, by inspecting a realization of $X_0, X_1, \ldots, X_{N-1}$, we can see that there are, say, between M and $M+1$ complete cycles of the sinusoid. Therefore we know that

$$\frac{M}{N\,\Delta_t} \le f_1 \le \frac{(M+1)}{N\,\Delta_t}. \tag{526e}$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

It follows that $|\tilde{f}_1 - f_1| \leq 1/(N\,\Delta_t)$ for any estimator $\tilde{f}_1$ satisfying the above. This implies that

$$\text{var}\,\{\tilde{f}_1\} \leq \frac{1}{N^2 \Delta_t^2} = O\left(N^{-2}\right).$$

If we can estimate f_1 to within order $O\left(N^{-2}\right)$ with only the constraint that $\tilde{f}_1$ satisfies Equation (526e), it is plausible that the fancier periodogram estimator is $O\left(N^{-3}\right)$. Rice and Rosenblatt (1988) show that the product of the amplitude of the sinusoid and the sample size must be quite large for the asymptotic theory to be meaningful.

If there is more than just a single frequency in Equation (512b) (i.e., $L > 1$), the form of $E\{\hat{S}^{(\text{P})}(\cdot)\}$ in Equation (524b) can be quite complicated due to the superposition of $2L$ Fejér kernels centered at the points $\{\pm f_l\}$ together with the constant term $\sigma_\epsilon^2\,\Delta_t$. In addition to interference due to the sidelobes, the widths of the central lobes of the superimposed kernels can contribute to the confusion. For example, if two of the f_l terms are close together and N is small enough, a plot of $E\{\hat{S}^{(\text{P})}(\cdot)\}$ might only indicate a single broad peak. Likewise, a frequency with a small D_l^2 might be effectively masked by a "sidelobe" from Fejér's kernel due to a nearby frequency with a large amplitude (see Section 10.8 and Figure 536).

In practical applications, we only observe $\hat{S}^{(\text{P})}(\cdot)$, which can differ substantially from $E\{\hat{S}^{(\text{P})}(\cdot)\}$ due to the presence of the white noise component. Our problem of identifying frequency components is much more difficult due to the distortions that arise from sampling variations. One effect of the white noise component is to introduce "spurious" peaks into the periodogram (see Figure 210b(a)). Recall that, in the null case when $L = 0$ and the white noise process is Gaussian, the $\hat{S}^{(\text{P})}(f)$ terms at the Fourier frequencies are independent χ^2 RVs with one or two degrees of freedom (see Section 6.6). Because they are independent RVs, in portions of the spectrum where the white noise component dominates, we can expect to see a rather choppy behavior with many "peaks" and "valleys." Furthermore, the χ_1^2 and χ_2^2 distributions have heavy tails (Figure 210a(a) shows the density function for the latter); in a random sample from such a distribution we can expect one or more observations that appear unusually large. These large observations could be mistaken for peaks due to a harmonic component. This fact points out the need for statistical tests to help us decide if a peak in the periodogram can be considered statistically significant (see Sections 10.9 and 10.10).

Comments and Extensions to Section 10.6

[1] Hannan (1973) looked at the $L = 1$ case of Equation (512b) and took the additive noise to be a strictly stationary process rather than just white noise. He found that

$$\lim_{N\to\infty} N\left(\hat{f}_1 - f_1\right) = 0 \text{ almost surely}$$

for $0 < f_1 < f_{\mathcal{N}}$. He also considered the case $f_1 = f_{\mathcal{N}}$ and found a stronger result: there is an integer N_0 – determined by the realization – such that $\hat{f}_1 = f_1$ for all $N \geq N_0$ with $\mathbf{P}[N_0 < \infty] = 1$ (he found this latter result also holds when f_1 is set to zero, but under the additional assumption that $\mu = 0$ in Equation (512b)).

[2] Suppose that $f_k' = k/(N'\,\Delta_t)$ maximizes the periodogram $\hat{S}^{(\text{P})}(\cdot)$ over a grid of equally spaced frequencies defined by $f_j' = j/(N'\,\Delta_t)$ for $1 \leq j < N'/2$ (the choice $N' = N$ yields the grid of Fourier frequencies, but $N' > N$ is also of interest if we pad $N' - N$ zeros to our time series for use with an FFT algorithm – see C&E [1] for Section 3.11). Suppose, however, that we are interested in obtaining the frequency maximizing $\hat{S}^{(\text{P})}(\cdot)$ over *all* frequencies. From Equation (170d), we can write

$$\hat{S}^{(\text{P})}(f) = \Delta_t \sum_{\tau=-(N-1)}^{N-1} \hat{s}_\tau^{(\text{P})} e^{-i2\pi f \tau\,\Delta_t} = \Delta_t \left[\hat{s}_0^{(\text{P})} + 2\sum_{\tau=1}^{N-1} \hat{s}_\tau^{(\text{P})} \cos\left(2\pi f \tau\,\Delta_t\right)\right],$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

where $\{\hat{s}_\tau^{(\mathrm{P})}\}$ is the biased estimator of the ACVS. If we let $\omega \stackrel{\text{def}}{=} 2\pi f\,\Delta_\mathrm{t}$ and

$$g(\omega) \stackrel{\text{def}}{=} \sum_{\tau=1}^{N-1} \hat{s}_\tau^{(\mathrm{P})} \cos(\omega\tau),$$

then a peak in $g(\cdot)$ at ω corresponds to a peak in $\hat{S}^{(\mathrm{P})}(\cdot)$ at $f = \omega/(2\pi\,\Delta_\mathrm{t})$. The first and second derivatives of $g(\cdot)$ are

$$g'(\omega) = -\sum_{\tau=1}^{N-1} \tau \hat{s}_\tau^{(\mathrm{P})} \sin(\omega\tau) \text{ and } g''(\omega) = -\sum_{\tau=1}^{N-1} \tau^2 \hat{s}_\tau^{(\mathrm{P})} \cos(\omega\tau). \tag{528a}$$

Note that a peak in $g(\cdot)$ corresponds to a root in $g'(\cdot)$. If $\omega^{(0)} \stackrel{\text{def}}{=} 2\pi f_k'\,\Delta_\mathrm{t}$ is not too far from the true location of the peak in $g(\cdot)$, we can apply the Newton–Raphson method to recursively compute

$$\omega^{(j)} = \omega^{(j-1)} - \frac{g'(\omega^{(j-1)})}{g''(\omega^{(j-1)})} \text{ for } j = 1, 2, \ldots, J,$$

stopping when $\left|\omega^{(J)} - \omega^{(J-1)}\right|$ is smaller than some specified tolerance. The value $\omega^{(J)}/(2\pi\,\Delta_\mathrm{t})$ is taken to be the location of the peak in $\hat{S}^{(\mathrm{P})}(\cdot)$ (Newton and Pagano, 1983; Newton, 1988, section 3.9.2).

In practice this scheme can fail if f_k' is not sufficiently close to the true peak location (Quinn et al., 2008; Quinn, 2012, section 9). Section 9.4 of Press et al. (2007) describes a useful method for finding the root of $g'(\cdot)$, call it $\omega^{(\mathrm{r})}$, using a Newton–Raphson scheme in combination with bisection. This method can be applied here if we can find two values of ω – call them $\omega^{(\mathrm{L})}$ and $\omega^{(\mathrm{U})}$ – that bracket the root; i.e., we have $\omega^{(\mathrm{L})} < \omega^{(\mathrm{r})} < \omega^{(\mathrm{U})}$ with $g'(\omega^{(\mathrm{L})}) > 0$ and $g'(\omega^{(\mathrm{U})}) < 0$. If N' is sufficiently large (usually $N' = 4N$ does the trick), these bracketing values can be taken to be $2\pi f_{k-1}'\,\Delta_\mathrm{t}$ and $2\pi f_{k+1}'\,\Delta_\mathrm{t}$. (There can also be problems if the equation for $g''(\omega)$ in Equation (528a) cannot be accurately computed, in which case *Brent's method* (Press et al., 2007, section 9.3) can be used to find the bracketed root of $g'(\cdot)$.)

[3] The relationship between the periodogram and the regression-based $\widetilde{S}(\cdot)$ provides motivation for the *Lomb–Scargle periodogram* (Lomb, 1976; Scargle, 1982; see also section 13.8 of Press et al., 2007, and Stoica et al., 2009). This periodogram handles an irregularly sampled time series $X(t_0)$, $X(t_1)$, $\ldots$, $X(t_{N-1})$, $t_0 < t_1 < \cdots < t_{N-1}$, in the context of a continuous parameter analog of the single-frequency model of Equation (525a):

$$X(t_n) = A\cos(2\pi f_1[t_n - c]) + B\sin(2\pi f_1[t_n - c]) + \epsilon_{t_n}, \tag{528b}$$

$n = 0, 1, \ldots, N-1$, where c is a constant (yet to be specified), and the errors $\{\epsilon_{t_n}\}$ are uncorrelated Gaussian RVs with zero mean and variance σ_ϵ^2. The least squares estimators $\hat{A}$ and $\hat{B}$ of A and B are the values minimizing the residual sum of squares

$$\begin{aligned} \mathrm{SS}_c(A, B) &\stackrel{\text{def}}{=} \|\boldsymbol{X} - \boldsymbol{H}_c\boldsymbol{\beta}\|^2 \\ &= \sum_{n=0}^{N-1} \left(X(t_n) - [A\cos(2\pi f_1[t_n - c]) + B\sin(2\pi f_1[t_n - c])]\right)^2, \end{aligned} \tag{528c}$$

where $\boldsymbol{X}^T \stackrel{\text{def}}{=} \left[X(t_0), X(t_1), \ldots, X(t_{N-1})\right]$, $\boldsymbol{\beta}^T \stackrel{\text{def}}{=} [A, B]$ and

$$\boldsymbol{H}_c^T \stackrel{\text{def}}{=} \begin{bmatrix} \cos(2\pi f_1[t_0 - c]) & \cos(2\pi f_1[t_1 - c]) & \cdots & \cos(2\pi f_1[t_{N-1} - c]) \\ \sin(2\pi f_1[t_0 - c]) & \sin(2\pi f_1[t_1 - c]) & \cdots & \sin(2\pi f_1[t_{N-1} - c]) \end{bmatrix}$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

(cf. Equations (514a) and (525b)). The Lomb–Scargle periodogram is defined as

$$\hat{S}^{(\mathrm{LS})}(f_1) = \frac{1}{2}\left[\sum_{n=0}^{N-1} X^2(t_n) - \mathrm{SS}_c(\hat{A}, \hat{B})\right] \tag{529a}$$

(cf. Equation (526b)). The above opens up the possibility of a dependence on a shift c in the t_n's, which is undesirable since, to quote Scargle (1982), "... the power spectrum is meant to measure harmonic content of signals without regard to phase." In fact there is no such dependence: Exercise [10.8a] is to show that $\mathrm{SS}_c(\hat{A}, \hat{B})$ is the same no matter what c is (however, in general $\hat{A}$ and $\hat{B}$ do depend on c). Setting c equal to

$$\tilde{c} = \frac{1}{4\pi f_1} \tan^{-1}\left[\frac{\sum_n \sin(4\pi f_1 t_n)}{\sum_n \cos(4\pi f_1 t_n)}\right] \tag{529b}$$

allows us to express the Lomb–Scargle periodogram succinctly as

$$\hat{S}^{(\mathrm{LS})}(f_1) = \frac{1}{2}\left(\frac{\left[\sum_n X(t_n)\cos(2\pi f_1[t_n - \tilde{c}])\right]^2}{\sum_n \cos^2(2\pi f_1[t_n - \tilde{c}])} + \frac{\left[\sum_n X(t_n)\sin(2\pi f_1[t_n - \tilde{c}])\right]^2}{\sum_n \sin^2(2\pi f_1[t_n - \tilde{c}])}\right). \tag{529c}$$

Verification of the above is the burden of Exercise [10.8b], while Exercise [10.8c] shows that $\tilde{c}$ leads to the corresponding least squares estimators $\hat{A}$ and $\hat{B}$ being uncorrelated (this point is expanded upon in C&E [2] for Section 10.9). There is, however, no real need to use Equations (529c) and (529b) since the Lomb–Scargle periodogram can just as well be computed via Equation (529a) using standard linear regression techniques with c set to zero.

In the case of equally spaced data $t_n = t_0 + n\,\Delta_t$, the Lomb–Scargle periodogram – after multiplication by Δ_t – is equal to the usual periodogram if f_1 is equal to $k/(2N\,\Delta_t)$ for some k satisfying $0 < k < N$. It should also be noted that the Lomb–Scargle periodogram arises from the single-frequency model of Equation (528b) and does not in general offer a proper decomposition of the sample variance of a time series across frequencies as does the usual periodogram; i.e., while we can take the usual periodogram to be a primitive SDF estimator, we should not in general regard the Lomb–Scargle periodogram as such. Finally, we note that formulation of the Lomb–Scargle periodogram in terms of Equation (528b) presumes that the mean of the time series is zero (in practice, this assumption is handled by centering the time series with its sample mean and taking $X(t_n)$ to be the centered series). Alternatively we can add a term μ to the model of Equation (528b) to handle a nonzero mean (cf. Equations (525a) and (513)). We can then estimate μ along with A and B via least squares (Cumming et al., 1999; Zechmeister and Kürster, 2009).

[4] The relationship between the periodogram and regression analysis also provides motivation for the *Laplace periodogram* (Li, 2008; see also section 5.5 of Li, 2014). Assuming the single-frequency model of Equation (525a) with f_1 set to a frequency of the form $f'_k = k/(2N\,\Delta_t)$, $0 < k < N$, and letting $\boldsymbol{\beta}^T = [A, B]$ as before, we can express the least squares estimator of $\boldsymbol{\beta}$ as

$$\hat{\boldsymbol{\beta}}^{(\mathrm{LS})} = \underset{A,B}{\operatorname{argmin}} \sum_{t=0}^{N-1} \left|X_t - A\cos(2\pi f'_k t\,\Delta_t) - B\sin(2\pi f'_k t\,\Delta_t)\right|^2. \tag{529d}$$

The usual periodogram and $\hat{\boldsymbol{\beta}}^{(\mathrm{LS})}$ are related by

$$\hat{S}^{(\mathrm{P})}(f'_k) = \frac{N}{4\,\Delta_t}\|\hat{\boldsymbol{\beta}}^{(\mathrm{LS})}\|^2 \tag{529e}$$

(see Equation (588a), which falls out from Exercise [10.6a]). Replacing the square modulus in Equation (529d) by just the modulus leads to the least absolute deviations estimator:

$$\hat{\boldsymbol{\beta}}^{(\mathrm{LAD})} = \underset{A,B}{\operatorname{argmin}} \sum_{t=0}^{N-1} \left|X_t - A\cos(2\pi f'_k t\,\Delta_t) - B\sin(2\pi f'_k t\,\Delta_t)\right|.$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

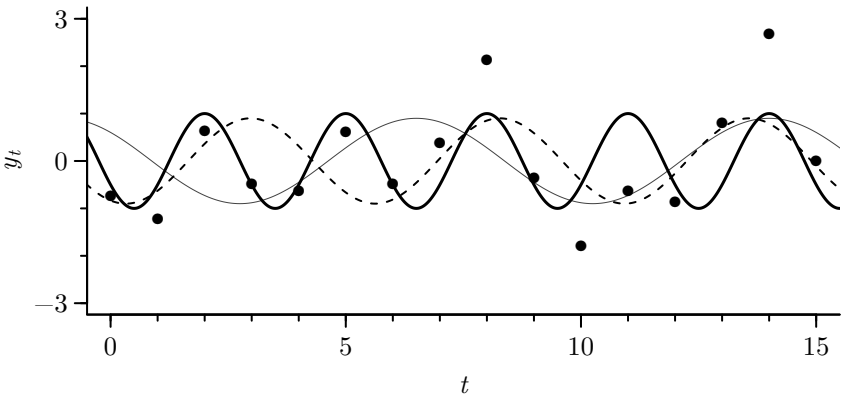

Figure 530 Sinusoids with periods 7.5 (thin solid curve), 5.3 (dashed) and 3 (thick solid) along with their summation at $t = 0, 1, \ldots, 15$ (dots).

In analogy to Equation (529e) the Laplace periodogram is defined by

$$\hat{S}^{(\mathrm{L})}(f'_k) = \frac{N}{4\,\Delta_\mathrm{t}} \|\hat{\boldsymbol{\beta}}^{(\mathrm{LAD})}\|^2$$

(there is, however, no need to restrict this definition to frequencies of the form f'_k).

Li (2008) gives two compelling arguments for the Laplace periodogram. First, for regression models such as Equation (525a), least squares estimators of the parameters have well-known optimality properties when the errors are Gaussian, but these estimators can perform poorly when the errors come from a heavy-tailed distribution. This fact is the inspiration for robust regression analysis, whose goal is to find estimators that work well in non-Gaussian settings. Along with other robust estimators, least absolute deviations estimators perform better than least squares estimators for certain non-Gaussian distributions, suggesting use of the Laplace periodogram to ascertain the harmonic content of a heavy-tailed time series. Second, recall that the periodogram is a primitive estimator of SDFs that are Fourier transforms of ACVSs. The Laplace periodogram is an estimator of the Fourier transform of a sequence that depends upon zero-crossing rates as defined by $\mathbf{P}[X_{t+\tau} X_t < 0]$, $\tau \in \mathbb{Z}$. These rates provide an interpretable – and robust – characterization of stationary processes alternative to that provided by the ACVS.

10.7 An Artificial Example from Kay and Marple

Let us illustrate some of the points in the previous sections by looking at an artificial example due to Kay and Marple (1981, p. 1386). Figure 530 shows three different sinusoids and their weighted summation

$$y_t = 0.9 \cos\left(2\pi[t+1]/7.5\right) + 0.9 \cos\left(2\pi[t+1]/5.3 + \pi/2\right) + \cos\left(2\pi[t+1]/3\right) \quad (530)$$

at indices $t = 0, 1, \ldots, 15$ (the use of $t + 1$ in the above is merely to replicate the Kay and Marple example using our standard zero-based indices). We can regard this time series as a realization of the model of Equation (511) with $\mu = 0$, $L = 3$ and $\Delta_\mathrm{t} = 1$. The frequencies of the three sinusoids are $1/7.5 = 0.133$ (thin solid curve), $1/5.3 = 0.189$ (dashed) and 1/3 (thick solid). This series of length 16 has about 2+ cycles of the first sinusoid, 3 cycles of the second and 5+ cycles of the third.

In Figure 531 we study the effect of grid size on our ability to detect the presence of harmonic components when examining the periodogram $\hat{S}^{(\mathrm{P})}(\cdot)$ for $y_0, y_1, \ldots, y_{15}$. The dots in the top plot show the value of $10 \log_{10} \hat{S}^{(\mathrm{P})}(\cdot)$ at the nine Fourier frequencies $f_k =$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

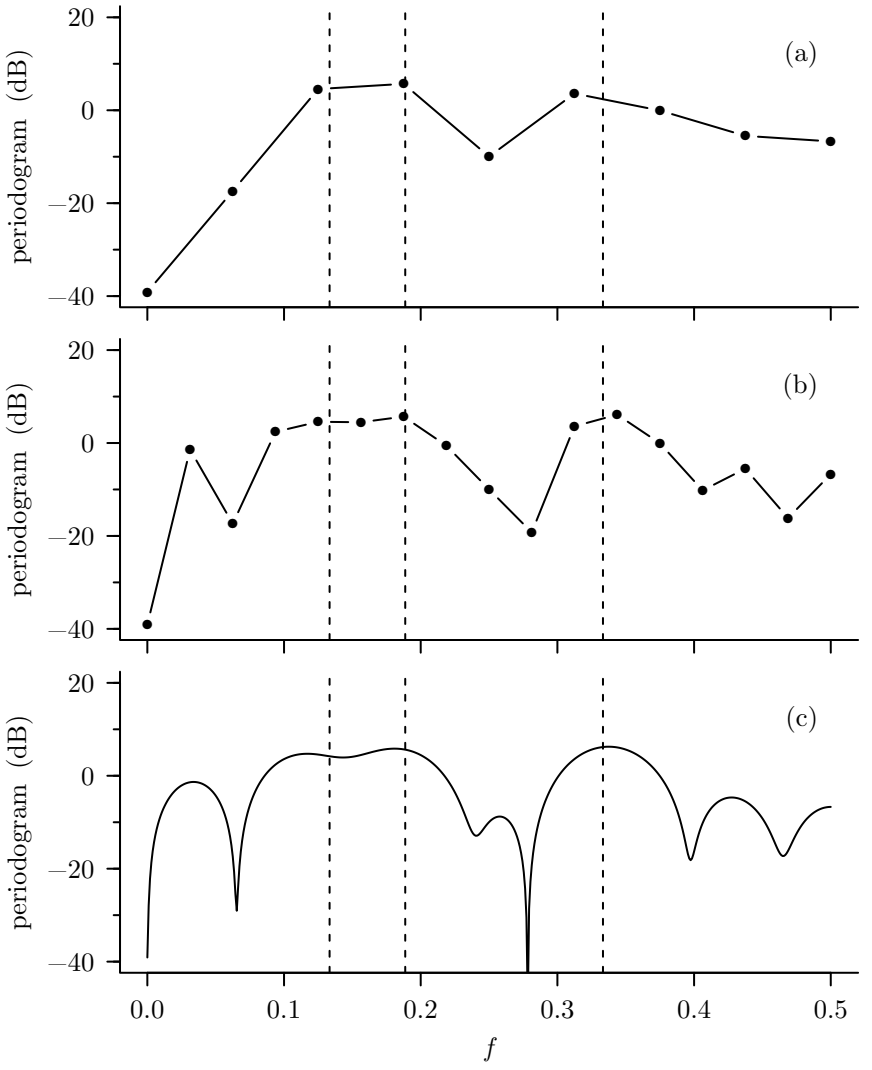

Figure 531 Effect of grid size on interpretation of the periodogram. For the time series of length 16 shown in Figure 530, the plots show the periodogram (on a decibel scale) versus frequency at the nine Fourier frequencies (dots in top plot, connected by lines), on a grid of frequencies twice as fine as the Fourier frequencies (17 dots in middle plot), and on a finely spaced grid of 513 frequencies (bottom).

$k/N = k/16$ for $k = 0, 1, \ldots, 8$. The three dashed vertical lines indicate the locations of the frequencies of the three sinusoids that form $\{y_t\}$. There are two broad peaks visible, the one with the lower frequency being larger in magnitude. Note that this plot fails to resolve the two lowest frequencies and that the apparent location of the highest frequency is shifted to the left. From this plot we might conclude that there are only two sinusoidal components in $\{y_t\}$.

Figure 531(b) shows $10 \log_{10} \hat{S}^{(\mathrm{P})}(\cdot)$ at the frequencies $k/32$ for $k = 0, 1, \ldots, 16$. There are still two major broad peaks visible (with maximum heights at points $k = 6$ and $k = 11$, respectively) and some minor bumps that are an order of magnitude smaller. Note now that the peak with the higher frequency is slightly larger than the one with the lower frequency. Moreover, there is a hint of some structure in the first broad bump (there is a slight dip between points $k = 4$ and 6). By going to a finer grid than that provided by the Fourier frequencies,

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

we are getting a better idea of the sinusoidal components in $\{y_t\}$.

In Figure 531(c) we have evaluated the periodogram on a still finer grid of 513 frequencies. The difference between adjacent frequencies is $1/1024 \approx 0.001$. This plot clearly shows two major broad peaks, the first of which (roughly between $f = 0.1$ and 0.2) has a small dip in the middle. The smaller peaks are evidently sidelobes due to the superposition of Fejér kernels. The appearance of nonzero terms in the periodogram at frequencies other than those of sinusoids in the harmonic process is due to leakage – see Sections 3.8 and 6.3 and the next section. Leakage can affect our ability to use the periodogram to determine what frequencies make up $\{y_t\}$. For example, the three highest peaks in the periodogram in Figure 531(c) occur at frequencies 0.117, 0.181 and 0.338 as compared to the true frequencies of 0.133, 0.189 and 0.333. These discrepancies are here entirely due to distortions caused by the six superimposed Fejér kernels (see Equation (524b)) since there is *no* observational noise present. Note, however, that the discrepancies are not large, particularly when we recall we are only dealing with 16 points from a function composed of three sinusoids.

We next study the effect of adding a realization of white noise to $y_0, y_1, \ldots, y_{15}$ to produce a realization $x_0, x_1, \ldots, x_{15}$ of Equation (512b). As in the previous section, we can use the decomposition

$$\text{var}\,\{X_t\} = \sum_{l=1}^{L} D_l^2/2 + \sigma_\epsilon^2 \;\text{ with }\; R \stackrel{\text{def}}{=} \frac{\sum_{l=1}^{L} D_l^2}{2\sigma_\epsilon^2},$$

where we consider R as a signal-to-noise ratio (see Equation (526d)). The plots in Figure 533 show the effect on the periodogram of adding white noise with variances of 0.01, 0.1 and 1.0 (top to bottom plots, respectively), yielding signal-to-noise ratios of 131, 13.1 and 1.31. Each of these periodograms has been evaluated on a grid of 513 frequencies for comparison with the noise-free case in the bottom plot of Figure 531. Note that the locations of the peaks shift around as the white noise variance varies. In Figures 533(a) and (b) the three largest peaks are attributable to the three sinusoids in $\{x_t\}$, whereas only the two largest are in (c). In both (b) and (c), there is a peak that could be mistaken for a sinusoidal component not present in Equation (530). For all three plots and the noise-free case, Table 533 lists peak frequencies that are reasonably close to frequencies in Equation (530).

In Figure 534 we study how the length of a series affects our ability to estimate the frequencies of the sinusoids that form it. Here we used Equation (530) to evaluate y_t for $t = 0, 1, \ldots, 127$. We added white noise with unit variance to each element of the extended series and calculated separate periodograms on a grid of 513 frequencies for the first 16, 32, 64 and 128 points in the contaminated series. These periodograms are shown in Figure 533(c) and the three plots of Figure 534. As the series length increases, it becomes easier to pick out the frequency components, and the center of the peaks in the periodogram get closer to the true frequencies of the sinusoids (again indicated by the thin vertical lines). Note also that the heights of the peaks increase directly in proportion to the number of points and that the widths of the peaks are comparable to the widths of the central lobes of the corresponding Fejér's kernels (these lobes are shown by the thick curves in the lower right-hand portion of each plot). Both of these indicate that we are dealing with a harmonic process as opposed to just a purely continuous stationary process with an SDF highly concentrated around certain frequencies.

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

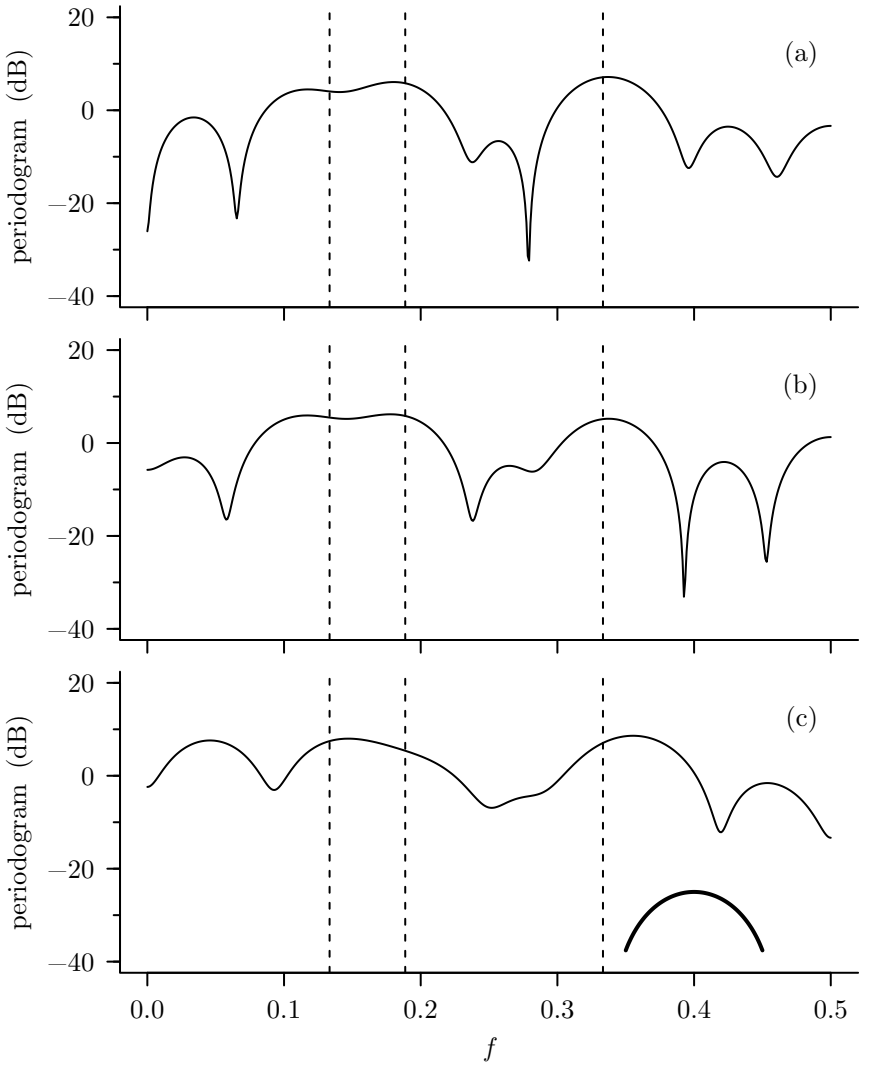

Figure 533 Effect of additive white noise on the periodogram. Here we add a realization of white noise to y_0, $y_1, \ldots, y_{15}$ of Equation (530) and compute the resulting periodogram $\hat{S}^{(\mathrm{P})}(\cdot)$ over a finely spaced grid of 513 frequencies. The variances of the additive white noise are 0.01, 0.1 and 1, respectively, in the top to bottom plots. The periodogram for the noise-free case is shown in the bottom plot of Figure 531. In plot (c), the thick curve centered at $f = 0.4$ depicts the central lobe of Fejér's kernel for sample size $N = 16$.

l	True f_l	531(c) $\sigma_\epsilon^2 = 0$	533(a) $\sigma_\epsilon^2 = 0.01$	533(b) $\sigma_\epsilon^2 = 0.1$	533(c) $\sigma_\epsilon^2 = 1$
1	0.133	0.117	0.117	0.117	0.147
2	0.189	0.181	0.181	0.178	—
3	0.333	0.338	0.337	0.338	0.355

Table 533 True and estimated f_l (the sample size N is fixed at 16).

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

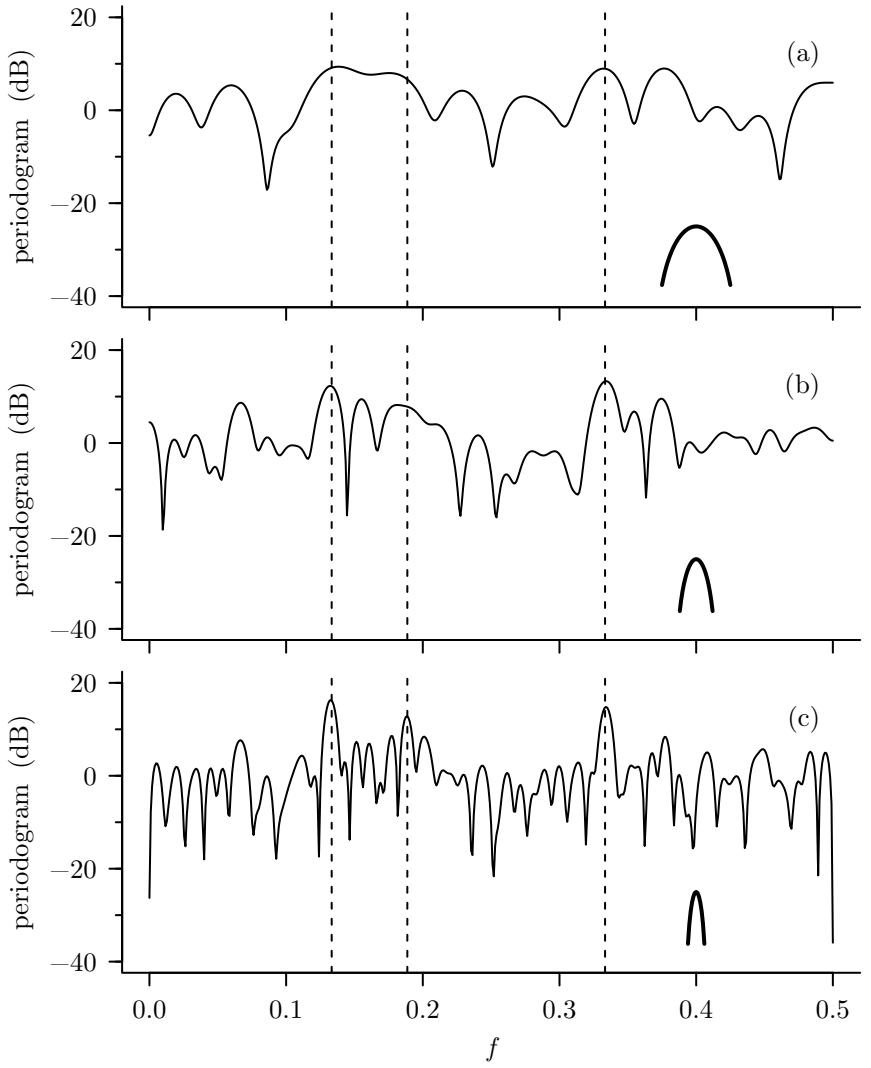

Figure 534 Effect of sample size on the periodogram. Here we add white noise with unit variance to $y_0, y_1, \ldots, y_{127}$ of Equation (530) and compute the resulting periodogram $\hat{S}^{(\mathrm{P})}(\cdot)$ using just the first 32 and 64 values of the series (a and b), and finally all 128 values (c). The periodogram for just the first 16 values is shown in the bottom plot of Figure 533. In each plot, the thick curve centered at $f = 0.4$ depicts the central lobe of Fejér's kernel for the appropriate sample size.

Comments and Extensions to Section 10.7

[1] Since the series in our example only had 16 points, is it reasonable to expect the agreement between estimated and actual frequencies to be as close as Table 533 indicates, or is there something peculiar about our example? Let us see what the approximation of Equation (526c) gives us for the variance of the estimated frequency for the sinusoid with the highest frequency using the three noise-contaminated series. Strictly speaking, this approximation assumes that the model involves only a single frequency rather than three; however, this approximation can be taken as valid if the frequencies are spaced "sufficiently" far apart (Priestley, 1981, p. 410). If we thus ignore the other two frequencies, the signal-to-noise ratio for the $f_3 = 1/3$ component alone is $R = 1/(2\sigma_\epsilon^2)$ since $D_3^2 = 1$. This implies that

$$\mathrm{var}\,\{\hat{f}_3\} \approx \frac{6\sigma_\epsilon^2}{N^3\pi^2} \doteq 0.00015 \times \sigma_\epsilon^2 \text{ for } N = 16.$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

For σ_ϵ^2 = 0.01, 0.1 and 1, the standard deviations of $\hat{f}_3$ are 0.001, 0.004 and 0.012. From the last row of Table 533, we see that $\left|\hat{f}_3 - f_3\right|$ is 0.004, 0.005 and 0.022 for these three cases. Thus the observed deviations are of the order of magnitude expected by statistical theory.

10.8 Tapering and the Identification of Frequencies

We have already discussed the idea of data tapers and examined their properties in Section 6.4. When we taper a sample of length N of the harmonic process of Equation (512b) with a data taper $\{h_t\}$, it follows from an argument similar to that used in the derivation of Equation (524b) that

$$E\{\hat{S}^{(\mathrm{D})}(f)\} = \sigma_\epsilon^2 \, \Delta_\mathrm{t} + \sum_{l=1}^{L} \frac{D_l^2}{4} \left[\mathcal{H}(f + f_l) + \mathcal{H}(f - f_l)\right], \tag{535a}$$

where $\mathcal{H}(\cdot)$ is the spectral window corresponding to the taper $\{h_t\}$ (see Equation (186e) and the discussion surrounding it). If the taper is the rectangular data taper, then $\hat{S}^{(\mathrm{D})}(\cdot) = \hat{S}^{(\mathrm{P})}(\cdot)$, $\mathcal{H}(\cdot) = \mathcal{F}(\cdot)$ (i.e., Fejér's kernel) and Equation (535a) reduces to Equation (524b).

It is easy to concoct an artificial example where use of a taper greatly enhances the identification of the frequencies in Equation (512b). Figure 536 shows three different direct spectral estimates $\hat{S}^{(\mathrm{D})}(\cdot)$ formed from one realization of length $N = 256$ from a special case of Equation (512b):

$$X_t = 0.0316 \cos\left(0.2943\pi t + \phi_1\right) + \cos\left(0.3333\pi t + \phi_2\right) + 0.0001 \cos\left(0.3971\pi t + \phi_3\right) + \epsilon_t, \tag{535b}$$

where $\{\epsilon_t\}$ is a white noise process with variance $\sigma_\epsilon^2 = 10^{-10} = -100$ dB. For the three plots in the figure, the corresponding data tapers are, from top to bottom, a rectangular data taper (cf. Figures 190(a) and 191(a) for $N = 64$), a Slepian data taper with $NW = 2$ (Figures 190(f) and 191(f) for $N = 64$) and a Slepian data taper with $NW = 4$ (Figures 190(g) and 191(g) for $N = 64$). The thick curve in the upper left-hand corner of each plot depicts the central lobe of the associated spectral window $\mathcal{H}(\cdot)$. For the rectangular taper (plot a), the sidelobes of Fejér's kernel that is associated with the middle term in Equation (535b) completely masks the sinusoid with the smallest amplitude (the third term), and there is only a slight hint of the sinusoid with the lowest frequency (first term). Thus the periodogram only clearly reveals the presence of a single harmonic component. Use of a Slepian taper with $NW = 2$ (plot b) yields a spectral window with sidelobes reduced enough to clearly reveal the presence of a second sinusoid, but the third sinusoid is still hidden (this Slepian data taper is roughly comparable to the popular Hanning data taper). Finally, all three sinusoidal components are revealed when a Slepian taper with $NW = 4$ is used (plot c), as is the level of the background continuum (the horizontal dashed lines indicates the true level σ_ϵ^2).

Note, however, that, in comparison to the $NW = 2$ Slepian data taper, the broader central lobe of the $NW = 4$ Slepian taper somewhat degrades the resolution of the two dominant sinusoidal components. An example can be constructed in which the $NW = 4$ Slepian taper fails to resolve two closely spaced frequencies, while a taper with a narrower central lobe (say, a Slepian taper with $NW = 1$ or 2 or a 100% cosine taper) succeeds in doing so (see Exercise [10.9b]). The selection of a suitable taper is clearly dependent upon the true model – a fact that accounts for the large number of tapers suggested as being appropriate for different problems (Harris, 1978). Rife and Vincent (1970) present an interesting study on the estimation of the magnitude and frequencies of two "tones" when using different families of tapers with differing central lobe and sidelobe properties. Tseng et al. (1981) propose a family of tapers that is specified by four parameters and is effective in the detection of three-tone signals.

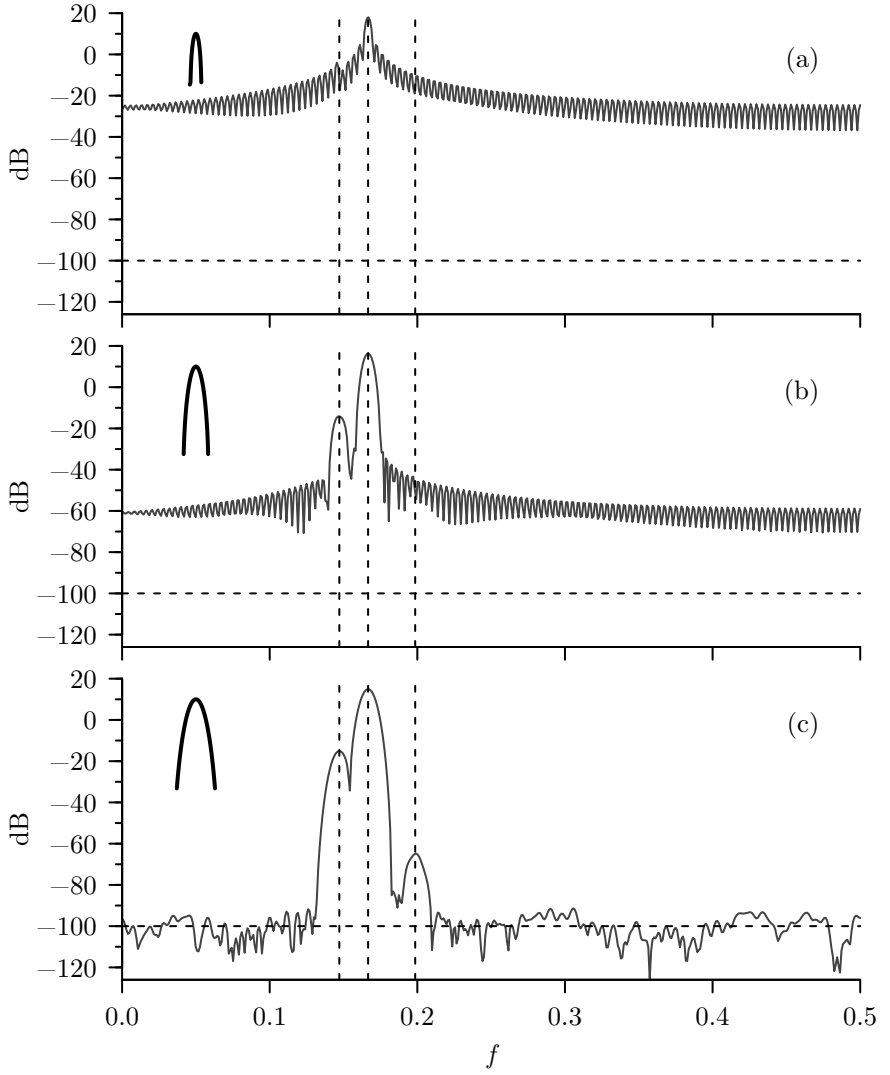

Figure 536 Effect of tapering on detection of harmonic components in process of Equation (535b). Plot (a) shows the periodogram, while the other two plots show direct spectral estimators based on (b) a Slepian taper with $NW = 2$ and (c) a Slepian taper with $NW = 4$. Each spectral estimate is plotted over a finely spaced grid of 513 frequencies. The vertical dashed lines mark the locations of the frequencies in Equation (535b), while the horizontal line shows the level σ_ϵ^2 of the background continuum. The central lobes of the associated spectral windows are shown in the upper left-hand corners.

Tapering is sometimes useful even when there is but a single sinusoidal term in a harmonic process. In this case, Equation (524b) tells us that

$$E\{\hat{S}^{(\mathrm{P})}(f)\} = \sigma_\epsilon^2 \, \Delta_\mathrm{t} + \frac{D_1^2}{4} \left[\mathcal{F}(f + f_1) + \mathcal{F}(f - f_1)\right]. \tag{536}$$

For f close to f_1, the second of the two Fejér kernels above is generally the dominant term, but the presence of the first kernel can distort $E\{\hat{S}^{(\mathrm{P})}(\cdot)\}$ such that its peak value is shifted away from f_1. The utility of tapering can also depend critically on what value the RV ϕ_1 assumes in the single-frequency model $X_t = D_1 \cos(2\pi f_1 t \, \Delta_\mathrm{t} + \phi_1) + \epsilon_t$. The burden of

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

Exercise [10.10a] is to show that the expected value of the periodogram conditional on ϕ_1 is

$$E\{\hat{S}^{(\mathrm{P})}(f) \mid \phi_1\} = \sigma_\epsilon^2 \, \Delta_\mathrm{t} + \frac{D_1^2}{4} \left[\mathcal{F}(f + f_1) + \mathcal{F}(f - f_1) + \mathcal{K}(f, f_1, \phi_1)\right], \qquad (537a)$$

where

$$\mathcal{K}(f, f_1, \phi_1) \stackrel{\mathrm{def}}{=} 2 \, \Delta_\mathrm{t} N \cos\left(2\pi[N-1] f_1 \, \Delta_\mathrm{t} + 2\phi_1\right) \mathcal{D}_N([f + f_1] \, \Delta_\mathrm{t}) \mathcal{D}_N([f - f_1] \, \Delta_\mathrm{t}),$$

and $\mathcal{D}_N(\cdot)$ is Dirichlet's kernel – see Equation (17c) for its definition and Equation (174c) for its connection to Fejér's kernel. For particular realizations of ϕ_1, the term involving Dirichlet's kernel can also push the peak value of the periodogram away from f_1. Distortions due to both kernels can potentially be lessened by tapering.

As a concrete example, consider the following single-frequency model with a fixed phase:

$$X_t = \cos\left(2\pi f_1 t + 5\pi/12\right) + \epsilon_t, \qquad (537b)$$

where $f_1 = 0.0725$, and the background continuum $\{\epsilon_t\}$ is white noise with variance $\sigma_\epsilon^2 = 10^{-4}$. This process has a signal-to-noise ratio of 37 dB, which is so high that plots of individual realizations are barely distinguishable from plots of pure sinusoids. We consider series of sample size $N = 64$, for which the cosine term in Equation (537b) traces out $Nf_1 = 4.64$ oscillations over $t = 0, 1, \ldots, N-1$. The dark curve in Figure 538a shows the corresponding unconditional expectation $E\{\hat{S}^{(\mathrm{P})}(\cdot)\}$ over an interval of frequencies centered about f_1. The peak of this curve – indicated by the dashed vertical line – is at 0.07251, which is quite close to f_1 (had the figure included a line depicting f_1, it would have been visually indistinguishable from the dashed line). The light curve shows the conditional expectation $E\{\hat{S}^{(\mathrm{P})}(\cdot) \mid \phi_1 = 5\pi/12\}$. The solid vertical line indicates that its peak frequency occurs at 0.07215, which is distinctly below f_1. For this particular setting for ϕ_1, the distortion away from f_1 is primarily due to $\mathcal{K}(f, f_1, \phi_1)$ in Equation (537a) rather than $\mathcal{F}(f + f_1)$.

To demonstrate the effectiveness of tapering, we generated 50 realizations of length $N = 64$ from the process described by Equation (537b). For each realization, we computed the periodogram $\hat{S}^{(\mathrm{P})}(\cdot)$ and a Hanning-based direct spectral estimate $\hat{S}^{(\mathrm{D})}(\cdot)$. We then found the locations of the frequencies at which $\hat{S}^{(\mathrm{P})}(\cdot)$ and $\hat{S}^{(\mathrm{D})}(\cdot)$ attained their peak values (see C&E [2] for Section 10.6). A scatter plot of the locations of the peak frequencies for $\hat{S}^{(\mathrm{D})}(\cdot)$ versus those for $\hat{S}^{(\mathrm{P})}(\cdot)$ is shown in Figure 538b for the 50 realizations. The dashed horizontal and vertical lines mark the location of the true frequency $f_1 = 0.0725$ of the sinusoidal component, while the solid vertical line marks the peak frequency of $E\{\hat{S}^{(\mathrm{P})}(\cdot) \mid \phi_1 = 5\pi/12\}$. We see that, whereas the peak frequencies derived from $\hat{S}^{(\mathrm{D})}(\cdot)$ are clustered nicely about f_1, those from $\hat{S}^{(\mathrm{P})}(\cdot)$ are biased low in frequency and are clustered around peak frequency for the conditional expectation of the periodogram. The mean square error in estimating f_1 via the periodogram is dominated by this bias and is considerably greater than that associated with the relatively bias-free $\hat{S}^{(\mathrm{D})}(\cdot)$. (We will return to this example later when we discuss the parametric approach to harmonic analysis – see the discussion concerning Figure 557 in Section 10.12. Exercise [10.10] expands upon this example to demonstrate that tapering can lead to an estimator of peak frequency that is better than the periodogram-based estimator in terms of not only bias, but also variance.)

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

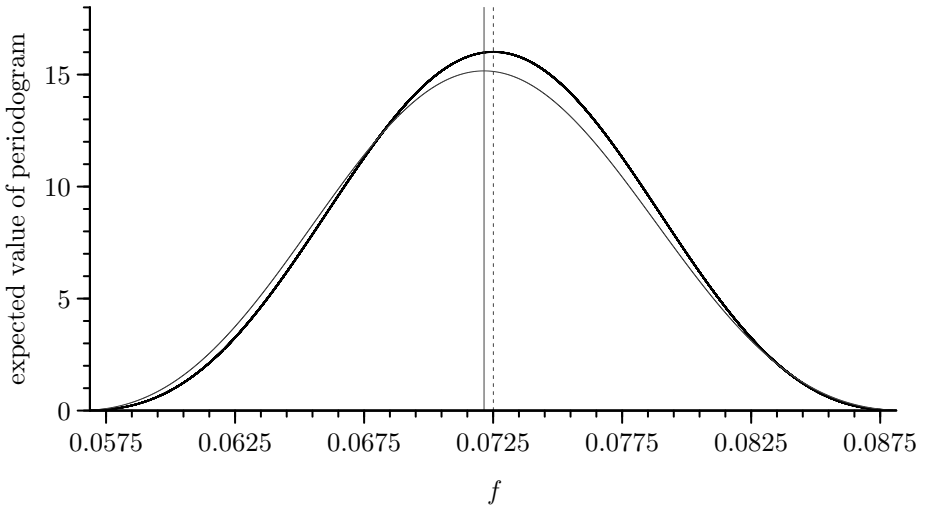

Figure 538a Unconditional expected value of periodogram (thick curve) for a harmonic regression model with a single frequency at $f_1 = 0.0725$, along with the expected value conditional on the random phase ϕ_1 being set to $5\pi/12$ (thin curve). The vertical dashed line marks the location of the peak for the unconditional expectation (this location is quite close to f_1). The solid line does the same for the conditional expectation.

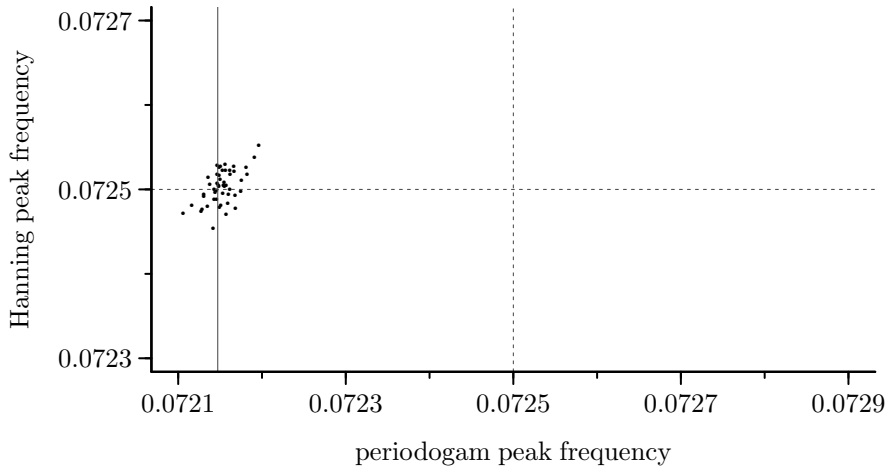

Figure 538b Scatter plot of peak frequencies computed from a Hanning-based direct spectral estimator versus those from the periodogram. The 50 plotted points are computed from 50 realizations of $\{X_t\}$ of Equation (537b), which has a single sinusoid with frequency $f_1 = 0.0725$ (dashed horizontal and vertical lines). There is a substantial downward bias in the periodogram-based peak frequencies, which are clustered around the peak frequency (solid vertical line) of the conditional expectation $E\{\hat{S}^{(\mathrm{P})}(\cdot) \mid \phi_1 = 5\pi/12\}$ shown in Figure 538a.

10.9 Tests for Periodicity – White Noise Case

As should be clear from the previous sections, we need statistical tests to help determine when a spectral peak is significantly larger than is likely to arise if there were no genuine periodic components. We consider two basic cases, where the decision is

[1] between white noise and spectral lines plus white noise (i.e., white noise versus a discrete spectrum) and

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

[2] between colored noise and spectral lines plus colored noise (i.e., colored noise versus a mixed spectrum).

Typical tests for the first case make use of the fact that, under the null hypothesis that $\{X_t\}$ is Gaussian white noise, the periodogram ordinates calculated at the Fourier frequencies (excluding the zero and Nyquist frequencies) are distributed as independent and identically distributed χ^2_2 RVs times a multiplicative constant (see Section 6.6). The tests then look for periodogram ordinates of a size incompatible with the null hypothesis. Implicitly it is assumed that any spectral lines occur at Fourier frequencies, a rather strong assumption we reconsider at the end of this section. For the remainder of this section we look at case [1] by using Equation (512b) as our model for the observed time series, where the process $\{\epsilon_t\}$ is assumed to be Gaussian white noise (we delve into case [2] in Section 10.10).

As usual, let $f_k = k/(N\,\Delta_{\mathrm{t}})$ denote the kth Fourier frequency, and let M be the number of Fourier frequencies satisfying $0 < f_k < f_{\mathcal{N}}$. Thus

$$M = \left\lfloor \frac{N-1}{2} \right\rfloor = \begin{cases} (N-2)/2, & N \text{ even;} \\ (N-1)/2, & N \text{ odd.} \end{cases} \tag{539a}$$

The null hypothesis is $D_1 = \cdots = D_L = 0$. We consider the periodogram terms $\hat{S}^{(\mathrm{P})}(f_k)$ for $k = 1, 2, \ldots, M$. Under the null hypothesis,

$$\frac{2\hat{S}^{(\mathrm{P})}(f_k)}{\sigma_\epsilon^2\,\Delta_{\mathrm{t}}} \stackrel{\mathrm{d}}{=} \chi^2_2,$$

where $\stackrel{\mathrm{d}}{=}$ means "equal in distribution" (see Equation (203a)). The probability density function for a chi-square RV with two degrees of freedom is $f(u) = \exp(-u/2)/2$, $u \geq 0$ (see Equation (204d) and Figure 210a(a)). For any $u_0 \geq 0$,

$$\mathbf{P}\left[\frac{2\hat{S}^{(\mathrm{P})}(f_k)}{\sigma_\epsilon^2\,\Delta_{\mathrm{t}}} \leq u_0\right] = \mathbf{P}\left[\chi^2_2 \leq u_0\right] = \int_0^{u_0} f(u)\,\mathrm{d}u = 1 - \mathrm{e}^{-u_0/2}.$$

Let

$$\gamma \stackrel{\mathrm{def}}{=} \max_{1 \leq k \leq M} \frac{2\hat{S}^{(\mathrm{P})}(f_k)}{\sigma_\epsilon^2\,\Delta_{\mathrm{t}}}.$$

Under the null hypothesis, γ is the maximum of the M independent and identically distributed χ^2_2 RVs; thus, for any u_0,

$$\mathbf{P}\left[\gamma > u_0\right] = 1 - \mathbf{P}\left[\frac{2\hat{S}^{(\mathrm{P})}(f_k)}{\sigma_\epsilon^2\,\Delta_{\mathrm{t}}} \leq u_0 \text{ for } 1 \leq k \leq M\right] = 1 - \left(1 - \mathrm{e}^{-u_0/2}\right)^M.$$

This equation gives us the distribution of γ under the null hypothesis. Under the alternative hypothesis that $D_l > 0$ for one or more terms in Equation (512b), γ will be large, so we use a one-sided test with a critical region of the form $\{\gamma : \gamma > u_0\}$, where u_0 is chosen so that $\mathbf{P}[\gamma > u_0] = \alpha$, the chosen level of the test.

The chief disadvantage of this test (due to Schuster, 1898) is that we must know σ_ϵ^2 in advance to be able to calculate γ. Since we usually must estimate σ_ϵ^2, Fisher (1929) derived an exact test for the null hypothesis based upon the statistic

$$g \stackrel{\mathrm{def}}{=} \max_{1 \leq k \leq M} \frac{\hat{S}^{(\mathrm{P})}(f_k)}{\sum_{j=1}^{M} \hat{S}^{(\mathrm{P})}(f_j)} \tag{539b}$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

(note that $\frac{1}{M} \leq g \leq 1$). The burden of Exercise [10.11] is to show that the denominator above is such that

$$\sum_{j=1}^{M} \hat{S}^{(\mathrm{P})}(f_j) = \frac{\Delta_t}{2} \left(\sum_{t=0}^{N-1} (X_t - \overline{X})^2 - \frac{\delta_N}{N} \left[\sum_{t=0}^{N-1} X_t (-1)^t \right]^2 \right), \tag{540a}$$

where $\delta_N = 1$ if N is even and is 0 otherwise. Note that, when N is odd so that the last summation vanishes, the right-hand side is proportional to the standard estimator of σ_ϵ^2 under the null hypothesis (cf. Equation (166b) with τ set to 0); when N is even, it is approximately so, with the approximation becoming better as N gets larger. We can regard g as the maximum of the sum of squares due to a single frequency f_k over the total sum of squares. Note also that σ_ϵ^2 acts as a proportionality constant in the distribution of both $\max \hat{S}^{(\mathrm{P})}(f_k)$ and $\sum \hat{S}^{(\mathrm{P})}(f_j)$, so that σ_ϵ^2 "ratios out" of Fisher's g statistic. (In practice, we can calculate $\sum_{j=1}^{M} \hat{S}^{(\mathrm{P})}(f_j)$ directly in the frequency domain – the time domain equivalent provided by Equation (540a) is of interest more for theory than for computations.)

Fisher (1929) found that the exact distribution of g under the null hypothesis is given by

$$\mathbf{P}\left[g > g_F\right] = \sum_{j=1}^{M'} (-1)^{j-1} \binom{M}{j} (1 - j g_F)^{M-1}, \tag{540b}$$

where M' is the largest integer satisfying both $M' < 1/g_F$ and $M' \leq M$. As a simplification, we can use just the $j = 1$ term in this summation:

$$\mathbf{P}\left[g > g_F\right] \approx (M-1)(1 - g_F)^{M-1}. \tag{540c}$$

Setting the right-hand side equal to α and solving for g_F gives us an approximation $\tilde{g}_F$ for the critical level for an α level test:

$$\tilde{g}_F = 1 - (\alpha/M)^{1/(M-1)}. \tag{540d}$$

For $\alpha = 0.05$ and $M = 5, 6, \ldots, 50$, the error in this approximation is within 0.1%; moreover, the error is in the direction of decreasing the probability; i.e., $\mathbf{P}\left[g > \tilde{g}_F\right] < 0.05$. The error in this approximation is also of about the same order for $\alpha = 0.01$, 0.02 and 0.1 (Nowroozi, 1967). Quinn and Hannan (2001, p. 15) suggest another approximation for the distribution for Fisher's g test:

$$\mathbf{P}\left[g > g_F\right] \approx 1 - \exp\left(-N \exp\left(-N g_F/2\right)/2\right)$$

(in contrast to Equation (540c), the above is correct asymptotically as $N \to \infty$). This leads to an approximation for g_F given by

$$\hat{g}_F = -2 \log\left(-2 \log\left(1 - \alpha\right)/N\right)/N. \tag{540e}$$

The results of Exercise [10.12] support using $\tilde{g}_F$ when N is small, but then switching to $\hat{g}_F$ when N is large (as a rule of thumb, $N = 3000$ is the boundary between small and large).

Anderson (1971) notes that Fisher's test is the uniformly most powerful symmetric invariant decision procedure against *simple* periodicities, i.e., where the alternative hypothesis is that there exists a periodicity at only one Fourier frequency. But how well does it perform when there is a *compound* periodicity, i.e., spectral lines at several frequencies? Siegel (1980)

studied this problem (for another approach, see Bøviken, 1983). He suggested a test statistic based on all large values of the rescaled periodogram

$$\tilde{S}^{(\mathrm{P})}(f_k) \stackrel{\mathrm{def}}{=} \frac{\hat{S}^{(\mathrm{P})}(f_k)}{\sum_{j=1}^{M} \hat{S}^{(\mathrm{P})}(f_j)} \tag{541a}$$

instead of only their maximum. For a value $0 < g_0 \leq g_F$ (where, as before, g_F is the critical value for Fisher's test) and for each $\tilde{S}^{(\mathrm{P})}(f_k)$ that exceeds g_0, Siegel sums the excess of $\tilde{S}^{(\mathrm{P})}(f_k)$ above g_0. This forms the statistic

$$\sum_{k=1}^{M} \left(\tilde{S}^{(\mathrm{P})}(f_k) - g_0 \right)_+ ,$$

where $(a)_+ = \max\{a, 0\}$ is the positive-part function. This statistic depends on the choice of g_0 and can be related to g_F by writing $g_0 = \lambda g_F$ with $0 < \lambda \leq 1$ to give

$$T_\lambda \stackrel{\mathrm{def}}{=} \sum_{k=1}^{M} \left(\tilde{S}^{(\mathrm{P})}(f_k) - \lambda g_F \right)_+ . \tag{541b}$$

Note that, when $\lambda = 1$, the event $g > g_F$ is identical to the event $T_1 > 0$; i.e., we have Fisher's test. Suppose that we choose g_F to correspond to the α significance level of Fisher's test so that

$$\mathbf{P}\left[T_1 > 0\right] = \mathbf{P}\left[\sum_{k=1}^{M} \left(\tilde{S}^{(\mathrm{P})}(f_k) - g_F \right)_+ > 0 \right] = \alpha.$$

Stevens (1939) showed that

$$\mathbf{P}\left[T_1 > 0\right] = \sum_{j=1}^{M'} (-1)^{j-1} \binom{M}{j} (1 - j g_F)^{M-1},$$

exactly as is required by Equation (540b).

We need to compute $\mathbf{P}\left[T_\lambda > t\right]$ for general λ in the range 0 to 1 so that we can find the critical points for significance tests. Siegel (1979) gives the formula

$$\mathbf{P}\left[T_\lambda > t\right] = \sum_{j=1}^{M} \sum_{k=0}^{j-1} (-1)^{j+k+1} \binom{M}{j} \binom{j-1}{k} \binom{M-1}{k} t^k (1 - j\lambda g_F - t)_+^{M-k-1}. \tag{541c}$$

Siegel (1980) gives critical values t_λ for T_λ for significance levels 0.01 and 0.05 with M ranging from 5 to 50 (corresponding to values of N ranging from 11 to 102) and with $\lambda = 0.4$, 0.6 and 0.8. For a value of $\lambda = 0.6$, Siegel found that his test statistic $T_{0.6}$ proved only slightly less powerful than Fisher's g (equivalently, T_1) against an alternative of simple periodicity, but that it outperformed Fisher's test against an alternative of *compound* periodicity (Siegel looked at periodic activity at both two and three frequencies).

Table 542 gives exact critical values for $T_{0.6}$ for $\alpha = 0.05$ and 0.01 and values of M corresponding to sample sizes N of integer powers of two ranging from 32 to 4096 (thus extending Siegel's table). For computing purposes, it is useful to rewrite Equation (541c) as

$$\mathbf{P}\left[T_\lambda > t\right] = t^{M-1} + \sum_{j=1}^{M} \sum_{k=0}^{\min\{j-1, M-2\}} (-1)^{j+k+1} \mathrm{e}^{Q_{j,k}}, \tag{541d}$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

		$\alpha = 0.05$			$\alpha = 0.01$		
N	M	Exact	$\tilde{t}_{0.6}$	$c\chi_0^2(\beta)$	Exact	$\tilde{t}_{0.6}$	$c\chi_0^2(\beta)$
16	7	0.225	0.253	0.220	0.268	0.318	0.273
32	15	0.140	0.146	0.138	0.164	0.173	0.165
64	31	0.0861	0.0861	0.0852	0.0969	0.0971	0.0974
128	63	0.0523	0.0515	0.0520	0.0562	0.0552	0.0564
256	127	0.0315	0.0310	0.0315	0.0322	0.0316	0.0323
512	255	0.0190	0.0187	0.0190	0.0184	0.0181	0.0184
1024	511	0.0114	0.0113	0.0114	0.0104	0.0104	0.0105
2048	1023	0.00688	0.00686	0.00688	0.00595	0.00598	0.00597
4096	2047	0.00416	0.00415	0.00416	0.00341	0.00344	0.00341

Table 542 Exact critical values $t_{0.6}$ for $T_{0.6}$ and two approximations thereof.

where

$$Q_{j,k} \stackrel{\text{def}}{=} \log \binom{M}{j} + \log \binom{j-1}{k} + \log \binom{M-1}{k} + k \log(t) + (M-k-1) \log(1 - j\lambda g_F - t)_+$$

(Walden, 1992). If we recall that, e.g., $\log(a!) = \log(\Gamma(a+1))$, we have

$$\log \binom{a}{b} = \log(\Gamma(a+1)) - \log(\Gamma(a-b+1)) - \log(\Gamma(b+1)),$$

and we can use standard approximations to the log of a gamma function with a large argument to evaluate $Q_{j,k}$ (see, for example, Press et al., 2007, p. 257). Use of Equation (541d) allows the construction of interpolation formulae that give accurate approximate critical values as a function of M. For $\alpha = 0.05$ the interpolation formula yields the approximation

$$t_{0.6} = 1.033 M^{-0.72356}, \tag{542a}$$

and for $\alpha = 0.01$ it is

$$t_{0.6} = 1.4987 M^{-0.79695}. \tag{542b}$$

Table 542 shows these approximations next to the exact values. The approximations are within 2% of the exact values for $31 \le M \le 2047$ (a range for which the formulae were designed to cover – they should be used with caution if M deviates significantly from this range).

Siegel (1979) gives an alternative way of finding the critical values for M large. He notes that

$$T_\lambda \stackrel{\text{d}}{=} c\chi_0^2(\beta) \tag{542c}$$

asymptotically, where the RV $\chi_0^2(\beta)$ obeys what Siegel calls the *noncentral chi-square distribution* with zero degrees of freedom and noncentrality parameter β. The first and second moments of the distribution in Equation (542c) are

$$E\{T_\lambda\} = (1 - \lambda g_F)^M \text{ and } E\{T_\lambda^2\} = \frac{2(1 - \lambda g_F)^{M+1} + (M-1)(1 - 2\lambda g_F)_+^{M+1}}{M+1},$$

and the parameters c and β in Equation (542c) are related by

$$c = \frac{\text{var}\{T_\lambda\}}{4E\{T_\lambda\}} = \frac{E\{T_\lambda^2\} - (E\{T_\lambda\})^2}{4E\{T_\lambda\}} \text{ and } \beta = \frac{E\{T_\lambda\}}{c}.$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

Hence it is merely necessary to find the critical values of $c\chi^2_0(\beta)$ with c and β given by the above. This can be done using the algorithm of Farebrother (1987) or, e.g., the function `qchisq` in the `R` language. Table 542 gives the approximate critical values using this method, all of which are within 2% of the exact values.

The tests we have looked at so far assume that any spectral lines occur at Fourier frequencies. Fisher's and Siegel's tests look particularly useful in theory for the case of large samples where the Fourier frequencies are closely spaced and thus are likely to be very close to any true lines (for large enough N, it will become obvious from the periodogram that there are significant peaks in the spectrum so that in practice the tests lose some of their value). What happens, however, if a spectral line occurs midway between two Fourier frequencies? Equation (524b) can be employed to show that the expected value of the periodogram at such an intermediate frequency can be substantially less than that obtained at a Fourier frequency (cf. Equation (524a); see also Whittle, 1952). Even though there must be some degradation in performance, Priestley (1981, p. 410) argues that Fisher's test will still be reasonably powerful, provided the signal-to-noise ratio $D_l^2/(2\sigma_\epsilon^2)$ is large (he also gives a summary of some other tests for periodicity, none of which appear particularly satisfactory).

Section 10.15 gives an example of applying Fisher's and Siegel's tests to the ocean noise time series of Figure 4(d).

Comments and Extensions to Section 10.9

[1] Fisher's and Siegel's white noise tests both make use of the periodogram, as does the cumulative periodogram test for white noise described in C&E [6] for Section 6.6. Whereas all three tests should have similar performance under the null hypothesis of white noise, they are designed to reject the null hypothesis when faced with different alternative hypotheses. In the case of Fisher's and Siegel's tests, the alternatives are time series with discrete spectra, while, for the cumulative periodogram test, the alternatives are often thought of as series with purely continuous spectra. This does not mean that the tests do not have some power against alternatives for which they were not specifically designed. Exercise [10.13] asks the reader to apply all three tests to various simulated time series, some of which the tests are designed for, and others of which they are not.

[2] For an irregularly sampled Gaussian time series, the Lomb–Scargle periodogram of Equation (529a) can be used to test the null hypothesis (NH) of white noise against the alternative hypothesis (AH) of a sinusoid buried in white noise. When the frequency f_1 is known a priori, the connection between this periodogram and least squares estimation of A and B in the model of Equation (528b) suggests using the standard F-test for the regression, which evaluates the joint significance of $\hat{A}$ and $\hat{B}$. This test takes the form

$$F = \frac{(\mathrm{SS}_{\mathrm{NH}} - \mathrm{SS}_{\mathrm{AH}})/(\mathrm{df}_{\mathrm{NH}} - \mathrm{df}_{\mathrm{AH}})}{\mathrm{SS}_{\mathrm{AH}}/\mathrm{df}_{\mathrm{AH}}},$$

where, recalling Equation (528c),

$$\mathrm{SS}_{\mathrm{AH}} = \mathrm{SS}_c(\hat{A}, \hat{B}) \text{ and } \mathrm{SS}_{\mathrm{NH}} = \sum_{n=0}^{N-1} X^2(t_n)$$

are the residual sum of squares under the alternative and null hypotheses, while $\mathrm{df}_{\mathrm{AH}} = N - 2$ and $\mathrm{df}_{\mathrm{NH}} = N$ are the associated degrees of freedom (Weisberg, 2014). Hence

$$F = \frac{\hat{S}^{(\mathrm{LS})}(f_1)}{\hat{\sigma}_\epsilon^2}, \text{ where } \hat{\sigma}_\epsilon^2 = \frac{\mathrm{SS}_c(\hat{A}, \hat{B})}{N-2} \tag{543}$$

is the usual least squares estimator of σ_ϵ^2 (Exercise [10.8a] shows that $\mathrm{SS}_c(\hat{A}, \hat{B})$ is the same for all c – hence setting c to $\tilde{c}$ of Equation (529b) to force $\hat{A}$ and $\hat{B}$ to be uncorrelated does not modify F). This F statistic is F-distributed with 2 and $N - 2$ degrees of freedom. Large values of F provide evidence

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

against the null hypothesis. Note that the upper $(1-\alpha)\times 100\%$ percentage point of the $F_{2,\nu}$ distribution can be surprisingly easily computed using the formula

$$\frac{\nu(1-\alpha^{2/\nu})}{2\alpha^{2/\nu}}. \tag{544a}$$

Complications arise in adapting the Lomb–Scargle approach when the frequency f_1 is unknown. One adaptation is to define a set of, say, M candidate frequencies for the sinusoids in the model of Equation (528b), compute the F-statistic of Equation (543) over the M frequencies and consider the maximum value of these to evaluate the hypothesis of white noise versus the alternative of a sinusoid with an unknown frequency buried in white noise. The distribution of this maximum under the null hypothesis is difficult to come by analytically. Zechmeister and Kürster (2009) discuss various approximations, noting that Monte Carlo simulations are a viable – but computationally intensive – choice for getting at the distribution.

10.10 Tests for Periodicity – Colored Noise Case

We now turn our attention to case [2] described at the beginning of Section 10.9 – tests for spectral lines in *colored* noise. As in the white noise case, periodogram ordinates can form the basis for a test in the mixed spectra case (see, e.g., Wen et al., 2012, and references therein), but other approaches have been explored (see, e.g., Priestley, 1981, sections 8.3–4, for a technique based upon the autocovariance function). For case [2] we turn our attention in this section to an appealing multitaper technique proposed in Thomson (1982).

We start by assuming the following version of Equation (518c) involving just a single frequency f_1:

$$X_t = D_1 \cos\left(2\pi f_1 t\,\Delta_{\mathrm{t}} + \phi_1\right) + \eta_t, \tag{544b}$$

where $\{\eta_t\}$ (the background continuum) is a zero mean stationary process with SDF $S_\eta(\cdot)$ that is not necessarily constant; i.e., $\{\eta_t\}$ can be colored noise. We also assume in this section that $\{\eta_t\}$ has a Gaussian distribution. We would usually consider ϕ_1 in Equation (544b) to be an RV, but Thomson's approach regards it as an unknown constant. This assumption turns $\{X_t\}$ into a process with a time varying mean value:

$$E\{X_t\} = D_1 \cos\left(2\pi f_1 t\,\Delta_{\mathrm{t}} + \phi_1\right) = C_1 \mathrm{e}^{\mathrm{i}2\pi f_1 t\,\Delta_{\mathrm{t}}} + C_1^* \mathrm{e}^{-\mathrm{i}2\pi f_1 t\,\Delta_{\mathrm{t}}}, \tag{544c}$$

where $C_1 \stackrel{\mathrm{def}}{=} D_1 \mathrm{e}^{\mathrm{i}\phi_1}/2$. Note that we can write

$$X_t = E\{X_t\} + \eta_t. \tag{544d}$$

The multitaper test for periodicity makes use of K orthogonal data tapers $\{h_{0,t}\}$, …, $\{h_{K-1,t}\}$, which, when used to form the basic multitaper estimator of Equation (352a), yield a spectral estimator with a standard bandwidth measure $B_{\overline{\mathcal{H}}}$ dictated by Equation (353e) (Thomson, 1982, focuses on the Slepian multitapers, but we also consider the sinusoidal tapers in what follows). Similar to Equations (186a) and (436a), let us consider

$$J_k(f) \stackrel{\mathrm{def}}{=} \Delta_{\mathrm{t}}^{1/2} \sum_{t=0}^{N-1} h_{k,t} X_t \mathrm{e}^{-\mathrm{i}2\pi f t\,\Delta_{\mathrm{t}}},$$

where the kth-order data taper $\{h_{k,t}\}$ is associated with spectral window

$$\mathcal{H}_k(f) \stackrel{\mathrm{def}}{=} \frac{1}{\Delta_{\mathrm{t}}} \left|H_k(f)\right|^2, \text{ for which } H_k(f) \stackrel{\mathrm{def}}{=} \Delta_{\mathrm{t}} \sum_{t=0}^{N-1} h_{k,t} \mathrm{e}^{-\mathrm{i}2\pi f t\,\Delta_{\mathrm{t}}}.$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

Equation (352a) makes it clear that $|J_k(f)|^2$ is equal to the kth eigenspectrum $\hat{S}_k^{(\mathrm{MT})}(f)$ in the multitaper scheme. Using Equation (544d), the expected value of the complex-valued quantity $J_k(f)$ is

$$\begin{aligned} E\{J_k(f)\} &= \Delta_{\mathrm{t}}^{1/2} \sum_{t=0}^{N-1} h_{k,t} \left(C_1 \mathrm{e}^{\mathrm{i}2\pi f_1 t\,\Delta_{\mathrm{t}}} + C_1^* \mathrm{e}^{-\mathrm{i}2\pi f_1 t\,\Delta_{\mathrm{t}}} \right) \mathrm{e}^{-\mathrm{i}2\pi f t\,\Delta_{\mathrm{t}}} \\ &= \frac{1}{\Delta_{\mathrm{t}}^{1/2}} \left[C_1 H_k(f - f_1) + C_1^* H_k(f + f_1) \right]. \end{aligned}$$

The frequency domain analog of Equation (544d) is thus

$$J_k(f) = E\{J_k(f)\} + \Delta_{\mathrm{t}}^{1/2} \sum_{t=0}^{N-1} h_{k,t} \eta_t \mathrm{e}^{-\mathrm{i}2\pi f t\,\Delta_{\mathrm{t}}}.$$

Note that the squared modulus of the term with η_t is a direct spectral estimator of $S_\eta(f)$ using the data taper $\{h_{k,t}\}$. At $f = f_1$, the above becomes

$$J_k(f_1) = E\{J_k(f_1)\} + \Delta_{\mathrm{t}}^{1/2} \sum_{t=0}^{N-1} h_{k,t} \eta_t \mathrm{e}^{-\mathrm{i}2\pi f_1 t\,\Delta_{\mathrm{t}}},$$

where

$$E\{J_k(f_1)\} = \frac{1}{\Delta_{\mathrm{t}}^{1/2}} \left[C_1 H_k(0) + C_1^* H_k(2f_1) \right]. \tag{545a}$$

For a Slepian or sinusoidal taper that contributes to a basic multitaper estimator with half-bandwidth measure $B_{\overline{\mathcal{H}}}/2$, the squared modulus of $H_k(\cdot)$ is highly concentrated within the interval $[-B_{\overline{\mathcal{H}}}/2, B_{\overline{\mathcal{H}}}/2]$. Thus we have, for $2f_1 > B_{\overline{\mathcal{H}}}/2$,

$$E\{J_k(f_1)\} \approx C_1 \frac{H_k(0)}{\Delta_{\mathrm{t}}^{1/2}} = \Delta_{\mathrm{t}}^{1/2} C_1 \sum_{t=0}^{N-1} h_{k,t}. \tag{545b}$$

The kth-order Slepian and sinusoidal data tapers are skew-symmetric about their midpoints for odd k and symmetric for even k (see the left-hand plots of Figures 360, 362, 394 and 396), For both multitaper schemes, the above summation is in fact zero for odd k, but it is real-valued and positive for even k.

We now assume that $2f_1 > B_{\overline{\mathcal{H}}}/2$ and take the approximation in Equation (545b) to be an equality in what follows. We can write the following model for $J_k(f_1)$:

$$J_k(f_1) = C_1 \frac{H_k(0)}{\Delta_{\mathrm{t}}^{1/2}} + \tilde{\epsilon}_k, \quad k = 0, \ldots, K-1, \tag{545c}$$

where

$$\tilde{\epsilon}_k \stackrel{\mathrm{def}}{=} \Delta_{\mathrm{t}}^{1/2} \sum_{t=0}^{N-1} h_{k,t} \eta_t \mathrm{e}^{-\mathrm{i}2\pi f_1 t\,\Delta_{\mathrm{t}}}.$$

Equation (545c) thus defines a complex-valued regression model: $J_k(f_1)$ is the observed response (dependent variable); C_1 is an unknown parameter; $H_k(0)/\Delta_{\mathrm{t}}^{1/2}$ is the kth predictor (independent variable); and $\tilde{\epsilon}_k$ is the error term. We have already assumed $\{\eta_t\}$ to be a zero mean real-valued Gaussian stationary process. We now add the assumptions that its SDF

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

$S_\eta(\cdot)$ is slowly varying in the interval $[f_1 - B_{\overline{\mathcal{H}}}/2, f_1 + B_{\overline{\mathcal{H}}}/2]$ and that f_1 is not too close to either zero or the Nyquist frequency. Since $\tilde{\epsilon}_k$ is just the kth eigencoefficient in a multitaper estimator of $S_\eta(f_1)$, we can appeal to the arguments used in the solution to Exercise [8.3] to make two reasonable assertions, one about $\text{cov}\,\{\tilde{\epsilon}_k, \tilde{\epsilon}_l\}$, and the other concerning $\text{cov}\,\{\tilde{\epsilon}_k, \tilde{\epsilon}_l^*\}$. First,

$$\text{cov}\,\{\tilde{\epsilon}_k, \tilde{\epsilon}_l\} = E\{\tilde{\epsilon}_k \tilde{\epsilon}_l^*\} = 0, \qquad k \neq l,$$

so that the $\tilde{\epsilon}_k$ are pairwise uncorrelated, and, when $k = l$,

$$\text{var}\,\{\tilde{\epsilon}_k\} = E\{|\tilde{\epsilon}_k|^2\} \stackrel{\text{def}}{=} \sigma_{\tilde{\epsilon}}^2,$$

so that each $\tilde{\epsilon}_k$ has the same variance; in addition, we have $\sigma_{\tilde{\epsilon}}^2 = S_\eta(f_1)$. Second,

$$\text{cov}\,\{\tilde{\epsilon}_k, \tilde{\epsilon}_l^*\} = E\{\tilde{\epsilon}_k \tilde{\epsilon}_l\} = 0, \qquad 0 \leq k, l \leq K-1.$$

These two assertions imply that the real and imaginary components of $\tilde{\epsilon}_k$, $k = 0, 1, \ldots, K-1$, are $2K$ real-valued and independent Gaussian RVs, each with the same semivariance $\sigma_{\tilde{\epsilon}}^2/2$. With all the above stipulations, $\tilde{\epsilon}_k$ matches the requirements of complex-valued regression theory as formulated in Miller (1973, p. 715).

We can now estimate C_1 by employing a version of least squares valid for complex-valued quantities. The estimator is the value, say $\hat{C}_1$, of C_1 that minimizes

$$\text{SS}(C_1) \stackrel{\text{def}}{=} \sum_{k=0}^{K-1} \left| J_k(f_1) - C_1 \frac{H_k(0)}{\Delta_t^{1/2}} \right|^2 .$$

It follows from Miller (1973, equation (3.2)) that

$$\hat{C}_1 = \Delta_t^{1/2} \frac{\sum_{k=0}^{K-1} J_k(f_1) H_k(0)}{\sum_{k=0}^{K-1} H_k^2(0)} = \Delta_t^{1/2} \frac{\sum_{k=0,2,\ldots}^{2\lfloor (K-1)/2 \rfloor} J_k(f_1) H_k(0)}{\sum_{k=0,2,\ldots}^{2\lfloor (K-1)/2 \rfloor} H_k^2(0)} \tag{546a}$$

(we can eliminate $k = 1, 3, \ldots$ from the summations because $H_k(0) = 0$ for odd k). Moreover, under the conditions on $\tilde{\epsilon}_k$ stated in the previous paragraph, we can state the following results, all based on theorem 8.1 of Miller (1973). First, $\hat{C}_1$ is a proper complex Gaussian RV (it is uncorrelated with its complex conjugate), with mean C_1 and variance $\sigma_{\tilde{\epsilon}}^2 \Delta_t / \sum_{k=0,2,\ldots}^{2\lfloor (K-1)/2 \rfloor} H_k^2(0)$. Second, an estimator of $\sigma_{\tilde{\epsilon}}^2$ is given by

$$\hat{\sigma}_{\tilde{\epsilon}}^2 = \frac{1}{K} \sum_{k=0}^{K-1} \left| J_k(f_1) - \hat{J}_k(f_1) \right|^2 , \tag{546b}$$

where $\hat{J}_k(f_1)$ is the fitted value for $J_k(f_1)$ given by

$$\hat{J}_k(f_1) = \hat{C}_1 \frac{H_k(0)}{\Delta_t^{1/2}}$$

(this is zero when k is odd). Third, the RV $2K\hat{\sigma}_{\tilde{\epsilon}}^2/\sigma_{\tilde{\epsilon}}^2$ follows a chi-square distribution with $2K - 2$ degrees of freedom. Finally, $\hat{C}_1$ and $2K\hat{\sigma}_{\tilde{\epsilon}}^2/\sigma_{\tilde{\epsilon}}^2$ are independent RVs.

Thomson's test for a periodicity in colored noise is based upon the usual F-test for the significance of a regression parameter. Under the null hypothesis that $C_1 = 0$, the RV $\hat{C}_1$ has

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

a complex Gaussian distribution with zero mean and variance $\sigma_{\tilde{\epsilon}}^2 \, \Delta_t / \sum_{k=0,2,\ldots}^{2\lfloor (K-1)/2 \rfloor} H_k^2(0)$; moreover, because $J_k(f_1) = \tilde{\epsilon}_k$ under the null hypothesis, it follows from our assumptions about $\tilde{\epsilon}_k$ that the real and imaginary components of $\hat{C}_1$ are uncorrelated and have the same semivariance. Since $|\hat{C}_1|^2$ is thus the sum of squares of two uncorrelated real-valued Gaussian RVs with zero means and equal semivariances, we have

$$\frac{2\left|\hat{C}_1\right|^2 \sum_{k=0,2,\ldots}^{2\lfloor (K-1)/2 \rfloor} H_k^2(0)}{\sigma_{\tilde{\epsilon}}^2 \, \Delta_t} \stackrel{\mathrm{d}}{=} \chi_2^2.$$

Because $\hat{C}_1$ and the χ_{2K-2}^2 RV $2K\hat{\sigma}_{\tilde{\epsilon}}^2 / \sigma_{\tilde{\epsilon}}^2$ are independent, it follows that the above χ_2^2 RV is independent of $2K\hat{\sigma}_{\tilde{\epsilon}}^2 / \sigma_{\tilde{\epsilon}}^2$ also. The ratio of independent χ^2 RVs with 2 and $2K-2$ degrees of freedom – each divided by their respective degrees of freedom – is F-distributed with 2 and $2K-2$ degrees of freedom. Hence, we obtain (after some reduction)

$$\frac{(K-1)\left|\hat{C}_1\right|^2 \sum_{k=0,2,\ldots}^{2\lfloor (K-1)/2 \rfloor} H_k^2(0)}{\Delta_t \sum_{k=0}^{K-1} \left| J_k(f_1) - \hat{J}_k(f_1) \right|^2} \stackrel{\mathrm{d}}{=} F_{2,2K-2} \tag{547a}$$

(Thomson, 1982, equation (13.10)). If in fact $C_1 \neq 0$, the above statistic should give a value exceeding a high percentage point of the $F_{2,2K-2}$ distribution (the 95% point, say; Equation (544a) gives a formula for computing any desired percentage point).

If the F-test at $f = f_1$ is significant (i.e., the null hypothesis that $C_1 = 0$ is rejected), Thomson (1982) suggests reshaping the spectrum around f_1 to give a better estimate of $S_\eta(\cdot)$, the SDF of the background continuum $\{\eta_t\}$. Since $\sigma_{\tilde{\epsilon}}^2 = S_\eta(f_1)$, Equation (546b) can be rewritten as

$$\hat{S}_\eta(f_1) = \frac{1}{K} \sum_{k=0}^{K-1} \left| J_k(f_1) - \hat{J}_k(f_1) \right|^2 = \frac{1}{K} \sum_{k=0}^{K-1} \left| J_k(f_1) - \hat{C}_1 \frac{H_k(0)}{\Delta_t^{1/2}} \right|^2 .$$

For f in the neighborhood of f_1, i.e., $f \in [f_1 - B_{\overline{\mathcal{H}}}/2, f_1 + B_{\overline{\mathcal{H}}}/2]$, the above generalizes to

$$\hat{S}_\eta(f) = \frac{1}{K} \sum_{k=0}^{K-1} \left| J_k(f) - \hat{C}_1 \frac{H_k(f - f_1)}{\Delta_t^{1/2}} \right|^2 . \tag{547b}$$

For f sufficiently outside of this neighborhood, we have $H_k(f - f_1) \approx 0$, and the right-hand side above reverts to the usual basic multitaper estimator of Equation (352a). To ensure a smooth transition into the basic estimator, we recommend computing the reshaped spectrum over all frequencies f such that

$$|f - f_1| \leq 1.25 B_{\overline{\mathcal{H}}}/2. \tag{547c}$$

Finally, here are two related comments are about practical application of Thomson's method. First, since the eigencoefficients $J_k(\cdot)$ are typically computed using an FFT, f_1 would need to coincide with a standard frequency after padding with zeros to obtain a fine grid. Taking f_1 to be the standard frequency at which $\hat{C}_1$ is a maximum is one way to *define* the frequency at which to carry out the F-test. The theory we have presented assumes f_1 to be known a priori, so caution is in order in interpreting the significance of a peak if in fact f_1 is determined a posteriori using the time series under study. A quote from Thomson (1990a) is pertinent here (edited slightly to fit in with our exposition):

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

It is important to remember that in typical time-series problems hundreds or thousands of uncorrelated estimates are being dealt with; consequently one will encounter numerous instances of the F-test giving what would normally be considered highly significant test values that, in actuality, will only be sampling fluctuations. A good rule of thumb is not to get excited by significance levels greater than $1/N$.

Our second comment concerns multiple applications of Thomson's test. Note that the regression model in Equation (545c) is defined pointwise at each frequency. There is thus no need to assume that the eigencoefficients are uncorrelated across frequencies; however, because of the localized nature of the F-test at f_1, a test at a second frequency f_2 will be approximately independent of the first test as long as f_2 is outside of the interval $[f_1 - B_{\overline{\mathcal{H}}}, f_1 + B_{\overline{\mathcal{H}}}]$. The assumption of a slowly varying spectrum around f_1 must also now hold around f_2. This approach extends to multiple lines if analogous requirements are satisfied. (Denison and Walden, 1999, discuss the subtleties that can arise when applying Thomson's test multiple times.)

Comments and Extensions to Section 10.10

[1] So far we have restricted f_1 to be nonzero, but let now consider the special case $f_1 = 0$ so that Equation (544b) becomes $X_t = D_1 + \eta_t$ (we set $\phi_1 = 0$ since the phase of a constant term is undefined). The parameter D_1 is now the expected value of the stationary process $\{X_t\}$; i.e., $E\{X_t\} = D_1$. With C_1 defined to be $D_1/2$, Equation (545a) becomes

$$E\{J_k(0)\} = \frac{2C_1 H_k(0)}{\Delta_{\mathrm{t}}^{1/2}} = \frac{D_1 H_k(0)}{\Delta_{\mathrm{t}}^{1/2}},$$

and the least squares estimator of C_1 (Equation (546a)) becomes

$$\hat{C}_1 = \frac{\Delta_{\mathrm{t}}^{1/2}}{2} \frac{\sum_{k=0,2,\ldots}^{2\lfloor (K-1)/2 \rfloor} J_k(0) H_k(0)}{\sum_{k=0,2,\ldots}^{2\lfloor (K-1)/2 \rfloor} H_k^2(0)}, \quad \text{and hence} \quad \hat{D}_1 = \Delta_{\mathrm{t}}^{1/2} \frac{\sum_{k=0,2,\ldots}^{2\lfloor (K-1)/2 \rfloor} J_k(0) H_k(0)}{\sum_{k=0,2,\ldots}^{2\lfloor (K-1)/2 \rfloor} H_k^2(0)}.$$

If $K = 1$, the above reduces to

$$\hat{D}_1 = \Delta_{\mathrm{t}}^{1/2} \frac{J_0(0) H_0(0)}{H_0^2(0)} = \frac{\sum_{t=0}^{N-1} h_{0,t} X_t}{\sum_{t=0}^{N-1} h_{0,t}}.$$

This is the alternative estimator to the sample mean given in Equation (195b). The reshaped SDF estimator appropriate for this case is

$$\hat{S}_\eta(f) = \left| J_0(f) - \hat{D}_1 \frac{H_0(f)}{\Delta_{\mathrm{t}}^{1/2}} \right|^2 = \Delta_{\mathrm{t}} \left| \sum_{t=0}^{N-1} h_{0,t} \left(X_t - \hat{D}_1 \right) \mathrm{e}^{-\mathrm{i}2\pi f t \, \Delta_{\mathrm{t}}} \right|^2,$$

in agreement with the "mean corrected" direct spectral estimator of Equation (196a) with $\hat{D}_1$ substituted for $\tilde{\mu}$.

[2] We noted briefly that Thomson's F-test for a single spectral line is valid for multiple lines as long as the lines are well separated. Denison et al. (1999) discuss how to adapt the test to handle closely spaced spectral lines.

For well-separated lines, Wei and Craigmile (2010) propose tests that are similar in spirit to Thomson's in operating locally (i.e., the test at f_1 is based on spectral estimates just in the neighborhood of f_1) but they also consider tests operating globally (Fisher's g test of Equation (539b) is a prime example of a global test because it assesses a peak using values of the periodogram with frequencies ranging globally from just above zero to just below Nyquist). One of their tests is a local F-test like Thomson's, but it gets around the usual assumption in multitaper estimation of the SDF being locally constant by allowing the log SDF to vary linearly. They also discuss the advantages and disadvantages of local and global tests in terms of power to detect periodicities under various scenarios.

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

10.11 Completing a Harmonic Analysis

In Sections 10.5 and 10.6 we outlined ways of searching for spectral lines immersed in white noise by looking for peaks in the periodogram. We considered tests for periodicity – both simple and compound – against an alternative hypothesis of white noise in Section 10.9 and Thomson's multitaper-based test for detecting periodicity in a white or colored background continuum in the previous section. Suppose that, based on one of these tests, we have concluded that there are one or more line components in a time series. To complete our analysis, we must estimate the amplitudes of the line components and the SDF of the background continuum. If we have evidence that the background continuum is white noise, we need only estimate the variance σ_ϵ^2 of the white noise since its SDF is just $\sigma_\epsilon^2 \, \Delta_\mathrm{t}$; for the case of a colored continuum, we must estimate the SDF after somehow compensating for the line components.

We begin by considering the white noise case in detail, which assumes Equation (515b) as its model (in what follows, we assume $\mu = 0$ – see C&E [1] for how to deal with an unknown process mean). Suppose we have estimates $\hat{f}_l$ of the line frequencies obtained from the periodogram (or from some other method). We can then form *approximate conditional* least squares (ACLS) estimates of A_l and B_l using Equation (515c) with f_l replaced by $\hat{f}_l$:

$$\hat{A}_l = \frac{2}{N} \sum_{t=0}^{N-1} X_t \cos\left(2\pi \hat{f}_l t \, \Delta_\mathrm{t}\right) \text{ and } \hat{B}_l = \frac{2}{N} \sum_{t=0}^{N-1} X_t \sin\left(2\pi \hat{f}_l t \, \Delta_\mathrm{t}\right). \tag{549a}$$

Here "conditional" refers to the use of the estimator $\hat{f}_l$ in lieu of knowing the actual f_l, and "approximate" acknowledges that, for non-Fourier frequencies, these estimators will not be exact least squares estimators (see Section 10.2). We estimate the size of the jumps at $\pm \hat{f}_l$ in the integrated spectrum for $\{X_t\}$ by

$$\frac{\hat{A}_l^2 + \hat{B}_l^2}{4} = \frac{\hat{S}^{(\mathrm{P})}(\hat{f}_l)}{N \, \Delta_\mathrm{t}} \tag{549b}$$

(see the discussion surrounding Equation (523a)). The estimator of the white noise variance σ_ϵ^2 is

$$\hat{\sigma}_\epsilon^2 = \frac{\mathrm{SS}(\{\hat{A}_l\}, \{\hat{B}_l\}, \{\hat{f}_l\})}{N - 2L}, \tag{549c}$$

where

$$\mathrm{SS}(\{A_l\}, \{B_l\}, \{f_l\}) \stackrel{\mathrm{def}}{=} \sum_{t=0}^{N-1} \left(X_t - \sum_{l=1}^{L} \left[A_l \cos\left(2\pi f_l t \, \Delta_\mathrm{t}\right) + B_l \sin\left(2\pi f_l t \, \Delta_\mathrm{t}\right) \right] \right)^2. \tag{549d}$$

The divisor $N - 2L$ is motivated by standard linear regression theory, i.e., the sample size minus the number of parameters estimated (Weisberg, 2014). The estimator of the SDF $S_\epsilon(\cdot)$ of the white noise is $\hat{S}_\epsilon(f) = \hat{\sigma}_\epsilon^2 \, \Delta_\mathrm{t}$.

A second approach would be to find *exact conditional* least squares (ECLS) estimates for A_l and B_l by locating those values of A_l and B_l actually minimizing $\mathrm{SS}(\{A_l\}, \{B_l\}, \{\hat{f}_l\})$. These estimates can be obtained by regressing X_t on $\cos\left(2\pi \hat{f}_l t \, \Delta_\mathrm{t}\right)$ and $\sin\left(2\pi \hat{f}_l t \, \Delta_\mathrm{t}\right)$, $l = 1, \ldots, L$. If we let $\hat{A}_l$ and $\hat{B}_l$ denote the minimizers, the estimator of the jumps takes the form of the left-hand side of Equation (549b), while the estimator of σ_ϵ^2 is again given by Equation (549c).

A third approach would be to find *exact unconditional* least squares (EULS) estimates for A_l, B_l and f_l by locating those values of A_l, B_l and f_l that minimize $\mathrm{SS}(\{A_l\}, \{B_l\}, \{f_l\})$.

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

With $\hat{A}_l$, $\hat{B}_l$ and $\hat{f}_l$ denoting the minimizers, the left-hand side of Equation (549b) would again estimate the jumps, but now linear least squares theory suggests the following estimator for the white noise variance:

$$\tilde{\sigma}_\epsilon^2 = \frac{\mathrm{SS}(\{\hat{A}_l\}, \{\hat{B}_l\}, \{\hat{f}_l\})}{N - 3L}; \tag{550a}$$

however, the fact that the estimators arise from nonlinear least squares dictates caution in use of the above.

In practice, we must use a nonlinear optimization routine to find EULS parameter estimates. Such routines typically require initial settings of the parameters, for which we can use the ACLS estimates. Bloomfield (2000) has an extensive discussion of EULS estimates. For most time series encountered in nature, ACLS estimates are adequate, but Bloomfield gives a pathological artificial example where EULS estimates are remarkably better than the ACLS and ECLS estimates.

Once we have obtained estimates of A_l, B_l and f_l by any of these three approaches, we can examine the adequacy of our fitted model by looking at the *residual process* $\{R_t\}$, where

$$R_t \stackrel{\mathrm{def}}{=} X_t - \sum_{l=1}^{L} \Big[\hat{A}_l \cos\left(2\pi \hat{f}_l t\, \Delta_{\mathrm{t}}\right) + \hat{B}_l \sin\left(2\pi \hat{f}_l t\, \Delta_{\mathrm{t}}\right) \Big]. \tag{550b}$$

If our fitted model is adequate, the series $\{R_t\}$ should resemble a realization of a white noise process because we can regard R_t as a proxy for the unknown ϵ_t. This suggests assessing $\{R_t\}$ using periodogram-based tests for white noise (e.g., Fisher's test of Equation (539b) or the normalized cumulative periodogram test that follows from Equation (215a)). Caution is in order here because $\{R_t\}$ tends to be deficient in power around the line frequencies, which is inconsistent with the hypothesis that the residuals are a realization of white noise. Figure 551 shows an example (Exercise [10.14a] expands upon it). We start with the time series whose periodogram is shown in Figure 534(c). We treat this series as if it were a realization of length $N = 128$ of the process of Equation (515b) with frequencies f_1, f_2 and f_3 dictated as per Equation (530) and with white noise variance $\sigma_\epsilon^2 = 1$. The EULS procedure yields a set of residuals $\{R_t\}$ whose periodogram is shown in Figure 551. The three dashed vertical lines indicate the EULS estimates of the line frequencies, which are in good agreement with the true frequencies (these are indicated by similar lines in Figure 534(c)). Whereas the periodogram for the original series has peaks at these frequencies, the periodogram of the residuals has pronounced troughs.

Now we consider the second possibility, namely, a colored continuum. Our model is now

$$X_t = \sum_{l=1}^{L} \left[A_l \cos\left(2\pi f_l t\, \Delta_{\mathrm{t}}\right) + B_l \sin\left(2\pi f_l t\, \Delta_{\mathrm{t}}\right) \right] + \eta_t,$$

where $\{\eta_t\}$ is a stationary process with zero mean and colored SDF $S_\eta(\cdot)$. As in the case of a white continuum, we could use either the ACLS, ECLS or EULS approach and then form the residual process $\{R_t\}$. If $\hat{A}_l$, $\hat{B}_l$ and $\hat{f}_l$ are good estimates, then $\{R_t\}$ should be a good approximation to $\{\eta_t\}$. We can thus estimate $S_\eta(\cdot)$ by computing an SDF estimate using these residuals, but again we note that such estimates tend to have a deficiency of power around the line frequencies.

A second approach is to reformulate the model in terms of complex exponentials, i.e., to write

$$X_t = \sum_{l=1}^{L} \left(C_l \mathrm{e}^{\mathrm{i}2\pi f_l t\, \Delta_{\mathrm{t}}} + C_l^* \mathrm{e}^{-\mathrm{i}2\pi f_l t\, \Delta_{\mathrm{t}}} \right) + \eta_t, \tag{550c}$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

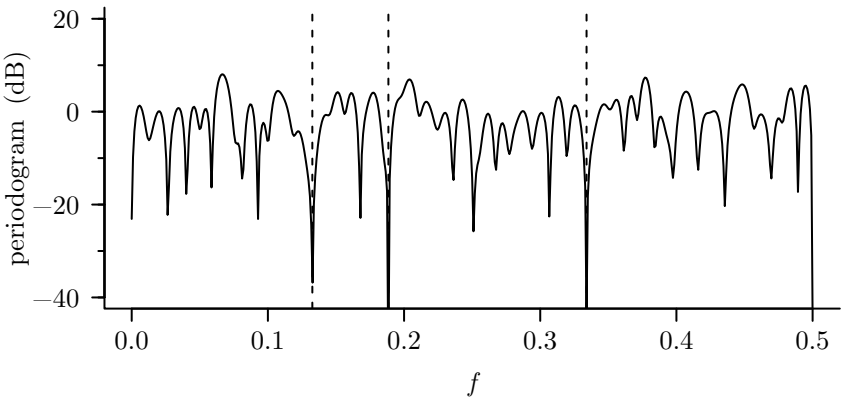

Figure 551 Periodogram for residuals from EULS estimates The residuals are derived from a time series whose periodogram (shown in Figure 534(c)) has three prominent peaks. Fitting an order $L = 3$ model of the form of Equation (515b) using the EULS method yields residuals whose periodogram, as this figure illustrates, now has pronounced troughs at the peak locations (indicated by the dashed lines)

and then to estimate C_l via the multitaper approach of the previous section. If the $\hat{f}_l$ terms are sufficiently separated and not too close to zero or the Nyquist frequency, as discussed, then the estimator of C_l is given by

$$\hat{C}_l = \Delta_t^{1/2} \frac{\sum_{k=0,2,\ldots}^{2\lfloor (K-1)/2 \rfloor} J_k(\hat{f}_l) H_k(0)}{\sum_{k=0,2,\ldots}^{2\lfloor (K-1)/2 \rfloor} H_k^2(0)} \tag{551a}$$

(cf. Equation (546a)), where $J_k(\hat{f}_l)$ and $H_k(0)$ are as defined previously. The size of the jumps in the integrated spectrum at $\pm\hat{f}_l$ would be estimated by $|\hat{C}_l|^2$. The SDF $S_\eta(\cdot)$ of the background continuum would be estimated by reshaping – in the neighborhoods of each $\hat{f}_l$ – the multitaper SDF estimate based upon $\{X_t\}$ (see Equation (547b)). This estimate of $S_\eta(\cdot)$ tends to be better behaved around the line frequencies than estimates based upon the residual process.

Finally we note that the multitaper approach can also be used with a white noise continuum. Here we need only estimate the variance σ_ϵ^2 of the white noise, and the appropriate estimator is now

$$\hat{\sigma}_\epsilon^2 = \frac{1}{N - 2L} \sum_{t=0}^{N-1} \left(X_t - \sum_{l=1}^{L} \left[\hat{C}_l e^{i2\pi \hat{f}_l t \, \Delta_t} + \hat{C}_l^* e^{-i2\pi \hat{f}_l t \, \Delta_t} \right] \right)^2 . \tag{551b}$$

This expression can be reduced somewhat by noting that

$$\hat{C}_l e^{i2\pi \hat{f}_l t \, \Delta_t} + \hat{C}_l^* e^{-i2\pi \hat{f}_l t \, \Delta_t} = 2\Re(\hat{C}_l) \cos(2\pi \hat{f}_l t \, \Delta_t) - 2\Im(\hat{C}_l) \sin(2\pi \hat{f}_l t \, \Delta_t),$$

where $\Re(z)$ and $\Im(z)$ refer, respectively, to the real and imaginary components of the complex number z. Exercise [10.14b] considers an example of this approach. (Rife and Boorstyn, 1974, 1976, present an interesting discussion about the estimation of complex exponentials in white noise.)

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

Comments and Extensions to Section 10.11

[1] In formulating the ACLS, ECLS and EULS estimators, we assumed for convenience that $\mu = 0$ in Equation (515b), which is not a viable assumption in most practical applications. When μ is unknown, we can use the sample mean $\overline{X}$ to center the time series, thus replacing X_t by $X_t - \overline{X}$ in Equations (549a) and (549d). Another approach is possible with the ECLS and EULS estimators by treating μ as an additional parameter to be estimated via least squares. In place of Equation (549d), we now consider SS($\mu, \{A_l\}, \{B_l\}, \{f_l\}$), which is defined to be

$$\sum_{t=0}^{N-1} \left(X_t - \mu - \sum_{l=1}^{L} [A_l \cos(2\pi f_l t\, \Delta_t) + B_l \sin(2\pi f_l t\, \Delta_t)] \right)^2 .$$

The minimizing values of SS($\mu, \{A_l\}, \{B_l\}, \{\hat{f}_l\}$) are the ECLS estimators for μ, A_l and B_l, and the ECLS estimator $\hat{\mu}$ is not in general equal to $\overline{X}$; likewise, the EULS estimators for μ, A_l, B_l and f_l are the minimizing values of SS($\mu, \{A_l\}, \{B_l\}, \{f_l\}$), and again $\hat{\mu}$ need not be equal to the sample mean. The divisors in the estimators $\hat{\sigma}_\epsilon^2$ and $\tilde{\sigma}_\epsilon^2$ of Equations (549c) and (550a) now become, respectively, $N - 2L - 1$ and $N - 3L - 1$. C&E [1] for Section 10.15 gives an example of these estimators of μ.

[2] For certain time series, we might know the frequencies f_l a priori, in which case we can use the known f_l in place of $\hat{f}_l$ in Equations (549a), (549b) and (549c). Because we no longer need to condition on an estimator of f_l, the methods we have previously described as yielding approximate conditional and exact conditional least squares estimators now give us approximate unconditional and exact unconditional least squares estimators.

[3] We considered estimators of the jumps in the integrated spectrum in this section, but we have yet to discuss their distributions. Here we derive the distribution for a jump estimator appropriate for the model of Equation (513) with (1) $\mu = 0$, (2) f set to $f_k = k/(N\,\Delta_t)$ for some integer k such that $1 \le k < N/2$ and (3) $\{\epsilon_t\}$ assumed to be Gaussian. Under these restrictive assumptions, the exact least squares estimators $\hat{A}$ and $\hat{B}$ for A and B are given by Equation (549a), and standard least square theory says that $\hat{A}$ and $\hat{B}$ are Gaussian RVs with means A and B and a covariance matrix given by $\sigma_\epsilon^2 (\boldsymbol{H}^T \boldsymbol{H})^{-1}$, where

$$\boldsymbol{H}^T = \begin{bmatrix} 1 & \cos(2\pi f_k\, \Delta_t) & \cos(4\pi f_k\, \Delta_t) & \cdots & \cos([N-1]2\pi f_k\, \Delta_t) \\ 0 & \sin(2\pi f_k\, \Delta_t) & \sin(4\pi f_k\, \Delta_t) & \cdots & \sin([N-1]2\pi f_k\, \Delta_t) \end{bmatrix}$$

(see, e.g., Weisberg, 2014). We can appeal to Exercise [1.3c] to obtain the elements of $\boldsymbol{H}^T \boldsymbol{H}$ (and hence $(\boldsymbol{H}^T \boldsymbol{H})^{-1}$), thus establishing that $\hat{A}$ and $\hat{B}$ are uncorrelated (and hence independent due to Gaussianity) and have a common variance of $2\sigma_\epsilon^2/N$. The true jump is $J = (A^2 + B^2)/4$, and the obvious jump estimator is $\hat{J} = (\hat{A}^2 + \hat{B}^2)/4$ (cf. the left-hand side of Equation (549b)), which is a quadratic form in Gaussian RVs. Now, if U and V are independent Gaussian RVs with means μ_U and μ_V and with unit variances, then the RV $\chi_2^2(\beta) = U^2 + V^2$ is said to obey a *noncentral chi-square distribution* with two degrees of freedom and with noncentrality parameter $\beta = \mu_U^2 + \mu_V^2$ (see, e.g., Anderson, 2003; a related RV arose in Equation (542c) concerning Siegel's test statistic). After a normalization to satisfy the unit variance criterion, we can claim that the normalized jump estimator $2N\hat{J}/\sigma_\epsilon^2$ is distributed as $\chi_2^2(\beta)$ with $\beta = N(A^2 + B^2)/(2\sigma_\epsilon^2) = NR$, where R is the signal-to-noise ratio of Equation (526d). We need to know σ_ϵ^2 to formulate this normalized estimator; in the usual case where σ_ϵ^2 is unknown, we can use the standard estimator $\hat{\sigma}_\epsilon^2 = \text{SS}(\hat{A}, \hat{B}, f_k)/(N-2)$, where SS is defined as per Equation (549d), and $(N-2)\hat{\sigma}_\epsilon^2/\sigma_\epsilon^2$ has the distribution of an ordinary χ_{N-2}^2 RV. Now the ratio of a $\chi_2^2(\beta)/2$ RV to an independent $\chi_{N-2}^2/(N-2)$ RV yields an RV obeying a *noncentral F-distribution* with 2 and $N-2$ degrees of freedom and with noncentrality parameter β (Anderson, 2003). Let $F_{2,N-2,\beta}(p)$ represent the $p \times 100\%$ percentage point for this distribution. Since the statistic $N\hat{J}/\hat{\sigma}_\epsilon^2$ is distributed as such, the distribution of $\hat{J}$ is dictated by

$$\mathbf{P}[\hat{J} \le \hat{\sigma}_\epsilon^2 F_{2,N-2,\beta}(p)/N] = p. \tag{552}$$

Unfortunately this distribution depends on the typically unknown signal-to-noise ratio R through β, a fact that complicates using the above to generate a confidence interval (CI) for the true jump J. Exercise [10.15] considers using Equation (552) with a plug-in approach for creating CIs, but, while this works reasonably well for certain settings for R, it can lead to disconcertingly conservative CIs for others.

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

10.12 A Parametric Approach to Harmonic Analysis

Autoregressive (AR) spectral estimation, as discussed in Chapter 9, is thought of by statisticians as a method for estimating SDFs, i.e., a spectrum with only a purely continuous component and no line components. However, this technique has been used in applied work to estimate the frequencies of line components, perhaps immersed in a background continuum (Ulrych, 1972b; Chen and Stegen, 1974; Satorius and Zeidler, 1978; Kane and de Paula, 1996; Sutcliffe et al., 2013; Tary et al., 2014). We investigate the AR approach to spectral line estimation in this section.

Consider a *deterministic* real sinusoid of the form $x_t = D\cos(2\pi f t\,\Delta_t + \phi)$, $t \in \mathbb{Z}$, where $0 < f \le 1/(2\,\Delta_t)$, $D > 0$ and $-\pi < \phi \le \pi$.

▷ **Exercise [553]** Show that $\{x_t\}$ gives rise to the second-order difference equation

$$x_t = 2\cos(2\pi f\,\Delta_t)x_{t-1} - x_{t-2}. \tag{553a}$$

Hint: write the cosines as complex exponentials. ◁

Equation (553a) does not depend on the fixed amplitude D or phase ϕ. These come into play by specifying initial conditions, which we can take to be setting any two adjacent values in $\{x_t\}$, e.g., $x_0 = D\cos(\phi)$ and $x_1 = D\cos(2\pi f\,\Delta_t + \phi)$. With these conditions, we can use Equation (553a) and the related equation $x_{t-2} = 2\cos(2\pi f\,\Delta_t)x_{t-1} - x_t$ to generate the remaining terms in $\{x_t\}$. (Exercise [10.16] explores an alternative way of creating a sinusoidal sequence satisfying Equation (553a) by specifying f, x_0 and x_1 rather than f, D and ϕ.)

Now consider a *randomly phased* sinusoid of the form $X_t = D\cos(2\pi f t\,\Delta_t + \phi)$, where ϕ is an RV uniformly distributed between $-\pi$ and π. The process $\{X_t\}$ satisfies

$$X_t = 2\cos(2\pi f\,\Delta_t)X_{t-1} - X_{t-2}, \tag{553b}$$

which is analogous to Equation (553a); here the initial conditions $X_0 = D\cos(\phi)$ and $X_1 = D\cos(2\pi f\,\Delta_t + \phi)$ involve the RV ϕ and the constant D. We call the above X_t a *pseudo*-AR(2) process. This terminology is meant to remind us that Equation (553b) lacks one aspect of the usual AR(2) process – the innovation term ϵ_t. If we write Equation (553b) as

$$X_t = \varphi_{2,1}X_{t-1} + \varphi_{2,2}X_{t-2},$$

it is easy to verify that the roots of the polynomial equation

$$1 - \varphi_{2,1}z^{-1} - \varphi_{2,2}z^{-2} = 0$$

are $z = \exp(\pm \mathrm{i}2\pi f\,\Delta_t)$, both of which are on the unit circle. We know that, for a randomly phased sinusoid, we hace $\operatorname{cov}\{X_{t+\tau}, X_t\} = D^2\cos(2\pi f\tau\,\Delta_t)/2$ (cf. Equation (37b) with $L = 1$ and with Δ_t inserted or Equation (519e) with no noise). Exercise [10.17] is to derive the same result directly from Equation (553b).

Just as a single sinusoid satisfies a second-order difference equation, the summation of p sinusoids satisfies a difference equation of order $2p$. If the roots $\{z_l\}$ of the polynomial equation

$$1 - \sum_{k=1}^{2p} \varphi_{2p,k}z^{-k} = 0 \tag{553c}$$

are all on the unit circle in conjugate pairs so that $z_l = \exp(\pm \mathrm{i}2\pi f_l\,\Delta_t)$, $l = 1, 2, \ldots, p$, then the pseudo-AR($2p$) equation

$$X_t = \sum_{k=1}^{2p} \varphi_{2p,k}X_{t-k} \tag{553d}$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

has a solution given by

$$X_t = \sum_{l=1}^{p} D_l \cos\left(2\pi f_l t\,\Delta_t + \phi_l\right) \tag{554a}$$

(cf. Equation (511) with μ set to zero; proof of the above constitutes Exercise [10.18]). As an example of how to obtain the $\varphi_{2p,k}$ coefficients given the f_l frequencies, consider Equation (553c) for the case of $p = 2$ sinusoids with $\Delta_t = 1$. We have

$$1 - \varphi_{4,1}e^{i2\pi f_1} - \varphi_{4,2}e^{i4\pi f_1} - \varphi_{4,3}e^{i6\pi f_1} - \varphi_{4,4}e^{i8\pi f_1} = 0$$

and

$$1 - \varphi_{4,1}e^{i2\pi f_2} - \varphi_{4,2}e^{i4\pi f_2} - \varphi_{4,3}e^{i6\pi f_2} - \varphi_{4,4}e^{i8\pi f_2} = 0.$$

The real and imaginary parts of the above lead to the matrix equation

$$\begin{bmatrix} \cos(2\pi f_1) & \cos(4\pi f_1) & \cos(6\pi f_1) & \cos(8\pi f_1) \\ \sin(2\pi f_1) & \sin(4\pi f_1) & \sin(6\pi f_1) & \sin(8\pi f_1) \\ \cos(2\pi f_2) & \cos(4\pi f_2) & \cos(6\pi f_2) & \cos(8\pi f_2) \\ \sin(2\pi f_2) & \sin(4\pi f_2) & \sin(6\pi f_2) & \sin(8\pi f_2) \end{bmatrix} \begin{bmatrix} \varphi_{4,1} \\ \varphi_{4,2} \\ \varphi_{4,3} \\ \varphi_{4,4} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \tag{554b}$$

from which $\varphi_{4,1}, \ldots, \varphi_{4,4}$ can be obtained (see, however, Exercise [10.19], which argues that we must have $\varphi_{4,4} = -1$ and that we can use a related three-dimensional matrix equation to solve for $\varphi_{4,1}$, $\varphi_{4,2}$ and $\varphi_{4,3}$).

So far we have only considered perfectly observed sinusoids. We now consider randomly phased sinusoids plus white noise. For p sinusoids observed with zero mean white noise $\{\alpha_t\}$, the observed process is the sum of a pseudo-AR($2p$) process and white noise:

$$\widetilde{X}_t = X_t + \alpha_t = \sum_{k=1}^{2p} \varphi_{2p,k} X_{t-k} + \alpha_t, \tag{554c}$$

where $\operatorname{cov}\{\alpha_{t+\tau}, X_t\} = 0$ for all integers τ; i.e., the noise $\{\alpha_t\}$ is assumed to be uncorrelated with the "signal" $\{X_t\}$. If we substitute $X_{t-k} = \widetilde{X}_{t-k} - \alpha_{t-k}$ into the right-hand side of the above, we get

$$\widetilde{X}_t - \sum_{k=1}^{2p} \varphi_{2p,k} \widetilde{X}_{t-k} = \alpha_t - \sum_{k=1}^{2p} \varphi_{2p,k} \alpha_{t-k}, \tag{554d}$$

which is an ARMA($2p$, $2p$) process whose AR and MA coefficients are identical. Such a result was first derived by Ulrych and Clayton (1976), but without the stipulation that sinusoids were randomly phased and hence stochastic. Kay and Marple (1981, p. 1403) report a method (essentially due to Pisarenko, 1973) for estimating the parameters $\varphi_{2p,k}$ in Equation (554c).

It is important to clearly distinguish between a pseudo-AR process (i.e., one without an innovation term) plus white noise and a standard AR process plus white noise. We have just seen how the former – for the case of p randomly phased sinusoids plus white noise – yields an ARMA($2p, 2p$) process with identical AR and MA coefficients. In the latter case, however, an ARMA process with an equal number of AR and MA coefficients is again obtained but now with *distinct* coefficients (Walker, 1960; Tong, 1975; Friedlander, 1982). To see this, let $\{Y_t\}$ be a standard AR(p) process with innovation variance σ_p^2 as in Equation (446a), and let $\{\xi_t\}$ be a white noise process with variance σ_ξ^2 that is uncorrelated with $\{Y_t\}$. Then the process defined by $\widetilde{Y}_t = Y_t + \xi_t$ has an SDF given by

$$\begin{aligned} S(f) &= \frac{\sigma_p^2\,\Delta_t}{\left|1 - \sum_{k=1}^{p} \phi_{p,k} e^{-i2\pi f k\,\Delta_t}\right|^2} + \sigma_\xi^2\,\Delta_t \\ &= \frac{\sigma_p^2\,\Delta_t + \sigma_\xi^2\,\Delta_t \left|1 - \sum_{k=1}^{p} \phi_{p,k} e^{-i2\pi f k\,\Delta_t}\right|^2}{\left|1 - \sum_{k=1}^{p} \phi_{p,k} e^{-i2\pi f k\,\Delta_t}\right|^2}. \end{aligned}$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

where we have evoked Equation (446b), along with the fact that the SDF for the white noise process is $\sigma_\xi^2 \, \Delta_t$ and the result of Exercise [4.12]. If we let

$$\sigma_p^2 + \sigma_\xi^2 \left| 1 - \sum_{k=1}^{p} \phi_{p,k} e^{-i2\pi f k \, \Delta_t} \right|^2 = \sigma_\zeta^2 \left| 1 - \sum_{k=1}^{p} \theta_{p,k} e^{-i2\pi f k \, \Delta_t} \right|^2$$

by appropriately defining σ_ζ^2 and $\theta_{p,k}$, we obtain

$$S(f) = \sigma_\zeta^2 \, \Delta_t \frac{\left| 1 - \sum_{k=1}^{p} \theta_{p,k} e^{-i2\pi f k \, \Delta_t} \right|^2}{\left| 1 - \sum_{k=1}^{p} \phi_{p,k} e^{-i2\pi f k \, \Delta_t} \right|^2}, \tag{555a}$$

which is the SDF for an ARMA(p,p) process with AR parameters $\phi_{p,k}$, MA parameters $\theta_{p,k}$ and innovation variance σ_ζ^2. The $\theta_{p,k}$ terms are no longer identical to the $\phi_{p,k}$ terms, but note that the $p+1$ values $\{\theta_{p,k}\}$ and σ_ζ^2 are determined by the $p+2$ values $\{\phi_{p,k}\}$, σ_p^2 and σ_ξ^2. (Proof that Equation (555a) gives the resulting SDF is the subject of Exercise [10.20].)

Since a standard AR(p) process plus white noise is an ARMA(p,p) process, the AR parameters can be estimated from the ARMA(p,p) nature of the observed process. This is done by Tong (1975) and Friedlander (1982). These methods, however, do not make use of the fact that there are only $p+2$ free parameters, but rather they treat the model as if there were $2p+2$ free parameters ($\phi_{p,1}, \ldots, \phi_{p,p}$; $\theta_{p,1}, \ldots, \theta_{p,p}$; σ_p^2 and σ_ξ^2); i.e., the MA parameters are not considered related to the other $p+2$ parameters. Jones (1980) and Tugnait (1986) present algorithms for solving the AR parameters that involve only the $p+2$ free parameters.

As a simple illustration of the AR approach to sinusoid estimation, we consider an example due to Makhoul (1981b) of a deterministic sinusoid with a frequency of $f = 1/6$ Hz and a sampling interval of $\Delta_t = 1$ sec. With these values substituted into Equation (553a), the appropriate difference equation becomes

$$x_t = 2\cos(\pi/3) x_{t-1} - x_{t-2} = x_{t-1} - x_{t-2}. \tag{555b}$$

For convenience let us assume that the amplitude of the sinusoid is $D = 1/\cos(\pi/6)$ and its phase is $\phi = -\pi/2$. The initial conditions are thus

$$x_0 = D\cos(\phi) = D\cos(-\pi/2) = 0$$

and

$$x_1 = D\cos(2\pi f + \phi) = D\cos(-\pi/6) = 1,$$

and subsequent values are

$$x_2 = x_1 - x_0 = 1, \qquad x_3 = x_2 - x_1 = 0,$$
$$x_4 = x_3 - x_2 = -1, \qquad x_5 = x_4 - x_3 = -1,$$

etc.

Suppose that we have only the samples x_0, x_1, x_2 and x_4 and that we seek to estimate the parameters $\phi_{2,1}$, $\phi_{2,2}$ and σ_2^2 of a *standard* AR(2) model. If we do so by using the forward/backward least squares (FBLS) estimator, we must find the values of $\phi_{2,1}$ and $\phi_{2,2}$ such that the forward/backward sum of squares

$$\sum_{t=2}^{3} \left[(x_t - \phi_{2,1} x_{t-1} - \phi_{2,2} x_{t-2})^2 + (x_{t-2} - \phi_{2,1} x_{t-1} - \phi_{2,2} x_t)^2 \right] \tag{555c}$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

is minimized (cf. Equation (477d)). If we call the solution points $\hat{\phi}_{2,1}$ and $\hat{\phi}_{2,2}$, we find that the minimum sum of squares is *zero* when $\hat{\phi}_{2,1} = 1$ and $\hat{\phi}_{2,2} = -1$; i.e., the fitted model is *exactly* the true model of Equation (555b). Note that the innovation term ϵ_t can be dropped from the fitted model – it has a mean of zero by assumption, and our estimate of the innovation variance would be zero, implying that $\epsilon_t = 0$ for all t.

If now we consider Burg's algorithm instead (Section 9.5), we must first fit an AR(1) model and estimate $\phi_{1,1}$ by minimizing the forward/backward sum of squares

$$\sum_{t=1}^{3} \left[(x_t - \phi_{1,1} x_{t-1})^2 + (x_{t-1} - \phi_{1,1} x_t)^2 \right] \tag{556a}$$

(see Equation (468a)). It follows from Exercise [9.16a] that the estimate is

$$\bar{\phi}_{1,1} = \frac{2 \sum_{t=1}^{3} x_t x_{t-1}}{x_0^2 + 2 \sum_{t=1}^{2} x_t^2 + x_3^2} = 1/2, \tag{556b}$$

after substitution of the sample values. We now use the Levinson–Durbin relationship

$$\phi_{2,1} = \phi_{1,1} - \phi_{2,2}\phi_{1,1} = \phi_{1,1}\,(1 - \phi_{2,2}).$$

If we now replace $\phi_{2,1}$ in Equation (555c) with the right-hand side of the above (with $\bar{\phi}_{1,1}$ replacing $\phi_{1,1}$), the next step in Burg's algorithm is the minimization of

$$\sum_{t=2}^{3} \left(\left[x_t - (1 - \phi_{2,2})\,\bar{\phi}_{1,1} x_{t-1} - \phi_{2,2} x_{t-2} \right]^2 + \left[x_{t-2} - (1 - \phi_{2,2})\,\bar{\phi}_{1,1} x_{t-1} - \phi_{2,2} x_t \right]^2 \right) \tag{556c}$$

with respect to $\phi_{2,2}$. With the data substituted we obtain $\bar{\phi}_{2,2} = -1$, and hence $\bar{\phi}_{2,1} = (1 - \bar{\phi}_{2,2})\bar{\phi}_{1,1} = 1$. These are the same estimates as obtained previously. (Makhoul, 1981b, appears to have miscalculated $\bar{\phi}_{1,1}$ and hence derived poor estimates of $\phi_{2,1}$ and $\phi_{2,2}$ from Burg's algorithm.)

As another example of the parametric approach to harmonic analysis, let us consider the stochastic process of Equation (537b) that we discussed previously in the context of tapering (see Figure 538b). This process consists of a single sinusoid with a frequency of $f_1 = 0.0725$ plus a small amount of white noise. We use the FBLS estimator to fit AR(16) models to the same 50 realizations entertained in Figure 538b. For each realization, we determine the frequency of the peak value in the AR SDF estimate (see C&E [3] for details on how to find this peak location). Figure 538b shows a scatter plot of peak locations as determined from a Hanning-based direct spectral estimator $\hat{S}^{(\mathrm{D})}(\cdot)$ (vertical axis) and the periodogram $\hat{S}^{(\mathrm{P})}(\cdot)$ (horizontal). Figure 557 has a similar scatter plot, but now we display the AR-based peak locations on the horizontal axis (the locations for $\hat{S}^{(\mathrm{D})}(\cdot)$ are again shown on the vertical axis). Note that there is no indication of bias in either estimator of f_1; however, the sample variance of the estimator based upon $\hat{S}^{(\mathrm{D})}(\cdot)$ is larger by a factor of 1.5. This example illustrates the usefulness of the parametric approach to harmonic analysis, but we must view its apparent superiority over the nonparametric approach with caution: first, we have used the standard Hanning taper rather than seeking an optimal taper for peak estimation for the model of Equation (537b) and, second, we set $p = 16$ by optimizing over AR orders 2 to 32 via a Monte Carlo experiment whose replication is the burden of Exercise [10.21]. (Use of the FBLS method is important here – the Yule–Walker and Burg estimators perform poorly on this example, with increases in sample variances by factors of, respectively, 5.1 and 676.6 over that for the FBLS estimator.)

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

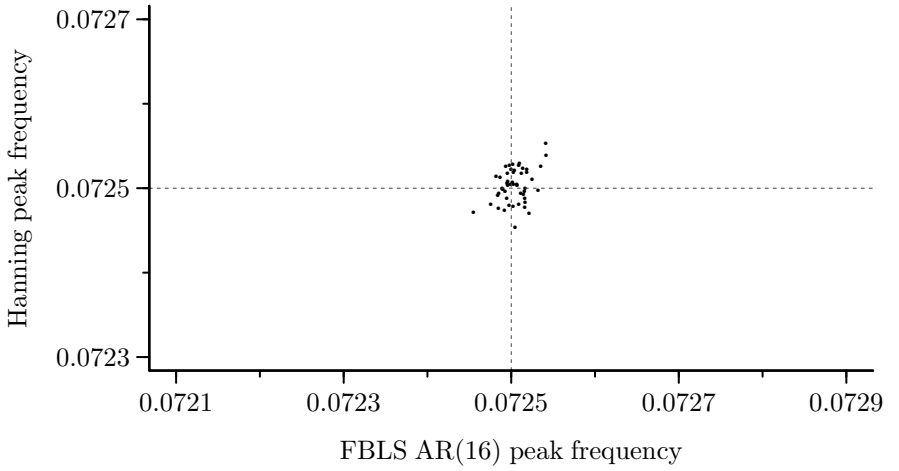

Figure 557 Scatter plot of locations of peak frequencies computed from a Hanning-based direct spectral estimator $\hat{S}^{(\mathrm{D})}(\cdot)$ (vertical axis) versus locations determined from an AR(16) SDF estimator with parameters estimated using the forward/backward least squares method (horizontal axis). Each point in the scatter plot corresponds to location estimators for one of 50 different realizations from the model of Equation (537b). The true location of the peak frequency is 0.0725 (indicated by the vertical and horizontal lines). (This figure should be compared with Figure 538b, which makes use of the same 50 realizations.)

Comments and Extensions to Section 10.12

[1] The discussion following Equation (446c) states that a necessary and sufficient condition for the existence of a stationary solution to the AR($2p$) process

$$Y_t = \sum_{k=1}^{2p} \phi_{2p,k} Y_{t-k} + \epsilon_t$$

(cf. Equation (446a)) is that *none* of the solutions of the polynomial equation

$$1 - \sum_{k=1}^{2p} \phi_{2p,k} z^{-k} = 0$$

lies on the unit circle in the complex plane. In this section we noted that a randomly phased harmonic process with p components can be represented as

$$X_t = \sum_{k=1}^{2p} \varphi_{2p,k} X_{t-k}$$

(Equation (553d)). This harmonic process is stationary, but its associated polynomial equation

$$1 - \sum_{k=1}^{2p} \varphi_{2p,k} z^{-k} = 0$$

(Equation (553c)) has *all* of its roots on the unit circle. This stark contrast in the conditions for stationarity is entirely due to the innovation term ϵ_t with nonzero variance and highlights the importance of carefully distinguishing between AR processes $\{Y_t\}$ and pseudo-AR processes $\{X_t\}$.

[2] For readers familiar with the notion of a backward shift operator and its use in expressing stationary nondeterministic ARMA processes (see, for example, Box et al., 2015), we point out the fact – unfortunately sometimes overlooked in the literature – that, since the roots of Equation (553c) are on the unit

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

circle, the operator common to the AR and MA components of Equation (554d) does not have an inverse and hence cannot be canceled out.

[3] We can accurately determine the locations of the peaks in an AR SDF in a manner analogous to that used to locate the peak frequencies in the direct spectral estimators (see C&E [2] for Section 10.6). Recall that the SDF for a stationary AR(p) process has the form

$$S(f) = \frac{\sigma_p^2 \, \Delta_t}{\left| 1 - \sum_{j=1}^{p} \phi_{p,j} e^{-i2\pi f j \, \Delta_t} \right|^2}$$

(this is Equation (446b)). With $\phi_{p,0}$ taken to be -1, the denominator becomes

$$D(f) = \left| \sum_{j=0}^{p} \phi_{p,j} e^{-i2\pi f j \, \Delta_t} \right|^2 .$$

We can regard $D(\cdot)$ as proportional to the periodogram for a "time series" $X_j = \phi_{p,j}$ of length $N = p+1$. Comparing the above to Equation (170d) and then taking into consideration Equations (170c) and (170b), we can write

$$D(f) = (p+1) \sum_{\tau=-p}^{p} \hat{s}_{\phi,\tau}^{(P)} e^{-i2\pi f \tau \, \Delta_t}, \quad \text{where } \hat{s}_{\phi,\tau}^{(P)} = \frac{1}{p+1} \sum_{j=0}^{p-|\tau|} \phi_{p,j+|\tau|} \phi_{p,j} .$$

Since a peak in $S(\cdot)$ corresponds to a valley in $D(\cdot)$, we can use the Newton–Raphson method described following Equation (528a) – with $\hat{s}_{\phi,\tau}^{(P)}$ substituted for $\hat{s}_{\tau}^{(P)}$ – to determine the valley location. This technique requires a good initial estimate, which we can usually obtain by evaluating $D(\cdot)$ on a fine enough grid of frequencies via a DFT of $\phi_{p,0}, \phi_{p,1}, \ldots, \phi_{p,p}$ padded with lots of zeros (for details, see the discussion surrounding Equation (449b); in creating Figure 557, padding with $256 - 17 = 239$ zeros sufficed, but $128 - 17 = 111$ did not). As we noted previously in the discussion following Equation (528a), the Newton–Raphson method can fail if the initial estimate is poor; in these cases, we can use a bisection technique (combined with Newton–Raphson) or Brent's method if the location can be bracketed. Since we are now searching for a minimum rather than a maximum, we must modify the bracketing conditions to be $g'(\omega^{(L)}) < 0$ and $g'(\omega^{(U)}) > 0$.

10.13 Problems with the Parametric Approach

For a process consisting of a sinusoid in additive noise, two basic problems have been identified with the AR spectral estimates. Chen and Stegen (1974) found the location of the peak in the estimate to depend on the phase of the sinusoid, while Fougere et al. (1976) found that the estimate can sometimes contain two adjacent peaks in situations where only a single peak should appear – this is known in the literature as *spectral line splitting* or *spontaneous line splitting*. Causes and cures for these problems have been the subject of many papers – see Kay and Marple (1981, p. 1396) for a nice summary. Here we give a perspective on the historic literature.

Toman (1965) looked at the effect on the periodogram of short truncation lengths T of the *continuous* parameter deterministic sinusoid $X(t) = D\cos(2\pi f_0 t - \pi/2) = D\sin(2\pi f_0 t)$, where $f_0 = 1$ Hz so that the corresponding period is 1 sec. The data segment is taken to be from $t = 0$ to $t = T$, and the periodogram is here defined as

$$\frac{1}{T} \left| \int_0^T X(t) e^{-i2\pi f t} \, dt \right|^2 .$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

Frequency (Hz)	Phase	$f_{\mathcal{N}}$	N	p	Peak Shift	Line Splitting
Ulrych, 1972a:						
1	$-\pi/2$	10	21	11	no	no
1	0	10	21	11	no	no
1	$-\pi/2$	10	12	11	no	no
Chen and Stegen, 1974:						
1	$-\pi/2$	10	24	2,8	no	no
1	$-\pi/2$	10	24	20	yes	yes
1	$-\pi/2$	10	10–60	8	oscillates	—
1	varied	10	15	8	oscillates	—
1	varied	10	12–65	8	oscillates	—
Fougere et al., 1976:						
1	$-\pi/2\,(\pi/9)\,\pi/2$	10	21	19	no	no
5	$-\pi/2\,(\pi/90)\,\pi/2$	10	6	5	yes	yes
$1.25\,(2)\,49.25$	$-\pi/4$	50	101	24	yes (all)	yes (all)

Table 559 Summary of simulation results for a single sinusoid with additive white noise (estimation by Burg's algorithm). The first column gives the frequency of the sinusoid; the second column is phase (with respect to a cosine term); $f_{\mathcal{N}}$ is the Nyquist frequency; N is the sample size; and p is the order of the autoregressive model. The notation $x\,(a)\,y$ means from x to y in steps of a.

Note that, because the period is unity, the ratio of the length of the data segment to the period is just T. Toman demonstrated that, for $0 \leq T \leq 0.58$, the location of the peak value of the periodogram is at $f = 0$; as T increases from 0.58 to 0.716, the location of the peak value of the periodogram increases from 0 to 1 Hz (the latter being the true sinusoidal frequency f_0); for T greater than 0.716, the peak frequency oscillates, giving 1 Hz at the maximum of each oscillation and converging to 1 Hz as $T \to \infty$. For the special case $T = 1$ corresponding to one complete cycle of the sinusoid, the peak occurred at 0.84 Hz.

Jackson (1967) clearly and concisely explained Toman's experimental results. He noted that, with $X(t) = D\cos(2\pi f_0 t + \phi)$,

$$\int_0^T X(t)\mathrm{e}^{-\mathrm{i}2\pi f t}\,\mathrm{d}t = \frac{D}{2}\mathrm{e}^{\mathrm{i}\phi}\mathrm{e}^{-\mathrm{i}\pi(f-f_0)T}\frac{\sin(\pi(f-f_0)T)}{\pi(f-f_0)} + \frac{D}{2}\mathrm{e}^{-\mathrm{i}\phi}\mathrm{e}^{-\mathrm{i}\pi(f+f_0)T}\frac{\sin(\pi(f+f_0)T)}{\pi(f+f_0)}.$$

For the special case considered by Toman, namely, $\phi = -\pi/2$, $f_0 = 1$ and $T = 1$, the above reduces to

$$\mathrm{i}\frac{D}{2}\mathrm{e}^{-\mathrm{i}\pi f}\left(\mathrm{sinc}\,(f-1) - \mathrm{sinc}\,(f+1)\right),$$

where, as usual, $\mathrm{sinc}\,(t) = \sin(\pi t)/(\pi t)$ is the sinc function. The periodogram is thus proportional to $\left|\mathrm{sinc}\,(f-1) - \mathrm{sinc}\,(f+1)\right|^2$. Jackson noted that the interference of the two sinc functions produces extrema at $\pm 0.84f$, as observed by Toman. For a single cycle of a cosine, i.e., $\phi = 0$, the maxima occur at $\pm 1.12f$. Jackson similarly explained Toman's other observations. His explanation is obviously in agreement with our example of interference from two Fejér's kernels (see the discussion concerning Equation (536)).

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

Ulrych (1972a) – see also Ulrych (1972b) – thought that such spectral shifts could be prevented by the use of Burg's algorithm. In his example – see Table 559 – neither shifting nor splitting was present even with additive white noise. Chen and Stegen (1974) carried out a more detailed investigation of Burg's algorithm using as input $1\,\text{Hz}$ sinusoids sampled at 20 samples per second with white noise superimposed. Table 559 summarizes some of their results. The particular parameter combinations used by Ulrych (1972a) happen to coincide with points in the oscillating location of the spectral peak that agree with the true value of $1\,\text{Hz}$; in general, however, this is not the case, so that Burg's method is indeed susceptible to these phenomena.

Fougere et al. (1976) noted that spectral line splitting is most likely to occur when

[1] the signal-to-noise ratio is high;
[2] the initial phase of sinusoidal components is some odd multiple of $45°$; and
[3] the time duration of the data sequence is such that the sinusoidal components have an odd number of quarter cycles.

Other historic papers of interest are Fougere (1977, 1985). Table 559 gives a summary of some of Fougere's results. In the $1\,\text{Hz}$ example, an AR(19) model with 21 data points produced no anomalous results for a range of input phases; on the other hand, an AR(24) model with 101 data points produced line splitting for a range of frequencies for the input sinusoids. Fougere (1977) concluded that "... splitting does not occur as a result of using an overly large number of filter weights [AR coefficients], as is frequently claimed in the literature." Chen and Stegen (1974) had reached the *opposite* conclusion based on the second of their examples given in Table 559.

Chen and Stegen (1974) considered that the frequency shifts they observed were due to essentially the same mechanism as observed by Jackson (1967). However, Fougere (1977) notes that "Jackson's worst cases were sine waves an even number of cycles long with either $90°$ or $0°$ initial phase." For these two cases Burg's spectra are neither split nor shifted but are extremely accurate.

Fougere (1977) put the problem with line splitting in Burg's algorithm down to the substitution of estimated $\phi_{k,k}$ from lower order fitted AR models into the higher order fitted models (as we have seen, the forward/backward sums of squares from the final model order are not necessarily minimized using Burg's algorithm due to the substitution from lower order fits). Fougere suggested setting $\phi_{k,k} = U \sin(\Phi_k)$ for $k = 1, 2, \ldots, p$, with U slightly less than 1, and solving for the Φ_k terms *simultaneously* using a nonlinear optimization scheme. Once Φ_k has been estimated by $\hat{\Phi}_k$, the corresponding estimates for $\phi_{k,k}$, namely, $\hat{\phi}_{k,k} = U \sin(\hat{\Phi}_k)$, must be less than unity as required for stability. Fougere demonstrated that, at least in the single sinusoid examples examined, line splitting could be eliminated by using this procedure. (Bell and Percival, 1991, obtained similar results with a "two-step" Burg algorithm, which estimates the reflection coefficients in pairs and hence can be regarded as a compromise between Burg's algorithm and Fougere's method.)

Kay and Marple (1979) diagnosed both a different cause and cure from the suggestions of Fougere. They considered that

> ... spectral line splitting is a result of estimation errors and is not inherent in the autoregressive approach. In particular, the interaction between positive and negative sinusoidal frequency components [Jackson's mechanism] in the Burg reflection coefficient and Yule–Walker autocorrelation estimates and the use of the biased estimator in the Yule–Walker approach are responsible for spectral line splitting.

Their cure (for one sinusoid) was the use of *complex-valued data* and the *unbiased* ACVS estimator in the Yule–Walker case. To follow their argument, consider a noise-free randomly

phased sinusoid with unit amplitude, $X_t = \cos(2\pi f t\,\Delta_t + \phi)$, and take the time series to be $X_0, X_1, \ldots, X_{N-1}$. The biased estimator of the ACVS is

$$\hat{s}_\tau^{(\mathrm{P})} = \frac{1}{N} \sum_{t=0}^{N-|\tau|-1} X_{t+|\tau|} X_t$$

since $E\{X_t\} = 0$ under the random phase assumption (this is Equation (170b)). Hence

$$\begin{aligned}\hat{s}_\tau^{(\mathrm{P})} &= \frac{1}{N} \sum_{t=0}^{N-|\tau|-1} \cos([2\pi f t\,\Delta_t + \phi] + 2\pi f|\tau|\,\Delta_t) \cos(2\pi f t\,\Delta_t + \phi) \\ &= \frac{N-|\tau|}{2N} \cos(2\pi f \tau\,\Delta_t) + \cos(2\pi[N-1] f\,\Delta_t + 2\phi)\, \frac{\sin(2\pi[N-|\tau|] f\,\Delta_t)}{2N \sin(2\pi f\,\Delta_t)} \quad (561a)\end{aligned}$$

(the proof of this result is Exercise [10.22]). This can be written in terms of the true ACVS $s_\tau = \cos(2\pi f \tau\,\Delta_t)/2$ as

$$\hat{s}_\tau^{(\mathrm{P})} = \frac{N-|\tau|}{N} s_\tau + \cos(2\pi[N-1] f\,\Delta_t + 2\phi)\, \frac{\sin(2\pi[N-|\tau|] f\,\Delta_t)}{2N \sin(2\pi f\,\Delta_t)}; \quad (561b)$$

i.e., $\hat{s}_\tau^{(\mathrm{P})}$ is composed of one term that is a biased estimator of s_τ and a second term that is phase dependent. Both these terms would contribute to an inaccurate spectral estimate using the Yule–Walker equations based on $\{\hat{s}_\tau^{(\mathrm{P})}\}$.

Suppose now that N is large enough that the phase dependent term is close to zero. Then

$$\hat{s}_\tau^{(\mathrm{P})} \approx \frac{N-|\tau|}{N} s_\tau = \left(1 - \frac{|\tau|}{N}\right) s_\tau. \quad (561c)$$

Kay and Marple (1979) argue that the estimator $\hat{s}_\tau^{(\mathrm{P})}$ "... corresponds more nearly to the ACVS of two sinusoids which beat together to approximate a linear tapering." Since an AR(p) spectral estimate fitted by the Yule–Walker method has an ACVS that is *identical* with the sample ACVS up to lag p, the spectral estimate should have two – rather than only one – spectral peaks. As a concrete example, they consider the case $f = 1/(4\,\Delta_t) = f_{\mathcal{N}}/2$ and $\phi = 0$ when N is odd, for which

$$\hat{s}_\tau^{(\mathrm{P})} = \left(1 - \frac{|\tau| - 1}{N}\right) s_\tau. \quad (561d)$$

Setting $N = 9$, line splitting occurs when $p = 8$ (verification of these results is part of Exercise [10.23]). Note that the model used by Kay and Marple is a noise-free sinusoid. This is reasonable for investigating line splitting, since the phenomenon has been seen to occur at *high* signal-to-noise ratios.

Kay and Marple (1979) argue that the phase-dependent term of Equation (561a), i.e., the term involving $\cos(2\pi[N-1] f\,\Delta_t + 2\phi)$, can be regarded as the interaction between complex exponentials since it can be written as

$$\frac{1}{4N} \sum_{t=0}^{N-|\tau|-1} \left(\mathrm{e}^{\mathrm{i}(2\pi f[2t+|\tau|]\,\Delta_t + 2\phi)} + \mathrm{e}^{-\mathrm{i}(2\pi f[2t+|\tau|]\,\Delta_t + 2\phi)} \right)$$

(the above falls out as part of the proof of Equation (561a) that Exercise [10.22] calls for). This fact led them to consider the *analytic series* associated with $\{X_t\}$. By definition this

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

series is $X_t + \mathrm{i}\mathcal{HT}\{X_t\}$, where $\mathcal{HT}$ is the *Hilbert transform* of $\{X_t\}$ (Papoulis and Pillai, 2002, section 9–3). For our purposes, we can regard the Hilbert transform of a series as a phase-shifted version of the series (the shift is $\pi/2$ for $f < 0$ and $-\pi/2$ for $f > 0$); thus, if $X_t = \cos(2\pi f t\,\Delta_{\mathrm{t}} + \phi)$, then $\mathcal{HT}\{X_t\} = \sin(2\pi f t\,\Delta_{\mathrm{t}} + \phi)$ if $f > 0$ and $\mathcal{HT}\{X_t\} = -\sin(2\pi f t\,\Delta_{\mathrm{t}} + \phi)$ if $f < 0$. Assuming $f > 0$ in our example, the analytic series is thus

$$\cos(2\pi f t\,\Delta_{\mathrm{t}} + \phi) + \mathrm{i}\sin(2\pi f t\,\Delta_{\mathrm{t}} + \phi) = \mathrm{e}^{\mathrm{i}(2\pi f t\,\Delta_{\mathrm{t}} + \phi)} \stackrel{\mathrm{def}}{=} Z_t,$$

say.

▷ **Exercise [562]** Show that, for $0 \le \tau \le N-1$,

$$\hat{s}^{(\mathrm{P})}_{Z,\tau} \stackrel{\mathrm{def}}{=} \frac{1}{N}\sum_{t=0}^{N-\tau-1} Z_{t+\tau}Z_t^* = \frac{N-\tau}{N}\mathrm{e}^{\mathrm{i}2\pi f\tau\,\Delta_{\mathrm{t}}}. \qquad \triangleleft$$

The above is the biased ACVS estimator appropriate for zero-mean complex-valued data (see Equation (231)). The corresponding unbiased estimator is

$$\hat{s}^{(\mathrm{U})}_{Z,\tau} = \frac{1}{N-\tau}\sum_{t=0}^{N-\tau-1} Z_{t+\tau}Z_t^* = \frac{N}{N-\tau}\hat{s}^{(\mathrm{P})}_{Z,\tau} = \mathrm{e}^{\mathrm{i}2\pi f\tau\,\Delta_{\mathrm{t}}}. \tag{562}$$

This is exactly the ACVS of a complex exponential with unit amplitude (cf. Equation (519j)). Since this estimate equals the exact or theoretical ACVS, the fact that the AR(p) process fitted by the Yule–Walker method has a theoretical ACVS that is identical with the estimated ACVS up to lag p will prevent the possibility of line splitting. Thus, for the Yule–Walker method, Kay and Marple (1979) recommend using complex-valued data (i.e., the analytic series) and the unbiased ACVS estimator; however, for time series other than the one considered here, this recommendation is tempered by the fact that, as opposed to $\{\hat{s}^{(\mathrm{P})}_{Z,\tau}\}$, the sequence $\{\hat{s}^{(\mathrm{U})}_{Z,\tau}\}$ in general need *not* correspond to a valid ACVS for a complex-valued process and hence might not be amenable for use with the Yule–Walker method.

Let us now turn to the case of Burg's algorithm. For the analytic series $\{Z_t\}$, we make use of the Burg estimate of $\phi_{1,1}$ appropriate for complex-valued series (see Section 9.13):

$$\bar{\phi}_{1,1} = \frac{2\sum_{t=0}^{N-2} Z_{t+1}Z_t^*}{|Z_0|^2 + 2\sum_{t=1}^{N-2}|Z_t|^2 + |Z_{N-1}|^2} = \frac{2(N-1)\mathrm{e}^{\mathrm{i}2\pi f\,\Delta_{\mathrm{t}}}}{2(N-1)} = \mathrm{e}^{\mathrm{i}2\pi f\,\Delta_{\mathrm{t}}},$$

where we have used the result in Equation (562) and the fact that $|Z_t|^2 = 1$ for all t. Kay and Marple (1979) argue that the estimates of the higher-order reflection coefficients, $\bar{\phi}_{2,2}$, $\bar{\phi}_{3,3}$, etc., are all zero. The ACVS estimator corresponding to Burg's algorithm is thus $\bar{\phi}^{\tau}_{1,1} = \exp(\mathrm{i}2\pi f\tau\,\Delta_{\mathrm{t}})$ for $\tau \ge 0$, again identical to the theoretical ACVS. Since the implied and theoretical ACVSs are identical, line splitting in Burg's algorithm is also arguably cured by using complex exponentials, i.e., by making the series "analytic."

Kay and Marple (1981) note that, for *multiple* sinusoids, the performance of Fougere's method is undocumented and that the use of complex-valued data in conjunction with Burg's algorithm can still result in line splitting. However, there is no evidence of line splitting in the FBLS method discussed in Section 9.7.

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

10.14 Singular Value Decomposition Approach

We stated in Equation (553d) that the summation of p *real-valued* sinusoids can be represented by a real-valued pseudo-AR($2p$) equation in which the parameters $\varphi_{2p,k}$ are the coefficients of the polynomial equation stated in Equation (553c) with roots occurring in pairs of the form $\exp(\pm \mathrm{i}2\pi f_j \Delta_t)$ for $j = 1, \ldots, p$. As might be anticipated, for the summation of p *complex-valued* sinusoids, i.e., complex exponentials,

$$\tilde{Z}_t = \sum_{l=1}^{p} D_l \mathrm{e}^{\mathrm{i}(2\pi f_l t\, \Delta_t + \phi_l)},$$

there is a corresponding representation in terms of a complex-valued pseudo-AR(p) equation where the parameters $\varphi_{p,k}$ are the coefficients of the polynomial equation

$$1 - \sum_{k=1}^{p} \varphi_{p,k} z^{-k} = 0 \;\text{ or, equivalently, }\; z^p - \sum_{k=1}^{p} \varphi_{p,k} z^{p-k} = 0$$

with roots of the form $z_j = \exp(\mathrm{i}2\pi f_j \Delta_t)$ for $j = 1, \ldots, p$. To see that we do indeed have

$$\tilde{Z}_t = \sum_{k=1}^{p} \varphi_{p,k} \tilde{Z}_{t-k},$$

we first write

$$\tilde{Z}_t = \sum_{l=1}^{p} C_l \mathrm{e}^{\mathrm{i}2\pi f_l t\, \Delta_t} = \sum_{l=1}^{p} C_l z_l^t, \;\text{ where }\; C_l \stackrel{\text{def}}{=} D_l \mathrm{e}^{\mathrm{i}\phi_l}.$$

The relationship between the roots $\{z_j\}$ and coefficients $\{\varphi_{p,k}\}$ is given by

$$\prod_{j=1}^{p} (z - z_j) = z^p - \sum_{k=1}^{p} \varphi_{p,k} z^{p-k}. \tag{563}$$

Now $\tilde{Z}_{t-k} = \sum_{l=1}^{p} C_l z_l^{t-k}$, and hence

$$\sum_{k=1}^{p} \varphi_{p,k} \tilde{Z}_{t-k} = \sum_{k=1}^{p} \varphi_{p,k} \left(\sum_{l=1}^{p} C_l z_l^{t-k} \right) = \sum_{l=1}^{p} C_l z_l^{t-p} \sum_{k=1}^{p} \varphi_{p,k} z_l^{p-k}.$$

However, because z_l is a root, we have

$$\prod_{j=1}^{p} (z_l - z_j) = z_l^p - \sum_{k=1}^{p} \varphi_{p,k} z_l^{p-k} = 0, \;\text{ i.e., }\; \sum_{k=1}^{p} \varphi_{p,k} z_l^{p-k} = z_l^p,$$

so we can now write

$$\sum_{k=1}^{p} \varphi_{p,k} \tilde{Z}_{t-k} = \sum_{l=1}^{p} C_l z_l^{t-p} z_l^p = \sum_{l=1}^{p} C_l z_l^t = \tilde{Z}_t$$

as required.

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

We now make use of this pseudo-AR(p) representation for the summation of p complex exponentials to design an estimation scheme for the frequencies f_l for a model with additive proper complex-valued white noise $\{\epsilon_t\}$ with zero mean; i.e.,

$$Z_t = \tilde{Z}_t + \epsilon_t = \sum_{l=1}^{p} C_l e^{i2\pi f_l t \,\Delta_t} + \epsilon_t = \sum_{k=1}^{p} \varphi_{p,k} \tilde{Z}_{t-k} + \epsilon_t$$

(cf. Equation (517a) with $\mu = 0$). To determine the f_l's, Tufts and Kumaresan (1982) consider the following model involving a predictive filter of order p' (with $p' > p$ and p assumed unknown) in both the forward and backward directions:

$$Z_t = \sum_{k=1}^{p'} \varphi_{p',k} Z_{t-k} \quad \text{and} \quad Z_t^* = \sum_{k=1}^{p'} \varphi_{p',k} Z_{t+k}^*,$$

which, given data $Z_0, Z_1, \ldots, Z_{N-1}$, yields the equations

$$\begin{aligned}
Z_{p'} &= \varphi_{p',1} Z_{p'-1} + \varphi_{p',2} Z_{p'-2} + \cdots + \varphi_{p',p'} Z_0 \\
&\vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \ddots \qquad\quad \vdots \\
Z_{N-1} &= \varphi_{p',1} Z_{N-2} + \varphi_{p',2} Z_{N-3} + \cdots + \varphi_{p',p'} Z_{N-p'-1} \\
Z_0^* &= \varphi_{p',1} Z_1^* + \varphi_{p',2} Z_2^* + \cdots + \varphi_{p',p'} Z_{p'}^* \\
&\vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \ddots \qquad\quad \vdots \\
Z_{N-p'-1}^* &= \varphi_{p',1} Z_{N-p'}^* + \varphi_{p',2} Z_{N-p'+1}^* + \cdots + \varphi_{p',p'} Z_{N-1}^*
\end{aligned}$$

(to see why complex conjugates are needed in the bottom $N - p'$ equations, consider the complex conjugate of Equation (502)). For the noiseless case when $Z_t = \tilde{Z}_t$, these equations would be exactly true if $p' = p$. For the moment we assume the noiseless case, but take $p' > p$. We can write the just-presented equations compactly as

$$\boldsymbol{A}\boldsymbol{\varphi} = \boldsymbol{Z}, \tag{564}$$

where $\boldsymbol{\varphi} \stackrel{\text{def}}{=} [\varphi_{p',1}, \ldots, \varphi_{p',p'}]^T$,

$$\boldsymbol{A} \stackrel{\text{def}}{=} \begin{bmatrix}
Z_{p'-1} & Z_{p'-2} & \cdots & Z_0 \\
\vdots & \vdots & \ddots & \vdots \\
Z_{N-2} & Z_{N-3} & \cdots & Z_{N-p'-1} \\
Z_1^* & Z_2^* & \cdots & Z_{p'}^* \\
\vdots & \vdots & \ddots & \vdots \\
Z_{N-p'}^* & Z_{N-p'+1}^* & \cdots & Z_{N-1}^*
\end{bmatrix} \quad \text{and} \quad \boldsymbol{Z} \stackrel{\text{def}}{=} \begin{bmatrix}
Z_{p'} \\ \vdots \\ Z_{N-1} \\ Z_0^* \\ \vdots \\ Z_{N-p'-1}^*
\end{bmatrix}$$

(note that $\boldsymbol{A}$ is a $2(N - p') \times p'$ matrix).

As is true for any matrix, we can express $\boldsymbol{A}$ in terms of its singular value decomposition (see, for example, Golub and Kahan, 1965, or Businger and Golub, 1969; for real-valued matrices see Jackson, 1972, or Golub and Van Loan, 2013, section 2.4). Two sets of orthonormal vectors $\{\boldsymbol{u}_j\}$ and $\{\boldsymbol{v}_j\}$ exist such that

$$\boldsymbol{A}^H \boldsymbol{u}_j = \gamma_j \boldsymbol{v}_j \quad \text{and} \quad \boldsymbol{A}\boldsymbol{v}_j = \lambda_j \boldsymbol{u}_j,$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

from which it follows that

$$\begin{aligned} \boldsymbol{A}\boldsymbol{A}^H\boldsymbol{u}_j &= \gamma_j\lambda_j\boldsymbol{u}_j, \qquad j = 1,\ldots,2(N-p'), \\ \boldsymbol{A}^H\boldsymbol{A}\boldsymbol{v}_j &= \gamma_j\lambda_j\boldsymbol{v}_j, \qquad j = 1,\ldots,p', \end{aligned} \tag{565a}$$

(see Exercise [10.24]). Here the superscript H denotes the operation of complex-conjugate transposition. Each $\boldsymbol{u}_j$ is an eigenvector of $\boldsymbol{A}\boldsymbol{A}^H$, while $\boldsymbol{v}_j$ is an eigenvector of $\boldsymbol{A}^H\boldsymbol{A}$. If we rank the γ_j and λ_j terms each in decreasing order of magnitude, it can be shown that, for some integer $r \le \min(2(N-p'), p')$,

$$\gamma_j = \begin{cases} \lambda_j > 0, & j \le r; \\ 0, & \text{otherwise;} \end{cases} \quad \text{and} \quad \lambda_j = 0 \ \text{ for } \ j > r;$$

i.e., there are r eigenvalues $\gamma_j\lambda_j = \lambda_j^2 > 0$ common to the two systems of equations in Equation (565a), and all the others are zero. The vectors $\boldsymbol{u}_j$ and $\boldsymbol{v}_j$ for $j = 1,\ldots,r$ are called the *left and right eigenvectors* or *singular vectors* of the matrix $\boldsymbol{A}$; the λ_j terms are called the *eigenvalues* or *singular values* of $\boldsymbol{A}$. The integer r is the rank of $\boldsymbol{A}$ (its importance becomes clearer when we look at the two cases of zero and nonzero additive noise). The matrix $\boldsymbol{A}$ can be written as

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^H, \tag{565b}$$

where $\boldsymbol{U}$ is a $2(N-p') \times r$ matrix whose columns are the eigenvectors $\boldsymbol{u}_j$; $\boldsymbol{V}$ is a $p' \times r$ matrix whose columns are the eigenvectors $\boldsymbol{v}_j$; and $\boldsymbol{\Lambda}$ is a diagonal matrix whose diagonal elements are the eigenvalues λ_j. The right-hand side of the above is called the *singular value decomposition* (SVD) of $\boldsymbol{A}$. By the orthonormality of the eigenvectors,

$$\boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{V}^H\boldsymbol{V} = \boldsymbol{I}_r, \ \ \boldsymbol{U}\boldsymbol{U}^H = \boldsymbol{I}_r \ \text{ if } \ r = 2(N-p') \ \text{ and } \ \boldsymbol{V}\boldsymbol{V}^H = \boldsymbol{I}_r \ \text{ if } \ r = p', \tag{565c}$$

where $\boldsymbol{I}_r$ is the $r \times r$ identity matrix (Jackson, 1972).

Let us now return to Equation (564) and premultiply both sides of this equation by $\boldsymbol{A}^H$ to obtain $\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{\varphi} = \boldsymbol{A}^H\boldsymbol{Z}$. If we let $\boldsymbol{R} = \boldsymbol{A}^H\boldsymbol{A}$ and $\boldsymbol{\beta} = \boldsymbol{A}^H\boldsymbol{Z}$, then we have

$$\boldsymbol{R}\boldsymbol{\varphi} = \boldsymbol{\beta},$$

where the (j,k)th element of $\boldsymbol{R}$ and the jth element of $\boldsymbol{\beta}$ are given by, respectively,

$$\sum_{t=p'}^{N-1} Z_{t-j}^* Z_{t-k} + \sum_{t=0}^{N-p'-1} Z_{t+j} Z_{t+k}^* \ \text{ and } \ \sum_{t=p'}^{N-1} Z_{t-j}^* Z_t + \sum_{t=0}^{N-p'-1} Z_{t+j} Z_t^*.$$

The system $\boldsymbol{R}\boldsymbol{\varphi} = \boldsymbol{\beta}$ is the least squares system of equations corresponding to minimizing the forward/backward sum of squares

$$\sum_{t=p'}^{N-1} \left| Z_t - \sum_{k=1}^{p'} \varphi_{p',k} Z_{t-k} \right|^2 + \sum_{t=0}^{N-p'-1} \left| Z_t^* - \sum_{k=1}^{p'} \varphi_{p',k} Z_{t+k}^* \right|^2.$$

If we use the singular value decomposition given in Equation (565b) and recall that $\boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{I}_r$ (see Equation (565c)), we can write

$$\boldsymbol{R} = \boldsymbol{A}^H\boldsymbol{A} = \left(\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{U}^H\right)\left(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^H\right) = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{I}_r\boldsymbol{\Lambda}\boldsymbol{V}^H = \boldsymbol{V}\boldsymbol{\Lambda}^2\boldsymbol{V}^H,$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

where $\boldsymbol{\Lambda}^n$ refers to the diagonal matrix with diagonal elements λ_j^n. Also, $\boldsymbol{\beta}$ can be written

$$\boldsymbol{\beta} = \boldsymbol{A}^H \boldsymbol{Z} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{U}^H \boldsymbol{Z},$$

so that $\boldsymbol{R}\boldsymbol{\varphi} = \boldsymbol{\beta}$ becomes

$$\boldsymbol{V}\boldsymbol{\Lambda}^2\boldsymbol{V}^H\boldsymbol{\varphi} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{U}^H\boldsymbol{Z}. \tag{566a}$$

The *generalized inverse* of $\boldsymbol{R} = \boldsymbol{V}\boldsymbol{\Lambda}^2\boldsymbol{V}^H$ is given by

$$\boldsymbol{R}^{\#} = \boldsymbol{V}\boldsymbol{\Lambda}^{-2}\boldsymbol{V}^H.$$

This generalized inverse always exists and gives the solution (using $\boldsymbol{V}^H\boldsymbol{V} = \boldsymbol{I}_r$ from Equation (565c))

$$\tilde{\boldsymbol{\varphi}} = \left(\boldsymbol{V}\boldsymbol{\Lambda}^{-2}\boldsymbol{V}^H\right)\left(\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{U}^H\right)\boldsymbol{Z} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{U}^H\boldsymbol{Z}. \tag{566b}$$

That this is indeed a solution to Equation (566a) can be checked by noting that

$$\boldsymbol{V}\boldsymbol{\Lambda}^2\boldsymbol{V}^H\tilde{\boldsymbol{\varphi}} = \boldsymbol{V}\boldsymbol{\Lambda}^2\boldsymbol{V}^H\boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{U}^H\boldsymbol{Z} = \boldsymbol{V}\boldsymbol{\Lambda}^2\boldsymbol{I}_r\boldsymbol{\Lambda}^{-1}\boldsymbol{U}^H\boldsymbol{Z} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{U}^H\boldsymbol{Z}$$

(again using $\boldsymbol{V}^H\boldsymbol{V} = \boldsymbol{I}_r$). This solution minimizes $\tilde{\boldsymbol{\varphi}}^H\tilde{\boldsymbol{\varphi}} = |\tilde{\boldsymbol{\varphi}}|^2$ (see, for example, Jackson, 1972). It is interesting to note that, if $r = p'$, then $\boldsymbol{R}\boldsymbol{R}^{\#} = \boldsymbol{R}^{\#}\boldsymbol{R} = \boldsymbol{V}\boldsymbol{V}^H = \boldsymbol{I}_{p'}$ from Equation (565c), and in this case Equation (566b) expresses the usual least squares solution for $\boldsymbol{\varphi}$.

When the additive noise $\{\epsilon_t\}$ is zero for all t, the process $\{Z_t\}$ is just the sum of p complex exponentials, and $r = p < p'$. From Equation (566b), we have

$$\tilde{\boldsymbol{\varphi}} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p]\begin{bmatrix}\lambda_1^{-1} & & & \\ & \lambda_2^{-1} & & \\ & & \ddots & \\ & & & \lambda_p^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{u}_1^H \\ \vdots \\ \vdots \\ \boldsymbol{u}_p^H\end{bmatrix}\boldsymbol{Z} = \left(\sum_{k=1}^{p}\frac{1}{\lambda_k}\boldsymbol{v}_k\boldsymbol{u}_k^H\right)\boldsymbol{Z} \tag{566c}$$

(Tufts and Kumaresan, 1982). To ensure that the rank of $\boldsymbol{A}$ is $r = p$, Hua and Sarkar (1988) point out that we should choose p' such that $p \le p' \le N - p$. With $\tilde{\boldsymbol{\varphi}}$ determined from Equation (566c), we form the corresponding polynomial equation, namely,

$$z^{p'} - \sum_{k=1}^{p'}\tilde{\varphi}_{p',k}z^{p'-k} = 0.$$

Since $r = p$ this expression has p roots z_j on the unit circle at the locations of the sinusoidal frequencies, and the other $p' - p$ roots lie inside the unit circle (Tufts and Kumaresan, 1982). The frequency f_j is found by inverting $z_j = \exp(\mathrm{i}2\pi f_j \Delta_t)$; i.e., $f_j = \Im(\log(z_j))/(2\pi\Delta_t)$, where $\Im(z)$ is the imaginary part of the complex number z, and we take the solution such that $|f_j| \le f_{\mathcal{N}}$.

Of course, a much more realistic situation is when the additive noise $\{\epsilon_t\}$ is nonzero. In this case $\boldsymbol{A}$ has full rank, i.e., $r = \min(2(N - p'), p')$, and, by a derivation identical to that used for Equation (566c), the solution using the generalized inverse is

$$\hat{\boldsymbol{\varphi}} = \left(\sum_{k=1}^{p'}\frac{1}{\hat{\lambda}_k}\hat{\boldsymbol{v}}_k\hat{\boldsymbol{u}}_k^H\right)\boldsymbol{Z}. \tag{566d}$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

Since the matrix $\boldsymbol{A}$ and vector $\boldsymbol{Z}$ of Equation (564) here include additive noise, the left and right eigenvectors and eigenvalues are in general different from those in Equation (566c), so we have added a hat to these quantities to emphasize this. The eigenvalues $\hat{\lambda}_{p+1}, \ldots, \hat{\lambda}_{p'}$ are due to the noise and will usually be small. The solution $\hat{\boldsymbol{\varphi}}$ thus becomes dominated by the noise components unless the sum in Equation (566d) is truncated at p, so we define

$$\hat{\boldsymbol{\varphi}}' = \left(\sum_{k=1}^{p} \frac{1}{\hat{\lambda}_k} \hat{\boldsymbol{v}}_k \hat{\boldsymbol{u}}_k^H \right) \boldsymbol{Z},$$

for which again the sinusoidal terms are the dominant influence. If prior knowledge of the exact number of sinusoids p is lacking, it might be possible to ascertain the number by looking for the point k in the eigenvalue sequence

$$\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \hat{\lambda}_k \geq \hat{\lambda}_{k+1} \geq \cdots \geq \hat{\lambda}_{p'}$$

at which the value is large before rapidly decreasing at $k+1$ (sometimes such a cutoff is clear, but often it is not). The polynomial equation

$$z^{p'} - \sum_{k=1}^{p'} \tilde{\varphi}'_{p',k} z^{p'-k} = 0$$

has in general p roots z'_k close to the unit circle, and these can be used to find the frequencies $f'_k = \Im(\log(z'_k))/(2\pi\,\Delta_t)$.

Comments and Extensions to Section 10.14

[1] Tufts and Kumaresan (1982) consider that the SVD approach is likely to be most useful when only a short time series is available and the sinusoids are more closely spaced than the reciprocal of the observation time.

[2] If minimization of the forward/backward sum of squares alone is used for the estimation of $\boldsymbol{\varphi}$, large values of p' can result in spurious spectral peaks. By using the SVD, the length of p' can be made considerably greater because of the effective gain in signal-to-noise ratio, thus increasing the resolution of the sinusoids. Tufts and Kumaresan found a good empirical choice to be $p' = 3N/4$. Since this implies that p' complex roots must be determined, it is clear that the method is only practicable for short series, as indeed was intended.

10.15 Examples of Harmonic Analysis

Willamette River Data

As a simple example of a harmonic analysis, let us consider a time series related to the flow of the Willamette River at Salem, Oregon (we used the first 128 values of this series as an example in Chapter 1). Figure 568a shows a plot of the $N = 395$ data points in the time series. Each point represents the log of the average daily water flow over a one month period from October, 1950, to August, 1983. The sampling interval is $\Delta_t = 1/12$ year, which yields a Nyquist frequency $f_{\mathcal{N}}$ of 6 cycles per year. There is a marked cyclical behavior of the data with a period of about 1 year. A harmonic process with one or more sinusoidal terms thus might be a reasonable model.

We start by entertaining the model of sinusoids plus white noise given by Equation (512b) and reformulated as Equation (515b). To handle the process mean μ, we center the time series by subtracting off its sample mean of 9.825 and consider Equation (515b) with $\mu = 0$ as a potential model for the centered series. To fit this model, we need first to determine

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

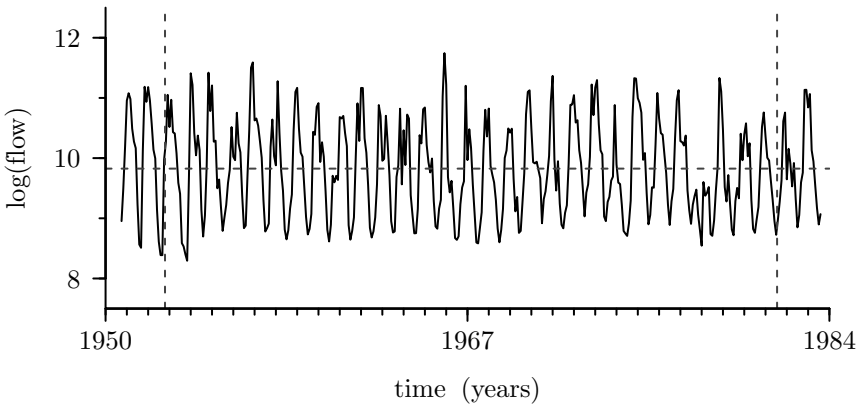

Figure 568a Willamette River data. There are $N = 395$ samples taken $\Delta_t = 1/12$ year apart (Figure 2(c) shows the first 128 values of this series). The horizontal dashed line indicates the sample mean for the series (9.825). The vertical dashed lines delineate a subseries of length 345 used in an SVD analysis.

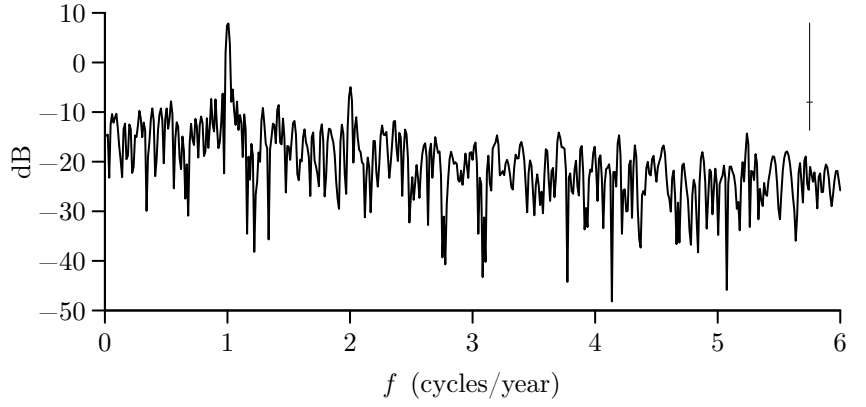

Figure 568b Periodogram for centered Willamette River data. Here the periodogram is plotted over a grid of frequencies more than twice as fine as the Fourier frequencies. The width of the crisscross in the upper right-hand corner gives the bandwidth measure $B_{\mathcal{H}}$ as per Equation (194), while its height gives the length of a 95% CI for $10 \log_{10}(S(f))$ as per Equation (205b).

the number of sinusoids L and the corresponding f_l terms. To do this we pad the centered data with 629 zeros to create a series of length 1024 and use an FFT routine to evaluate the periodogram $\hat{S}^{(\mathrm{P})}(\cdot)$ at the frequencies $j/(1024\,\Delta_t)$ for $j = 1, \ldots, 512$ (see C&E [1] for Section 3.11; we know from Exercise [6.15b] that $\hat{S}^{(\mathrm{P})}(0)$ is zero due to centering and hence the $j = 0$ term is of no interest). This grid of frequencies is more than twice as fine as that associated with the Fourier frequencies for $N = 395$, and ensures that we do not miss any important peaks in the periodogram. Figure 568b is a plot of $\hat{S}^{(\mathrm{P})}(\cdot)$ at these 512 frequencies. The largest peak – about 15 dB above the rest of the periodogram – occurs at $86/(1024\,\Delta_t) \doteq 1.00781$ cycles/year, which is the second closest frequency to 1 cycle/year of the form $j/(1024\,\Delta_t)$ because $85/(1024\,\Delta_t) \doteq 0.996$ cycle/year deviates only by 0.004 from 1 cycle/year.

The second largest peak in the periodogram is at $171/(1024\,\Delta_t) \doteq 2.00391$ cycles/year, which is the closest frequency to 2 cycles/year because $170/(1024\,\Delta_t) \doteq 1.992$ cycles/year. An explanation for the presence of this peak is that the river flow data have an annual variation

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

that is periodic but not purely sinusoidal. If we describe this variation by a real-valued periodic function $g_p(\cdot)$ with a period of 1 year, the results of Section 3.1 say that we can write

$$g_p(t) = \frac{a_0}{2} + \sum_{l=1}^{\infty} [a_l \cos(2\pi l t) + b_l \sin(2\pi l t)]$$

(cf. Equation (48a) with $T = 1$). The frequency that is associated with the lth cosine and sine in the summation is just $f'_l = l$ cycles/year. The lowest such frequency, namely $f'_1 = 1$ cycle/year, is called the *fundamental frequency* for $g_p(\cdot)$ and is equal to the reciprocal of its period. All the other frequencies that are involved in representing $g_p(\cdot)$ are multiples of the fundamental frequency; i.e., $f'_l = l f'_1 = l$ cycles/year. The frequency f'_{l+1} is called the lth harmonic of the fundamental frequency f'_1 so that, for example, $f'_2 = 2$ cycles/year is the first harmonic. If we now sample $g_p(\cdot)$ at points $\Delta_t = 1/12$ year apart, we obtain a periodic sequence given by

$$g_t \stackrel{\text{def}}{=} g_p(t\,\Delta_t) = \frac{a_0}{2} + \sum_{l=1}^{\infty} [a_l \cos(2\pi l t\,\Delta_t) + b_l \sin(2\pi l t\,\Delta_t)]\,. \tag{569a}$$

Because this choice of Δ_t produces an aliasing effect, the above can be rewritten as

$$g_t = \mu' + \sum_{l=1}^{6} [A'_l \cos(2\pi l t\,\Delta_t) + B'_l \sin(2\pi l t\,\Delta_t)]\,, \tag{569b}$$

where μ', A'_l and B'_l are related to a_l and b_l (see Exercise [10.25]). Note that the summation in the above equation resembles the nonrandom portion of Equation (515b) if we let $L = 6$ and equate f_l with $l = f'_l$. Given the nature of this time series, it is thus not unreasonable to find a peak in the periodogram at 2 cycles/year, the first harmonic for a fundamental frequency of 1 cycle/year (recall that we also encountered harmonics in the example from tidal analysis in Section 10.4). While A'_l and B'_l for $l = 3, 4, \ldots, 6$ need not be nonzero, the visual lack of peaks in the periodogram at $f = 3, 4, \ldots, 6$ suggests that they are not significantly different from zero.

Besides the peaks corresponding to annual and semiannual periods, there are numerous other lesser peaks in the periodogram that might or might not be due to other sinusoidal components. There is a possibility that the periodogram is masking components such as the second or higher harmonics of 1 cycle/year due to leakage from the two dominant peaks. Exercise [10.26] indicates that in fact this is not the case by considering a direct spectral estimate $\hat{S}^{(\mathrm{D})}(\cdot)$ based upon a Slepian data taper.

As a starting point, let us assume that there is only one frequency of importance in the time series so that the following simple harmonic model is appropriate for the centered data:

$$X_t = A_1 \cos(2\pi f_1 t\,\Delta_t) + B_1 \sin(2\pi f_1 t\,\Delta_t) + \epsilon_t. \tag{569c}$$

Although it would be reasonable to consider f_1 to be 1 cycle/year based on physical arguments, let us assume – for illustrative purposes – that it is unknown and use the periodogram to estimate it. If we search the periodogram in Figure 568b for its peak value (see C&E [2] for Section 10.6), we find that it occurs at $\hat{f}_1 \doteq 1.00319$ cycles/year. Conditional on this periodogram-based estimate of f_1, we can now estimate the parameters A_1 and B_1 using the ACLS estimators of Equation (549a) or the ECLS estimators, which minimize $\mathrm{SS}(A_1, B_1, \hat{f}_l)$ of Equation (549d). Alternatively we can find the EULS estimators A_1, B_1 and f_1 by minimizing $\mathrm{SS}(A_1, B_1, f_l)$. The upper part of Table 570 shows the resulting estimates, along with

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

Method	$\hat{f}_1$	$\hat{A}_1$	$\hat{B}_1$	Jump	$\hat{f}_2$	$\hat{A}_2$	$\hat{B}_2$	Jump	SS
ACLS	1.00319	−0.3033	0.8443	0.2012					87.40164
ECLS	1.00319	−0.3029	0.8447	0.2013					87.40157
EULS	1.00326	−0.3086	0.8427	0.2013					87.39881
ACLS	1.00319	−0.3033	0.8443	0.2012	2.00339	0.0365	0.1937	0.00971	79.58391
ECLS	1.00319	−0.3030	0.8451	0.2015	2.00339	0.0364	0.1956	0.00989	79.58305
EULS	1.00333	−0.3147	0.8411	0.2016	2.00244	0.0553	0.1915	0.00994	79.54593

Table 570 Parameter estimates of f_l, A_l and B_l for single- and two-frequency models of the form of Equation (512b) with $\mu = 0$ for the centered Willamette River time series, along with estimated jumps in the integrated spectrum at $\pm\hat{f}_l$ and the corresponding residual sum of squares (SS). Three types of least squares estimates are considered: approximate conditional (ACLS), exact conditional (ECLS) and exact unconditional (EULS). The top part of the table shows estimates for the single-frequency model; the bottom, for the model with two frequencies. The ACLS and ECLS estimates are conditional on estimates of f_l determined from the periodogram of the centered series.

the corresponding estimates of the jump in the integrated spectrum at $\hat{f}_1$ (computed using the left-hand side of Equation (549b)) and the corresponding residual sum of squares (we can use the latter to estimate the white noise variance σ_ϵ^2 via Equation (549c) for the ACLS and ECLS estimates or Equation (550a) in the EULS case). The estimates of A_1, B_1, f_1 and the jump in the integrated spectrum from all three approaches are in good agreement. Theory demands that the value of the SS associated with the ACLS method cannot be lower than that for the ECLS method; likewise, the ECLS value cannot be lower than the one for EULS. The values in the table adhere to this ordering. For comparison the sum of squares of the centered time series $\sum_t X_t^2$ is 246.3938. The coefficient of determination, namely,

$$1 - \frac{\mathrm{SS}(\hat{A}_1, \hat{B}_1, \hat{f}_l)}{\sum_t X_t^2} \doteq \begin{cases} 0.6452766, & \text{for ACLS;} \\ 0.6452769, & \text{for ECLS;} \\ 0.6452881, & \text{for EULS,} \end{cases}$$

is the proportion of variability in the time series explained by the sinusoidal terms in Equation (569c) and is the same for all three methods to four significant digits.

Figure 571a(a) is a plot of the fitted model for the uncentered river flow data based upon the EULS estimates using the centered series, i.e.,

$$\overline{X} + \hat{A}_1 \cos(2\pi \hat{f}_1 t\, \Delta_t) + \hat{B}_1 \sin(2\pi \hat{f}_1 t\, \Delta_t)$$

while Figure 571a(b) shows the corresponding EULS residuals:

$$R_t = X_t - \hat{A}_1 \cos(2\pi \hat{f}_1 t\, \Delta_t) - \hat{B}_1 \sin(2\pi \hat{f}_1 t\, \Delta_t)$$

(the sum of the series in (a) and (b) is equal to the original time series shown in Figure 568a; because the ACLS and ECLS parameter estimates are so similar to the EULS estimates, figures corresponding to Figure 571a for the ACLS and ECLS cases are visually identical to the EULS case – we concentrate on the EULS case in what follows and leave the ACLS and ECLS cases as exercises for the reader). If Equation (569c) is adequate as a model for the river flow data, the residuals should resemble a white noise process. To examine whether $\{R_t\}$ can be considered to be white noise, we plot its periodogram in Figure 571b. Note that the only substantial difference between this periodogram and that of $\{X_t\}$ in Figure 568b is near $f = 1$ cycle/year, as is reasonable to expect. The peak close to $f = 2$ cycles/year is the most

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

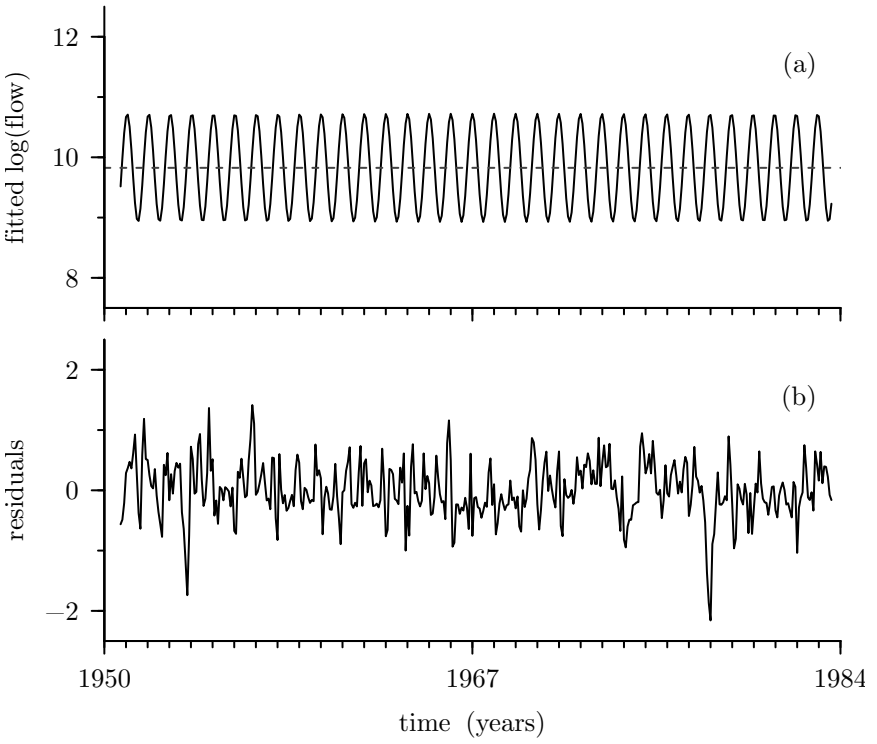

Figure 571a Mean-adjusted fitted single-frequency harmonic model for Willamette River data based on EULS parameter estimates (top plot), along with corresponding residuals (bottom).

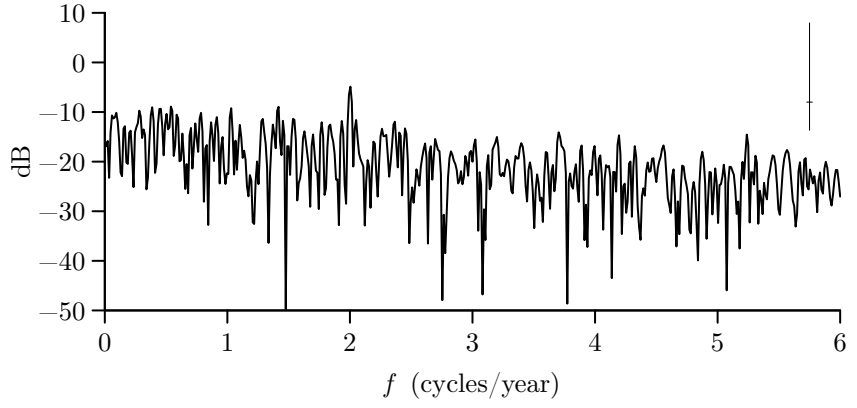

Figure 571b As in Figure 568b, but now the periodogram is based on the EULS residuals of Figure 571a(b).

prominent one in the periodogram of the residuals. Although this peak does occur close to a frequency that, given the nature of this series, is reasonable, it is only about 5 dB above the rest of the periodogram. Since we are in a situation where Fisher's g statistic (Equation (539b)) can help decide whether this peak is significant, we compute it for these residuals and obtain $g \doteq 0.088$. Under the null hypothesis of white noise, the upper 1% critical level for the test statistic is approximately 0.049 (from the approximation in Equation (540d), or using the entry for $m = 200$ in Table 1(b) of Shimshoni, 1971, which is reasonable since here

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

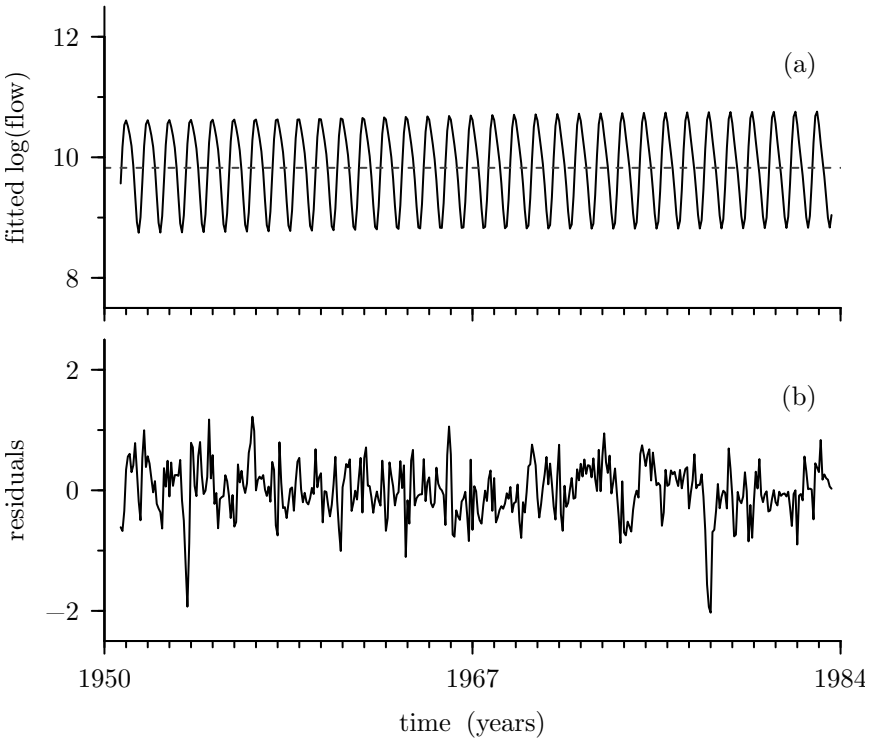

Figure 572 As in Figure 571a, but now for two-frequency harmonic model.

$m = (N-1)/2 = 197$). Since g exceeds this value, we have evidence that the peak is not just due to random variation. (There is, however, reason to use this statistic here with caution: the periodogram for $\{R_t\}$ decreases about 10 dB as f increases from 0 to 6 cycles/year, indicating that the presumed null hypothesis of white noise is questionable, a point we return to in our discussion of Figure 573.)

Let us now assume that there are two frequencies of importance so that our model for the centered time series becomes

$$X_t = \sum_{l=1}^{2} A_l \cos\left(2\pi f_l t\, \Delta_t\right) + B_l \sin\left(2\pi f_l t\, \Delta_t\right) + \epsilon_t. \tag{572}$$

The lower part of Table 570 shows that the ACLS, ECLS and EULS parameter estimates for this two-frequency model are in good agreement. For the ACLS and ECLS methods, we use the same periodogram-based estimate of f_1 as in the single-frequency model, and we estimate f_2 also from the periodogram for $\{X_t\}$, obtaining $\hat{f}_2 \doteq 2.00339$. The table shows that the estimated jumps in integrated spectrum are approximately the same for all three methods and that their SS values are similar, with the small differences amongst them obeying the pattern dictated by theory (the SS for the EULS method does not exceed the one for ECLS, and the latter does not exceed the one for ACLS). Overall there is a reduction in SS of about 10% over the single-frequency model; however, the coefficient of determination is now 0.677 for all three methods, which is a modest 3% increase in explained variability as compared to 0.645 for the single-frequency model.

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

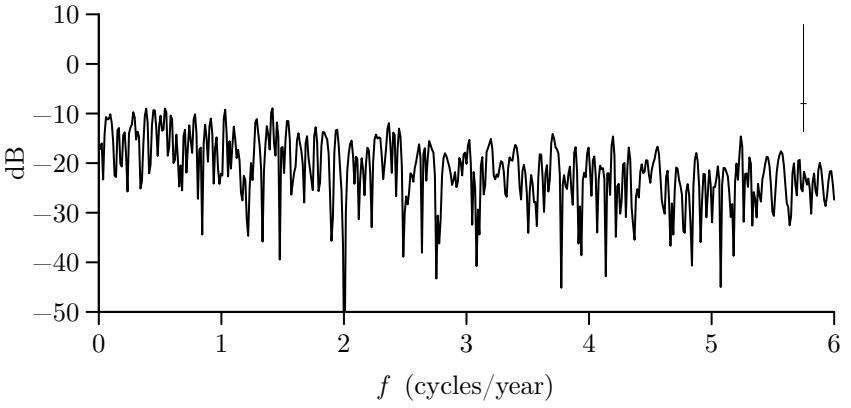

Figure 573 As in Figures 568b and 571b, but now the periodogram is based on the EULS residuals of Figure 572(b).

Figure 572 shows the mean-adjusted fitted two-frequency model

$$\overline{X} + \sum_{l=1}^{2} \left(\hat{A}_l \cos\left(2\pi \hat{f}_l t\, \Delta_t\right) + \hat{B}_l \sin\left(2\pi \hat{f}_l t\, \Delta_t\right) \right)$$

(top plot) and associated residuals

$$R_t = X_t - \sum_{l=1}^{2} \left(\hat{A}_l \cos\left(2\pi \hat{f}_l t\, \Delta_t\right) + \hat{B}_l \sin\left(2\pi \hat{f}_l t\, \Delta_t\right) \right)$$

(bottom) based upon the EULS method (plots for the ACLS and EULS methods are virtually identical). There are subtle visual differences between the two plots here and the corresponding ones in Figure 571a for the single-frequency model. Again, if the two-frequency model were adequate, $\{R_t\}$ should be close to a realization from a white noise process (we note in passing that there are two substantive downshoots in the residuals, perhaps attributable to periods of drought). Figure 573 shows the periodogram for the residuals. As is to be expected, the only substantial difference between this periodogram and that for single-frequency residuals in Figure 571b is near $f = 2$ cycles/year, where there is now a downshoot instead of a peak. There are no prominent peaks in the periodogram. Application of Fisher's g statistic now yields $g \doteq 0.039$, which fails to exceed either the 1% critical level (0.049) or the 5% critical level (0.041); however, the null hypothesis for Fisher's test is white noise, and failure to reject this hypothesis does not imply we should entertain it uncritically, particularly since the periodogram in Figure 573 visibly decreases with increasing frequency, in violation of the pattern we would expect to see from a realization of white noise.

We can test whether the residuals $\{R_t\}$ of Figure 572(b) can be considered to be drawn from a white noise process by applying the normalized cumulative periodogram test for white noise (described following Equation (215a)). Figure 574 shows the graphical equivalent of the level $\alpha = 0.05$ test. The test statistic is $D \doteq 0.334$, which, since it is greater than the critical level $\tilde{D}(0.05) \doteq 0.096$ of Equation (215b), says that we can reject the null hypothesis that the residuals come from a white noise process. This is reflected in the fact that, in Figure 574, the cumulative periodogram (the solid curve) crosses the upper slanted dashed line. The frequency associated with D is 2.46, which is indicated by a solid vertical line. To the left of this line are two dashed vertical lines. These indicate the frequencies $\hat{f}_1$ and $\hat{f}_2$ near which troughs occur

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

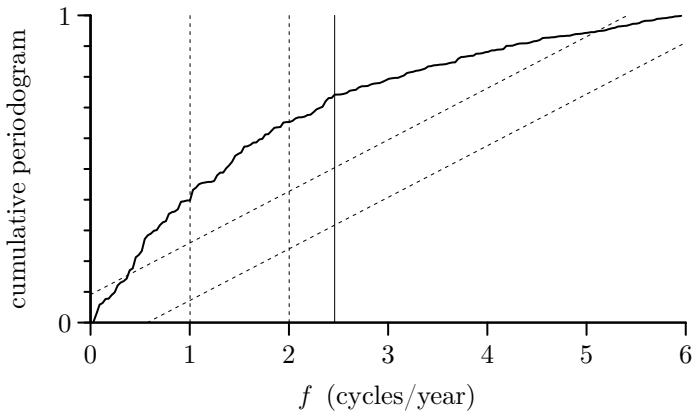

Figure 574 Normalized cumulative periodogram test for white noise as applied to EULS residuals obtained by fitting the model of Equation (572) to the centered Willamette River data.

in Figure 573 (plots similar to Figure 574 for the ACLS and ECLS residuals are visually identical to this figure).

Given that we cannot regard the background continuum as a white noise process, let us now consider a model of the form of Equation (550c), which takes the continuum to be a zero mean stationary process with an SDF $S_\eta(\cdot)$ that need not be that of white noise. Although we already have a good idea of the appropriate frequencies f_l to use in this model, let us illustrate the use of the F-test for periodicity discussed in Section 10.10. For this test, we select $K = 5$ Slepian multitapers with $NW = 4/\Delta_t$ (Exercise [10.28] is to redo what follows using $K = 5$ and $K = 6$ sinusoidal multitapers instead). We compute the F-test of Equation (547a) at frequencies $j/(1024\,\Delta_t)$ with $j = 1, 2, \ldots, 512$ and plot it in Figure 575(a). At a given fixed frequency, the F-test is F-distributed with 2 and $2K - 2 = 8$ degrees of freedom under the null hypothesis of no spectral line. We reject the null hypothesis for large values of the F-test. The two dashed horizontal lines on the plot indicate the levels of the upper 99% and 99.9% percentage points of the $F_{2,8}$ distribution (8.6 and 18.5, respectively, obtained by using Equation (544a) with $\nu = 8$ and α set to 0.01 and 0.001). The two largest excursions in the F-test occur at the same frequencies as the two largest peaks of the periodogram in Figure 568b, namely, 1.00781 and 2.00391 cycles/year. The largest excursion (at 1.00781 cycles/year) exceeds the 99.9% point and is clearly significant; the F-test at 2.00391 cycles/year exceeds the 99% point (but not the 99.9%). There are also two other frequencies at which the F-test exceeds the 99% point, namely, 2.461 and 4.195 cycles/year. The periodograms in Figures 568b, 571b and 573 do not clearly indicate a line component at either frequency, so how can we explain this result? The rule of thumb from Thomson (1990a) dictates not getting excited by significance levels greater than $1/N$ (see the discussion at the end of Section 10.10). Here $1/N = 1/395 \doteq 0.0025$, which translates into a critical value of 13.8 (indicated by the dotted horizontal line). Since only the excursions at 1.00781 and 2.00391 cycles/year exceed 13.8, the rule of thumb says that all the other peaks in the F-test are due to sampling fluctuations.

Based upon the results of the F-test, we now proceed under the assumption that there are line components at the frequencies $\hat{f}_1 \doteq 1.00781$ and $\hat{f}_2 \doteq 2.00391$ cycles/year so that $L = 2$ in Equation (550c). The associated complex-valued amplitudes for these frequencies, namely, C_1 and C_2, can be estimated using Equation (546a) to obtain

$$\begin{array}{ll} \hat{C}_1 \doteq -0.2912 - \mathrm{i}0.3122 & |\hat{C}_1|^2 \doteq 0.1823 \\ \hat{C}_2 \doteq 0.0232 - \mathrm{i}0.0984 & |\hat{C}_2|^2 \doteq 0.0102. \end{array} \quad \text{yielding}$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

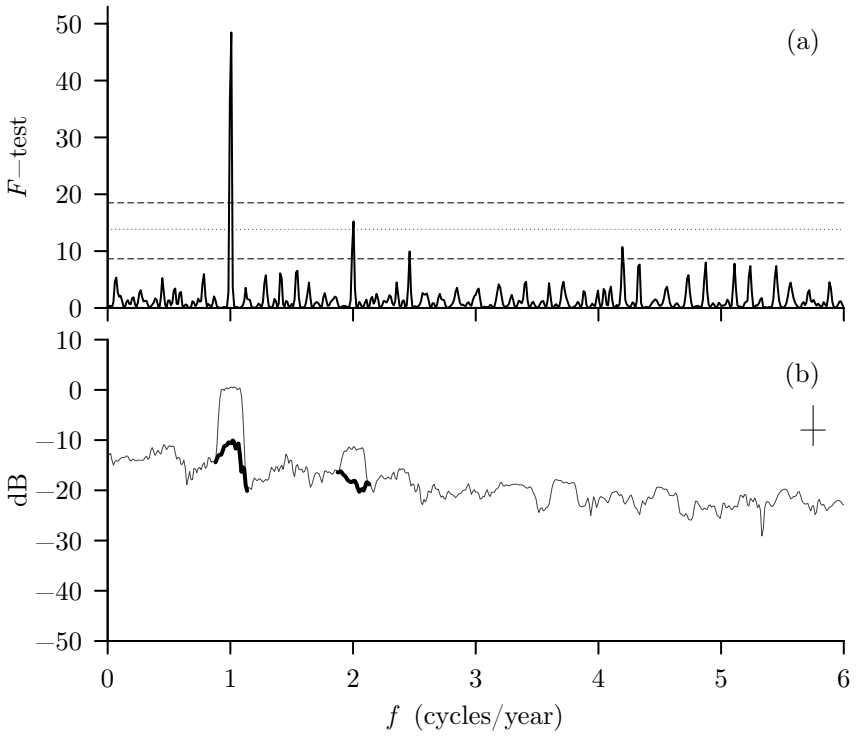

Figure 575 Thomson's F-test (top plot) and Slepian multitaper SDF estimate (bottom) for centered Willamette River data (see text for details).

The sizes of the jumps (0.1823 and 0.0102) in the integrated spectrum at $\hat{f}_1 \doteq 1.00781$ and $\hat{f}_2 \doteq 2.00391$ cycles/year are within 10% of the ACLS, ECLS and EULS estimates previously determined at frequencies close to these (see the bottom three rows of Table 570).

The thin curve in Figure 575(b) shows the basic multitaper SDF estimate $\hat{S}^{(\mathrm{MT})}(\cdot)$ of Equation (352a) for the centered Willamette River data. This estimate uses the same set of Slepian tapers as in the F-test; i.e., $K = 5$ and $NW = 4/\Delta_t$. As determined by Equation (353e), the bandwidth for the estimator is $B_{\overline{\mathcal{H}}} \doteq 0.19847$, which is the width of the criss-cross on the plot (as usual, the height is the length of a 95% CI for $10 \log_{10}(S(f))$). The peaks at 1 and 2 cycles/year are spread out by about this amount, and the distance between the two peaks is greater than $B_{\overline{\mathcal{H}}}$, indicating that the F-tests at the two frequencies are approximately independent. To obtain an estimate of $S_\eta(\cdot)$, we reshape $\hat{S}^{(\mathrm{MT})}(\cdot)$ using Equation (547b) at all frequencies dictated by Equation (547c), i.e., $|f - f_1| \le 1.25 B_{\overline{\mathcal{H}}}/2$ or $|f - f_2| \le 1.25 B_{\overline{\mathcal{H}}}/2$. Since f_1 corresponds to $86/(1024\,\Delta_t)$ and f_2 corresponds to $171/(1024\,\Delta_t)$, this amounts to reshaping $\hat{S}^{(\mathrm{MT})}(\cdot)$ at the frequencies $j/(1024\,\Delta_t)$ for $j = 76, 77, \ldots, 96$ and $j = 161, 162, \ldots, 181$. The two short thick curves indicate the reshaped portions of $\hat{S}^{(\mathrm{MT})}(\cdot)$. Our estimate of $S_\eta(\cdot)$ is thus given by these curves in addition to the unreshaped portions of $\hat{S}^{(\mathrm{MT})}(\cdot)$.

Let us now analyze the Willamette River data using a parametric approach. We fit AR models of orders $p = 1, 2, \ldots, 150$ using Burg's algorithm (Exercise [10.29] invites the reader to use the Yule–Walker and FBLS methods instead). Figure 576 shows a plot of the AICC order selection criterion versus p – see Section 9.11 for details. The minimum value occurs at $p = 27$ (marked by a vertical line) and suggests that an AR(27) process might be appropriate, but we also consider the case $p = 150$ for comparison. The top two plots in Figure 577 show the resulting SDF estimates over the grid of frequencies $j/(2^{16}\,\Delta_t)$, $0 \le j \le 2^{15} = 32,768$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

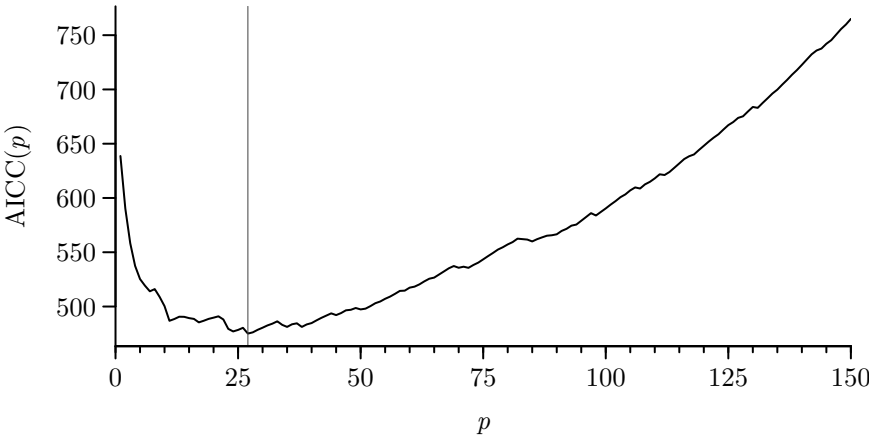

Figure 576 AICC order selection criterion for Willamette River data (parameter estimation via Burg's algorithm).

(we need a grid this fine for accurate portrayal of the peaks in the AR(150) SDF). In both estimates, the two largest peaks occur near 1 and 2 cycles/year, in agreement with periodogram shown in Figure 568b. Table 578 lists the observed peak frequencies $\hat{f}_j$, peak values $\hat{S}(\hat{f}_j)$ and peak widths for these SDFs and – for comparison – the periodogram. The peak widths at the observed peak frequency are measured by $f' + f''$, where $f' > 0$ and $f'' > 0$ are the smallest values such that

$$10 \log_{10}\left(\frac{\hat{S}(\hat{f}_j)}{\hat{S}(\hat{f}_j + f')}\right) = 10 \log_{10}\left(\frac{\hat{S}(\hat{f}_j)}{\hat{S}(\hat{f}_j - f'')}\right) = 3\,\text{dB}.$$

This table and the plots show some interesting qualitative differences and similarities amongst the SDF estimates:

[1] the peak near 2 cycles/year in the AR(27) SDF is displaced in frequency somewhat low compared to those based on the other two SDF estimates, while the peak near 1 cycle/year has about the same location for all three estimates;
[2] the widths of both AR(150) peaks are smaller than those for the corresponding AR(27) peaks by more than an order of magnitude;
[3] the widths of the peaks near 1 cycle/year in the AR SDFs are much smaller than the width of the peak in the periodogram (about 25 times smaller in the AR(150) case and two times in the AR(27)), while at 2 cycles/year the width is smaller only for $p = 150$;
[4] the widths of the peaks at 1 and 2 cycles/year for a particular AR estimate differ by about an order of magnitude, whereas the periodogram widths are similar (0.0265 and 0.0269);
[5] the heights of the two peaks in the AR(27) SDF are about an order of magnitude smaller than the corresponding peaks in the AR(150) SDF; and
[6] the AR(27) SDF is much smoother looking than the AR(150) SDF and the periodogram.

Item [1] illustrates the danger of too low an AR order (a loss of accuracy of peak location), while item [6], of too high an AR order (introduction of spurious peaks). Items [2] and [4] indicate an inherent problem with assessing peak widths in AR SDF estimates: whereas the peak widths in the periodogram can be attributed simply to the width of the central lobe of the associated spectral window, there is evidently no corresponding simple relationship for the AR estimates. If we believe that the true frequencies for this time series are at 1 and 2 cycles/year, items [1] and [3] indicate that the narrowness of the AR peaks does not translate into increased

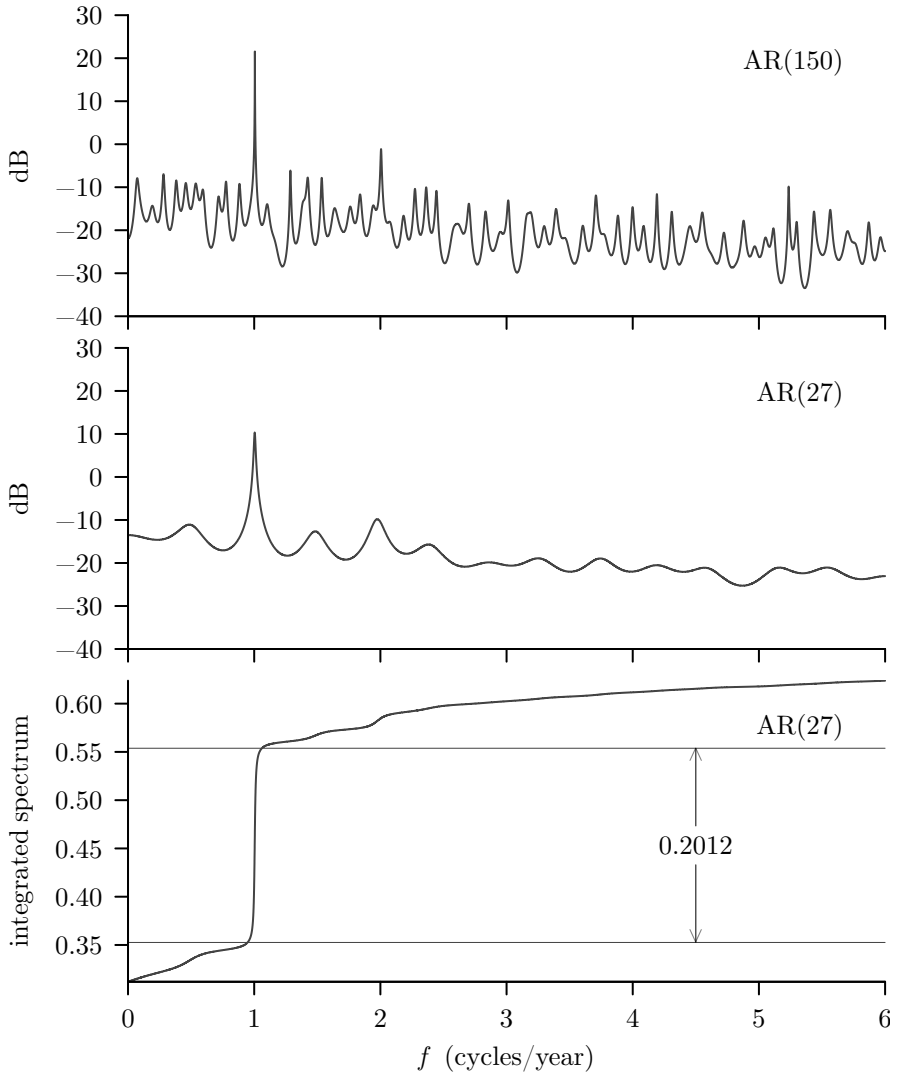

Figure 577 Burg-based AR(150) and AR(27) SDF estimates (top two plots) and integrated AR(27) spectrum (bottom plot) for Willamette River data (see text for details).

accuracy of the peak location. Indeed, as is pointed out by Burg (1975, p. 64) the narrowness of AR peaks calls for care in determining peak locations (see C&E [3] for Section 10.12).

Item [5] raises the question of how we can relate the height of a peak in an AR SDF at frequency f_l to the amplitude D_l of a sinusoidal component at that frequency in the assumed model

$$X_t = \sum_{l=1}^{L} D_l \cos\left(2\pi f_l \, \Delta_{\mathrm{t}} + \phi_l\right) + \epsilon_t.$$

In the case of the periodogram, we have the approximate relationship

$$E\{\hat{S}^{(\mathrm{P})}(f_l)\} \approx \sigma_\epsilon^2 \, \Delta_{\mathrm{t}} + N \, \Delta_{\mathrm{t}} D_l^2/4$$

(since $\mathcal{F}(0) = N \, \Delta_{\mathrm{t}}$, this follows from Equation (524b) under the assumption that f_l is well separated from the other frequencies in the model and is not too close to zero or the Nyquist

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

SDF Estimate	$\hat{f}_1$	3 dB Width	$\hat{S}(\hat{f}_1)$	$\hat{f}_2$	3 dB Width	$\hat{S}(\hat{f}_2)$
Burg AR(150)	1.0055	0.0009	21.6 dB	2.0055	0.0091	−1.1 dB
Burg AR(27)	1.0042	0.0126	10.3 dB	1.9757	0.1225	−9.8 dB
Periodogram	1.0032	0.0265	8.2 dB	2.0034	0.0269	−5.0 dB

Table 578 Characteristics of autoregressive SDFs and periodogram for Willamette River data (top two plots of Figure 577 and Figure 568b). The top two rows of numbers concern the two dominant peaks in the AR(p) SDFs. The first column indicates the AR estimator; the second, the location of the peak near 1 cycle/year; the third, the width of the peak 3 dB down from the peak value; and the fourth, the peak value itself. Columns 5, 6 and 7 show similar values for the peak near 2 cycles/year. The final row gives the corresponding values for the periodogram $\hat{S}^{(\mathrm{P})}(\cdot)$.

frequency; see also Figure 534). Pignari and Canavero (1991) concluded, via both theoretical calculations and Monte Carlo simulations, that "... the power estimation of sinusoidal signals by means of the Burg method is scarcely reliable, even for very high signal-to-noise ratios for which noise corruption of data is unimportant" (in our notation, the power in a sinusoid is proportional to D_l^2).

Burg (1985) advocated plotting the integrated spectrum $S^{(\mathrm{I})}(\cdot)$ as a way of determining the power in sinusoidal components. Since the integrated spectrum is related to the SDF $S(\cdot)$ by

$$S^{(\mathrm{I})}(f) = \int_{-f_{\mathcal{N}}}^{f} S(u)\,\mathrm{d}u,$$

the power in the frequency interval $[f', f'']$ is given by $S^{(\mathrm{I})}(f'') - S^{(\mathrm{I})}(f')$. As an example, the bottom plot in Figure 577 shows the integrated spectrum – obtained via numerical integration – corresponding to the AR(27) SDF estimate for the Willamette River data. The vertical distance between the two thin horizontal lines (centered about the peak near 1 cycle/year) in this plot is 0.2012, which is the ACLS-based estimate of the size of the jump in the integrated spectrum due to the line component near 1 cycle/year (see Table 570). Note that the increase in the AR(27) integrated spectrum around 1 cycle/year is roughly consistent with this value, but that precise determination of the power would be tricky, pending future research on exactly how to pick the interval $[f', f'']$ to bracket a peak. (Burg, 1975, pp. 65–6, notes that, because the integral over $f \in [-f_{\mathcal{N}}, f_{\mathcal{N}}]$ of a Burg-based AR SDF estimate must in theory be equal to the sample variance of the time series, numerical integration of the SDF over the grid of frequencies used for plotting should ideally yield a value close to the sample variance – if not, then the grid might be too coarse to accurately display the estimated SDF. Numerical integration of the AR(150) and AR(27) SDFs in Figure 577 yields, respectively, 0.6237824 and 0.6237827, both of which agree well with the sample variance for the centered Willamette River data, namely, 0.6237818. This result suggests that the frequency grid is adequate for displaying the overall patterns in the SDFs, but does not guarantee the accurate portrayal of sharp peaks.)

Let us now use the centered Willamette River time series to illustrate the SVD approach to harmonic analysis. In Section 10.14 we formulated the SVD method for complex-valued time series, whereas the Willamette River is real-valued. While a real-valued series is a special case of complex-valued series, here we render the series $\{X_t\}$ into complex-valued form by creating a corresponding "analytic" series (Exercise [10.31] explores using the real-valued series as is and also using a complex-valued series created by complex demodulation). The

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

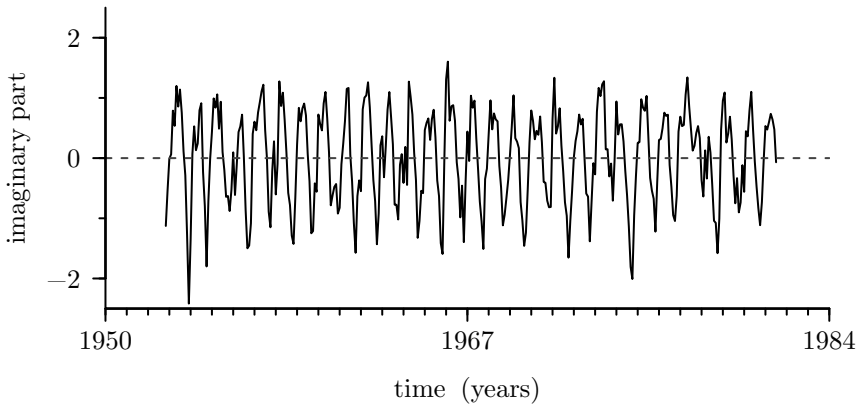

Figure 579 Imaginary part of analytic series used in SVD analysis. The series has $N = 345$ values, and the corresponding real part (prior to centering) is delineated in Figure 568a by two vertical dashed lines.

analytic series is

$$X_t + \mathrm{i}\,\mathcal{HT}\{X_t\},$$

where $\mathcal{HT}$ is the Hilbert transform of $\{X_t\}$ (see Section 10.13 above and section 9–3 of Papoulis and Pillai, 2002). By definition, the transfer function (frequency response function) of the Hilbert transform, $G(\cdot)$ say, must have unit gain, a phase angle of $-\pi/2$ for f between 0 and 1/2 and a phase angle of $\pi/2$ for f between 0 and $-1/2$. Thus $G(\cdot)$ takes the form

$$G(f) = \begin{cases} \mathrm{i}, & -1/2 < f < 0; \\ -\mathrm{i}, & 0 < f < 1/2 \end{cases} \tag{579a}$$

(Oppenheim and Schafer, 2010, equation (12.62b)). The corresponding impulse response sequence $\{g_u\}$ for the Hilbert transform is

$$g_u = \int_{-1/2}^{1/2} G(f) \mathrm{e}^{\mathrm{i}2\pi f u}\,\mathrm{d}f = \begin{cases} 2/(\pi u), & u \text{ odd}; \\ 0, & u \text{ even} \end{cases} \tag{579b}$$

(verification of the above is the subject of Exercise [10.30a]). We can obtain a realizable approximation to this ideal (infinitely long) impulse response by multiplying the g_u sequence by a set of convergence factors (see Sections 3.8 and 5.8); a reasonable choice (amongst many others) is a Hanning lag window:

$$c_u = \begin{cases} 1 - 0.5 + 0.5\cos(\pi u/25), & |u| \le 25; \\ 0, & \text{otherwise} \end{cases} \tag{579c}$$

(see Equation (343c) with a set to $1/4$ and m set to 25). The convergence factors reduce the impulse response sequence smoothly to a total length of 51 nonzero coefficients (i.e., 25 each side of $u = 0$). This approximation creates no phase errors, but slight gain errors do occur, mainly near zero and Nyquist frequencies (see Exercise [10.30b]). The result of filtering the series $\{X_t\}$ of length $N = 395$ with the filter $\{c_u g_u\}$ of width 51 is a series of length 345. This series is the imaginary part of the analytic series and is plotted in Figure 579; the corresponding real part $X_{25}, X_{26}, \ldots, X_{369}$ is shown (prior to centering) in Figure 568a between the two vertical dashed lines.

In applying the methods of Section 10.14 to the complex (analytic) series $\{Z_t\}$, we need to set p', which represents the number of values prior to Z_t needed to predict it well. Our

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

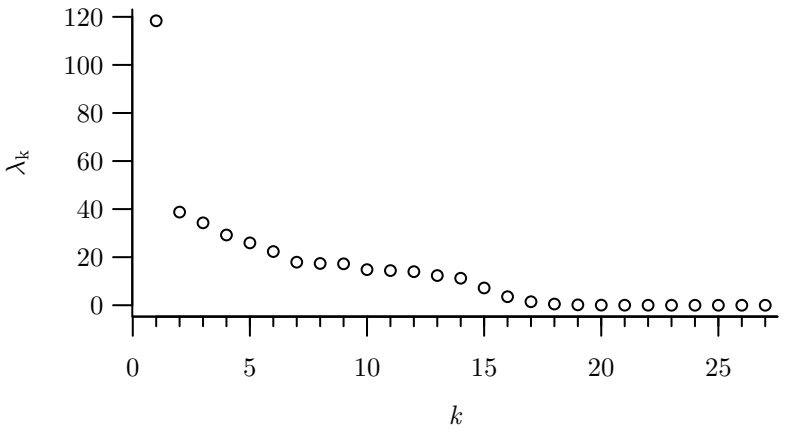

Figure 580a Eigenvalues λ_k of orders $k = 1, \ldots, 27$ in SVD of analytic series created from Willamette River data.

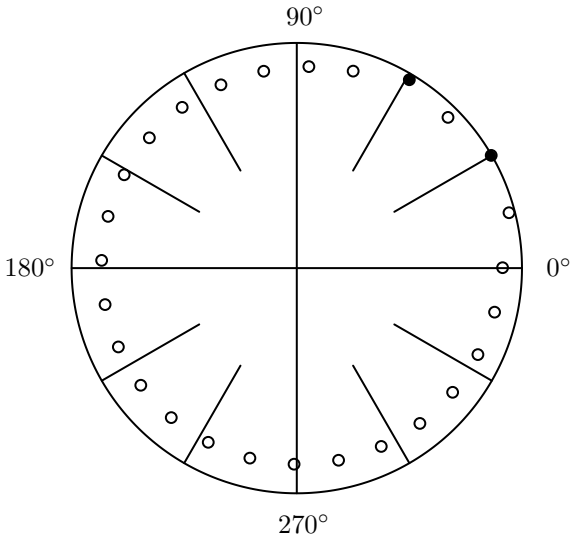

Figure 580b Complex roots z'_k (small circles) in SVD analysis of Willamette River data. The large circle is the unit circle in the complex plane. The lines indicate phase angles on the unit circle that are multiples of 30°. There are 27 roots in all, and the two filled circles indicate the two closest to the unit circle – these have phase angles close to 30° and 60°.

parametric analysis of $\{X_t\}$ suggests $p' = 27$, and experimentation with other choices indicates this to be suitable. Our previous investigations indicate two sinusoids with frequencies close to one and two cycles/year, which suggests truncating at $p = 2$ in the SVD method; on the other hand, a plot of the eigenvalues (Figure 580a) suggests $p = 1$. Neither of these choices gives results in reasonable agreement with our previous analyses, but setting $p = 3$ does. The resulting estimated roots are plotted in polar form in Figure 580b and were found using the function `polyroot` in the `R` language (the SVD itself was computed using the function `svd`). The two complex roots z'_k closest to the unit circle are at $(0.8635, 0.5007)$ and $(0.4997, 0.8366)$, having magnitudes 0.9981 and 0.9745, respectively, and phase angles quite close to 30° and 60°. Recalling that $\Delta t = 1/12$ year, the corresponding frequencies f'_k are

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

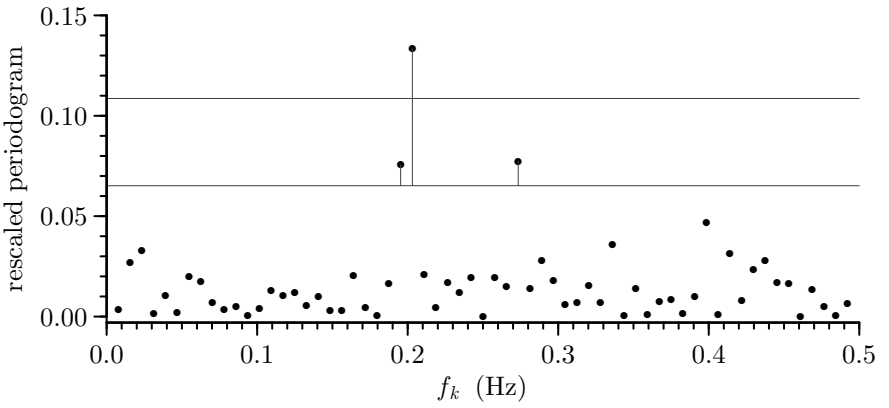

Figure 581 Application of Fisher's and Siegel's tests to ocean noise data (see text for details).

1.0035 and 1.9717 cycles/year, similar to those obtained in our other analyses. There is, however, a third root whose magnitude (0.9740) is just below that of the root with the second largest magnitude. This root has a phase angle close to 15° and is associated with 0.4876 cycles/year, a frequency that has not been flagged by our previous analyses and hence must be regarded with caution.

Ocean Noise Data

Here we give an example of the application of Fisher's and Siegel's tests using the ocean noise data in Figure 4(d). In Section 1.2 we noted that the sample ACS for this time series shows a tendency to oscillate with a period of 5 sec. What do the tests for simple or compound periodicity in white noise tell us about this series? To begin, we calculate $\hat{S}^{(\mathrm{P})}(f_k)$ over the grid of Fourier frequencies appropriate for a series of length $N = 128$, for which, as per Equation (539a), the number of Fourier frequencies satisfying $0 < f_k < f_{\mathcal{N}} = 1/2$ Hz is $M = (N-2)/2 = 63$. We then form the rescaled periodogram $\tilde{S}^{(\mathrm{P})}(f_k)$ of Equation (541a). Figure 581 shows $\tilde{S}^{(\mathrm{P})}(f_k)$ versus f_k for $k = 1, 2, \ldots, M$ (solid circles). The critical value g_F for Fisher's test at $\alpha = 0.05$ obtained via Equation (540b) is marked as the upper horizontal line (its values is 0.1086, while its approximation $\tilde{g}_F$ given by Equation (540d) is 0.1088); the lower line marks $0.6g_F$ used in Siegel's test. The maximum value of $\tilde{S}^{(\mathrm{P})}(f_k)$ is 0.1336 and substantially exceeds g_F, so that the null hypothesis of white noise is rejected at the 5% level using Fisher's test. Siegel's test statistic $T_{0.6}$ is formed from the sum of the positive excesses of the $\tilde{S}^{(\mathrm{P})}(f_k)$ terms over $0.6g_F \doteq 0.0652$. These excesses are shown by the three vertical lines in Figure 581; their sum is 0.0909. For $\alpha = 0.05$ and $M = 63$, the critical value for Siegel's test as stated in Table 542 is 0.0523 (its approximations 0.0515 and 0.0518 via $\tilde{t}_{0.6}$ and $c\chi_0^2(\beta)$ are slightly smaller). Hence the null hypothesis of white noise is rejected at the 5% level using Siegel's test. The first two values of $\tilde{S}^{(\mathrm{P})}(f_k)$ exceeding $0.6g_F$ are adjacent and probably are both associated with a frequency of around 0.2 Hz, which corresponds to a period of 5 sec. There appears to be an additional sinusoid present with a frequency of 0.273 Hz, corresponding to a period of about 3.7 sec.

Now consider carrying out the tests at the 1% level ($\alpha = 0.01$). The critical value g_F is now 0.1316 (its approximation $\tilde{g}_F$ is also 0.1316). The maximum value of $\tilde{S}^{(\mathrm{P})}(f_k)$ of 0.1336 just exceeds g_F, so Fisher's test rejects the null hypothesis of white noise also at the 1% level. The sum of positive excesses of $\tilde{S}^{(\mathrm{P})}(f_k)$ over $0.6g_F \doteq 0.0789$ is 0.0547. For $\alpha = 0.01$ and $M = 63$, Table 542 says that the critical value for Siegel's test is 0.0562 (its approximations via $\tilde{t}_{0.6}$ and $c\chi_0^2(\beta)$ are quite close: 0.0552 and 0.0564). Hence we fail to reject the null hypothesis of white noise (but just barely!) at the 1% significance level using Siegel's test.

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

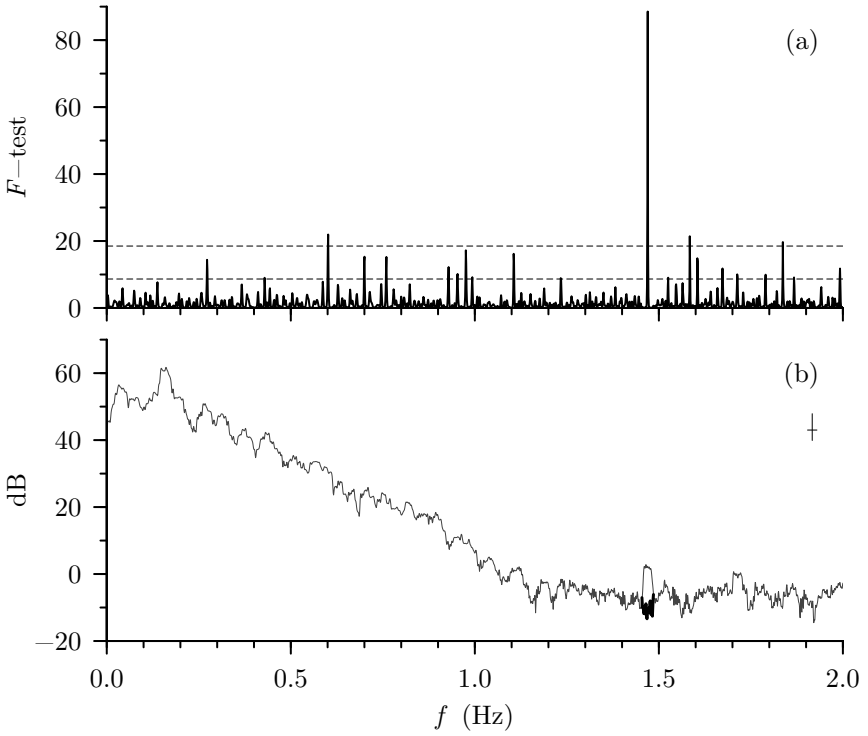

Figure 582 Thomson's F-test (top plot) and Slepian multitaper SDF estimate (bottom) for centered ocean wave data (see text for details).

The practical conclusion from these tests is that there is almost certainly one, and probably two sinusoids, of periods about 5 and 3.7 sec, present in the ocean noise data (this finding is not inconsistent with results from the normalized cumulative periodogram test for white noise discussed in C&E [6] for Section 6.8).

Ocean Wave Data

In our analysis of the ocean wave data in Section 6.8, we noted a small peak at $f = 1.469$ Hz in the direct spectral estimates shown in Figures 226(c) and (d). Here we assess the significance of that peak using Thomson's multitaper-based F-test for periodicity (see Section 10.10). As we did for the Willamette River data, we use $K = 5$ Slepian multitapers with $NW = 4/\Delta_t$ and compute the F-test of Equation (547a) over a grid of frequencies $j/(4096\,\Delta_t)$, $j = 1$, 2, ..., 2048 (since the sample size for the ocean wave data is $N = 1024$, the spacing between Fourier frequencies is $1/(1024\,\Delta_t)$, so the grid we are using has a much finer spacing). Figure 582 plots the results in a manner similar to Figure 575 for the Willamette River case. Figure 582(a) shows the F-test and has two dashed horizontal lines, which indicate the levels of the upper 99% and 99.9% percentage points of the $F_{2,8}$ distribution needed to evaluate the test (here the rule-of-thumb critical level from Thomson, 1990a, corresponds to a significance level of $1/N = 1/1024 \approx 0.001$ and hence is virtually the same as the critical level for the upper 99.9% percentage point). Recalling that we reject the null hypothesis of no spectral line for large values of the F-test, there is one obvious extreme rejection, which is associated with a level of significance of 3.5×10^{-6} and with frequency $\hat{f}_1 = 1505/(4096\,\Delta_t) \doteq 1.470$ Hz, in close agreement with the location of the peak in Figure 226 (the SDF estimates there are computed over the grid of Fourier frequencies, and the Fourier frequency closest to $\hat{f}_1$ is in

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

Method	Time Series	Model	μ Estimator	μ Estimate	SS
ECLS	uncentered	one frequency	least squares	9.82573	87.40153
ECLS	centered	one frequency	$\overline{X}$	9.82542	87.40157
EULS	uncentered	one frequency	least squares	9.82576	87.39877
EULS	centered	one frequency	$\overline{X}$	9.82542	87.39881
ECLS	uncentered	two frequencies	least squares	9.82564	79.58303
ECLS	centered	two frequencies	$\overline{X}$	9.82542	79.58305
EULS	uncentered	two frequencies	least squares	9.82566	79.54591
EULS	centered	two frequencies	$\overline{X}$	9.82542	79.54593

Table 583 Comparison of two ways to handle the mean μ in models of the form of Equation (512b) for the Willamette River time series. The first way is to center the series using the sample mean, set $\mu = 0$ in Equation (512b) and then estimate the remaining model parameters via least squares; the second is to use the uncentered series and estimate μ along with the other parameters. The next-to-last column in the table lists the resulting estimates of the mean for different combinations of methods and models; the last column lists the associated SSs.

fact $376/(1024\,\Delta_t) \doteq 1.469$ Hz). Figure 582(b) shows the corresponding multitaper SDF estimate (thin curve) and the reshaped estimate done using Equation (547b) over the interval of frequencies around $\hat{f}_1$ dictated by Equation (547c) (thick shorter curve). The revised estimate of the background continuum flows well with the SDF estimate before and after the interval about $\hat{f}_1$.

There are, however, three other frequencies at which the F-test exceeds the critical level associated with a level of significance 0.001, namely, $\hat{f}_2 \doteq 0.602$ Hz, $\hat{f}_3 \doteq 1.584$ Hz and $\hat{f}_4 \doteq 1.837$ Hz. The SDF estimate in Figure 582(b) does not offer visual confirmation of peaks at these frequencies. If we consider F-tests based upon $K = 5$ and $K = 6$ sinusoidal multitapers, we find that the ones at $\hat{f}_2$ and $\hat{f}_4$ no longer exceed the critical level associated with 0.001, which suggests caution in declaring significant peaks at these two frequencies; however, both sinusoidal-based F-tests at $\hat{f}_3$ still exceed the 0.001 critical level, in agreement with the Slepian-based test. A look at the reshaped Slepian-based SDF estimate around $\hat{f}_3$ (not shown in Figure 582(b)) reveals an unappealing gouge in the estimated background continuum with a depth of about 10 dB. This result suggests considering an F-test based upon a multitapering scheme with a much wider bandwidth. If we consider the F-test based upon $K = 12$ sinusoidal multitapers, indeed the test at $\hat{f}_3$ no longer exceeds the critical level associated with 0.001. In conclusion, after considering the F-test with four different sets of multitapers, the only frequency at which we reject the null hypothesis of no spectral line in all cases is $\hat{f}_1$, and the observed critical level is always orders of magnitude smaller than 0.001.

Comments and Extensions to Section 10.15

[1] In our analysis of the Willamette River data, we centered the time series by subtracting off its sample mean $\overline{X}$ and then fit single- and two-frequency models of the form of Equation (512b) with μ set to zero using the ACLS, ECLS and EULS methods. For ECLS and EULS, an alternative to centering the series is to retain μ in the model and to estimate it using least squares along with the other model parameters. Including μ impacts the estimates of the other parameters, but almost undetectably: the new values for $\hat{f}_1$, $\hat{A}_1$, $\hat{B}_1$, $\hat{f}_2$, $\hat{A}_2$ and $\hat{B}_2$ round to the same values listed in Table 570 for estimates based on the centered data (however, for the two-frequency model with the EULS method, the estimated jump at $\hat{f}_2 \doteq 2.0024$ changes slightly from 0.00993 to 0.00994). Table 583 lists the least squares estimates of μ and the resulting SSs for the ECLS and EULS methods in combination with the single- and two-frequency models, along with corresponding results extracted from Table 570 based on centering the

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

time series. The least squares estimates of μ and the sample mean differ only in the third decimal place. Least squares estimation of μ must give a SS that is no greater than the one associated with setting μ equal to $\overline{X}$, a pattern that holds true in Table 583 in all four method/model combinations; however, in each case, the difference in SSs for two ways of handling μ amounts to a change in the fifth decimal place. We conclude that handling the nonzero mean for the Willamette River data by centering the series is as effective as dealing with it by including μ in the model.

10.16 Summary of Harmonic Analysis

In this section we collect together the most useful results of this chapter. It is assumed throughout that $\mu = 0$; D_l and f_l are real-valued constants; C_l is a complex-valued constant; and the ϕ_l terms are independent real-valued RVs uniformly distributed on $(-\pi, \pi]$. The spectral classifications used are as defined in Section 4.4.

[1] *Real-valued discrete parameter harmonic processes*

(a) Mixed spectrum:

$$X_t = \sum_{l=1}^{L} D_l \cos(2\pi f_l t\,\Delta_t + \phi_l) + \eta_t \qquad \text{(see (518c))}$$

$$= \sum_{l=-L}^{L} C_l e^{i2\pi f_l t\,\Delta_t} + \eta_t, \qquad \text{(see (519b))}$$

where $\{\eta_t\}$ is a real-valued (possibly colored) noise process (independent of each ϕ_l) with zero mean and variance σ_η^2, $C_0 = 0$, $f_0 = 0$, and, for $l = 1, \ldots, L$, $C_l = D_l \exp(i\phi_l)/2$, $C_{-l} = C_l^*$ and $f_{-l} = -f_l$. Then, $E\{X_t\} = 0$,

$$\text{var}\{X_t\} = \sum_{l=-L}^{L} E\{|C_l|^2\} + \text{var}\{\eta_t\} = \sum_{l=1}^{L} D_l^2/2 + \sigma_\eta^2 \qquad \text{(see (519c))}$$

and

$$\text{cov}\{X_{t+\tau}, X_t\} = \sum_{l=1}^{L} D_l^2 \cos(2\pi f_l \tau\,\Delta_t)/2 + \text{cov}\{\eta_{t+\tau}, \eta_t\}. \qquad \text{(see (519e))}$$

(b) Discrete spectrum:
If the noise process $\{\eta_t\}$ is white, i.e., $\eta_t = \epsilon_t$, then the mixed spectrum reduces to a discrete spectrum, and

$$X_t = \sum_{l=1}^{L} D_l \cos(2\pi f_l t\,\Delta_t + \phi_l) + \epsilon_t \qquad \text{(see (512b))}$$

$$= \sum_{l=1}^{L} (A_l \cos(2\pi f_l t\,\Delta_t) + B_l \sin(2\pi f_l t\,\Delta_t)) + \epsilon_t, \qquad \text{(see (515b))}$$

where $\{\epsilon_t\}$ is a real-valued white noise process with zero mean and variance σ_ϵ^2, independent of each ϕ_l. We have $E\{X_t\} = 0$,

$$\text{var}\{X_t\} = \sum_{l=1}^{L} D_l^2/2 + \sigma_\epsilon^2 \quad \text{and} \quad \text{cov}\{X_{t+\tau}, X_t\} = \sum_{l=1}^{L} D_l^2 \cos(2\pi f_l \tau\,\Delta_t)/2, \quad |\tau| > 0.$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

(c) Known frequencies, discrete spectrum:
If each f_l is any of the Fourier frequencies not equal to 0 or $f_{\mathcal{N}}$,

$$\hat{A}_l = \frac{2}{N}\sum_{t=0}^{N-1} X_t \cos(2\pi f_l t\,\Delta_t); \quad \hat{B}_l = \frac{2}{N}\sum_{t=0}^{N-1} X_t \sin(2\pi f_l t\,\Delta_t) \qquad \text{(see (515c))}$$

give *exact* least squares estimates of A_l and B_l. If now each ϕ_l is treated as a constant so that A_l and B_l are constants and Equation (515b) is a multiple linear regression model, then

$$E\{\hat{A}_l\} = A_l, \;\; E\{\hat{B}_l\} = B_l \;\text{ and }\; \text{var}\,\{\hat{A}_l\} = \text{var}\,\{\hat{B}_l\} = \frac{2\sigma_\epsilon^2}{N}. \qquad \text{(see (516a))}$$

Moreover, if $k \neq l$, cov $\{\hat{A}_k, \hat{B}_l\}$, cov $\{\hat{A}_l, \hat{B}_l\}$, cov $\{\hat{A}_k, \hat{A}_l\}$ and cov $\{\hat{B}_k, \hat{B}_l\}$ are all zero, and σ_ϵ^2 is estimated by Equation (516b). If the frequencies f_l are not all Fourier frequencies, then $\hat{A}_l$ and $\hat{B}_l$ give *approximate* least squares estimates, and

$$E\{\hat{A}_l\} = A_l + O\left(\frac{1}{N}\right) \;\text{ and }\; E\{\hat{B}_l\} = B_l + O\left(\frac{1}{N}\right). \qquad \text{(see (516c))}$$

(d) Unknown frequencies, discrete spectrum:
When the frequencies f_l are unknown, the standard approach is to look for peaks in the periodogram. If f_l is a Fourier frequency (not equal to 0 or $f_{\mathcal{N}}$) present in $\{X_t\}$, then

$$E\{\hat{S}^{(\mathrm{P})}(f_l)\} = \left(N\frac{D_l^2}{4} + \sigma_\epsilon^2\right)\Delta_t, \qquad \text{(see (524a))}$$

while, if f_l is a Fourier frequency not in $\{X_t\}$, then

$$E\{\hat{S}^{(\mathrm{P})}(f_l)\} = \sigma_\epsilon^2\,\Delta_t.$$

When f_l is not necessarily one of the Fourier frequencies, it follows that

$$E\{\hat{S}^{(\mathrm{P})}(f)\} = \sigma_\epsilon^2\,\Delta_t + \sum_{l=1}^{L}\frac{D_l^2}{4}\left[\mathcal{F}(f+f_l) + \mathcal{F}(f-f_l)\right], \qquad \text{(see (524b))}$$

where $\mathcal{F}(\cdot)$ is Fejér's kernel. If a data taper $\{h_t\}$ is applied to the process, then the expectation of the direct spectral estimator is given by

$$E\{\hat{S}^{(\mathrm{D})}(f)\} = \sigma_\epsilon^2\,\Delta_t + \sum_{l=1}^{L}\frac{D_l^2}{4}\left[\mathcal{H}(f+f_l) + \mathcal{H}(f-f_l)\right], \qquad \text{(see (535a))}$$

where $\mathcal{H}(\cdot)$ is the spectral window corresponding to $\{h_t\}$. If the taper is the default rectangular taper, then $\mathcal{H}(\cdot) = \mathcal{F}(\cdot)$.

[2] *Complex-valued discrete parameter harmonic processes*

(a) Mixed spectrum:

$$Z_t = \sum_{l=1}^{L} D_l e^{i(2\pi f_l t\,\Delta_t + \phi_l)} + \eta_t = \sum_{l=1}^{L} C_l e^{i2\pi f_l t\,\Delta_t} + \eta_t, \qquad \text{(see (519g))}$$

where $\{\eta_t\}$ is a proper complex-valued (possibly colored) noise process with zero mean and variance σ_η^2, independent of each ϕ_l, and $C_l = D_l \exp(i\phi_l)$. Then, $E\{Z_t\} = 0$,

$$\text{var}\,\{Z_t\} = \sum_{l=1}^{L} E\{|C_l|^2\} + \text{var}\,\{\eta_t\} = \sum_{l=1}^{L} D_l^2 + \sigma_\eta^2 \qquad \text{(see (519h))}$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

and

$$\text{cov}\{Z_{t+\tau}, Z_t\} = \sum_{l=1}^{L} D_l^2 e^{i2\pi f_l \tau \Delta_t} + \text{cov}\{\eta_{t+\tau}, \eta_t\}. \qquad \text{(see (519i))}$$

(b) Discrete spectrum:
If the noise process is white, i.e., $\eta_t = \epsilon_t$ then the mixed spectrum reduces to a discrete spectrum, and

$$Z_t = \sum_{l=1}^{L} D_l e^{i(2\pi f_l t \Delta_t + \phi_l)} + \epsilon_t, \qquad \text{(see (517a))}$$

where $\{\epsilon_t\}$ is a real-valued white noise process with zero mean and variance σ_ϵ^2, independent of each ϕ_l. We have $E\{Z_t\} = 0$,

$$\text{var}\{Z_t\} = \sum_{l=1}^{L} D_l^2 + \sigma_\epsilon^2 \text{ and } \text{cov}\{Z_{t+\tau}, Z_t\} = \sum_{l=1}^{L} D_l^2 e^{i2\pi f_l \tau \Delta_t}, \quad |\tau| > 0.$$

(c) Known frequencies, discrete spectrum:
Let us assume a process with a discrete spectrum. If each f_l is any of the Fourier frequencies not equal to 0 or $f_{\mathcal{N}}$,

$$\hat{C}_l = \frac{1}{N} \sum_{t=0}^{N-1} Z_t e^{-i2\pi f_l t \Delta_t} \qquad \text{(see Exercise [10.2])}$$

is the exact least squares estimate of C_l. With each ϕ_l treated as a constant,

$$E\{\hat{C}_l\} = C_l, \quad \text{var}\{\hat{C}_l\} = \frac{\sigma_\epsilon^2}{N} \text{ and } \text{cov}\{\hat{C}_k, \hat{C}_l\} = 0, \quad k \neq l.$$

(d) Unknown frequencies, discrete spectrum:
If f_l is a Fourier frequency (not equal to 0 or $f_{\mathcal{N}}$) present in $\{Z_t\}$, then

$$E\{\hat{S}^{(\text{P})}(f_l)\} = \left(N D_l^2 + \sigma_\epsilon^2\right) \Delta_t,$$

while, if f_l is a Fourier frequency not in $\{Z_t\}$,

$$E\{\hat{S}^{(\text{P})}(f_l)\} = \sigma_\epsilon^2 \Delta_t.$$

[3] *Tests for periodicity for real-valued process*

(a) White noise:
This case is discussed in detail in Section 10.9. The sample size N is usually taken to be odd so that $N = 2M + 1$ for some integer M (Section 10.15 demonstrates how to accommodate an even sample size). The null hypothesis is $D_1 = \cdots = D_L = 0$. Fisher's exact test for *simple* periodicity uses the statistic

$$g \stackrel{\text{def}}{=} \max_{1 \le k \le M} \frac{\hat{S}^{(\text{P})}(f_k)}{\sum_{j=1}^{M} \hat{S}^{(\text{P})}(f_j)}, \qquad \text{(see (539b))}$$

where $\hat{S}^{(\text{P})}(f_k)$ is a periodogram term. The exact distribution of g under the null hypothesis is given by Equation (540b). Critical values g_F for Fisher's test for $\alpha = 0.01, 0.02, 0.05$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

and 0.1 can be adequately approximated using Equation (540d). To test for the presence of *compound* periodicities, Siegel derived the statistic based on excesses over a threshold,

$$T_\lambda \stackrel{\text{def}}{=} \sum_{k=1}^{M} \left(\tilde{S}^{(\mathrm{P})}(f_k) - \lambda g_F \right)_+, \qquad \text{(see (541b))}$$

where $\tilde{S}^{(\mathrm{P})}(f_k) = \hat{S}^{(\mathrm{P})}(f_k) \big/ \sum_{j=1}^{M} \hat{S}^{(\mathrm{P})}(f_j)$, $0 < \lambda \leq 1$, and g_F is the critical value for Fisher's test. The exact distribution of T_λ under the null hypothesis is given by Equation (541c). When $\lambda = 0.6$, critical values for M from about 20 to 2000 can be approximated using Equation (542a) for $\alpha = 0.05$, and Equation (542b) for $\alpha = 0.01$.

(b) Colored noise:
This case is discussed in detail in Section 10.10. To test for periodicity at a particular frequency f_1 for a real-valued process incorporating colored *Gaussian* noise, the recommended statistic is

$$\frac{(K-1)\left|\hat{C}_1\right|^2 \sum_{k=0}^{K-1} H_k^2(0)}{\Delta_\mathrm{t} \sum_{k=0}^{K-1} \left|J_k(f_1) - \hat{J}_k(f_1)\right|^2}, \qquad \text{(see (547a))}$$

where $H_k(0) = \Delta_\mathrm{t} \sum_{t=0}^{N-1} h_{k,t}$,

$$J_k(f_1) = \Delta_\mathrm{t}^{1/2} \sum_{t=0}^{N-1} h_{k,t} X_t \mathrm{e}^{-\mathrm{i}2\pi f_1 t \Delta_\mathrm{t}}, \quad \hat{J}_k(f_1) = \hat{C}_1 \frac{H_k(0)}{\Delta_\mathrm{t}^{1/2}}$$

and $\hat{C}_1$ is given by Equation (546a). Here $\{h_{k,t}\}$ is a kth-order Slepian or sinusoidal data taper. This statistic has a simple $F_{2,2K-2}$ distribution under the null hypothesis of no periodicity at the frequency f_1.

10.17 Exercises

[10.1] Verify that the approximations in Equation (515a) are valid.

[10.2] For the complex-valued model given by Equation (518a), show that $\operatorname{cov}\{\hat{C}_k, \hat{C}_l\} = 0$ for $k \neq l$, where $\hat{C}_k$ and $\hat{C}_l$ are the least squares estimators of, respectively, C_k and C_l; i.e.,

$$\hat{C}_k = \frac{1}{N} \sum_{t=0}^{N-1} Z_t \mathrm{e}^{-\mathrm{i}2\pi f_k t \Delta_\mathrm{t}} \text{ and } \hat{C}_l = \frac{1}{N} \sum_{t=0}^{N-1} Z_t \mathrm{e}^{-\mathrm{i}2\pi f_l t \Delta_\mathrm{t}}.$$

[10.3] Prove that

$$\operatorname{cov}\{Z_{t+\tau}, Z_t\} = \sum_{l=1}^{L} D_l^2 \mathrm{e}^{\mathrm{i}2\pi f_l \tau \Delta_\mathrm{t}} + \operatorname{cov}\{\eta_{t+\tau}, \eta_t\}$$

for $\{Z_t\}$ given by Equation (519g).

[10.4] Verify Equation (524b).

[10.5] Consider a time series $X_0, X_1, \ldots, X_{N-1}$ obeying Equation (512b) specialized to the case $\mu = 0$, $L = 1$ and $\Delta_\mathrm{t} = 1$:

$$X_t = D_1 \cos(2\pi f_k' t + \phi_1) + \epsilon_t,$$

where $f_k' = \frac{2k+1}{2N}$, and k is a nonnegative integer such that $k < (N-1)/2$; i.e., f_k' falls halfway between the Fourier frequencies k/N and $(k+1)/N$. Equation (526d) says that the signal-to-noise ratio for $\{X_t\}$ is $R = D_1^2/(2\sigma_\epsilon^2)$, where, as usual, $\sigma_\epsilon^2 = \operatorname{var}\{\epsilon_t\}$.

(a) Assuming that N and R are sufficiently large and that f_k' is not too close to zero or Nyquist frequency, argue that the ratio

$$r(f_k') \stackrel{\text{def}}{=} \frac{E\{\hat{S}^{(\mathrm{P})}(f_k' + \frac{1}{2N})\}}{E\{\hat{S}^{(\mathrm{P})}(f_k')\}} \approx \frac{4}{\pi^2} \doteq 0.405$$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

by appealing to Equation (524b) (Whittle, 1952; see also Priestley, 1981, p. 403; because $f'_k + \frac{1}{2N}$ is a Fourier frequency, the above supports the claim that evaluation of the periodogram at just the Fourier frequencies tends to underestimate its peak value when the dominant frequency in $\{X_t\}$ is halfway between Fourier frequencies).

(b) Compute and plot the normalized ratio $r(f'_k)/(4/\pi^2)$ versus f'_k, $k = 0, 1, \ldots, 511$, for the case $N = 1024$ and $R = 1$, Repeat for $R = 0.1$, 10, and 100. How well does the approximation $r(f'_k) \approx 4/\pi^2$ work here?

[10.6] (a) Verify Equation (526a) and the closely related equation

$$\hat{S}^{(\mathrm{P})}(f'_k) = \frac{N}{4\,\Delta_\mathrm{t}} \left(\hat{A}^2 + \hat{B}^2\right). \tag{588a}$$

(b) Equation (526a) excludes the zero and Nyquist frequencies. For these two cases, develop expressions that relate the periodogram to a residual sum of squares from an appropriate regression model.

[10.7] For the time series $\{y_t\}$ of length $N = 16$ given by Equation (530), compute and plot the periodogram $\hat{S}^{(\mathrm{P})}(f'_j)$ versus $f'_j = j/1024$ for $j = 1, 2, \ldots, 511$ (note that this ignores the zero and Nyquist frequencies since $0 < f'_j < 1/2$). With f_1 replaced by f'_j in Equation (526b), do the same for the regression-based $\widetilde{S}(f'_j)$. At what frequencies do we have $\widetilde{S}(f'_j) = \hat{S}^{(\mathrm{P})}(f'_j)$? How does

$$\sum_{j=1}^{511} \widetilde{S}(f'_j) \text{ compare with } \sum_{j=1}^{511} \hat{S}^{(\mathrm{P})}(f'_j)?$$

[10.8] This exercise looks into three key properties of the Lomb–Scargle periodogram.

(a) Show that the residual sum of squares given by Equation (528c) does not depend on the constant c. Hint: recall the standard result that the residual sum of squares can be expressed as $\boldsymbol{X}^T[\boldsymbol{I}_N - \boldsymbol{H}_c(\boldsymbol{H}_c^T\boldsymbol{H}_c)^{-1}\boldsymbol{H}_c^T]\boldsymbol{X}$ (see, e.g., Weisberg, 2014), and show that $\boldsymbol{H}_c = \boldsymbol{H}_0\boldsymbol{C}$ for a suitably defined 2×2 matrix $\boldsymbol{C}$. (As usual, $\boldsymbol{I}_N$ is the $N \times N$ identity matrix.)

(b) Verify Equation (529c).

(c) Show that the least squares estimators $\hat{A}$ and $\hat{B}$ minimizing $\mathrm{SS}_{\tilde{c}}(A, B)$ are uncorrelated.

[10.9] (a) Reformulate Figure 536 so that it shows direct spectral estimates based upon the Hanning and Slepian tapers with $NW = 6$ and $NW = 8$ rather than the default and Slepian tapers with $NW = 2$ and $NW = 4$. Comment on how these three new estimates compare with the ones shown in Figure 536.

(b) Generate a realization of length $N = 256$ from the following process, which, like the process of Equation (535b), is a special case of Equation (512b):

$$X_t = \cos(0.311\pi t + \phi_1) + \cos(0.3167\pi t + \phi_2) + \epsilon_t, \tag{588b}$$

where $\{\epsilon_t\}$ is a white noise process with variance $\sigma_\epsilon^2 = 10^{-6} = -60\,\mathrm{dB}$. Using both linear and decibel scales, compute and plot direct spectral estimates for the time series based upon the Hanning data taper and Slepian data tapers with $NW = 1$, 2 and 4. How well do these estimates indicate the presence of the two sinusoidal components in the process, and how well do they portray the background continuum?

[10.10] (a) Verify Equation (537a).

(b) If U and V are two RVs with a bivariate distribution, the law of total expectation says that $E\{U\} = E\{E\{U \mid V\}\}$ (for a precise statement, see, e.g., Chung, 1974, section 9.1, Priestley, 1981, p. 75, or Rao, 1973, section 2b.3). Use this law with Equation (537a) to verify that $E\{\hat{S}^{(\mathrm{P})}(f)\}$ is as stated in Equation (536).

(c) For $\sigma^2 = 10^{-4}$, $\Delta_\mathrm{t} = 1$, $D_1 = 1$, $f_1 = 0.0725$ and $N = 64$, determine – via computations – the value $f \in [f_1 - \frac{1}{N}, f_1 + \frac{1}{N}]$ such that $E\{\hat{S}^{(\mathrm{P})}(f) \mid \phi_1\}$ of Equation (537a) is maximized when the phase is set to each of 360 different values, namely, $\phi_1 = -\pi + \frac{k\pi}{180}$, $k = 0, 1, \ldots, 359$. Plot these peak frequencies versus ϕ_1 for comparison with f_1, paying particular attention to the settings $\phi_1 = -\pi/3$, 0 and $5\pi/12$. Comment briefly.

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

(d) Create a figure similar to Figure 538b, but this time for the single-frequency model $X_t = \cos\left(2\pi f_1 t - \pi/3\right) + \epsilon_t$, which is the same as the model of Equation (537b), but has a different fixed phase. Do the same for the model $X_t = \cos(2\pi f_1 t) + \epsilon_t$. Comment briefly on how these two new scatter plots compare with the one in Figure 538b.

(e) Create a figure similar to Figure 538b, but this time drawing 1000 realizations from the single-frequency model $X_t = \cos(2\pi f_1 t + \phi_1) + \epsilon_t$, where each realization of X_t, $t = 0, 1, \ldots, 63$, is created using a different realization of the uniformly distributed ϕ_1. Briefly compare this new scatter plot with the one in Figure 538b and the ones considered in part (d).

[10.11] Verify Equation (540a).

[10.12] Equations (540d) and (540e) give approximations $\tilde{g}_F$ and $\hat{g}_F$ for the critical level g_F of Fisher's g test. For significance levels $\alpha = 0.01$, 0.02, 0.05 and 0.1 in combination with sample sizes $N = 21$, 101, 501, 1001, 2001, 3001 and 4001, compare these approximations by computing $\mathbf{P}[g > \tilde{g}_F]$ and $\mathbf{P}[g > \hat{g}_F]$ using Equation (540b) to see which is closer to the selected α. Comment upon your findings.

[10.13] This exercise compares the performance of three tests for white noise: the cumulative periodogram test (see Equation (215a) and the discussion following it), Fisher's test (Equation (539b)) and Siegel's test with $\lambda = 0.6$ (Equation (541b)). For each of the nine power-of-two sample sizes N used in Table 542, generate a realization of a portion $X_{j,0}, X_{j,1}, \ldots, X_{j,N-1}$ of each of the following four stationary processes:

(a) $\{X_{1,t}\}$, a Gaussian white noise process with zero mean and unit variance;

(b) $\{X_{2,t}\}$, an AR(2) process dictated by Equation (34) – see Exercise [597] for a description of how to generate a realization from this process;

(c) $\{X_{3,t}\}$, a harmonic process with additive white noise $\{\epsilon_t\}$ dictated by $X_{3,t} = 0.5\cos(\pi t/2 + \phi) + \epsilon_t$, where ϕ is uniformly distributed over $(-\pi, \pi]$, while $\{\epsilon_t\}$ is Gaussian with zero mean and unit variance (this is a special case of Equation (512b) with $\mu = 0$, $L = 1$, $D_1 = 0.1$, $f_1 = 1/4$, $\Delta_t = 1$ and $\phi_1 = \phi$); and

(d) $\{X_{4,t}\}$, another harmonic process with additive white noise $\{\epsilon_t\}$, but now given by $X_{4,t} = 0.25\cos(0.2943\pi t + \phi_1) + 0.25\cos(0.3333\pi t + \phi_2) + \epsilon_t$, where ϕ_1 and ϕ_2 are independent and uniformly distributed over $(-\pi, \pi]$, while $\{\epsilon_t\}$ is again Gaussian with zero mean and unit variance (this is a special case of Equation (512b) with $L = 2$).

For each realization of $\{X_{j,t}\}$ for the nine sample sizes, compute the cumulative periodogram test statistic D, Fisher's test statistic g and Siegel's test statistic $T_{0.6}$, and use these three statistics individually to evaluate the null hypothesis of white noise at both the $\alpha = 0.05$ and $\alpha = 0.01$ levels of significance (use Equation (215b) to determine the critical levels for D; Equation (540d), for g; and the "Exact" columns of Table 542, for $T_{0.6}$). Repeat everything a large number N_R of times (use different realizations for all four $\{X_{j,t}\}$ each time, and take "large" to mean something from 1000 up to 100,000). For each $\{X_{j,t}\}$ for the nine samples sizes, count the number of times out of N_R that each of the three white noise tests rejected the null hypothesis of white noise at the two settings for α. Comment upon your findings.

[10.14] (a) Using the time series $\{X_t\}$ whose periodogram is displayed in Figure 534(c) (downloadable in the file `ts-km-noisy-128.txt` from the "Data" part of the website for the book or recreatable using `R` code from the site), compute ACLS and ECLS estimates of A_l, B_l, the jumps in the integrated spectrum and σ_ϵ^2 for the model

$$X_t = \sum_{l=1}^{3} \left[A_l \cos(2\pi \hat{f}_l t) + B_l \sin(2\pi \hat{f}_l t)\right] + \epsilon_t, \tag{589}$$

where $\hat{f}_1 = 0.1326075$, $\hat{f}_2 = 0.1883504$ and $\hat{f}_3 = 0.334064$ (these are the maximizing values – rounded to seven decimal places – for the three largest peaks in the periodogram – see the discussion in C&E [2] for Section 10.6); as usual, σ_ϵ^2 is the variance of the white noise process $\{\epsilon_t\}$. Compute the periodogram over the grid of frequencies $k/1024$, $k = 0, 1, \ldots, 512$, for the residuals $\{R_t\}$ corresponding to the ACLS estimates (see Equation (550b)), and plot it for comparison with the periodogram shown in Figure 551, which is for the residuals for

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

EULS estimates. Do the same for the residuals for the ECLS estimates. For both sets of residuals, perform Fisher's test (Equation (539b)) and the normalized cumulative periodogram test (described following Equation (215a)). Comment upon your findings.

(b) For the same time series as considered in part (a), compute the jumps $|\hat{C}_l|^2$ in the integrated spectrum based upon the multitaper estimates $\hat{C}_l$ of Equation (551a) by employing the following eight sets of tapers: Slepian tapers with $NW = 2$ in combination first with $K = 1$ and then with $K = 3$; Slepian tapers with $NW = 4$ and $K = 1$, 3 and 5; and sinusoidal tapers with $K = 1$, 3 and 5. Compare the jump estimates with the ones found in part (a) and with what Equation (530) suggests. Use the multitaper approach to estimate the white noise variance via Equation (551b). Comment on how these eight estimates of σ_ϵ^2 compare with the ACLS and ECLS estimates found in part (a).

(c) The time series in the file `ts-km-noisy-16384.txt` that is accessible from the "Data" part of the website for the book (or recreatable using `R` code from the site) was created in a manner similar to the series of interest in parts (a) and (b), but is of length $N = 2^{14} = 16384$ rather than $N = 2^7 = 128$. Using this longer series, compute ACLS and ECLS estimates of the jumps in the integrated spectrum and σ_ϵ^2 for the model of Equation (589), but with $\hat{f}_1$, $\hat{f}_2$ and $\hat{f}_3$ replaced by $f_1 = 1/7.5$, $f_2 = 1/5.3$ and $f_3 = 1/3$ as suggested by Equation (530). Compute the jumps $|\hat{C}_l|^2$ in the integrated spectrum based upon the multitaper estimates $\hat{C}_l$ of Equation (551a) by employing the same eight sets of tapers as used in part (b). Compute the corresponding estimates of the white noise variance via Equation (551b). Comment on how all these estimates compare with the ones obtained in parts (a) and (b) for the shorter time series.

[10.15] Here we consider creating a 95% CI for a jump J in the integrated spectrum based upon the jump estimator $\hat{J}$ whose distribution is stated by Equation (552). Generate a large number, say M, of time series from the model

$$X_t = D\cos(2\pi f t + \phi) + \epsilon_t, \quad t = 0, 1, \ldots, N-1,$$

where $D^2 = 1$; $f = 1/8$; ϕ is an RV uniformly distributed over the interval $(-\pi, \pi]$ (a different RV independently chosen for each of the M time series); $\{\epsilon_t\}$ is Gaussian white noise with mean zero and variance $\sigma_\epsilon^2 = 1$ (independently chosen for each of the M time series); and N is taken to be first 128, then 512, 2048, 8192 and finally 32,768. Noting this model to be a special case of Equations (512b) and (513) with $\mu = 0$, $A^2 + B^2 = D^2 = 1$ and $\Delta_t = 1$,

(a) compute the least squares estimators $\hat{A}$ and $\hat{B}$ of A and B for each $\{X_t\}$;

(b) form the corresponding jump estimator $\hat{J} = (\hat{A}^2 + \hat{B}^2)/4$;

(c) count the number of cases out of M for which $\hat{J}$ is trapped by the interval

$$\left[\frac{\hat{\sigma}_\epsilon^2 F_{2,N-2,\beta}(0.025)}{N}, \frac{\hat{\sigma}_\epsilon^2 F_{2,N-2,\beta}(0.975)}{N}\right] \tag{590a}$$

with β set to $N(A^2 + B^2)/(2\sigma_\epsilon^2) = N/2$;

(d) count the number of cases out of M for which $J = (A^2 + B^2)/4 = 1/4$ is trapped by the interval whose end points are given by

$$\hat{J} \pm \frac{\hat{\sigma}_\epsilon^2 \left[F_{2,N-2,\beta}(0.975) - F_{2,N-2,\beta}(0.025)\right]}{2N}, \tag{590b}$$

with β set, as before, to $N/2$; and

(e) count the number of cases out of M for which $J = 1/4$ is trapped by the interval whose end points are given by Equation (590b) with β now set to the obvious estimator $\hat{\beta} = N(\hat{A}^2 + \hat{B}^2)/(2\hat{\sigma}_\epsilon^2)$.

Repeat (a) through (e) with $D^2 = A^2 + B^2$ changed from 1 to 0.1 and then also to 10. Finally repeat everything, but with f first changed to $1/4$ and then to $3/8$. Comment upon your findings.

[10.16] Given $f \in (0, 1/(2\Delta_t)]$, determine what conditions we need to impose on the real-valued constants x_0 and x_1 so that we can write $x_0 = D\cos(\phi)$ and $x_1 = D\cos(2\pi f \Delta_t + \phi)$ for some $D > 0$

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

and $\phi \in (-\pi, \pi]$. As four concrete examples, suppose $\Delta_t = 1$, $f = 1/16$, $x_0 = \pm 2$ and $x_1 = \pm 2$, and determine the corresponding D and ϕ in each case if there be such. Use Equation (553a) to form x_t, $t = 2, 3, \ldots, 16$, for the four cases. Plot x_t, $t = 0, 1, \ldots, 16$, versus t for each case. As four additional examples, repeat everything, but now with f assumed to be $1/2$.

[10.17] Using Equation (553b) and the initial conditions $X_0 = D\cos(\phi)$ and $X_1 = D\cos(2\pi f\,\Delta_t + \phi)$, show that $\text{cov}\{X_{t+\tau}, X_t\} = D^2\cos(2\pi f \tau\,\Delta_t)/2$.

[10.18] Show that Equation (554a) provides a solution to Equation (553d) under the stipulation that the coefficients $\{\varphi_{2p,k}\}$ in the difference equation satisfy Equation (553c) when z is set to $z_l = \exp(\pm i2\pi f_l\,\Delta_t)$, $l = 1, \ldots, p$.

[10.19] Given that the roots $\{z_j\}$ of polynomial Equation (553c) are all on the unit circle and occur in conjugate pairs, show that we must have $\varphi_{2p,2p} = -1$. For the case $p = 2$, merge this result with Equation (554b) to form a three-dimensional matrix equation that can be used to solve for $\varphi_{4,1}$, $\varphi_{4,2}$ and $\varphi_{4,3}$.

[10.20] Show that an AR(p) SDF plus a white noise SDF yields an ARMA(p, p) SDF. As a concrete example find the coefficients for the ARMA(2,2) process obtained by adding white noise with unit variance to the AR(2) process defined by Equation (34). Plot the SDFs for the AR(2) and ARMA(2,2) processes for comparison.

[10.21] Figure 557 shows peak frequencies estimated using 50 realizations – each of sample size $N = 64$ – from the process of Equation (537b) (the actual peak frequency is $f_1 = 0.0725$). The peak frequencies shown on the horizontal axis are estimated using the AR forward/backward least squares (FBLS) method with order $p = 16$; i.e., the coefficients from a fitted AR(16) model are used to compute the corresponding SDF estimate, and the peak frequency is the location of the largest value in this estimate. The "Data" part of the website for the book has 1000 downloadable realizations from this process (the first 50 of these are the ones used in Figure 557). For each realization, use the FBLS AR method of orders $p = 2, 3, \ldots, 32$ to obtain 31 different estimates of the peak frequency. For a given p, compute the sample variance and MSE of the 1000 estimates, and plot these quantities versus p (use a logarithm scale for the vertical axis). Comment upon your findings.

[10.22] Verify Equation (561a).

[10.23]
(a) Verify Equation (561d), which assumes that the sample size N is odd. Derive a corresponding expression for $\hat{s}_\tau^{(\mathrm{P})}$ when N is even.
(b) Using the biased ACVS estimator $\hat{s}_\tau^{(\mathrm{P})}$ of Equation (561d), set $N = 9$, and compute and plot AR(p) SDF estimates based upon the Yule–Walker method for $p = 2, 3, \ldots, 8$ assuming that $\Delta_t = 1$. Comment upon your findings.
(c) Repeat part (b), but now use the expression you found for $\hat{s}_\tau^{(\mathrm{P})}$ in part (a) for even sample sizes with $N = 10$ and $p = 2, 3, \ldots, 9$.
(d) Repeat part (b), but now focusing on the cases $N = 15$ with $p = 14$; $N = 31$ with $p = 30$; $N = 63$ with $p = 62$; and, finally, $N = 127$ with $p = 126$.
(e) Repeat part (b), but now set $N = 127$, and consider $p = 74$, 76, 78 and 80.
(f) For the time series considered in parts (a) to (e), i.e., $x_t = \cos(\pi t/2)$, $t = 0, 1, \ldots, N-1$, with $\Delta_t = 1$, show that, for all $N \geq 3$, use of either the forward/backward least squares method or Burg's algorithm yields estimates of the spectrum that, in contrast to the Yule–Walker method, do not suffer from line splitting. Hint: show that $x_t = \varphi_{2,1}x_{t-1} + \varphi_{2,2}x_{t-2}$ (with $\varphi_{2,1}$ and $\varphi_{2,2}$ appropriately chosen), and study the conclusions drawn from Equations (555c) and (556a).

[10.24] With $m \stackrel{\text{def}}{=} 2(N - p')$ and $n = p'$, derive Equation (565b) by considering the eigenvectors and eigenvalues of the $(m+n) \times (m+n)$ matrix

$$\boldsymbol{B} \stackrel{\text{def}}{=} \begin{bmatrix} 0 & \boldsymbol{A} \\ \boldsymbol{A}^H & 0 \end{bmatrix}.$$

[10.25] Show that Equation (569a) can be rewritten as Equation (569b). What happens if $\Delta_t = 1/5$ instead of $1/12$?

[10.26] Using the Willamette River data (downloadable from the "Data" part of the website for the book), compute a direct spectral estimate based upon an $NW = 4/\Delta_t$ Slepian data taper. Plot the estimate

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press

in a manner similar to Figure 568b to verify that the periodogram in that figure does not suffer from any apparent leakage (particularly at the second and higher harmonics of 1 cycle/year).

[10.27] Substitute appropriate estimates for the unknown quantities in the expression for the variance of a periodogram estimate of frequency in Equation (526c) to determine roughly whether the periodogram-based estimated frequencies $\hat{f}_1 \doteq 1.00319$ cycles/year and $\hat{f}_2 \doteq 2.00339$ cycles/year for the Willamette River data include – within their two standard deviation limits – the values 1 cycle/year and 2 cycles/year (see Table 570 and the discussion surrounding it). Are the corresponding AR(p) estimates in Table 578 also within these two standard deviation limits?

[10.28] Figure 575 shows Thomson's F-test and corresponding reshaped multitaper estimate for the centered Willamette River data using $K = 5$ Slepian multitapers with $NW = 4/\Delta_t$. Redo this figure using $K = 5$ sinusoidal multitapers and then using $K = 6$ tapers. Comment on how well the three F-tests agree with one other.

[10.29] The top two plots of Figure 577 show Burg-based AR(150) and AR(27) SDF estimates for the Willamette River data of Figure 568a after centering, and the top two rows of Table 578 summarize some properties of these estimates. Create similar plots and summaries using (a) the Yule–Walker method and (b) the forward/backward least squares method for estimation of the AR SDFs. Comment on how these SDF estimates compare with those based on Burg's algorithm.

[10.30] (a) Verify Equation (579b).

(b) Compute and plot the phase function and the squared gain function for the filter $\{c_u g_u\}$, where $\{g_u\}$ and $\{c_u\}$ are given by, respectively, Equations (579b) and (579c). How well do these functions match up the ones corresponding to the transfer function of Equation (579a)?

[10.31] For our SVD analysis of the Willamette River time series in Section 10.15, we used a complex-valued analytic series whose real and imaginary parts were, respectively, a portion of the original series and a corresponding filtered version of the series (the filter was an approximation to the Hilbert transform). Here we explore two alternatives to the analytic series (the Willamette River data are downloadable from the "Data" part of the website for the book).

(a) A way to create a complex-valued time series that is usefully related to a real-valued series is via complex demodulation and low-pass filtering (see Tukey, 1961, Hasan, 1983, Bloomfield, 2000 or Stoica and Moses, 2005). To do so here, take the centered Willamette River series $\{X_t\}$ and form the demodulated series $Z_t = X_t e^{-i2\pi f_0 t \Delta_t}$, $t = 0, 1, \ldots, N-1$, where $f_0 = 1.5$ cycles/year (since $\Delta_t = 1/12$ year, the unitless demodulating frequency is $f_0 \Delta_t = 1/8$). Next filter $\{Z_t\}$ using a low-pass filter $\{c_u g_{I,u}\}$ with 51 nonzero coefficients to obtain

$$Z'_t = \sum_{u=-25}^{25} c_u g_{I,u} Z_{t-u}, \quad t = 25, 26, \ldots, N-26,$$

where $\{g_{I,u}\}$ is the ideal low-pass filter of Equation (152b) with W set to $1/8$, while $\{c_u\}$ are the convergence factors specified by Equation (579c). Finally use the remodulated filtered series $\{Z'_t e^{i2\pi f_0 t \Delta_t}\}$ in place of the analytic series to create plots analogous to Figures 580a and 580b. How well do the two roots closest to the unit circle gotten using the analytic series agree with those from $\{Z'_t e^{i2\pi f_0 t \Delta_t}\}$? What is the rationale behind the choices $f_0 = 1.5$ cycles/year and $W = 1/8$ and behind the choice to remodulate (hint: recall Exercise [5.13a])?

(b) Since a real-valued series is a special case of a complex-valued one, there is nothing preventing the application of the theory presented in Section 10.15 to a real-valued series. Do so by taking the centered Willamette River series $\{X_t\}$ and subjecting it to the same SVD analysis as was done using the analytic series, but set $p = 6$ rather than $p = 3$. Comment upon your findings.

https://doi.org/10.1017/9781139235723.011 Published online by Cambridge University Press