

## Chapter 4

---

### Spectral Analysis and Filtering

In this chapter, we focus on the *frequency domain* approach to time series analysis. We argue that the concept of regularity of a series can best be expressed in terms of periodic variations of the underlying phenomenon that produced the series. Many of the examples in Section 1.1 are time series that are driven by periodic components. For example, the speech recording in Figure 1.3 contains a complicated mixture of frequencies related to the opening and closing of the glottis. The monthly SOI displayed in Figure 1.5 contains two periodicities, a seasonal periodic component of 12 months and an El Niño component of about three to seven years. Of fundamental interest is the return period of the El Niño phenomenon, which can have profound effects on local climate.

An important part of analyzing data in the frequency domain, as well as the time domain, is the investigation and exploitation of the properties of the time-invariant linear filter. This special linear transformation is used similarly to linear regression in conventional statistics, and we use many of the same terms in the time series context.

We also introduce coherency as a tool for relating the common periodic behavior of two series. Coherency is a frequency based measure of the correlation between two series at a given frequency, and we show later that it measures the performance of the best linear filter relating the two series.

Many frequency scales will often coexist, depending on the nature of the problem. For example, in the Johnson & Johnson data set in Figure 1.1, the predominant frequency of oscillation is one cycle per year (4 quarters), or  $\omega = .25$  cycles per observation. The predominant frequency in the SOI and fish populations series in Figure 1.5 is also one cycle per year, but this corresponds to 1 cycle every 12 months, or  $\omega = .083$  cycles per observation. Throughout the text, we measure frequency,  $\omega$ , at cycles per time point rather than the alternative  $\lambda = 2\pi\omega$  that would give radians per point. Of descriptive interest is the *period* of a time series, defined as the number of points in a cycle, i.e.,  $1/\omega$ . Hence, the predominant period of the Johnson & Johnson series is  $1/.25$  or 4 quarters per cycle, whereas the predominant period of the SOI series is 12 months per cycle.

## 4.1 Cyclical Behavior and Periodicity

We have already encountered the notion of periodicity in numerous examples in Chapters 1, 2 and 3. The general notion of periodicity can be made more precise by introducing some terminology. In order to define the rate at which a series oscillates, we first define a *cycle* as one complete period of a sine or cosine function defined over a unit time interval. As in (1.5), we consider the periodic process

$$x_t = A \cos(2\pi\omega t + \phi) \quad (4.1)$$

for  $t = 0, \pm 1, \pm 2, \dots$ , where  $\omega$  is a *frequency* index, defined in cycles per unit time with  $A$  determining the height or *amplitude* of the function and  $\phi$ , called the *phase*, determining the start point of the cosine function. We can introduce random variation in this time series by allowing the amplitude and phase to vary randomly.

As discussed in Example 2.10, for purposes of data analysis, it is easier to use a trigonometric identity<sup>4.1</sup> and write (4.1) as

$$x_t = U_1 \cos(2\pi\omega t) + U_2 \sin(2\pi\omega t), \quad (4.2)$$

where  $U_1 = A \cos \phi$  and  $U_2 = -A \sin \phi$  are often taken to be normally distributed random variables. In this case, the amplitude is  $A = \sqrt{U_1^2 + U_2^2}$  and the phase is  $\phi = \tan^{-1}(-U_2/U_1)$ . From these facts we can show that if, and only if, in (4.1),  $A$  and  $\phi$  are independent random variables, where  $A^2$  is chi-squared with 2 degrees of freedom, and  $\phi$  is uniformly distributed on  $(-\pi, \pi)$ , then  $U_1$  and  $U_2$  are independent, standard normal random variables (see Problem 4.3).

If we assume that  $U_1$  and  $U_2$  are uncorrelated random variables with mean 0 and variance  $\sigma^2$ , then  $x_t$  in (4.2) is stationary with mean  $E(x_t) = 0$  and, writing  $c_t = \cos(2\pi\omega t)$  and  $s_t = \sin(2\pi\omega t)$ , autocovariance function

$$\begin{aligned} \gamma_x(h) &= \text{cov}(x_{t+h}, x_t) = \text{cov}(U_1 c_{t+h} + U_2 s_{t+h}, U_1 c_t + U_2 s_t) \\ &= \text{cov}(U_1 c_{t+h}, U_1 c_t) + \text{cov}(U_1 c_{t+h}, U_2 s_t) \\ &\quad + \text{cov}(U_2 s_{t+h}, U_1 c_t) + \text{cov}(U_2 s_{t+h}, U_2 s_t) \\ &= \sigma^2 c_{t+h} c_t + 0 + 0 + \sigma^2 s_{t+h} s_t = \sigma^2 \cos(2\pi\omega h), \end{aligned} \quad (4.3)$$

using Footnote 4.1 and noting that  $\text{cov}(U_1, U_2) = 0$ . From (4.3), we see that

$$\text{var}(x_t) = \gamma_x(0) = \sigma^2.$$

Thus, if we observe  $U_1 = a$  and  $U_2 = b$ , an estimate of  $\sigma^2$  is the sample variance of these two observations, which in this case is simply  $S^2 = \frac{a^2+b^2}{2-1} = a^2 + b^2$ .

The random process in (4.2) is function of its frequency,  $\omega$ . For  $\omega = 1$ , the series makes one cycle per time unit; for  $\omega = .50$ , the series makes a cycle every two time units; for  $\omega = .25$ , every four units, and so on. In general, for data that occur at discrete time points, we will need at least two points to determine a cycle, so the

<sup>4.1</sup>  $\cos(\alpha \pm \beta) = \cos(\alpha) \cos(\beta) \mp \sin(\alpha) \sin(\beta)$ .

highest frequency of interest is .5 cycles per point. This frequency is called the *folding frequency* and defines the highest frequency that can be seen in discrete sampling. Higher frequencies sampled this way will appear at lower frequencies, called *aliases*; an example is the way a camera samples a rotating wheel on a moving automobile in a movie, in which the wheel appears to be rotating at a different rate, and sometimes backwards (the *wagon wheel effect*). For example, most movies are recorded at 24 frames per second (or 24 Hertz). If the camera is filming a wheel that is rotating at 24 Hertz, the wheel will appear to stand still.

Consider a generalization of (4.2) that allows mixtures of periodic series with multiple frequencies and amplitudes,

$$x_t = \sum_{k=1}^q [U_{k1} \cos(2\pi\omega_k t) + U_{k2} \sin(2\pi\omega_k t)], \quad (4.4)$$

where  $U_{k1}, U_{k2}$ , for  $k = 1, 2, \dots, q$ , are uncorrelated zero-mean random variables with variances  $\sigma_k^2$ , and the  $\omega_k$  are distinct frequencies. Notice that (4.4) exhibits the process as a sum of uncorrelated components, with variance  $\sigma_k^2$  for frequency  $\omega_k$ . As in (4.3), it is easy to show (Problem 4.4) that the autocovariance function of the process is

$$\gamma_x(h) = \sum_{k=1}^q \sigma_k^2 \cos(2\pi\omega_k h), \quad (4.5)$$

and we note the autocovariance function is the sum of periodic components with weights proportional to the variances  $\sigma_k^2$ . Hence,  $x_t$  is a mean-zero stationary process with variance

$$\gamma_x(0) = \text{var}(x_t) = \sum_{k=1}^q \sigma_k^2, \quad (4.6)$$

exhibiting the overall variance as a sum of variances of each of the component parts.

As in the simple case, if we observe  $U_{k1} = a_k$  and  $U_{k2} = b_k$  for  $k = 1, \dots, q$ , then an estimate of the  $k$ th variance component,  $\sigma_k^2$ , of  $\text{var}(x_t)$ , would be the sample variance  $S_k^2 = a_k^2 + b_k^2$ . In addition, an estimate of the total variance of  $x_t$ , namely,  $\gamma_x(0)$  would be the sum of the sample variances,

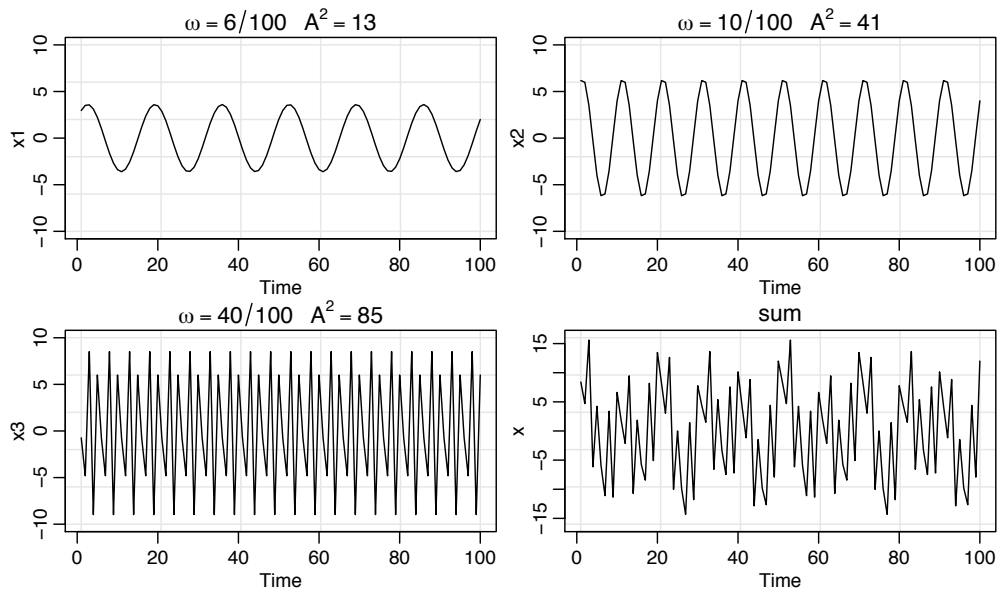
$$\hat{\gamma}_x(0) = \hat{\text{var}}(x_t) = \sum_{k=1}^q (a_k^2 + b_k^2). \quad (4.7)$$

Hold on to this idea because we will use it in Example 4.2.

#### Example 4.1 A Periodic Series

Figure 4.1 shows an example of the mixture (4.4) with  $q = 3$  constructed in the following way. First, for  $t = 1, \dots, 100$ , we generated three series

$$\begin{aligned} x_{t1} &= 2 \cos(2\pi t 6/100) + 3 \sin(2\pi t 6/100) \\ x_{t2} &= 4 \cos(2\pi t 10/100) + 5 \sin(2\pi t 10/100) \\ x_{t3} &= 6 \cos(2\pi t 40/100) + 7 \sin(2\pi t 40/100) \end{aligned}$$



**Fig. 4.1.** Periodic components and their sum as described in Example 4.1.

These three series are displayed in Figure 4.1 along with the corresponding frequencies and squared amplitudes. For example, the squared amplitude of  $x_{t1}$  is  $A^2 = 2^2 + 3^2 = 13$ . Hence, the maximum and minimum values that  $x_{t1}$  will attain are  $\pm\sqrt{13} = \pm 3.61$ .

Finally, we constructed

$$x_t = x_{t1} + x_{t2} + x_{t3}$$

and this series is also displayed in Figure 4.1. We note that  $x_t$  appears to behave as some of the periodic series we saw in Chapters 1 and 2. The systematic sorting out of the essential frequency components in a time series, including their relative contributions, constitutes one of the main objectives of spectral analysis. The R code to reproduce Figure 4.1 is

```
x1 = 2*cos(2*pi*1:100*6/100) + 3*sin(2*pi*1:100*6/100)
x2 = 4*cos(2*pi*1:100*10/100) + 5*sin(2*pi*1:100*10/100)
x3 = 6*cos(2*pi*1:100*40/100) + 7*sin(2*pi*1:100*40/100)
x = x1 + x2 + x3
par(mfrow=c(2,2))
plot.ts(x1, ylim=c(-10,10), main=expression(omega==6/100~~~A^2==13))
plot.ts(x2, ylim=c(-10,10), main=expression(omega==10/100~~~A^2==41))
plot.ts(x3, ylim=c(-10,10), main=expression(omega==40/100~~~A^2==85))
plot.ts(x, ylim=c(-16,16), main="sum")
```

The model given in (4.4) along with the corresponding autocovariance function given in (4.5) are population constructs. Although, in (4.7), we hinted as to how we would estimate the variance components, we now discuss the practical aspects of how, given data  $x_1, \dots, x_n$ , to actually estimate the variance components  $\sigma_k^2$  in (4.6).

**Example 4.2 Estimation and the Periodogram**

For any time series sample  $x_1, \dots, x_n$ , where  $n$  is odd, we may write, *exactly*

$$x_t = a_0 + \sum_{j=1}^{(n-1)/2} [a_j \cos(2\pi t j/n) + b_j \sin(2\pi t j/n)], \tag{4.8}$$

for  $t = 1, \dots, n$  and suitably chosen coefficients. If  $n$  is even, the representation (4.8) can be modified by summing to  $(n/2 - 1)$  and adding an additional component given by  $a_{n/2} \cos(2\pi t \frac{1}{2}) = a_{n/2}(-1)^t$ . The crucial point here is that (4.8) is exact for any sample. Hence (4.4) may be thought of as an approximation to (4.8), the idea being that many of the coefficients in (4.8) may be close to zero.

Using the regression results from Chapter 2, the coefficients  $a_j$  and  $b_j$  are of the form  $\sum_{t=1}^n x_t z_{tj} / \sum_{t=1}^n z_{tj}^2$ , where  $z_{tj}$  is either  $\cos(2\pi t j/n)$  or  $\sin(2\pi t j/n)$ . Using Problem 4.1,  $\sum_{t=1}^n z_{tj}^2 = n/2$  when  $j/n \neq 0, 1/2$ , so the regression coefficients in (4.8) can be written as ( $a_0 = \bar{x}$ ),

$$a_j = \frac{2}{n} \sum_{t=1}^n x_t \cos(2\pi t j/n) \quad \text{and} \quad b_j = \frac{2}{n} \sum_{t=1}^n x_t \sin(2\pi t j/n).$$

We then define the *scaled periodogram* to be

$$P(j/n) = a_j^2 + b_j^2, \tag{4.9}$$

and it is of interest because it indicates which frequency components in (4.8) are large in magnitude and which components are small. *The scaled periodogram is simply the sample variance at each frequency component and consequently is an estimate of  $\sigma_j^2$  corresponding to the sinusoid oscillating at a frequency of  $\omega_j = j/n$ . These particular frequencies are called the *Fourier* or *fundamental frequencies*. Large values of  $P(j/n)$  indicate which frequencies  $\omega_j = j/n$  are predominant in the series, whereas small values of  $P(j/n)$  may be associated with noise. The periodogram was introduced in Schuster (1898) and used in Schuster (1906) for studying the periodicities in the sunspot series (shown in Figure 4.22).*

Fortunately, it is not necessary to run a large regression to obtain the values of  $a_j$  and  $b_j$  because they can be computed quickly if  $n$  is a highly composite integer. Although we will discuss it in more detail in Section 4.3, the *discrete Fourier transform (DFT)* is a complex-valued weighted average of the data given by<sup>4.2</sup>

$$\begin{aligned} d(j/n) &= n^{-1/2} \sum_{t=1}^n x_t \exp(-2\pi i t j/n) \\ &= n^{-1/2} \left( \sum_{t=1}^n x_t \cos(2\pi t j/n) - i \sum_{t=1}^n x_t \sin(2\pi t j/n) \right), \end{aligned} \tag{4.10}$$

<sup>4.2</sup> Euler's formula:  $e^{i\alpha} = \cos(\alpha) + i \sin(\alpha)$ . Consequently,  $\cos(\alpha) = \frac{e^{i\alpha} + e^{-i\alpha}}{2}$ , and  $\sin(\alpha) = \frac{e^{i\alpha} - e^{-i\alpha}}{2i}$ . Also,  $\frac{1}{i} = -i$  because  $-i \times i = 1$ . If  $z = a + ib$  is complex, then  $|z|^2 = z z^* = (a + ib)(a - ib) = a^2 + b^2$ ; the \* denotes conjugation.

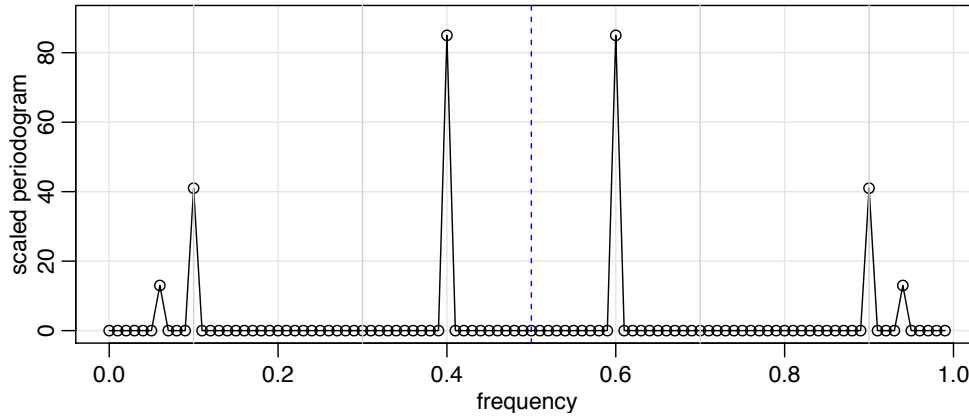


Fig. 4.2. The scaled periodogram (4.12) of the data generated in Example 4.1.

for  $j = 0, 1, \dots, n - 1$ , where the frequencies  $j/n$  are the Fourier or fundamental frequencies. Because of a large number of redundancies in the calculation, (4.10) may be computed quickly using the *fast Fourier transform (FFT)*. Note that

$$|d(j/n)|^2 = \frac{1}{n} \left( \sum_{t=1}^n x_t \cos(2\pi t j/n) \right)^2 + \frac{1}{n} \left( \sum_{t=1}^n x_t \sin(2\pi t j/n) \right)^2 \quad (4.11)$$

and it is this quantity that is called the *periodogram*. We may calculate the scaled periodogram, (4.9), using the periodogram as

$$P(j/n) = \frac{4}{n} |d(j/n)|^2. \quad (4.12)$$

The scaled periodogram of the data,  $x_t$ , simulated in Example 4.1 is shown in Figure 4.2, and it clearly identifies the three components  $x_{t1}$ ,  $x_{t2}$ , and  $x_{t3}$  of  $x_t$ . Note that

$$P(j/n) = P(1 - j/n), \quad j = 0, 1, \dots, n - 1,$$

so there is a mirroring effect at the folding frequency of  $1/2$ ; consequently, the periodogram is typically not plotted for frequencies higher than the folding frequency. In addition, note that the heights of the scaled periodogram shown in the figure are

$$P\left(\frac{6}{100}\right) = P\left(\frac{94}{100}\right) = 13, \quad P\left(\frac{10}{100}\right) = P\left(\frac{90}{100}\right) = 41, \quad P\left(\frac{40}{100}\right) = P\left(\frac{60}{100}\right) = 85,$$

and  $P(j/n) = 0$  otherwise. These are exactly the values of the squared amplitudes of the components generated in Example 4.1.

Assuming the simulated data,  $\mathbf{x}$ , were retained from the previous example, the R code to reproduce Figure 4.2 is

```
P = Mod(2*fft(x)/100)^2; Fr = 0:99/100
plot(Fr, P, type="o", xlab="frequency", ylab="scaled periodogram")
```

Different packages scale the FFT differently, so it is a good idea to consult the documentation. R computes it without the factor  $n^{-1/2}$  and with an additional factor of  $e^{2\pi i \omega_j}$  that can be ignored because we will be interested in the squared modulus.

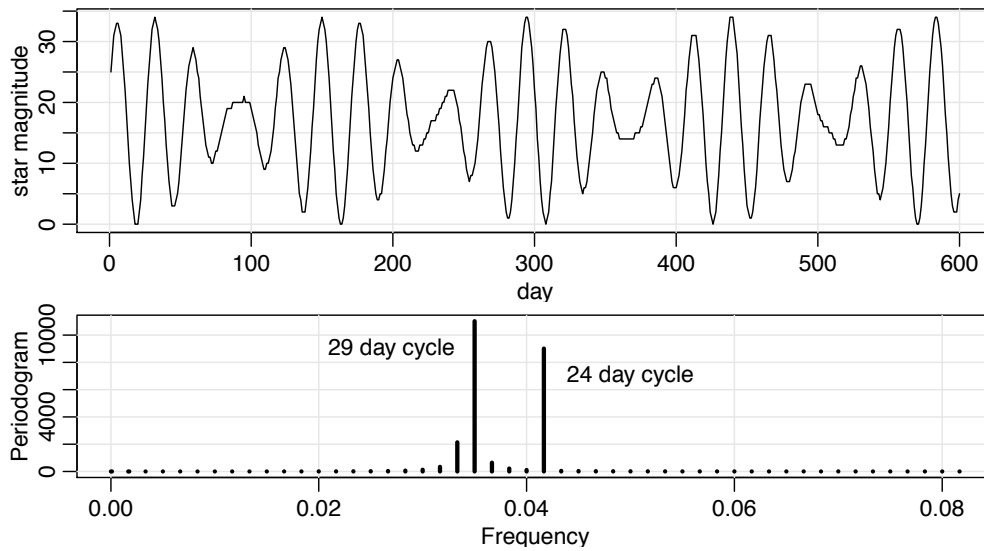


Fig. 4.3. Star magnitudes and part of the corresponding periodogram.

If we consider the data  $x_t$  in Example 4.1 as a color (waveform) made up of primary colors  $x_{t1}, x_{t2}, x_{t3}$  at various strengths (amplitudes), then we might consider the periodogram as a prism that decomposes the color  $x_t$  into its primary colors (spectrum). Hence the term *spectral analysis*. The following is an example using actual data.

### Example 4.3 Star Magnitude

The data in Figure 4.3 are the magnitude of a star taken at midnight for 600 consecutive days. The data are taken from the classic text, *The Calculus of Observations, a Treatise on Numerical Mathematics*, by E.T. Whittaker and G. Robinson, (1923, Blackie & Son, Ltd.).

The periodogram for frequencies less than .08 is also displayed in the figure; the periodogram ordinates for frequencies higher than .08 are essentially zero. Note that the 29 ( $\approx 1/.035$ ) day cycle and the 24 ( $\approx 1/.041$ ) day cycle are the most prominent periodic components of the data.

We can interpret this result as we are observing an *amplitude modulated* signal. For example, suppose we are observing signal-plus-noise,  $x_t = s_t + v_t$ , where  $s_t = \cos(2\pi\omega t)\cos(2\pi\delta t)$ , and  $\delta$  is very small. In this case, the process will oscillate at frequency  $\omega$ , but the amplitude will be modulated by  $\cos(2\pi\delta t)$ . Since  $2\cos(\alpha)\cos(\delta) = \cos(\alpha + \delta) + \cos(\alpha - \delta)$ , the periodogram of data generated as  $x_t$  will have two peaks close to each other at  $\alpha \pm \delta$ . Try this on your own:

```
t = 1:200
plot.ts(x <- 2*cos(2*pi*.2*t)*cos(2*pi*.01*t)) # not shown
lines(cos(2*pi*.19*t)+cos(2*pi*.21*t), col=2) # the same
Px = Mod(fft(x))^2; plot(0:199/200, Px, type='o') # the periodogram
```

The R code to reproduce Figure 4.3 is

```
n = length(star)
par(mfrow=c(2,1), mar=c(3,3,1,1), mgp=c(1.6,.6,0))
plot(star, ylab="star magnitude", xlab="day")
```

```

Per = Mod(fft(star-mean(star)))^2/n
Freq = (1:n -1)/n
plot(Freq[1:50], Per[1:50], type='h', lwd=3, ylab="Periodogram",
      xlab="Frequency")
u = which.max(Per[1:50])      # 22  freq=21/600=.035 cycles/day
uu = which.max(Per[1:50][-u]) # 25  freq=25/600=.041 cycles/day
1/Freq[22]; 1/Freq[26]      # period = days/cycle
text(.05, 7000, "24 day cycle"); text(.027, 9000, "29 day cycle")
### another way to find the two peaks is to order on Per
y = cbind(1:50, Freq[1:50], Per[1:50]); y[order(y[,3]),]

```

## 4.2 The Spectral Density

In this section, we define the fundamental frequency domain tool, the spectral density. In addition, we discuss the spectral representations for stationary processes. Just as the Wold decomposition ([Theorem B.5](#)) theoretically justified the use of regression for analyzing time series, the spectral representation theorems supply the theoretical justifications for decomposing stationary time series into periodic components appearing in proportion to their underlying variances. This material is enhanced by the results presented in [Appendix C](#).

### Example 4.4 A Periodic Stationary Process

Consider a periodic stationary random process given by (4.2), with a fixed frequency  $\omega_0$ , say,

$$x_t = U_1 \cos(2\pi\omega_0 t) + U_2 \sin(2\pi\omega_0 t), \quad (4.13)$$

where  $U_1$  and  $U_2$  are uncorrelated zero-mean random variables with equal variance  $\sigma^2$ . The number of time periods needed for the above series to complete one cycle is exactly  $1/\omega_0$ , and the process makes exactly  $\omega_0$  cycles per point for  $t = 0, \pm 1, \pm 2, \dots$ . Recalling (4.3) and using [Footnote 4.2](#), we have

$$\begin{aligned} \gamma(h) &= \sigma^2 \cos(2\pi\omega_0 h) = \frac{\sigma^2}{2} e^{-2\pi i\omega_0 h} + \frac{\sigma^2}{2} e^{2\pi i\omega_0 h} \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i\omega h} dF(\omega) \end{aligned}$$

using Riemann–Stieltjes integration (see [Section C.4.1](#)), where  $F(\omega)$  is the function defined by

$$F(\omega) = \begin{cases} 0 & \omega < -\omega_0, \\ \sigma^2/2 & -\omega_0 \leq \omega < \omega_0, \\ \sigma^2 & \omega \geq \omega_0. \end{cases}$$

The function  $F(\omega)$  behaves like a cumulative distribution function for a discrete random variable, except that  $F(\infty) = \sigma^2 = \text{var}(x_t)$  instead of one. In fact,  $F(\omega)$  is a cumulative distribution function, not of probabilities, but rather of variances, with  $F(\infty)$  being the total variance of the process  $x_t$ . Hence, we term  $F(\omega)$  the *spectral distribution function*. This example is continued in [Example 4.9](#).



A representation such as the one given in **Example 4.4** always exists for a stationary process. For details, see **Theorem C.1** and its proof; Riemann–Stieltjes integration is described in **Section C.4.1**.

**Property 4.1 Spectral Representation of an Autocovariance Function**

If  $\{x_t\}$  is stationary with autocovariance  $\gamma(h) = \text{cov}(x_{t+h}, x_t)$ , then there exists a unique monotonically increasing function  $F(\omega)$ , called the spectral distribution function, with  $F(-\infty) = F(-1/2) = 0$ , and  $F(\infty) = F(1/2) = \gamma(0)$  such that

$$\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi i \omega h} dF(\omega). \quad (4.14)$$

An important situation we use repeatedly is the case when the autocovariance function is absolutely summable, in which case the spectral distribution function is absolutely continuous with  $dF(\omega) = f(\omega) d\omega$ , and the representation (4.14) becomes the motivation for the property given below.

**Property 4.2 The Spectral Density**

If the autocovariance function,  $\gamma(h)$ , of a stationary process satisfies

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty, \quad (4.15)$$

then it has the representation

$$\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi i \omega h} f(\omega) d\omega \quad h = 0, \pm 1, \pm 2, \dots \quad (4.16)$$

as the inverse transform of the spectral density,

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \omega h} \quad -1/2 \leq \omega \leq 1/2. \quad (4.17)$$

This spectral density is the analogue of the probability density function; the fact that  $\gamma(h)$  is non-negative definite ensures

$$f(\omega) \geq 0$$

for all  $\omega$ . It follows immediately from (4.17) that

$$f(\omega) = f(-\omega)$$

verifying the spectral density is an even function. Because of the evenness, we will typically only plot  $f(\omega)$  for  $0 \leq \omega \leq 1/2$ . In addition, putting  $h = 0$  in (4.16) yields

$$\gamma(0) = \text{var}(x_t) = \int_{-1/2}^{1/2} f(\omega) d\omega,$$

which expresses the total variance as the integrated spectral density over all of the frequencies. We show later on, that a linear filter can isolate the variance in certain frequency intervals or *bands*.

It should now be clear that the autocovariance and the spectral distribution functions contain the same information. That information, however, is expressed in different ways. The autocovariance function expresses information in terms of lags, whereas the spectral distribution expresses the same information in terms of cycles. Some problems are easier to work with when considering lagged information and we would tend to handle those problems in the time domain. Nevertheless, other problems are easier to work with when considering periodic information and we would tend to handle those problems in the spectral domain.

We note that the autocovariance function,  $\gamma(h)$ , in (4.16) and the spectral density,  $f(\omega)$ , in (4.17) are Fourier transform pairs. In particular, this means that if  $f(\omega)$  and  $g(\omega)$  are two spectral densities for which

$$\gamma_f(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} f(\omega)e^{2\pi i\omega h} d\omega = \int_{-\frac{1}{2}}^{\frac{1}{2}} g(\omega)e^{2\pi i\omega h} d\omega = \gamma_g(h) \quad (4.18)$$

for all  $h = 0, \pm 1, \pm 2, \dots$ , then

$$f(\omega) = g(\omega). \quad (4.19)$$

Finally, the absolute summability condition, (4.15), is not satisfied by (4.5), the example that we have used to introduce the idea of a spectral representation. The condition, however, is satisfied for ARMA models.

It is illuminating to examine the spectral density for the series that we have looked at in earlier discussions.

#### Example 4.5 White Noise Series

As a simple example, consider the theoretical power spectrum of a sequence of uncorrelated random variables,  $w_t$ , with variance  $\sigma_w^2$ . A simulated set of data is displayed in the top of Figure 1.8. Because the autocovariance function was computed in Example 1.16 as  $\gamma_w(h) = \sigma_w^2$  for  $h = 0$ , and zero, otherwise, it follows from (4.17), that

$$f_w(\omega) = \sigma_w^2$$

for  $-1/2 \leq \omega \leq 1/2$ . Hence the process contains equal power at all frequencies. This property is seen in the realization, which seems to contain all different frequencies in a roughly equal mix. In fact, the name white noise comes from the analogy to white light, which contains all frequencies in the color spectrum at the same level of intensity. The top of Figure 4.4 shows a plot of the white noise spectrum for  $\sigma_w^2 = 1$ . The R code to reproduce the figure is given at the end of Example 4.7.

Since the linear process is an essential tool, it is worthwhile investigating the spectrum of such a process. In general, a linear filter uses a set of specified coefficients, say  $a_j$ , for  $j = 0, \pm 1, \pm 2, \dots$ , to transform an input series,  $x_t$ , producing an output series,  $y_t$ , of the form

$$y_t = \sum_{j=-\infty}^{\infty} a_j x_{t-j}, \quad \sum_{j=-\infty}^{\infty} |a_j| < \infty. \quad (4.20)$$

The form (4.20) is also called a *convolution* in some statistical contexts. The coefficients are collectively called the *impulse response function*, and the Fourier transform

$$A(\omega) = \sum_{j=-\infty}^{\infty} a_j e^{-2\pi i \omega j}, \quad (4.21)$$

is called the *frequency response function*. If, in (4.20),  $x_t$  has spectral density  $f_x(\omega)$ , we have the following result.

### Property 4.3 Output Spectrum of a Filtered Stationary Series

For the process in (4.20), if  $x_t$  has spectrum  $f_x(\omega)$ , then the spectrum of the filtered output,  $y_t$ , say  $f_y(\omega)$ , is related to the spectrum of the input  $x_t$  by

$$f_y(\omega) = |A(\omega)|^2 f_x(\omega), \quad (4.22)$$

where the frequency response function  $A(\omega)$  is defined in (4.21).

*Proof:* The autocovariance function of the filtered output  $y_t$  in (4.20) is

$$\begin{aligned} \gamma_y(h) &= \text{cov}(x_{t+h}, x_t) \\ &= \text{cov}\left(\sum_r a_r x_{t+h-r}, \sum_s a_s x_{t-s}\right) \\ &= \sum_r \sum_s a_r \gamma_x(h-r+s) a_s \\ &\stackrel{(1)}{=} \sum_r \sum_s a_r \left[ \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i \omega (h-r+s)} f_x(\omega) d\omega \right] a_s \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left( \sum_r a_r e^{-2\pi i \omega r} \right) \left( \sum_s a_s e^{2\pi i \omega s} \right) e^{2\pi i \omega h} f_x(\omega) d\omega \\ &\stackrel{(2)}{=} \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i \omega h} \underbrace{|A(\omega)|^2 f_x(\omega)}_{f_y(\omega)} d\omega, \end{aligned}$$

where we have, (1) replaced  $\gamma_x(\cdot)$  by its representation (4.16), and (2) substituted  $A(\omega)$  from (4.21). The result holds by exploiting the uniqueness of the Fourier transform.  $\square$

The use of Property 4.3 is explored further in Section 4.7. If  $x_t$  is ARMA, its spectral density can be obtained explicitly using the fact that it is a linear process, i.e.,  $x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$ , where  $\sum_{j=0}^{\infty} |\psi_j| < \infty$ . The following property is a direct consequence of Property 4.3, by using the additional facts that the spectral density of white noise is  $f_w(\omega) = \sigma_w^2$ , and by Property 3.1,  $\psi(z) = \theta(z)/\phi(z)$ .

**Property 4.4 The Spectral Density of ARMA**

If  $x_t$  is ARMA( $p, q$ ),  $\phi(B)x_t = \theta(B)w_t$ , its spectral density is given by

$$f_x(\omega) = \sigma_w^2 \frac{|\theta(e^{-2\pi i\omega})|^2}{|\phi(e^{-2\pi i\omega})|^2} \quad (4.23)$$

where  $\phi(z) = 1 - \sum_{k=1}^p \phi_k z^k$  and  $\theta(z) = 1 + \sum_{k=1}^q \theta_k z^k$ .

**Example 4.6 Moving Average**

As an example of a series that does not have an equal mix of frequencies, we consider a moving average model. Specifically, consider the MA(1) model given by

$$x_t = w_t + .5w_{t-1}.$$

A sample realization is shown in the top of [Figure 3.2](#) and we note that the series has less of the higher or faster frequencies. The spectral density will verify this observation.

The autocovariance function is displayed in [Example 3.5](#), and for this particular example, we have

$$\gamma(0) = (1 + .5^2)\sigma_w^2 = 1.25\sigma_w^2; \quad \gamma(\pm 1) = .5\sigma_w^2; \quad \gamma(\pm h) = 0 \text{ for } h > 1.$$

Substituting this directly into the definition given in (4.17), we have

$$\begin{aligned} f(\omega) &= \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i\omega h} = \sigma_w^2 \left[ 1.25 + .5 \left( e^{-2\pi i\omega} + e^{2\pi i\omega} \right) \right] \\ &= \sigma_w^2 [1.25 + \cos(2\pi\omega)]. \end{aligned} \quad (4.24)$$

We can also compute the spectral density using [Property 4.4](#), which states that for an MA,  $f(\omega) = \sigma_w^2 |\theta(e^{-2\pi i\omega})|^2$ . Because  $\theta(z) = 1 + .5z$ , we have

$$\begin{aligned} |\theta(e^{-2\pi i\omega})|^2 &= |1 + .5e^{-2\pi i\omega}|^2 = (1 + .5e^{-2\pi i\omega})(1 + .5e^{2\pi i\omega}) \\ &= 1.25 + .5 \left( e^{-2\pi i\omega} + e^{2\pi i\omega} \right) \end{aligned}$$

which leads to agreement with (4.24).

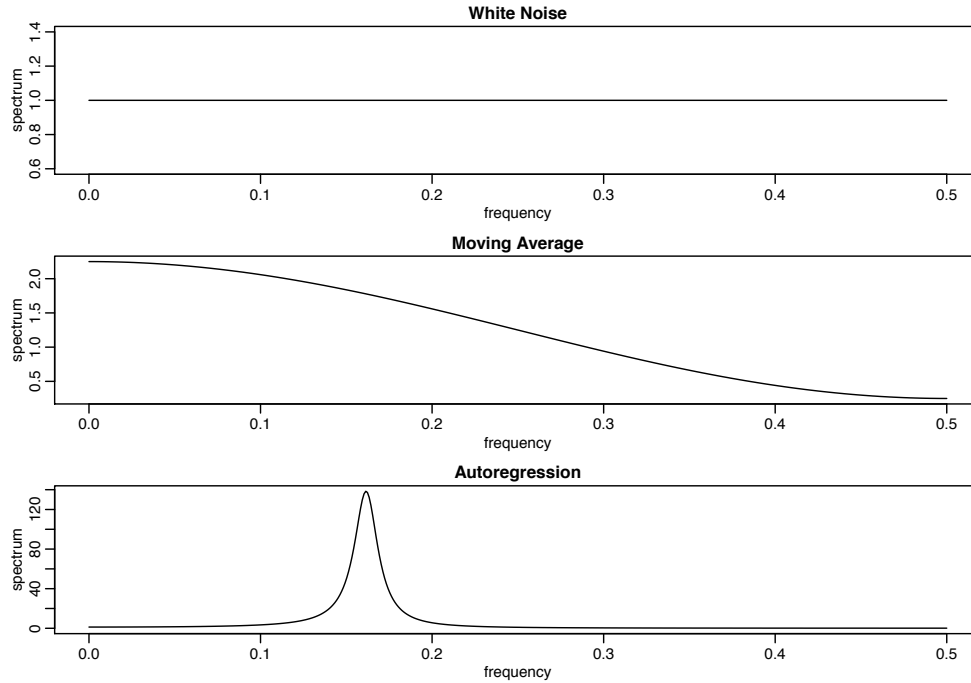
Plotting the spectrum for  $\sigma_w^2 = 1$ , as in the middle of [Figure 4.4](#), shows the lower or slower frequencies have greater power than the higher or faster frequencies.

**Example 4.7 A Second-Order Autoregressive Series**

We now consider the spectrum of an AR(2) series of the form

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} = w_t,$$

for the special case  $\phi_1 = 1$  and  $\phi_2 = -.9$ . [Figure 1.9](#) shows a sample realization of such a process for  $\sigma_w = 1$ . We note the data exhibit a strong periodic component that makes a cycle about every six points.



**Fig. 4.4.** Theoretical spectra of white noise (top), a first-order moving average (middle), and a second-order autoregressive process (bottom).

To use **Property 4.4**, note that  $\theta(z) = 1$ ,  $\phi(z) = 1 - z + .9z^2$  and

$$\begin{aligned} |\phi(e^{-2\pi i\omega})|^2 &= (1 - e^{-2\pi i\omega} + .9e^{-4\pi i\omega})(1 - e^{2\pi i\omega} + .9e^{4\pi i\omega}) \\ &= 2.81 - 1.9(e^{2\pi i\omega} + e^{-2\pi i\omega}) + .9(e^{4\pi i\omega} + e^{-4\pi i\omega}) \\ &= 2.81 - 3.8 \cos(2\pi\omega) + 1.8 \cos(4\pi\omega). \end{aligned}$$

Using this result in (4.23), we have that the spectral density of  $x_t$  is

$$f_x(\omega) = \frac{\sigma_w^2}{2.81 - 3.8 \cos(2\pi\omega) + 1.8 \cos(4\pi\omega)}.$$

Setting  $\sigma_w = 1$ , the bottom of **Figure 4.4** displays  $f_x(\omega)$  and shows a strong power component at about  $\omega = .16$  cycles per point or a period between six and seven cycles per point and very little power at other frequencies. In this case, modifying the white noise series by applying the second-order AR operator has concentrated the power or variance of the resulting series in a very narrow frequency band.

The spectral density can also be obtained from first principles, without having to use **Property 4.4**. Because  $w_t = x_t - x_{t-1} + .9x_{t-2}$  in this example, we have

$$\begin{aligned} \gamma_w(h) &= \text{cov}(w_{t+h}, w_t) \\ &= \text{cov}(x_{t+h} - x_{t+h-1} + .9x_{t+h-2}, x_t - x_{t-1} + .9x_{t-2}) \\ &= 2.81\gamma_x(h) - 1.9[\gamma_x(h+1) + \gamma_x(h-1)] + .9[\gamma_x(h+2) + \gamma_x(h-2)] \end{aligned}$$

Now, substituting the spectral representation (4.16) for  $\gamma_x(h)$  in the above equation yields

$$\begin{aligned}\gamma_w(h) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} [2.81 - 1.9(e^{2\pi i\omega} + e^{-2\pi i\omega}) + .9(e^{4\pi i\omega} + e^{-4\pi i\omega})] e^{2\pi i\omega h} f_x(\omega) d\omega \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} [2.81 - 3.8 \cos(2\pi\omega) + 1.8 \cos(4\pi\omega)] e^{2\pi i\omega h} f_x(\omega) d\omega.\end{aligned}$$

If the spectrum of the white noise process,  $w_t$ , is  $g_w(\omega)$ , the uniqueness of the Fourier transform allows us to identify

$$g_w(\omega) = [2.81 - 3.8 \cos(2\pi\omega) + 1.8 \cos(4\pi\omega)] f_x(\omega).$$

But, as we have already seen,  $g_w(\omega) = \sigma_w^2$ , from which we deduce that

$$f_x(\omega) = \frac{\sigma_w^2}{2.81 - 3.8 \cos(2\pi\omega) + 1.8 \cos(4\pi\omega)}$$

is the spectrum of the autoregressive series.

To reproduce Figure 4.4, use `arma.spec` from `astsa`:

```
par(mfrow=c(3,1))
arma.spec(log="no", main="White Noise")
arma.spec(ma=.5, log="no", main="Moving Average")
arma.spec(ar=c(1,-.9), log="no", main="Autoregression")
```

#### Example 4.8 Every Explosion has a Cause (cont)

In Example 3.4, we discussed the fact that explosive models have causal counterparts. In that example, we also indicated that it was easier to show this result in general in the spectral domain. In this example, we give the details for an AR(1) model, but the techniques used here will indicate how to generalize the result.

As in Example 3.4, we suppose that  $x_t = 2x_{t-1} + w_t$ , where  $w_t \sim \text{iid } N(0, \sigma_w^2)$ . Then, the spectral density of  $x_t$  is

$$f_x(\omega) = \sigma_w^2 |1 - 2e^{-2\pi i\omega}|^{-2}. \quad (4.25)$$

But,  $|1 - 2e^{-2\pi i\omega}| = |1 - 2e^{2\pi i\omega}| = |(2e^{2\pi i\omega})(\frac{1}{2}e^{-2\pi i\omega} - 1)| = 2|1 - \frac{1}{2}e^{-2\pi i\omega}|$ . Thus, (4.25) can be written as

$$f_x(\omega) = \frac{1}{4}\sigma_w^2 |1 - \frac{1}{2}e^{-2\pi i\omega}|^{-2},$$

which implies that  $x_t = \frac{1}{2}x_{t-1} + v_t$ , with  $v_t \sim \text{iid } N(0, \frac{1}{4}\sigma_w^2)$  is an equivalent form of the model.

We end this section by mentioning another spectral representation that deals with the process directly. In nontechnical terms, the result suggests that (4.4) is approximately true for any stationary time series, and this gives an additional theoretical justification for decomposing time series into harmonic components.

**Example 4.9 A Periodic Stationary Process (cont)**

In [Example 4.4](#), we considered the periodic stationary process given in [\(4.13\)](#), namely,  $x_t = U_1 \cos(2\pi\omega_0 t) + U_2 \sin(2\pi\omega_0 t)$ . Using [Footnote 4.2](#), we may write this as

$$x_t = \frac{1}{2}(U_1 + iU_2)e^{-2\pi i\omega_0 t} + \frac{1}{2}(U_1 - iU_2)e^{2\pi i\omega_0 t},$$

where we recall that  $U_1$  and  $U_2$  are uncorrelated, mean-zero, random variables each with variance  $\sigma^2$ . If we call  $Z = \frac{1}{2}(U_1 + iU_2)$ , then  $Z^* = \frac{1}{2}(U_1 - iU_2)$ , where \* denotes conjugation. In this case,  $E(Z) = \frac{1}{2}[E(U_1) + iE(U_2)] = 0$  and similarly  $E(Z^*) = 0$ . For mean-zero complex random variables, say  $X$  and  $Y$ ,  $\text{cov}(X, Y) = E(XY^*)$ . Thus

$$\begin{aligned} \text{var}(Z) &= E(|Z|^2) = E(ZZ^*) = \frac{1}{4}E[(U_1 + iU_2)(U_1 - iU_2)] \\ &= \frac{1}{4}[E(U_1^2) + E(U_2^2)] = \frac{\sigma^2}{2}. \end{aligned}$$

Similarly,  $\text{var}(Z^*) = \sigma^2/2$ . Moreover, since  $Z^{**} = Z$ ,

$$\text{cov}(Z, Z^*) = E(ZZ^{**}) = \frac{1}{4}E[(U_1 + iU_2)(U_1 + iU_2)] = \frac{1}{4}[E(U_1^2) - E(U_2^2)] = 0.$$

Hence, [\(4.13\)](#) may be written as

$$x_t = Z e^{-2\pi i\omega_0 t} + Z^* e^{2\pi i\omega_0 t} = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i\omega t} dZ(\omega),$$

where  $Z(\omega)$  is a complex-valued random process that makes uncorrelated jumps at  $-\omega_0$  and  $\omega_0$  with mean-zero and variance  $\sigma^2/2$ . Stochastic integration is discussed further in [Section C.4.2](#). This notion generalizes to all stationary series in the following property (also, see [Theorem C.2](#)).

**Property 4.5 Spectral Representation of a Stationary Process**

*If  $x_t$  is a mean-zero stationary process, with spectral distribution  $F(\omega)$  as given in [Property 4.1](#), then there exists a complex-valued stochastic process  $Z(\omega)$ , on the interval  $\omega \in [-1/2, 1/2]$ , having stationary uncorrelated non-overlapping increments, such that  $x_t$  can be written as the stochastic integral (see [Section C.4.2](#))*

$$x_t = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i\omega t} dZ(\omega),$$

where, for  $-1/2 \leq \omega_1 \leq \omega_2 \leq 1/2$ ,

$$\text{var} \{Z(\omega_2) - Z(\omega_1)\} = F(\omega_2) - F(\omega_1).$$

**4.3 Periodogram and Discrete Fourier Transform**

We are now ready to tie together the periodogram, which is the sample-based concept presented in [Section 4.1](#), with the spectral density, which is the population-based concept of [Section 4.2](#).

**Definition 4.1** Given data  $x_1, \dots, x_n$ , we define the **discrete Fourier transform (DFT)** to be

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^n x_t e^{-2\pi i \omega_j t} \quad (4.26)$$

for  $j = 0, 1, \dots, n-1$ , where the frequencies  $\omega_j = j/n$  are called the **Fourier or fundamental frequencies**.

If  $n$  is a highly composite integer (i.e., it has many factors), the DFT can be computed by the fast Fourier transform (FFT) introduced in Cooley and Tukey (1965). Also, different packages scale the FFT differently, so it is a good idea to consult the documentation. R computes the DFT defined in (4.26) without the factor  $n^{-1/2}$ , but with an additional factor of  $e^{2\pi i \omega_j}$  that can be ignored because we will be interested in the squared modulus of the DFT. Sometimes it is helpful to exploit the inversion result for DFTs, which shows the linear transformation is one-to-one. For the *inverse DFT* we have,

$$x_t = n^{-1/2} \sum_{j=0}^{n-1} d(\omega_j) e^{2\pi i \omega_j t} \quad (4.27)$$

for  $t = 1, \dots, n$ . The following example shows how to calculate the DFT and its inverse in R for the data set  $\{1, 2, 3, 4\}$ ; note that R writes a complex number  $z = a + ib$  as  $a+bi$ .

```
(dft = fft(1:4)/sqrt(4))
[1] 5+0i -1+1i -1+0i -1-1i
(idft = fft(dft, inverse=TRUE)/sqrt(4))
[1] 1+0i 2+0i 3+0i 4+0i
(Re(idft)) # keep it real
[1] 1 2 3 4
```

We now define the periodogram as the squared modulus of the DFT.

**Definition 4.2** Given data  $x_1, \dots, x_n$ , we define the **periodogram** to be

$$I(\omega_j) = |d(\omega_j)|^2 \quad (4.28)$$

for  $j = 0, 1, 2, \dots, n-1$ .

Note that  $I(0) = n\bar{x}^2$ , where  $\bar{x}$  is the sample mean. Also,  $\sum_{t=1}^n \exp(-2\pi i t \frac{j}{n}) = 0$  for  $j \neq 0$ ,<sup>4.3</sup> so we can write the DFT as

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^n (x_t - \bar{x}) e^{-2\pi i \omega_j t} \quad (4.29)$$

for  $j \neq 0$ . Thus,

<sup>4.3</sup>  $\sum_{t=1}^n z^t = z \frac{1-z^n}{1-z}$  for  $z \neq 1$ . In this case,  $z^n = e^{-2\pi i j} = 1$ .



$$\begin{aligned}
I(\omega_j) &= |d(\omega_j)|^2 = n^{-1} \sum_{t=1}^n \sum_{s=1}^n (x_t - \bar{x})(x_s - \bar{x}) e^{-2\pi i \omega_j (t-s)} \\
&= n^{-1} \sum_{h=-(n-1)}^{n-1} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}) e^{-2\pi i \omega_j h} \\
&= \sum_{h=-(n-1)}^{n-1} \hat{\gamma}(h) e^{-2\pi i \omega_j h}
\end{aligned} \tag{4.30}$$

for  $j \neq 0$ , where we have put  $h = t - s$ , with  $\hat{\gamma}(h)$  as given in (1.36).<sup>4.4</sup> In view of (4.30), the periodogram,  $I(\omega_j)$ , is the sample version of  $f(\omega_j)$  given in (4.17). That is, we may think of the periodogram as the *sample spectral density* of  $x_t$ .

At first, (4.30) seems to be an obvious way to estimate a spectral density (4.17); i.e., simply put a hat on  $\gamma(h)$  and sum as far as the sample size will allow. However, after further consideration, it turns out that this is not a very good estimator because it uses some bad estimates of  $\gamma(h)$ . For example, there is only one pair of observations,  $(x_1, x_n)$  for estimating  $\gamma(n-1)$ , and only two pairs  $(x_1, x_{n-1})$ , and  $(x_2, x_n)$  that can be used to estimate  $\gamma(n-2)$ , and so on. We will discuss this problem further as we progress, but an obvious improvement over (4.30) would be something like  $\hat{f}(\omega) = \sum_{|h| \leq m} \hat{\gamma}(h) e^{-2\pi i \omega h}$ , where  $m$  is much smaller than  $n$ .

It is sometimes useful to work with the real and imaginary parts of the DFT individually. To this end, we define the following transforms.

**Definition 4.3** Given data  $x_1, \dots, x_n$ , we define **the cosine transform**

$$d_c(\omega_j) = n^{-1/2} \sum_{t=1}^n x_t \cos(2\pi \omega_j t) \tag{4.31}$$

and the **sine transform**

$$d_s(\omega_j) = n^{-1/2} \sum_{t=1}^n x_t \sin(2\pi \omega_j t) \tag{4.32}$$

where  $\omega_j = j/n$  for  $j = 0, 1, \dots, n-1$ .

We note that  $d(\omega_j) = d_c(\omega_j) - i d_s(\omega_j)$  and hence

$$I(\omega_j) = d_c^2(\omega_j) + d_s^2(\omega_j). \tag{4.33}$$

We have also discussed the fact that spectral analysis can be thought of as an analysis of variance. The next example examines this notion.

<sup>4.4</sup> Note that (4.30) can be used to obtain  $\hat{\gamma}(h)$  by taking the inverse DFT of  $I(\omega_j)$ . This approach was used in Example 1.31 to obtain a two-dimensional ACF.

**Example 4.10 Spectral ANOVA**

Let  $x_1, \dots, x_n$  be a sample of size  $n$ , where for ease,  $n$  is odd. Then, recalling [Example 4.2](#),

$$x_t = a_0 + \sum_{j=1}^m [a_j \cos(2\pi\omega_j t) + b_j \sin(2\pi\omega_j t)], \quad (4.34)$$

where  $m = (n-1)/2$ , is exact for  $t = 1, \dots, n$ . In particular, using multiple regression formulas, we have  $a_0 = \bar{x}$ ,

$$a_j = \frac{2}{n} \sum_{t=1}^n x_t \cos(2\pi\omega_j t) = \frac{2}{\sqrt{n}} d_c(\omega_j)$$

$$b_j = \frac{2}{n} \sum_{t=1}^n x_t \sin(2\pi\omega_j t) = \frac{2}{\sqrt{n}} d_s(\omega_j).$$

Hence, we may write

$$(x_t - \bar{x}) = \frac{2}{\sqrt{n}} \sum_{j=1}^m [d_c(\omega_j) \cos(2\pi\omega_j t) + d_s(\omega_j) \sin(2\pi\omega_j t)]$$

for  $t = 1, \dots, n$ . Squaring both sides and summing we obtain

$$\sum_{t=1}^n (x_t - \bar{x})^2 = 2 \sum_{j=1}^m [d_c^2(\omega_j) + d_s^2(\omega_j)] = 2 \sum_{j=1}^m I(\omega_j)$$

using the results of [Problem 4.1](#). Thus, we have partitioned the sum of squares into harmonic components represented by frequency  $\omega_j$  with the periodogram,  $I(\omega_j)$ , being the mean square regression. This leads to the ANOVA table for  $n$  odd:

Source	df	SS	MS
$\omega_1$	2	$2I(\omega_1)$	$I(\omega_1)$
$\omega_2$	2	$2I(\omega_2)$	$I(\omega_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\omega_m$	2	$2I(\omega_m)$	$I(\omega_m)$
Total	$n - 1$	$\sum_{t=1}^n (x_t - \bar{x})^2$	

The following is an R example to help explain this concept. We consider  $n = 5$  observations given by  $x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 2, x_5 = 1$ . Note that the data complete one cycle, but not in a sinusoidal way. Thus, we should expect the  $\omega_1 = 1/5$  component to be relatively large but not exhaustive, and the  $\omega_2 = 2/5$  component to be small.

```
x = c(1, 2, 3, 2, 1)
c1 = cos(2*pi*1:5*1/5); s1 = sin(2*pi*1:5*1/5)
c2 = cos(2*pi*1:5*2/5); s2 = sin(2*pi*1:5*2/5)
```

```

omega1 = cbind(c1, s1); omega2 = cbind(c2, s2)
anova(lm(x~omega1+omega2)) # ANOVA Table
      Df Sum Sq Mean Sq
omega1  2  2.74164  1.37082
omega2  2   .05836   .02918
Residuals 0   .00000
Mod(fft(x))^2/5 # the periodogram (as a check)
[1] 16.2  1.37082 .029179 .029179  1.37082
# I(0)  I(1/5)  I(2/5)  I(3/5)  I(4/5)

```

Note that  $I(0) = n\bar{x}^2 = 5 \times 1.8^2 = 16.2$ . Also, the sum of squares associated with the residuals (SSE) is zero, indicating an exact fit.

**Example 4.11 Spectral Analysis as Principal Component Analysis**

It is also possible to think of spectral analysis as a principal component analysis. In Section C.5, we show that the spectral density may be thought of as the approximate eigenvalues of the covariance matrix of a stationary process. If  $X = (x_1, \dots, x_n)$  are  $n$  values of a mean-zero time series,  $x_t$  with spectral density  $f_x(\omega)$ , then

$$\text{cov}(X) = \Gamma_n = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{bmatrix}.$$

For  $n$  sufficiently large, the eigenvalues of  $\Gamma_n$  are

$$\lambda_j \approx f(\omega_j) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i h j/n},$$

with approximate eigenvectors

$$g_j^* = \frac{1}{\sqrt{n}}(e^{-2\pi i 0j/n}, e^{-2\pi i 1j/n}, \dots, e^{-2\pi i (n-1)j/n}),$$

for  $j = 0, 1, \dots, n-1$ . If we let  $G$  be the complex matrix with columns  $g_j$ , then the complex vector  $Y = G^*X$  has elements that are the DFTs,

$$y_j = \frac{1}{\sqrt{n}} \sum_{t=1}^n x_t e^{-2\pi i t j/n}$$

for  $j = 0, 1, \dots, n-1$ . In this case, the elements of  $Y$  are asymptotically uncorrelated complex random variables, with mean-zero and variance  $f(\omega_j)$ . Also,  $X$  may be recovered as  $X = GY$ , so that  $x_t = \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} y_j e^{2\pi i t j/n}$ .

We are now ready to present some large sample properties of the periodogram. First, let  $\mu$  be the mean of a stationary process  $x_t$  with absolutely summable autocovariance function  $\gamma(h)$  and spectral density  $f(\omega)$ . We can use the same argument as in (4.30), replacing  $\bar{x}$  by  $\mu$  in (4.29), to write

$$I(\omega_j) = n^{-1} \sum_{h=-(n-1)}^{n-1} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \mu)(x_t - \mu) e^{-2\pi i \omega_j h} \quad (4.35)$$

where  $\omega_j$  is a non-zero fundamental frequency. Taking expectation in (4.35) we obtain

$$E[I(\omega_j)] = \sum_{h=-(n-1)}^{n-1} \left( \frac{n-|h|}{n} \right) \gamma(h) e^{-2\pi i \omega_j h}. \quad (4.36)$$

For any given  $\omega \neq 0$ , choose a sequence of fundamental frequencies  $\omega_{j:n} \rightarrow \omega$ <sup>4.5</sup> from which it follows by (4.36) that, as  $n \rightarrow \infty$ <sup>4.6</sup>

$$E[I(\omega_{j:n})] \rightarrow f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i h \omega}. \quad (4.37)$$

In other words, under absolute summability of  $\gamma(h)$ , the spectral density is the long-term average of the periodogram.

Additional asymptotic properties may be established under the condition that the autocovariance function satisfies

$$\theta = \sum_{h=-\infty}^{\infty} |h| |\gamma(h)| < \infty. \quad (4.38)$$

First, we note that straight-forward calculations lead to

$$\text{cov}[d_c(\omega_j), d_c(\omega_k)] = n^{-1} \sum_{s=1}^n \sum_{t=1}^n \gamma(s-t) \cos(2\pi \omega_j s) \cos(2\pi \omega_k t), \quad (4.39)$$

$$\text{cov}[d_c(\omega_j), d_s(\omega_k)] = n^{-1} \sum_{s=1}^n \sum_{t=1}^n \gamma(s-t) \cos(2\pi \omega_j s) \sin(2\pi \omega_k t), \quad (4.40)$$

$$\text{cov}[d_s(\omega_j), d_s(\omega_k)] = n^{-1} \sum_{s=1}^n \sum_{t=1}^n \gamma(s-t) \sin(2\pi \omega_j s) \sin(2\pi \omega_k t), \quad (4.41)$$

where the variance terms are obtained by setting  $\omega_j = \omega_k$  in (4.39) and (4.41). In Appendix C, Section C.2, we show the terms in (4.39)–(4.41) have interesting properties under assumption that (4.38) holds. In particular, for  $\omega_j, \omega_k \neq 0$  or  $1/2$ ,

$$\text{cov}[d_c(\omega_j), d_c(\omega_k)] = \begin{cases} f(\omega_j)/2 + \varepsilon_n & \omega_j = \omega_k, \\ \varepsilon_n & \omega_j \neq \omega_k, \end{cases} \quad (4.42)$$

<sup>4.5</sup> By this we mean  $\omega_{j:n} = j_n/n$ , where  $\{j_n\}$  is a sequence of integers chosen so that  $j_n/n$  is the closest Fourier frequency to  $\omega$ ; consequently,  $|j_n/n - \omega| \leq \frac{1}{2n}$ .

<sup>4.6</sup> From Definition 4.2 we have  $I(0) = n\bar{x}^2$ , so the analogous result of (4.37) for the case  $\omega = 0$  is  $E[I(0)] - n\mu^2 = n \text{var}(\bar{x}) \rightarrow f(0)$  as  $n \rightarrow \infty$ .

$$\text{cov}[d_s(\omega_j), d_s(\omega_k)] = \begin{cases} f(\omega_j)/2 + \varepsilon_n & \omega_j = \omega_k, \\ \varepsilon_n & \omega_j \neq \omega_k, \end{cases} \quad (4.43)$$

and

$$\text{cov}[d_c(\omega_j), d_s(\omega_k)] = \varepsilon_n, \quad (4.44)$$

where the error term  $\varepsilon_n$  in the approximations can be bounded,

$$|\varepsilon_n| \leq \theta/n, \quad (4.45)$$

and  $\theta$  is given by (4.38). If  $\omega_j = \omega_k = 0$  or  $1/2$  in (4.42), the multiplier  $1/2$  disappears; note that  $d_s(0) = d_s(1/2) = 0$ , so (4.43) does not apply in these cases.

**Example 4.12 Covariance of Sine and Cosine Transforms**

For the three-point moving average series of Example 1.9 and  $n = 256$  observations, the theoretical covariance matrix of the vector  $D = (d_c(\omega_{26}), d_s(\omega_{26}), d_c(\omega_{27}), d_s(\omega_{27}))'$  using (4.39)–(4.41) is

$$\text{cov}(D) = \begin{pmatrix} .3752 & -.0009 & -.0022 & -.0010 \\ -.0009 & .3777 & -.0009 & .0003 \\ -.0022 & -.0009 & .3667 & -.0010 \\ -.0010 & .0003 & -.0010 & .3692 \end{pmatrix}.$$

The diagonal elements can be compared with half the theoretical spectral values of  $\frac{1}{2}f(\omega_{26}) = .3774$  for the spectrum at frequency  $\omega_{26} = 26/256$ , and of  $\frac{1}{2}f(\omega_{27}) = .3689$  for the spectrum at  $\omega_{27} = 27/256$ . Hence, the cosine and sine transforms produce nearly uncorrelated variables with variances approximately equal to one half of the theoretical spectrum. For this particular case, the uniform bound is determined from  $\theta = 8/9$ , yielding  $|\varepsilon_{256}| \leq .0035$  for the bound on the approximation error.

If  $x_t \sim \text{iid}(0, \sigma^2)$ , then it follows from (4.38)–(4.44), and a central limit theorem<sup>4.7</sup> that

$$d_c(\omega_{j:n}) \sim \text{AN}(0, \sigma^2/2) \quad \text{and} \quad d_s(\omega_{j:n}) \sim \text{AN}(0, \sigma^2/2) \quad (4.46)$$

jointly and independently, and independent of  $d_c(\omega_{k:n})$  and  $d_s(\omega_{k:n})$  provided  $\omega_{j:n} \rightarrow \omega_1$  and  $\omega_{k:n} \rightarrow \omega_2$  where  $0 < \omega_1 \neq \omega_2 < 1/2$ . We note that in this case,  $f_x(\omega) = \sigma^2$ . In view of (4.46), it follows immediately that as  $n \rightarrow \infty$ ,

$$\frac{2I(\omega_{j:n})}{\sigma^2} \xrightarrow{d} \chi_2^2 \quad \text{and} \quad \frac{2I(\omega_{k:n})}{\sigma^2} \xrightarrow{d} \chi_2^2 \quad (4.47)$$

with  $I(\omega_{j:n})$  and  $I(\omega_{k:n})$  being asymptotically independent, where  $\chi_v^2$  denotes a chi-squared random variable with  $v$  degrees of freedom. If the process is also Gaussian, then the above statements are true for any sample size.

Using the central limit theory of Section C.2, it is fairly easy to extend the results of the iid case to the case of a linear process.

<sup>4.7</sup> If  $\{Y_j\} \sim \text{iid}(0, \sigma^2)$  and  $\{a_j\}$  are constants for which  $\sum_{j=1}^n a_j^2 / \max_{1 \leq j \leq n} a_j^2 \rightarrow \infty$  as  $n \rightarrow \infty$ , then  $\sum_{j=1}^n a_j Y_j \sim \text{AN}(0, \sigma^2 \sum_{j=1}^n a_j^2)$ . AN is read *asymptotically normal*; see Definition A.5.

**Property 4.6 Distribution of the Periodogram Ordinates**

If

$$x_t = \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}, \quad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty \quad (4.48)$$

where  $w_t \sim iid(0, \sigma_w^2)$ , and (4.38) holds, then for any collection of  $m$  distinct frequencies  $\omega_j \in (0, 1/2)$  with  $\omega_{j:n} \rightarrow \omega_j$

$$\frac{2I(\omega_{j:n})}{f(\omega_j)} \xrightarrow{d} iid \chi_2^2 \quad (4.49)$$

provided  $f(\omega_j) > 0$ , for  $j = 1, \dots, m$ .

This result is stated more precisely in **Theorem C.7**. Other approaches to large sample normality of the periodogram ordinates are in terms of cumulants, as in Brillinger (1981), or in terms of mixing conditions, such as in Rosenblatt (1956a). Here, we adopt the approach used by Hannan (1970), Fuller (1996), and Brockwell and Davis (1991).

The distributional result (4.49) can be used to derive an approximate *confidence interval for the spectrum* in the usual way. Let  $\chi_\nu^2(\alpha)$  denote the lower  $\alpha$  probability tail for the chi-squared distribution with  $\nu$  degrees of freedom; that is,

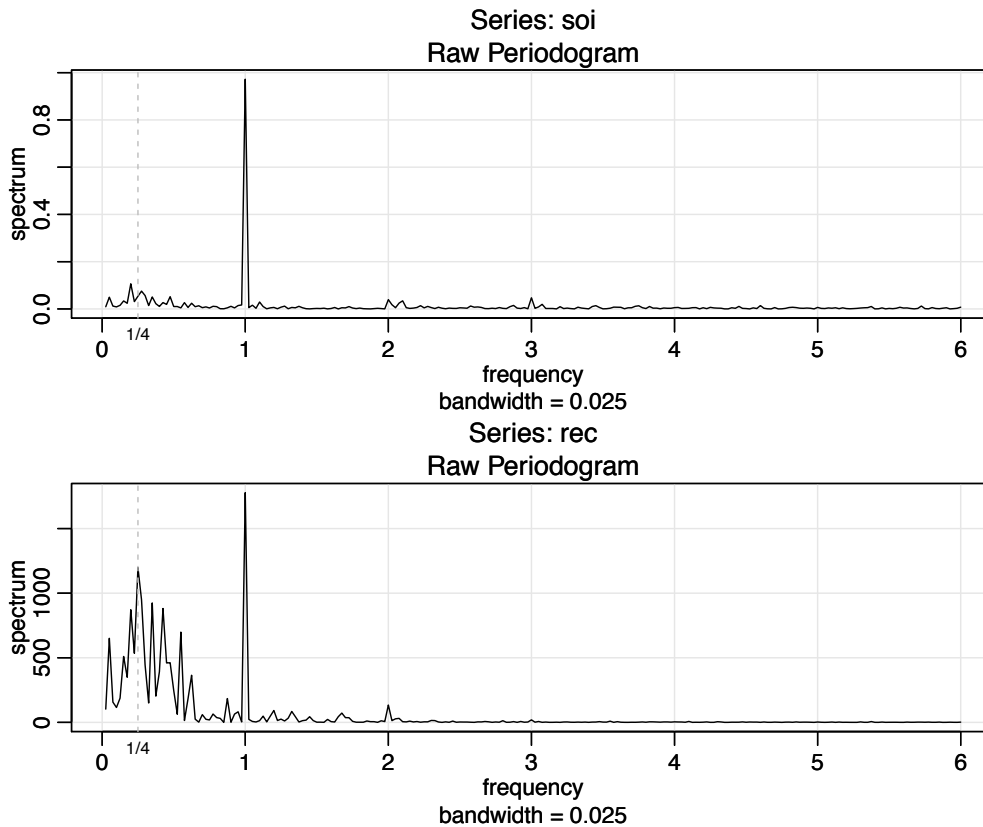
$$\Pr\{\chi_\nu^2 \leq \chi_\nu^2(\alpha)\} = \alpha. \quad (4.50)$$

Then, an approximate  $100(1-\alpha)\%$  confidence interval for the spectral density function would be of the form

$$\frac{2 I(\omega_{j:n})}{\chi_2^2(1 - \alpha/2)} \leq f(\omega) \leq \frac{2 I(\omega_{j:n})}{\chi_2^2(\alpha/2)}. \quad (4.51)$$

Often, trends are present that should be eliminated before computing the periodogram. Trends introduce extremely low frequency components in the periodogram that tend to obscure the appearance at higher frequencies. For this reason, it is usually conventional to center the data prior to a spectral analysis using either mean-adjusted data of the form  $x_t - \bar{x}$  to eliminate the zero or d-c component or to use detrended data of the form  $x_t - \hat{\beta}_1 - \hat{\beta}_2 t$  to eliminate the term that will be considered a half cycle by the spectral analysis. Note that higher order polynomial regressions in  $t$  or nonparametric smoothing (linear filtering) could be used in cases where the trend is nonlinear.

As previously indicated, it is often convenient to calculate the DFTs, and hence the periodogram, using the fast Fourier transform algorithm. The FFT utilizes a number of redundancies in the calculation of the DFT when  $n$  is highly composite; that is, an integer with many factors of 2, 3, or 5, the best case being when  $n = 2^p$  is a factor of 2. Details may be found in Cooley and Tukey (1965). To accommodate this property, we can pad the centered (or detrended) data of length  $n$  to the next highly composite integer  $n'$  by adding zeros, i.e., setting  $x_{n+1}^c = x_{n+2}^c = \dots = x_{n'}^c = 0$ , where  $x_t^c$



**Fig. 4.5.** Periodogram of SOI and Recruitment,  $n = 453$  ( $n' = 480$ ), where the frequency axis is labeled in multiples of  $\Delta = 1/12$ . Note the common peaks at  $\omega = 1\Delta = 1/12$ , or one cycle per year (12 months), and some larger values near  $\omega = \frac{1}{4}\Delta = 1/48$ , or one cycle every four years (48 months).

denotes the centered data. This means that the fundamental frequency ordinates will be  $\omega_j = j/n'$  instead of  $j/n$ . We illustrate by considering the periodogram of the SOI and Recruitment series shown in **Figure 1.5**. Recall that they are monthly series and  $n = 453$  months. To find  $n'$  in R, use the command `nextn(453)` to see that  $n' = 480$  will be used in the spectral analyses by default.

#### Example 4.13 Periodogram of SOI and Recruitment Series

**Figure 4.5** shows the periodograms of each series, where the frequency axis is labeled in multiples of  $\Delta = 1/12$ . As previously indicated, the centered data have been padded to a series of length 480. We notice a narrow-band peak at the obvious yearly (12 month) cycle,  $\omega = 1\Delta = 1/12$ . In addition, there is considerable power in a wide band at the lower frequencies that is centered around the four-year (48 month) cycle  $\omega = \frac{1}{4}\Delta = 1/48$  representing a possible El Niño effect. This wide band activity suggests that the possible El Niño cycle is irregular, but tends to be around four years on average. We will continue to address this problem as we move to more sophisticated analyses.

Noting  $\chi_2^2(.025) = .05$  and  $\chi_2^2(.975) = 7.38$ , we can obtain approximate 95% confidence intervals for the frequencies of interest. For example, the periodogram

of the SOI series is  $I_S(1/12) = .97$  at the yearly cycle. An approximate 95% confidence interval for the spectrum  $f_S(1/12)$  is then

$$[2(.97)/7.38, 2(.97)/.05] = [.26, 38.4],$$

which is too wide to be of much use. We do notice, however, that the lower value of .26 is higher than any other periodogram ordinate, so it is safe to say that this value is significant. On the other hand, an approximate 95% confidence interval for the spectrum at the four-year cycle,  $f_S(1/48)$ , is

$$[2(.05)/7.38, 2(.05)/.05] = [.01, 2.12],$$

which again is extremely wide, and with which we are unable to establish significance of the peak.

We now give the R commands that can be used to reproduce Figure 4.5. To calculate and graph the periodogram, we used the `mvspec` command in available from `astsa`. We note that the value of  $\Delta$  is the reciprocal of the value of `frequency` for the data of a time series object. If the data are not a time series object, `frequency` is set to 1. Also, we set `log="no"` because the periodogram is plotted on a  $\log_{10}$  scale by default. Figure 4.5 displays a `bandwidth`. We will discuss bandwidth in the next section, so ignore this for the time being.

```
par(mfrow=c(2,1))
soi.per = mvspec(soi, log="no")
abline(v=1/4, lty=2)
rec.per = mvspec(rec, log="no")
abline(v=1/4, lty=2)
```

The confidence intervals for the SOI series at the yearly cycle,  $\omega = 1/12 = 40/480$ , and the possible El Niño cycle of four years  $\omega = 1/48 = 10/480$  can be computed in R as follows:

```
soi.per$spec[40] # 0.97223; soi pgram at freq 1/12 = 40/480
soi.per$spec[10] # 0.05372; soi pgram at freq 1/48 = 10/480
# conf intervals - returned value:
U = qchisq(.025,2) # 0.05063
L = qchisq(.975,2) # 7.37775
2*soi.per$spec[10]/L # 0.01456
2*soi.per$spec[10]/U # 2.12220
2*soi.per$spec[40]/L # 0.26355
2*soi.per$spec[40]/U # 38.40108
```

The preceding example made it clear that the periodogram as an estimator is susceptible to large uncertainties, and we need to find a way to reduce the variance. Not surprisingly, this result follows if consider (4.49) and the fact that, for any  $n$ , the periodogram is based on only two observations. Recall that the mean and variance of the  $\chi_\nu^2$  distribution are  $\nu$  and  $2\nu$ , respectively. Thus, using (4.49), we have  $I(\omega) \sim \frac{1}{2}f(\omega)\chi_2^2$ , implying

$$E[I(\omega)] \approx f(\omega) \quad \text{and} \quad \text{var}[I(\omega)] \approx f^2(\omega).$$

Consequently,  $\text{var}[I(\omega)] \not\rightarrow 0$  as  $n \rightarrow \infty$  and thus the periodogram is not a consistent estimator of the spectral density. The solution to this dilemma can be resolved by smoothing the periodogram.



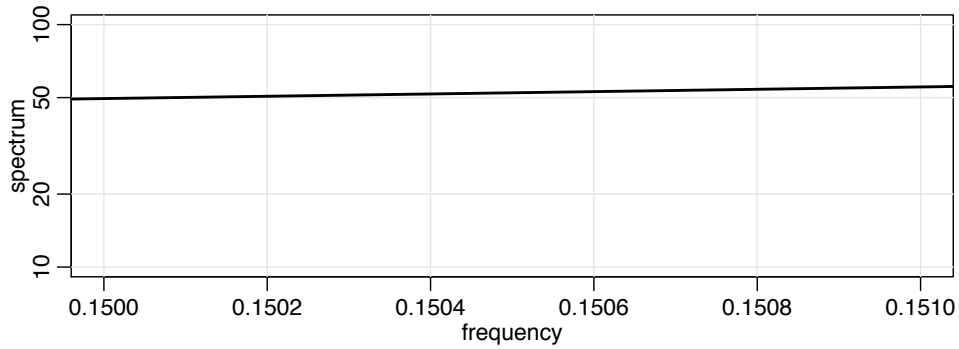


Fig. 4.6. A small section (near the peak) of the AR(2) spectrum shown in Figure 4.4.

### 4.4 Nonparametric Spectral Estimation

To continue the discussion that ended the previous section, we introduce a frequency band,  $\mathcal{B}$ , of  $L \ll n$  contiguous fundamental frequencies, centered around frequency  $\omega_j = j/n$ , which is chosen close to a frequency of interest,  $\omega$ . For frequencies of the form  $\omega^* = \omega_j + k/n$ , let

$$\mathcal{B} = \left\{ \omega^* : \omega_j - \frac{m}{n} \leq \omega^* \leq \omega_j + \frac{m}{n} \right\}, \tag{4.52}$$

where

$$L = 2m + 1 \tag{4.53}$$

is an odd number, chosen such that the spectral values in the interval  $\mathcal{B}$ ,

$$f(\omega_j + k/n), \quad k = -m, \dots, 0, \dots, m$$

are approximately equal to  $f(\omega)$ . This structure can be realized for large sample sizes, as shown formally in Section C.2. Values of the spectrum in this band should be relatively constant for the smoothed spectra defined below to be good estimators. For example, to see a small section of the AR(2) spectrum (near the peak) shown in Figure 4.4, use

```
arma.spec(ar=c(1,-.9), xlim=c(.15,.151), n.freq=100000)
```

which is displayed in Figure 4.6.

We now define an averaged (or smoothed) periodogram as the average of the periodogram values, say,

$$\tilde{f}(\omega) = \frac{1}{L} \sum_{k=-m}^m I(\omega_j + k/n), \tag{4.54}$$

over the band  $\mathcal{B}$ . Under the assumption that the spectral density is fairly constant in the band  $\mathcal{B}$ , and in view of (4.49) we can show that under appropriate conditions,<sup>4.8</sup>

<sup>4.8</sup> The conditions, which are sufficient, are that  $x_t$  is a linear process, as described in Property 4.6, with  $\sum_j \sqrt{|j|} |\psi_j| < \infty$ , and  $w_t$  has a finite fourth moment.

for large  $n$ , the periodograms in (4.54) are approximately distributed as independent  $f(\omega)\chi_2^2/2$  random variables, for  $0 < \omega < 1/2$ , as long as we keep  $L$  fairly small relative to  $n$ . This result is discussed formally in Section C.2. Thus, under these conditions,  $L\bar{f}(\omega)$  is the sum of  $L$  approximately independent  $f(\omega)\chi_2^2/2$  random variables. It follows that, for large  $n$ ,

$$\frac{2L\bar{f}(\omega)}{f(\omega)} \sim \chi_{2L}^2 \quad (4.55)$$

where  $\sim$  means *is approximately distributed as*.

In this scenario, where we smooth the periodogram by simple averaging, it seems reasonable to call the width of the frequency interval defined by (4.52),

$$B = \frac{L}{n}, \quad (4.56)$$

the *bandwidth*.<sup>4.9</sup> The concept of bandwidth, however, becomes more complicated with the introduction of spectral estimators that smooth with unequal weights. Note that (4.56) implies the degrees of freedom can be expressed as

$$2L = 2Bn, \quad (4.57)$$

or twice the *time-bandwidth product*. The result (4.55) can be rearranged to obtain an approximate  $100(1 - \alpha)\%$  confidence interval of the form

$$\frac{2L\bar{f}(\omega)}{\chi_{2L}^2(1 - \alpha/2)} \leq f(\omega) \leq \frac{2L\bar{f}(\omega)}{\chi_{2L}^2(\alpha/2)} \quad (4.58)$$

for the true spectrum,  $f(\omega)$ .

Many times, the visual impact of a spectral density plot will be improved by plotting the logarithm of the spectrum instead of the spectrum (the log transformation is the variance stabilizing transformation in this situation). This phenomenon can occur when regions of the spectrum exist with peaks of interest much smaller than some of the main power components. Taking logs in (4.58), we obtain an interval for the logged spectrum given by

$$\left[ \log \bar{f}(\omega) - a_L, \log \bar{f}(\omega) + b_L \right] \quad (4.59)$$

where

<sup>4.9</sup> There are many definitions of bandwidth and an excellent discussion may be found in Percival and Walden (1993, §6.7). The bandwidth value used in R for `spec.pgram` is based on Grenander (1951). The basic idea is that bandwidth can be related to the standard deviation of the weighting distribution. For the uniform distribution on the frequency range  $-m/n$  to  $m/n$ , the standard deviation is  $L/n\sqrt{12}$  (using a continuity correction). Consequently, in the case of (4.54), R will report a bandwidth of  $L/n\sqrt{12}$ , which amounts to dividing our definition by  $\sqrt{12}$ . Note that in the extreme case  $L = n$ , we would have  $B = 1$  indicating that everything was used in the estimation. In this case, R would report a bandwidth of  $1/\sqrt{12} \approx .29$ , which seems to miss the point.

$$a_L = -\log 2L + \log \chi_{2L}^2(1 - \alpha/2) \quad \text{and} \quad b_L = \log 2L - \log \chi_{2L}^2(\alpha/2)$$

do not depend on  $\omega$ .

If zeros are appended before computing the spectral estimators, we need to adjust the degrees of freedom (because you do not get more information by padding) and an approximation is to replace  $2L$  by  $2Ln/n'$ . Hence, we define the *adjusted degrees of freedom* as

$$df = \frac{2Ln}{n'} \quad (4.60)$$

and use it instead of  $2L$  in the confidence intervals (4.58) and (4.59). For example, (4.58) becomes

$$\frac{df \bar{f}(\omega)}{\chi_{df}^2(1 - \alpha/2)} \leq f(\omega) \leq \frac{df \bar{f}(\omega)}{\chi_{df}^2(\alpha/2)}. \quad (4.61)$$

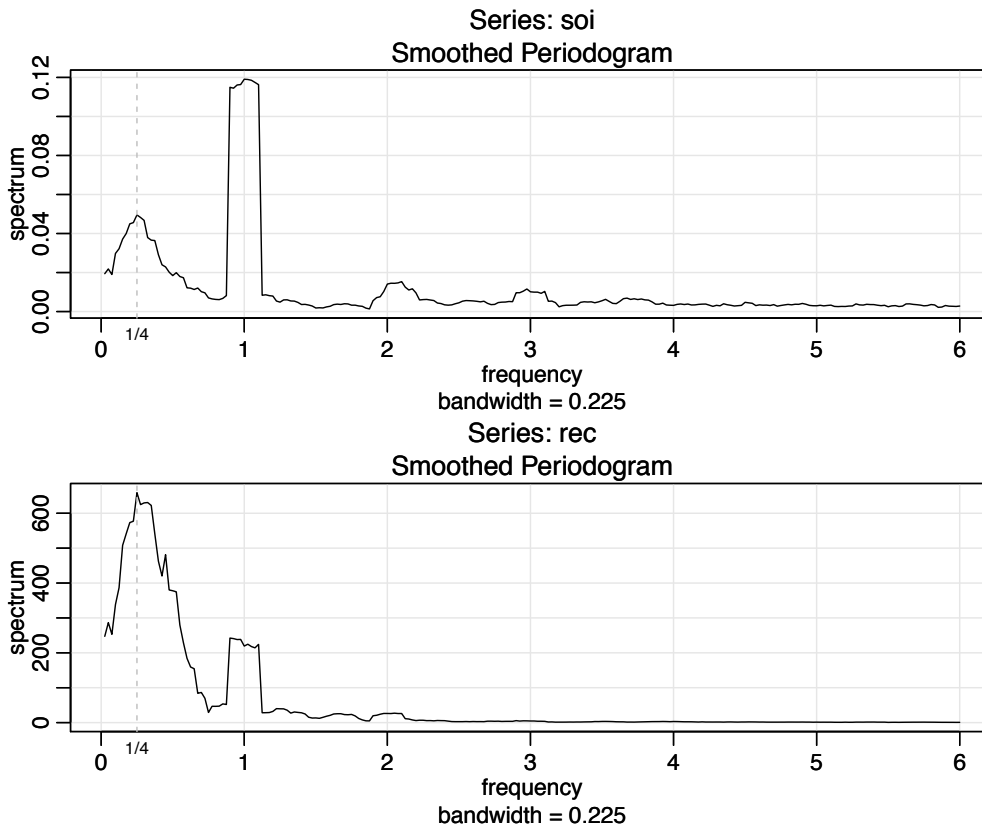
A number of assumptions are made in computing the approximate confidence intervals given above, which may not hold in practice. In such cases, it may be reasonable to employ resampling techniques such as one of the parametric bootstraps proposed by Hurvich and Zeger (1987) or a nonparametric *local bootstrap* proposed by Paparoditis and Politis (1999). To develop the bootstrap distributions, we assume that the contiguous DFTs in a frequency band of the form (4.52) all came from a time series with identical spectrum  $f(\omega)$ . This, in fact, is exactly the same assumption made in deriving the large-sample theory. We may then simply resample the  $L$  DFTs in the band, with replacement, calculating a spectral estimate from each bootstrap sample. The sampling distribution of the bootstrap estimators approximates the distribution of the nonparametric spectral estimator. For further details, including the theoretical properties of such estimators, see Paparoditis and Politis (1999).

Before proceeding further, we consider computing the average periodograms for the SOI and Recruitment series.

#### Example 4.14 Averaged Periodogram for SOI and Recruitment

Generally, it is a good idea to try several bandwidths that seem to be compatible with the general overall shape of the spectrum, as suggested by the periodogram. We will discuss this problem in more detail after the example. The SOI and Recruitment series periodograms, previously computed in Figure 4.5, suggest the power in the lower El Niño frequency needs smoothing to identify the predominant overall period. Trying values of  $L$  leads to the choice  $L = 9$  as a reasonable value, and the result is displayed in Figure 4.7.

The smoothed spectra shown provide a sensible compromise between the noisy version, shown in Figure 4.5, and a more heavily smoothed spectrum, which might lose some of the peaks. An undesirable effect of averaging can be noticed at the yearly cycle,  $\omega = 1\Delta$ , where the narrow band peaks that appeared in the periodograms in Figure 4.5 have been flattened and spread out to nearby frequencies. We also notice, and have marked, the appearance of *harmonics* of the yearly cycle, that is, frequencies of the form  $\omega = k\Delta$  for  $k = 1, 2, \dots$ . Harmonics typically occur when a periodic non-sinusoidal component is present; see Example 4.15.



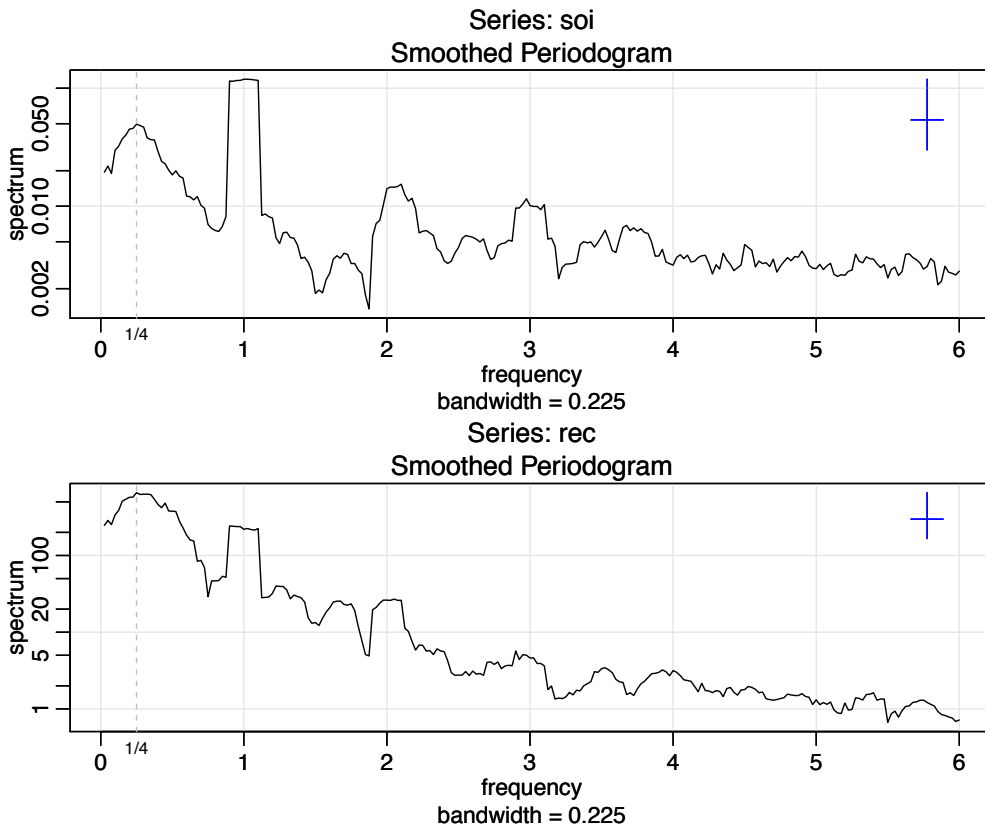
**Fig. 4.7.** The averaged periodogram of the SOI and Recruitment series  $n = 453$ ,  $n' = 480$ ,  $L = 9$ ,  $df = 17$ , showing common peaks at the four year period,  $\omega = \frac{1}{4}\Delta = 1/48$  cycles/month, the yearly period,  $\omega = 1\Delta = 1/12$  cycles/month and some of its harmonics  $\omega = k\Delta$  for  $k = 2, 3$ .

Figure 4.7 can be reproduced in R using the following commands. To compute averaged periodograms, use the Daniell kernel, and specify  $m$ , where  $L = 2m + 1$  ( $L = 9$  and  $m = 4$  in this example). We will explain the kernel concept later in this section, specifically just prior to Example 4.16.

```
soi.ave = mvspec(soi, kernel('daniell',4)), log='no')
abline(v=c(.25,1,2,3), lty=2)
soi.ave$bandwidth # = 0.225
# Repeat above lines using rec in place of soi on line 3
```

The displayed bandwidth (.225) is adjusted for the fact that the frequency scale of the plot is in terms of cycles per year instead of cycles per month. Using (4.56), the bandwidth in terms of months is  $9/480 = .01875$ ; the displayed value is simply converted to years,  $.01875 \times 12 = .225$ .

The adjusted degrees of freedom are  $df = 2(9)(453)/480 \approx 17$ . We can use this value for the 95% confidence intervals, with  $\chi_{df}^2(.025) = 7.56$  and  $\chi_{df}^2(.975) = 30.17$ . Substituting into (4.61) gives the intervals in Table 4.1 for the two frequency bands identified as having the maximum power. To examine the two peak power possibilities, we may look at the 95% confidence intervals and see whether the lower limits are substantially larger than adjacent baseline spectral levels. For example,



**Fig. 4.8.** Figure 4.7 with the average periodogram ordinates plotted on a  $\log_{10}$  scale. The display in the upper right-hand corner represents a generic 95% confidence interval where the middle tick mark is the width of the bandwidth.

the El Niño frequency of 48 months has lower limits that exceed the values the spectrum would have if there were simply a smooth underlying spectral function without the peaks. The relative distribution of power over frequencies is different, with the SOI having less power at the lower frequency, relative to the seasonal periods, and the Recruitment series having more power at the lower or El Niño frequency.

The entries in Table 4.1 for SOI can be obtained in R as follows:

```
df = soi.ave$df           # df = 16.9875 (returned values)
U = qchisq(.025, df)     # U = 7.555916
L = qchisq(.975, df)     # L = 30.17425
soi.ave$spec[10]         # 0.0495202
soi.ave$spec[40]         # 0.1190800
# intervals
df*soi.ave$spec[10]/L   # 0.0278789
df*soi.ave$spec[10]/U   # 0.1113333
df*soi.ave$spec[40]/L   # 0.0670396
df*soi.ave$spec[40]/U   # 0.2677201
# repeat above commands with soi replaced by rec
```

Finally, Figure 4.8 shows the averaged periodograms in Figure 4.7 plotted on a  $\log_{10}$  scale. This is the default can be obtained by removing the statement `log="no"`.

**Table 4.1.** Confidence Intervals for the Spectra of the SOI and Recruitment Series

Series	$\omega$	Period	Power	Lower	Upper
SOI	1/48	4 years	.05	.03	.11
	1/12	1 year	.12	.07	.27
Recruits $\times 10^2$	1/48	4 years	6.59	3.71	14.82
	1/12	1 year	2.19	1.24	4.93

Notice that the default plot also shows a generic confidence interval of the form (4.59) (with  $\log$  replaced by  $\log_{10}$ ) in the upper right-hand corner. To use it, imagine placing the middle tick mark (the width of which is the bandwidth) on the averaged periodogram ordinate of interest; the resulting bar then constitutes an approximate 95% confidence interval for the spectrum at that frequency. We note that displaying the estimates on a log scale tends to emphasize the harmonic components.

### Example 4.15 Harmonics

In the previous example, we saw that the spectra of the annual signals displayed minor peaks at the harmonics; that is, the signal spectra had a large peak at  $\omega = 1\Delta = 1/12$  cycles/month (the one-year cycle) and minor peaks at its harmonics  $\omega = k\Delta$  for  $k = 2, 3, \dots$  (two-, three-, and so on, cycles per year). This will often be the case because most signals are not perfect sinusoids (or perfectly cyclic). In this case, the harmonics are needed to capture the non-sinusoidal behavior of the signal. As an example, consider the signal formed in Figure 4.9 from a (fundamental) sinusoid oscillating at two cycles per unit time along with the second through sixth harmonics at decreasing amplitudes. In particular, the signal was formed as

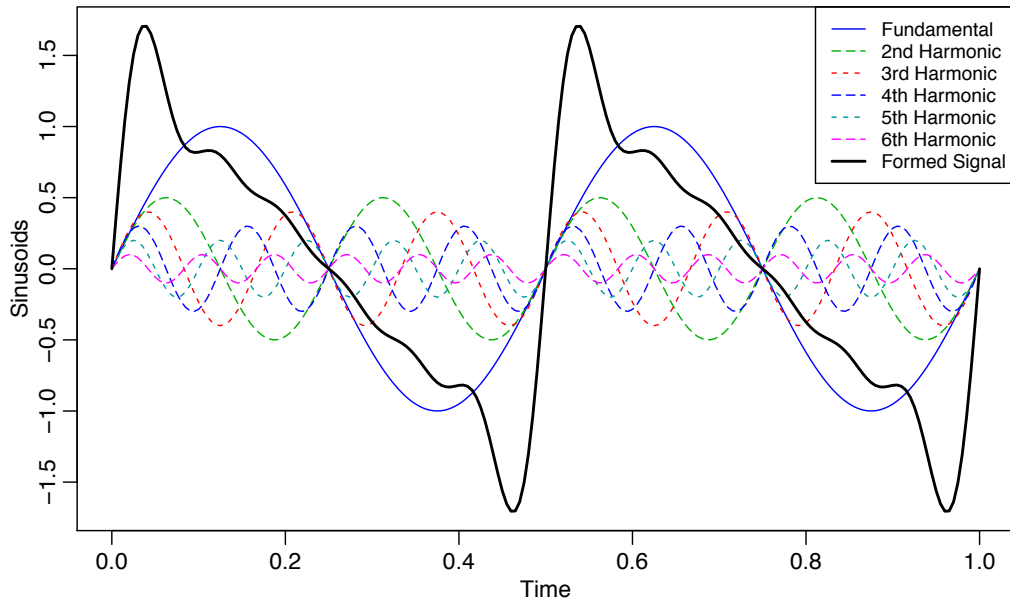
$$x_t = \sin(2\pi 2t) + .5 \sin(2\pi 4t) + .4 \sin(2\pi 6t) \\ + .3 \sin(2\pi 8t) + .2 \sin(2\pi 10t) + .1 \sin(2\pi 12t) \quad (4.62)$$

for  $0 \leq t \leq 1$ . Notice that the signal is non-sinusoidal in appearance and rises quickly then falls slowly.

A figure similar to Figure 4.9 can be generated in R as follows.

```
t = seq(0, 1, by=1/200)
amps = c(1, .5, .4, .3, .2, .1)
x = matrix(0, 201, 6)
for (j in 1:6){ x[,j] = amps[j]*sin(2*pi*t*2*j) }
x = ts(cbind(x, rowSums(x)), start=0, deltat=1/200)
ts.plot(x, lty=c(1:6, 1), lwd=c(rep(1,6), 2), ylab="Sinusoids")
names = c("Fundamental", "2nd Harmonic", "3rd Harmonic", "4th Harmonic", "5th
Harmonic", "6th Harmonic", "Formed Signal")
legend("topright", names, lty=c(1:6, 1), lwd=c(rep(1,6), 2))
```

Example 4.14 points out the necessity for having some relatively systematic procedure for deciding whether peaks are significant. The question of deciding whether a single peak is significant usually rests on establishing what we might think of as a baseline level for the spectrum, defined rather loosely as the shape that one would



**Fig. 4.9.** A signal (thick solid line) formed by a fundamental sinusoid (thin solid line) oscillating at two cycles per unit time and its harmonics as specified in (4.62).

expect to see if no spectral peaks were present. This profile can usually be guessed by looking at the overall shape of the spectrum that includes the peaks; usually, a kind of baseline level will be apparent, with the peaks seeming to emerge from this baseline level. If the lower confidence limit for the spectral value is still greater than the baseline level at some predetermined level of significance, we may claim that frequency value as a statistically significant peak. To be consistent with our stated indifference to the upper limits, we might use a one-sided confidence interval.

An important aspect of interpreting the significance of confidence intervals and tests involving spectra is that typically, more than one frequency will be of interest, so that we will potentially be interested in *simultaneous statements* about a whole collection of frequencies. For example, it would be unfair to claim in Table 4.1 the two frequencies of interest as being statistically significant and all other potential candidates as nonsignificant at the overall level of  $\alpha = .05$ . In this case, we follow the usual statistical approach, noting that if  $K$  statements  $S_1, S_2, \dots, S_k$  are made at significance level  $\alpha$ , i.e.,  $P\{S_k\} = 1 - \alpha$ , then the overall probability all statements are true satisfies the *Bonferroni inequality*

$$P\{\text{all } S_k \text{ true}\} \geq 1 - K\alpha. \quad (4.63)$$

For this reason, it is desirable to set the significance level for testing each frequency at  $\alpha/K$  if there are  $K$  potential frequencies of interest. If, a priori, potentially  $K = 10$  frequencies are of interest, setting  $\alpha = .01$  would give an overall significance level of bound of .10.

The use of the confidence intervals and the necessity for smoothing requires that we make a decision about the **bandwidth B** over which the spectrum will be essentially

constant. Taking too broad a band will tend to smooth out valid peaks in the data when the constant variance assumption is not met over the band. Taking too narrow a band will lead to confidence intervals so wide that peaks are no longer statistically significant. Thus, we note that there is a conflict here between variance properties or *bandwidth stability*, which can be improved by increasing  $B$  and *resolution*, which can be improved by decreasing  $B$ . A common approach is to try a number of different bandwidths and to look qualitatively at the spectral estimators for each case.

To address the problem of resolution, it should be evident that the flattening of the peaks in Figure 4.7 and Figure 4.8 was due to the fact that simple averaging was used in computing  $\bar{f}(\omega)$  defined in (4.54). There is no particular reason to use simple averaging, and we might improve the estimator by employing a weighted average, say

$$\hat{f}(\omega) = \sum_{k=-m}^m h_k I(\omega_j + k/n), \quad (4.64)$$

using the same definitions as in (4.54) but where the weights  $h_k > 0$  satisfy

$$\sum_{k=-m}^m h_k = 1.$$

In particular, it seems reasonable that the resolution of the estimator will improve if we use weights that decrease as distance from the center weight  $h_0$  increases; we will return to this idea shortly. To obtain the averaged periodogram,  $\bar{f}(\omega)$ , in (4.64), set  $h_k = L^{-1}$ , for all  $k$ , where  $L = 2m + 1$ . The asymptotic theory established for  $\bar{f}(\omega)$  still holds for  $\hat{f}(\omega)$  provided that the weights satisfy the additional condition that if  $m \rightarrow \infty$  as  $n \rightarrow \infty$  but  $m/n \rightarrow 0$ , then

$$\sum_{k=-m}^m h_k^2 \rightarrow 0.$$

Under these conditions, as  $n \rightarrow \infty$ ,

- (i)  $E(\hat{f}(\omega)) \rightarrow f(\omega)$
- (ii)  $\left(\sum_{k=-m}^m h_k^2\right)^{-1} \text{cov}(\hat{f}(\omega), \hat{f}(\lambda)) \rightarrow f^2(\omega)$  for  $\omega = \lambda \neq 0, 1/2$ .

In (ii), replace  $f^2(\omega)$  by 0 if  $\omega \neq \lambda$  and by  $2f^2(\omega)$  if  $\omega = \lambda = 0$  or  $1/2$ .

We have already seen these results in the case of  $\bar{f}(\omega)$ , where the weights are constant,  $h_k = L^{-1}$ , in which case  $\sum_{k=-m}^m h_k^2 = L^{-1}$ . The distributional properties of (4.64) are more difficult now because  $\hat{f}(\omega)$  is a weighted linear combination of asymptotically independent  $\chi^2$  random variables. An approximation that seems to work well is to replace  $L$  by  $\left(\sum_{k=-m}^m h_k^2\right)^{-1}$ . That is, define

$$L_h = \left(\sum_{k=-m}^m h_k^2\right)^{-1} \quad (4.65)$$



and use the approximation<sup>4.10</sup>

$$\frac{2L_h \hat{f}(\omega)}{f(\omega)} \sim \chi_{2L_h}^2. \quad (4.66)$$

In analogy to (4.56), we will define the bandwidth in this case to be

$$B = \frac{L_h}{n}. \quad (4.67)$$

Using the approximation (4.66) we obtain an approximate  $100(1 - \alpha)\%$  confidence interval of the form

$$\frac{2L_h \hat{f}(\omega)}{\chi_{2L_h}^2(1 - \alpha/2)} \leq f(\omega) \leq \frac{2L_h \hat{f}(\omega)}{\chi_{2L_h}^2(\alpha/2)} \quad (4.68)$$

for the true spectrum,  $f(\omega)$ . If the data are padded to  $n'$ , then replace  $2L_h$  in (4.68) with  $df = 2L_h n/n'$  as in (4.60).

An easy way to generate the weights in R is by repeated use of the *Daniell kernel*. For example, with  $m = 1$  and  $L = 2m + 1 = 3$ , the Daniell kernel has weights  $\{h_k\} = \{\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\}$ ; applying this kernel to a sequence of numbers,  $\{u_t\}$ , produces

$$\hat{u}_t = \frac{1}{3}u_{t-1} + \frac{1}{3}u_t + \frac{1}{3}u_{t+1}.$$

We can apply the same kernel again to the  $\hat{u}_t$ ,

$$\hat{\hat{u}}_t = \frac{1}{3}\hat{u}_{t-1} + \frac{1}{3}\hat{u}_t + \frac{1}{3}\hat{u}_{t+1},$$

which simplifies to

$$\hat{\hat{u}}_t = \frac{1}{9}u_{t-2} + \frac{2}{9}u_{t-1} + \frac{3}{9}u_t + \frac{2}{9}u_{t+1} + \frac{1}{9}u_{t+2}.$$

The *modified Daniell kernel* puts half weights at the end points, so with  $m = 1$  the weights are  $\{h_k\} = \{\frac{1}{4}, \frac{2}{4}, \frac{1}{4}\}$  and

$$\hat{u}_t = \frac{1}{4}u_{t-1} + \frac{1}{2}u_t + \frac{1}{4}u_{t+1}.$$

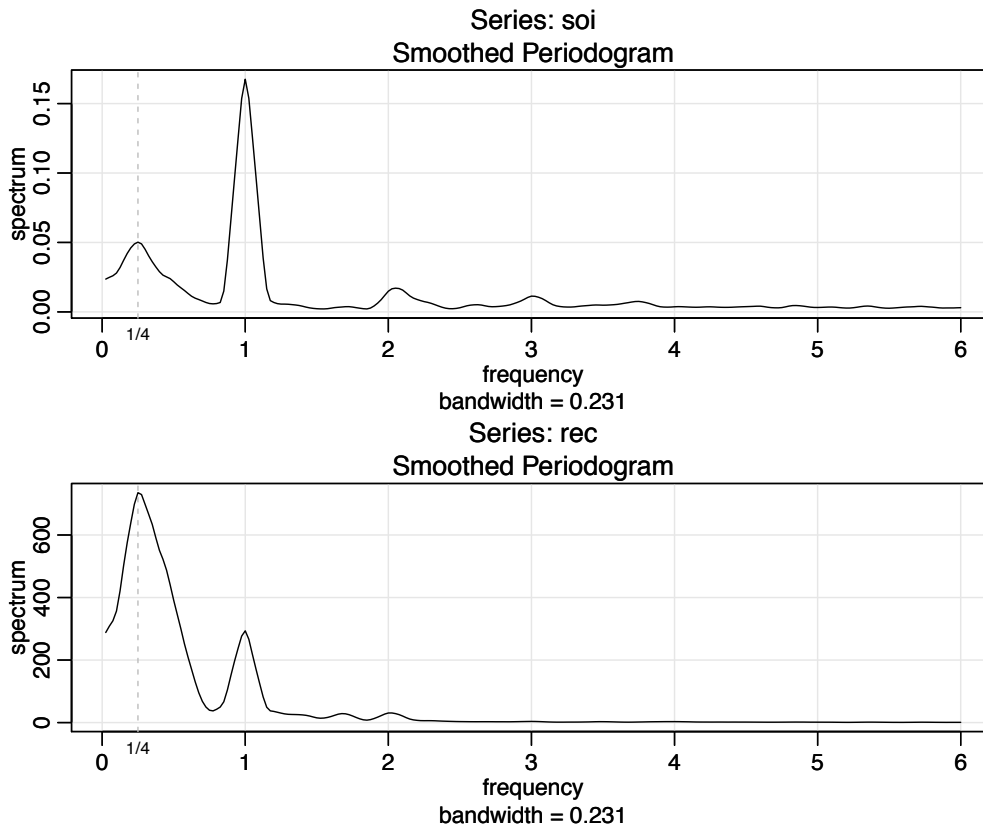
Applying the same kernel again to  $\hat{u}_t$  yields

$$\hat{\hat{u}}_t = \frac{1}{16}u_{t-2} + \frac{4}{16}u_{t-1} + \frac{6}{16}u_t + \frac{4}{16}u_{t+1} + \frac{1}{16}u_{t+2}.$$

These coefficients can be obtained in R by issuing the `kernel` command. For example, `kernel("modified.daniell", c(1,1))` would produce the coefficients of the last example. The other kernels that are currently available in R are the Dirichlet kernel and the Fejér kernel, which we will discuss shortly.

It is interesting to note that these kernel weights form a probability distribution. If  $X$  and  $Y$  are independent discrete uniforms on the integers  $\{-1, 0, 1\}$  each with probability  $\frac{1}{3}$ , then the convolution  $X + Y$  is discrete on the integers  $\{-2, -1, 0, 1, 2\}$  with corresponding probabilities  $\{\frac{1}{9}, \frac{2}{9}, \frac{3}{9}, \frac{2}{9}, \frac{1}{9}\}$ .

<sup>4.10</sup> The approximation proceeds as follows: If  $\hat{f} \sim c\chi_\nu^2$ , where  $c$  is a constant, then  $E\hat{f} \approx c\nu$  and  $\text{var}\hat{f} \approx f^2 \sum_k h_k^2 \approx c^2 2\nu$ . Solving,  $c \approx f \sum_k h_k^2 / 2 = f/2L_h$  and  $\nu \approx 2 \left( \sum_k h_k^2 \right)^{-1} = 2L_h$ .



**Fig. 4.10.** Smoothed (tapered) spectral estimates of the SOI and Recruitment series; see Example 4.16 for details.

#### Example 4.16 Smoothed Periodogram for SOI and Recruitment

In this example, we estimate the spectra of the SOI and Recruitment series using the smoothed periodogram estimate in (4.64). We used a modified Daniell kernel twice, with  $m = 3$  both times. This yields  $L_h = 1/\sum_{k=-m}^m h_k^2 = 9.232$ , which is close to the value of  $L = 9$  used in Example 4.14. In this case, the bandwidth is  $B = 9.232/480 = .019$  and the modified degrees of freedom is  $df = 2L_h 453/480 = 17.43$ . The weights,  $h_k$ , can be obtained and graphed in R as follows:

```
kernel("modified.daniell", c(3,3))
coef[-6] = 0.006944 = coef[ 6]
coef[-5] = 0.027778 = coef[ 5]
coef[-4] = 0.055556 = coef[ 4]
coef[-3] = 0.083333 = coef[ 3]
coef[-2] = 0.111111 = coef[ 2]
coef[-1] = 0.138889 = coef[ 1]
coef[ 0] = 0.152778
plot(kernel("modified.daniell", c(3,3))) # not shown
```

The resulting spectral estimates can be viewed in Figure 4.10 and we notice that the estimates more appealing than those in Figure 4.7. Figure 4.10 was generated in R as follows; we also show how to obtain the associated bandwidth and degrees of freedom.

```

k = kernel("modified.daniell", c(3,3))
soi.smo = mvspec(soi, kernel=k, taper=.1, log="no")
abline(v=c(.25,1), lty=2)
## Repeat above lines with rec replacing soi in line 3
df = soi.smo$df           # df = 17.42618
soi.smo$bandwidth        # B = 0.2308103

```

Note that a *taper* was applied in the estimation process; we discuss tapering in the next part. Reissuing the `mvspec` commands with `log="no"` removed will result in a figure similar to [Figure 4.8](#). Finally, we mention that the modified Daniell kernel is used by default and an easier way to obtain `soi.smo` is to issue the command:

```
soi.smo = mvspec(soi, taper=.1, spans=c(7,7))
```

Notice that `spans` is a vector of odd integers, given in terms of  $L = 2m + 1$  instead of  $m$ .

There have been many attempts at dealing with the problem of smoothing the periodogram in an automatic way; an early reference is Wahba (1980). It is apparent from [Example 4.16](#) that the smoothing bandwidth for the broadband El Niño behavior (near the 4 year cycle), should be much larger than the bandwidth for the annual cycle (the 1 year cycle). Consequently, it is perhaps better to perform automatic adaptive smoothing for estimating the spectrum. We refer interested readers to Fan and Kreutzberger (1998) and the numerous references within.

#### TAPERING

We are now ready to introduce the concept of *tapering*; a more detailed discussion may be found in Bloomfield (2000, §9.5). Suppose  $x_t$  is a mean-zero, stationary process with spectral density  $f_x(\omega)$ . If we replace the original series by the tapered series

$$y_t = h_t x_t, \quad (4.69)$$

for  $t = 1, 2, \dots, n$ , use the modified DFT

$$d_y(\omega_j) = n^{-1/2} \sum_{t=1}^n h_t x_t e^{-2\pi i \omega_j t}, \quad (4.70)$$

and let  $I_y(\omega_j) = |d_y(\omega_j)|^2$ , we obtain (see [Problem 4.17](#))

$$E[I_y(\omega_j)] = \int_{-\frac{1}{2}}^{\frac{1}{2}} W_n(\omega_j - \omega) f_x(\omega) d\omega \quad (4.71)$$

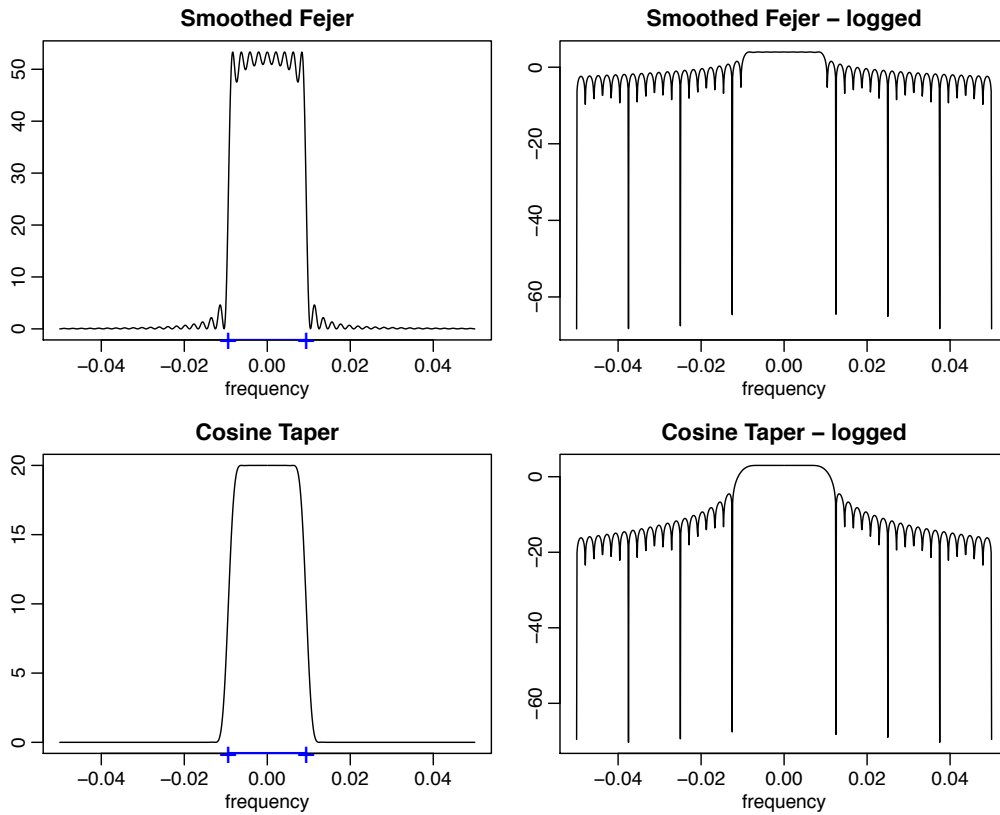
where

$$W_n(\omega) = |H_n(\omega)|^2 \quad (4.72)$$

and

$$H_n(\omega) = n^{-1/2} \sum_{t=1}^n h_t e^{-2\pi i \omega t}. \quad (4.73)$$

The value  $W_n(\omega)$  is called a *spectral window* because, in view of (4.71), it is determining which part of the spectral density  $f_x(\omega)$  is being “seen” by the estimator



**Fig. 4.11.** Averaged Fejér window (top row) and the corresponding cosine taper window (bottom row) for  $L = 9$ ,  $n = 480$ . The extra tic marks on the horizontal axis of the left-hand plots exhibit the predicted bandwidth,  $B = 9/480 = .01875$ .

$I_y(\omega_j)$  on average. In the case that  $h_t = 1$  for all  $t$ ,  $I_y(\omega_j) = I_x(\omega_j)$  is simply the periodogram of the data and the window is

$$W_n(\omega) = \frac{\sin^2(n\pi\omega)}{n \sin^2(\pi\omega)} \tag{4.74}$$

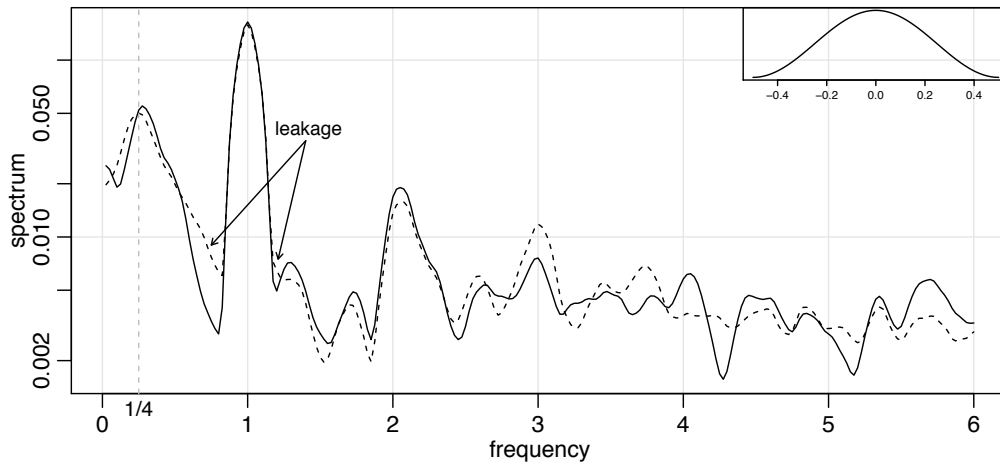
with  $W_n(0) = n$ , which is known as the Fejér or modified Bartlett kernel. If we consider the averaged periodogram in (4.54), namely

$$\bar{f}_x(\omega) = \frac{1}{L} \sum_{k=-m}^m I_x(\omega_j + k/n),$$

the window,  $W_n(\omega)$ , in (4.71) will take the form

$$W_n(\omega) = \frac{1}{nL} \sum_{k=-m}^m \frac{\sin^2[n\pi(\omega + k/n)]}{\sin^2[\pi(\omega + k/n)]}. \tag{4.75}$$

Tapers generally have a shape that enhances the center of the data relative to the extremities, such as a cosine bell of the form



**Fig. 4.12.** Smoothed spectral estimates of the SOI without tapering (dashed line) and with full tapering (solid line); see Example 4.17. The insert shows a full cosine bell taper, (4.76), with horizontal axis  $(t - \bar{t})/n$ , for  $t = 1, \dots, n$ .

$$h_t = .5 \left[ 1 + \cos \left( \frac{2\pi(t - \bar{t})}{n} \right) \right], \quad (4.76)$$

where  $\bar{t} = (n + 1)/2$ , favored by Blackman and Tukey (1959). The shape of this taper is shown in the insert to Figure 4.12. In Figure 4.11, we have plotted the shapes of two windows,  $W_n(\omega)$ , for  $n = 480$  and  $L = 9$ , when (i)  $h_t \equiv 1$ , in which case, (4.75) applies, and (ii)  $h_t$  is the cosine taper in (4.76). In both cases the predicted bandwidth should be  $B = 9/480 = .01875$  cycles per point, which corresponds to the “width” of the windows shown in Figure 4.11. Both windows produce an integrated average spectrum over this band but the untapered window in the top panels shows considerable ripples over the band and outside the band. The ripples outside the band are called sidelobes and tend to introduce frequencies from outside the interval that may contaminate the desired spectral estimate within the band. For example, a large dynamic range for the values in the spectrum introduces spectra in contiguous frequency intervals several orders of magnitude greater than the value in the interval of interest. This effect is sometimes called *leakage*. Figure 4.11 emphasizes the suppression of the sidelobes in the Fejér kernel when a cosine taper is used.

#### Example 4.17 The Effect of Tapering the SOI Series

The estimates in Example 4.16 were obtained by tapering the upper and lower 10% of the data. In this example, we examine the effect of tapering on the estimate of the spectrum of the SOI series (the results for the Recruitment series are similar). Figure 4.12 shows two spectral estimates plotted on a log scale. The dashed line in Figure 4.12 shows the estimate without any tapering. The solid line shows the result with full tapering. Notice that the tapered spectrum does a better job in separating the yearly cycle ( $\omega = 1$ ) and the El Niño cycle ( $\omega = 1/4$ ).

The following R session was used to generate Figure 4.12. We note that, by default, `mvspec` does not taper. For full tapering, we use the argument `taper=.5` to

instruct `mvspec` to taper 50% of each end of the data; any value between 0 and .5 is acceptable. In Example 4.16, we used `taper=.1`.

```
s0 = mvspec(soi, spans=c(7,7), plot=FALSE)           # no taper
s50 = mvspec(soi, spans=c(7,7), taper=.5, plot=FALSE) # full taper
plot(s0$freq, s50$spec, log="y", type="l", ylab="spectrum",
      xlab="frequency")                               # solid line
lines(s0$freq, s0$spec, lty=2)                       # dashed line
```

We close this section with a brief discussion of *lag window* estimators. First, consider the periodogram,  $I(\omega_j)$ , which was shown in (4.30) to be

$$I(\omega_j) = \sum_{|h|<n} \hat{\gamma}(h) e^{-2\pi i \omega_j h}.$$

Thus, (4.64) can be written as

$$\begin{aligned} \hat{f}(\omega) &= \sum_{|k|\leq m} h_k I(\omega_j + k/n) = \sum_{|k|\leq m} h_k \sum_{|h|<n} \hat{\gamma}(h) e^{-2\pi i (\omega_j + k/n)h} \\ &= \sum_{|h|<n} g\left(\frac{h}{n}\right) \hat{\gamma}(h) e^{-2\pi i \omega_j h}. \end{aligned} \quad (4.77)$$

where  $g\left(\frac{h}{n}\right) = \sum_{|k|\leq m} h_k \exp(-2\pi i k h/n)$ . Equation (4.77) suggests estimators of the form

$$\tilde{f}(\omega) = \sum_{|h|\leq r} w\left(\frac{h}{r}\right) \hat{\gamma}(h) e^{-2\pi i \omega h} \quad (4.78)$$

where  $w(\cdot)$  is a weight function, called the *lag window*, that satisfies

- (i)  $w(0) = 1$
- (ii)  $|w(x)| \leq 1$  and  $w(x) = 0$  for  $|x| > 1$ ,
- (iii)  $w(x) = w(-x)$ .

Note that if  $w(x) = 1$  for  $|x| < 1$  and  $r = n$ , then  $\tilde{f}(\omega_j) = I(\omega_j)$ , the periodogram. This result indicates the problem with the periodogram as an estimator of the spectral density is that it gives too much weight to the values of  $\hat{\gamma}(h)$  when  $h$  is large, and hence is unreliable [e.g, there is only one pair of observations used in the estimate  $\hat{\gamma}(n-1)$ , and so on]. The smoothing window is defined to be

$$W(\omega) = \sum_{h=-r}^r w\left(\frac{h}{r}\right) e^{-2\pi i \omega h}, \quad (4.79)$$

and it determines which part of the periodogram will be used to form the estimate of  $f(\omega)$ . The asymptotic theory for  $\hat{f}(\omega)$  holds for  $\tilde{f}(\omega)$  under the same conditions and provided  $r \rightarrow \infty$  as  $n \rightarrow \infty$  but with  $r/n \rightarrow 0$ . That is,

$$E\{\tilde{f}(\omega)\} \rightarrow f(\omega), \quad (4.80)$$

$$\frac{n}{r} \text{cov}(\tilde{f}(\omega), \tilde{f}(\lambda)) \rightarrow f^2(\omega) \int_{-1}^1 w^2(x) dx \quad \omega = \lambda \neq 0, 1/2. \quad (4.81)$$

In (4.81), replace  $f^2(\omega)$  by 0 if  $\omega \neq \lambda$  and by  $2f^2(\omega)$  if  $\omega = \lambda = 0$  or  $1/2$ .

Many authors have developed various windows and Brillinger (2001, Ch 3) and Brockwell and Davis (1991, Ch 10) are good sources of detailed information on this topic.

## 4.5 Parametric Spectral Estimation

The methods of the previous section lead to what is generally referred to as *non-parametric spectral estimators* because no assumption is made about the parametric form of the spectral density. In Property 4.4, we exhibited the spectrum of an ARMA process and we might consider basing a spectral estimator on this function, substituting the parameter estimates from an ARMA( $p, q$ ) fit on the data into the formula for the spectral density  $f_x(\omega)$  given in (4.23). Such an estimator is called a parametric spectral estimator. For convenience, a parametric spectral estimator is obtained by fitting an AR( $p$ ) to the data, where the order  $p$  is determined by one of the model selection criteria, such as AIC, AICc, and BIC, defined in (2.15)–(2.17). Parametric autoregressive spectral estimators will often have superior resolution in problems when several closely spaced narrow spectral peaks are present and are preferred by engineers for a broad variety of problems (see Kay, 1988). The development of autoregressive spectral estimators has been summarized by Parzen (1983).

If  $\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p$  and  $\hat{\sigma}_w^2$  are the estimates from an AR( $p$ ) fit to  $x_t$ , then based on Property 4.4, a parametric spectral estimate of  $f_x(\omega)$  is attained by substituting these estimates into (4.23), that is,

$$\hat{f}_x(\omega) = \frac{\hat{\sigma}_w^2}{|\hat{\phi}(e^{-2\pi i\omega})|^2}, \quad (4.82)$$

where

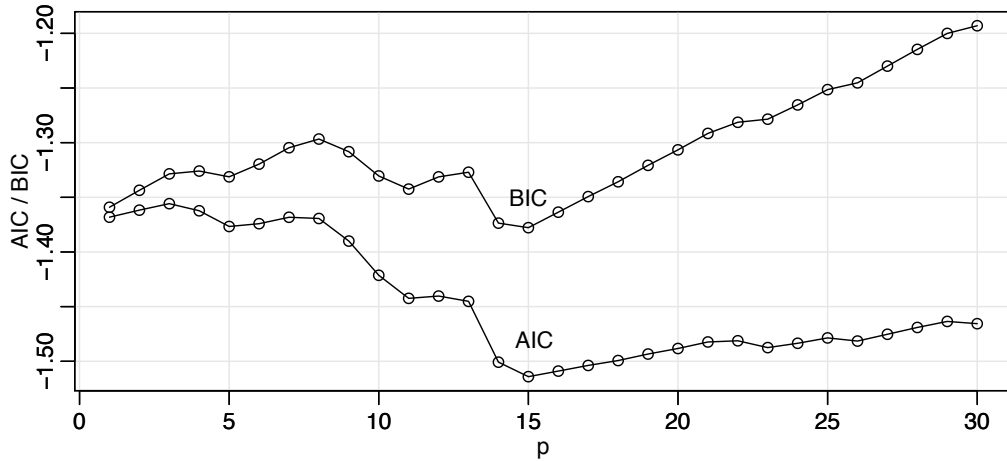
$$\hat{\phi}(z) = 1 - \hat{\phi}_1 z - \hat{\phi}_2 z^2 - \dots - \hat{\phi}_p z^p. \quad (4.83)$$

The asymptotic distribution of the autoregressive spectral estimator has been obtained by Berk (1974) under the conditions  $p \rightarrow \infty$ ,  $p^3/n \rightarrow 0$  as  $p, n \rightarrow \infty$ , which may be too severe for most applications. The limiting results imply a confidence interval of the form

$$\frac{\hat{f}_x(\omega)}{(1 + Cz_{\alpha/2})} \leq f_x(\omega) \leq \frac{\hat{f}_x(\omega)}{(1 - Cz_{\alpha/2})}, \quad (4.84)$$

where  $C = \sqrt{2p/n}$  and  $z_{\alpha/2}$  is the ordinate corresponding to the upper  $\alpha/2$  probability of the standard normal distribution. If the sampling distribution is to be checked, we suggest applying the bootstrap estimator to get the sampling distribution of  $\hat{f}_x(\omega)$  using a procedure similar to the one used for  $p = 1$  in Example 3.36. An alternative for higher order autoregressive series is to put the AR( $p$ ) in state-space form and use the bootstrap procedure discussed in Section 6.7.

An interesting fact about rational spectra of the form (4.23) is that any spectral density can be approximated, arbitrarily close, by the spectrum of an AR process.



**Fig. 4.13.** Model selection criteria AIC and BIC as a function of order  $p$  for autoregressive models fitted to the SOI series.

#### Property 4.7 AR Spectral Approximation

Let  $g(\omega)$  be the spectral density of a stationary process. Then, given  $\epsilon > 0$ , there is a time series with the representation

$$x_t = \sum_{k=1}^p \phi_k x_{t-k} + w_t$$

where  $w_t$  is white noise with variance  $\sigma_w^2$ , such that

$$|f_x(\omega) - g(\omega)| < \epsilon \quad \text{for all } \omega \in [-1/2, 1/2].$$

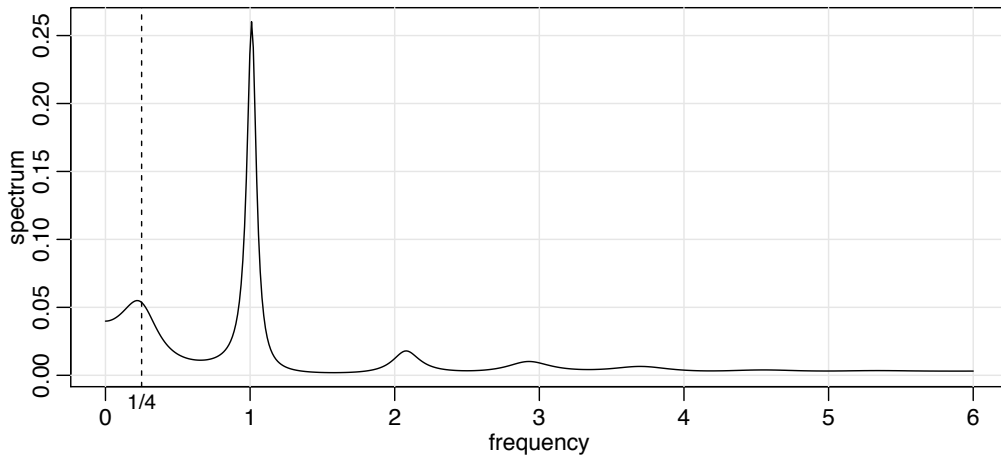
Moreover,  $p$  is finite and the roots of  $\phi(z) = 1 - \sum_{k=1}^p \phi_k z^k$  are outside the unit circle.

One drawback of the property is that it does not tell us how large  $p$  must be before the approximation is reasonable; in some situations  $p$  may be extremely large. Property 4.7 also holds for MA and for ARMA processes in general, and a proof of the result may be found in Section C.6. We demonstrate the technique in the following example.

#### Example 4.18 Autoregressive Spectral Estimator for SOI

Consider obtaining results comparable to the nonparametric estimators shown in Figure 4.7 for the SOI series. Fitting successively higher order AR( $p$ ) models for  $p = 1, 2, \dots, 30$  yields a minimum BIC and a minimum AIC at  $p = 15$ , as shown in Figure 4.13. We can see from Figure 4.13 that BIC is very definite about which model it chooses; that is, the minimum BIC is very distinct. On the other hand, it is not clear what is going to happen with AIC; that is, the minimum is not so clear, and there is some concern that AIC will start decreasing after  $p = 30$ . Minimum AICc selects the  $p = 15$  model, but suffers from the same uncertainty as AIC. The spectrum is shown in Figure 4.14, and we note the strong peaks near the four year





**Fig. 4.14.** Autoregressive spectral estimator for the SOI series using the AR(15) model selected by AIC, AICc, and BIC.

and one year cycles as in the nonparametric estimates obtained in Section 4.4. In addition, the harmonics of the yearly period are evident in the estimated spectrum.

To perform a similar analysis in R, the command `spec.ar` can be used to fit the best model via AIC and plot the resulting spectrum. A quick way to obtain the AIC values is to run the `ar` command as follows.

```

spaic = spec.ar(soi, log="no")           # min AIC spec
abline(v=frequency(soi)*1/52, lty=3)   # El Nino peak
(soi.ar = ar(soi, order.max=30))       # estimates and AICs
dev.new()
plot(1:30, soi.ar$aic[-1], type="o")    # plot AICs

```

No likelihood is calculated here, so the use of the term AIC is loose. To generate Figure 4.13 we used the following code to (loosely) obtain AIC, AICc, and BIC. Because AIC and AICc are nearly identical in this example, we only graphed AIC and BIC+1; we added 1 to the BIC to reduce white space in the graphic.

```

n = length(soi)
AIC = rep(0, 30) -> AICc -> BIC
for (k in 1:30){
  sigma2 = ar(soi, order=k, aic=FALSE)$var.pred
  BIC[k] = log(sigma2) + (k*log(n)/n)
  AICc[k] = log(sigma2) + ((n+k)/(n-k-2))
  AIC[k] = log(sigma2) + ((n+2*k)/n) }
IC = cbind(AIC, BIC+1)
ts.plot(IC, type="o", xlab="p", ylab="AIC / BIC")

```

Finally, it should be mentioned that any parametric spectrum, say  $f(\omega; \theta)$ , depending on the vector parameter  $\theta$  can be estimated via the Whittle likelihood (Whittle, 1961), using the approximate properties of the discrete Fourier transform derived in Appendix C. We have that the DFTs,  $d(\omega_j)$ , are approximately complex normally distributed with mean zero and variance  $f(\omega_j; \theta)$  and are approximately independent for  $\omega_j \neq \omega_k$ . This implies that an approximate log likelihood can be written in the form

$$\ln L(x; \theta) \approx - \sum_{0 < \omega_j < 1/2} \left( \ln f_x(\omega_j; \theta) + \frac{|d(\omega_j)|^2}{f_x(\omega_j; \theta)} \right), \quad (4.85)$$

where the sum is sometimes expanded to include the frequencies  $\omega_j = 0, 1/2$ . If the form with the two additional frequencies is used, the multiplier of the sum will be unity, except for the purely real points at  $\omega_j = 0, 1/2$  for which the multiplier is  $1/2$ . For a discussion of applying the Whittle approximation to the problem of estimating parameters in an ARMA spectrum, see Anderson (1978). The Whittle likelihood is especially useful for fitting long memory models that will be discussed in Chapter 5.

## 4.6 Multiple Series and Cross-Spectra

The notion of analyzing frequency fluctuations using classical statistical ideas extends to the case in which there are several jointly stationary series, for example,  $x_t$  and  $y_t$ . In this case, we can introduce the idea of a correlation indexed by frequency, called the *coherence*. The results in Section C.2 imply the covariance function

$$\gamma_{xy}(h) = E[(x_{t+h} - \mu_x)(y_t - \mu_y)]$$

has the representation

$$\gamma_{xy}(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} f_{xy}(\omega) e^{2\pi i \omega h} d\omega \quad h = 0, \pm 1, \pm 2, \dots, \quad (4.86)$$

where the *cross-spectrum* is defined as the Fourier transform

$$f_{xy}(\omega) = \sum_{h=-\infty}^{\infty} \gamma_{xy}(h) e^{-2\pi i \omega h} \quad -1/2 \leq \omega \leq 1/2, \quad (4.87)$$

assuming that the cross-covariance function is absolutely summable, as was the case for the autocovariance. The cross-spectrum is generally a complex-valued function, and it is often written as

$$f_{xy}(\omega) = c_{xy}(\omega) - iq_{xy}(\omega), \quad (4.88)$$

where

$$c_{xy}(\omega) = \sum_{h=-\infty}^{\infty} \gamma_{xy}(h) \cos(2\pi \omega h) \quad (4.89)$$

and

$$q_{xy}(\omega) = \sum_{h=-\infty}^{\infty} \gamma_{xy}(h) \sin(2\pi \omega h) \quad (4.90)$$

are defined as the *cospectrum* and *quadspectrum*, respectively. Because of the relationship  $\gamma_{yx}(h) = \gamma_{xy}(-h)$ , it follows, by substituting into (4.87) and rearranging, that

$$f_{yx}(\omega) = f_{xy}^*(\omega), \tag{4.91}$$

with \* denoting conjugation. This result, in turn, implies that the cospectrum and quadspectrum satisfy

$$c_{yx}(\omega) = c_{xy}(\omega) \tag{4.92}$$

and

$$q_{yx}(\omega) = -q_{xy}(\omega). \tag{4.93}$$

An important example of the application of the cross-spectrum is to the problem of predicting an output series  $y_t$  from some input series  $x_t$  through a linear filter relation such as the three-point moving average considered below. A measure of the strength of such a relation is the *squared coherence* function, defined as

$$\rho_{y \cdot x}^2(\omega) = \frac{|f_{yx}(\omega)|^2}{f_{xx}(\omega)f_{yy}(\omega)}, \tag{4.94}$$

where  $f_{xx}(\omega)$  and  $f_{yy}(\omega)$  are the individual spectra of the  $x_t$  and  $y_t$  series, respectively. Although we consider a more general form of this that applies to multiple inputs later, it is instructive to display the single input case as (4.94) to emphasize the analogy with conventional squared correlation, which takes the form

$$\rho_{yx}^2 = \frac{\sigma_{yx}^2}{\sigma_x^2 \sigma_y^2},$$

for random variables with variances  $\sigma_x^2$  and  $\sigma_y^2$  and covariance  $\sigma_{yx} = \sigma_{xy}$ . This motivates the interpretation of squared coherence and the squared correlation between two time series at frequency  $\omega$ .

**Example 4.19 Three-Point Moving Average**

As a simple example, we compute the cross-spectrum between  $x_t$  and the three-point moving average  $y_t = (x_{t-1} + x_t + x_{t+1})/3$ , where  $x_t$  is a stationary input process with spectral density  $f_{xx}(\omega)$ . First,

$$\begin{aligned} \gamma_{xy}(h) &= \text{cov}(x_{t+h}, y_t) = \frac{1}{3} \text{cov}(x_{t+h}, x_{t-1} + x_t + x_{t+1}) \\ &= \frac{1}{3} [\gamma_{xx}(h+1) + \gamma_{xx}(h) + \gamma_{xx}(h-1)] \\ &= \frac{1}{3} \int_{-\frac{1}{2}}^{\frac{1}{2}} (e^{2\pi i \omega} + 1 + e^{-2\pi i \omega}) e^{2\pi i \omega h} f_{xx}(\omega) d\omega \\ &= \frac{1}{3} \int_{-\frac{1}{2}}^{\frac{1}{2}} [1 + 2 \cos(2\pi \omega)] f_{xx}(\omega) e^{2\pi i \omega h} d\omega, \end{aligned}$$

where we have use (4.16). Using the uniqueness of the Fourier transform, we argue from the spectral representation (4.86) that

$$f_{xy}(\omega) = \frac{1}{3} [1 + 2 \cos(2\pi \omega)] f_{xx}(\omega)$$

so that the cross-spectrum is real in this case. Using **Property 4.3**, the spectral density of  $y_t$  is

$$f_{yy}(\omega) = \frac{1}{9} |e^{2\pi i\omega} + 1 + e^{-2\pi i\omega}|^2 f_{xx}(\omega) = \frac{1}{9} [1 + 2 \cos(2\pi\omega)]^2 f_{xx}(\omega).$$

Substituting into (4.94) yields,

$$\rho_{y \cdot x}^2(\omega) = \frac{\left| \frac{1}{3} [1 + 2 \cos(2\pi\omega)] f_{xx}(\omega) \right|^2}{f_{xx}(\omega) \cdot \frac{1}{9} [1 + 2 \cos(2\pi\omega)]^2 f_{xx}(\omega)} = 1;$$

that is, the squared coherence between  $x_t$  and  $y_t$  is unity over all frequencies. This is a characteristic inherited by more general linear filters; see **Problem 4.30**. However, if some noise is added to the three-point moving average, the coherence is not unity; these kinds of models will be considered in detail later.

#### **Property 4.8 Spectral Representation of a Vector Stationary Process**

If  $x_t = (x_{t1}, x_{t2}, \dots, x_{tp})'$  is a  $p \times 1$  stationary process with autocovariance matrix  $\Gamma(h) = E[(x_{t+h} - \mu)(x_t - \mu)'] = \{\gamma_{jk}(h)\}$  satisfying

$$\sum_{h=-\infty}^{\infty} |\gamma_{jk}(h)| < \infty \quad (4.95)$$

for all  $j, k = 1, \dots, p$ , then  $\Gamma(h)$  has the representation

$$\Gamma(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i\omega h} f(\omega) d\omega \quad h = 0, \pm 1, \pm 2, \dots, \quad (4.96)$$

as the inverse transform of the spectral density matrix,  $f(\omega) = \{f_{jk}(\omega)\}$ , for  $j, k = 1, \dots, p$ . The matrix  $f(\omega)$  has the representation

$$f(\omega) = \sum_{h=-\infty}^{\infty} \Gamma(h) e^{-2\pi i\omega h} \quad -1/2 \leq \omega \leq 1/2. \quad (4.97)$$

The spectral matrix  $f(\omega)$  is Hermitian,  $f(\omega) = f^*(\omega)$ , where  $*$  means to conjugate and transpose.

#### **Example 4.20 Spectral Matrix of a Bivariate Process**

Consider a jointly stationary bivariate process  $(x_t, y_t)$ . We arrange the autocovariances in the matrix

$$\Gamma(h) = \begin{pmatrix} \gamma_{xx}(h) & \gamma_{xy}(h) \\ \gamma_{yx}(h) & \gamma_{yy}(h) \end{pmatrix}.$$

The spectral matrix would be given by

$$f(\omega) = \begin{pmatrix} f_{xx}(\omega) & f_{xy}(\omega) \\ f_{yx}(\omega) & f_{yy}(\omega) \end{pmatrix},$$

where the Fourier transform (4.96) and (4.97) relate the autocovariance and spectral matrices.

The extension of spectral estimation to vector series is fairly obvious. For the vector series  $x_t = (x_{t1}, x_{t2}, \dots, x_{tp})'$ , we may use the vector of DFTs, say  $d(\omega_j) = (d_1(\omega_j), d_2(\omega_j), \dots, d_p(\omega_j))'$ , and estimate the spectral matrix by

$$\bar{f}(\omega) = L^{-1} \sum_{k=-m}^m I(\omega_j + k/n) \quad (4.98)$$

where now

$$I(\omega_j) = d(\omega_j) d^*(\omega_j) \quad (4.99)$$

is a  $p \times p$  complex matrix. The series may be tapered before the DFT is taken in (4.98) and we can use weighted estimation,

$$\hat{f}(\omega) = \sum_{k=-m}^m h_k I(\omega_j + k/n) \quad (4.100)$$

where  $\{h_k\}$  are weights as defined in (4.64). The estimate of squared coherence between two series,  $y_t$  and  $x_t$  is

$$\hat{\rho}_{y \cdot x}^2(\omega) = \frac{|\hat{f}_{yx}(\omega)|^2}{\hat{f}_{xx}(\omega)\hat{f}_{yy}(\omega)}. \quad (4.101)$$

If the spectral estimates in (4.101) are obtained using equal weights, we will write  $\bar{\rho}_{y \cdot x}^2(\omega)$  for the estimate.

Under general conditions, if  $\rho_{y \cdot x}^2(\omega) > 0$  then

$$|\hat{\rho}_{y \cdot x}(\omega)| \sim AN \left( |\rho_{y \cdot x}(\omega)|, (1 - \rho_{y \cdot x}^2(\omega))^2 / 2L_h \right) \quad (4.102)$$

where  $L_h$  is defined in (4.65); the details of this result may be found in Brockwell and Davis (1991, Ch 11). We may use (4.102) to obtain approximate confidence intervals for the squared coherence,  $\rho_{y \cdot x}^2(\omega)$ .

We may also test the null hypothesis that  $\rho_{y \cdot x}^2(\omega) = 0$  if we use  $\bar{\rho}_{y \cdot x}^2(\omega)$  for the estimate with  $L > 1$ ,<sup>4.11</sup> that is,

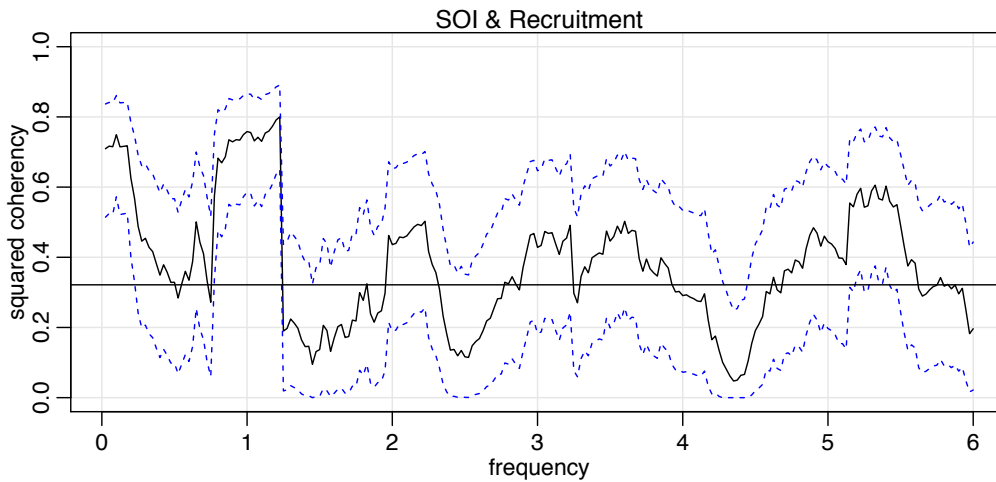
$$\bar{\rho}_{y \cdot x}^2(\omega) = \frac{|\bar{f}_{yx}(\omega)|^2}{\bar{f}_{xx}(\omega)\bar{f}_{yy}(\omega)}. \quad (4.103)$$

In this case, under the null hypothesis, the statistic

$$F = \frac{\bar{\rho}_{y \cdot x}^2(\omega)}{(1 - \bar{\rho}_{y \cdot x}^2(\omega))} (L - 1) \quad (4.104)$$

has an approximate  $F$ -distribution with 2 and  $2L - 2$  degrees of freedom. When the series have been extended to length  $n'$ , we replace  $2L - 2$  by  $df - 2$ , where  $df$  is defined in (4.60). Solving (4.104) for a particular significance level  $\alpha$  leads to

<sup>4.11</sup> If  $L = 1$  then  $\bar{\rho}_{y \cdot x}^2(\omega) \equiv 1$ .



**Fig. 4.15.** Squared coherency between the SOI and Recruitment series;  $L = 19$ ,  $n = 453$ ,  $n' = 480$ , and  $\alpha = .001$ . The horizontal line is  $C_{.001}$ .

$$C_{\alpha} = \frac{F_{2,2L-2}(\alpha)}{L-1 + F_{2,2L-2}(\alpha)} \quad (4.105)$$

as the approximate value that must be exceeded for the original squared coherence to be able to reject  $\rho_{y \cdot x}^2(\omega) = 0$  at an a priori specified frequency.

#### Example 4.21 Coherence Between SOI and Recruitment

Figure 4.15 shows the squared coherence between the SOI and Recruitment series over a wider band than was used for the spectrum. In this case, we used  $L = 19$ ,  $df = 2(19)(453/480) \approx 36$  and  $F_{2,df-2}(.001) \approx 8.53$  at the significance level  $\alpha = .001$ . Hence, we may reject the hypothesis of no coherence for values of  $\bar{\rho}_{y \cdot x}^2(\omega)$  that exceed  $C_{.001} = .32$ . We emphasize that this method is crude because, in addition to the fact that the  $F$ -statistic is approximate, we are examining the squared coherence across all frequencies with the Bonferroni inequality, (4.63), in mind. Figure 4.15 also exhibits confidence bands as part of the R plotting routine. We emphasize that these bands are only valid for  $\omega$  where  $\rho_{y \cdot x}^2(\omega) > 0$ .

In this case, the two series are obviously strongly coherent at the annual seasonal frequency. The series are also strongly coherent at lower frequencies that may be attributed to the El Niño cycle, which we claimed had a 3 to 7 year period. The peak in the coherence, however, occurs closer to the 9 year cycle. Other frequencies are also coherent, although the strong coherence is less impressive because the underlying power spectrum at these higher frequencies is fairly small. Finally, we note that the coherence is persistent at the seasonal harmonic frequencies.

This example may be reproduced using the following R commands.

```
sr = mvspec(cbind(soi,rec), kernel("daniell",9), plot=FALSE)
sr$df # df = 35.8625
f = qf(.999, 2, sr$df-2) # = 8.529792
C = f/(18+f) # = 0.321517
plot(sr, plot.type = "coh", ci.lty = 2)
abline(h = C)
```

## 4.7 Linear Filters

Some of the examples of the previous sections have hinted at the possibility the distribution of power or variance in a time series can be modified by making a linear transformation. In this section, we explore that notion further by showing how linear filters can be used to extract signals from a time series. These filters modify the spectral characteristics of a time series in a predictable way, and the systematic development of methods for taking advantage of the special properties of linear filters is an important topic in time series analysis.

Recall **Property 4.3** that stated if

$$y_t = \sum_{j=-\infty}^{\infty} a_j x_{t-j}, \quad \sum_{j=-\infty}^{\infty} |a_j| < \infty,$$

and  $x_t$  has spectrum  $f_{xx}(\omega)$ , then  $y_t$  has spectrum

$$f_{yy}(\omega) = |A_{yx}(\omega)|^2 f_{xx}(\omega),$$

where

$$A_{yx}(\omega) = \sum_{j=-\infty}^{\infty} a_j e^{-2\pi i \omega j}$$

is the *frequency response function*. This result shows that the filtering effect can be characterized as a frequency-by-frequency multiplication by the squared magnitude of the frequency response function.

### Example 4.22 First Difference and Moving Average Filters

We illustrate the effect of filtering with two common examples, the first difference filter

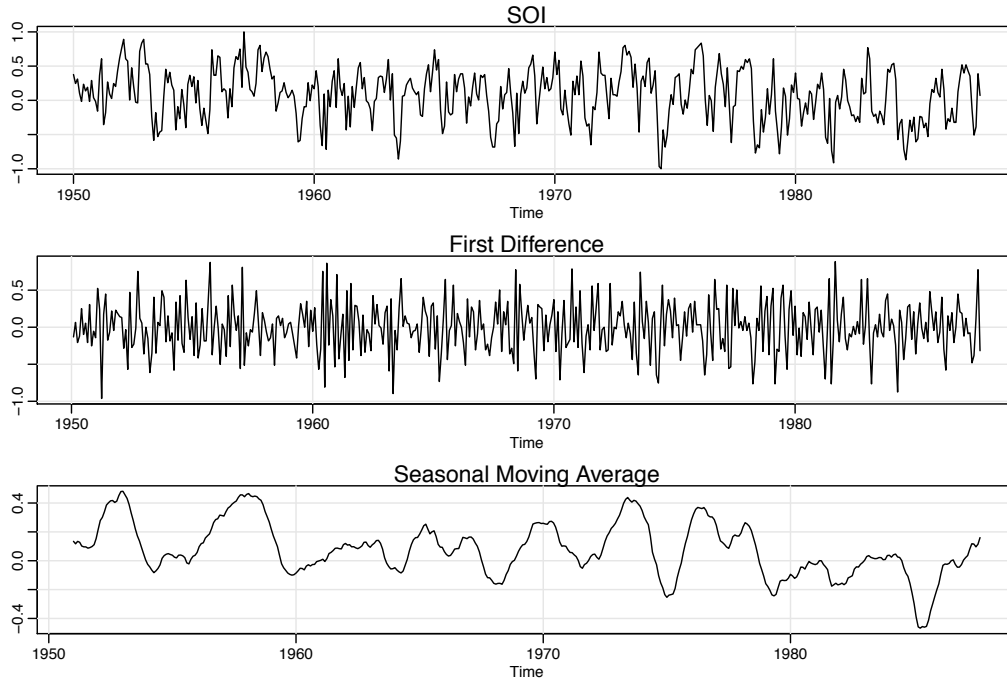
$$y_t = \nabla x_t = x_t - x_{t-1}$$

and the annual symmetric moving average filter,

$$y_t = \frac{1}{24} (x_{t-6} + x_{t+6}) + \frac{1}{12} \sum_{r=-5}^5 x_{t-r},$$

which is a modified Daniell kernel with  $m = 6$ . The results of filtering the SOI series using the two filters are shown in the middle and bottom panels of **Figure 4.16**. Notice that the effect of differencing is to roughen the series because it tends to retain the higher or faster frequencies. The centered moving average smooths the series because it retains the lower frequencies and tends to attenuate the higher frequencies. In general, differencing is an example of a *high-pass filter* because it retains or passes the higher frequencies, whereas the moving average is a *low-pass filter* because it passes the lower or slower frequencies.

Notice that the slower periods are enhanced in the symmetric moving average and the seasonal or yearly frequencies are attenuated. The filtered series makes



**Fig. 4.16.** SOI series (top) compared with the differenced SOI (middle) and a centered 12-month moving average (bottom).

about 9 cycles in the length of the data (about one cycle every 52 months) and the moving average filter tends to enhance or *extract* the El Niño signal. Moreover, by low-pass filtering the data, we get a better sense of the El Niño effect and its irregularity.

Now, having done the filtering, it is essential to determine the exact way in which the filters change the input spectrum. We shall use (4.21) and (4.22) for this purpose. The first difference filter can be written in the form (4.20) by letting  $a_0 = 1$ ,  $a_1 = -1$ , and  $a_r = 0$  otherwise. This implies that

$$A_{y,x}(\omega) = 1 - e^{-2\pi i\omega},$$

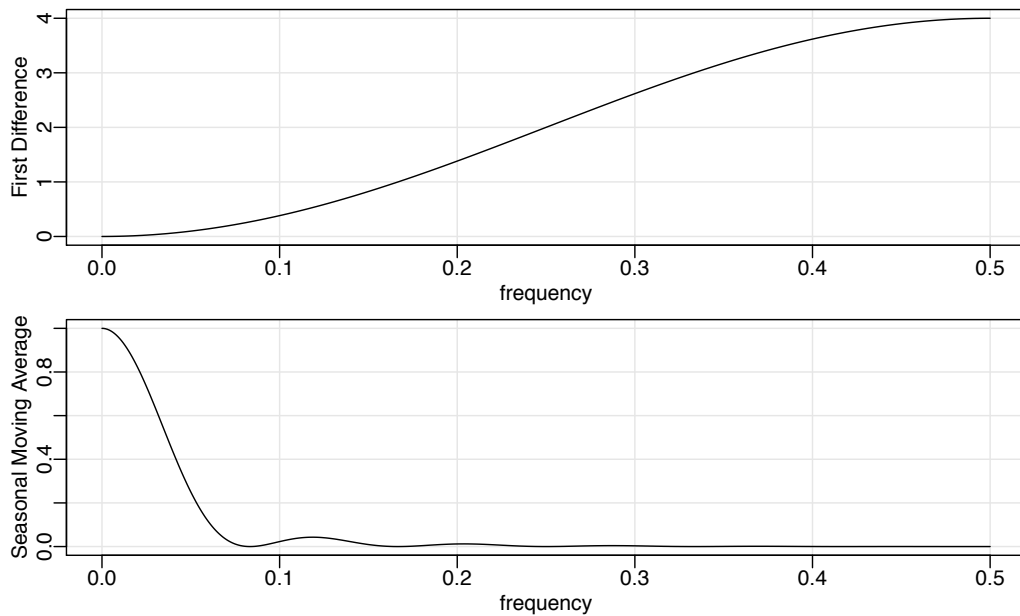
and the squared frequency response becomes

$$|A_{y,x}(\omega)|^2 = (1 - e^{-2\pi i\omega})(1 - e^{2\pi i\omega}) = 2[1 - \cos(2\pi\omega)]. \quad (4.106)$$

The top panel of Figure 4.17 shows that the first difference filter will attenuate the lower frequencies and enhance the higher frequencies because the multiplier of the spectrum,  $|A_{y,x}(\omega)|^2$ , is large for the higher frequencies and small for the lower frequencies. Generally, the slow rise of this kind of filter does not particularly recommend it as a procedure for retaining only the high frequencies.

For the centered 12-month moving average, we can take  $a_{-6} = a_6 = 1/24$ ,  $a_k = 1/12$  for  $-5 \leq k \leq 5$  and  $a_k = 0$  elsewhere. Substituting and recognizing the cosine terms gives





**Fig. 4.17.** Squared frequency response functions of the first difference (top) and twelve-month moving average (bottom) filters.

$$A_{yx}(\omega) = \frac{1}{12} \left[ 1 + \cos(12\pi\omega) + 2 \sum_{k=1}^5 \cos(2\pi\omega k) \right]. \quad (4.107)$$

Plotting the squared frequency response of this function as in the bottom of Figure 4.17 shows that we can expect this filter to cut most of the frequency content above .05 cycles per point, and nearly all of the frequency content above  $1/12 \approx .083$ . In particular, this drives down the yearly components with periods of 12 months and enhances the El Niño frequency, which is somewhat lower. The filter is not completely efficient at attenuating high frequencies; some power contributions are left at higher frequencies, as shown in the function  $|A_{yx}(\omega)|^2$ .

The following R session shows how to filter the data, perform the spectral analysis of a filtered series, and plot the squared frequency response curves of the difference and moving average filters.

```
par(mfrow=c(3,1), mar=c(3,3,1,1), mgp=c(1.6,.6,0))
plot(soi) # plot data
plot(diff(soi)) # plot first difference
k = kernel("modified.daniell", 6) # filter weights
plot(soif <- kernapply(soi, k)) # plot 12 month filter
dev.new()
spectrum(soif, spans=9, log="no") # spectral analysis (not shown)
abline(v=12/52, lty="dashed")
dev.new()
##-- frequency responses --##
par(mfrow=c(2,1), mar=c(3,3,1,1), mgp=c(1.6,.6,0))
w = seq(0, .5, by=.01)
FRdiff = abs(1-exp(2i*pi*w))^2
plot(w, FRdiff, type='l', xlab='frequency')
```

```

u = cos(2*pi*w)+cos(4*pi*w)+cos(6*pi*w)+cos(8*pi*w)+cos(10*pi*w)
FRma = ((1 + cos(12*pi*w) + 2*u)/12)^2
plot(w, FRma, type='l', xlab='frequency')

```

The two filters discussed in the previous example were different in that the frequency response function of the first difference was complex-valued, whereas the frequency response of the moving average was purely real. A short derivation similar to that used to verify (4.22) shows, when  $x_t$  and  $y_t$  are related by the linear filter relation (4.20), the cross-spectrum satisfies

$$f_{yx}(\omega) = A_{yx}(\omega)f_{xx}(\omega),$$

so the frequency response is of the form

$$A_{yx}(\omega) = \frac{f_{yx}(\omega)}{f_{xx}(\omega)} \quad (4.108)$$

$$= \frac{c_{yx}(\omega)}{f_{xx}(\omega)} - i \frac{q_{yx}(\omega)}{f_{xx}(\omega)}, \quad (4.109)$$

where we have used (4.88) to get the last form. Then, we may write (4.109) in polar coordinates as

$$A_{yx}(\omega) = |A_{yx}(\omega)| \exp\{-i \phi_{yx}(\omega)\}, \quad (4.110)$$

where the *amplitude* and *phase* of the filter are defined by

$$|A_{yx}(\omega)| = \frac{\sqrt{c_{yx}^2(\omega) + q_{yx}^2(\omega)}}{f_{xx}(\omega)} \quad (4.111)$$

and

$$\phi_{yx}(\omega) = \tan^{-1} \left( -\frac{q_{yx}(\omega)}{c_{yx}(\omega)} \right). \quad (4.112)$$

A simple interpretation of the phase of a linear filter is that it exhibits time delays as a function of frequency in the same way as the spectrum represents the variance as a function of frequency. Additional insight can be gained by considering the simple delaying filter

$$y_t = Ax_{t-D},$$

where the series gets replaced by a version, amplified by multiplying by  $A$  and delayed by  $D$  points. For this case,

$$f_{yx}(\omega) = Ae^{-2\pi i \omega D} f_{xx}(\omega),$$

and the amplitude is  $|A|$ , and the phase is

$$\phi_{yx}(\omega) = -2\pi \omega D,$$

or just a linear function of frequency  $\omega$ . For this case, applying a simple time delay causes phase delays that depend on the frequency of the periodic component being delayed. Interpretation is further enhanced by setting

$$x_t = \cos(2\pi\omega t),$$

in which case

$$y_t = A \cos(2\pi\omega t - 2\pi\omega D).$$

Thus, the output series,  $y_t$ , has the same period as the input series,  $x_t$ , but the amplitude of the output has increased by a factor of  $|A|$  and the phase has been changed by a factor of  $-2\pi\omega D$ .

### Example 4.23 Difference and Moving Average Filters

We consider calculating the amplitude and phase of the two filters discussed in [Example 4.22](#). The case for the moving average is easy because  $A_{yx}(\omega)$  given in [\(4.107\)](#) is purely real. So, the amplitude is just  $|A_{yx}(\omega)|$  and the phase is  $\phi_{yx}(\omega) = 0$ . In general, symmetric ( $a_j = a_{-j}$ ) filters have zero phase. The first difference, however, changes this, as we might expect from the example above involving the time delay filter. In this case, the squared amplitude is given in [\(4.106\)](#). To compute the phase, we write

$$\begin{aligned} A_{yx}(\omega) &= 1 - e^{-2\pi i\omega} = e^{-i\pi\omega}(e^{i\pi\omega} - e^{-i\pi\omega}) \\ &= 2ie^{-i\pi\omega} \sin(\pi\omega) = 2 \sin^2(\pi\omega) + 2i \cos(\pi\omega) \sin(\pi\omega) \\ &= \frac{c_{yx}(\omega)}{f_{xx}(\omega)} - i \frac{q_{yx}(\omega)}{f_{xx}(\omega)}, \end{aligned}$$

so

$$\phi_{yx}(\omega) = \tan^{-1} \left( -\frac{q_{yx}(\omega)}{c_{yx}(\omega)} \right) = \tan^{-1} \left( \frac{\cos(\pi\omega)}{\sin(\pi\omega)} \right).$$

Noting that

$$\cos(\pi\omega) = \sin(-\pi\omega + \pi/2)$$

and that

$$\sin(\pi\omega) = \cos(-\pi\omega + \pi/2),$$

we get

$$\phi_{yx}(\omega) = -\pi\omega + \pi/2,$$

and the phase is again a linear function of frequency.

The above tendency of the frequencies to arrive at different times in the filtered version of the series remains as one of two annoying features of the difference type filters. The other weakness is the gentle increase in the frequency response function. If low frequencies are really unimportant and high frequencies are to be preserved, we would like to have a somewhat sharper response than is obvious in [Figure 4.17](#). Similarly, if low frequencies are important and high frequencies are not, the moving average filters are also not very efficient at passing the low frequencies and attenuating the high frequencies. Improvement is possible by designing better and longer filters, but we do not discuss this here.

We will occasionally use results for multivariate series  $x_t = (x_{t1}, \dots, x_{tp})'$  that are comparable to the simple property shown in [\(4.22\)](#). Consider the *matrix filter*

$$y_t = \sum_{j=-\infty}^{\infty} A_j x_{t-j}, \quad (4.113)$$

where  $\{A_j\}$  denotes a sequence of  $q \times p$  matrices such that  $\sum_{j=-\infty}^{\infty} \|A_j\| < \infty$  and  $\|\cdot\|$  denotes any matrix norm,  $x_t = (x_{t1}, \dots, x_{tp})'$  is a  $p \times 1$  stationary vector process with mean vector  $\mu_x$  and  $p \times p$ , matrix covariance function  $\Gamma_{xx}(h)$  and spectral matrix  $f_{xx}(\omega)$ , and  $y_t$  is the  $q \times 1$  vector output process. Then, we can obtain the following property.

**Property 4.9 Output Spectral Matrix of Filtered Vector Series**

The spectral matrix of the filtered output  $y_t$  in (4.113) is related to the spectrum of the input  $x_t$  by

$$f_{yy}(\omega) = \mathcal{A}(\omega) f_{xx}(\omega) \mathcal{A}^*(\omega), \quad (4.114)$$

where the matrix frequency response function  $\mathcal{A}(\omega)$  is defined by

$$\mathcal{A}(\omega) = \sum_{j=-\infty}^{\infty} A_j \exp(-2\pi i \omega j). \quad (4.115)$$

## 4.8 Lagged Regression Models

One of the intriguing possibilities offered by the coherence analysis of the relation between the SOI and Recruitment series discussed in Example 4.21 would be extending classical regression to the analysis of lagged regression models of the form

$$y_t = \sum_{r=-\infty}^{\infty} \beta_r x_{t-r} + v_t, \quad (4.116)$$

where  $v_t$  is a stationary noise process,  $x_t$  is the observed input series, and  $y_t$  is the observed output series. We are interested in estimating the filter coefficients  $\beta_r$  relating the adjacent lagged values of  $x_t$  to the output series  $y_t$ .

In the case of SOI and Recruitment series, we might identify the El Niño driving series, SOI, as the input,  $x_t$ , and  $y_t$ , the Recruitment series, as the output. In general, there will be more than a single possible input series and we may envision a  $q \times 1$  vector of driving series. This multivariate input situation is covered in Chapter 7. The model given by (4.116) is useful under several different scenarios, corresponding to different assumptions that can be made about the components.

We assume that the inputs and outputs have zero means and are jointly stationary with the  $2 \times 1$  vector process  $(x_t, y_t)'$  having a spectral matrix of the form

$$f(\omega) = \begin{pmatrix} f_{xx}(\omega) & f_{xy}(\omega) \\ f_{yx}(\omega) & f_{yy}(\omega) \end{pmatrix}. \quad (4.117)$$

Here,  $f_{xy}(\omega)$  is the cross-spectrum relating the input  $x_t$  to the output  $y_t$ , and  $f_{xx}(\omega)$  and  $f_{yy}(\omega)$  are the spectra of the input and output series, respectively. Generally, we

observe two series, regarded as input and output and search for regression functions  $\{\beta_t\}$  relating the inputs to the outputs. We assume all autocovariance functions satisfy the absolute summability conditions of the form (4.38).

Then, minimizing the mean squared error

$$MSE = E \left( y_t - \sum_{r=-\infty}^{\infty} \beta_r x_{t-r} \right)^2 \quad (4.118)$$

leads to the usual orthogonality conditions

$$E \left[ \left( y_t - \sum_{r=-\infty}^{\infty} \beta_r x_{t-r} \right) x_{t-s} \right] = 0 \quad (4.119)$$

for all  $s = 0, \pm 1, \pm 2, \dots$ . Taking the expectations inside leads to the normal equations

$$\sum_{r=-\infty}^{\infty} \beta_r \gamma_{xx}(s-r) = \gamma_{yx}(s) \quad (4.120)$$

for  $s = 0, \pm 1, \pm 2, \dots$ . These equations might be solved, with some effort, if the covariance functions were known exactly. If data  $(x_t, y_t)$  for  $t = 1, \dots, n$  are available, we might use a finite approximation to the above equations with  $\hat{\gamma}_{xx}(h)$  and  $\hat{\gamma}_{yx}(h)$  substituted into (4.120). If the regression vectors are essentially zero for  $|s| \geq M/2$ , and  $M < n$ , the system (4.120) would be of full rank and the solution would involve inverting an  $(M-1) \times (M-1)$  matrix.

A frequency domain approximate solution is easier in this case for two reasons. First, the computations depend on spectra and cross-spectra that can be estimated from sample data using the techniques of Section 4.5. In addition, no matrices will have to be inverted, although the frequency domain ratio will have to be computed for each frequency. In order to develop the frequency domain solution, substitute the representation (4.96) into the normal equations, using the convention defined in (4.117). The left side of (4.120) can then be written in the form

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_{r=-\infty}^{\infty} \beta_r e^{2\pi i \omega (s-r)} f_{xx}(\omega) d\omega = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i \omega s} B(\omega) f_{xx}(\omega) d\omega,$$

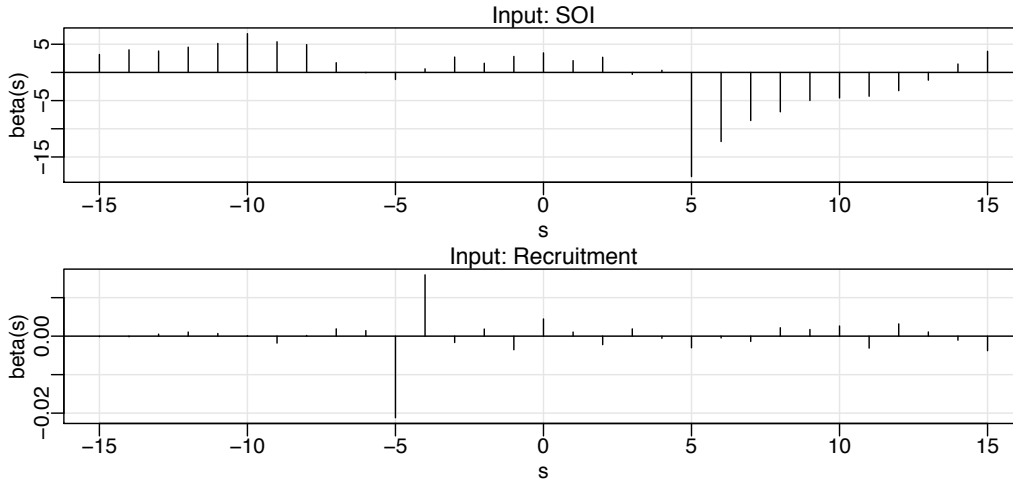
where

$$B(\omega) = \sum_{r=-\infty}^{\infty} \beta_r e^{-2\pi i \omega r} \quad (4.121)$$

is the Fourier transform of the regression coefficients  $\beta_t$ . Now, because  $\gamma_{yx}(s)$  is the inverse transform of the cross-spectrum  $f_{yx}(\omega)$ , we might write the system of equations in the frequency domain, using the uniqueness of the Fourier transform, as

$$B(\omega) f_{xx}(\omega) = f_{yx}(\omega), \quad (4.122)$$

which then become the analogs of the usual normal equations. Then, we may take



**Fig. 4.18.** Estimated impulse response functions relating SOI to Recruitment (top) and Recruitment to SOI (bottom)  $L = 15$ ,  $M = 32$ .

$$\hat{B}(\omega_k) = \frac{\hat{f}_{yx}(\omega_k)}{\hat{f}_{xx}(\omega_k)} \quad (4.123)$$

as the estimator for the Fourier transform of the regression coefficients, evaluated at some subset of fundamental frequencies  $\omega_k = k/M$  with  $M \ll n$ . Generally, we assume smoothness of  $B(\cdot)$  over intervals of the form  $\{\omega_k + \ell/n; \ell = -m, \dots, 0, \dots, m\}$ , with  $L = 2m + 1$ . The inverse transform of the function  $\hat{B}(\omega)$  would give  $\hat{\beta}_t$ , and we note that the discrete time approximation can be taken as

$$\hat{\beta}_t = M^{-1} \sum_{k=0}^{M-1} \hat{B}(\omega_k) e^{2\pi i \omega_k t} \quad (4.124)$$

for  $t = 0, \pm 1, \pm 2, \dots, \pm(M/2 - 1)$ . If we were to use (4.124) to define  $\hat{\beta}_t$  for  $|t| \geq M/2$ , we would end up with a sequence of coefficients that is periodic with a period of  $M$ . In practice we define  $\hat{\beta}_t = 0$  for  $|t| \geq M/2$  instead. **Problem 4.32** explores the error resulting from this approximation.

#### Example 4.24 Lagged Regression for SOI and Recruitment

The high coherence between the SOI and Recruitment series noted in **Example 4.21** suggests a lagged regression relation between the two series. A natural direction for the implication in this situation is implied because we feel that the sea surface temperature or SOI should be the input and the Recruitment series should be the output. With this in mind, let  $x_t$  be the SOI series and  $y_t$  the Recruitment series.

Although we think naturally of the SOI as the input and the Recruitment as the output, two input-output configurations are of interest. With SOI as the input, the model is

$$y_t = \sum_{r=-\infty}^{\infty} a_r x_{t-r} + w_t$$

whereas a model that reverses the two roles would be

$$x_t = \sum_{r=-\infty}^{\infty} b_r y_{t-r} + v_t,$$

where  $w_t$  and  $v_t$  are white noise processes. Even though there is no plausible environmental explanation for the second of these two models, displaying both possibilities helps to settle on a parsimonious transfer function model.

Based on the script `LagReg` in `astsa`, the estimated regression or impulse response function for SOI, with  $M = 32$  and  $L = 15$  is

```
LagReg(soi, rec, L=15, M=32, threshold=6)
      lag s   beta(s)
[1,]     5 -18.479306
[2,]     6 -12.263296
[3,]     7  -8.539368
[4,]     8  -6.984553
The prediction equation is
rec(t) = alpha + sum_s[ beta(s)*soi(t-s) ], where alpha = 65.97
MSE = 414.08
```

Note the negative peak at a lag of five points in the top of [Figure 4.18](#); in this case, SOI is the input series. The fall-off after lag five seems to be approximately exponential and a possible model is

$$y_t = 66 - 18.5x_{t-5} - 12.3x_{t-6} - 8.5x_{t-7} - 7x_{t-8} + w_t.$$

If we examine the inverse relation, namely, a regression model with the Recruitment series  $y_t$  as the input, the bottom of [Figure 4.18](#) implies a much simpler model,

```
LagReg(rec, soi, L=15, M=32, inverse=TRUE, threshold=.01)
      lag s   beta(s)
[1,]     4  0.01593167
[2,]     5 -0.02120013
The prediction equation is
soi(t) = alpha + sum_s[ beta(s)*rec(t+s) ], where alpha = 0.41
MSE = 0.07
```

depending on only two coefficients, namely,

$$x_t = .41 + .016y_{t+4} - .02y_{t+5} + v_t.$$

Multiplying both sides by  $50B^5$  and rearranging, we have

$$(1 - .8B)y_t = 20.5 - 50B^5x_t + \epsilon_t.$$

Finally, we check whether the noise,  $\epsilon_t$ , is white. In addition, at this point, it simplifies matters if we rerun the regression with autocorrelated errors and reestimate the coefficients. The model is referred to as an ARMAX model (the X stands for exogenous; see [Section 5.6](#) and [Section 6.6.1](#)):

```
fish = ts.intersect(R=rec, RL1=lag(rec,-1), SL5=lag(soi,-5))
(u = lm(fish[,1]~fish[,2:3], na.action=NULL))
acf2(resid(u)) # suggests ar1
sarima(fish[,1], 1, 0, 0, xreg=fish[,2:3]) # armax model
```

```

Coefficients:
      ar1  intercept      RL1      SL5
      0.4487  12.3323  0.8005 -21.0307
s.e.  0.0503   1.5746  0.0234   1.0915
sigma^2 estimated as 49.93

```

Our final parsimonious fitted model is (with rounding)

$$y_t = 12 + .8y_{t-1} - 21x_{t-5} + \epsilon_t, \quad \text{and} \quad \epsilon_t = .45\epsilon_{t-1} + w_t,$$

where  $w_t$  is white noise with  $\sigma_w^2 = 50$ . This example is also examined in [Chapter 5](#) and the fitted values for the final model can be viewed [Figure 5.12](#).

The example shows we can get a clean estimator for the transfer functions relating the two series if the coherence  $\hat{\rho}_{xy}^2(\omega)$  is large. The reason is that we can write the minimized mean squared error (4.118) as

$$MSE = E \left[ \left( y_t - \sum_{r=-\infty}^{\infty} \beta_r x_{t-r} \right) y_t \right] = \gamma_{yy}(0) - \sum_{r=-\infty}^{\infty} \beta_r \gamma_{xy}(-r),$$

using the result about the orthogonality of the data and error term in the Projection theorem. Then, substituting the spectral representations of the autocovariance and cross-covariance functions and identifying the Fourier transform (4.121) in the result leads to

$$\begin{aligned} MSE &= \int_{-\frac{1}{2}}^{\frac{1}{2}} [f_{yy}(\omega) - B(\omega)f_{xy}(\omega)] d\omega \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} f_{yy}(\omega)[1 - \rho_{yx}^2(\omega)] d\omega, \end{aligned} \quad (4.125)$$

where  $\rho_{yx}^2(\omega)$  is just the squared coherence given by (4.94). The similarity of (4.125) to the usual mean square error that results from predicting  $y$  from  $x$  is obvious. In that case, we would have

$$E(y - \beta x)^2 = \sigma_y^2(1 - \rho_{xy}^2)$$

for jointly distributed random variables  $x$  and  $y$  with zero means, variances  $\sigma_x^2$  and  $\sigma_y^2$ , and covariance  $\sigma_{xy} = \rho_{xy}\sigma_x\sigma_y$ . Because the mean squared error in (4.125) satisfies  $MSE \geq 0$  with  $f_{yy}(\omega)$  a non-negative function, it follows that the coherence satisfies

$$0 \leq \rho_{xy}^2(\omega) \leq 1$$

for all  $\omega$ . Furthermore, [Problem 4.33](#) shows the squared coherence is one when the output are linearly related by the filter relation (4.116), and there is no noise, i.e.,  $v_t = 0$ . Hence, the multiple coherence gives a measure of the association or correlation between the input and output series as a function of frequency.

The matter of verifying that the  $F$ -distribution claimed for (4.104) will hold when the sample coherence values are substituted for theoretical values still remains. Again,



the form of the  $F$ -statistic is exactly analogous to the usual  $t$ -test for no correlation in a regression context. We give an argument leading to this conclusion later using the results in [Section C.3](#). Another question that has not been resolved in this section is the extension to the case of multiple inputs  $x_{t1}, x_{t2}, \dots, x_{tq}$ . Often, more than just a single input series is present that can possibly form a lagged predictor of the output series  $y_t$ . An example is the cardiovascular mortality series that depended on possibly a number of pollution series and temperature. We discuss this particular extension as a part of the multivariate time series techniques considered in [Chapter 7](#).

## 4.9 Signal Extraction and Optimum Filtering

A model closely related to regression can be developed by assuming again that

$$y_t = \sum_{r=-\infty}^{\infty} \beta_r x_{t-r} + v_t, \quad (4.126)$$

but where the  $\beta$ s are known and  $x_t$  is some unknown random *signal* that is uncorrelated with the *noise* process  $v_t$ . In this case, we observe only  $y_t$  and are interested in an estimator for the signal  $x_t$  of the form

$$\hat{x}_t = \sum_{r=-\infty}^{\infty} a_r y_{t-r}. \quad (4.127)$$

In the frequency domain, it is convenient to make the additional assumptions that the series  $x_t$  and  $v_t$  are both mean-zero stationary series with spectra  $f_{xx}(\omega)$  and  $f_{vv}(\omega)$ , often referred to as the *signal spectrum* and *noise spectrum*, respectively. Often, the special case  $\beta_t = \delta_t$ , in which  $\delta_t$  is the Kronecker delta, is of interest because [\(4.126\)](#) reduces to the simple *signal plus noise* model

$$y_t = x_t + v_t \quad (4.128)$$

in that case. In general, we seek the set of filter coefficients  $a_t$  that minimize the mean squared error of estimation, say,

$$MSE = E \left[ \left( x_t - \sum_{r=-\infty}^{\infty} a_r y_{t-r} \right)^2 \right]. \quad (4.129)$$

This problem was originally solved by Kolmogorov (1941) and by Wiener (1949), who derived the result in 1941 and published it in classified reports during World War II.

We can apply the orthogonality principle to write

$$E \left[ \left( x_t - \sum_{r=-\infty}^{\infty} a_r y_{t-r} \right) y_{t-s} \right] = 0$$

for  $s = 0, \pm 1, \pm 2, \dots$ , which leads to

$$\sum_{r=-\infty}^{\infty} a_r \gamma_{yy}(s-r) = \gamma_{xy}(s),$$

to be solved for the filter coefficients. Substituting the spectral representations for the autocovariance functions into the above and identifying the spectral densities through the uniqueness of the Fourier transform produces

$$A(\omega) f_{yy}(\omega) = f_{xy}(\omega), \quad (4.130)$$

where  $A(\omega)$  and the optimal filter  $a_t$  are Fourier transform pairs for  $B(\omega)$  and  $\beta_t$ . Now, a special consequence of the model is that (see [Problem 4.30](#))

$$f_{xy}(\omega) = B^*(\omega) f_{xx}(\omega) \quad (4.131)$$

and

$$f_{yy}(\omega) = |B(\omega)|^2 f_{xx}(\omega) + f_{vv}(\omega), \quad (4.132)$$

implying the optimal filter would be Fourier transform of

$$A(\omega) = \frac{B^*(\omega)}{\left(|B(\omega)|^2 + \frac{f_{vv}(\omega)}{f_{xx}(\omega)}\right)}, \quad (4.133)$$

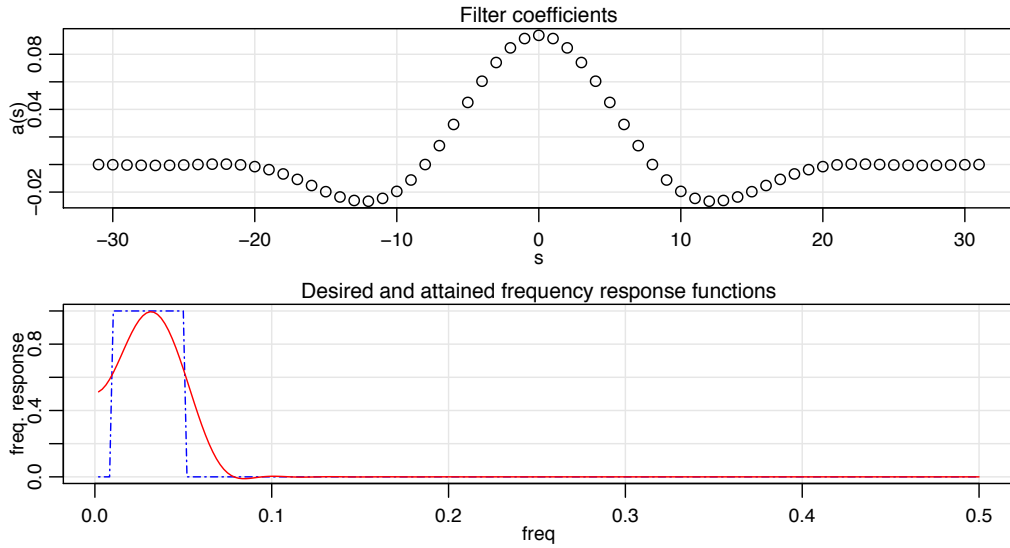
where the second term in the denominator is just the inverse of the *signal to noise ratio*, say,

$$\text{SNR}(\omega) = \frac{f_{xx}(\omega)}{f_{vv}(\omega)}. \quad (4.134)$$

The result shows the optimum filters can be computed for this model if the signal and noise spectra are both known or if we can assume knowledge of the signal-to-noise ratio  $\text{SNR}(\omega)$  as function of frequency. In [Chapter 7](#), we show some methods for estimating these two parameters in conjunction with random effects analysis of variance models, but we assume here that it is possible to specify the signal-to-noise ratio a priori. If the signal-to-noise ratio is known, the optimal filter can be computed by the inverse transform of the function  $A(\omega)$ . It is more likely that the inverse transform will be intractable and a finite filter approximation like that used in the previous section can be applied to the data. In this case, we will have

$$a_t^M = M^{-1} \sum_{k=0}^{M-1} A(\omega_k) e^{2\pi i \omega_k t} \quad (4.135)$$

as the estimated filter function. It will often be the case that the form of the specified frequency response will have some rather sharp transitions between regions where the signal-to-noise ratio is high and regions where there is little signal. In these cases, the shape of the frequency response function will have ripples that can introduce



**Fig. 4.19.** Filter coefficients (top) and frequency response functions (bottom) for designed SOI filters.

frequencies at different amplitudes. An aesthetic solution to this problem is to introduce tapering as was done with spectral estimation in (4.69)–(4.76). We use below the tapered filter  $\tilde{a}_t = h_t a_t$  where  $h_t$  is the cosine taper given in (4.76). The squared frequency response of the resulting filter will be  $|\tilde{A}(\omega)|^2$ , where

$$\tilde{A}(\omega) = \sum_{t=-\infty}^{\infty} a_t h_t e^{-2\pi i \omega t}. \tag{4.136}$$

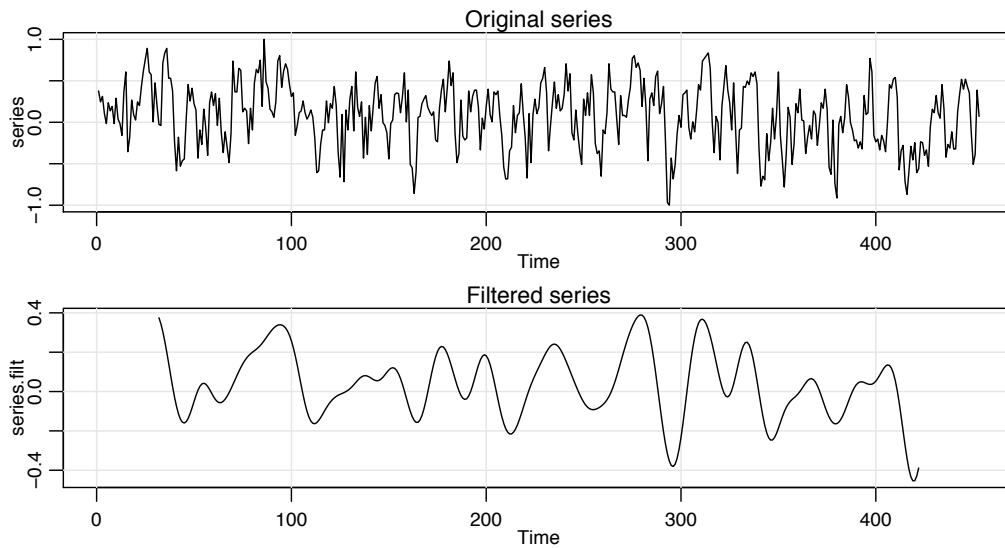
The results are illustrated in the following example that extracts the El Niño component of the sea surface temperature series.

**Example 4.25 Estimating the El Niño Signal via Optimal Filters**

Figure 4.7 shows the spectrum of the SOI series, and we note that essentially two components have power, the El Niño frequency of about .02 cycles per month (the four-year cycle) and a yearly frequency of about .08 cycles per month (the annual cycle). We assume, for this example, that we wish to preserve the lower frequency as signal and to eliminate the higher order frequencies, and in particular, the annual cycle. In this case, we assume the simple signal plus noise model

$$y_t = x_t + v_t,$$

so that there is no convolving function  $\beta_t$ . Furthermore, the signal-to-noise ratio is assumed to be high to about .06 cycles per month and zero thereafter. The optimal frequency response was assumed to be unity to .05 cycles per point and then to decay linearly to zero in several steps. Figure 4.19 shows the coefficients as specified by (4.135) with  $M = 64$ , as well as the frequency response function given by (4.136), of the cosine tapered coefficients; recall Figure 4.11, where we demonstrated the



**Fig. 4.20.** Original SOI series (top) compared to filtered version showing the estimated El Niño temperature signal (bottom).

need for tapering to avoid severe ripples in the window. The constructed response function is compared to the ideal window in [Figure 4.19](#).

[Figure 4.20](#) shows the original and filtered SOI index, and we see a smooth extracted signal that conveys the essence of the underlying El Niño signal. The frequency response of the designed filter can be compared with that of the symmetric 12-month moving average applied to the same series in [Example 4.22](#). The filtered series, shown in [Figure 4.16](#), shows a good deal of higher frequency chatter riding on the smoothed version, which has been introduced by the higher frequencies that leak through in the squared frequency response, as in [Figure 4.17](#).

The analysis can be replicated using the script [SigExtract](#).

```
SigExtract(soi, L=9, M=64, max.freq=.05)
```

The design of finite filters with a specified frequency response requires some experimentation with various target frequency response functions and we have only touched on the methodology here. The filter designed here, sometimes called a low-pass filter reduces the high frequencies and keeps or passes the low frequencies. Alternately, we could design a high-pass filter to keep high frequencies if that is where the signal is located. An example of a simple high-pass filter is the first difference with a frequency response that is shown in [Figure 4.17](#). We can also design band-pass filters that keep frequencies in specified bands. For example, seasonal adjustment filters are often used in economics to reject seasonal frequencies while keeping both high frequencies, lower frequencies, and trend (see, for example, Grether and Nerlove, 1970).

The filters we have discussed here are all symmetric two-sided filters, because the designed frequency response functions were purely real. Alternatively, we may design recursive filters to produce a desired response. An example of a recursive filter is one that replaces the input  $x_t$  by the filtered output

$$y_t = \sum_{k=1}^p \phi_k y_{t-k} + x_t - \sum_{k=1}^q \theta_k x_{t-k}. \quad (4.137)$$

Note the similarity between (4.137) and the ARMA( $p, q$ ) model, in which the white noise component is replaced by the input. Transposing the terms involving  $y_t$  and using the basic linear filter result in [Property 4.3](#) leads to

$$f_y(\omega) = \frac{|\theta(e^{-2\pi i\omega})|^2}{|\phi(e^{-2\pi i\omega})|^2} f_x(\omega), \quad (4.138)$$

where

$$\phi(e^{-2\pi i\omega}) = 1 - \sum_{k=1}^p \phi_k e^{-2\pi i k \omega}$$

and

$$\theta(e^{-2\pi i\omega}) = 1 - \sum_{k=1}^q \theta_k e^{-2\pi i k \omega}.$$

Recursive filters such as those given by (4.138) distort the phases of arriving frequencies, and we do not consider the problem of designing such filters in any detail.

## 4.10 Spectral Analysis of Multidimensional Series

Multidimensional series of the form  $x_s$ , where  $s = (s_1, s_2, \dots, s_r)'$  is an  $r$ -dimensional vector of spatial coordinates or a combination of space and time coordinates, were introduced in [Section 1.6](#). The example given there, shown in [Figure 1.18](#), was a collection of temperature measurements taking on a rectangular field. These data would form a two-dimensional process, indexed by row and column in space. In that section, the multidimensional autocovariance function of an  $r$ -dimensional stationary series was given as  $\gamma_x(h) = E[x_{s+h}x_s]$ , where the multidimensional lag vector is  $h = (h_1, h_2, \dots, h_r)'$ .

The multidimensional *wavenumber spectrum* is given as the Fourier transform of the autocovariance, namely,

$$f_x(\omega) = \sum_h \cdots \sum_h \gamma_x(h) e^{-2\pi i \omega' h}. \quad (4.139)$$

Again, the inverse result

$$\gamma_x(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \cdots \int_{-\frac{1}{2}}^{\frac{1}{2}} f_x(\omega) e^{2\pi i \omega' h} d\omega \quad (4.140)$$

holds, where the integral is over the multidimensional range of the vector  $\omega$ . The wavenumber argument is exactly analogous to the frequency argument, and we have the corresponding intuitive interpretation as the cycling rate  $\omega_i$  per distance traveled  $s_i$  in the  $i$ -th direction.

Two-dimensional processes occur often in practical applications, and the representations above reduce to

$$f_x(\omega_1, \omega_2) = \sum_{h_1=-\infty}^{\infty} \sum_{h_2=-\infty}^{\infty} \gamma_x(h_1, h_2) e^{-2\pi i(\omega_1 h_1 + \omega_2 h_2)} \quad (4.141)$$

and

$$\gamma_x(h_1, h_2) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{-\frac{1}{2}}^{\frac{1}{2}} f_x(\omega_1, \omega_2) e^{2\pi i(\omega_1 h_1 + \omega_2 h_2)} d\omega_1 d\omega_2 \quad (4.142)$$

in the case  $r = 2$ . The notion of linear filtering generalizes easily to the two-dimensional case by defining the impulse response function  $a_{s_1, s_2}$  and the spatial filter output as

$$y_{s_1, s_2} = \sum_{u_1} \sum_{u_2} a_{u_1, u_2} x_{s_1 - u_1, s_2 - u_2}. \quad (4.143)$$

The spectrum of the output of this filter can be derived as

$$f_y(\omega_1, \omega_2) = |A(\omega_1, \omega_2)|^2 f_x(\omega_1, \omega_2), \quad (4.144)$$

where

$$A(\omega_1, \omega_2) = \sum_{u_1} \sum_{u_2} a_{u_1, u_2} e^{-2\pi i(\omega_1 u_1 + \omega_2 u_2)}. \quad (4.145)$$

These results are analogous to those in the one-dimensional case, described by [Property 4.3](#).

The multidimensional DFT is also a straightforward generalization of the univariate expression. In the two-dimensional case with data on a rectangular grid,  $\{x_{s_1, s_2}; s_1 = 1, \dots, n_1, s_2 = 1, \dots, n_2\}$ , we will write, for  $-1/2 \leq \omega_1, \omega_2 \leq 1/2$ ,

$$d(\omega_1, \omega_2) = (n_1 n_2)^{-1/2} \sum_{s_1=1}^{n_1} \sum_{s_2=1}^{n_2} x_{s_1, s_2} e^{-2\pi i(\omega_1 s_1 + \omega_2 s_2)} \quad (4.146)$$

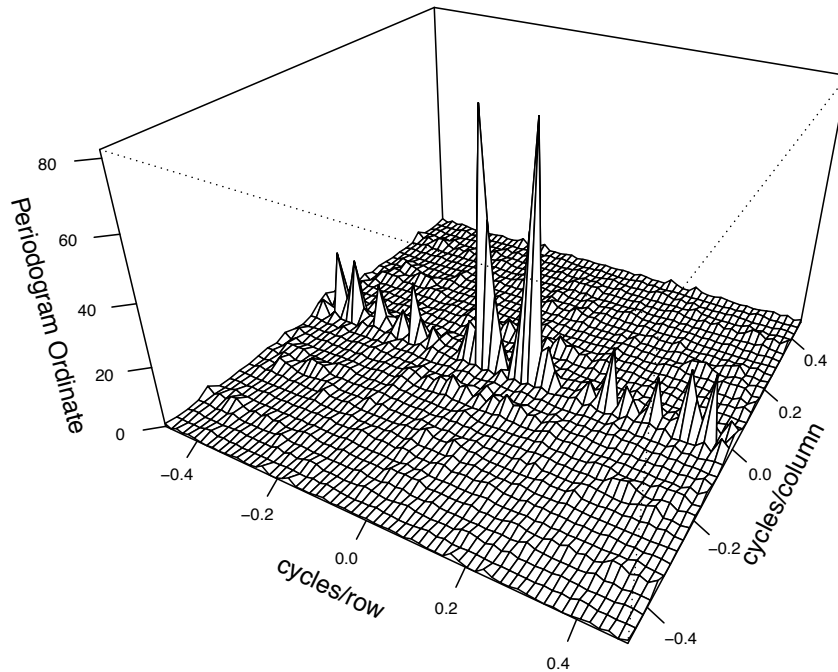
as the two-dimensional DFT, where the frequencies  $\omega_1, \omega_2$  are evaluated at multiples of  $(1/n_1, 1/n_2)$  on the spatial frequency scale. The two-dimensional wavenumber spectrum can be estimated by the smoothed *sample wavenumber spectrum*

$$\bar{f}_x(\omega_1, \omega_2) = (L_1 L_2)^{-1} \sum_{\ell_1, \ell_2} |d(\omega_1 + \ell_1/n_1, \omega_2 + \ell_2/n_2)|^2, \quad (4.147)$$

where the sum is taken over the grid  $\{-m_j \leq \ell_j \leq m_j; j = 1, 2\}$ , where  $L_1 = 2m_1 + 1$  and  $L_2 = 2m_2 + 1$ . The statistic

$$\frac{2L_1 L_2 \bar{f}_x(\omega_1, \omega_2)}{f_x(\omega_1, \omega_2)} \sim \chi_{2L_1 L_2}^2 \quad (4.148)$$

can be used to set confidence intervals or make approximate tests against a fixed assumed spectrum  $f_0(\omega_1, \omega_2)$ .



**Fig. 4.21.** Two-dimensional periodogram of soil temperature profile showing peak at .0625 cycles/row. The period is 16 rows, and this corresponds to  $16 \times 17 \text{ ft} = 272 \text{ ft}$ .

#### Example 4.26 Soil Surface Temperatures

As an example, consider the periodogram of the two-dimensional temperature series shown in [Figure 1.18](#) and analyzed by Bazza et al. (1988). We recall the spatial coordinates in this case will be  $(s_1, s_2)$ , which define the spatial coordinates rows and columns so that the frequencies in the two directions will be expressed as cycles per row and cycles per column. [Figure 4.21](#) shows the periodogram of the two-dimensional temperature series, and we note the ridge of strong spectral peaks running over rows at a column frequency of zero. An obvious periodic component appears at frequencies of .0625 and  $-.0625$  cycles per row, which corresponds to 16 rows or about 272 ft. On further investigation of previous irrigation patterns over this field, treatment levels of salt varied periodically over columns. This analysis is extended in [Problem 4.24](#), where we recover the salt treatment profile over rows and compare it to a signal, computed by averaging over columns.

[Figure 4.21](#) may be reproduced in R as follows. In the code for this example, the periodogram is computed in one step as `per`; the rest of the code is simply manipulation to obtain a nice graphic.

```
per = Mod(fft(soiltemp-mean(soiltemp))/sqrt(64*36))^2
per2 = cbind(per[1:32,18:2], per[1:32,1:18])
per3 = rbind(per2[32:2,],per2)
par(mar=c(1,2.5,0,0)+.1)
persp(-31:31/64, -17:17/36, per3, phi=30, theta=30, expand=.6,
      ticktype="detailed", xlab="cycles/row", ylab="cycles/column",
      zlab="Periodogram Ordinate")
```

Another application of two-dimensional spectral analysis of agricultural field trials is given in McBratney and Webster (1981), who used it to detect ridge and furrow patterns in yields. The requirement for regular, equally spaced samples on fairly large grids has tended to limit enthusiasm for strict two-dimensional spectral analysis. An exception is when a propagating signal from a given velocity and azimuth is present so predicting the wavenumber spectrum as a function of velocity and azimuth becomes feasible (see Shumway et al., 1999).

## Problems

### Section 4.1

**4.1** Verify that for any positive integer  $n$  and  $j, k = 0, 1, \dots, \lfloor n/2 \rfloor$ , where  $\lfloor \cdot \rfloor$  denotes the greatest integer function:

(a) Except for  $j = 0$  or  $j = n/2$ ,<sup>4.12</sup>

$$\sum_{t=1}^n \cos^2(2\pi t j/n) = \sum_{t=1}^n \sin^2(2\pi t j/n) = n/2.$$

(b) When  $j = 0$  or  $j = n/2$ ,

$$\sum_{t=1}^n \cos^2(2\pi t j/n) = n \text{ but } \sum_{t=1}^n \sin^2(2\pi t j/n) = 0.$$

(c) For  $j \neq k$ ,

$$\sum_{t=1}^n \cos(2\pi t j/n) \cos(2\pi t k/n) = \sum_{t=1}^n \sin(2\pi t j/n) \sin(2\pi t k/n) = 0.$$

Also, for any  $j$  and  $k$ ,

$$\sum_{t=1}^n \cos(2\pi t j/n) \sin(2\pi t k/n) = 0.$$

<sup>4.12</sup> *Hint: We'll do part of the problem.*

$$\begin{aligned} \sum_{t=1}^n \cos^2(2\pi t j/n) &= \frac{1}{4} \sum_{t=1}^n (e^{2\pi i t j/n} + e^{-2\pi i t j/n})(e^{2\pi i t j/n} + e^{-2\pi i t j/n}) \\ &= \frac{1}{4} \sum_{t=1}^n (e^{4\pi i t j/n} + 1 + 1 + e^{-4\pi i t j/n}) = \frac{n}{2}. \end{aligned}$$



**4.2** Repeat the simulations and analyses in [Example 4.1](#) and [Example 4.2](#) with the following changes:

- (a) Change the sample size to  $n = 128$  and generate and plot the same series as in [Example 4.1](#):

$$\begin{aligned}x_{t1} &= 2 \cos(2\pi \cdot 06 t) + 3 \sin(2\pi \cdot 06 t), \\x_{t2} &= 4 \cos(2\pi \cdot 10 t) + 5 \sin(2\pi \cdot 10 t), \\x_{t3} &= 6 \cos(2\pi \cdot 40 t) + 7 \sin(2\pi \cdot 40 t), \\x_t &= x_{t1} + x_{t2} + x_{t3}.\end{aligned}$$

What is the major difference between these series and the series generated in [Example 4.1](#)? (Hint: The answer is *fundamental*. But if your answer is the series are longer, you may be punished severely.)

- (b) As in [Example 4.2](#), compute and plot the periodogram of the series,  $x_t$ , generated in (a) and comment.  
(c) Repeat the analyses of (a) and (b) but with  $n = 100$  (as in [Example 4.1](#)), and adding noise to  $x_t$ ; that is

$$x_t = x_{t1} + x_{t2} + x_{t3} + w_t$$

where  $w_t \sim \text{iid } N(0, 25)$ . That is, you should simulate and plot the data, and then plot the periodogram of  $x_t$  and comment.

**4.3** With reference to equations (4.1) and (4.2), let  $Z_1 = U_1$  and  $Z_2 = -U_2$  be independent, standard normal variables. Consider the polar coordinates of the point  $(Z_1, Z_2)$ , that is,

$$A^2 = Z_1^2 + Z_2^2 \quad \text{and} \quad \phi = \tan^{-1}(Z_2/Z_1).$$

- (a) Find the joint density of  $A^2$  and  $\phi$ , and from the result, conclude that  $A^2$  and  $\phi$  are independent random variables, where  $A^2$  is a chi-squared random variable with 2 df, and  $\phi$  is uniformly distributed on  $(-\pi, \pi)$ .  
(b) Going in reverse from polar coordinates to rectangular coordinates, suppose we assume that  $A^2$  and  $\phi$  are independent random variables, where  $A^2$  is chi-squared with 2 df, and  $\phi$  is uniformly distributed on  $(-\pi, \pi)$ . With  $Z_1 = A \cos(\phi)$  and  $Z_2 = A \sin(\phi)$ , where  $A$  is the positive square root of  $A^2$ , show that  $Z_1$  and  $Z_2$  are independent, standard normal random variables.

**4.4** Verify (4.5).

### Section 4.2

**4.5** A time series was generated by first drawing the white noise series  $w_t$  from a normal distribution with mean zero and variance one. The observed series  $x_t$  was generated from

$$x_t = w_t - \theta w_{t-1}, \quad t = 0, \pm 1, \pm 2, \dots,$$

where  $\theta$  is a parameter.

- (a) Derive the theoretical mean value and autocovariance functions for the series  $x_t$  and  $w_t$ . Are the series  $x_t$  and  $w_t$  stationary? Give your reasons.
- (b) Give a formula for the power spectrum of  $x_t$ , expressed in terms of  $\theta$  and  $\omega$ .

**4.6** A first-order autoregressive model is generated from the white noise series  $w_t$  using the generating equations

$$x_t = \phi x_{t-1} + w_t,$$

where  $\phi$ , for  $|\phi| < 1$ , is a parameter and the  $w_t$  are independent random variables with mean zero and variance  $\sigma_w^2$ .

- (a) Show that the power spectrum of  $x_t$  is given by

$$f_x(\omega) = \frac{\sigma_w^2}{1 + \phi^2 - 2\phi \cos(2\pi\omega)}.$$

- (b) Verify the autocovariance function of this process is

$$\gamma_x(h) = \frac{\sigma_w^2 \phi^{|h|}}{1 - \phi^2},$$

$h = 0, \pm 1, \pm 2, \dots$ , by showing that the inverse transform of  $\gamma_x(h)$  is the spectrum derived in part (a).

**4.7** In applications, we will often observe series containing a signal that has been delayed by some unknown time  $D$ , i.e.,

$$x_t = s_t + A s_{t-D} + n_t,$$

where  $s_t$  and  $n_t$  are stationary and independent with zero means and spectral densities  $f_s(\omega)$  and  $f_n(\omega)$ , respectively. The delayed signal is multiplied by some unknown constant  $A$ . Show that

$$f_x(\omega) = [1 + A^2 + 2A \cos(2\pi\omega D)]f_s(\omega) + f_n(\omega).$$

**4.8** Suppose  $x_t$  and  $y_t$  are stationary zero-mean time series with  $x_t$  independent of  $y_s$  for all  $s$  and  $t$ . Consider the product series

$$z_t = x_t y_t.$$

Prove the spectral density for  $z_t$  can be written as

$$f_z(\omega) = \int_{-\frac{1}{2}}^{\frac{1}{2}} f_x(\omega - \nu) f_y(\nu) d\nu.$$

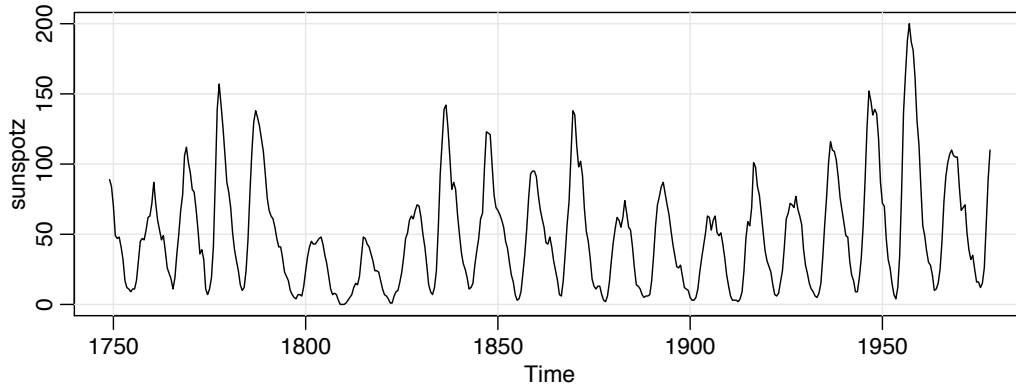


Fig. 4.22. Smoothed 12-month sunspot numbers (`sunspotz`) sampled twice per year.

### Section 4.3

**4.9** Figure 4.22 shows the biyearly smoothed (12-month moving average) number of sunspots from June 1749 to December 1978 with  $n = 459$  points that were taken twice per year; the data are contained in `sunspotz`. With Example 4.13 as a guide, perform a periodogram analysis identifying the predominant periods and obtaining confidence intervals for the identified periods. Interpret your findings.

**4.10** The levels of salt concentration known to have occurred over rows, corresponding to the average temperature levels for the soil science data considered in Figure 1.18 and Figure 1.19, are in `salt` and `saltemp`. Plot the series and then identify the dominant frequencies by performing separate spectral analyses on the two series. Include confidence intervals for the dominant frequencies and interpret your findings.

**4.11** Let the observed series  $x_t$  be composed of a periodic signal and noise so it can be written as

$$x_t = \beta_1 \cos(2\pi\omega_k t) + \beta_2 \sin(2\pi\omega_k t) + w_t,$$

where  $w_t$  is a white noise process with variance  $\sigma_w^2$ . The frequency  $\omega_k$  is assumed to be known and of the form  $k/n$  in this problem. Suppose we consider estimating  $\beta_1$ ,  $\beta_2$  and  $\sigma_w^2$  by least squares, or equivalently, by maximum likelihood if the  $w_t$  are assumed to be Gaussian.

(a) Prove, for a fixed  $\omega_k$ , the minimum squared error is attained by

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = 2n^{-1/2} \begin{pmatrix} d_c(\omega_k) \\ d_s(\omega_k) \end{pmatrix},$$

where the cosine and sine transforms (4.31) and (4.32) appear on the right-hand side.

(b) Prove that the error sum of squares can be written as

$$\text{SSE} = \sum_{t=1}^n x_t^2 - 2I_x(\omega_k)$$

so that the value of  $\omega_k$  that minimizes squared error is the same as the value that maximizes the periodogram  $I_x(\omega_k)$  estimator (4.28).

- (c) Under the Gaussian assumption and fixed  $\omega_k$ , show that the  $F$ -test of no regression leads to an  $F$ -statistic that is a monotone function of  $I_x(\omega_k)$ .

**4.12** Prove the convolution property of the DFT, namely,

$$\sum_{s=1}^n a_s x_{t-s} = \sum_{k=0}^{n-1} d_A(\omega_k) d_x(\omega_k) \exp\{2\pi\omega_k t\},$$

for  $t = 1, 2, \dots, n$ , where  $d_A(\omega_k)$  and  $d_x(\omega_k)$  are the discrete Fourier transforms of  $a_t$  and  $x_t$ , respectively, and we assume that  $x_t = x_{t+n}$  is periodic.

### Section 4.4

**4.13** Analyze the chicken price data ([chicken](#)) using a nonparametric spectral estimation procedure. Aside from the obvious annual cycle discovered in [Example 2.5](#), what other interesting cycles are revealed?

**4.14** Repeat [Problem 4.9](#) using a nonparametric spectral estimation procedure. In addition to discussing your findings in detail, comment on your choice of a spectral estimate with regard to smoothing and tapering.

**4.15** Repeat [Problem 4.10](#) using a nonparametric spectral estimation procedure. In addition to discussing your findings in detail, comment on your choice of a spectral estimate with regard to smoothing and tapering.

**4.16 Cepstral Analysis.** The periodic behavior of a time series induced by echoes can also be observed in the spectrum of the series; this fact can be seen from the results stated in [Problem 4.7](#). Using the notation of that problem, suppose we observe  $x_t = s_t + A s_{t-D} + n_t$ , which implies the spectra satisfy  $f_x(\omega) = [1 + A^2 + 2A \cos(2\pi\omega D)] f_s(\omega) + f_n(\omega)$ . If the noise is negligible ( $f_n(\omega) \approx 0$ ) then  $\log f_x(\omega)$  is approximately the sum of a periodic component,  $\log[1 + A^2 + 2A \cos(2\pi\omega D)]$ , and  $\log f_s(\omega)$ . Bogart et al. (1962) proposed treating the detrended log spectrum as a pseudo time series and calculating its spectrum, or *cepstrum*, which should show a peak at a *quefreny* corresponding to  $1/D$ . The cepstrum can be plotted as a function of quefreny, from which the delay  $D$  can be estimated.

For the speech series presented in [Example 1.3](#), estimate the pitch period using cepstral analysis as follows. The data are in [speech](#).

- Calculate and display the log-periodogram of the data. Is the periodogram periodic, as predicted?
- Perform a cepstral (spectral) analysis on the detrended logged periodogram, and use the results to estimate the delay  $D$ . How does your answer compare with the analysis of [Example 1.27](#), which was based on the ACF?

**4.17** Use [Property 4.2](#) to verify (4.71). Then verify (4.74) and (4.75).

**4.18** Consider two time series

$$x_t = w_t - w_{t-1},$$

$$y_t = \frac{1}{2}(w_t + w_{t-1}),$$

formed from the white noise series  $w_t$  with variance  $\sigma_w^2 = 1$ .

- Are  $x_t$  and  $y_t$  jointly stationary? Recall the cross-covariance function must also be a function only of the lag  $h$  and cannot depend on time.
- Compute the spectra  $f_y(\omega)$  and  $f_x(\omega)$ , and comment on the difference between the two results.
- Suppose sample spectral estimators  $\bar{f}_y(\cdot)$  are computed for the series using  $L = 3$ . Find  $a$  and  $b$  such that

$$P\left\{a \leq \bar{f}_y(\cdot) \leq b\right\} = .90.$$

This expression gives two points that will contain 90% of the sample spectral values. Put 5% of the area in each tail.

### Section 4.5

**4.19** Often, the periodicities in the sunspot series are investigated by fitting an autoregressive spectrum of sufficiently high order. The main periodicity is often stated to be in the neighborhood of 11 years. Fit an autoregressive spectral estimator to the sunspot data using a model selection method of your choice. Compare the result with a conventional nonparametric spectral estimator found in [Problem 4.9](#).

**4.20** Analyze the chicken price data ([chicken](#)) using a parametric spectral estimation procedure. Compare the results to [Problem 4.13](#).

**4.21** Fit an autoregressive spectral estimator to the Recruitment series and compare it to the results of [Example 4.16](#).

**4.22** Suppose a sample time series with  $n = 256$  points is available from the first-order autoregressive model. Furthermore, suppose a sample spectrum computed with  $L = 3$  yields the estimated value  $\bar{f}_x(1/8) = 2.25$ . Is this sample value consistent with  $\sigma_w^2 = 1$ ,  $\phi = .5$ ? Repeat using  $L = 11$  if we just happen to obtain the same sample value.

**4.23** Suppose we wish to test the noise alone hypothesis  $H_0 : x_t = n_t$  against the signal-plus-noise hypothesis  $H_1 : x_t = s_t + n_t$ , where  $s_t$  and  $n_t$  are uncorrelated zero-mean stationary processes with spectra  $f_s(\omega)$  and  $f_n(\omega)$ . Suppose that we want the test over a band of  $L = 2m + 1$  frequencies of the form  $\omega_{j;n} + k/n$ , for  $k = 0, \pm 1, \pm 2, \dots, \pm m$  near some fixed frequency  $\omega$ . Assume that both the signal and noise spectra are approximately constant over the interval.

- (a) Prove the approximate likelihood-based test statistic for testing  $H_0$  against  $H_1$  is proportional to

$$T = \sum_k |d_x(\omega_{j:n} + k/n)|^2 \left( \frac{1}{f_n(\omega)} - \frac{1}{f_s(\omega) + f_n(\omega)} \right).$$

- (b) Find the approximate distributions of  $T$  under  $H_0$  and  $H_1$ .  
 (c) Define the false alarm and signal detection probabilities as  $P_F = P\{T > K|H_0\}$  and  $P_d = P\{T > k|H_1\}$ , respectively. Express these probabilities in terms of the signal-to-noise ratio  $f_s(\omega)/f_n(\omega)$  and appropriate chi-squared integrals.

### Section 4.6

**4.24** Analyze the coherency between the temperature and salt data discussed in **Problem 4.10**. Discuss your findings.

**4.25** Consider two processes

$$x_t = w_t \quad \text{and} \quad y_t = \phi x_{t-D} + v_t$$

where  $w_t$  and  $v_t$  are independent white noise processes with common variance  $\sigma^2$ ,  $\phi$  is a constant, and  $D$  is a fixed integer delay.

- (a) Compute the coherency between  $x_t$  and  $y_t$ .  
 (b) Simulate  $n = 1024$  normal observations from  $x_t$  and  $y_t$  for  $\phi = .9$ ,  $\sigma^2 = 1$ , and  $D = 0$ . Then estimate and plot the coherency between the simulated series for the following values of  $L$  and comment:  
 (i)  $L = 1$ , (ii)  $L = 3$ , (iii)  $L = 41$ , and (iv)  $L = 101$ .

### Section 4.7

**4.26** For the processes in **Problem 4.25**:

- (a) Compute the phase between  $x_t$  and  $y_t$ .  
 (b) Simulate  $n = 1024$  observations from  $x_t$  and  $y_t$  for  $\phi = .9$ ,  $\sigma^2 = 1$ , and  $D = 1$ . Then estimate and plot the phase between the simulated series for the following values of  $L$  and comment:  
 (i)  $L = 1$ , (ii)  $L = 3$ , (iii)  $L = 41$ , and (iv)  $L = 101$ .

**4.27** Consider the bivariate time series records containing monthly U.S. production (**prod**) as measured by the Federal Reserve Board Production Index and the monthly unemployment series (**unemp**).

- (a) Compute the spectrum and the log spectrum for each series, and identify statistically significant peaks. Explain what might be generating the peaks. Compute the coherence, and explain what is meant when a high coherence is observed at a particular frequency.

(b) What would be the effect of applying the filter

$$u_t = x_t - x_{t-1} \quad \text{followed by} \quad v_t = u_t - u_{t-12}$$

to the series given above? Plot the predicted frequency responses of the simple difference filter and of the seasonal difference of the first difference.

(c) Apply the filters successively to one of the two series and plot the output. Examine the output after taking a first difference and comment on whether stationarity is a reasonable assumption. Why or why not? Plot after taking the seasonal difference of the first difference. What can be noticed about the output that is consistent with what you have predicted from the frequency response? Verify by computing the spectrum of the output after filtering.

**4.28** Determine the theoretical power spectrum of the series formed by combining the white noise series  $w_t$  to form

$$y_t = w_{t-2} + 4w_{t-1} + 6w_t + 4w_{t+1} + w_{t+2}.$$

Determine which frequencies are present by plotting the power spectrum.

**4.29** Let  $x_t = \cos(2\pi\omega t)$ , and consider the output

$$y_t = \sum_{k=-\infty}^{\infty} a_k x_{t-k},$$

where  $\sum_k |a_k| < \infty$ . Show

$$y_t = |A(\omega)| \cos(2\pi\omega t + \phi(\omega)),$$

where  $|A(\omega)|$  and  $\phi(\omega)$  are the amplitude and phase of the filter, respectively. Interpret the result in terms of the relationship between the input series,  $x_t$ , and the output series,  $y_t$ .

**4.30** Suppose  $x_t$  is a stationary series, and we apply two filtering operations in succession, say,

$$y_t = \sum_r a_r x_{t-r} \quad \text{then} \quad z_t = \sum_s b_s y_{t-s}.$$

(a) Show the spectrum of the output is

$$f_z(\omega) = |A(\omega)|^2 |B(\omega)|^2 f_x(\omega),$$

where  $A(\omega)$  and  $B(\omega)$  are the Fourier transforms of the filter sequences  $a_t$  and  $b_t$ , respectively.

(b) What would be the effect of applying the filter

$$u_t = x_t - x_{t-1} \quad \text{followed by} \quad v_t = u_t - u_{t-12}$$

to a time series?

(c) Plot the predicted frequency responses of the simple difference filter and of the seasonal difference of the first difference. Filters like these are called seasonal adjustment filters in economics because they tend to attenuate frequencies at multiples of the monthly periods. The difference filter tends to attenuate low-frequency trends.

**4.31** Suppose we are given a stationary zero-mean series  $x_t$  with spectrum  $f_x(\omega)$  and then construct the derived series

$$y_t = ay_{t-1} + x_t, \quad t = \pm 1, \pm 2, \dots$$

- (a) Show how the theoretical  $f_y(\omega)$  is related to  $f_x(\omega)$ .  
 (b) Plot the function that multiplies  $f_x(\omega)$  in part (a) for  $a = .1$  and for  $a = .8$ . This filter is called a recursive filter.

### Section 4.8

**4.32** Consider the problem of approximating the filter output

$$y_t = \sum_{k=-\infty}^{\infty} a_k x_{t-k}, \quad \sum_{-\infty}^{\infty} |a_k| < \infty,$$

by

$$y_t^M = \sum_{|k| < M/2} a_k^M x_{t-k}$$

for  $t = M/2 - 1, M/2, \dots, n - M/2$ , where  $x_t$  is available for  $t = 1, \dots, n$  and

$$a_t^M = M^{-1} \sum_{k=0}^{M-1} A(\omega_k) \exp\{2\pi i \omega_k t\}$$

with  $\omega_k = k/M$ . Prove

$$E\{(y_t - y_t^M)^2\} \leq 4\gamma_x(0) \left( \sum_{|k| \geq M/2} |a_k| \right)^2.$$

**4.33** Prove the squared coherence  $\rho_{y,x}^2(\omega) = 1$  for all  $\omega$  when

$$y_t = \sum_{r=-\infty}^{\infty} a_r x_{t-r},$$

that is, when  $x_t$  and  $y_t$  can be related exactly by a linear filter.

**4.34** The data set [climhyd](#), contains 454 months of measured values for six climatic variables: (i) air temperature [[Temp](#)], (ii) dew point [[DewPt](#)], (iii) cloud cover [[CldCvr](#)], (iv) wind speed [[WndSpd](#)], (v) precipitation [[Precip](#)], and (vi) inflow [[Inflow](#)], at Lake Shasta in California; the data are displayed in [Figure 7.3](#). We would like to look at possible relations among the weather factors and between the weather factors and the inflow to Lake Shasta.



- (a) First transform the inflow and precipitation series as follows:  $I_t = \log i_t$ , where  $i_t$  is inflow, and  $P_t = \sqrt{p_t}$ , where  $p_t$  is precipitation. Then, compute the square coherencies between all the weather variables and transformed inflow and argue that the strongest determinant of the inflow series is (transformed) precipitation. [Tip: If  $\mathbf{x}$  contains multiple time series, then the easiest way to display all the squared coherencies is to plot the coherencies suppressing the confidence intervals, e.g., `mvspec(x, spans=c(7,7), taper=.5, plot.type="coh", ci=-1)`.]
- (b) Fit a lagged regression model of the form

$$I_t = \beta_0 + \sum_{j=0}^{\infty} \beta_j P_{t-j} + w_t,$$

using thresholding, and then comment of the predictive ability of precipitation for inflow.

### Section 4.9

**4.35** Consider the *signal plus noise* model

$$y_t = \sum_{r=-\infty}^{\infty} \beta_r x_{t-r} + v_t,$$

where the signal and noise series,  $x_t$  and  $v_t$  are both stationary with spectra  $f_x(\omega)$  and  $f_v(\omega)$ , respectively. Assuming that  $x_t$  and  $v_t$  are independent of each other for all  $t$ , verify (4.131) and (4.132).

**4.36** Consider the model

$$y_t = x_t + v_t,$$

where

$$x_t = \phi x_{t-1} + w_t,$$

such that  $v_t$  is Gaussian white noise and independent of  $x_t$  with  $\text{var}(v_t) = \sigma_v^2$ , and  $w_t$  is Gaussian white noise and independent of  $v_t$ , with  $\text{var}(w_t) = \sigma_w^2$ , and  $|\phi| < 1$  and  $\text{E}x_0 = 0$ . Prove that the spectrum of the observed series  $y_t$  is

$$f_y(\omega) = \sigma^2 \frac{|1 - \theta e^{-2\pi i \omega}|^2}{|1 - \phi e^{-2\pi i \omega}|^2},$$

where

$$\theta = \frac{c \pm \sqrt{c^2 - 4}}{2}, \quad \sigma^2 = \frac{\sigma_v^2 \phi}{\theta},$$

and

$$c = \frac{\sigma_w^2 + \sigma_v^2(1 + \phi^2)}{\sigma_v^2 \phi}.$$

**4.37** Consider the same model as in the preceding problem.

(a) Prove the optimal smoothed estimator of the form

$$\hat{x}_t = \sum_{s=-\infty}^{\infty} a_s y_{t-s}$$

has

$$a_s = \frac{\sigma_w^2}{\sigma^2} \frac{\theta^{|s|}}{1 - \theta^2}.$$

(b) Show the mean square error is given by

$$E\{(x_t - \hat{x}_t)^2\} = \frac{\sigma_v^2 \sigma_w^2}{\sigma^2(1 - \theta^2)}.$$

(c) Compare mean square error of the estimator in part (b) with that of the optimal finite estimator of the form

$$\hat{x}_t = a_1 y_{t-1} + a_2 y_{t-2}$$

when  $\sigma_v^2 = .053$ ,  $\sigma_w^2 = .172$ , and  $\phi_1 = .9$ .

### Section 4.10

**4.38** Consider the two-dimensional linear filter given as the output (4.143).

- (a) Express the two-dimensional autocovariance function of the output, say,  $\gamma_y(h_1, h_2)$ , in terms of an infinite sum involving the autocovariance function of  $x_s$  and the filter coefficients  $a_{s_1, s_2}$ .
- (b) Use the expression derived in (a), combined with (4.142) and (4.145) to derive the spectrum of the filtered output (4.144).

*The following problems require supplemental material from Appendix C*

**4.39** Let  $w_t$  be a Gaussian white noise series with variance  $\sigma_w^2$ . Prove that the results of **Theorem C.4** hold without error for the DFT of  $w_t$ .

**4.40** Show that condition (4.48) implies (C.19) by showing

$$n^{-1/2} \sum_{h \geq 0} h |\gamma(h)| \leq \sigma_w^2 \sum_{k \geq 0} |\psi_k| \sum_{j \geq 0} \sqrt{j} |\psi_j|.$$

**4.41** Prove **Lemma C.4**.

**4.42** Finish the proof of **Theorem C.5**.

**4.43** For the zero-mean complex random vector  $\mathbf{z} = x_c - ix_s$ , with  $\text{cov}(\mathbf{z}) = \Sigma = C - iQ$ , with  $\Sigma = \Sigma^*$ , define

$$w = 2\text{Re}(a^* \mathbf{z}),$$

where  $a = a_c - ia_s$  is an arbitrary non-zero complex vector. Prove

$$\text{cov}(w) = 2a^* \Sigma a.$$

Recall  $*$  denotes the complex conjugate transpose.

## Chapter 5

---

### Additional Time Domain Topics

In this chapter, we present material that may be considered special or advanced topics in the time domain. [Chapter 6](#) is devoted to one of the most useful and interesting time domain topics, state-space models. Consequently, we do not cover state-space models or related topics—of which there are many—in this chapter. This chapter contains sections of independent topics that may be read in any order. Most of the sections depend on a basic knowledge of ARMA models, forecasting and estimation, which is the material that is covered in [Chapter 3](#). A few sections, for example the section on long memory models, require some knowledge of spectral analysis and related topics covered in [Chapter 4](#). In addition to long memory, we discuss unit root testing, GARCH models, threshold models, lagged regression or transfer functions, and selected topics in multivariate ARMAX models.

#### 5.1 Long Memory ARMA and Fractional Differencing

The conventional ARMA( $p, q$ ) process is often referred to as a short-memory process because the coefficients in the representation

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j},$$

obtained by solving

$$\phi(z)\psi(z) = \theta(z),$$

are dominated by exponential decay. As pointed out in [Section 3.2](#) and [Section 3.3](#), this result implies the ACF of the short memory process satisfies  $\rho(h) \rightarrow 0$  exponentially fast as  $h \rightarrow \infty$ . When the sample ACF of a time series decays slowly, the advice given in [Chapter 3](#) has been to difference the series until it seems stationary. Following this advice with the glacial varve series first presented in [Example 3.33](#) leads to the first difference of the logarithms of the data being represented as a first-order moving average. In [Example 3.41](#), further analysis of the residuals leads to fitting an ARIMA(1, 1, 1) model,