Page 121

CHAPTER 7

Spectral Analysis

Spectral analysis is the name given to methods of estimating the spectral density function, or spectrum, of a given time series.

Before about 1900, research workers such as A.Schuster were essentially concerned with looking for 'hidden periodicities' in data at one or two specific frequencies. Spectral analysis as we know it today is concerned with estimating the spectrum over the whole range of frequencies. The techniques are widely used by many scientists, particularly in electrical engineering, physics, meteorology and marine science.

We are mainly concerned with purely indeterministic processes, which have a continuous spectrum, but the techniques can also be used for deterministic processes to pick out periodic components in the presence of noise.

7.1 Fourier Analysis

Traditional spectral analysis is essentially a modification of Fourier analysis so as to make it suitable for stochastic rather than deterministic functions of time. **Fourier analysis** (e.g. Priestley, 1981) is essentially concerned with approximating a function by a sum of sine and cosine terms, called the Fourier series representation. Suppose that a function $f(t)$ is defined on $(-\pi, \pi]$1 and satisfies the so-called Dirichlet conditions. These conditions ensure that $f(t)$ is reasonably 'well behaved', meaning that, over the range $(-\pi, \pi]$, $f(t)$ is absolutely integrable, has a finite number of discontinuities, and has a finite number of maxima and minima. Then $f(t)$ may be approximated by the Fourier series

$$\frac{a_0}{2} + \sum_{r=1}^{k} (a_r \cos rt + b_r \sin rt)$$

where

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t)\, \mathrm{d}t$$

$$a_r = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos rt\, \mathrm{d}t \qquad r = 1, 2, \ldots$$

$$b_r = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin rt\, \mathrm{d}t \qquad r = 1, 2, \ldots$$

1 The different-shaped brackets indicate that the lower limit $-\pi$ is *not* included in the interval, while the square bracket indicates that the upper limit $+\pi$ *is* included.

Page 122
It can be shown that this Fourier series converges to $f(t)$ as $k \to \infty$ except at points of discontinuity, where it converges to halfway2 up the step change.
In order to apply Fourier analysis to discrete time series, we need to consider the Fourier series representation of $f(t)$ when $f(t)$ is defined only on the integers 1, 2,..., N. Rather than write down the formula, we demonstrate that the required Fourier series emerges naturally by considering a simple sinusoidal model.

7.2 A Simple Sinusoidal Model

Suppose we suspect that a given time series, with observations made at unit time intervals, contains a deterministic sinusoidal component at a known frequency ω, together with a random error term. Then we will consider the model

$$X_t = \mu + \alpha \cos \omega t + \beta \sin \omega t + Z_t \tag{7.1}$$

where Zt denotes a purely random process, and μ, α, β are parameters to be estimated from the data. The observations will be denoted by ($x1$, $x2$,..., xN). The algebra in the next few sections is somewhat simplified if we confine ourselves to the case where N is *even*. There is no real difficulty in extending the results to the case where N is odd (e.g. Anderson, 1971), and indeed many of the later estimation formulae apply for both odd and even N, but some results require one to consider odd N and even N separately. Thus, if N happens to be odd and a spectral analysis is required, computation can be made somewhat simpler by removing the first observation so as to make N even. If N is reasonably large, little information is lost.
Expected values for the model in Equation (7.1) can be represented in matrix notation by

$$E(\boldsymbol{X}) = A\,\boldsymbol{\theta}$$

where

$$\begin{aligned} \boldsymbol{X}^T &= (X_1, \ldots, X_N) \\ \boldsymbol{\theta}^T &= (\mu, \alpha, \beta) \\ A &= \begin{pmatrix} 1 & \cos\omega & \sin\omega \\ 1 & \cos 2\omega & \sin 2\omega \\ \cdots & \cdots & \cdots \\ 1 & \cos N\omega & \sin N\omega \end{pmatrix} \end{aligned}$$

As this model is linear in the parameters μ, α and β, it is an example of a general linear model. In that case the least squares estimate of θ, which minimizes $\sum_{t=1}^{N} (x_t - \mu - \alpha \cos \omega t - \beta \sin \omega t)^2$, is 'well known' to be

$$\hat{\boldsymbol{\theta}} = (A^T A)^{-1} A^T \boldsymbol{x}$$

2 Mathematicians say that this is the average of the limit from below and the limit from above, sometimes written as $\frac{1}{2}[f(t-0) + f(t+0)]$.

Page 123
where

$$x^T = (x_1, \ldots, x_N)$$

The above formulae hold for any value of the frequency ω, but they only make practical sense for values of ω that are not too high or too low. As noted in Section 6.2, the highest frequency we can uniquely fit to the data is the Nyquist frequency, given by $\omega=\pi$, which completes one cycle every two observations. At the other end of the spectrum, the lowest frequency we can reasonably fit completes one cycle in the whole length of the time series. These upper and lower limits will be explained further in Section 7.2.1 below. By equating the cycle length $2\pi/\omega$ to N, we find that the lowest frequency is given by $2\pi/N$. The formulae for the least squares estimates of $\hat{\theta}$ turn out to be particularly simple if ω, is restricted to one of the values

$$\omega_p = 2\pi p/N \quad p = 1, \ldots, N/2$$

which lie in equal steps from the lowest frequency $2\pi/N$ to the Nyquist frequency π. In this case, it turns out that (ATA) is a diagonal matrix in view of the following 'well-known' trigonometric results (all summations are for t=1 to N):

$$\sum \cos \omega_p t = \sum \sin \omega_p t = 0 \tag{7.2}$$

$$\sum \cos \omega_p t \cos \omega_q t = \begin{cases} 0 & p \neq q \\ N & p = q = N/2 \\ N/2 & p = q \neq N/2 \end{cases} \tag{7.3}$$

$$\sum \sin \omega_p t \sin \omega_q t = \begin{cases} 0 & p \neq q \\ 0 & p = q = N/2 \\ N/2 & p = q \neq N/2 \end{cases} \tag{7.4}$$

$$\sum \cos \omega_p t \sin \omega_q t = 0 \qquad \text{for all } p, q \tag{7.5}$$

With (ATA) diagonal, it is easy to evaluate the least squares estimate of θ, as the inverse $(ATA)-1$ will also be diagonal. For ω_p such that $p\neq N/2$, we find (Exercise 7.2)

$$\begin{aligned} \hat{\mu} &= \sum x_t/N = \bar{x} \\ \hat{\alpha} &= 2\left[\sum x_t \cos \omega_p t\right]/N \\ \hat{\beta} &= 2\left[\sum x_t \sin \omega_p t\right]/N \end{aligned} \tag{7.6}$$

If $p=N/2$, we ignore the term in β sin ωt, which is zero for all t, and find

$$\begin{aligned} \hat{\mu} &= \bar{x} \\ \hat{\alpha} &= \sum(-1)^t x_t/N \end{aligned} \tag{7.7}$$

The model in Equation (7.1) is essentially the one used before about 1900 to search for hidden periodicities, but this model has now gone out of fashion. However, it can still be useful if there is reason to suspect that a time series

Page 124
does contain a deterministic periodic component at a known frequency and it is desired to isolate this component (e.g. Bloomfield, 2000, Chapters 2, 3).
Readers who are familiar with the analysis of variance (ANOVA) technique will be able to work out that the total corrected sum of squared deviations, namely,

$$\sum_{t=1}^{N}(x_t - \bar{x})^2$$

can be partitioned into two components, namely, the residual sum of squares and the sum of squares 'explained' by the periodic component at frequency ω_p. This latter component is given by

$$\sum_{t=1}^{N}(\hat{\alpha}\cos\omega_p t + \hat{\beta}\sin\omega_p t)^2$$

which, after some algebra (Exercise 7.2), can be shown to be

$$\begin{array}{ll}(\hat{\alpha}^2+\hat{\beta}^2)N/2 & p \neq N/2 \\ \hat{\alpha}^2 N & p = N/2\end{array} \qquad (7.8)$$

using Equations (7.2)–(7.5).
7.2.1 The highest (Nyquist) frequency and the lowest (fundamental) frequency
When fitting the simple sinusoidal model in Equation (7.1), we restricted the frequency ω to one of the values ($2\pi/N$, $4\pi/N$,..., π), assuming that N is even. Here we examine the practical rationale for the upper and lower limits, namely, π and $2\pi/N$.
In Section 6.2, we pointed out that, for a discrete-time process measured at unit intervals of time, there is no loss of generality in restricting the spectral distribution function to the range (0, π). We now demonstrate that the upper bound π, called the **Nyquist frequency,** is indeed the highest frequency about which we can get meaningful information from a set of data.
First, we give a more general form for the Nyquist frequency. If observations are taken at equal intervals of time of length Δt, then the Nyquist (angular) frequency is given by $\omega_N = \pi/\Delta t$. The equivalent frequency expressed in cycles per unit time is $f_N = \omega_N/2\pi = 1/2\Delta t$.
Consider the following example. Suppose that temperature readings are taken every day in a certain town at noon. It is clear that these observations will tell us nothing about temperature variation *within* a day. In particular, they will not tell us whether nights are hotter or cooler than days. With only one observation per day, the Nyquist frequency is $\omega_N = \pi$ radians per day or $f_N = \frac{1}{2}$ cycle per day (or 1 cycle per 2 days). This is lower than the frequencies, which correspond to variation within a day. For example, variation with a period of 1 day has (angular) frequency $\omega = 2\pi$ radians per day or f=1 cycle per day. In order to get information about variation within

Page 125
a day at these higher frequencies, we must increase the sampling rate and take two or more observations per day.
A similar example is provided by yearly sales figures. These will obviously give no information about any seasonal effects, whereas monthly or quarterly observations *will* give information about seasonality.
At the other end of the spectrum, we will now explain why there is a *lowest* frequency below which it is not sensible to try to fit to a set of data. If we had just 6 months of temperature readings from winter to summer, the analyst would not be able to decide, from the data alone, whether there is an upward trend in the observations or whether winters are colder than summers. However, with 1 year's data, it *would* become clear that winters are colder than summers. Thus if we are interested in variation at the low frequency of 1 cycle per year, then we should have at least 1 year's data, in which case the lowest frequency we can fit is at 1 cycle per year. With weekly observations, for example, 1 year's data have N=52, Δt=1 week, and the lowest angular frequency of $2\pi/N\Delta t$ corresponds to a frequency of $1/N\Delta t$ cycles per week. (Note that all time units must be expressed in terms of the same period, here a week.) The lowest frequency is therefore 1/52 cycles per week, which can now be converted to 1 cycle per year.
The lowest frequency, namely, $2\pi/N\Delta$, is sometimes called the **fundamental Fourier frequency,** because the Fourier series representation of the data is normally evaluated at the frequencies $\omega_p = 2\pi p/N\Delta$ for p=1,..., N/2, which are all integer multiples of the fundamental frequency. These integer multiples are often called **harmonics**. The phrase *fundamental frequency* is perhaps more typically, and more helpfully, used when a function, $f(t)$ say**,** is **periodic** with period T so that $f(t+nT)=f(t)$ for all integer values of n. Then f=1/T, or $\omega = 2\pi/T$, is called the fundamental frequency and the Fourier series representation of $f(t)$ is a sum over integer multiples, or harmonics, of the fundamental frequency. When $T= N\Delta$=(the length of the observed time series), the fundamental frequencies coincide. This raises a practical point, in regard to choosing the length of a time series. Suppose, for example, that you are collecting weekly data and are particularly interested in annual variation. As noted above, you should collect at least 1 year's data. If you collect exactly 52 weeks of data3, then the fundamental frequency will be at exactly 1 cycle per year. We will see that this makes it much easier to interpret the results of a spectral analysis. The fundamental frequency is at 1 cycle per year and the harmonics are at 2 cycles per year, 3 cycles per year and so on. However, if you have say an extra 12 weeks of data making 64 weeks, then it will be much harder to interpret the results at frequencies $\omega_p = 2\pi p/N\Delta$. Wherever possible, you should choose the length of the time series so that the harmonics cover the frequencies of particular interest. The easiest option is to collect observations covering an integer multiple of the lowest wavelength of particular interest. This ensures
3 For simplicity, ignore day 365, and day 366 if a leap year.

Page 126
that this frequency is an integer multiple of the fundamental frequency. Thus, to investigate annual variation, 2 years of data is good and 3 or 4 years of data even better.
The reader will notice that the Nyquist frequency does *not* depend on *N,* but rather only on the sampling frequency, whereas the lowest frequency *does* depend on *N*. Put another way, the lower the frequency we are interested in, the longer the time period over which we need to take measurements, whereas the higher the frequency we are interested in, the more frequently must we take observations.

7.3 Periodogram Analysis

Early attempts at discovering hidden periodicities in a given time series basically consisted of repeating the analysis of Section 7.2 at all the frequencies 2*π*/*N,* 4*π*/*N*,..., *π*. In view of Equations (7.3)–(7.5), the different terms are orthogonal and we end up with the finite Fourier series representation of the *{xt},* namely

$$x_t = a_0 + \sum_{p=1}^{(N/2)-1} [a_p \cos(2\pi pt/N) + b_p \sin(2\pi pt/N)] + a_{N/2} \cos \pi t \tag{7.9}$$

for *t*=1, 2,..., *N*, where the coefficients *{ap, bp}* are of the same form as Equations (7.6) and (7.7), namely

$$\begin{aligned} a_0 &= \bar{x} \\ a_{N/2} &= \sum (-1)^t x_t / N \\ \left.\begin{aligned} a_p &= 2\left[\sum x_t \cos(2\pi pt/N)\right]/N \\ b_p &= 2\left[\sum x_t \sin(2\pi pt/N)\right]/N \end{aligned}\right\} & \quad p = 1, \ldots, (N/2) - 1 \end{aligned} \tag{7.10}$$

An analysis along these lines is sometimes called a **Fourier analysis** or a **harmonic analysis**. The Fourier series representation in Equation (7.9) has *N* parameters to describe *N* observations and so can be made to fit the data exactly (just as a polynomial of degree *N*−1 involving *N* parameters can be found that goes exactly through *N* observations in polynomial regression). This explains why there is no error term in Equation (7.9) in contrast to Equation (7.1). Also note that there is no term in sin *πt* in Equation (7.9) as sin *πt* is zero for all integer *t*.
It is worth stressing that the Fourier series coefficients in Equation (7.10) at a given frequency ω are exactly the same as the least squares estimates for the coefficients of the model in Equation (7.1).
The overall effect of the Fourier analysis of the data is to partition the variability of the series into components at frequencies 2*π*/*N,* 4*π*/*N*,..., *π*. The component at frequency $\omega_p = 2\pi p/N$ is called the *p*th harmonic. For

Page 127

$p \neq N/2$, it can be useful to write the pth harmonic in the equivalent form

$$a_p \cos \omega_p t + b_p \sin \omega_p t = R_p \cos(\omega_p t + \phi_p) \tag{7.11}$$

where

$$R_p = \sqrt{(a_p^2 + b_p^2)} \tag{7.12}$$

is the **amplitude** of the *p*th harmonic, and

$$\phi_p = \tan^{-1}(-b_p/a_p) \tag{7.13}$$

is the **phase** of the *p*th harmonic.

We have already noted in Section 7.2 that, for $p \neq N/2$, the contribution of the *p*th harmonic to the total sum of squares is given by $N(a_p^2 + b_p^2)/2$ Using Equation (7.12), this is equal to $NR_p^2/2$. Extending this result using Equations (7.2)–(7.5) and (7.9), we have, after some algebra (Exercise 7.3), that

$$\sum_{t=1}^{N}(x_t - \bar{x})^2 = N \sum_{p=1}^{(N/2)-1} R_p^2/2 + N a_{N/2}^2$$

Dividing through by *N* we have

$$\sum (x_t - \bar{x})^2/N = \sum_{p=1}^{(N/2)-1} R_p^2/2 + a_{N/2}^2 \tag{7.14}$$

which is known as **Parseval's theorem**. The left-hand side of Equation (7.14) is effectively the variance[4] of the observations. Thus $R_p^2/2$ is the contribution of the pth harmonic to the variance, and Equation (7.14) shows how the total variance is partitioned.

If we plot $R_p^2/2$ against $\omega_p = 2\pi p/N$, we obtain a line spectrum. A different type of line spectrum occurs in the physical sciences when light from molecules in a gas discharge tube is viewed through a spectroscope. The light has energy at discrete frequencies and this energy can be seen as bright lines. However, most time series have continuous spectra, and then it is inappropriate to plot a line spectrum. If we regard $R_p^2/2$ as the contribution to variance in the range $\omega_p \pm \pi/N$, we can plot a histogram whose height in the range $\omega_p \pm \pi/N$ is such that

$$\begin{aligned} R_p^2/2 &= \text{area of histogram rectangle} \\ &= \text{height of histogram} \times 2\pi/N \end{aligned}$$

Thus the height of the histogram at ω_p, denoted by $I(\omega_p)$, is given by

$$I(\omega_p) = NR_p^2/4\pi \tag{7.15}$$

As usual, Equation (7.15) does not apply for $p = N/2$; we may regard $a_{N/2}^2$ as the contribution to variance in the range $[(N-1)\pi/N, \pi]$ so that

$$I(\pi) = Na_{N/2}^2/\pi$$

4 The divisor is N rather than the more usual $(N-1)$, but this makes little difference for large *N*.

Page 128

The plot of $I(\omega)$ against ω is usually called the **periodogram,** even though $I(\omega)$ is a function of frequency rather than period. It follows from Parseval's theorem in Equation (7.14) that the total area under the periodogram is equal to the variance of the time series.

Note that the formula for R_p^2 and hence for $I(\omega_p)$, can be written in several equivalent ways that look quite different. For example, after some algebra, it can be shown that

$$I(\omega_p) = \frac{1}{\pi N}\left|\sum_{t=1}^{N} x_t e^{it\omega_p}\right|^2 \tag{7.16}$$

or we can replace $e^{it\omega}$ with $e^{-it\omega}$ in Equation (7.16). The usual way to actually calculate the periodogram directly from the data uses the expression

$$I(\omega_p) = \left[\left(\sum x_t \cos 2\pi pt/N\right)^2 + \left(\sum x_t \sin 2\pi pt/N\right)^2\right]/N\pi \tag{7.17}$$

Equation (7.17) also applies for *p*=*N*/2.

Other authors define the periodogram in what appear to be slightly different ways, but the differences usually arise from allowing negative frequencies or using the cyclic frequency $f = \omega/2\pi$, rather than ω. The expressions generally turn out to be some other multiple of *I(ωp)* or R_p^2. For example, Hannan (1970, Equation (3.8)) and Koopmans (1995, Equation (8.7)) give expressions that correspond to $\frac{1}{2}\times$expression (7.16). As to terminology, Anderson (1971, Section 4.3.2) describes the graph of R_p^2 against the *period N/p,* as the periodogram, and suggests the term **spectrogram** to describe the graph of R_p^2 against frequency. Jenkins and Watts (1968) define a similar expression to Equation (7.17) in terms of the variable $f = \omega/2\pi$, but call it the 'sample spectrum'. As always, when comparing terms and formulae from different sources, the reader needs to take great care.

The periodogram appears to be a natural way of estimating the power spectral density function, but Section 7.3.2 shows that, for a process with a **continuous** spectrum, it provides a poor estimate and needs to be modified. First, we derive the relationship between the periodogram of a given time series and the corresponding autocovariance function (acv.f.).

7.3.1 The relationship between the periodogram and the acv.f.

The periodogram ordinate $I(\omega_p)$ and the autocovariance coefficient *ck* are both quadratic forms of the data *{xt}*. It is therefore natural to enquire how they are related. In fact, we will show that the periodogram is the finite Fourier transform of *{ck}*.

Using Equation (7.2), we may rewrite Equation (7.17) for *p*≠*N*/2 as

$$I(\omega_p) = \left\{\left[\sum (x_t - \bar{x})\cos \omega_p t\right]^2 + \left[\sum (x_t - \bar{x})\sin \omega_p t\right]^2\right\}/N\pi$$

Page 129

$$= \sum_{s,t=1}^{N} (x_t - \bar{x})(x_s - \bar{x})(\cos\omega_p t \cos\omega_p s + \sin\omega_p t \sin\omega_p s)/N\pi$$

However, (see Equation (4.1))

$$\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})/N = c_k$$

and $\cos\omega_p t \cos\omega_p (t+k) + \sin\omega_p t \sin\omega_p (t+k) = \cos\omega_p (t+k-t) = \cos\omega_p k$

so that

$$I(\omega_p) = \left(c_0 + 2\sum_{k=1}^{N-1} c_k \cos\omega_p k\right)/\pi \tag{7.18}$$

$$= \sum_{k=-(N-1)}^{N-1} c_k e^{-i\omega_p k}/\pi \tag{7.19}$$

The formula in Equation (7.19) is an expression called a **discrete finite Fourier transform** (assuming that *ck*=0 for |*k*| ≥*N*). Any reader not familiar with the Fourier transform, is recommended to read Appendix A—see especially Equation (A.5).

7.3.2 Properties of the periodogram

When the periodogram is expressed in the form of Equation (7.18), it appears to be the 'obvious' estimate of the power spectrum

$$f(\omega) = \left(\gamma_0 + 2\sum_{k=1}^{\infty} \gamma_k \cos\omega k\right)/\pi$$

simply replacing *γk* by its estimate *ck* for values of *k* up to (*N*−1), and putting subsequent estimates of *γk* equal to zero. However, although we find

$$\underset{N\to\infty}{E}[I(\omega)] \to f(\omega) \tag{7.20}$$

so that the periodogram is asymptotically unbiased, we see below that the variance of $I(\omega)$ does not decrease as *N* increases. Thus $I(\omega)$ is *not a consistent estimator* for $f(\omega)$. An example of a periodogram is given later in Figure 7.5(c), and it can be seen that the graph fluctuates wildly. The lack of consistency is perhaps not too surprising when one realizes that the Fourier series representation in Equation (7.9) requires one to evaluate *N* parameters from *N* observations, however long the series may be. Thus in Section 7.4 we will consider alternative ways of estimating a power spectrum that are essentially ways of *smoothing* the periodogram.

We complete this section by proving that $I(\omega)$ is not a consistent estimator for $f(\omega)$ in the case where the observations are assumed to be independent *N*(*μ, σ*2) variates, so that they form a discrete-time purely random process with a uniform spectrum. This result can be extended to other stationary

Page 130
processes with continuous spectra, but this does not need to be demonstrated here. If the periodogram estimator does not 'work' for a uniform spectrum, it cannot be expected to 'work' for more complicated spectra. Given the above assumptions, Equation (7.10) shows that *ap* and *bp* are linear combinations of normally distributed random variables and so will themselves be normally distributed. Using Equations (7.2)–(7.4), it can be shown (Exercise 7.4) that *ap* and *bp* each have mean zero and variance $2\sigma 2/N$ for $p \neq N/2$. Furthermore we have

$$\begin{aligned} \mathrm{Cov}(a_p, b_p) &= 4\mathrm{Cov}\left[\left(\sum X_t \cos\omega_p t\right), \left(\sum X_t \sin\omega_p t\right)\right]/N^2 \\ &= 4\sigma^2\left(\sum \cos\omega_p t \sin\omega_p t\right)/N^2 \end{aligned}$$

since the observations are assumed to be independent. Thus, using Equation (7.5), we see that *ap* and *bp* are uncorrelated. Since *(ap, bp)* are bivariate normal, zero correlation implies that *ap* and *bp* are independent. The variables *ap* and *bp* can be standardized by dividing by $\sqrt{2}\sigma/\sqrt{N}$ to give standard *N*(0, 1) variables. Now a result from distribution theory says that if *Y*1, *Y*2 are independent *N*(0, 1) variables, then $(Y_1^2 + Y_2^2)$ has a *X*2 distribution with two degrees of freedom, which is written χ_2^2. Thus

$$\frac{N(a_p^2 + b_p^2)}{2\sigma^2} = \frac{I(\omega_p)2\pi}{\sigma^2}$$

is χ_2^2. Put another way, this means that $2I(\omega)/f(\omega)$ is χ_2^2 when $f(\omega) = \sigma^2/\pi$ as in this case, although this result does, in fact, generalize to spectra that are not constant. Now the variance of a *X*2 distribution with *v* degrees of freedom is 2*v*, so that

$$\mathrm{Var}[I(\omega_p)2\pi/\sigma^2] = 4$$

and

$$\mathrm{Var}[I(\omega_p)] = \sigma^4/\pi^2$$

As this variance is a constant, it does *not* tend to zero as $N \to \infty$, and hence $I(\omega_p)$ is not a consistent estimator for $f(\omega_p)$. Furthermore it can be shown that neighbouring periodogram ordinates are asymptotically independent, which further explains the very irregular form of an observed periodogram. This all means that the periodogram needs to be modified in order to obtain a good estimate of a continuous spectrum.

7.4 Some Consistent Estimation Procedures

This section describes several alternative ways of estimating a spectrum. The different methods will be compared in Section 7.6. Each method provides a *consistent* estimator for the (power) spectral density function, in contrast to the (raw) periodogram. However, although the periodogram is itself an inconsistent estimator, the procedures described in this section are essentially based on smoothing the periodogram in some way.

Throughout the section we will assume that any obvious trend and seasonal

Page 131
variation have been removed from the data. If this is not done, the results of the spectral analysis are likely to be dominated by these effects, making any other effects difficult or impossible to see. Trend produces a peak at zero frequency, while seasonal variation produces peaks at the seasonal frequency and at integer multiples of the seasonal frequency—the seasonal **harmonics** (see Section 7.2.1). For a nonstationary series, the estimated spectrum of the detrended, deseasonalized data will depend to some extent on the method chosen to remove trend and seasonality. We assume throughout that a 'good' method is used to do this.

The methods described in this chapter are essentially *non-parametric* in that no model fitting is involved. It is possible to use a model-based approach and an alternative, parametric approach, called **autoregressive spectrum estimation,** will be introduced later in Section 13.7.1.

7.4.1 Transforming the truncated acv.f.

One type of estimation procedure consists of taking a Fourier transform of the truncated weighted sample acv.f. From Equation (7.18), we know that the periodogram is the discrete finite Fourier transform of the complete sample acv.f. However, it is clear that the precision of the values of *ck* decreases as *k* increases, because the coefficients are based on fewer and fewer terms. Thus, it would seem intuitively reasonable to give less weight to the values of *ck* as *k* increases. An estimator, which has this property is

$$\hat{f}(\omega) = \frac{1}{\pi}\left\{\lambda_0 c_0 + 2\sum_{k=1}^{M} \lambda_k c_k \cos \omega k\right\} \tag{7.21}$$

where *{λk}* are a set of weights called the **lag window,** and *M*(**<***N*) is called the **truncation point**. Comparing Equation (7.21) with (7.18) we see that values of *ck* for *M*<*k*<*N* are no longer used, while values of *ck* for *k*≤*M* are weighted by a factor *λk*. The latter are chosen so as to get smaller as *k* approaches *M*.

In order to use the above estimator, the analyst must choose a suitable lag window and a suitable truncation point. The two best-known lag windows are as follows.

Tukey window

$$\lambda_k = \frac{1}{2}\left(1 + \cos\frac{\pi k}{M}\right) \quad k = 0, 1, \ldots, M$$

This window is sometimes called the Tukey-Hanning or Blackman-Tukey window.

Page 132
Parzen window

$$\lambda_k = \begin{cases} 1 - 6\left(\frac{k}{M}\right)^2 + 6\left(\frac{k}{M}\right)^3 & 0 \le k \le M/2 \\ 2(1 - k/M)^3 & M/2 \le k \le M \end{cases}$$

These two windows are illustrated in Figure 7.1 with *M*=20.

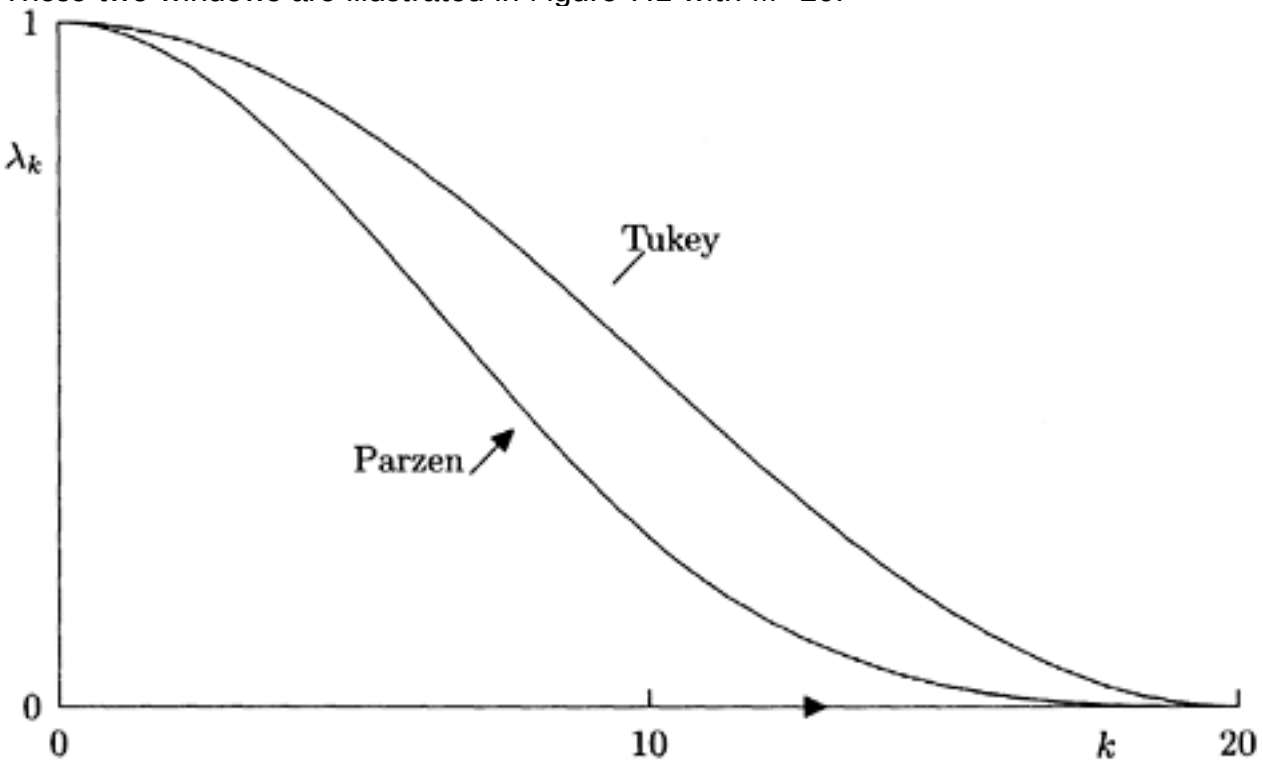


Figure 7.1 The Tukey and Parzen lag windows with *M*=20.

The Tukey and Parzen windows give very similar estimated spectra for a given time series, although the Parzen window has a slight advantage in that it cannot give negative estimates. Many other lag windows have been suggested and 'window carpentry' was a popular research topic in the 1950s. Ways of comparing different windows will be discussed in Section 7.6. The well-known **Bartlett window,** with $\lambda k = 1 - k/M$ for $k = 0, 1, \ldots, M$, is very simple but is now rarely used as its properties are inferior to those of the Tukey and Parzen windows.

The choice of the truncation point *M* is more difficult and it is not easy to give clear-cut advice. It has to be chosen subjectively so as to *balance 'resolution' against 'variance'*. The smaller the value of *M,* the smaller will be the variance of $\hat{f}(\omega)$ but the larger will be the bias. If *M* is too small, important features of $f(\omega)$ may be smoothed out, but if *M* is too large the behaviour of $\hat{f}(\omega)$ becomes more like that of the periodogram with erratic variation. Thus a compromise value must be chosen. A useful rough guide is to choose *M* to be about $2\sqrt{N}$, so that if, for example, *N* is 200, then *M* will be round about the value 28. This choice of *M* ensures the asymptotic situation that as $N \to \infty$, so also does $M \to \infty$ but in such a way that $M/N \to 0$. A somewhat larger value of *M* is required for the Parzen window than for the Tukey window. Other writers have suggested $\sqrt{N}$ rather than

Page 133

$2\sqrt{N}$, while results from density estimation suggest that a different power of *N* may be appropriate. Percival and Walden (1993, Chapter 6) point out that an appropriate value of *M* depends on the properties of the underlying process and give more detailed guidance. However, my advice is to try three or four different values of *M*. A low value will give an idea where the large peaks in $f(\omega)$ are, but the curve is likely to be too smooth. A high value is likely to produce a curve showing a large number of peaks, some of which may be spurious. A compromise can then be achieved with an in-between value of *M*.

In principle, Equation (7.21) may be evaluated at any value of ω in (0, π), but it is usually evaluated at equal intervals at $\omega = \pi j/Q$ for *j*=0, 1..., *Q,* where *Q* is chosen sufficiently large to show up all features of $\hat{f}(\omega)$. Often *Q* is chosen to be equal to *M*. The graph of $\hat{f}(\omega)$ against ω can then be plotted and examined. An example is given later in Figure 7.5, for the data plotted in Figure 1.2, using the Tukey window with *M*=24.

7.4.2 Hanning

This procedure, named after Julius Von Hann, is equivalent to the use of the Tukey window as described in Section 7.4.1, but adopts a different computational procedure. The estimated spectrum is calculated in two stages. First, a truncated unweighted cosine transform of the acv.f. of the data is taken to give

$$\hat{f}_1(\omega) = \frac{1}{\pi}\left(c_0 + 2\sum_{k=1}^{M} c_k \cos\omega k\right) \tag{7.22}$$

This is the same as Equation (7.21) except that the lag window is taken to be unity (i.e. λk=1). The estimates given by Equation (7.22) are calculated at $\omega = \pi j/M$ for *j*=0, 1,..., *M*. These estimates are then smoothed using the weights $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ to give the Hanning estimates

$$\hat{f}(\omega) = \frac{1}{4}\hat{f}_1(\omega - \pi/M) + \frac{1}{2}\hat{f}_1(\omega) + \frac{1}{4}\hat{f}_1(\omega + \pi/M) \tag{7.23}$$

at $\omega = \pi j/M$ for *j*=1, 2 ,..., (*M*−1). At zero frequency, and at the Nyquist frequency π, we take

$$\hat{f}(0) = \frac{1}{2}[\hat{f}_1(0) + \hat{f}_1(\pi/M)]$$

$$\hat{f}(\pi) = \frac{1}{2}[\hat{f}_1(\pi) + \hat{f}_1(\pi(M-1)/M)]$$

It can easily be shown algebraically that this procedure is equivalent to the use of the Tukey window. Substituting Equation (7.22) into (7.23) we find

$$\hat{f}(\omega) = \frac{1}{\pi}\left\{c_o + 2\sum_{k=1}^{M} c_k\left[\frac{1}{4}\cos(\omega - \pi/M)k + \frac{1}{2}\cos\omega k + \frac{1}{4}\cos(\omega + \pi/M)k\right]\right\}$$

Page 134

and, using $\cos(\omega - \pi/M)k + \cos(\omega + \pi/M)k = 2\cos\omega k \cos(\pi k/M)$, a comparison with Equation (7.21) shows that the lag window is indeed the Tukey window.

There is relatively little difference in the computational efficiency of Hanning and the straightforward use of the Tukey window. Both methods should yield the same estimates and so it does not matter which of the two procedures is used in practice.

7.4.3 Hamming

This technique is very similar to Hanning and has a very similar title, which sometimes leads to confusion. In fact Hamming is named after a quite different person, namely R.W.Hamming. The technique is nearly identical to Hanning except that the weights $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ in Equation (7.23) are changed to (0.23, 0.54, 0.23). At the frequencies ω=0 and ω=π, the weights are 0.54 (at the 'end' frequency) and 0.46. The procedure gives similar estimates to those produced by Hanning.

7.4.4 Smoothing the periodogram

The methods of Sections 7.4.1–7.4.3 are based on transforming the truncated sample acv.f. An alternative type of approach is to smooth the periodogram ordinates in some way, the simplest approach being to group the periodogram ordinates in sets of size m and find their average value. The latter approach is based on a suggestion made by P.J.Daniell as long ago as 1946. However, the use of lag window estimators was standard for many years because less computation was involved. Nowadays, some form of smoothed periodogram is used much more widely, particularly with the advent of the fast Fourier transform—see Section 7.4.5.

The basic idea of the simple smoothed periodogram can be expressed in the following formula:

$$\hat{f}(\omega) = \frac{1}{m}\sum_{j} I(\omega_j) \tag{7.24}$$

where $\omega_j = 2\pi j/N$ and j varies over m consecutive integers so that the ω_j are symmetric about the frequency of interest, namely, ω. In order to estimate $f(\omega)$ at the end-points ω=0 and ω=π, Equation (7.24) has to be modified in an obvious way, treating the periodogram as being symmetric about 0 and π. Then, taking m to be odd with $m^*=(m-1)/2$, we have

$$\hat{f}(0) = I(0) + 2\sum_{j=1}^{m^*} I(2\pi j/N)/m$$

$$\hat{f}(\pi) = \left[I(\pi) + 2\sum_{j=1}^{m^*} I(\pi - 2\pi j/N)\right]\Big/ m$$

Page 135

The expression for $\hat{f}(0)$ can be simplified as the first term $I(0)$ is zero.
Now we know that the periodogram is asymptotically unbiased but inconsistent for the true spectrum. Since neighbouring periodogram ordinates are asymptotically uncorrelated, it is clear that the variance of Equation (7.24) will be of order $1/m$. It is also clear that the estimator in Equation (7.24) may be biased since

$$E[\hat{f}(\omega)] \simeq \frac{1}{m}\sum_{j} f(\omega_j)$$

which is only equal to $f(\omega)$ if the spectrum is linear over the relevant interval. However, the bias will be 'small' provided that $f(\omega)$ is a reasonably smooth function and m is not too large compared with N.
The consequence of the above remarks is that the choice of group size m is rather like the choice of the truncation point M in Section 7.4.1 in that it has to be chosen so as to balance resolution against variance. However, the choice is different in that changes in m and in M act in opposite directions. An increase in m has a similar effect to a reduction in M. The larger the value of m the smaller will be the variance of the resulting estimate but the larger will be the bias. If m is too large, then interesting features of $f(\omega)$, such as peaks, may be smoothed out. Of course, as N increases, we can in turn allow m to increase, just as we allowed M to increase with N in Section 7.4.1.
There is relatively little advice in the literature on the choice of m. As in Section 7.4.1, it seems advisable to try several values for m. A 'high' value should give some idea as to whether large peaks in $f(\omega)$ exist, but the curve is likely to be too smooth and some real peaks may be hidden. A 'low' value is likely to produce a much more uneven curve showing many peaks, some of which will be spurious. A compromise between the effects of bias and variance can then be made. In earlier editions, I suggested trying values near $N/40$, but I now think that $2\sqrt{N}$ is probably a better guideline.
Although the procedure described in this section is computationally quite different to that of Section 7.4.1, there are in fact close theoretical links between the two procedures. In Section 7.3.1 we derived the relationship between the periodogram and the sample acv.f., and, if we substitute Equation (7.18) into (7.24), we can express the smoothed periodogram estimate of the spectrum in terms of the sample acv.f. in a similar form to Equation (7.21). After some algebra (Exercise 7.5), it can be shown that the truncation point is $(N-1)$ and the lag window is given by

$$\lambda_k = \begin{cases} 1 & k = 0 \\ \sin(m\pi k/N)/[m\sin(\pi k/N)] & k = 1, 2, \ldots, N-1 \end{cases}$$

Thus, the formula uses values of ck right up to $k=(N-1)$, rather than having a truncation point much lower than N. Moreover, the lag window has the undesirable property that it does not tend to zero as k tends to N. The smoothed periodogram effectively uses a rectangular window in the frequency domain and the resulting lag window shows that a sudden cut-off in the frequency domain can give rise to 'nasty' effects in the time domain

Page 136
(and vice versa). The smoothed periodogram often works reasonably well, but its window properties suggest that it may be possible to find a way of smoothing the periodogram, using a non-uniform averaging procedure, that has better time-domain properties. In fact, the simple smoothed periodogram is rarely used today, but rather a windowed form of averaging is used instead. Various alternative smoothing procedures have been suggested, with the idea of giving more weight to the periodogram ordinate at the particular frequency of interest and progressively less weight to periodogram ordinates further away. The analyst can think of this as applying a window in the frequency domain rather than in the time domain, but in a way that corresponds to the use of a lag window as in Section 7.4.1. It is possible to use a triangular (Bartlett window) or aim for a bell-shaped curve, perhaps by applying a simple smoother, such as Hanning, more than once. These approaches will not be considered here and the reader is referred, for example, to Hayes (1996, Chapter 8) or Bloomfield (2000, Chapter 8).
Historically, the smoothed periodogram was not much used until the 1990s because it apparently requires much more computational effort than transforming the truncated acv.f. Calculating the periodogram using Equation (7.17) at ω_p for p=1, 2,..., N/2 would require about $N2$ arithmetic operations (each one a multiplication and an addition), whereas using Equation (7.21) fewer than MN operations are required to calculate the $\{ck\}$ so that the total number of operations is only of order $M(N+M)$ if the spectrum is evaluated at M frequencies. Two factors have led to the increasing use of the smoothed periodogram. First, the advent of high-speed computers means that it is unnecessary to restrict attention to the method requiring fewest calculations. The second factor has been the widespread use of an algorithm called the **fast Fourier transform,** which makes it much quicker to compute the periodogram. This procedure will now be described.

7.4.5 The fast Fourier transform (FFT)

The computational procedure described in this section is usually abbreviated to FFT5 and we adopt this abbreviation. For long series, the technique can substantially reduce the time required to perform a Fourier analysis of a set of data on a computer, and can also give more accurate results.
The history of the FFT dates back to the early 1900s. However, it was the work of J.W.Cooley, J.W.Tukey and G.Sande in about 1965 coupled with the arrival of faster computers that stimulated the application of the technique to time-series analysis. Much of the early work was published in the various Transactions of the IEEE, but more recent coverage is given, for example, by Bendat and Piersol (2000), Bloomfield (2000) and Priestley (1981). We will only give a broad outline of the technique here.
The FFT requires that the value of N should be composite, meaning that

5 Some authors have used this abbreviation to denote the *finite Fourier transform*.

Page 137

N is not a prime number and so can be factorized. The basic idea of the FFT will be illustrated for the case when *N* can be factorized in the form *N*=*rs,* where *r* and *s* are integers. If we assume that *N* is even, then at least one of the factors, say *r,* will be even. Using complex numbers for mathematical simplicity, the Fourier coefficients from Equation (7.10) can be expressed in the form

$$a_p + \mathrm{i}b_p = 2\left[\sum x_t \, \mathrm{e}^{2\pi \mathrm{i} p t/N}\right]/N \tag{7.25}$$

for *p*=0, 1, 2 ,..., (*N*/2)−1. For mathematical convenience, we denote the observations by *x*0, *x*1,..., *xN*−1, so that the summation in Equation (7.25) is from *t*=0 to *N*−1. Now we can write *t* in the form

$$t = rt_1 + t_0$$

where *t*1=0, 1,..., *s*−1, and *t*0=0, 1,..., r−1, as *t* goes from 0 to *N*−1, in view of the fact that *N*=*rs*. Similarly we can decompose *p* in the form

$$p = sp_1 + p_0$$

where *p*1=0, 1,..., (*r*/2)−1, and *P*0=0, 1,..., *s*−1, as *p* goes from 0 to (*N*/2)−1. Then the summation in Equation (7.25) may be written

$$\sum_{t_0=0}^{r-1} \mathrm{e}^{2\pi \mathrm{i} p t_0/N} \sum_{t_1=0}^{s-1} x_t \, \mathrm{e}^{2\pi \mathrm{i} p r t_1/N}$$

However,

$$\mathrm{e}^{2\pi \mathrm{i} p r t_1/N} = \mathrm{e}^{2\pi \mathrm{i}(sp_1+p_0)rt_1/N} = \mathrm{e}^{2\pi \mathrm{i} p_0 r t_1/N}$$

since $\mathrm{e}^{2\pi \mathrm{i} s p_1 r t_1/N} = \mathrm{e}^{2\pi \mathrm{i} p_1 t_1} = 1$ for all p_1, t_1. Thus $\sum_{t_1=0}^{s-1} x_t \, \mathrm{e}^{2\pi \mathrm{i} p r t_1/N}$ does not depend on *p*1 and is therefore a function of *t*0 and *P*0 only, say *A*(*p*0, *t*0). Then Equation (7.25) may be written

$$a_p + \mathrm{i}b_p = 2\left[\sum_{t_0=0}^{r-1} A(p_0, t_0) \, \mathrm{e}^{2\pi \mathrm{i} p t_0/N}\right] \Big/ N$$

Now there are *N*=*rs* functions of type *A*(*p*0, *t*0) to be calculated, each requiring *s* complex multiplications and additions. There are *N*/2 values of *(ap+ibp)* to be calculated, each requiring *r* further complex multiplications and additions. This gives a grand total of $Ns + \frac{N}{2}r = N(s + r/2)$ calculations instead of the *N*×*N*/2=*N*2/2 calculations required to use Equation (7.25) directly. By a suitable choice of *s* and *r,* we can usually arrange for (*s*+*r*/2) to be (much) less than *N*/2.

Much bigger reductions in computing can be made by an extension of the above procedure when *N* is highly composite (i.e. has many small factors). In particular, if *N* is of the form 2*k*, then we find that the number of operations is of order *Nk* (or *N* log2 *N*) instead of *N*2/2. Substantial gains can also be made when *N* has several factors (e.g. *N*=2p3q5r...).

In practice it is unlikely that N will naturally be of a simple form such as 2*k,* unless the value of *N* can be chosen before measurement starts. However, there

Page 138
are other things we can do. It may be possible to make *N* highly composite by the simple expedient of omitting a few observations from the beginning or end of the series. For example, with 270 observations, we can omit the last 14 to make *N*=256=28. More generally we can *increase* the length of the series by adding zeros to the (mean-corrected) sample record until the value of the revised *N* becomes a suitable integer. Then a procedure called **tapering** or **data windowing** (e.g. Percival and Walden, 1993; Priestley, 1981) is often recommended6 to avoid a discontinuity at the end of the data. Suppose, for example, that we happen to have 382 observations. This value of *N* is *not* highly composite and we might proceed as follows:

- Remove any linear trend from the data, and keep the residuals (which should have mean zero) for subsequent analysis. If there is no trend, simply subtract the overall mean from each observation.
- Apply a linear taper to about 5% of the data at each end. In this example, if we denote the detrended mean-corrected data by *x*0, *x*1,..., *x*381, then the tapered series is given by

$$x_t^* = \begin{cases} (t+1)x_t/20 & t = 0, 1, \ldots, 18 \\ (382-t)x_t/20 & t = 363, \ldots, 381 \\ x_t & t = 19, 20, \ldots, 362 \end{cases}$$

- Add 512−382=130 zeros at one end of the tapered series, so that *N*=512=29.
- Carry out an FFT on the data, calculate the Fourier coefficients *ap*+*ibp* and average the values of $(a_p^2 + b_p^2)$ in groups of about 10.

In fact with *N* as low as 382, the computational advantage of the FFT is limited and we could equally well calculate the periodogram directly, which avoids the need for tapering and adding zeros. The FFT really comes into its own when there are several thousand observations.

It is also worth explaining that the FFT is still useful when the analyst prefers to look at the autocorrelation function (ac.f.) *before* carrying out a spectral analysis, either because inspecting the ac.f. is thought to be an invaluable preliminary exercise or because the analyst prefers to transform the truncated weighted acv.f. rather than smooth the periodogram. It can be quicker to calculate the sample acv.f. by performing *two* FFTs (e.g. Priestley, 1981, Section 7.6), rather than directly as a sum of lagged products. The procedure is as follows. Compute the Fourier coefficients *(ap, bp)* with an FFT of the mean-corrected data at $\omega_p = 2\pi p/N$ for *p*=0, 1,..., *N*−1 rather than for *p*=0, 1,..., *N*/2 as we usually do. The extra coefficients are normally redundant for real-valued processes since *aN*−*k*=*ak* and *bN*−*k*=−*bk*. However, for calculating the autocovariances, we can compute $R_p^2 = a_p^2 + b_p^2$ at these values of *p* and then fast Fourier retransform the sequence (R_p^2) to

6 Note that some researchers view tapering with suspicion as the data are modified—see, for example, the discussion in Percival and Walden (1993, p. 215).

Page 139

get the mean lagged products. We will not give the algebra here. For several thousand observations, this can be much faster than calculating them directly. However, when using the FFT in this way, the analyst should take care to add enough zeros to the data (*without* tapering) to make sure that **non-circular** sums of lagged products are calculated, as defined by Equation (4.1) and used throughout this book. **Circular** coefficients result if zeros are not added where, for example, the circular autocovariance coefficient at lag 1 is

$$c_1^* = \left[\sum_{t=1}^{N}(x_t - \bar{x})(x_{t+1} - \bar{x})\right] \Big/ N$$

where *xN*+1 is taken to be equal to *x*1 to make the series 'circular'. Note that, if *x*1=$\bar{x}$, then the circular and non-circular coefficients at lag 1 are the same. If we use mean-corrected data, which will have mean zero, then adding zeros will make circular and non-circular coefficients be the same. In order to calculate all the non-circular autocovariance coefficients of a set of *N* mean-corrected observations, the analyst should add *N* zeros, to make 2*N* 'observations' in all.

7.5 Confidence Intervals for the Spectrum

The methods of Section 7.4 all produce *point* estimates of the spectral density function, and hence give no indication of their likely accuracy. This section shows how to find appropriate confidence intervals.

In Section 7.3.2, we showed that data from a white noise process, with constant spectrum $f(\omega) = \sigma^2/\pi$, yields a periodogram ordinate $I(\omega)$ at frequency ω, which is such that $2I(\omega)/f(\omega)$ is distributed as χ^2_2. Note that this distibution does not depend on *N*, which explains why $I(\omega)$ is not a consistent estimator for $f(\omega)$. Wide confidence intervals would result if $I(\omega)$ was used as an estimator.

Suppose instead that we use the estimator of Section 7.4.1, namely

$$\hat{f}(\omega) = \left[\sum_{k=-M}^{M} \lambda_k c_k \cos \omega k\right] \Big/ \pi$$

Then, it can be shown (Jenkins and Watts, 1968, Section 6.4.2) that $\nu\hat{f}(\omega)/f(\omega)$ is asymptotically distributed as an approximate χ^2_ν random variable, where

$$\nu = 2N \Big/ \sum_{k=-M}^{M} \lambda_k^2 \qquad (7.26)$$

is called the number of *degrees of freedom* of the lag window. It follows that

$$P(\chi^2_{\nu,1-\alpha/2} < \nu\hat{f}(\omega)/f(\omega) < \chi^2_{\nu,\alpha/2}) = 1 - \alpha$$

so that a 100(1−*a*)% confidence interval for $f(\omega)$ is given by

$$\frac{\nu\hat{f}(\omega)}{\chi^2_{\nu,\alpha/2}} \quad \text{to} \quad \frac{\nu\hat{f}(\omega)}{\chi^2_{\nu,1-\alpha/2}}$$

Page 140
Some simple algebra shows that the degrees of freedom for the Tukey and Parzen windows turn out to be 2.67*N*/*M* and 3.71*N*/*M,* respectively. Although relying on asymptotic results, Neave (1972a) has shown that the above formulae are also quite accurate for short series.
For the smoothed periodogram estimator of Section 7.4.4, there is no need to apply Equation (7.26), because smoothing the periodogram in groups of size *m* is effectively the same as averaging independent χ^2_2 random variables. Thus, it is clear that the smoothed periodogram will have *v*=2*m* degrees of freedom, and we can then apply the same formula for the confidence interval as given above.

7.6 Comparison of Different Estimation Procedures

Several factors need to be considered when comparing the different estimation procedures that were introduced in Section 7.4. Although we concentrate on the theoretical properties of the different procedures, the analyst will also need to consider practical questions such as computing time and the availability of suitable computer software. Alternative comparative discussions are given by Jenkins and Watts (1968), Neave (1972b), Priestley (1981, Section 7.5) and Bloomfield (2000).
It is helpful to introduce a function called the **spectral window** or **kernel,** which is defined to be the Fourier transform of the lag window $\{\lambda_k\}$ introduced in Equation (7.21). Assuming that λ_k is zero for $k>M$, and symmetric, so that $\lambda_{-k} = \lambda_k$ then the spectral window is defined by

$$K(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \lambda_k e^{-ik\omega} \tag{7.27}$$

for $(-\pi<\omega<\pi)$[7]. The corresponding inverse Fourier transform is given by

$$\lambda_k = \int_{-\pi}^{\pi} K(\omega) e^{i\omega k} d\omega \tag{7.28}$$

All the estimation procedures for the spectrum that we have studied so far can be put in the general form

$$\begin{aligned} \hat{f}(\omega_0) &= \frac{1}{\pi} \sum_{k=-N+1}^{N-1} \lambda_k c_k e^{-i\omega_0 k} \\ &= \frac{1}{\pi} \sum \left[\int_{-\pi}^{\pi} K(\omega) e^{i\omega k} d\omega \right] c_k e^{-i\omega_0 k} \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} K(\omega) \left[\sum c_k e^{ik(\omega-\omega_0)} \right] d\omega \end{aligned}$$

7 Note that we cannot avoid negative frequencies here, as the spectral window looks at differences in frequency from some specified frequency. If the lag window is symmetric about zero, then so is $K(\omega)$.

Page 141

$$= \int_{-\pi}^{\pi} K(\omega) I(\omega_0 - \omega) \mathrm{d}\omega \tag{7.29}$$

using Equation (7.19). Equation (7.29) shows that all the estimation procedures are essentially smoothing the periodogram using the weight function $K(\omega)$. The value of the lag window at lag zero is usually specified to be one, so that from Equation (7.28) we have

$$\lambda_0 = 1 = \int_{-\pi}^{\pi} K(\omega) \mathrm{d}\omega$$

which is a desirable property for a smoothing function.
Taking expectations in Equation (7.29) we have asymptotically that

$$E[\hat{f}(\omega_0)] = \int_{-\pi}^{\pi} K(\omega) f(\omega_0 - \omega) \mathrm{d}\omega \tag{7.30}$$

Thus the spectral window is a weight function expressing the contribution of the spectral density function at each frequency to the expectation of $\hat{f}(\omega_0)$. The name 'window' arises from the fact that $K(\omega)$ determines the part of the periodogram that is 'seen' by the estimator.

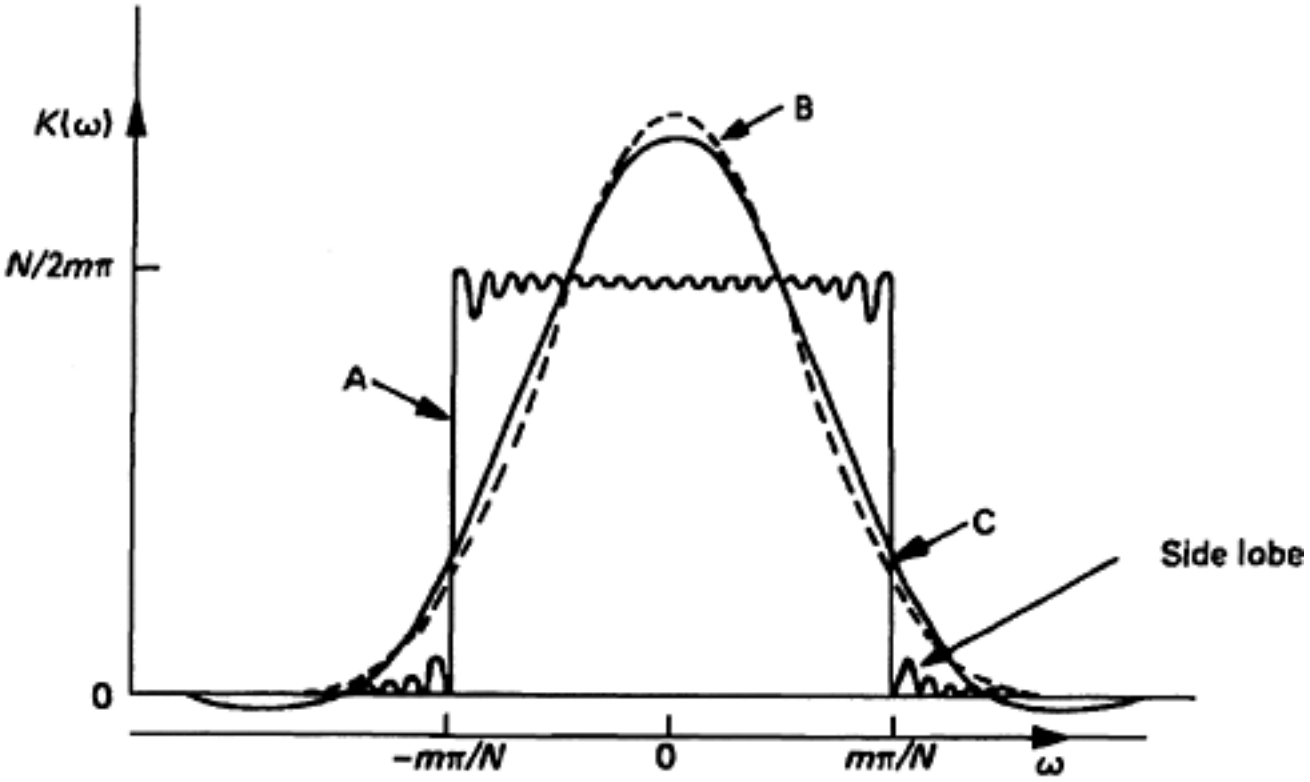


Figure 7.2 The spectral windows for three common methods of spectral analysis: A, smoothed periodogram (*m*=20); B, Parzen (*M*=93); C, Tukey (*M*=67); all with *N*=1000.

Page 142

Examples of the spectral windows for three common methods of spectral analysis are shown in Figure 7.2. Taking N=1000, the spectral window for the smoothed periodogram with m=20 is shown as line A. The other two windows are the Parzen and Tukey windows, denoted by lines B and C. The values of the truncation point M were chosen to be 93 for the Parzen window and 67 for the Tukey window. These values of M were chosen so that all three windows gave estimators with equal variance. Formulae for variances will be given later in this section.

Inspecting Figure 7.2, we see that the Parzen and Tukey windows look very similar, although the Parzen window has the advantage of being non-negative and of having smaller side lobes. The shape of the periodogram window is quite different. It is approximately rectangular with a sharp cut-off and is close to the 'ideal' band-pass filter, which would be exactly rectangular but which is unattainable in practice. The periodogram window also has the advantage of being non-negative.

In comparing different windows, we should consider both the bias and the variance of the estimator. This is sometimes called the *variance-bias trade-off* question, as well as *balancing resolution against variance*. By taking a wider window, we generally get a lower variance but a larger bias and some sort of compromise has to be made in practice. This is often achieved by using trial and error, as, for example, in the choice of the truncation point for a lag window as discussed earlier in Section 7.4.1. It is not easy to get general formulae for the bias produced by the different procedures. However, it is intuitively clear from Equation (7.30) and from earlier remarks that the wider the window, the larger will be the bias. In particular, it is clear that all the smoothing procedures will tend to lower peaks and raise troughs.

As regards variance, we noted in Section 7.5 that $\nu\hat{f}(\omega)/f(\omega)$ is approximately distributed as χ^2_ν, where ν=2m, for the smoothed periodogram, and, using Equation (7.26), 3.71N/M and 8N/3M for the Parzen and Tukey windows, respectively. Since

$$\mathrm{Var}(\chi^2_\nu) = 2\nu$$

and

$$\mathrm{Var}[\nu\hat{f}(\omega)/f(\omega)] = \nu^2\mathrm{Var}[\hat{f}(\omega)/f(\omega)]$$

we find $\mathrm{Var}[\hat{f}(\omega)/f(\omega)]$ turns out to be 1/m, 2M/3.71N, and 3M/4N, respectively, for the three windows. Equating these expressions gives the values of M chosen for Figure 7.2.

When comparing the different estimators, the notion of a **bandwidth** may be helpful. Roughly speaking, the bandwidth is the width of the spectral window, as might be expected. Various formal definitions are given in the literature, but we adopt the one given by Jenkins and Watts (1968), namely, the width of the 'ideal' rectangular window that would give an estimator with the same variance. The window of the smoothed periodogram is so close to being rectangular for m 'large' that it is clear from Figure 7.2 that the bandwidth will be approximately 2$m\pi/N$ (as area must be unity and height

Page 143
is $N/2m\pi$). The bandwidths for the Bartlett, Parzen and Tukey windows turn out to be $3/2M$, $2\pi(1.86/M)$ and $8\pi/3M$, respectively. When plotting a graph of an estimated spectrum, it is a good idea to indicate the bandwidth that has been used.
The choice of bandwidth is equivalent to the choice of m or M, depending on the method used. This choice is an important step in spectral analysis, though it is important to remember that the effects of changing m and M act in opposite directions. For the Bartlett, Parzen and Tukey windows, the bandwidth is inversely proportional to M. Figure 7.3 shows how the window changes as M varies, using the Bartlett window as a representative example. As M gets larger, the window gets narrower, the bias gets smaller but the variance of the resulting estimator gets larger. For the smoothed periodogram, the reverse happens. The bandwidth is directly proportional to m, and as m gets larger, the window gets wider, the bias increases but the variance reduces. For the unsmoothed periodogram, with m=1, the window is very tall and narrow giving an estimator with large variance as we have already shown. All in all, the choice of bandwidth is rather like the choice of class interval when constructing a histogram.

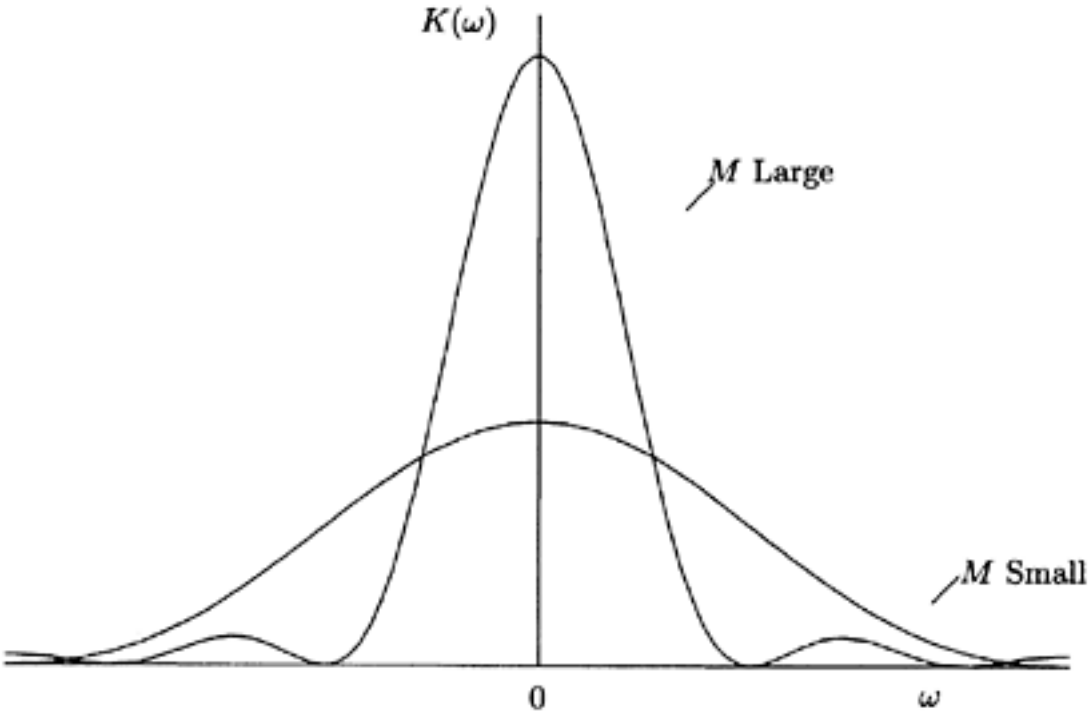


Figure 7.3 The Bartlett spectral window for different values of M.
We are now in a position to give guidance on the relative merits of the different estimation procedures. As regards theoretical properties, it is arguable that the smoothed periodogram has the better-shaped spectral window in that it is approximately rectangular, although there are some side lobes. For the transformed acv. f., the Parzen and Tukey windows are preferred to the Bartlett window. Computationally, the smoothed periodogram can be much slower for large N unless the FFT is used. However, if the FFT *is* used, then the smoothed periodogram can be faster. Moreover it is

Page 144
possible to calculate the ac.f. quickly using *two* FFTs. One drawback to the use of the FFT is that it may require data-tapering, whose use is still somewhat controversial. Of course, for small N, computing time is a relatively unimportant consideration. As regards computer software, it is much easier to write a program for the Parzen or Tukey windows, but programs and algorithms for the FFT are becoming readily available. Thus the use of the smoothed periodogram has become more general, either with equal weights as discussed in Section 7.4.4, or with a more 'bell-shaped' set of weights.
All the above assumes that a *non-parametric* approach is used in that no model fitting is attempted prior to carrying out a spectral analysis. As noted earlier, an alternative approach gaining ground is to use a *parametric* approach, fitting an autoregressive (AR) or ARMA model to the data. The spectrum of the fitted model is then used to estimate the spectrum. This approach will be described later in Section 13.7.1.

7.7 Analysing a Continuous Time Series

Up to now, we have been concerned with the spectral analysis of time series recorded at *discrete* time intervals. However, time series are sometimes recorded as a *continuous trace*. For example, variables such as air temperature, humidity and the moisture content of tobacco emerging from a processing plant are often recorded by machines that give continuous-time readings. For series that contain components at very high frequencies, such as those arising in acoustics and speech processing, it may be possible to analyse such records mechanically using tuned filters, but the more usual procedure is to **digitize** the series by reading off the values of the trace at discrete intervals. If values are taken at equal time intervals of length Δt, we have converted a continuous time series into a standard discrete-time time series and can use the methods already described.
In sampling a continuous time series, the main question is how to choose the sampling interval Δt. It is clear that sampling leads to some loss of information and that this loss gets worse as Δt increases. However, sampling costs increase as Δt gets smaller and so a compromise value must be sought.
For the sampled series, the Nyquist frequency is $\pi/\Delta t$ radians per unit time, and we can get no information about variation at higher frequencies. Thus we clearly want to choose Δt so that variation in the continuous series is negligible at frequencies higher than $\pi/\Delta t$. In fact most measuring instruments are **bandlimited** in that they do not respond to frequencies higher than a certain maximum frequency. If this maximum frequency, say ω_{max}, is known or can be guessed, then the choice of Δt straightforward in that it should be less than π/ω_{max}. However, if Δt is chosen to be too large, then a phenomenon called **aliasing** may occur. This can be illustrated by the following theorem.

Page 145

Theorem 7.1 Suppose that a continuous time series, with spectrum $f_c(\omega)$ for $0<\omega<\infty$, is sampled at equal time intervals of length Δt. The resulting discrete time series will have a somewhat different spectrum, say $f_d(\omega)$ defined over $0<\omega<\pi/\Delta t$. We will see that the two spectra will only be equal if $f_c(\omega)$ is zero for $\omega>\pi/\Delta t$. More generally, it can be shown that $f_d(\omega)$ and $f_c(\omega)$ are related by

$$f_d(\omega) = \sum_{s=0}^{\infty} f_c(\omega + 2\pi s/\Delta t) + \sum_{s=1}^{\infty} f_c(-\omega + 2\pi s/\Delta t) \tag{7.31}$$

Proof The proof will be given for the case $\Delta t=1$. The extension to other values of Δt is straightforward. Suppose that the acv.f.s of the continuous and sampled series are given by $\gamma(\tau)$ and γ_k, respectively. Here $\gamma(\tau)$ is defined for all τ, while γ_k is only defined for integer k. Of course if τ takes an integer value, say k, then the two functions are equal as in

$$\gamma(k) = \gamma_k \tag{7.32}$$

Now from Equation (6.18) we have

$$\gamma(\tau) = \int_0^{\infty} f_c(\omega) \cos \omega\tau \, d\omega$$

while, from Equation (6.9), we have

$$\gamma_k = \int_0^{\pi} f_d(\omega) \cos \omega k \, d\omega$$

Thus, using Equation (7.32), we have

$$\int_0^{\pi} f_d(\omega) \cos \omega k \, d\omega = \int_0^{\infty} f_c(\omega) \cos \omega k \, d\omega$$

for k=0, ±1, ±2,.... The next step is to split the infinite integral into sections of length 2π, and then of π, using $\cos \omega k = \cos (\omega + 2\pi s)k = \cos (2\pi s - \omega)k$ for all integer s. We get

$$\begin{aligned}
\int_0^{\infty} f_c(\omega) \cos \omega k \, d\omega &= \sum_{s=0}^{\infty} \int_{2\pi s}^{2\pi(s+1)} f_c(\omega) \cos \omega k \, d\omega \\
&= \sum_{s=0}^{\infty} \int_0^{2\pi} f_c(\omega + 2\pi s) \cos \omega k \, d\omega \\
&= \sum_{s=0}^{\infty} \int_0^{\pi} \{f_c(\omega + 2\pi s) + f_c[2\pi(s+1) - \omega]\} \cos \omega k \, d\omega \\
&= \int_0^{\pi} \left\{ \sum_{s=0}^{\infty} f_c(\omega + 2\pi s) + \sum_{s=1}^{\infty} f_c(2\pi s - \omega) \right\} \cos \omega k \, d\omega
\end{aligned}$$

and the result follows. □

< previous page | page_146 | next page >

Page 146
The implications of this theorem may now be considered. First, as noted earlier, if the continuous series contains no variation at frequencies above the Nyquist frequency, so that $f_c(\omega) = 0 \text{ for } \omega > \pi/\Delta t$ then $f_d(\omega) = f_c(\omega)$. In this case no information is lost by sampling. However, the more general result is that sampling *will* have an effect in that variation at frequencies above the Nyquist frequency in the continuous series will be 'folded back' to produce apparent variation in the sampled series at a frequency lower than the Nyquist frequency. If we denote the Nyquist frequency $\pi/\Delta t$ by ω_N, then the frequencies $\omega,\ 2\omega_N - \omega, 2\omega_N + \omega, 4\omega_N - \omega, \cdots$ are called aliases of one another. Variation at all these frequencies in the continuous series will appear as variation at frequency ω in the sampled series.
From a practical point of view, aliasing will cause trouble unless Δt is chosen to be sufficiently small so that $f_c(\omega) \simeq 0 \text{ for } \omega > \pi/\Delta t$. If we have no advance knowledge about $f_c(\omega)$, then we have to guesstimate a value for Δt. If the resulting estimate of $f_d(\omega)$ approaches zero near the Nyquist frequency $\pi/\Delta t$, then our choice of Δt is almost certainly sufficiently small. However, if $f_d(\omega)$ does not approach zero near the Nyquist frequency, then it is probably wise to try a smaller value of Δt. Alternatively, if the analyst is only interested in the low-frequency components, then it may be easier to filter the continuous series so as to remove the high-frequency components and remove the need for selecting a small value of Δt.

7.8 Examples and Discussion

Spectral analysis can be a useful exploratory diagnostic tool in the analysis of many types of time series. With the aid of examples, this section discusses how to interpret an estimated spectrum, and tries to indicate when spectral analysis is likely to be most useful and when it is likely to be unhelpful. We also discuss some of the practical problems arising in spectral analysis.
We begin with an example to give the reader some 'feel' for the sorts of spectrum shapes that may arise. Figure 7.4 shows four sections of trace, labelled A, B, C and D, which were produced by four different processes (generated in a control engineering laboratory). Figure 7.4 also shows the corresponding spectra calculated from longer series than the short sections shown here. The spectra are labelled J, K, L and M, but are given in *random* order. Note that the four traces use the same scale, the length produced in one second being shown on trace D. The four spectra are also plotted using the same scales (linear in both directions). The peak in spectrum L is at 15 cycles per second (or 15 Hz).
Before reading on, the reader is invited to guess which series goes with which spectrum. This is not easy, especially for the novice time-series analyst, but even experienced analysts may have difficulty. Try to assess which series are smoother (and hence have less high frequency variation) and which oscillate quicker. Spotting a deterministic sinusoidal perturbation in the presence of noise is more difficult than you might expect.

< previous page | page_146 | next page >

Page 147

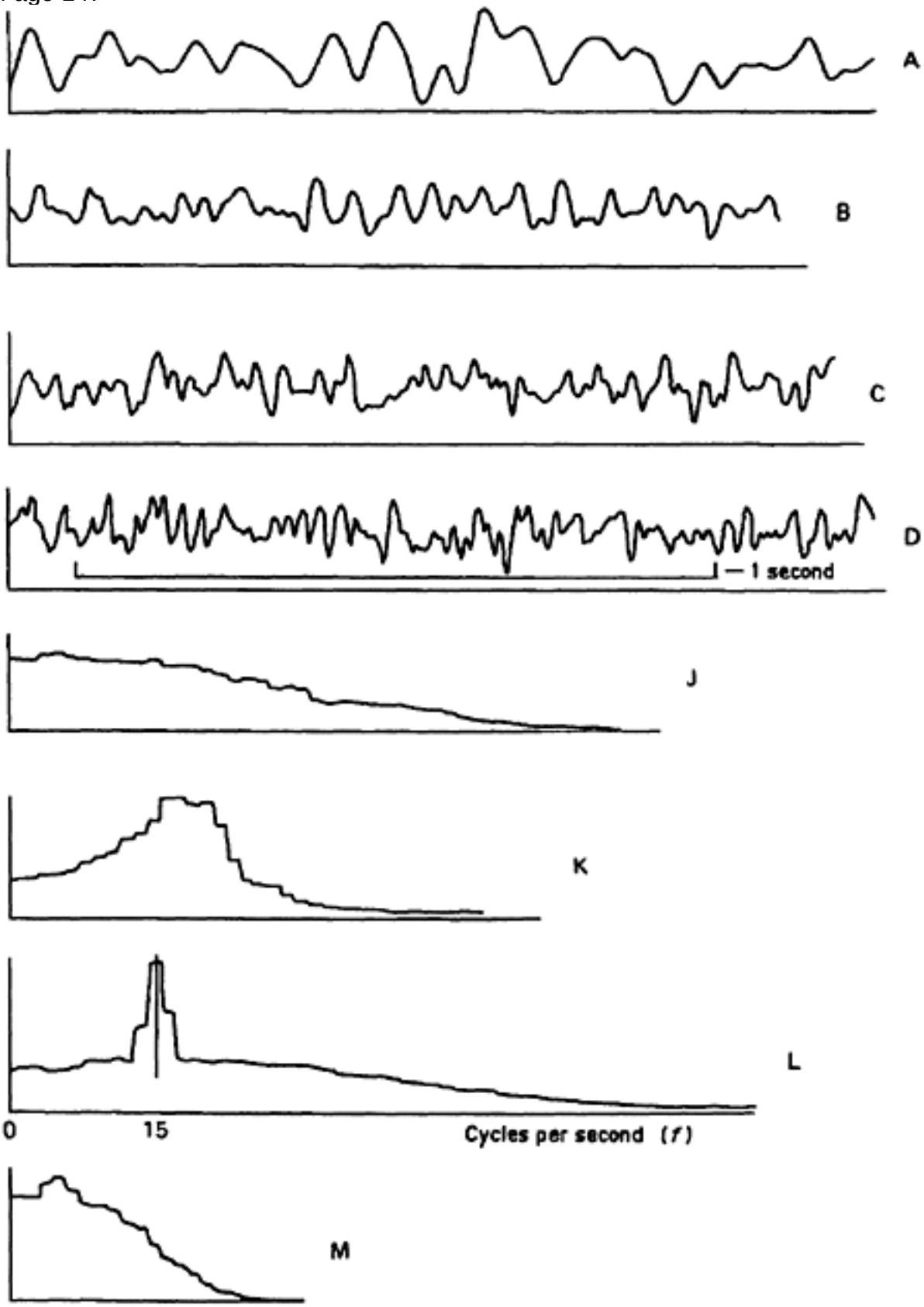


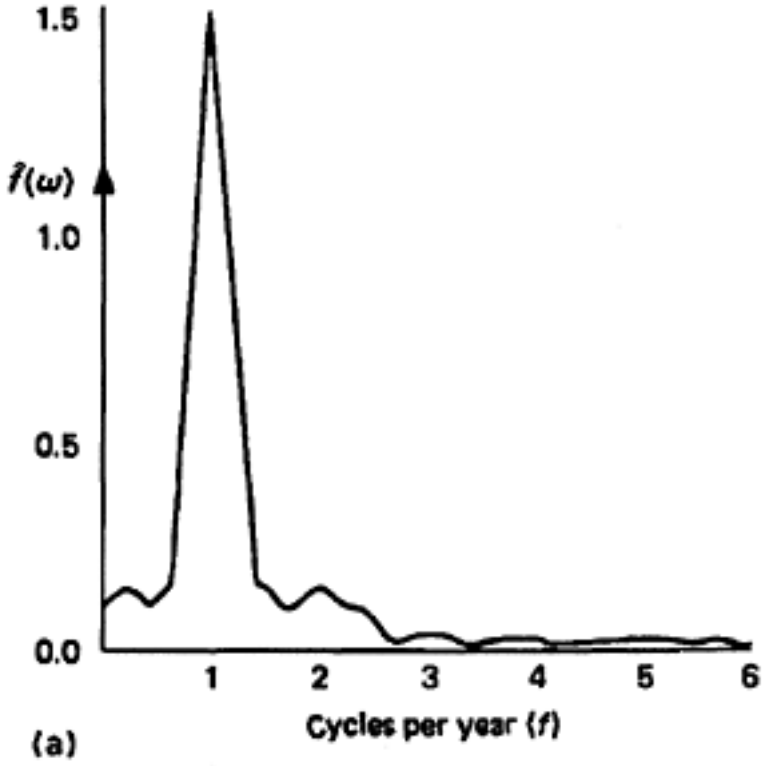
Figure 7.4 Four time series and their spectra. The spectra are given in random order.

< previous page page_148 next page >

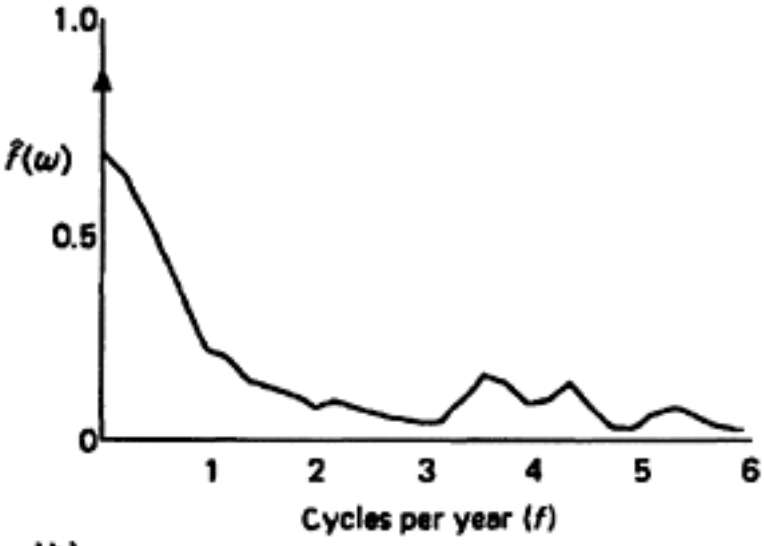
Page 148
The easiest series to assess is trace A, which is much smoother than the other three traces. This means that its spectrum is concentrated at low frequency. Of the four spectra, spectrum M cuts off at the lowest frequency and is largest at zero frequency. This is the spectrum for trace A.
The other three spectra are harder to distinguish. Trace B is somewhat smoother than C or D and corresponds to spectrum K, which 'cuts off' at a lower frequency than J or L. Trace C corresponds to spectrum J, while trace D contains a deterministic sinusoidal component at 15 cycles per second, which contributes 20% of the total power. Thus D corresponds to spectrum L.
From a visual inspection of traces C and D, it is difficult or impossible to decide which goes with spectrum J and which with spectrum L. For this type of data, spectral analysis is invaluable in assessing the frequency properties. The reader may find it surprising that the deterministic component in trace D is so hard to see, but remember that 80% of the power is some sort of noise spread over a wide frequency range and this makes the deterministic component, constituting only 20% of the power, hard to see.
The above example contrasts with the air temperature series at Recife, plotted in Figure 1.2. There the regular seasonal variation is quite obvious from a visual inspection of the time plot, but in this case the deterministic component accounts for about 85% of the total variation. If we nevertheless carry out a spectral analysis of the air temperature series, we get the spectrum shown in Figure 7.5(a) with a large peak at a frequency of one cycle per year. However, it is arguable that the spectral analysis is not really necessary here, as the seasonal effect is so obvious anyway. In fact, if the analyst has a series containing an obvious trend or seasonal component, then it is advisable to remove such variation from the data **before** carrying out a spectral analysis, as any other effects will be relatively small and may not be visible in the spectrum of the raw data.
Figure 7.5(b) shows the spectrum of the Recife air temperature data when the seasonal variation has been removed. The variance is concentrated at low frequencies, indicating either a trend, which is not apparent in Figure 1.2, or short-term correlation as in a first-order AR process with a positive coefficient (cf. Figure 6.4 (a)). The latter seems the more likely explanation here, given that there is no contextual reason to expect a trend in temperature (other than global warming, which is relatively small compared with other effects). As noted earlier, the corresponding periodogram in Figure 7.5(c) shows a graph that oscillates up and down very quickly and is not helpful for interpreting the properties of the data. This demonstrates again that the periodogram has to be smoothed to get a consistent estimate of the underlying spectrum.

< previous page page_148 next page >

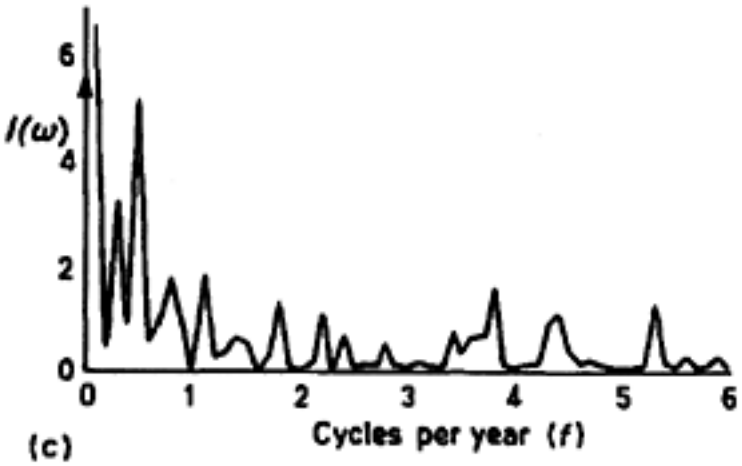
Page 149


1.5
$\hat{f}(\omega)$
1.0
0.5
0.0
1
2
3
4
5
6
Cycles per year (f)

(a)


1.0
$\hat{f}(\omega)$
0.5
0
1
2
3
4
5
6
Cycles per year (f)

(b)


6
$I(\omega)$
4
2
0
0
1
2
3
4
5
6
Cycles per year (f)

(c)

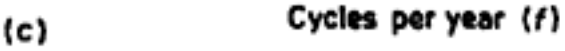


Figure 7.5 Spectra for average monthly air temperature readings at Recife, (a) for the raw data; (b) for the seasonally adjusted data using the Tukey window with M=24; (c) the periodogram of the seasonally adjusted data is shown for comparison.

< previous page | **page_149** | **next page >**

< previous page | page_150 | next page >

Page 150

Removing trend and seasonality is a simple form of the general procedure usually called **prewhitening**. As the name suggests, this aims to construct a series having properties which are closer to those of white noise. In spectral analysis, this is useful because it is easier to estimate the spectrum of a series having a relatively flat spectrum, than one with sharp peaks and troughs. Prewhitening is often carried out by making a linear transformation of the raw data. Then the spectrum of the transformed series can be found, after which the spectrum of the original series can be found, if desired, by using the properties of the linear transformation8 used to carry out the prewhitening. In spectral analysis, this procedure is often limited to removing trend and seasonality, though in other applications more sophisticated model fitting is often used9.

Having estimated the spectrum of a given time series, how do we interpret the results? There are various features to look for. First, are there any peaks in the spectrum and, if so, at what frequency? Can we find a contextual reason for a peak at this frequency? Second, what is the general shape of the spectrum? In particular, does the spectrum get larger as the frequency tends to zero? This often happens with economic variables and indicates a business cycle with a very long period or an underlying long-term non-stationarity in the mean that has not been removed by prior filtering. Economists, who expect to find a clear peak at low frequency, will usually be disappointed, especially if looking for business cycles with a period around 5–7 years. There is usually little evidence of anything so clear-cut. Rather the low-frequency variation is typically spread over a range of frequencies.

The general shape of the spectrum could in principle be helpful in indicating an appropriate parametric model. For example, the shape of the spectrum of various ARMA models could be found and listed in a similar way to that used for specifying ac.f.s for different ARMA models. The use of the correlogram is a standard diagnostic tool in the Box-Jenkins procedure for identifying an appropriate ARIMA process, but the observed spectrum has rarely been used in this way. Why is this? Spectral analysis, as described in this chapter, is essentially a non-parametric procedure in which a finite set of observations is used to estimate a function defined over the whole range from (0, π). The function is not constrained to any particular functional form and so one is effectively trying to estimate more items than in a correlogram analysis, where the analyst may only look at values for a few low lags. Being non-parametric, spectral analysis is in one sense more general than inference based on a particular parametric class of models, but the downside is that it is likely to be less accurate if a parametric model really is appropriate. In my experience, spectral analysis is typically used when there is a suspicion that

8 The frequency response function of the linear transformation is defined later in Chapter 9 and leads directly to the required spectrum.

9 For example, with two time series it is advisable to prewhiten the series by removing as much autocorrelation as possible before calculating quantities called cross-correlations—see Chapter 8.

< previous page | page_150 | next page >

< previous page | page_151 | next page >

cyclic variation may be present at some unknown frequency, and the spectrum shape is rarely used for diagnosing a parametric model.

Spectral analysis is arguably at its most useful for series of the type shown in Figure 7.4, where there is no obvious trend or 'seasonal' variation. Such series arise mostly in the physical sciences. In economics, spectral techniques have perhaps not proved as useful as was first hoped, although there have been a few successes. Attempts have also been made to apply spectral analysis to marketing data, but it can be argued (Chatfield, 1974) that marketing series are usually too short and the seasonal variation too large for spectral analysis to give useful results. In meteorology and oceanography, spectral analysis can be very useful (e.g. Craddock, 1965; Snodgrass et al., 1966) but, even in these sciences, spectral analysis may produce no worthwhile results, other than those that are obvious anyway. It is often the case that, once obvious cyclic effects have been removed (e.g. annual variation from monthly rainfall data; daily variation from hourly temperature data), the spectrum will show no clear peaks, but rather a tendency to get larger as the frequency tends to zero. The spectrum in Figure 7.5(b) is a case in point. The two examples in Percival and Walden (1993, Chapter 6), featuring ocean wave data and ice profile data, yield similar results. Sometimes a small peak is observed but tests usually show that this has dubious significance.

We conclude this section by commenting on some practical aspects of spectral analysis. Most aspects, such as the choice of truncation point, have already been discussed and will be further clarified in Example 7.1 below.

One problem that has not been discussed, is whether to plot the estimated spectrum on a linear or logarithm scale. An advantage of using a logarithmic scale is that the asymptotic variance of the estimated spectrum is then independent of the level of the spectrum, and so confidence intervals for the spectrum are of constant width on a logarithmic scale. For spectra showing large variations in power, a logarithmic scale also makes it possible to show more detail over a wide range. A similar idea is used by engineers when measuring sound in decibels, as the latter take values on a logarithmic scale. Jenkins and Watts (1968, p. 266) suggest that spectrum estimates should always be plotted on a logarithmic scale. However, Anderson (1971, p. 547) points out that this exaggerates the visual effects of variations where the spectrum is small. It may be easier to interpret a spectrum plotted on an arithmetic scale, as the area under the graph corresponds to power and this makes it easier to assess the relative importance of different peaks. Thus, while it is often useful to plot $\hat{f}(\omega)$ on a logarithmic scale in the initial stages of a spectral analysis, especially when trying different truncation points and testing the significance of peaks, it is often better to plot the final version of the estimated spectrum on a linear scale in order to get a clearer interpretation of the final result.

It is also generally easier to interpret a spectrum if the frequency scale is measured in cycles per unit time *(f)* rather than radians per unit time ω. This has been done in Figures 7.4 and 7.5. A linear transformation of frequency

< previous page | page_151 | next page >

Page 152
does not affect the *relative* heights of the spectrum at different frequencies, though it does change the absolute heights by a constant multiple.
Another point worth mentioning is the possible presence in estimated spectra of **harmonics**. As noted earlier, when a series has a strong cyclic component at some frequency ω, then the estimated spectrum may additionally show related peaks at 2ω, 3ω,.... These multiples of the fundamental frequency are called harmonics and generally speaking simply indicate that the main cyclical component is not exactly sinusoidal in character.
Finally, a question that is often asked is how large a value of *N* is required to get a reasonable estimate of the spectrum. It is often recommended that between 100 and 200 observations is the minimum. With smaller values of *N,* only very large peaks can be found. However, if the data are prewhitened to make the spectrum fairly flat, then reasonable estimates may be obtained even with values of *N* around 100, as we have shown in Figure 7.5(b). However, much longer series are to be preferred and are the norm when spectral analysis is contemplated.
Example 7.1 As an example, we analyse part of trace D of Figure 7.4. Although a fairly long trace was available, I decided just to analyse a section lasting for 1 second to illustrate the problems of analysing a fairly short series. This set of data will also illustrate the problems of analysing a continuous trace as opposed to a discrete time series.
The first problem was to digitize the data, and this required the choice of a suitable sampling interval. Inspection of the original trace showed that variation seemed to be 'fairly smooth' over a length of 1 mm, corresponding to 1/100 second, but to ensure that there was no aliasing a sampling interval of 1/200 second was chosen giving *N*=200 observations.
For such a short series, there is little to be gained by using the FFT. I therefore decided to transform the truncated acv.f. using Equation (7.21), with the Tukey window. Several truncation points were tried, and the results for *M*=20, 40 and 80 are shown in Figure 7.6. Equation (7.21) was evaluated at 51 points at $\omega = \pi j/50$, for *j*=0, 1,..., 50, where ω is measured in radians per unit time. Now in this example 'unit time' is 1/200 second and so the values of ω in radians per second are ω=200*πj*/50, for *j*=0, 1,..., 50. If we now convert the frequencies into cycles per second using $f = \omega/2\pi$, we find that the spectrum is evaluated at *f*=2*j,* for *j*=0, 1,..., 50. The Nyquist frequency is given by *f N*=100 cycles per second, which completes one cycle every two observations.

< previous page | page_153 | next page >

Page 153

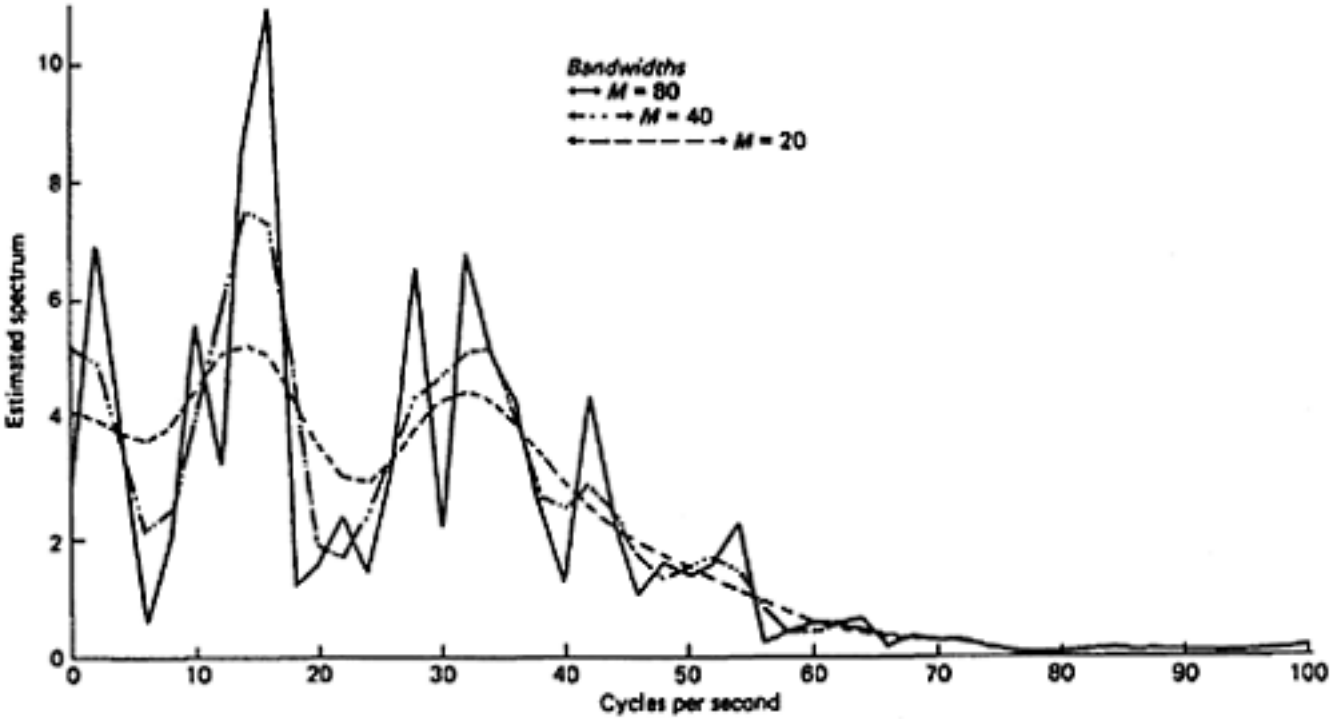


Figure 7.6 Estimated spectra for graph D of Figure 7.4 using the Tukey window with (a) M=80; (b) M=40; (c) M=20.

Looking at Figure 7.6, it can be seen that above about 70 cycles per second, the estimates produced by the three values of M are all very small and cannot be distinguished on the graph. As the estimated spectrum approaches zero as the frequency approaches the Nyquist frequency, it seems clear that no information has been lost by aliasing so that our choice of sampling interval is sufficiently small. Indeed we could have made the sampling interval somewhat larger without losing much information. For lower frequencies, the estimated spectrum is judged rather too smooth with M=20, and much too erratic when M=80. The value M=40 looks about right, although M=30 might be even better. The subjective nature of this choice is clear. Using M=40 or 80, there is a clear peak in the spectrum at about 15 cycles per second (15 Hz). This matches the peak in spectrum L of Figure 7.4. However, Figure 7.6 also reveals a smaller unexpected peak at around 30 cycles per second. This looks like a harmonic of the deterministic sinusoidal component at 15 cycles per second, and may reduce in size if a longer series of observations were to be analysed.

We also estimated the spectrum using a Parzen window with a truncation point of M=56. This value was chosen so that the degrees of freedom of the window, namely, 13.3, were almost the same as for the Tukey window with M=40. The results were so close to those produced by the Tukey window that there was no point in plotting them. The largest difference in the spectrum estimates was 0.33 at 12 cycles per second, but most of the estimates differed only in the second decimal place. Thus the Tukey and Parzen windows give much the same estimates when equivalent values of M are used.

< previous page | page_153 | next page >

Page 154
The reader should note that the bandwidths for different values of *M* are indicated in Figure 7.6. The bandwidth for the Tukey window is 8*π*/3*M* in radians per unit time. As 'unit time' is 1/200 second, the bandwidth is 1600*π*/3*M* in radians per second or 800/3*M* in cycles per second.
Confidence intervals can be calculated as described in Section 7.5. For a sample of only 200 observations, they are disturbingly wide. For example, when *M*=40, the degrees of freedom are 2.67*N*/*M*=13.3. For convenience this is rounded off to the nearest integer, namely, *v*=13. The peak in the estimated spectrum is at 14 cycles per second, where $\hat{f}(\omega)$=7.5. Here the 95% confidence interval is (3.9 to 19.5). Clearly a longer series is desirable to make the confidence intervals acceptably narrow.

Exercises

7.1 *Revision of Fourier series*. Show that the Fourier series, which represents the function

$$f(x) = x^2 \qquad \text{for } -\pi \le x \le \pi$$

is given by

$$f(x) = \frac{\pi^2}{3} - 4\left(\frac{\cos x}{1} - \frac{\cos 2x}{2^2} + \frac{\cos 3x}{3^2} - \cdots\right)$$

7.2 Derive Equations (7.6) and (7.8).
7.3 Derive Parseval's theorem, given by Equation (7.14).
7.4 If *X*1,..., *XN* are independent *N*(*μ, σ*2) variates show that

$$a_p = 2\left[\sum X_t \cos(2\pi pt/N)\right]/N$$

is *N*(0, 2*σ*2/*N*) for *p*=1, 2,..., (*N*/2)−1.
7.5 Derive the lag window for smoothing the periodogram in sets of size m. For algebraic simplicity take *m* odd, with $m = 2m^* + 1$, so that

$$\hat{f}(\omega_p) = \frac{1}{m}\sum_{j=-m^*}^{m^*} I\left(\omega_p + \frac{2\pi j}{N}\right)$$

(Hint: The answer is given in Section 7.4.4. The algebra is rather messy. Use Equation (7.18) and the following two trigonometric results:

$$\cos[2\pi k(p+j)/N] + \cos[2\pi k(p-j)/N] = 2\cos[2\pi kp/N]\cos[2\pi kj/N]$$

$$\sin A - \sin B = 2\sin[(A-B)/2]\cos[(A+B)/2]\,)$$