

Fourier Analysis of Time Series

WILEY SERIES IN PROBABILITY AND STATISTICS
APPLIED PROBABILITY AND STATISTICS SECTION

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

Editors: *Vic Barnett, Noel A. C. Cressie, Nicholas I. Fisher,
Iain M. Johnstone, J. B. Kadane, David G. Kendall, David W. Scott,
Bernard W. Silverman, Adrian F. M. Smith, Jozef L. Teugels;
Ralph A. Bradley, Emeritus, J. Stuart Hunter, Emeritus*

A complete list of the titles in this series appears at the end of this volume.

Fourier Analysis of Time Series An Introduction

Second Edition

PETER BLOOMFIELD

North Carolina State University
Raleigh, North Carolina



A Wiley-Interscience Publication

JOHN WILEY & SONS, INC.

New York / Chichester / Weinheim / Brisbane / Singapore / Toronto

This book is printed on acid-free paper. ☺

Copyright © 2000 by John Wiley & Sons, Inc. All rights reserved.

Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4744. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, (212) 850-6011, fax (212) 850-6008, E-Mail: PERMREQ@WILEY.COM.

For ordering and customer service, call 1-800-CALL WILEY.

Library of Congress Cataloging in Publication Data is available.

ISBN 0-471-88948-2

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

TO DIANA AND ANNALEE

Preface

There are two groups of users of an applied book on time series analysis such as this. The first consists of students (undergraduate or graduate) who encounter a course on time series methods in their study of statistics or its allied fields. The second consists of workers in the many fields in which time series data arise. The fact that this book was written with both groups in mind has imposed noticeable constraints on the contents and presentation, but they have proved entirely beneficial.

In the interests of the second group, the statistical level of the presentation has been kept low. The minimum statistical knowledge needed to follow the essential sections corresponds to a single introductory course in statistics. Greater knowledge of statistics, combined with some experience in the analysis of observational data, would, of course, allow the reader both an easier passage and the opportunity for greater gain along the way.

The interests of students are best served, at least on their first contact with time series, by tying the presentation to examples. All the methods described in this book are introduced in the context of specific sets of data, so that the motivation behind a method is evident as it is developed. The abstract properties of a procedure are discussed only when the motivation has been solidly established.

Many people have difficulty when they first encounter Fourier analysis or the Fourier transform. The discrete Fourier transform is described in

Chapter 4 and is used in one form or another throughout most of the remaining chapters. It is elementary from a mathematical point of view, involving nothing more advanced than the summation of finite series, not even calculus. However, the properties of the discrete transform are analogous to those of other types of Fourier transform. Careful study of Chapter 4 and the expenditure of some time on its exercises will convince the most faint-hearted that Fourier transforms can be fun!

Complex numbers are used extensively in deriving the properties of the discrete Fourier transform, but purely for notational simplicity. To follow the algebra it is necessary only to know the rules for obtaining the sum and product of two complex numbers.

The topics discussed are as follows:

- (i) Harmonic regression: least squares regression on one or more sinusoids (Chapters 2 and 3);
- (ii) Harmonic analysis: the discrete Fourier transform and periodogram analysis (Chapters 4, 5, and 6);
- (iii) Complex demodulation: local harmonic analysis, and the complex time series (Chapter 7); and
- (iv) Spectrum analysis (Chapters 8, 9, and 10).

The order of the discussion is dictated by the increasing complexity of the statistical concepts involved. At all stages of the book, the reader is urged to stop and consider the appropriateness of applying a particular method to the time series being studied. In several cases, some preprocessing is carried out to make the analysis more appropriate. This is all designed to make the point that any data-analytic procedure based on the sine and cosine functions has a better chance of yielding useful conclusions if the data show some kind of oscillation, preferably as uniform or as regular as possible.

Exercises will be found at the ends of many sections. Some are algebraic manipulations designed to make the reader more familiar with the tools of discrete Fourier analysis, and to build the reader's confidence. Others are used to indicate some of the directions in which the theory of time series analysis has revealed useful results. However, the most useful exercise, one that should be omitted by no serious reader, is the analysis of data. All of the time series used as examples in this book are readily obtained, for instance on the Internet. Footnotes give complete directions for retrieving each set of data as of this writing. However, the reader who tries the methods on data arising in his or her own field will gain the added benefit of seeing these data from a new point of view.

All of the analysis of examples was carried out in the S-PLUS environment. Appendices to some chapters discuss particular functions that were used. Other data analysis software such as SAS or MATLAB may be used to carry out essentially similar analyses.

Bookmarks for the data and S-PLUS `dump()`s of the versions used in this edition will be found at <http://www.stat.ncsu.edu/~bloomfld/FATS/>. More extensive samples of code are available at the same site, and other material will be posted there from time to time. This site is also accessible through the Wiley statistics page, <http://www.wiley.com/statistics>.

This second edition differs from the first in various ways:

- The material on harmonic regression has been split into two chapters, the first dealing with the case of known frequencies, and the second dealing with unknown frequencies that need to be estimated.
- The discussion of the fast Fourier transform has been abridged.
- Discussion of complex demodulation has been extended to include construction of the complex analog of an observed time series.
- Throughout, frequencies are expressed in cycles per unit time rather than the less useful radians per unit time, and the definitions of periodograms and spectra have been changed accordingly.
- One time series used in examples has been omitted, and some new ones have been introduced.

I was encouraged to write this book by Geoffrey Watson, who saw clearly the need for an introductory text on Fourier methods not encumbered by an abundance of mathematical, probabilistic, or statistical detail. With sadness, I note his passing in 1998. The time series community lost another of its most eminent members with the death of E. J. Hannan in 1994. Both are missed, as colleagues and as friends.

Many years have gone into the preparation of this second edition, with the encouragement of many and against the advice of a few; all shall remain nameless—you know who you are—I respect all the opinions that have been offered on the subject. Many thanks to all who have offered corrections or other comments—without you, this would not have been possible; all remaining or newly introduced errors are wholly mine.

PETER BLOOMFIELD

Raleigh, North Carolina
November 1999

Contents

1	<i>Introduction</i>	1
1.1	<i>Fourier Analysis</i>	2
1.2	<i>Historical Development of Fourier Methods</i>	5
1.3	<i>Why Use Trigonometric Functions?</i>	7
2	<i>Fitting Sinusoids</i>	9
2.1	<i>Curve-Fitting Approach</i>	9
2.2	<i>Least Squares Fitting of Sinusoids</i>	11
2.3	<i>Multiple Periodicities</i>	17
2.4	<i>Orthogonality of Sinusoids</i>	19
2.5	<i>Effect of Discrete Time: Aliasing</i>	21
2.6	<i>Some Statistical Results</i>	23
	<i>Appendix</i>	24
3	<i>The Search for Periodicity</i>	25
3.1	<i>Fitting the Frequency</i>	25
3.2	<i>Fitting Multiple Frequencies</i>	28
3.3	<i>Some More Statistical Results</i>	30
	<i>Appendix</i>	34

4	Harmonic Analysis	37
4.1	<i>Fourier Frequencies</i>	37
4.2	<i>Discrete Fourier Transform</i>	40
4.3	<i>Decomposing the Sum of Squares</i>	44
4.4	<i>Special Functions</i>	45
4.5	<i>Smooth Functions</i>	53
5	The Fast Fourier Transform	57
5.1	<i>Computational Cost of Fourier Transforms</i>	57
5.2	<i>Two-Factor Case</i>	58
5.3	<i>Application to Harmonic Analysis of Data</i>	61
6	Examples of Harmonic Analysis	63
6.1	<i>Variable Star Data</i>	63
6.2	<i>Leakage Reduction by Data Windows</i>	66
6.3	<i>Tapering the Variable Star Data</i>	72
6.4	<i>Wolf's Sunspot Numbers</i>	76
6.5	<i>Nonsinusoidal Oscillations</i>	78
6.6	<i>Amplitude and Phase Fluctuations</i>	81
6.7	<i>Transformations</i>	83
6.8	<i>Periodogram of a Noise Series</i>	87
6.9	<i>Fisher's Test for Periodicity</i>	91
	<i>Appendix</i>	95
7	Complex Demodulation	97
7.1	<i>Introduction</i>	97
7.2	<i>Smoothing: Linear Filtering</i>	100
7.3	<i>Designing a Filter</i>	105
7.4	<i>Least Squares Filter Design</i>	110
7.5	<i>Demodulating the Sunspot Series</i>	118
7.6	<i>Complex Time Series</i>	124
7.7	<i>Sunspots: The Complex Series</i>	126
	<i>Appendix</i>	130
8	The Spectrum	133
8.1	<i>Periodogram Analysis of Wheat Prices</i>	133
8.2	<i>Analysis of Segments of a Series</i>	140

8.3	<i>Smoothing the Periodogram</i>	142
8.4	<i>Autocovariances and Spectrum Estimates</i>	147
8.5	<i>Alternative Representations</i>	149
8.6	<i>Choice of a Spectral Window</i>	155
8.7	<i>Examples of Smoothing the Periodogram</i>	157
8.8	<i>Reroughing the Spectrum</i>	160
	<i>Appendix</i>	164
9	<i>Some Stationary Time Series Theory</i>	167
9.1	<i>Stationary Time Series</i>	167
9.2	<i>Continuous Spectra</i>	173
9.3	<i>Time Averaging and Ensemble Averaging</i>	175
9.4	<i>Periodogram and Continuous Spectra</i>	176
9.5	<i>Approximate Mean and Variance</i>	177
9.6	<i>Properties of Spectral Windows</i>	190
9.7	<i>Aliasing and the Spectrum</i>	195
10	<i>Analysis of Multiple Series</i>	201
10.1	<i>Cross Periodogram</i>	202
10.2	<i>Estimating the Cross Spectrum</i>	204
10.3	<i>Theoretical Cross Spectrum</i>	211
10.4	<i>Distribution of the Cross Periodogram</i>	214
10.5	<i>Distribution of Estimated Cross Spectra</i>	218
10.6	<i>Alignment</i>	226
	<i>Appendix</i>	230
11	<i>Further Topics</i>	233
11.1	<i>Time Domain Analysis</i>	233
11.2	<i>Spatial Series</i>	234
11.3	<i>Multiple Series</i>	236
11.4	<i>Higher Order Spectra</i>	238
11.5	<i>Nonquadratic Spectrum Estimates</i>	239
11.6	<i>Incomplete and Irregular Data</i>	242
	<i>References</i>	247
	<i>Author Index</i>	255

1

Introduction

A *time series* is a collection of numerical observations arranged in a natural order. Usually each observation is associated with a particular instant or interval of time, and it is this that provides the ordering. The observations could equally well be associated with points along a line, but whenever they are ordered by a single variable it is referred to conventionally as “time.” It will generally be assumed that the time values are equally spaced.

There are many situations in which more than one aspect of a phenomenon is observed at each time point, and this gives rise to *multiple time series*. In other situations each observation is associated with the values of a number of variables (or, in other words, with a point in a space of dimension higher than 1). Such data are called *spatial series*, the most common form being observations associated with points in the plane. Both of these types of data are generalizations of the simple time series, and each requires the appropriate extension of simple time series methods. However, many of the considerations that arise in the analysis of such series arise also in the analysis of simple time series, to which the bulk of this book is devoted.

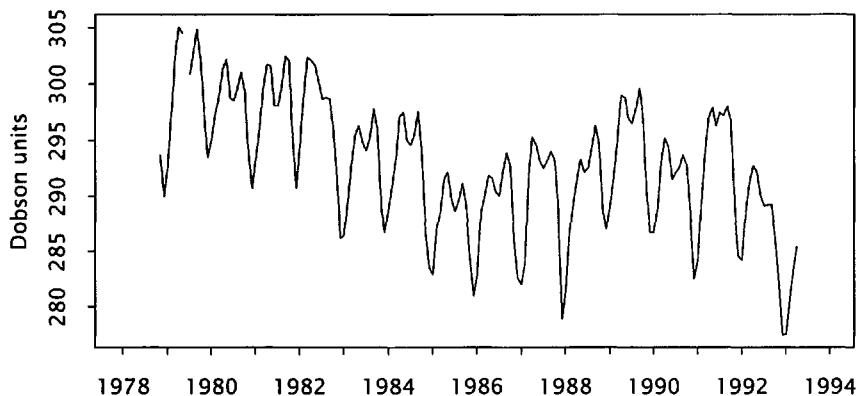


Fig. 1.1 Monthly average total ozone levels, 65° S to 65° N.

1.1 FOURIER ANALYSIS

In its narrowest sense, the *Fourier analysis* or *harmonic analysis* of a time series is a decomposition of the series into a sum of sinusoidal components (the coefficients of which are the *discrete Fourier transform* of the series). However, the term is used in a wider sense to describe any data-analysis procedure that describes or measures the fluctuations in a time series by comparing them with sinusoids.

Figure 1.1 shows a set of data that are used in Chapter 2 to illustrate harmonic regression. They are monthly average global ozone column levels, from November 1978 to April 1993. The measurements were made by the Total Ozone Mapping Spectrometer (TOMS) on the Nimbus-7 satellite (Stolarski et al., 1991). The TOMS instrument produces a global map of ozone levels each day. These were averaged by month from 65° S to 65° N, the part of the globe for which largely complete data are available.¹ The figure shows a clear annual variation superimposed on a small downward trend. Sinusoids may be used to approximate the seasonal variation and by subtracting these components the long term trend is made more visible. Other sources of variability on shorter time scales also become

¹These ozone data were obtained from NASA's TOMS website, <http://jwocky.gsfc.nasa.gov/>. Zonal, hemispheric, and global means for Version 7 of the ozone levels are in ftp://toms.gsfc.nasa.gov/pub/nimbus7/data/zonal_means/ozone/zm_month.n7t.

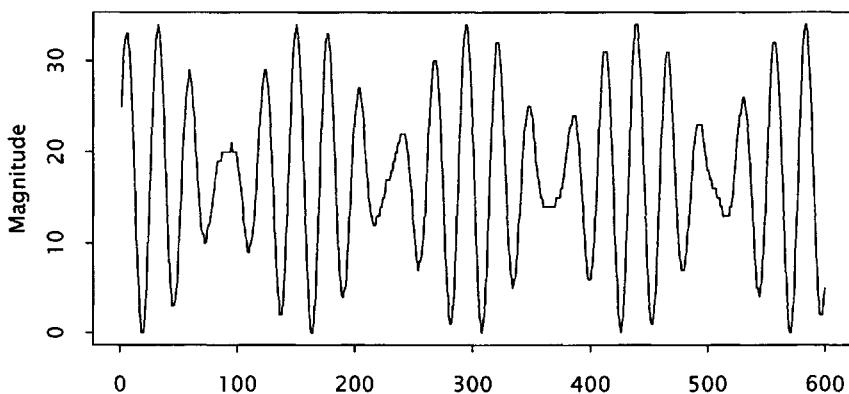


Fig. 1.2 Magnitude of a variable star at midnight on 600 successive nights (from Whittaker and Robinson, 1944, p. 349).

apparent when the strong annual cycle is removed.

Another set of data that is used as an example in Chapters 3 and 6 is shown in Figure 1.2. It represents

the magnitude (*i.e.*, a measure of the brightness) of a variable star at midnight on 600 successive days. (These magnitudes were obtained by reading off from a curve, on which all of the star's brightness were plotted: they have been reduced to a scale suitable for periodogram analysis.)

(From Whittaker and Robinson, 1944, p. 349).² This example differs from the previous one in that here there is no guidance as to what periods might be present. This time series is shown in Chapter 3 to consist approximately of the sum of two sinusoidal components, and in this case Fourier analysis in its narrower sense (with some slight modifications discussed in Chapter 6) provides an accurate and economical description of the data.

²The variable star series is available online at the OzData archive <http://www.maths.uq.oz.au/~gks/data/index.html> maintained at the University of Queensland, Australia. The series is described in the entry <http://www.maths.uq.oz.au/~gks/data/general/star.html> and the data may be retrieved from <http://www.maths.uq.oz.au/~gks/data/general/star.txt>. The series is also contained in the time series archive <http://www.york.ac.uk/depts/mathsls/data/ts/welcome.htm> maintained by the University of York, England, where it will be found in the entry <http://www.york.ac.uk/depts/mathsls/data/ts/ts26.dat>.

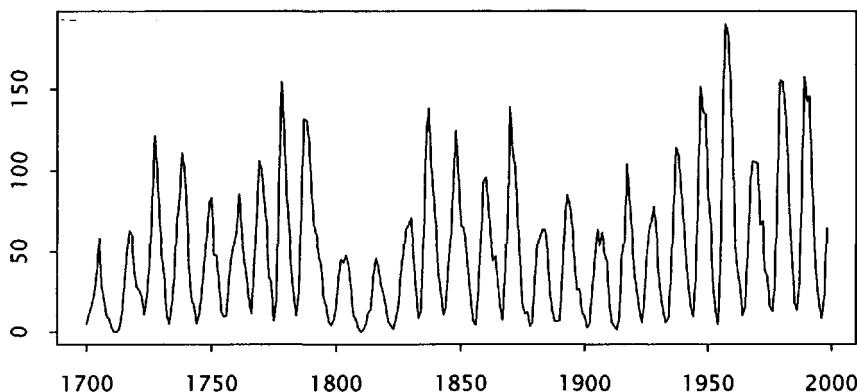


Fig. 1.3 Annual sunspot numbers, 1700 to the present (from the Solar-Terrestrial Physics site at the National Geophysical Data Center).

Figure 1.3 shows a third set of data used as an example, the annual sunspot numbers from 1700 to the present.³ Although these data contain a clear succession of peaks occurring about every 11 years, they are not regular enough to be represented by any one sinusoid. A local form of harmonic analysis known as *complex demodulation* may be used to describe the oscillations in these data.

Figure 1.4 shows the index of wheat prices in Western Europe⁴ presented by Beveridge (1921). Again there is a succession of peaks in the data, but no tendency for them to occur in any regular fashion is evident. Harmonic analysis reveals that there are few if any persistent sinusoidal components in these data. Nevertheless, the oscillations in this series (or a suitably transformed series) may be described in sinusoidal terms by a *spectrum analysis*. This is a method that describes the *tendency* for oscillations of a given frequency to appear in the data, without requiring them to persist throughout.

³Annual sunspot numbers are maintained online at the Solar-Terrestrial Physics site of the U.S. Department of Commerce's National Geophysical Data Center <http://www.ngdc.noaa.gov/stp/stp.html>. The series is in the entry ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA/SUNSPOT_NUMBERS/YEARLY.PLT.

⁴The wheat price index may be found at the University of York's time series archive, in the entry <http://www.york.ac.uk/depts/mathematics/data/ts/ts04.dat>. Beveridge also created a "trend-free" version of the index, which is also in the archive: <http://www.york.ac.uk/depts/mathematics/data/ts/ts05.dat>.

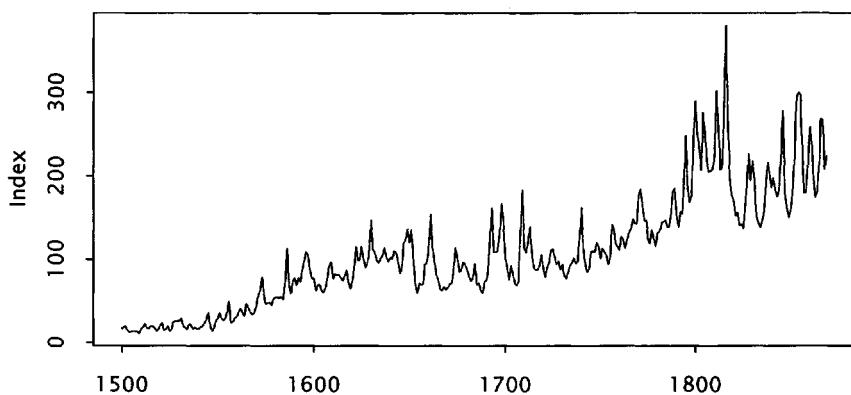


Fig. 1.4 Index of wheat prices in Western Europe, 1500 to 1869 (average from 1700 to 1745 = 100).

All of these methods fall under the heading of Fourier analysis in its wider sense. In the following chapters the methods are described in detail, and applying them to various sets of data shows how they may be used to draw various inferences about the data being analyzed.

1.2 HISTORICAL DEVELOPMENT OF FOURIER METHODS

The simplest periodic data are those consisting of a single cosine wave

$$x_t = R \cos(2\pi ft + \phi),$$

observed with very little error. For example, early Egyptian and Greek astronomers had sufficiently accurate observations of the apparent motion of the sun to know that the length of the year is approximately 365.25 days. Julius Caesar relied on the advice of the Greek astronomer Sosigenes when, in 46 B. C., he enshrined this value in a reformed, or *Julian*, calendar (Richards, 1998).

Data that contain more than one periodic component are more difficult to analyze. In 1772, Lagrange proposed a method based on the use of rational functions to identify such components, and used it to analyze the orbit of a comet (see Lagrange, 1873). Like related methods proposed by Dale (1914a,b) and Prony (Hamming, 1987, Section 39.4), the method is tedious to apply to any but the shortest of series, and it is very sensitive

to errors or other disturbances in the data.

The first procedure to be used at all widely, and the first to be feasible for moderate numbers of data points, was described in 1847 by Buys-Ballot (see, e.g., Whittaker and Robinson, 1944). This is a tabular method and is most easily used to detect a periodic component for which one cycle covers a whole number of observations. A more sophisticated version, described by Stewart and Dodgson (1879), may be used with some computational effort to improve the estimate of an approximately known period. The procedure is discussed in detail by Whittaker and Robinson (1944) in Chapters 10 and 13, and Anderson (1994) on pages 106–112.

The Fourier analysis of a series of numbers may be carried out by a similar tabular technique, described by Schuster (1897). This method was used in the second half of the nineteenth century to find periodic components of known periods in tidal data, meteorological series, and astronomical series. However, the computations involved were repetitive and tedious. Again, the method is easiest if the period of the component covers a whole number of observations. Thomson (1876, 1878) described and built an instrument for carrying out this analysis mechanically and claimed that it would reduce the time needed for an analysis by a factor of 10. Stokes (1879), in a comment on the report of Stewart and Dodgson (1879), pointed out that Thomson's harmonic analyzer could also be used to determine the unknown period of a component in a series. The method is related to complex demodulation.

However, when Fourier analysis is used to search for periodic components of unknown period in empirical data, the results may be very misleading. For instance, Knott (1897) claimed to have discovered components in a series of Japanese earthquakes with periods related to the lunar cycle. Schuster (1897) then showed that their magnitudes were not large enough to be statistically significant. Schuster (1898) provided additional discussion of the Fourier analysis of empirical data, introduced the *periodogram*, and made many penetrating comments on its use. In some further papers (Schuster, 1900, 1906) he applied these ideas to the analysis of various sets of data, including the sunspot series (Figure 1.3). Beveridge (1921, 1922) gave an extensive periodogram analysis of the wheat-price index (Figure 1.4), which at the time constituted a major computational venture.

The probabilistic and statistical theory of time series was developed during the 1920s and 1930s (see Wold, 1954), and the concept of the *spectrum* of a series was introduced. This concept was in fact the subject of a remarkable note by Einstein (1914) (see also Brillinger, 1993; Yaglom, 1987). The spectrum measures the *probable amplitude* of the oscillations in a series as function of frequency, and therefore shifts attention away

from the search for unknown periodicities in a series toward the usually more informative study of the relative amplitudes at *all* frequencies.

In the 1940s and 1950s there was great interest in the problem of *estimating* the spectrum of a series. Daniell (1946) pointed out that a smoothed form of the periodogram is a suitable estimate, as had Einstein (1914). This was followed up by Bartlett (1948) and Kendall (1948), while Hamming and Tukey (1949) proposed a slightly different procedure. There followed a rapid development of the theory and practice of spectrum estimation, to which major contributions were made by Grenander and Rosenblatt (1953, 1957), Parzen (1957a,b), and Blackman and Tukey (1959). The stimulus for this development was the increase in the use of Fourier methods in many fields, principally electrical engineering in its diverse forms, and the parallel increase in the availability of computers to carry out the extensive computations involved.

The next major advance was in the computation of Fourier transforms of data. Cooley and Tukey (1965) described an algorithm that significantly reduces the computational effort involved. The *fast Fourier transform*, as it became known, together with advances in computer technology, has made feasible the routine Fourier analysis of extensive sets of data (see, e.g., Brigham, 1988). The spectrum estimation techniques described in Chapters 8 and 10 are designed to take full advantage of these computational advances.

1.3 WHY USE TRIGONOMETRIC FUNCTIONS?

The essence of Fourier analysis is the representation of a set of data in terms of sinusoidal functions. At this point it seems appropriate to justify the choice of these particular functions, since many other families of periodic functions share at least some of the properties of the sinusoids. Any of these families could be used in a similar way, and in some situations special considerations may make a nonsinusoidal family more suitable. For instance, square waves are used in the Haar transform and the related Haar wavelets, which have found fruitful application in signal processing, especially of images. However, sinusoids have some characteristic properties which give them a distinguished role.

The most basic property of the sinusoids that makes them generally suitable for the analysis of time series is their simple behavior under a change of time scale. A sinusoid of *frequency* f (in cycles per unit time) or *period* $1/f$ (in units of time) may be written as

$$c(t) = R \cos 2\pi(ft + \phi), \quad (1.1)$$

where R is the *amplitude* and ϕ is the *phase*. If the time variable is changed to $u = (t - a)/b$, which incorporates a change of both origin and scale, $c(t)$ becomes

$$\begin{aligned} d(u) &= c(a + bu) \\ &= R \cos 2\pi(fbu + \phi + fa) \\ &= R' \cos 2\pi(f'u + \phi'), \end{aligned}$$

say, where $R' = R$, $f' = fb$, and $\phi' = \phi + fa$. Thus the amplitude is unchanged, the frequency is multiplied by b (the *reciprocal* of the change in the time scale), and the phase is altered by an amount involving the change of time origin and the frequency of the sinusoid. Since the time origin associated with a set of data is often arbitrary, the simplicity of these relationships is useful. In particular, since the amplitude of the sinusoid depends on neither the origin nor the scale of the time variable, it may be regarded as an absolute quantity with no arbitrariness in its definition.

Another interesting property is that a sum of sinusoids with a common frequency is another sinusoid with the same frequency. In fact, since

$$R \cos 2\pi(ft + \phi) = R \cos 2\pi ft \cos 2\pi\phi - R \sin 2\pi ft \sin 2\pi\phi,$$

any sinusoid with frequency f is a linear combination of the two *basis functions* $\cos 2\pi ft$ and $\sin 2\pi ft$, and the converse is also true (see, e.g., Section 2.2).

In interpreting the sinusoid (1.1) it is often helpful to rewrite it as

$$c(t) = R \cos 2\pi f(t - t_0),$$

where $t_0 = -\phi/f$ is a time at which the curve reaches its maximum. The effect of the change of time scale from t to u is now to change the sinusoid to

$$d(u) = R' \cos 2\pi f'(u - u_0),$$

where $R' = R$ and $f' = fb$ as before, and $u_0 = (t_0 - a)/b$ is the time of the same maximum represented on the transformed scale.

A further useful feature of the sinusoids is their behavior under *sampling* (i.e., observing a function of the continuous variable t at an equally spaced set of values t_0, t_1, \dots), for if the sampling interval is Δ , the sinusoids

$$R \cos 2\pi(ft + \phi) \quad \text{and} \quad R \cos 2\pi(f't + \phi)$$

are indistinguishable if $f - f'$ is a multiple of $1/\Delta$. This phenomenon, known as *aliasing*, is discussed further in Section 2.5.

2

Fitting Sinusoids

The simplest use of sinusoids in data analysis is to describe and isolate the periodic part of a series, when the periods are known. This amounts to estimating their amplitudes and phases, although it is easier to approach the problem in slightly different terms.

2.1 CURVE-FITTING APPROACH

As an example, consider the ozone data of Figure 1.1 (p. 2). These data show a pronounced annual cycle, though not of a simple sinusoidal form. Nevertheless, the dominant part of the annual cycle may be expected to be represented in the form

$$s_t = \mu + R \cos 2\pi(ft + \phi) \quad (2.1)$$

where the frequency is $f = 1/12$ cycles per month. The data $\{x_0, x_1, \dots, x_{n-1}\}$ will be modeled as

$$x_t = s_t + e_t,$$

where e_t is the *residual* at time t (i.e., whatever is needed to make the equality exact). The model is regarded as good (and is said to *fit* the data well) if the residuals are generally small. The term μ is an added constant. Since a cosine wave oscillates about 0, while the data oscillate around the value 300, such a term is clearly needed if the residuals are to be at all

small.

The unknown parameters are μ , R , and ϕ ; the next section will describe how to find values for them that make the residuals as small as possible in a certain specific sense. The situation where f is also unknown will be covered in the following chapter.

For the purposes of this chapter, the size of the residuals as a group will be measured by the sum of their squared values, although other choices are possible. Thus the problem is to find μ , R , and ϕ to minimize

$$S(\mu, R, \phi) = \sum_{t=0}^n \{x_t - \mu - R \cos 2\pi(ft + \phi)\}^2,$$

the term between braces being precisely the t th residual for given values of μ , R , and ϕ . This is an example of the method of *least squares*. Least squares methods are widely used and have many computational and theoretical conveniences. However, they also have certain deficiencies, one of which will be mentioned briefly in Section 6.3.

It is easily seen that least squares problems are simplest when the model is a linear function of the unknown parameters, since then the function to be minimized is quadratic in the same quantities. Equation (2.1) is nonlinear in ϕ , but it may be rewritten as

$$s_t = \mu + A \cos 2\pi ft + B \sin 2\pi ft,$$

where $A = R \cos 2\pi\phi$ and $B = -R \sin 2\pi\phi$. Furthermore, given any values of A and B , the corresponding values of R and ϕ may be found. The parameters will therefore be taken to be μ , A , and B , and the model is now linear. The elementary problem of finding μ , A , and B will be solved in the next section.

Exercise 2.1 The Least Squares Straight Line

Suppose that (x_1, y_1) , (x_2, y_2) , \dots , (x_n, y_n) are points in the plane. It is sometimes useful to model such a set of points by a straight line, $y = a + bx$. The *least squares straight line* has parameter values \hat{a} and \hat{b} that minimize the sum of squared residuals,

$$S(a, b) = \sum_{i=1}^n (y_i - a - bx_i)^2.$$

- (i) Verify that, provided the x -values are not all the same,

$$\hat{b} = \frac{\sum_{i=1}^n y_i(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

and

$$\hat{a} = \bar{y} - \hat{b}\bar{x},$$

where $\bar{x} = (x_1 + x_2 + \dots + x_n)/n$ and \bar{y} is similarly defined.

- (ii) Find the corresponding formulas for the coefficients of the least squares parabola, $y = a + bx + cx^2$.

Note that both problems are simplified by writing $x' = x - \bar{x}$, and changing the equations to $y = a' + b'x'$ and $y = a' + b'x' + c'x'^2$, respectively.

2.2 LEAST SQUARES FITTING OF SINUSOIDS

This section covers how to estimate the parameters of a sinusoid, with or without an added constant. The frequency f is regarded as known and is not varied to improve the fit. In the next chapter the method is extended to include the estimation of f . First consider the simple two-parameter model of a sinusoid with no added constant. The model is

$$x_t = A \cos 2\pi ft + B \sin 2\pi ft + e_t,$$

and the principle of least squares leads to the minimization of

$$S(A, B) = \sum_{t=0}^{n-1} (x_t - A \cos 2\pi ft - B \sin 2\pi ft)^2,$$

keeping f fixed. Now

$$\begin{aligned} \frac{\partial T}{\partial A} &= -2 \sum \cos 2\pi ft (x_t - A \cos 2\pi ft - B \sin 2\pi ft), \\ \frac{\partial T}{\partial B} &= -2 \sum \sin 2\pi ft (x_t - A \cos 2\pi ft - B \sin 2\pi ft), \end{aligned}$$

12 FITTING SINUSOIDS

and the equations that result from equating these to zero have the solution

$$\begin{aligned} A = \hat{A} &= \frac{1}{\Delta} \left\{ \sum x_t \cos 2\pi ft \sum (\sin 2\pi ft)^2 \right. \\ &\quad \left. - \sum x_t \sin 2\pi ft \sum \cos 2\pi ft \sin 2\pi ft \right\}, \\ B = \hat{B} &= \frac{1}{\Delta} \left\{ \sum x_t \sin 2\pi ft \sum (\cos 2\pi ft)^2 \right. \\ &\quad \left. - \sum x_t \cos 2\pi ft \sum \cos 2\pi ft \sin 2\pi ft \right\}, \end{aligned} \quad (2.2)$$

where

$$\Delta = \sum (\cos 2\pi ft)^2 \sum (\sin 2\pi ft)^2 - \left(\sum \cos 2\pi ft \sin 2\pi ft \right)^2.$$

The sums involving only trigonometrical functions may be evaluated using the results of Exercise 2.2 to give

$$\begin{aligned} \sum (\cos 2\pi ft)^2 &= \frac{n}{2} \{1 + D_n(2f) \cos 2\pi(n-1)f\}, \\ \sum \cos 2\pi ft \sin 2\pi ft &= \frac{n}{2} D_n(2f) \sin 2\pi(n-1)f, \\ \sum (\sin 2\pi ft)^2 &= \frac{n}{2} \{1 - D_n(2f) \cos 2\pi(n-1)f\}, \end{aligned}$$

and hence

$$\Delta = \frac{n^2}{4} \{1 - D_n(2f)^2\},$$

where

$$D_n(f) = \frac{\sin \pi f n}{n \sin \pi f} \quad (2.3)$$

is a version of the *Dirichlet kernel* (Titchmarsh, 1939, p. 402). The sums involving $\{x_t\}$ usually have to be evaluated directly.

To find R and ϕ , the amplitude and phase, the equations

$$A = R \cos 2\pi\phi \quad \text{and} \quad B = -R \sin 2\pi\phi$$

must be solved. Since R is nonnegative, it follows that $R = \sqrt{A^2 + B^2}$. The basic equation for ϕ is $\tan 2\pi\phi = -B/A$. However, the solution $2\pi\phi = \arctan -B/A$ is incomplete, since this gives the same value for $(-A, -B)$ as for (A, B) . A full solution which gives an answer in the in-

terval $(-1/2, 1/2]$ is

$$2\pi\phi = \begin{cases} \arctan(-B/A), & A > 0, \\ \arctan(-B/A) - \pi, & A < 0, B > 0, \\ \arctan(-B/A) + \pi, & A < 0, B \leq 0, \\ -\pi/2, & A = 0, B > 0, \\ \pi/2, & A = 0, B < 0, \\ \text{arbitrary}, & A = 0, B = 0. \end{cases}$$

(The FORTRAN function ATAN2(-B, A) returns the required value while in S-PLUS it is atan(-b, a). SAS has the function atan, but this is precisely arctan, and the above extension must be used to give the correct answer when $A \leq 0$.)

Since the ozone series oscillates around a positive value, an appropriate model to describe the dominant part of the seasonal oscillation is the three-parameter “sinusoid plus constant” model given in Section 2.1,

$$x_t = \mu + A \cos 2\pi ft + B \sin 2\pi ft + e_t. \quad (2.4)$$

The equations for the least squares estimates of μ , A , and B are

$$\begin{aligned} \sum(x_t - \mu - A \cos 2\pi ft - B \sin 2\pi ft) &= 0, \\ \sum \cos 2\pi ft(x_t - \mu - A \cos 2\pi ft - B \sin 2\pi ft) &= 0, \quad \text{and} \\ \sum \sin 2\pi ft(x_t - \mu - A \cos 2\pi ft - B \sin 2\pi ft) &= 0. \end{aligned} \quad (2.5)$$

These too may be solved explicitly (see Exercise 2.3). For the ozone data with $f = 1/12$ cycles per month, the solution is

$$\begin{aligned} \hat{x}_t &= 293.017 - 4.274 \cos 2\pi ft + 1.16 \sin 2\pi ft \\ &= 293.017 + 4.429 \cos 2\pi(ft - 0.458). \end{aligned}$$

The fitted seasonal curve is shown in Figure 2.1. The residuals, which may be interpreted as deviations from this seasonal curve, are also shown. The deviations exhibit a pronounced semi-annual wave, which suggests that the model (2.4) should be extended to include such a wave. Fitting models with more than one sinusoidal component is discussed in the following section.

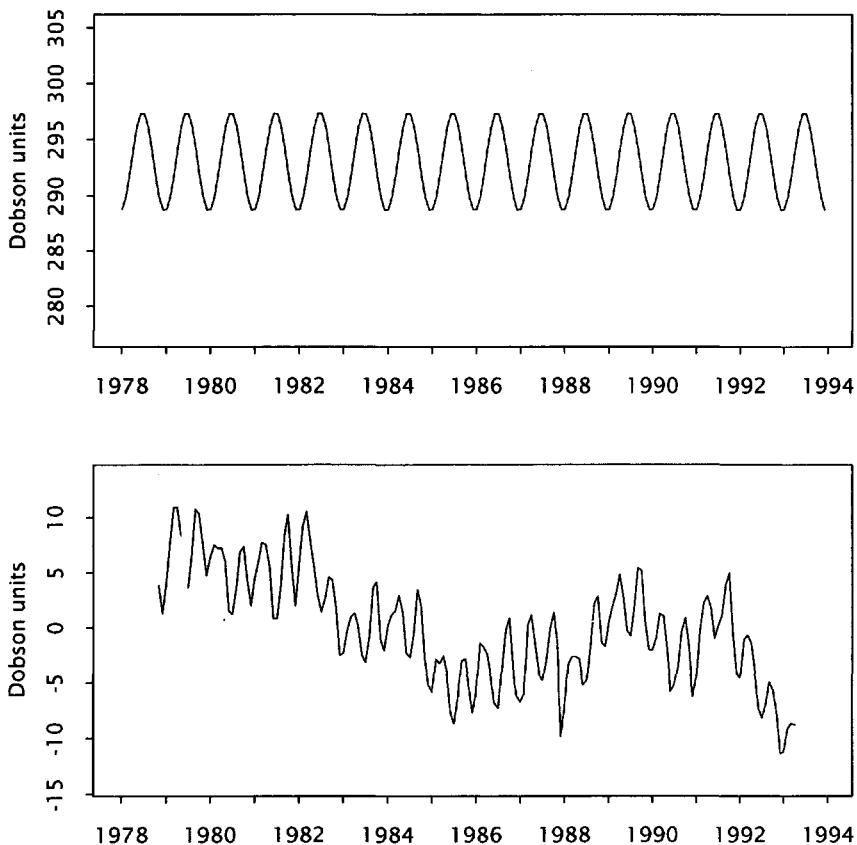


Fig. 2.1 Three-parameter seasonal curve fitted to the ozone series (upper graph), and residuals (lower graph).

Exercise 2.2 Some Trigonometric Identities

(i) Show that

$$\begin{aligned} \sum_{t=0}^{n-1} e^{2\pi ift} &= \frac{e^{2\pi ifn} - 1}{e^{2\pi if} - 1} \\ &= e^{\pi if(n-1)} \frac{e^{\pi ifn} - e^{-\pi ifn}}{e^{\pi if} - e^{-\pi if}}. \end{aligned}$$

(ii) Use the Euler relation

$$e^{ix} = \cos x + i \sin x$$

and its inverse

$$\cos x = \frac{1}{2} (e^{ix} + e^{-ix}), \quad \sin x = \frac{1}{2i} (e^{ix} - e^{-ix})$$

to deduce that

$$\sum_{t=0}^{n-1} \cos 2\pi f t = \cos\{\pi f(n-1)\} \frac{\sin \pi f n}{\sin \pi f} = \cos\{\pi f(n-1)\} n D_n(f),$$

$$\sum_{t=0}^{n-1} \sin 2\pi f t = \sin\{\pi f(n-1)\} \frac{\sin \pi f n}{\sin \pi f} = \sin\{\pi f(n-1)\} n D_n(f),$$

where $D_n(f)$ was defined in (2.3) (strictly, only for $f \neq 0, \pm 1, \dots$; by continuity, $D_n(0) = 1$, etc.). Note that $(n-1)/2 = \bar{t}$ is the average of the time values $0, 1, \dots, n-1$, and that the results may be written

$$\sum_{t=0}^{n-1} \cos 2\pi f t = \cos(2\pi f \bar{t}) n D_n(f),$$

$$\sum_{t=0}^{n-1} \sin 2\pi f t = \sin(2\pi f \bar{t}) n D_n(f).$$

(iii) Use the addition formulas

$$\sin(x+y) = \sin x \cos y + \cos x \sin y,$$

$$\cos(x+y) = \cos x \cos y - \sin x \sin y$$

and their inverses

$$\cos x \cos y = \frac{1}{2} \{ \cos(x+y) + \cos(x-y) \},$$

$$\cos x \sin y = \frac{1}{2} \{ \sin(x+y) + \sin(x-y) \},$$

$$\sin x \sin y = \frac{1}{2} \{ \cos(x-y) - \cos(x+y) \},$$

to evaluate

$$\sum \cos 2\pi f t \cos 2\pi f' t,$$

$$\sum \cos 2\pi f t \sin 2\pi f' t,$$

$$\sum \sin 2\pi f t \sin 2\pi f' t.$$

Note the special forms of the answers when $f = f'$.

16 FITTING SINUSOIDS

Exercise 2.3 The Three-parameter “Sinusoid Plus Constant” Model

The derivatives of

$$S(\mu, A, B) = \sum_{t=0}^{n-1} (x_t - \mu - A \cos 2\pi ft - B \sin 2\pi ft)^2$$

with respect to μ , A , and B are

$$\frac{\partial S}{\partial \mu} = -2 \sum_{t=0}^{n-1} (x_t - \mu - A \cos 2\pi ft - B \sin 2\pi ft),$$

$$\frac{\partial S}{\partial A} = -2 \sum_{t=0}^{n-1} \cos 2\pi ft (x_t - \mu - A \cos 2\pi ft - B \sin 2\pi ft), \text{ and}$$

$$\frac{\partial S}{\partial B} = -2 \sum_{t=0}^{n-1} \sin 2\pi ft (x_t - \mu - A \cos 2\pi ft - B \sin 2\pi ft),$$

respectively. Simplify these expressions using the results of Exercise 2.2, and solve them for the least squares estimates $\hat{\mu}$, \hat{A} , and \hat{B} of μ , A , and B , respectively.

Exercise 2.4 The Sum of Squares Associated with a Frequency

- (i) For the two-parameter model, show that the residual sum of squares $S(\hat{A}, \hat{B})$ may be written as

$$\begin{aligned} S(\hat{A}, \hat{B}) &= \sum x_t^2 - \frac{1}{\Delta} \left\{ \left(\sum x_t \cos 2\pi ft \right)^2 \sum (\sin 2\pi ft)^2 \right. \\ &\quad - 2 \sum x_t \cos 2\pi ft \sum x_t \sin 2\pi ft \sum \cos 2\pi ft \sin 2\pi ft \\ &\quad \left. + \left(\sum x_t \sin 2\pi ft \right)^2 \sum (\cos 2\pi ft)^2 \right\}, \end{aligned}$$

and use the results of Exercise 2.2 to evaluate the purely trigonometric sums.

NOTE: This equation may be interpreted as

sum of squares of residuals

= sum of squares of original data

- sum of squares associated with frequency f .

- (ii) Find the corresponding expression for the sum of squares associated

with f in the three-parameter model, defined by the equation

sum of squares of residuals

- = sum of squares of original data
- sum of squares associated with μ
- sum of squares associated with frequency f ,

where

$$\text{sum of squares associated with } \mu = n\bar{x}^2.$$

2.3 MULTIPLE PERIODICITIES

Figure 2.1 made it clear that the seasonal behavior in the ozone series could not be described by a single sinusoid with the annual frequency (1/12 cycles per month). If a semi-annual wave is added to the model (2.4), the resulting five-parameter model is

$$\begin{aligned}x_t = & \mu + A_1 \cos 2\pi ft + B_1 \sin 2\pi ft \\& + A_2 \cos 4\pi ft + B_2 \sin 4\pi ft + e_t,\end{aligned}$$

which may be fitted using an extension of the equations (2.5) (see Exercise 2.5). The fitted curve is

$$\begin{aligned}\hat{x}_t = & 293 - 4.205 \cos 2\pi ft + 1.075 \sin 2\pi ft \\& - 2.614 \cos 4\pi ft + 1.477 \sin 4\pi ft \\= & 293 + 4.34 \cos 2\pi(ft - 0.46) \\& + 3.003 \cos 2\pi(2ft - 0.418).\end{aligned}$$

The fitted seasonal curve, with both components, and the corresponding residuals are shown in Figure 2.2. The new deviations show no obvious periodicity, and subtracting out the seasonal part of the original data has allowed the other variations, notably the downward trend but also some variations on a two- to three-year time scale, to be seen more clearly.

Exercise 2.5 Fitting Two Sinusoids

To fit the general model with two sinusoids,

$$\begin{aligned}x_t = & \mu + A_1 \cos 2\pi f_1 t + B_1 \sin 2\pi f_1 t \\& + A_2 \cos 2\pi f_2 t + B_2 \sin 4\pi f_2 t + e_t,\end{aligned}$$

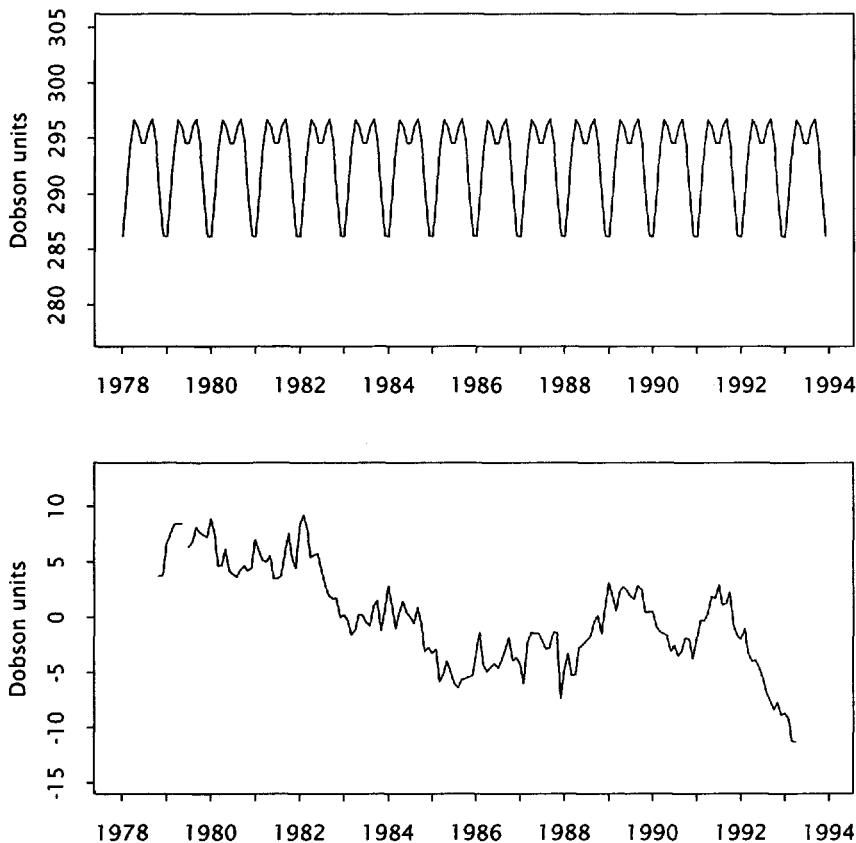


Fig. 2.2 Five-parameter seasonal curve fitted to the ozone series (upper graph), and residuals (lower graph).

the sum of squares function

$$S(\mu, A_1, B_1, A_2, B_2) = \sum_{t=0}^{n-1} (x_t - \mu - A_1 \cos 2\pi f_1 t - B_1 \sin 2\pi f_1 t - A_2 \cos 2\pi f_2 t - B_2 \sin 2\pi f_2 t)^2$$

must be minimized. Its derivatives are

$$\begin{aligned} \frac{\partial S}{\partial \mu} = & -2 \sum_{t=0}^{n-1} (x_t - \mu - A_1 \cos 2\pi f_1 t - B_1 \sin 2\pi f_1 t \\ & - A_2 \cos 2\pi f_2 t - B_2 \sin 2\pi f_2 t), \end{aligned}$$

$$\begin{aligned}\frac{\partial S}{\partial A_j} = -2 \sum_{t=0}^{n-1} & \cos 2\pi f_j t (x_t - \mu - A_1 \cos 2\pi f_1 t - B_1 \sin 2\pi f_1 t \\ & - A_2 \cos 2\pi f_2 t - B_2 \sin 2\pi f_2 t), \quad j = 1, 2\end{aligned}$$

$$\begin{aligned}\frac{\partial S}{\partial B_j} = -2 \sum_{t=0}^{n-1} & \sin 2\pi f_j t (x_t - \mu - A_1 \cos 2\pi f_1 t - B_1 \sin 2\pi f_1 t \\ & - A_2 \cos 2\pi f_2 t - B_2 \sin 2\pi f_2 t), \quad j = 1, 2.\end{aligned}$$

Simplify these expressions using the results of Exercise 2.2, and specialize them to the case $f_2 = 2f_1$.

2.4 ORTHOGONALITY OF SINUSOIDS

Comparing the coefficients in the three- and five-parameter models fitted to the ozone series reveals an interesting detail: the constant term and the coefficients of the annual wave changed when the semi-annual wave was added to the model, but by very little, even though the semi-annual term itself was comparable in size with the annual term. In general, coefficients in a fitted equation do not change only when the newly introduced variables are *orthogonal* to those in the existing equation. This suggests that the cosine and sine at the semi-annual frequency are not orthogonal to the constant term or the annual terms, but are nearly so. This near orthogonality is explored in this section, and a condition is found under which it becomes exact orthogonality.

The clue lies in the coefficients of the equations derived in Exercises 2.3 and 2.5, which define the respective fitted coefficients. These equations may be approximated by

$$\begin{aligned}n\hat{\mu} &= \sum x_t, \\ \frac{n}{2}\hat{A} &= \sum x_t \cos 2\pi ft, \\ \frac{n}{2}\hat{B} &= \sum x_t \cos 2\pi ft\end{aligned}$$

and

$$\left. \begin{aligned}n\hat{\mu} &= \sum x_t, \\ \frac{n}{2}\hat{A}_j &= \sum x_t \cos 2\pi f_j t, \\ \frac{n}{2}\hat{B}_j &= \sum x_t \cos 2\pi f_j t,\end{aligned} \right\} \quad j = 1, 2,$$

respectively, where the coefficients of the omitted terms include one of the form $nD_n(f_1 - f_2) \cos 2\pi(f_1 - f_2)\bar{t}$, and the others are similar. Now

$$|nD_n(f_1 - f_2) \cos 2\pi(f_1 - f_2)\bar{t}| \leq \frac{1}{|\sin \pi(f_1 - f_2)|},$$

and hence these terms are all small compared with the terms that were retained, which are all n or $n/2$, provided that none of f_1 , f_2 , and $f_1 \pm f_2$ is close to 0 or 1. However, it is clear that the solutions for $\hat{\mu}$, \hat{A} , and \hat{B} are the same for both models in this approximate form, and this explains the near equality seen in the exact solutions.

The omitted terms do in fact vanish exactly if f_1 and f_2 are of the form j_1/n and j_2/n for integers j_1 and j_2 , since in this case

$$nD_n(f_1 - f_2) \cos 2\pi(f_1 - f_2)\bar{t} = \frac{\sin \pi(j_1 - j_2) \cos 2\pi(j_1 - j_2)\bar{t}/n}{\sin \pi(j_1 - j_2)/n} = 0,$$

and the other terms also have the sine of an integer multiple of π in the numerator. In this case, all the sine and cosine terms are orthogonal to one another, and to the constant term whose coefficient is μ . The solutions for the three-parameter model become

$$\begin{aligned}\hat{\mu} &= \bar{x} = \frac{1}{n} \sum x_t, \\ \hat{A} &= \frac{2}{n} \sum x_t \cos 2\pi ft, \\ \hat{B} &= \frac{2}{n} \sum x_t \sin 2\pi ft,\end{aligned}$$

and the sum of squares associated with f is just

$$\frac{2}{n}(\hat{A}^2 + \hat{B}^2)$$

(see Exercise 2.4, p. 16).

Frequencies that are integer multiples of $1/n$ are distinguished by the fact that the corresponding periods n/j are repeated a whole number of times in the span of the data. They are said to be *harmonic* with respect to the span of the data, and are known as the *Fourier frequencies*. The annual and semi-annual frequencies fail to be harmonic for the ozone data because the data available for analysis cover some years and a fraction. Had they covered a whole number of years, the annual and semi-annual frequencies would have been harmonic, and the near constancy noted at the start of this section would have been exact.

2.5 EFFECT OF DISCRETE TIME: ALIASING

Thus far, no restrictions have been discussed that might need to be imposed on the frequency f of any sinusoid being fitted to the data. Since the units of frequency are cycles per unit time, it is natural to require that frequency be nonnegative. This may be justified by arguing that, since $\cos(-x) = \cos x$ and $\sin(-x) = -\sin x$, any cosine wave with negative frequency $-f$ may be written as

$$A \cos 2\pi(-f)t + B \sin 2\pi(-f)t = A \cos 2\pi ft + (-B) \sin 2\pi ft,$$

a cosine wave with a positive frequency. Thus the frequencies f and $-f$ are indistinguishable; they are said to be *aliases* of each other.

The equal spacing in time of the observations introduces a further aliasing. Suppose that the *sampling interval* is Δ , so that the t th observation is made at time $t\Delta$. If the data consist of a pure cosine wave at frequency f (for the sake of argument, with unit amplitude and zero phase), the t th observation will be

$$x_t = \cos 2\pi ft\Delta.$$

If f is increased from zero, this wave oscillates more and more rapidly until at $f = 1/2\Delta$ it is

$$x_t = \cos \pi t = (-1)^t,$$

which is clearly the most rapid oscillation that may be observed. Suppose that f is increased further, say, to a value satisfying $1/2\Delta < f < 1/\Delta$. Let $f' = 1/\Delta - f$. Then

$$\begin{aligned} x_t &= \cos 2\pi ft\Delta \\ &= \cos 2\pi(1/\Delta - f')t\Delta \\ &= \cos 2\pi(t - f't\Delta) \\ &= \cos 2\pi f't\Delta. \end{aligned}$$

In the same way $\sin 2\pi ft\Delta = -\sin 2\pi f't\Delta$. Thus the frequencies f and f' are also indistinguishable and hence are aliases of each other. The argument may be extended to any positive frequency, no matter how large.

The result is that every frequency not in the interval $0 \leq f \leq 1/2\Delta$ has an alias in that interval, termed its *principal alias*. To avoid indeterminacy, frequencies will be restricted to this range. Figure 2.3 shows a number of frequencies with the same principal alias. The frequency $1/2\Delta$ is known as the *Nyquist frequency*. It is also called the *folding frequency*, since higher frequencies are effectively folded down into the interval $[0, 1/2\Delta]$.

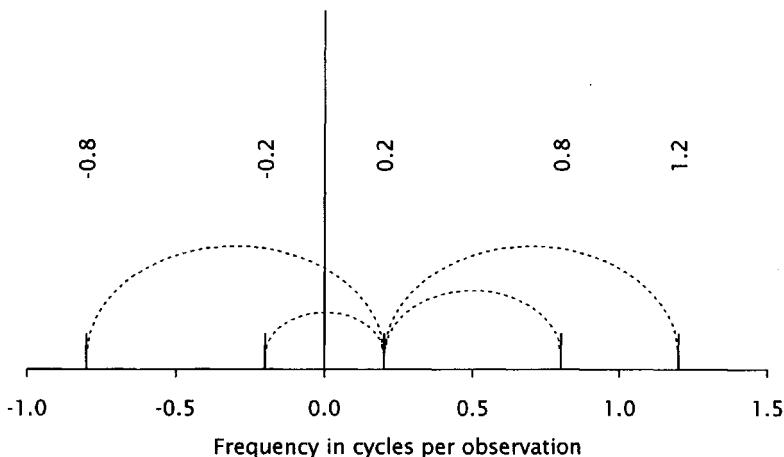


Fig. 2.3 Some frequencies with the same principal alias, $f = 0.2$ cycles per observation $= 0.2/\Delta$ cycles per unit time.

Since the sampling interval is Δ , the *sampling frequency* is $1/\Delta$ observations per unit time. Thus the Nyquist frequency is one-half the sampling frequency; in other words, there are two samples per cycle of the Nyquist frequency, the highest frequency that may be observed.

The phenomenon of aliasing is important not only in the choice of frequencies to be fitted to data; it must also be borne in mind when designing a scheme to *observe* a time series. Suppose that $x(u)$ is a function of the continuous time parameter u , and that it is desired to sample $x(u)$ to obtain information about frequencies in some interval, say (f_0, f_1) . Then the Nyquist frequency should usually be chosen to be larger than f_1 so that all such frequencies are directly observable. However, if $x(u)$ contains oscillations with frequencies greater than f_1 , the sampling frequency should be chosen so that these are not aliased into the interval of interest. In fact, it is preferable when possible to remove these frequencies from the function before sampling, so that this problem cannot arise.

It should be noted that aliasing is a relatively simple phenomenon. In general, when one takes a discrete sequence of observations on a continuous function, information is lost. It is an advantage of the trigonometric functions that this loss of information is manifest in the easily understood form of aliasing.

In the chapters to follow the sampling interval will typically be adopted as the unit of time. Then $\Delta = 1$, and the Nyquist frequency is simply $1/2$.

Except where otherwise stated, this convention will be followed tacitly.

2.6 SOME STATISTICAL RESULTS

This chapter has described how to obtain estimates of the coefficients of one or more sinusoidal components in a series. With the added assumption that the errors in the series are statistical or random in nature, the accuracy of those estimates may be described.

Suppose that the data x_0, x_1, \dots, x_{n-1} were generated by the model

$$x_t = \mu + A \cos 2\pi ft + B \sin 2\pi ft + a_t,$$

where $\{a_t\}$ are random errors or disturbances, and satisfy

$$\begin{aligned} E(a_t) &= 0, \\ E(a_t a_{t'}) &= \begin{cases} \sigma^2, & t = t' \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

The assumption that the errors at different times are uncorrelated is a nontrivial restriction, often violated in practice. In the case of the ozone series, there is clearly more in the residuals shown in Figure 2.2 than uncorrelated errors. A more realistic assumption will be discussed in Chapter 9.

Statistical statements about the estimates discussed above may be obtained, based on these statistical assumptions. As least squares estimates, $\hat{\mu}$, \hat{A} , and \hat{B} are all unbiased. Their variances and covariances are, in general, complicated (see Exercise 2.6), but for the Fourier frequencies they simplify to

$$\begin{aligned} \text{var}(\hat{\mu}) &= \frac{\sigma^2}{n}, \\ \text{var}(\hat{A}) &= \text{var}(\hat{B}) = \frac{2\sigma^2}{n}, \\ \text{cov}(\hat{\mu}, \hat{A}) &= \text{cov}(\hat{\mu}, \hat{B}) = \text{cov}(\hat{A}, \hat{B}) = 0. \end{aligned}$$

If the errors a_0, a_1, \dots, a_{n-1} are additionally assumed to be independent, then by the central limit theorem (see, e.g., Feller, 1968, pp. 244 and 254) $\hat{\mu}$, \hat{A} , and \hat{B} , as linear functions of the a_t s, would be expected to be approximately normally distributed with the stated means and variances. This may be verified by showing that the sequences of coefficients in these linear functions satisfy the relevant requirements.

Exercise 2.6 Means, Variances, and Covariances of the Estimates

- (i) For the two-parameter model with no constant term, least squares estimates of A and B are given by (2.2). Use these equations to write \hat{A} and \hat{B} as A and B plus corresponding linear functions of the a_t s, respectively. Verify the unbiasedness of \hat{A} and \hat{B} , and derive their variances and covariances. Note the simplification that results when f is a Fourier frequency.
- (ii) Extend the results to the three-parameter model in which the constant term μ is added to the equation.

Appendix

The coefficients in the fitted equations mentioned above were computed using general-purpose least squares fitting programs, rather than by solving the equations. Typical code using S-PLUS (see, e.g., Spector, 1994), in this instance for the ozone time series, is:

```
a <- (2 * pi * time(toms))
x <- cbind(cos = cos(a), sin = sin(a), cos2 = cos(2 * a),
           sin2 = sin(2 * a))
ls.print(lsfit(x, toms))
```

Here `toms` is an S-PLUS time series object containing the ozone observations together with the information that the series contains 12 values per year and that the first value is for month 11 in 1978. The function `time()` uses this information, its value being a vector of the same length as the time series with entries that are the times, in years, associated with the values of the series. These values begin at $1978 + 10/12$ and increment by $1/12$. The second statement creates a matrix `x`, whose columns are the required cosines and sines, and the third statement carries out the fitting and prints the results, here for the five-parameter model:

Residual Standard Error = 4.5272, Multiple R-Square = 0.417
 $N = 173$, F-statistic = 30.0358 on 4 and 168 df, p-value = 0

	coef	std.err	t.stat	p.value
Intercept	293.0001	0.3445	850.6130	0.0000
cos	-4.2052	0.4880	-8.6180	0.0000
sin	1.0748	0.4864	2.2099	0.0285
cos2	-2.6142	0.4862	-5.3764	0.0000
sin2	1.4769	0.4878	3.0279	0.0029

3

The Search for Periodicity

The problem of fitting sinusoids of known frequency to a time series was discussed in Chapter 2. There are many situations where this can be a useful step in data analysis, typically involving the identification and study of annual, weekly, or diurnal variations. In some series, however, there may be clear oscillations of fixed but unknown frequency, such as those in the variable star series shown in Figure 1.2 (p. 3). A statistical approach to the analysis of such series is to view the frequency as an additional unknown, and estimate it together with the other parameters of the oscillation. Estimating frequency is however a more difficult problem than estimating amplitude and phase, as the model is inherently nonlinear in f .

3.1 FITTING THE FREQUENCY

The 600 days of variable star data (Figure 1.2, p. 3) show 21 peaks, and this suggests that any periodicity should have a period of around $600/21 \approx 28.6$ days. Thus the t th data value should contain a component of the form

$$A \cos 2\pi f t + B \sin 2\pi f t$$

with

$$f \approx 1/28.6 = 0.035 \text{cycles per day.}$$

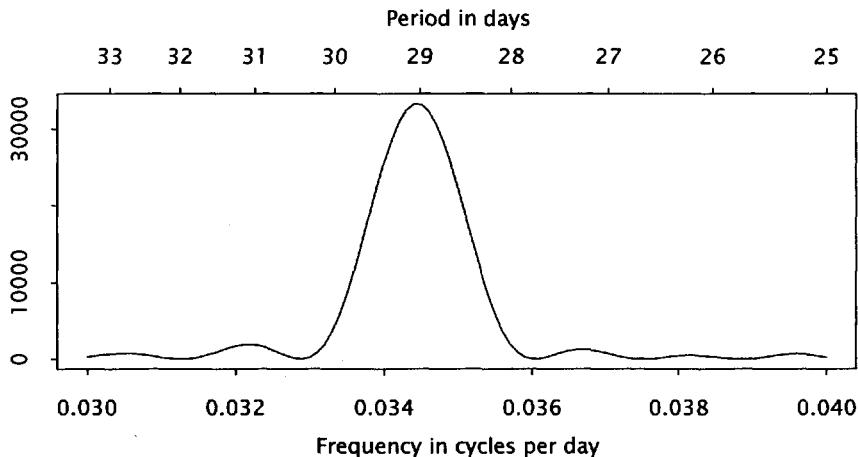


Fig. 3.1 The sum of squares of the variable star data associated with f , $0.03 \leq f \leq 0.04$ cycles per day.

The problem of fitting the parameters A and B by least squares, for fixed f , was covered in the previous chapter. The method of least squares may be extended to estimating f by minimizing

$$S(\mu, A, B, f) = \sum_{t=0}^{n-1} (x_t - \mu - A \cos 2\pi f t - B \sin 2\pi f t)^2.$$

The minimum for fixed f is $T(f) = S(\hat{\mu}(f), \hat{A}(f), \hat{B}(f), f)$, where μ , A , and B have been replaced by the estimates derived in the previous chapter. The frequency f is now estimated by the value that minimizes $T(f)$, or, equivalently, that maximizes the sum of squares associated with f (see Exercise 2.4, p. 16). The latter is graphed for f close to 0.035 cycles per day in Figure 3.1. The frequency at which the function is maximized, and the fitted equation, are

$$\begin{aligned}\hat{f} &= 0.034442 = 1/29.0343 \\ \hat{x}_t &= 17.084 + 7.024 \cos 2\pi f t + 7.82 \sin 2\pi f t \\ &= 17.084 + 10.511 \cos 2\pi(f t - 0.134).\end{aligned}$$

The fitted model and residuals are shown in Figure 3.2.

The graph of the sum of squares associated with f shows subsidiary peaks, or *sidelobes*, occurring on either side of the main peak and separated from it by troughs in which the value is indistinguishable from zero. It will be seen in Chapter 4 that this is typical of such graphs. Small

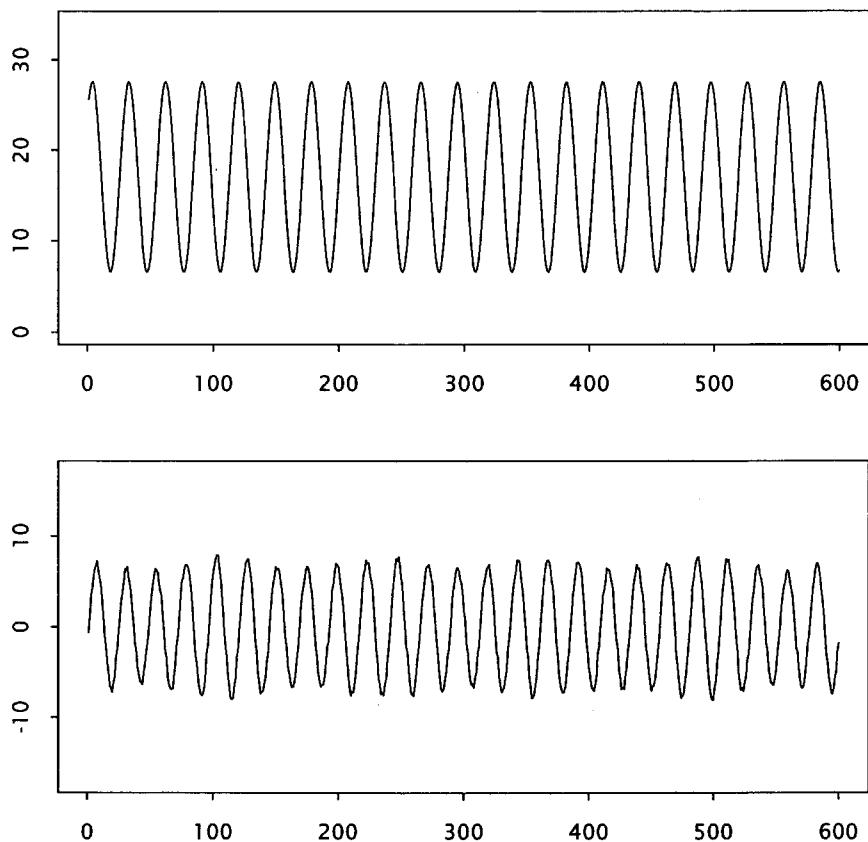


Fig. 3.2 Fitted sinusoid with period 29.034 days (upper curve), and residuals (lower curve).

sidelobes such as these do not indicate the presence of other periodic components.

The location of the maximum of the function was found numerically (see the Appendix to this chapter). The derivative with respect to f is highly nonlinear and has many zeros; indeed, Figure 3.1 shows that there are eleven zeros in the interval $0.03 \leq f \leq 0.04$ cycles per day. This makes analytic solution impossible and also renders methods based on the gradient treacherous. For instance, Newton's method could easily lead to one of the other stationary points, either a local maximum or, worse, a local minimum.

The residuals from the current fit, shown in the lower panel of Figure 3.2, have a very pronounced periodicity with a period of around 24 days, or a frequency of approximately 0.042 cycles per day. Thus the original data must have contained at least two periodic components. The estimation of a number of frequencies is described in the next section; in particular, it will be shown that the presence of a second periodic component, especially one with a similar frequency, can noticeably distort the estimates of frequency and of amplitude and phase.

3.2 FITTING MULTIPLE FREQUENCIES

It emerged at the end of the preceding section that the data being used as an example actually contain more than one periodic component. The fitting of a number of components is described in this section, using the same variable star series as an example.

The simplest procedure would be to repeat the analysis of Section 3.1 but searching now for a maximum near $f = 0.042$ cycles per day. The best frequency and the fitted equation are

$$\begin{aligned}\hat{f} &= 0.041724 = 1/23.967 \\ \hat{x}_t &= 17.115 - 2.847 \cos 2\pi \hat{f}t + 7.234 \sin 2\pi \hat{f}t \\ &= 17.115 + 7.774 \cos 2\pi(\hat{f}t - 0.31).\end{aligned}$$

However, these estimates were obtained by the least squares fitting of the model

$$x_t = \mu + A_2 \cos 2\pi f_2 t + B_2 \sin 2\pi f_2 t + e_t,$$

where e_t is the t th residual. The idea behind a least squares method is to make these residuals as small as possible. In the present case the residual term necessarily contains the strong periodic component found in the preceding section, and it does not make sense to try to make this small. A better approach is to include the second component in the model, as in Section 2.3:

$$\begin{aligned}x_t &= \mu + A_1 \cos 2\pi f_1 t + B_1 \sin 2\pi f_1 t \\ &\quad + A_2 \cos 2\pi f_2 t + B_2 \sin 2\pi f_2 t + e_t.\end{aligned}$$

The sum of squares function to be minimized is

$$\begin{aligned}S(\mu, A_1, B_1, A_2, B_2, f_1, f_2) &= \sum_{t=0}^{n-1} (x_t - \mu - A_1 \cos 2\pi f_1 t - B_1 \sin 2\pi f_1 t \\ &\quad - A_2 \cos 2\pi f_2 t - B_2 \sin 2\pi f_2 t)^2.\end{aligned}$$

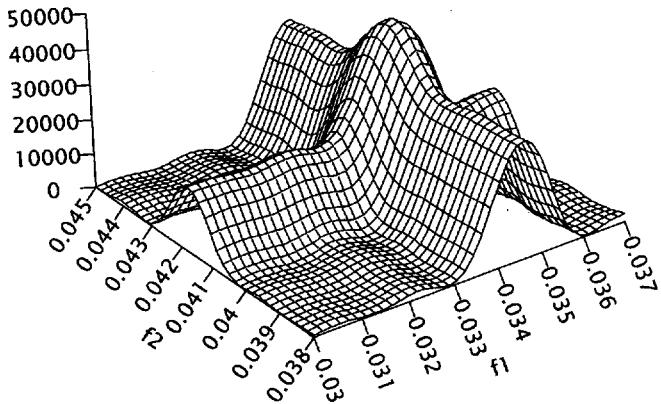


Fig. 3.3 The sum of squares associated with f_1 and f_2 , $0.030 \leq f_1 \leq 0.037$, $0.038 \leq f_2 \leq 0.045$ cycles per day.

The most natural extension of the method used in Section 3.1 to find a single frequency is as follows: First note that for fixed f_1 and f_2 the model is linear in the remaining parameters. Hence the conditionally best values of these may be found by conventional methods and substituted in the function S . This gives a new function

$$T(f_1, f_2) = S(\hat{\mu}, \hat{A}_1, \hat{B}_1, \hat{A}_2, \hat{B}_2, f_1, f_2),$$

where $\hat{\mu}$, \hat{A}_1 , \hat{B}_1 , \hat{A}_2 , and \hat{B}_2 are all functions of f_1 and f_2 . The function $T(f_1, f_2)$ may then be minimized numerically. An equivalent optimization problem is to maximize the sum of squares associated with the pair of frequencies (f_1, f_2) ,

$$U(f_1, f_2) = \sum_{t=0}^{n-1} (x_t - \bar{x})^2 - T(f_1, f_2).$$

Figure 3.3 shows $U(f_1, f_2)$ for (f_1, f_2) in the square $0.030 \leq f_1 \leq 0.037$, $0.038 \leq f_2 \leq 0.045$ cycles per day. The graph shows two orthogonal ridges, intersecting in a peak. The higher ridge is associated with $f_1 \approx 0.034$ cycles per day, and the lower ridge with $f_2 \approx 0.042$ cycles per day. Note that by analogy with the functions examined in the preceding section, $T(f_1, f_2)$ may be expected to have many stationary points, and therefore to require some care in finding its minimum. Indeed, the figure has 15 local maxima in the set of frequencies graphed, and an additional 12 local minima.

The least squares estimates and the fitted equation are

$$\begin{aligned}\hat{f}_1 &= 0.034482 = 1/29.0003 \\ \hat{f}_2 &= 0.041666 = 1/24.0001 \\ \hat{x}_t &= 17.086 + 6.074 \cos 2\pi \hat{f}_1 t + 7.983 \sin 2\pi \hat{f}_1 t \\ &\quad - 1.833 \cos 2\pi \hat{f}_2 t + 6.843 \sin 2\pi \hat{f}_2 t \\ &= 17.086 + 10.031 \cos 2\pi (\hat{f}_2 t - 0.146) \\ &\quad + 7.085 \cos 2\pi (\hat{f}_2 t - 0.292).\end{aligned}$$

The fitted curve and the residuals are shown in Figure 3.4. The residuals now show no obvious periodic behavior.

The residual sum of squares is 54.7, and the root mean square value of the residuals is 0.3019, which are remarkably small values. The fact that the data were reported as integers means that they should contain errors at least as large as that caused by rounding off a number to the nearest integer. Since the error incurred by such rounding is roughly uniformly distributed from $-1/2$ to $1/2$, the mean squared error would be $1/12$. Thus rounding alone would cause a residual root mean square of around $\sqrt{1/12} = 0.2887$. Hence the unrounded data must have been almost exactly the sum of two pure sinusoids. A further curious feature of these data is that the two fitted periods are surprisingly close to whole numbers of days for the periods of a variable star. The apparent construction of this series will be mentioned in Chapter 6.

3.3 SOME MORE STATISTICAL RESULTS

The statistical properties of estimated frequencies may be studied under the same assumptions about the model as before, namely,

$$x_t = \mu + A \cos 2\pi f t + B \sin 2\pi f t + a_t,$$

where

$$\begin{aligned}E(a_t) &= 0, \\ E(a_t a_{t'}) &= \begin{cases} \sigma^2, & t = t' \\ 0, & \text{otherwise,} \end{cases}\end{aligned}$$

and independence of the errors. Since the frequency f enters nonlinearly into the model, the sampling distribution of its fitted value is relatively difficult to obtain. The problem was first studied by Whittle (1952) and

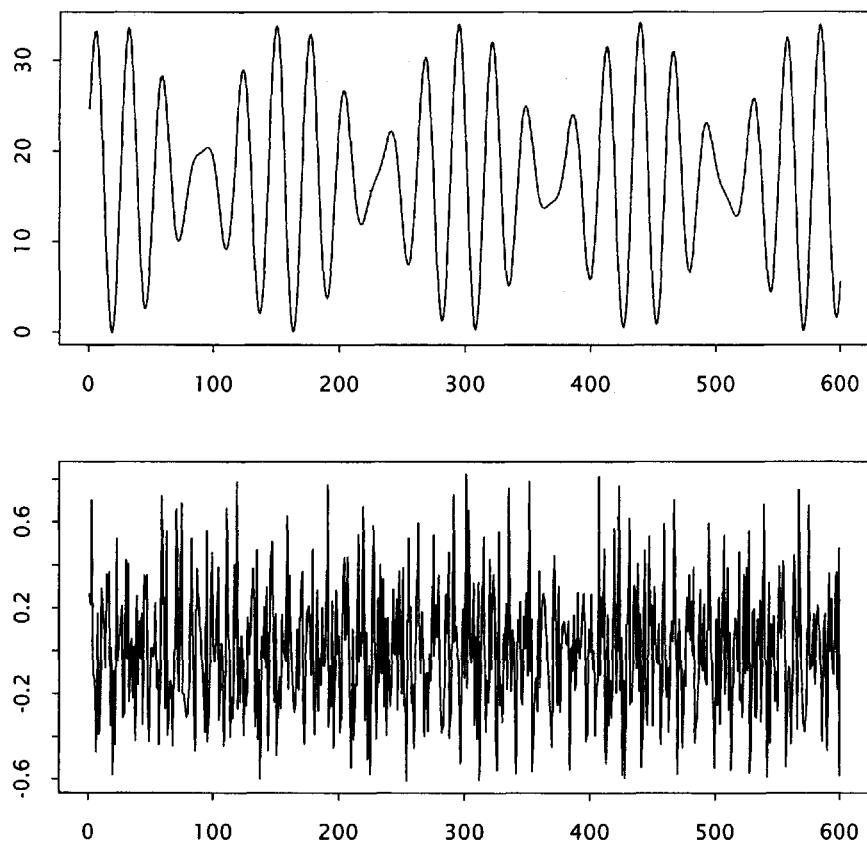


Fig. 3.4 Two-component model fitted to the variable star series. Upper graph is the fitted curve, lower graph shows the residuals (note expanded scale of axis).

later by Walker (1971). The principal results for the estimate \hat{f} are that

$$E(\hat{f}) = f + \text{ terms involving } \frac{1}{n}$$

and

$$\text{var}(\hat{f}) = \frac{6\sigma^2}{\pi^2 n^3 (A^2 + B^2)} + \text{smaller terms.}$$

At first sight, the n^{-3} behavior of $\text{var}(\hat{f})$ is surprising, since the variance of an estimated parameter usually behaves like the variances of \hat{A} and \hat{B} , that is like n^{-1} . However, we may easily demonstrate that a higher power is appropriate.

Consider the case in which $R^2 = A^2 + B^2$ is large compared with σ^2 . Thus the data consist of clear oscillations, with small errors superimposed. Then the number of cycles in the n data points may be counted accurately, and the only uncertainty involves the magnitude of the odd fraction of a cycle at each end of the data. If, for instance, the data show m complete cycles, but not $m+1$, the period $1/f$ is known to lie between $n/(m+1)$ and n/m , or $m/n \leq f \leq (m+1)/n$. Thus any estimate of f should lie within $1/n$ of the true value, and this implies that its variance is of order $1/n^2$ or better. The extra power of n is achieved by the relatively sophisticated least squares estimate \hat{f} .

The variance σ^2 of the errors is their mean square value. The corresponding quantity for the signal is

$$\begin{aligned}\text{ave}(A \cos 2\pi ft + B \sin 2\pi ft)^2 &= \text{ave}\{R \cos 2\pi(ft + \phi)\}^2 \\ &= R^2 \text{ave} \cos 2\pi(ft + \phi)^2 \\ &= \frac{R^2}{2}.\end{aligned}$$

The quantity

$$\frac{R^2/2}{\sigma^2} = \frac{\text{mean square value of signal}}{\text{mean square value of noise}},$$

called the *signal-to-noise ratio*, or snr, indicates how well the signal shows up in the noise. The variance of \hat{f} may be rewritten as

$$\text{var}(\hat{f}) \approx \frac{3}{\pi^2 n^3 \text{snr}},$$

which shows, perhaps surprisingly, that a *long series* is more effective in reducing the variance than a *strong signal*.

For the variable star series these formulas yield standard errors of 1.60×10^{-6} cycles per day for \hat{f}_1 and 2.27×10^{-6} cycles per day for \hat{f}_2 . These values show that the frequencies are in theory capable of very sharp resolution. It should be noted, however, that this result depends heavily on the assumptions made, especially independence of the errors.

The variances of \hat{A} and \hat{B} are increased by replacing f by its esti-

mate \hat{f} . The results are

$$\begin{aligned}\text{var}(\hat{A}) &\approx \frac{2\sigma^2}{n} \left(\frac{A^2 + 4B^2}{R^2} \right), \\ \text{var}(\hat{B}) &\approx \frac{2\sigma^2}{n} \left(\frac{4A^2 + B^2}{R^2} \right), \\ \text{cov}(\hat{A}, \hat{B}) &\approx \frac{2\sigma^2}{n} \left(\frac{-3AB}{R^2} \right), \\ \text{cov}(\hat{A}, \hat{f}) &\approx \frac{\sigma^2}{\pi n^2} \left(\frac{-6B}{R^2} \right), \\ \text{cov}(\hat{B}, \hat{f}) &\approx \frac{\sigma^2}{\pi n^2} \left(\frac{6A}{R^2} \right).\end{aligned}$$

However, if the model is reparametrized as

$$x_t = \mu + A' \cos 2\pi f(t - \bar{t}) + B' \sin 2\pi f(t - \bar{t}) + a_t$$

where $\bar{t} = (n - 1)/2$, then the results become

$$\text{var}(\hat{A}') \approx \text{var}(\hat{B}') \approx \frac{2\sigma^2}{n},$$

$$\text{cov}(\hat{A}', \hat{B}') \approx \text{cov}(\hat{A}', \hat{f}) \approx \text{cov}(\hat{B}', \hat{f}) \approx 0,$$

(see Exercise 3.1). That is, when the cosine and sine coefficients A and B are defined relative to a time origin at the midpoint of the series, rather than at its start, their variances and covariances are unchanged by estimating the frequency.

Furthermore, estimates concerning different frequencies are, to this order of approximation, uncorrelated. Since, as Walker (1971) shows, \hat{A} , \hat{B} , and \hat{f} are all approximately normally distributed, a confidence interval for each parameter is readily found.

In the light of the (approximate) standard errors of the two estimated frequencies, it is instructive to recall that their estimates when fitted jointly differed by many standard errors from the values obtained when fitting them separately. Since the two frequencies are fairly close, they interact or interfere with each other, and this effect dominates the statistical error, unless it is removed as in joint fitting. Pisarenko (1973) has shown that when a series contains two very similar frequencies the above formulas for variances and covariances may not be valid. Although Pisarenko's results are for frequencies closer than those in the present data, they suggest that the standard errors given above should be treated with some caution.

Exercise 3.1 Reparametrization

- (i) Rewrite the model

$$x_t = \mu + A \cos 2\pi f t + B \sin 2\pi f t + a_t$$

as

$$x_t = \mu + A' \cos 2\pi f(t - \bar{t}) + B' \sin 2\pi f(t - \bar{t}) + a_t$$

by expanding $\cos\{2\pi f(t - \bar{t}) + 2\pi f\bar{t}\}$ and $\sin\{2\pi f(t - \bar{t}) + 2\pi f\bar{t}\}$ using the trigonometric addition formulas. Obtain equations for A' and B' in terms of A , B , and f , by identifying the coefficients of $\cos 2\pi f t$ and $\sin 2\pi f t$.

- (ii) Use Taylor's expansion to find linear approximations to $\hat{A}' - A'$ and $\hat{B}' - B'$ in terms of $\hat{A} - A$, $\hat{B} - B$, and $\hat{f} - f$. Justify the linear expansion by showing that the omitted quadratic terms are of smaller order, using the approximate moments from this section.
- (iii) Obtain the variances and covariances of \hat{A}' , \hat{B}' , and \hat{f} from the linear approximation and the variances and covariances of \hat{A} , \hat{B} , and \hat{f} .

Appendix

Estimates of one or two frequencies were found using the nonlinear least squares function `nls` from S-PLUS. The commands for the third optimization, of both frequencies simultaneously, were

```
f1 <- 0.034442
f2 <- 0.041724
t <- time(star)
a <- 2 * pi * f1 * t
x <- cbind(cos(a), sin(a))
a <- 2 * pi * f2 * t
x <- cbind(x, cos(a), sin(a))
l <- lsfit(x, star)
n <- nls(star ~ mu + a1 * cos(2 * pi * f1 * t)
         + b1 * sin(2 * pi * f1 * t)
         + a2 * cos(2 * pi * f2 * t)
         + b2 * sin(2 * pi * f2 * t),
         data = data.frame(star, t),
         start = list(mu = l$coef[1],
                     a1 = l$coef[2], b1 = l$coef[3]),
```

```
a2 = l$coef[4], b2 = l$coef[5],  
f1 = f1, f2 = f2))
```

The values 0.034442 and 0.041724 cycles per day are starting values for the search, and were taken to be the values from fitting the two frequencies separately. The (linear) least squares fitting function `lsfit()` is used to find good starting values for the other parameters. Models with more frequencies may be fitted in the same way, but the computational effort may be expected to increase rapidly with their number.

4

Harmonic Analysis

Fitting a sinusoid of known frequency to a time series was discussed in Chapter 2, and the ideas were extended to fitting the frequency as well in Chapter 3. Such procedures are useful when periodicity clearly exists in the data and needs to be described exactly. This chapter covers a method for analyzing an *arbitrary* set of data into periodic components, whether or not the data appear periodic.

For some sequences of values such an analysis is of little meaning. However, other sets of data, although they may not appear to be periodic, do, in fact, contain interesting periodic components. Harmonic analysis is irreplaceable in detecting such components.

4.1 FOURIER FREQUENCIES

The result that underlies harmonic analysis is the *orthogonality* property of sinusoids with frequencies restricted to the *Fourier frequencies* $f_j = j/n$. This property was mentioned in Section 2.4, but will be exploited more fully in this chapter. Because of aliasing (see Section 2.5) the only frequencies that need to be considered are those satisfying $0 \leq f_j \leq 1/2$. Note that if n is even, then $1/2 = f_{n/2}$ is a Fourier frequency, but not if n is odd. The following identities are a direct consequence of the results

found in Exercise 2.2 (p. 13):

$$\begin{aligned} \sum_{t=0}^{n-1} \cos 2\pi f_j t &= 0, \quad j \neq 0 \\ \sum_{t=0}^{n-1} \sin 2\pi f_j t &= 0, \\ \sum_{t=0}^{n-1} \cos 2\pi f_j t \cos 2\pi f_{j'} t &= \begin{cases} n/2, & j = j' \neq 0 \text{ or } n/2, \\ n, & j = j' = 0 \text{ or } n/2, \\ 0, & j \neq j', \end{cases} \quad (4.1) \\ \sum_{t=0}^{n-1} \cos 2\pi f_j t \sin 2\pi f_{j'} t &= 0, \\ \sum_{t=0}^{n-1} \sin 2\pi f_j t \cos 2\pi f_{j'} t &= \begin{cases} n/2, & j = j' \neq 0 \text{ or } n/2, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

The special results for $j = j' = 0$ or $n/2$ hold essentially because the sine terms vanish identically at these frequencies. The first two relations are special cases of the next two with j set equal to 0 and j' replaced by j . The results state that the cosines and sines of the Fourier frequencies are *orthogonal* with respect to summation over the integers $0, 1, \dots, n - 1$. (Their orthogonality with respect to integration as well as summation is also important. See Exercise 4.1.)

Now suppose that x_0, x_1, \dots, x_{n-1} are any n numbers, and let

$$\begin{aligned} A(f) &= \frac{2}{n} \sum_{t=0}^{n-1} x_t \cos 2\pi f t, \\ B(f) &= \frac{2}{n} \sum_{t=0}^{n-1} x_t \sin 2\pi f t. \end{aligned} \quad (4.2)$$

Then the orthogonality relations (4.1) imply that

$$\begin{aligned} x_t &= A(0) + 2 \sum_{0 < j < n/2} \{A(f_j) \cos 2\pi f_j t + B(f_j) \sin 2\pi f_j t\} \\ &\quad [+ A(f_{n/2}) \cos 2\pi f_{n/2} t], \quad t = 0, 1, \dots, n - 1, \end{aligned} \quad (4.3)$$

the term in square brackets being included only if n is even (see Exercise 4.2). Thus an *arbitrary* sequence of numbers may be represented as a sum of periodic components. Notice that, whether n is even or odd, there are n coefficients in the sum.

This is, of course, not the only way in which a representation in terms of cosines and sines may be found. If the n data values are regarded as

a single point in n -dimensional space, a similar representation holds for any set of frequencies whose cosines and sines form a basis for that space. However, the Fourier frequencies are a natural set to use, being equally spaced over the range of frequencies we wish to use. Also, because of the orthogonality relations (4.1), the coefficients $\{A(0), A(f_1), B(f_1), \dots\}$ are easily calculated. This is not true for most other sets of frequencies.

Equation (4.3) has been shown to hold only for t in the range $0, 1, \dots, n-1$. Any other value of t may be written as $kn + t'$ for some integer k and t' in this range. If $kn + t'$ is substituted for t in the right-hand side of (4.3), the k s drop out because of the form of the Fourier frequencies. Thus the value found is simply $x_{t'}$. Hence if t is an unrestricted integer variable, the sum defines a periodic sequence with period n , consisting of the values x_0, x_1, \dots, x_{n-1} repeated cyclically.

The j th Fourier frequency has period $1/f_j = n/j$. A sinusoid with this frequency executes j complete cycles in the span of the data, thus providing a useful interpretation of the index j . One consequence of this is that few of the periods are integers (although of course all are rational). The harmonic analysis (4.3) represents the decomposition of a series into components each of which is repeated a whole number of times in the span of the data.

Exercise 4.1 Orthogonality of Sinusoids with Respect to Integration

- (i) Show that if t is an integer, then

$$\int_0^1 \cos 2\pi f t \, df = \begin{cases} 1, & \text{if } t = 0, \\ 0, & \text{otherwise,} \end{cases} \quad (4.4)$$

and

$$\int_0^1 \sin 2\pi f t \, df = 0, \quad \text{all } t. \quad (4.5)$$

- (ii) Hence show that for integers t and t' ,

$$\int_0^1 \cos 2\pi f t \cos 2\pi f t' \, df = \begin{cases} 1 & t = t' = 0, \\ \frac{1}{2}, & t = \pm t' \neq 0, \\ 0, & t \neq \pm t', \end{cases}$$

$$\int_0^1 \cos 2\pi f t \sin 2\pi f t' \, df = 0, \quad \text{all } t, t'$$

$$\int_0^1 \sin 2\pi f t \sin 2\pi f t' \, df = \begin{cases} \frac{1}{2}, & t = t' \neq 0, \\ -\frac{1}{2}, & t = -t' \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Exercise 4.2 Inverse Relations

Since $f_j t = jt/n = f_t j$, the orthogonality relations (4.1) may also be written (for t and t' in the range from 0 to $n - 1$) as

$$1 + 2 \sum_{0 < j < n/2} \cos 2\pi f_j t \cos 2\pi f_j t' [+ \cos 2\pi f_{n/2} t \cos 2\pi f_{n/2} t']$$

$$= \begin{cases} n/2, & t = t' \neq 0 \text{ or } n/2, \\ n/2, & t = n - t' \neq n/2, \\ n, & t = t' = 0 \text{ or } n/2, \\ 0, & \text{otherwise,} \end{cases}$$

$$2 \sum_{0 < j < n/2} \cos 2\pi f_j t \sin 2\pi f_j t' = 0, \quad \text{all } t, t'$$

$$2 \sum_{0 < j < n/2} \sin 2\pi f_j t \sin 2\pi f_j t' = \begin{cases} n/2, & t = t' \neq 0 \text{ or } n/2, \\ -n/2, & t = n - t' \neq n/2, \\ 0, & \text{otherwise;} \end{cases}$$

here the term in square brackets is omitted if n is odd. Use these relations to verify the inversion formula (4.3).

Exercise 4.3 Integral Inverse Formula

The integral orthogonality property derived in Exercise 4.1 may be used to show that

$$\begin{aligned} x_t &= \int_0^1 \{A(f) \cos 2\pi f t + B(f) \sin 2\pi f t\} df \\ &= 2 \int_0^{\frac{1}{2}} \{A(f) \cos 2\pi f t + B(f) \sin 2\pi f t\} df, \quad t = 0, 1, \dots, n - 1. \end{aligned}$$

Verify this equation, and show that if the right-hand side is evaluated for other values of t , the value is 0. Note the contrast with the discrete inverse formula (4.3).

4.2 DISCRETE FOURIER TRANSFORM

The theory of Section 4.1 is simpler when written in complex form. In particular, all formulas are the same for even n as for odd n , and there are no separate forms for cosine *versus* sine functions. There are in fact

occasions when a pair of real-valued time series are most naturally regarded as the real and imaginary parts of a single complex-valued time series. The deviations of the instantaneous axis of rotation of the earth provide one example, studied by Brillinger (1973). The deviation is measured by a pair of Cartesian coordinates which may be treated theoretically as a single complex number.

In general, however, observed data are strictly real-valued. These may always be regarded as complex numbers with zero imaginary parts, although this may seem an unnecessary complication. However, certain algebraic simplifications that arise make the required stretch of the imagination seductively appealing.

The Euler relation

$$e^{ix} = \cos x + i \sin x$$

and its inverse

$$\cos x = \frac{1}{2} \{e^{ix} + e^{-ix}\}, \quad \sin x = \frac{1}{2i} \{e^{ix} - e^{-ix}\}$$

were used in Chapter 2, and show that the complex exponential function e^{ix} is intimately related to the cosines and sines. The orthogonality relations corresponding to (4.1) are

$$\begin{aligned} \sum_{t=0}^{n-1} e^{2\pi if_j t} \overline{e^{2\pi if_k t}} &= \sum_{t=0}^{n-1} e^{2\pi if_j t} e^{-2\pi if_k t} \\ &= \begin{cases} n & \text{if } j \equiv k \pmod{n}, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{4.6}$$

The result for $j \equiv k \pmod{n}$ is immediate, for then each summand is 1. The other values follow from the identity derived in Exercise 2.2 (p. 13).

Now suppose that x_0, x_1, \dots, x_{n-1} are any n complex numbers. Let

$$d(f) = \frac{1}{n} \sum_{t=0}^{n-1} x_t e^{-2\pi ift}. \tag{4.7}$$

Notice that $d(f) = A(f)/2 - iB(f)/2$, where $A(f)$ and $B(f)$ were defined in the previous section. The data may be recovered from $d(f)$ by the equation

$$x_t = \sum_j d(f_j) e^{2\pi if_j t}, \tag{4.8}$$

which follows from the inverse form of (4.6) (see Exercise 4.4). The summation may be over $j = 0, 1, \dots, n-1$ or $-n/2 < j \leq n/2$. Both ranges

give the same answer because $d(f)$ is periodic with period 1. The range $-n/2 < j \leq n/2$ is perhaps the more natural in the light of the results of the previous section; it is the same range, with negative frequencies added because $e^{2\pi ift}$ and $e^{2\pi i(-f)t}$ are not identical. Note that if the symmetric range $-n/2 \leq j \leq n/2$ were used instead, the difference between n even and odd would reappear, and the terms for $j = \pm n/2$ would have to be given weight 1/2. The range of j will usually be taken to be $0, 1, \dots, n-1$, however, to emphasize the similarity between (4.7) and (4.8).

By analogy with the usual (integral) Fourier transform, $d(f)$ is called the *discrete Fourier transform* of $\{x_0, x_1, \dots, x_{n-1}\}$, which are obtained from $d(f)$ by the *inverse transform* (4.8). In the same way, $A(f)$ and $B(f)$ of Section 4.1 are referred to as the cosine and sine transforms, respectively.

The introduction of negative frequencies (or frequencies between $1/2$ and 1) into the transform and its inverse brings no new information to the analysis of real-valued data. From (4.7),

$$\begin{aligned} d(1-f) &= d(-f) \\ &= \frac{1}{n} \sum_{t=0}^{n-1} x_t e^{2\pi ift}, \end{aligned}$$

and if all the x 's are real, then

$$\begin{aligned} \frac{1}{n} \sum_{t=0}^{n-1} x_t e^{2\pi ift} &= \frac{1}{n} \sum_{t=0}^{n-1} \overline{x_t e^{-2\pi ift}} \\ &= \overline{d(f)}. \end{aligned}$$

Thus for real-valued data, the values of $d(f)$ for negative frequencies are determined by the values for $0 \leq f \leq 1/2$.

The discrete Fourier transform, as a complex function, possesses two natural representations. The first is in terms of its real and imaginary parts, which are just $A(f)/2$ and $-B(f)/2$, where $A(f)$ and $B(f)$ were defined in the previous section. The second is in terms of its *magnitude* $R(f)$ and *phase* and $\phi(f)$, defined by $R(f) = |d(f)|$ and

$$d(f) = R(f)e^{i\phi(f)}.$$

The magnitude $R(f)$ measures how strongly the oscillation with frequency f is represented in the data. It is often displayed in the form

$$I(f) = nR(f)^2 = n|d(f)|^2,$$

where $I(f)$ is known as the *periodogram*.

If f is restricted to be a Fourier frequency, the cosine and sine transforms $A(f_j)$ and $B(f_j)$ are precisely the cosine and sine coefficients obtained in fitting a cosine wave by least squares (see Section 2.4). The discrete Fourier transform has a similar interpretation in the context of fitting the complex exponential function: the sum of squares function

$$S(d) = \sum_{t=0}^{n-1} |x_t - d e^{2\pi i f t}|^2$$

is minimized at $d = d(f)$, regardless of whether f is a Fourier frequency.

The Fourier transform has an elementary but important property known as *linearity* or *superposability*, which will be used in many places below. Suppose that the data $\{x_t\}$ are the sum (or *superposition*) of $\{y_t\}$ and $\{z_t\}$. Then

$$\begin{aligned} \sum_t x_t e^{-2\pi i f t} &= \sum_t (y_t + z_t) e^{-2\pi i f t} \\ &= \sum_t y_t e^{-2\pi i f t} + \sum_t z_t e^{-2\pi i f t} \end{aligned}$$

or, in an obvious notation,

$$d_x(f) = d_y(f) + d_z(f).$$

In words, the transform of the sum is the sum of the transforms.

In the rest of this book, the complex discrete Fourier transform will be used extensively. However, unless stated otherwise, the series being analyzed will *always* be assumed to be real.

Exercise 4.4 Complex Inverse Relations

The inverse version of (4.6) is

$$\sum_j e^{2\pi i f_j t} \overline{e^{2\pi i f_j t'}} = \begin{cases} n & \text{if } t \equiv t' \pmod{n}, \\ 0 & \text{otherwise,} \end{cases}$$

where the range of summation is any set of n consecutive integers j . Verify this, and use it to derive the inverse transform (4.8).

Exercise 4.5 Other Inversion Formulas

- (i) The complex analog of (4.4) and (4.5) is

$$\int_0^1 e^{2\pi i f t} \overline{e^{2\pi i f t'}} df = \begin{cases} 1, & t = t', \\ 0, & \text{otherwise.} \end{cases}$$

Verify this, and use it to confirm the complex analog of the result in Exercise 4.3,

$$x_t = \int_0^1 d(f) e^{2\pi i f t} df, \quad t = 0, 1, \dots, n - 1.$$

Note that, as before, the right-hand side evaluates to 0 for other values of t .

- (ii) The orthogonality relations (4.6) may be generalized to

$$\sum_{t=0}^{n-1} e^{2\pi i (f_j + \theta)t} \overline{e^{2\pi i (f_k + \theta)t}} = \begin{cases} n & \text{if } j \equiv k \pmod{n}, \\ 0 & \text{otherwise.} \end{cases}$$

Verify this, and deduce that

$$x_t = \sum_j d(f_j + \theta) e^{2\pi i (f_j + \theta)t}, \quad t = 0, 1, \dots, n - 1.$$

Here, as before, the range of summation of j is any set of n consecutive integers, such as $j = 0, 1, \dots, n - 1$, and consequently it is sufficient to consider $f_0 = 0 < \theta < f_1 = 1/n$.

- What values are obtained if the sum is evaluated for other values of t ?
- Show that in the particular case $\theta = 1/2n$, the values x_0, x_1, \dots, x_{n-1} are repeated cyclically, but with alternate blocks opposite in sign.

Parts (i) and (ii) both provide inversion formulas for obtaining x_0, x_1, \dots, x_{n-1} from the transform $d(f)$, $0 \leq f < 1$ or $-1/2 < f \leq 1/2$. They give different answers when evaluated for t outside this range, which may be useful when the observed data are to be embedded in a longer sequence, perhaps of zeros or of cyclic repetitions of the data. Note that the formulas may be averaged, still giving the values x_0, x_1, \dots, x_{n-1} for $t = 0, 1, \dots, n - 1$ but embedded in a variety of longer series.

4.3 DECOMPOSING THE SUM OF SQUARES

The orthogonality relations of Sections 4.1 and 4.2 imply identities between the sums of squares of the original data and of the transforms. For the cosine and sine transforms $A(f)$ and $B(f)$ this takes the form

$$\sum_{t=0}^{n-1} x_t^2 = nA(0)^2 + 2n \sum_{0 < j < n/2} \{A(f_j)^2 + B(f_j)^2\} \quad [+ nA(f_{n/2})^2], \quad (4.9)$$

where, as always, the term in square brackets is included only if n is even. The analog for the discrete Fourier transform $d(f)$ is

$$\sum_{t=0}^{n-1} |x_t|^2 = n \sum_j |d(f_j)|^2 = \sum_j I(f_j), \quad (4.10)$$

which again shows the algebraic simplicity of the complex version.

This partitioning of a sum of squares is usually known as an *analysis of variance*. In the present case, it may be interpreted as a parallel to the representation

$$x_t = \sum_j d(f_j) e^{2\pi i f_j t} \quad (4.11)$$

of the data. In (4.11) the data are exhibited as a sum of various periodic components. In (4.10) the *sum of squares* of the data is decomposed into corresponding parts.

Exercise 4.6 Decomposing the Sum of Squares

Relation (4.9) may be derived either from (4.2) and the inverse orthogonality relations (Exercise 4.2), or from (4.3) and the original orthogonal relations (4.1). Verify this, and find the corresponding two ways to derive (4.10).

4.4 HARMONIC ANALYSIS OF SOME SPECIAL FUNCTIONS

In this section, harmonic analysis of some particular functions is used to display some of the properties of the discrete Fourier transform.

(i) Cosine Wave

In terms of Fourier analysis, the simplest place to begin is with a cosine wave, $R \cos 2\pi(f_0 t + \phi)$. Since the transform is a linear operation, R occurs only as a scale factor throughout, and may be taken to be 1. Now

$$\begin{aligned} \cos 2\pi(f_0 t + \phi) &= \frac{1}{2} [e^{2\pi i(f_0 t + \phi)} + e^{-2\pi i(f_0 t + \phi)}] \\ &= \frac{1}{2} (e^{2\pi i\phi} e^{2\pi i f_0 t} + e^{-2\pi i\phi} e^{-2\pi i f_0 t}), \end{aligned}$$

whence it is sufficient to find the transform of $e^{2\pi i f_0 t}$ (once again, $e^{2\pi i \phi}$ is just a scale factor). But the complex transform of $e^{2\pi i f_0 t}$ is

$$d(f) = \frac{1}{n} \sum_{t=0}^{n-1} e^{2\pi i f_0 t} e^{-2\pi i f t}.$$

The sum may be evaluated using the result of Exercise 2.2 (p. 13) to give

$$d(f) = e^{2\pi i (f_0 - f)\bar{t}} n D_n(f - f_0),$$

where $\bar{t} = (n - 1)/2$ is the average of the time values $0, 1, \dots, n - 1$ and

$$D_n(f) = \frac{\sin \pi f n}{n \sin \pi f}$$

is the Dirichlet kernel (see Section 2.2).

The transform vanishes if $f - f_0$ is a multiple of $1/n$, that is, if f and f_0 differ by a *Fourier frequency*, except at $f = f_0$, where it takes the value 1. Since the original cosine wave involves $-f_0$ as well as f_0 , its transform vanishes only if $f \pm f_0$ are both nonzero Fourier frequencies. This is true if f and f_0 are different Fourier frequencies, but also if they each differ from a Fourier frequency by $1/2n$, that is, if they fall midway between Fourier frequencies. The first factor has modulus 1, whence the amplitude of the transform is (the absolute value of) the second factor, and the phase is the phase of the first factor (+ 1/2 if the second is negative).

When $f \pm f_0$ is not a Fourier frequency, the transform is nonzero. The appearance of a nonzero value in the transform at a frequency f because of the presence of a sinusoid at a different frequency f_0 is called *leakage*. Since harmonic analysis is carried out to *separate* the effects of different frequencies, leakage is an undesirable phenomenon. The transform is typically calculated at the Fourier frequencies, and leakage then occurs whenever there are oscillations in the data at other frequencies. This may sometimes be avoided, by choosing n to be a multiple of the periods of all oscillations; for instance, by analyzing a whole number of years' worth of data when only the annual wave, the semiannual wave, etc., are expected. This is not always possible or desirable, and techniques for controlling leakage when it is unavoidable are discussed in Section 6.2.

The Dirichlet kernel arises frequently in the study of harmonic analysis. If n is moderately large and f is small, it is approximately

$$\frac{\sin \pi f n}{\pi f n}.$$

The functions $g_0(x) = (\sin \pi x)/(\pi x)$ and $D_n(x/n)$ are shown in Figure 4.1 for $n = 5$ and 20. The principal features are the peak of height 1

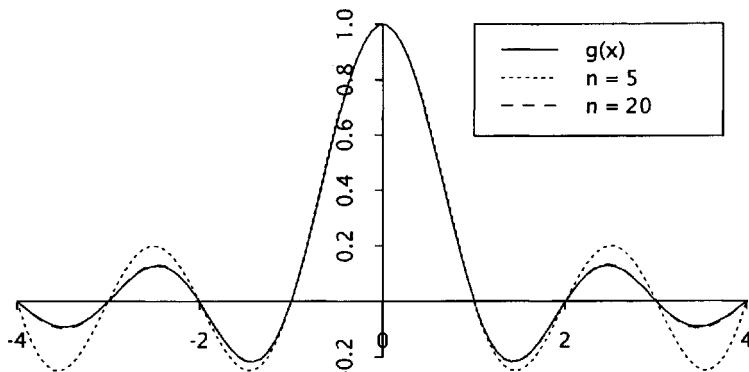


Fig. 4.1 The function $g_0(x) = (\sin \pi x)/(\pi x)$ and the Dirichlet kernels $D_n(x/n)$, $n = 5, 20$.

at $x = 0$, the zeroes at $x = \pm 1, \pm 2, \dots$, and the relatively slowly decaying minor peaks or *sidelobes*.

Thus the amplitude of the transform $d(f)$ of the complex wave $e^{2\pi i f t_0}$ consists approximately of the function $(\sin \pi x)/(\pi x)$ centered at f_0 and scaled so that its zeros are separated by $1/n$, the frequency separation of adjacent Fourier frequencies. The transform is large only for f close to f_0 , and as f moves away from f_0 its magnitude decays proportionally to $1/|f - f_0|$.

The transform of the original real cosine wave consists of contributions from f_0 and $-f_0$. When f is positive, the term centered at $-f_0$ is relatively small unless f is close to 0 or $1/2$ cycles per unit time.

(ii) Single Impulse

The transform of a sequence that consists of a single nonzero value of 1 at $t = t_0$ is

$$d(f) = \frac{1}{n} e^{-2\pi i f t_0}.$$

It has constant amplitude, and the phase is $(\text{mod } 1)$ a linear function of frequency. Notice the duality between this example and the last. They emphasize the phenomenon that if a function is highly *localized* (i.e., large over only a small part of its domain), then its transform is very *diffuse* (nearly constant in amplitude and linear in phase) and conversely.

(iii) Step Function

Let $\{x_t\}$ be the step function

$$x_t = \begin{cases} 1, & 0 \leq t < m, \\ 0, & m \leq t < n, \end{cases}$$

for some m in the range $1, 2, \dots, n - 2$. Then

$$\begin{aligned} d(f) &= \frac{1}{n} \sum_{t=0}^{n-1} x_t e^{-2\pi ift} \\ &= \frac{1}{n} \sum_{t=0}^{m-1} e^{-2\pi ift} \\ &= e^{-2\pi if(m-1)/2} \frac{\sin \pi f m / 2}{n \sin \pi f / 2} \\ &= e^{-2\pi if(m-1)/2} \frac{m}{n} D_m(f). \end{aligned}$$

The amplitude of $d(f)$ is thus the (absolute value of the) Dirichlet function $(m/n)|D_m(f)|$, which decays roughly as $1/f$ from a maximum of $m/n < 1$ at $f = 0$. As in examples (i) and (ii), the phase of the transform is a linear function of frequency.

(iv) Straight Line

A general linear function $a + bt$ may be written as a linear combination of a constant and the simplest linear function, which for the present purpose is

$$x_t = t - \bar{t} = t - \frac{n-1}{2}, \quad t = 0, 1, \dots, n-1.$$

The transform of this function is

$$d(f) = \frac{i}{2n} e^{-2\pi if\bar{t}} \frac{n \sin \pi f \cos \pi f n - \cos \pi f \sin \pi f n}{(\sin \pi f)^2}$$

for $f \neq 0$, and $d(0) = 0$. The amplitude of $d(f)$ is the absolute value of the second factor, and for large n and small f this is approximately $n^2 g_1(nf)$, where

$$g_1(x) = \frac{\pi x \cos \pi x - \sin \pi x}{(\pi x)^2}.$$

This function and $d_n(x/n)/n^2$ are shown in Figure 4.2 for $n = 5$ and 20. Note that $\pi g_1(x)$ is the derivative of $(\sin \pi x)/(\pi x)$, while the linear function of which it is approximately the transform is the integral of

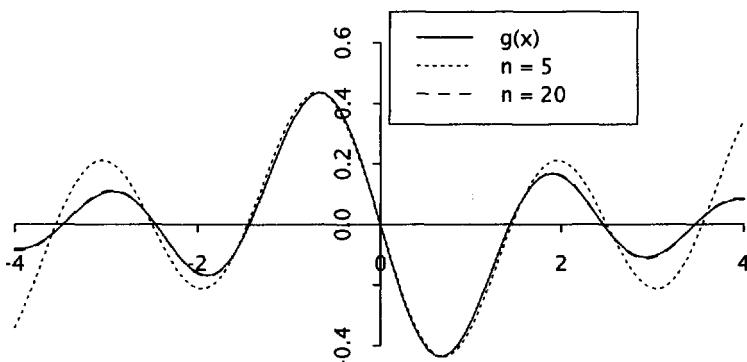


Fig. 4.2 The functions $g_1(x) = \frac{\pi x \cos \pi x - \sin \pi x}{(\pi x)^2}$ and $d_n(x/n)/n^2$, $n = 5, 20$.

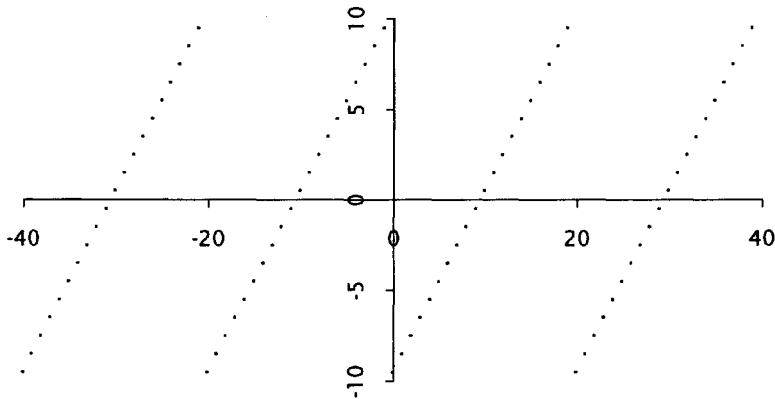


Fig. 4.3 The sawtooth function, $n = 20$.

the step function. As in the case of the step function, the decay of the sidelobes for large $|f|$ is like $1/|f|$.

It should be recalled that when the inverse transform is evaluated for values other than $0, 1, \dots, n - 1$, the *periodic extension* of the original series is found. In the present case, this is the sawtooth function shown in Figure 4.3, and not a continued straight line, as might have been expected.

(v) Shifts, Symmetry, and Linear Phase

In all the above examples, the phase of the transform depends linearly on the frequency f . This is a consequence of certain symmetries in the corresponding series.

Consider first the effect of *shifting* (cyclically permuting) a series. Suppose that $y_t = x_{t+h}$ for some (integer) h , the subscripts being interpreted $(\text{mod } n)$. For example, if $h = 1$ then $y_t = x_{t+1}$ for $t = 0, 1, \dots, n-1$ and $y_{n-1} = x_0$. The transform of the shifted series, for a general shift h , is

$$\begin{aligned} d_y(f) &= \frac{1}{n} \sum_t y_t e^{-2\pi i f t} \\ &= \frac{1}{n} \sum_t x_{t+h} e^{-2\pi i f t} \\ &= \frac{1}{n} \sum_t x_t e^{-2\pi i f (t-h)} \\ &= \frac{1}{n} e^{2\pi i f h} \sum_t x_t e^{-2\pi i f t} \\ &= e^{2\pi i f h} d_x(f). \end{aligned}$$

The amplitudes of the transforms are the same, and the phases differ by a linear function of $2\pi f$, the coefficient h being precisely the shift.

Now consider a periodic series symmetric about $t = 0$ (a series with *even symmetry*), that is, with $x_{-t} = x_{n-t} = x_t$, $t = 1, 2, \dots, n-1$. In this case,

$$\begin{aligned} d(f) &= \frac{1}{n} \sum_t x_t e^{-2\pi i f t} \\ &= \frac{1}{n} \sum_t x_{n-t} e^{-2\pi i f t} \\ &= \frac{1}{n} \sum_t x_{n-t} e^{2\pi i f (n-t)} \\ &= \frac{1}{n} \sum_{t'} x_{t'} e^{2\pi i f t'} \\ &= \overline{d(f)}, \end{aligned}$$

and hence the transform $d(f)$ is real. The converse is also true. In the same way, for an antisymmetric series (a series with *odd symmetry*), one for which $x_{-t} = -x_t$, the transform $d(f)$ is purely imaginary, and conversely.

In Example (iii), the step function, the amplitude is a symmetric function of f and the phase is $-f(m-1)/2$. If the series were shifted by $-(m-1)/2$, the transform would be entirely real and would thus be symmetric. This implies that the series has even symmetry about $t = (m-1)/2$. Strictly, this argument is valid only for odd m , as the shift must be an integer. However, the conclusion is also valid for even m , since a series may be symmetric about either an integer time or a half-integer time, and the whole argument is easily extended to the latter case.

In Example (iv), a shift of $-(n-1)/2$, or equivalently of

$$n - (n-1)/2 = (n+1)/2,$$

makes the transform purely imaginary, whence this series has odd symmetry about $t = (n-1)/2$. Both symmetries are easily verified from the definitions of the corresponding series.

(vi) Periodic Series

Suppose that the series x_0, x_1, \dots, x_{n-1} is periodic with period h , an integer, and that n is an integer multiple of h , say $n = kh$. The discrete Fourier transform $d(f)$ satisfies

$$\begin{aligned} d(f) &= \frac{1}{n} \sum_{t=0}^{n-1} x_t e^{-2\pi i f t} \\ &= \frac{1}{n} \sum_{t=0}^{h-1} \sum_{u=0}^{k-1} x_{t+hu} e^{-2\pi i f(t+hu)} \\ &= \frac{1}{n} \sum_{t=0}^{h-1} x_t e^{-2\pi i f t} \sum_{u=0}^{k-1} e^{-2\pi i fh u} \\ &= \frac{1}{n} \sum_{t=0}^{h-1} x_t e^{-2\pi i f t} e^{-\pi i f h(k-1)} k D_k(fh) \\ &= d_h(f) e^{-\pi i f h(k-1)} D_k(fh), \end{aligned}$$

where

$$d_h(f) = \frac{1}{h} \sum_{t=0}^{h-1} x_t e^{-2\pi i f t}$$

is the transform of the first full cycle of the series. Thus $d(j/h) = d_h(j/h)$ for any integer j , but $d(f) = 0$ for any other Fourier frequency. That is, the discrete Fourier transform, when evaluated at the Fourier

frequencies, is nonzero only at $f = 1/h$ and its multiples. This frequency is called the *fundamental frequency* of the wave, and its multiples are *harmonics*.

The argument shows that the transform of a periodic series must vanish at certain Fourier frequencies, and not that it is necessarily nonzero at the remainder. For instance, if the periodic series were just $\cos 2\pi t/h = \cos 2\pi f_h t$, only the fundamental frequency would have a nonzero coefficient, and all the harmonics would disappear from the representation.

Exercise 4.7 Quadratic Function

Find the transform of the series

$$x_t = \left(t - \frac{n-1}{2} \right)^2, \quad t = 0, 1, \dots, n-1,$$

and an approximation for large n , small f , of the form

$$d(f) \approx e^{-\pi i f(n-1)} g_q(nf).$$

Should $g_q(x)$ be real or imaginary? At what rate does $g_q(x)$ decay for large $|x|$?

Exercise 4.8 Triangle Wave

The series

$$x_t = \begin{cases} t, & 0 \leq t \leq n/2, \\ n-t, & n/2 \leq t < n \end{cases}$$

defines a triangle (or hat) function that extends periodically to be a triangle wave. Find its transform and an approximation as in Exercise 4.7. Find the rate of decay of the approximation.

Exercise 4.9 Truncated Sinusoid

The step function of Example (iii) is a special case of a truncated sinusoid,

$$x_t = \begin{cases} \cos 2\pi(f_0 t + \phi), & 0 \leq t < m, \\ 0, & m \leq t < n. \end{cases}$$

Find the transform of this series, and note that, in general, it is nonzero at all of the Fourier frequencies, even if the frequency f_0 is itself a Fourier frequency.

4.5 SMOOTH FUNCTIONS

Consider a sequence $\{x_0, x_1, \dots, x_{n-1}\}$ for which the cosine and sine coefficients are large only for low frequencies. This means that the sequence may be represented as a sum of terms each of which is smooth, and hence the sequence itself must be smooth. Since the cosines and sines of higher frequencies oscillate rapidly, no smooth sequence can contain them with large amplitudes. Thus smooth sequences must be those whose transforms are small except at low frequencies.

This qualitative argument may be made more precise in several ways. The simplest is to define the (circular) *roughness coefficient* of the sequence to be

$$\frac{\sum_{t=0}^{n-1} (x_t - x_{t-1})^2}{\sum_{t=0}^{n-1} x_t^2},$$

where the definition is made circular by replacing x_{-1} by x_{n-1} . This definition reflects the idea that, in a smooth sequence of points, the differences between successive values are all relatively small, and hence the numerator will be small. The denominator is essentially a scaling factor, since the roughness of a sequence is a dimensionless quantity and should not depend on the magnitudes of the numbers. In many cases the roughness coefficient should also not depend on the *level* of the data (that is, to be unaffected by adding the same constant to each x_t , $t = 0, 1, \dots, n-1$). This is achieved by replacing x_t by $x_t - \bar{x}$, with very little effect on what follows.

This roughness coefficient was defined, although not with this name, by Ernst Abbe in 1863 (see Kendall, 1971). The noncircular version

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

is also known as the von Neumann ratio and the Durbin-Watson statistic. It has been studied by von Neumann et al. (1941), von Neumann (1941, 1942) and Hart and von Neumann (1942), and by Durbin and Watson (1950, 1951, 1971). In each case the principal goal was to find the probability distribution of the ratio when the series $\{x_t\}$ has some Gaussian distribution. However, the coefficient is of interest here more as a general indicator of the roughness (or smoothness) of a sequence.

The noncircular definition is the more natural, since a series that is otherwise smooth should not usually be penalized for a difference between its end values. However, the circular definition is more convenient and the difference is slight, at least for large n .

The roughness coefficient may be represented in terms of the discrete Fourier transform of the sequence as follows. The complex version of the discrete Fourier transform is used for convenience. Its inverse is

$$x_t = \sum_{j=0}^{n-1} d(f_j) e^{2\pi i f_j t},$$

whence

$$\begin{aligned} x_t - x_{t-1} &= \sum_{j=0}^{n-1} d(f_j) [e^{2\pi i f_j t} - e^{2\pi i f_j (t-1)}] \\ &= \sum_{j=0}^{n-1} d(f_j) e^{2\pi i f_j t} (1 - e^{-2\pi i f_j}). \end{aligned}$$

The results of Section 4.3 allow the roughness coefficient to be rewritten as

$$\begin{aligned} \frac{\sum_{j=0}^{n-1} |1 - e^{-2\pi i f_j}|^2 |d(f_j)|^2}{\sum_{j=0}^{n-1} |d(f_j)|^2} &= \frac{\sum_{j=0}^{n-1} 2(1 - \cos 2\pi f_j) |d(f_j)|^2}{\sum_{j=0}^{n-1} |d(f_j)|^2} \\ &= \frac{\sum_{j=0}^{n-1} 4(\sin \pi f_j)^2 |d(f_j)|^2}{\sum_{j=0}^{n-1} |d(f_j)|^2}. \end{aligned}$$

Since $\sin \pi f$ increases from 0 to 1 as f increases from 0 to $1/2$, it is clear that, for a fixed value of the denominator, the roughness is reduced by increasing the amplitude of low-frequency terms at the expense of high-frequency terms.

The smoothest function in this sense is the constant function, for which the roughness coefficient vanishes. The roughest functions are those for which the transform vanishes except at the frequency or frequencies closest to $1/2$, that is

$$\begin{aligned} x_t &= A \cos \pi t \\ &= A(-1)^t, \quad A \neq 0, \quad t = 0, 1, \dots, n-1 \end{aligned}$$

if n is even, and

$$\begin{aligned}x_t &= R \cos 2\pi \left\{ \left(\frac{1}{2} - \frac{1}{2n} \right) t + \phi \right\} \\&= R(-1)^t \cos 2\pi \left(\frac{t}{2n} - \phi \right), \quad R > 0, \quad t = 0, 1, \dots, n-1\end{aligned}$$

if n is odd. In each case the series oscillates rapidly. The cosine term in the case of odd n (which completes one half of a cycle in the span of the series) is present because of the circularity of the definition of the roughness coefficient.

If x_t is replaced by $x_t - \bar{x}$ in the definition of the coefficient, the only modification is that $d(0)$ vanishes. The roughness coefficient is thus undefined for the constant function, and the smoothest functions are of the form $A \cos 2\pi(t/n + \phi)$, which executes one complete cycle in the span of the data.

Exercise 4.10 Smooth Sequence

Suppose that

$$x_t = f \left(\frac{t + \frac{1}{2}}{n} \right),$$

where $f(x)$ is a continuous function with a continuous derivative on the interval $(0, 1)$. Then for large n , $\{x_t\}$ is certainly a smooth sequence. Show that the roughness coefficient is approximately

$$\frac{n^{-2} \int_0^1 f'(x)^2 dx}{\int_0^1 f(x)^2 dx}.$$

Exercise 4.11 Segments of a Series

Suppose that a series is divided into two segments, which either start with the same value or end with the same value. Show that the roughness coefficient of the whole series lies between the coefficients of the two segments.

5

The Fast Fourier Transform

Despite the similarity of the names, the fast Fourier transform is neither a variant of nor an alternative to the discrete Fourier transform described in Chapter 4. It is instead an algorithm (or, rather, a class of related algorithms) for *computing* the discrete Fourier transform of a data series at all of the Fourier frequencies, using relatively few arithmetic operations.

The background of the fast Fourier transform and its computational advantages are described in Section 5.1. The simplest case, in which the series length factorizes into two factors, is discussed in Section 5.2. Section 5.3 describes how the algorithm may be used to carry out harmonic analysis of observed time series data.

5.1 COMPUTATIONAL COST OF FOURIER TRANSFORMS

The simplest way to compute the discrete Fourier transform of a set of data x_0, x_1, \dots, x_{n-1} of length n is to evaluate the sums

$$nd(f_j) = \sum_{t=0}^{n-1} x_t e^{-2\pi i f_j t}, \quad j = 0, 1, \dots, n - 1, \quad (5.1)$$

in turn. The complex formulation is used for convenience, and complex addition and multiplication will be used to count the computational cost of the algorithm. The first sum, for $j = 0$, requires just $n - 1$ additions,

since the exponent is 0 for each summand. Each of the remaining sums requires $n - 1$ multiplications and $n - 1$ additions, since the first term in each is x_0 . Thus the total cost is $(n - 1)^2$ multiplications and $n(n - 1)$ additions, or roughly n^2 of each. The values of

$$e^{-2\pi if_j t} = e^{-2\pi i(jt/n)} \quad (5.2)$$

are also needed for all j and t in the range $0, 1, \dots, n - 1$. Since (5.2) depends only on the residue of jt (mod n), there are only n distinct values, which could be tabulated to avoid repeated calculation.

The fast Fourier transform gives the same values (5.1), but the number of additions and multiplications used is of the order $n \log_2 n$. In comparison with the simple method, the cost is reduced by a factor of the order of $(\log_2 n)/n$. For the relatively short series length of 1024, Gentleman and Sande (1966) found that the execution times for ALGOL programs using a refinement of the simple method and a fast Fourier transform were 59.1 seconds and 2.0 seconds, respectively (on an IBM 7094 computer). The advantage of the fast Fourier transform increases with n and makes feasible the transformation of very long series.

The most frequently used algorithm was first discussed in detail by Cooley and Tukey (1965), although the basic idea was known much earlier. For a history of the development of the algorithm see Cooley et al. (1967). An important variation (the Sande-Tukey algorithm) was introduced by Gentleman and Sande (1966). Both of these variations depend on the series length n having many small factors (being *highly composite*). A simple version of each occurs when n is a power of 2. A different algorithm described by Good (1958, 1971) requires that n possess mutually prime factors, and thus cannot be used in the case $n = 2^k$. Bluestein has described an algorithm that does not depend on the factorization of n (see Brigham, 1988, p. 195).

5.2 TWO-FACTOR CASE

The basic idea behind the fast Fourier transform may be seen in the case where n has two factors $n = n_1 n_2$, when the Cooley-Tukey and Sande-Tukey algorithms are in fact the same. For any t in the range $0, 1, \dots, n - 1$, t_1 and t_2 may be found such that

$$t = t_1 n_2 + t_2, \quad 0 \leq t_1 < n_1, \quad 0 \leq t_2 < n_2.$$

Specifically, t_1 is the integer part of t/n_2 and t_2 is the remainder. The numbers t_1 and t_2 are the *characters* (generalized digits) of t in the finite

mixed-radix arithmetic with radices n_1 and n_2 . Similarly, for any such t_1 and t_2 , the value $t_1 n_2 + t_2$ lies in the range $0, 1, \dots, n - 1$. Thus each x_t is associated unambiguously with the corresponding pair (t_1, t_2) . If the data are written in a table with n_1 rows and n_2 columns, filling the rows of the table consecutively, x_t will fall in the $(t_1 + 1)$ st row and the $(t_2 + 1)$ st column, here called the (t_1, t_2) position:

$$\begin{array}{ccccccc} & & & \xrightarrow{\quad\longrightarrow\quad} & & & \\ x_0 & x_1 & \cdots & x_{n_2-1} & & & \\ x_{n_2} & x_{n_2+1} & \cdots & x_{2n_2-1} & & & \\ \vdots & \vdots & & \vdots & & & \\ x_{n-n_2} & x_{n-n_2+1} & \cdots & x_{n-1}. & & & \end{array}$$

The entry in the (t_1, t_2) position will be denoted y_{t_1, t_2} ,

$$y_{t_1, t_2} = x_{t_1 n_2 + t_2}.$$

The reversed arithmetic with radices (n_2, n_1) is used to represent the index j of the transform. Specifically, j is written $j_2 n_1 + j_1$ with $0 \leq j_2 < n_2$ and $0 \leq j_1 < n_1$.

For any integer α let $W_\alpha = e^{-2\pi i/\alpha}$. Then the transform of $\{x_0, x_1, \dots, x_{n-1}\}$ may be written as

$$\begin{aligned} nd(f_j) &= \sum_{t=0}^{n-1} x_t W_n^{jt} \\ &= \sum_{t_1=0}^{n_1-1} \sum_{t_2=0}^{n_2-1} x_{t_1 n_2 + t_2} W_n^{(j_2 n_1 + j_1)(t_1 n_2 + t_2)} \\ &= \sum_{t_1=0}^{n_1-1} \sum_{t_2=0}^{n_2-1} y_{t_1, t_2} W_n^{j_2 n_1 t_2 + j_1 t_1 n_2 + j_1 t_2}, \end{aligned}$$

since

$$W_n^{j_2 n_1 t_1 n_2} = W_n^{j_2 t_1 n} = 1.$$

The sums may be rearranged as

$$\begin{aligned} nd(f_j) &= \sum_{t_2=0}^{n_2-1} W_n^{j_2 n_1 t_2} W_n^{j_1 t_2} \sum_{t_1=0}^{n_1-1} y_{t_1, t_2} W_n^{j_1 t_1 n_2} \\ &= \sum_{t_2=0}^{n_2-1} W_n^{j_2 t_2} \left\{ W_n^{j_1 t_2} \sum_{t_1=0}^{n_1-1} y_{t_1, t_2} W_n^{j_1 t_1} \right\}, \end{aligned}$$

since

$$W_n^{n_1} = W_{n_2} \quad \text{and} \quad W_n^{n_2} = W_{n_1}.$$

Now

$$z_{j_1, t_2} = \sum_{t_1=0}^{n_1-1} y_{t_1, t_2} W_n^{j_1 t_1}$$

is (proportional to) the j_1 th term in the transform of the n_1 numbers $y_{0, t_2}, y_{1, t_2}, \dots, y_{n_1-1, t_2}$. Similarly

$$nd(f_j) = \sum_{t_2=0}^{n_2-1} W_n^{j_2 t_2} \{W_n^{j_1 t_2} z_{j_1, t_2}\}$$

is (proportional to) the j_2 th term in the transform of the n_2 numbers $W_n^0 z_{j_1, 0}, W_n^{j_1} z_{j_1, 1}, \dots, W_n^{j_1(n_2-1)} z_{j_1, n_2-1}$. Thus the overall transform of the series of length $n = n_1 n_2$ may be accomplished by a number of transforms of subseries of lengths n_1 and n_2 , together with multiplication by the intermediate “twiddle factors” $W_n^{j_1 t_2}$.

The final transform $nd(f_j)$ may also be arranged in a table with n_1 rows and n_2 columns. In computer programs, it is usually the *same* table as the data were stored in at the start of the process. However, since the mixed radix representation of j used n_1 and n_2 in the reverse order, consecutive entries fall in the same *column*:

$$\begin{array}{cccc} d(f_0) & d(f_{n_1}) & \dots & d(f_{n-n_1}) \\ d(f_1) & d(f_{n_1+1}) & \dots & d(f_{n-n_1+1}) \\ \vdots & \vdots & & \vdots \\ d(f_{n_1-1}) & d(f_{2n_1-1}) & \dots & d(f_{n-1}). \end{array}$$

Hence the transform must be copied out in a different order, a process known as “unscrambling.”

In computing the transform for all n of the Fourier frequencies, the process involves n_2 transforms of subseries of length n_1 and n_1 transforms of subseries of length n_2 , as well as $n = n_1 n_2$ twiddle factor multiplications. If straightforward summation is used to evaluate these shorter transforms, the total computational cost is roughly $n_2 n_1^2 + n_1 n_2^2 + n_1 n_2 = n_1 n_2 (n_1 + n_2 + 1)$ multiplications and $n_2 n_1^2 + n_1 n_2^2 = n_1 n_2 (n_1 + n_2)$ additions, or roughly $n_2 n_1^2 + n_1 n_2^2 = n_1 n_2 (n_1 + n_2)$ of each. This may be compared with $n^2 = n_1 n_2 (n_1 n_2)$, the approximate number of multiplications and additions used in the simple method. The computational cost is reduced by the factor $(n_1 + n_2)/n_1 n_2$. For instance, if $n_1 = n_2 = 5$ the factor is 0.4, and if $n_1 = n_2 = 10$ it is 0.2.

In the general algorithm it is assumed that n may be factorized further as $n_1 n_2 \dots n_k$. In this case the computational cost is reduced roughly by the factor $(n_1 + n_2 + \dots + n_k)/n$. If each factor is 2, this becomes $2k/n =$

$(2 \log_2 n)/n$, which is the result usually quoted. Bergland (1968) gives precise formulas for the number of real additions and multiplications required by some different versions of the algorithm.

5.3 APPLICATION TO HARMONIC ANALYSIS OF DATA

The efficiency of the fast Fourier transform algorithm has made possible the routine harmonic analysis of extensive sets of data. However, the gains may be small if n has large prime factors. Sometimes a small part of the data may be discarded to leave a more convenient number. However, this is often unacceptable, since it inevitably involves some loss of information. One solution is to use the *chirp-Z* algorithm, introduced by Bluestein (see Brigham, 1988, p. 195). This algorithm works by rewriting the basic transform sum

$$nd(f) = \sum_{t=0}^{n-1} x_t e^{-2\pi i f t}$$

as a convolution, and using the fast Fourier transform to carry out the convolution numerically (see Section 7.2).

Another solution is to *pad* the data with zeros. This has the side effect of changing the grid of frequencies at which the transform is calculated, making it finer. Data are in fact sometimes padded to achieve just this refinement of the grid of frequencies. The padded data $x'_0, x'_1, \dots, x'_n, x'_{n+1}, \dots, x'_{n'}$ are defined as

$$x'_t = \begin{cases} x_t & 0 \leq t < n \\ 0 & n \leq t < n'. \end{cases}$$

The discrete Fourier transform $d'(f)$ of these padded data satisfies

$$\begin{aligned} n'd'(f) &= \sum_{t=0}^{n'-1} x'_t e^{-2\pi i f t} \\ &= \sum_{t=0}^{n-1} x_t e^{-2\pi i f t} \\ &= nd(f) \end{aligned}$$

and therefore differs trivially from $d(f)$. However, it is normally evaluated on the grid of Fourier frequencies for the padded data, namely, $f = 0, 1/n', 2/n', \dots, (n' - 1)/n'$.

This refinement of the grid is sometimes sought for graphical purposes, for instance, allowing $d(f)$ and functions derived from it to be plotted

more smoothly. The extension of the data by zeros can be beneficial in certain applications, such as the computation of *autocovariances* (see Section 8.4). However, it emphasizes the problem of *leakage*: the phenomenon in which the presence of a particular harmonic component causes the transform to be nonzero at other frequencies. Example (i) of Chapter 4 (p. 45) showed that any frequency other than a Fourier frequency causes leakage into all Fourier frequencies. In the present case, however, a sinusoid in the original data becomes a truncated sinusoid in the extended data, and it was shown in Exercise 4.9 (p. 52) that the transform of this is in general nonzero at the Fourier frequencies, for *any* frequency in the original data. Thus *any* harmonic component in the original data will give rise to leakage in the extended data. This makes the control of leakage more obviously desirable, although it is in fact essential even in the analysis of nonextended data. Techniques for leakage reduction are discussed in Section 6.2.

Exercise 5.1 Correcting for the Mean

Suppose that the series x_0, \dots, x_{n-1} shows relatively small fluctuations y_t about a large average value, x (that is, $x_t = x + y_t$, where $|y_t| \ll |x|$). The series is analyzed by extending it with zeros to a length $n' > n$, followed by calculation of the discrete Fourier transform (at the Fourier frequencies for n').

- (i) Show that leakage from the zero-frequency component x is present in all terms of the transform. How large would x have to be for this leakage to dominate the transform of $\{y_t\}$?
- (ii) Show that, if the mean of the *extended* series is subtracted from each term in the *extended* series, this leakage is unaffected.
- (iii) Show that, if the mean of the *original* series is subtracted out *before* the extension by zeros, the leakage is removed completely.
- (iv) Show that these considerations hold also for any periodic component in the data whose frequency is a Fourier frequency for the original, unextended data.

6

Examples of Harmonic Analysis

The theory of harmonic analysis was described in Chapter 4, and a numerical algorithm for carrying it out with relatively little computational effort was given in Chapter 5. This chapter presents some examples of harmonic analysis of data series and illustrates the insights that such analyses can (and cannot) provide about the data. Sections 6.2 and 6.7 mention two ways in which the data may be modified before the analysis to make the interesting features of the data more apparent.

6.1 VARIABLE STAR DATA

As a first example consider the variable star data (Figure 1.2, p. 3), which were used in Chapter 3 to illustrate the model of hidden periodicities. In that chapter it was shown that these data consist of two strong sinusoidal components plus residuals that are of precisely the order of magnitude expected to arise from rounding off the data to integers. Harmonic analysis will now be used to see what structure, if any, there is to these residuals. Harmonic analysis can detect whether any behavior of interest occurs at frequencies other than the two discovered in Section 3.2.

Figure 6.1 shows the amplitude $R(f)$ of the discrete Fourier transform for the variable star data, graphed on a logarithmic scale for $0 < f \leq 1/2$. The logarithm of the *periodogram* $I(f) = n|d(f)|^2 = nR(f)^2$ is a linear

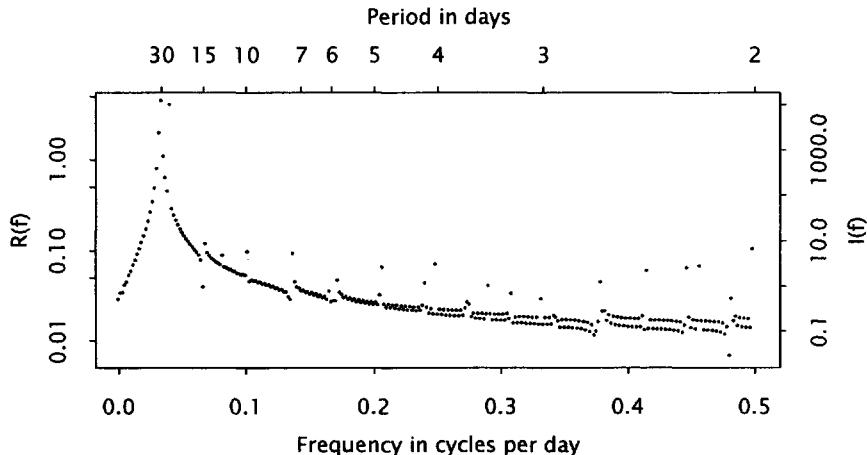


Fig. 6.1 Amplitude function for variable star data at frequencies $f_j = j/600$ cycles per day, $j = 1, \dots, 299$.

function of the logarithm of $R(f)$, and this graph may also be interpreted as a graph of $I(f)$ against f by a simple change of axis labeling. The right-hand axis of Figure 6.1 is labeled accordingly. Recall that the discrete Fourier transform is usually computed on the grid of frequencies $f = f_j = j/n$, $0 \leq j < n$, but graphs are often restricted to $0 \leq f \leq 1/2$ because of the symmetries described in Section 4.2. In the present graph, the point at $f = 0$ was omitted since $d(0)$ is just the series mean, and comparing it with the amplitudes of sinusoidal components is usually of no interest.

The logarithmic scale is preferable for plotting because of the variation in the order of magnitude of the amplitude between different frequencies. Had the logarithmic scale not been used, no scaling of the axis could accommodate the details both at the largest and the smallest amplitudes.

The peaks corresponding to the two components are the dominant features of the graph. The largest ordinates are at $j = 21$ and 25 , corresponding to $f = 0.0350$ and 0.0417 cycles per day, respectively. These two frequencies would have provided adequate starting frequencies for the nonlinear least squares fitting of Section 3.2. This would be the usual procedure for determining the number of components and the approximate values of the corresponding frequencies.

The moderately large values of the amplitude close to the peaks are caused by *leakage* from the peaks. The results of Section 3.2 showed that no other components with amplitudes as large as 1.0 could be present in

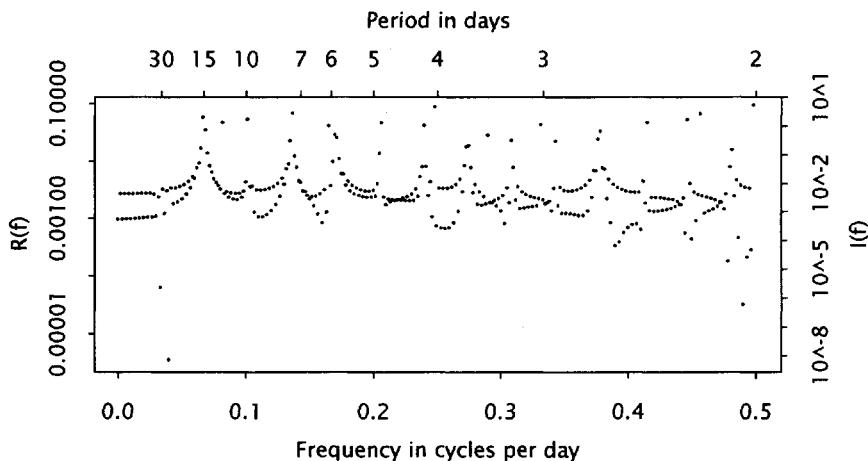


Fig. 6.2 Amplitude and periodogram functions for the final residuals of the variable star data.

the data, since the series may be written as a sum of the two components plus residuals with a average squared value of around 1/12.

The simplest way to remove the leakage is to remove its sources, the two strong components. (Another way is described in the next section.) Figure 6.2 shows the amplitude function of the residuals from the two-component model fitted in Section 3.2, graphed as before. Note first that all trace of the two main peaks has gone. There are, in fact, slight "holes" where they have been removed.

Comparison of Figures 6.1 and 6.2 shows that the leakage in the former extends to essentially *all* frequencies. Figure 6.2 shows a succession of peaks, some accompanied by their own leakage profiles, separated by troughs in which the values of the amplitude are smaller by two or more orders of magnitude. These peaks are mostly visible also in Figure 6.1, although they are not all as well defined. A few are completely submerged in the leakage.

All of these subsidiary peaks, the largest of which are at amplitudes of a little less than 0.1, occur at multiples of the frequencies of the two strong components. This observation suggests that the original signal is the sum of two *periodic* terms, each of which is almost, but not exactly, sinusoidal (see Example (vi) of Chapter 4, p. 51). Table 6.1 displays the frequencies of all local maxima with amplitudes larger than 0.014, and their relationships with the frequencies of the main peaks.

Table 6.1 Frequencies associated with subsidiary peaks in Figure 6.2.

Frequency	Multiple of f_{21}	Multiple of f_{25}
41	2	
50		2
62	3	
83	4	
100		4
103	5	
125	6	5
145	7	
150		6
166	8	
175		7
186	9	
200		8
207	10	
228	11	
250		10
269	13	
275		11
290	14	
300		12

6.2 LEAKAGE REDUCTION BY DATA WINDOWS: TAPERS AND FADERS

It was demonstrated in Section 6.1 that leakage from strong components in a set of data may be reduced by removing its source, the strong components themselves. This section introduces an alternative method that leaves the sources intact in the series, but dramatically reduces the magnitude of the leakage.

Consider first the simple case of data consisting of a pure sinusoid:

$$x_t = Re^{2\pi i(f_0 t + \phi)}, \quad t = 0, 1, \dots, n - 1.$$

For simplicity, take $R = 1$ and $\phi = 0$; the general case may be recovered by multiplying by $Re^{2\pi i\phi}$. The transform of this series is

$$\begin{aligned} d(f) &= \frac{1}{n} \sum_{t=0}^{n-1} x_t e^{-2\pi i f t} \\ &= \exp \left\{ \frac{2\pi i(n-1)(f_0 - f)}{2} \right\} \frac{\sin\{\pi n(f_0 - f)\}}{n \sin\{\pi(f_0 - f)\}} \end{aligned}$$

(see Exercise 2.2, p. 13). The algebra to follow is simpler when written in terms of a slightly modified transform in which time is measured relative to $\bar{t} = (n - 1)/2$, the average of the data times instead of that of the first observation. The modified transform is

$$d^c(f) = \frac{1}{n} \sum_{t=0}^{n-1} x_t e^{-2\pi i f(t-\bar{t})} \quad (6.1)$$

$$\begin{aligned} &= e^{2\pi i f \bar{t}} d(f) \\ &= e^{2\pi i f_0 \bar{t}} \frac{\sin\{\pi n(f_0 - f)\}}{n \sin\{\pi(f_0 - f)\}}. \end{aligned} \quad (6.2)$$

For large n and $f_0 \neq f$,

$$\begin{aligned} d^c(f - 1/n) &= e^{2\pi i f_0 \bar{t}} \frac{\sin\{\pi n(f_0 - f) + \pi\}}{n \sin\{\pi(f_0 - f) + \pi/n\}} \\ &\approx e^{2\pi i f_0 \bar{t}} \frac{\sin\{\pi n(f_0 - f) + \pi\}}{n \sin\{\pi(f_0 - f)\}} \\ &= -d^c(f), \end{aligned}$$

and, similarly, $d^c(f + 1/n) \approx -d^c(f)$. Thus, if $d_H^c(f)$ is defined by

$$d_H^c(f) = \frac{1}{4} d^c(f - 1/n) + \frac{1}{2} d^c(f) + \frac{1}{4} d^c(f + 1/n), \quad (6.3)$$

it follows that $d_H^c(f) \approx 0$ to this order of approximation. More precisely,

$$\begin{aligned} d_H^c(f) &= \frac{1}{4n} e^{2\pi i f_0 \bar{t}} \sin\{\pi n(f_0 - f)\} \\ &\times \left[\frac{2}{\sin\{\pi(f_0 - f)\}} - \frac{1}{\sin\{\pi(f_0 - f) - \pi/n\}} - \frac{1}{\sin\{\pi(f_0 - f) + \pi/n\}} \right]. \end{aligned}$$

If n is large, then

$$\sin\{\pi n(f_0 - f) - \pi/n\} \approx \sin\{\pi n(f_0 - f)\} - \frac{\pi}{n} \cos\{\pi n(f_0 - f)\}$$

and

$$\sin\{\pi n(f_0 - f) + \pi/n\} \approx \sin\{\pi n(f_0 - f)\} + \frac{\pi}{n} \cos\{\pi n(f_0 - f)\},$$

and hence, after some simplification,

$$\begin{aligned} d_H^c(f) &\approx -\frac{\pi^2}{2n^3} e^{2\pi i f_0 \bar{t}} \sin\{\pi n(f_0 - f)\} \\ &\times \frac{[\cos\{\pi(f_0 - f)\}]^2}{\sin\{\pi(f_0 - f)\} ([\sin\{\pi(f_0 - f)\}]^2 - (\pi^2/n^2)[\cos\{\pi(f_0 - f)\}]^2)} \\ &\approx -\frac{\pi^2}{2n^3} e^{2\pi i f_0 \bar{t}} \sin\{\pi n(f_0 - f)\} \frac{[\cos\{\pi(f_0 - f)\}]^2}{[\sin\{\pi(f_0 - f)\}]^3}. \end{aligned} \quad (6.4)$$

Operation (6.3) is known as *hanning*, and the subscript H distinguishes its results. The leakage in the hanned version of d^c , approximated by (6.4), is better than that for d^c itself, given exactly by (6.2), in two ways. First, it contains a factor n^{-3} rather than n^{-1} ; this means that it is much smaller in long series. Second, the factor $[\sin\{\pi(f_0 - f)\}]^{-3}$ decays like $|f_0 - f|^{-3}$ rather than $|f_0 - f|^{-1}$, and thus for a given value of n the leakage diminishes more rapidly as f moves away from f_0 , the position of the peak. Hence the leakage is reduced in magnitude and is contained more closely around $f = f_0$.

If (6.1) and (6.3) are used to define $d_H^c(f)$ for an arbitrary series $\{x_t\}$, then (see Exercise 6.1)

$$d_H^c(f) = \frac{1}{n} e^{2\pi i f t} \sum_{t=0}^{n-1} x_t e^{-2\pi i f t} \frac{1 - \cos\left\{2\pi\left(t + \frac{1}{2}\right)/n\right\}}{2}, \quad (6.5)$$

implying that

$$d_H(f) = e^{-2\pi i f t} d_H^c(f) \quad (6.6)$$

is the transform of the series

$$y_t = x_t \times \frac{1 - \cos\left\{2\pi\left(t + \frac{1}{2}\right)/n\right\}}{2}, \quad t = 0, 1, \dots, n-1. \quad (6.7)$$

The function $d_H(f)$ defined in (6.6) will be called the *hanned transform* of $\{x_t\}$, although it should be pointed out that it is *not* found from $d(f)$ by hanning as in (6.3). Equation (6.7) shows that an alternative way to compute $d_H(f)$ is to multiply the data x_0, x_1, \dots, x_{n-1} by the *data window* (or *fader*)

$$w_t = \frac{1}{2} \left[1 - \cos \left\{ \frac{2\pi\left(t + \frac{1}{2}\right)}{n} \right\} \right], \quad t = 0, 1, \dots, n-1 \quad (6.8)$$

and transform the *windowed* or *tapered* data.

When data are analyzed without first tapering them, a rectangular data window (or *boxcar*) is, in effect, used by default:

$$v_t = \begin{cases} 1, & 0 \leq t < n, \\ 0, & \text{otherwise.} \end{cases}$$

The raised cosine sequence $\{w_t\}$ defined in (6.8) may be thought of as a smooth approximation to the boxcar. Now the transform of a tapered sinusoid is just the transform of the data window centered at the frequency of the sinusoid (see Exercise 6.2). Hence leakage in the transform of a

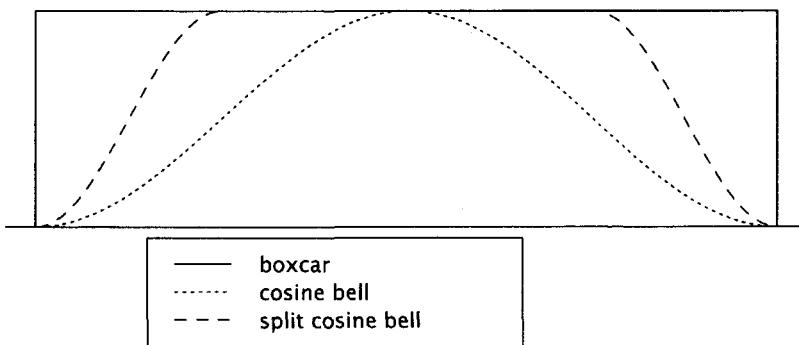


Fig. 6.3 The boxcar, the cosine bell window, and the split cosine bell window, $p = 0.5$.

tapered sinusoid is caused by the sidelobes of the transform of the data window (which is the Dirichlet kernel for the boxcar). It was shown in Sections 4.4 and 4.5 that the transform of a smooth series decays more rapidly than that of a rough series. Thus the smoothness of the data window $\{w_t\}$ leads to its good leakage characteristics.

Consequently, any smooth approximation to the boxcar is a candidate for use as a data window. Specifically, if $w(x)$ is a smooth function defined for $0 \leq x \leq 1$, with $w(0) = w(1) = 0$, then $w_t = w\{(2t + 1)/2n\}$ could be used as a data window. In many cases it is desirable to leave the bulk of the data unmodified, and just taper the ends. A convenient window may be constructed by separating the two halves of the raised cosine bell (6.8) and inserting a stretch of 1s. This gives the *split cosine bell* window

$$w_p(x) = \begin{cases} \frac{1}{2}\{1 - \cos 2\pi x/p\}, & 0 \leq x < p/2, \\ 1, & p/2 \leq x < 1 - p/2, \\ \frac{1}{2}\{1 - \cos 2\pi(1-x)/p\}, & 1 - p/2 \leq x \leq 1, \end{cases} \quad (6.9)$$

where p , the proportion of the data that is tapered, is some desired value. Tukey (1967) has suggested that 10% or 20% may be suitable. It may be seen (see Exercise 6.5) that this window gives an intermediate reduction in leakage. Figure 6.3 shows the boxcar and cosine bell data windows and the split cosine bell with 50% tapering. The transforms of these windows (computed with centered times as in equation 6.1) are shown in Figure 6.4. Other candidate functions for use as data windows are discussed by Harris

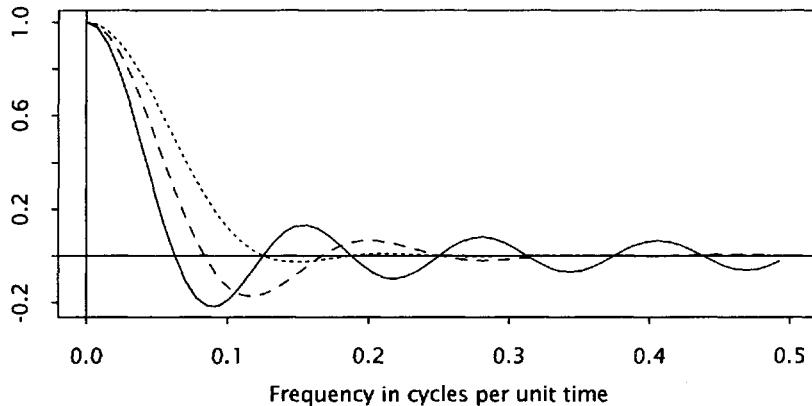


Fig. 6.4 Transforms of the data windows shown in Figure 6.3 for a series of length 16.

(1978) and Marple (1987). Kaiser and Schafer (1980) present a simple and attractive window based on Bessel functions.

Exercise 6.1 Hanning

Verify that (6.1) and (6.3) imply (6.5).

Exercise 6.2 Transform of a Tapered Sinusoid

Suppose that $\{w_0, w_1, \dots, w_{n-1}\}$ is a sequence of numbers. In the present context, they may be interpreted as a data window, but the results of this exercise and the next do not depend on this interpretation. The transform of this sequence is

$$d_w(f) = \frac{1}{n} \sum_{t=0}^{n-1} w_t e^{-2\pi i f t}.$$

Let

$$z_t = w_t e^{2\pi i f_0 t}, t = 0, 1, \dots, n-1$$

be a corresponding tapered sinusoid. Verify that the transform of $\{z_t\}$ is

$$\begin{aligned} d_z(f) &= \frac{1}{n} \sum_{t=0}^{n-1} z_t e^{-2\pi i f t} = \frac{1}{n} \sum_{t=0}^{n-1} w_t e^{-2\pi i (f-f_0)t} \\ &= d_w(f-f_0). \end{aligned}$$

Exercise 6.3 Transform of a Product of Series

The following fundamental result is easily verified using the result of the preceding exercise. Suppose that $\{x_t\}$ and $\{y_t\}$ are any two series, with discrete Fourier transforms $d_x(f)$ and $d_y(f)$, respectively. Let $z_t = x_t y_t, t = 0, 1, \dots, n - 1$. If x_t is written as the inverse of its transform

$$x_t = \sum d(f_j) e^{2\pi i f_j t},$$

then z_t is a sum of sinusoids with coefficients $\{d(f_j)\}$, each “tapered” by $\{y_t\}$. From this, deduce that the transform of $\{z_t\}$ is

$$d_z(f_j) = \sum_k d_x(f_k) d_y(f_j - f_k) = \sum_k d_x(f_k) d_y(f_{j-k}),$$

the *convolution* of the transforms of $\{x_t\}$ and $\{y_t\}$. In words, *the transform of a product is the convolution of the transforms*. Note that the convolution is defined circularly, since $d(f)$ is a periodic function of f with period 1.

Verify the more general result

$$\begin{aligned} d_z(f) &= \sum_k d_x(f_k) d_y(f - f_k) \\ &= \sum_k d_y(f_k) d_x(f - f_k). \end{aligned}$$

Exercise 6.4 Continuation: The Dual Result

The duality between the (complex) discrete Fourier transform and its inverse makes the dual result easy to prove. If

$$z_t = \sum_u x_u y_{t-u}, t = 0, 1, \dots, n - 1,$$

where now $t - u$ is interpreted $(\text{mod } n)$, verify that if f is a Fourier frequency then

$$d_x(f) = n d_x(f) d_y(f),$$

or, in words, *the transform of a convolution is (proportional to) the product of the transforms*.

Exercise 6.5 Split Cosine Bell Data Window

Suppose that $\{u_t\}$ is the data window for a series of length n obtained from the split cosine bell window $w_p(x)$ defined as in (6.9). Find its

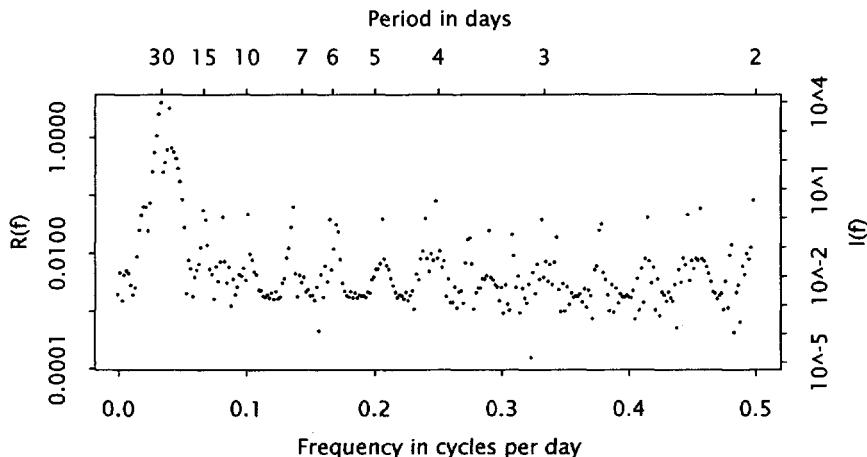


Fig. 6.5 Periodogram of the variable star data, tapered 25%.

transform.¹ Show that the rate of decay of the transform lies between that of the boxcar and that of the (undivided) cosine bell window that results from hanning.

6.3 TAPERING THE VARIABLE STAR DATA

Figure 6.5 is the periodogram (plotted on a logarithmic scale) of the variable star data, with the mean subtracted and with 25% of the data tapered. Comparison with Figure 6.2 (p. 65) shows that the leakage from the main peaks has been eliminated except in their immediate neighborhood. The closest subsidiary peaks now stand out from the background. There is a price to be paid for this reduction in leakage, namely, that the peaks are slightly broader or more rounded than they were (see Exercise 6.6), although this change is not obvious in the graph. For these data, it appears that 25% tapering is sufficient to contain the leakage from the main peaks to a usefully narrow band of frequencies. It should be pointed out that tapering requires far less thought and effort than the model-fitting procedure that led, as one result, to Figure 6.2. However, it gives no precise information about the locations of the peaks.

¹As a result of Exercise 6.2 the transform $d(f)$ is needed for all frequencies f , not just the Fourier frequencies $f_j = j/n$.

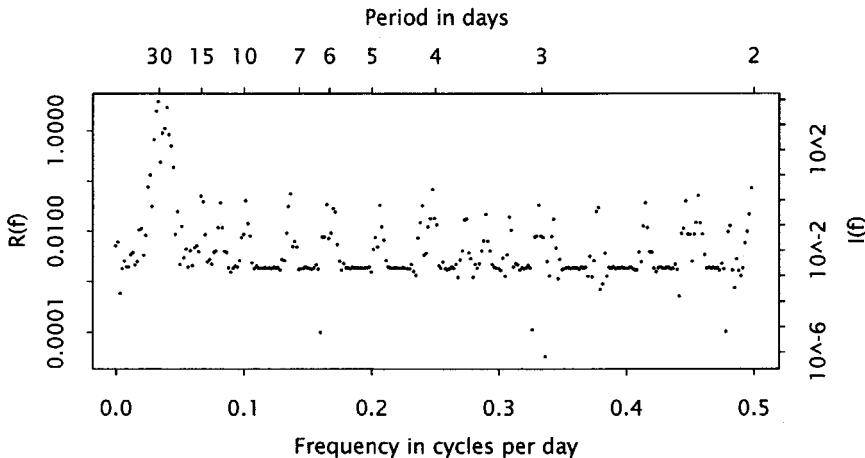


Fig. 6.6 Periodogram of the variable star data, tapered 50%.

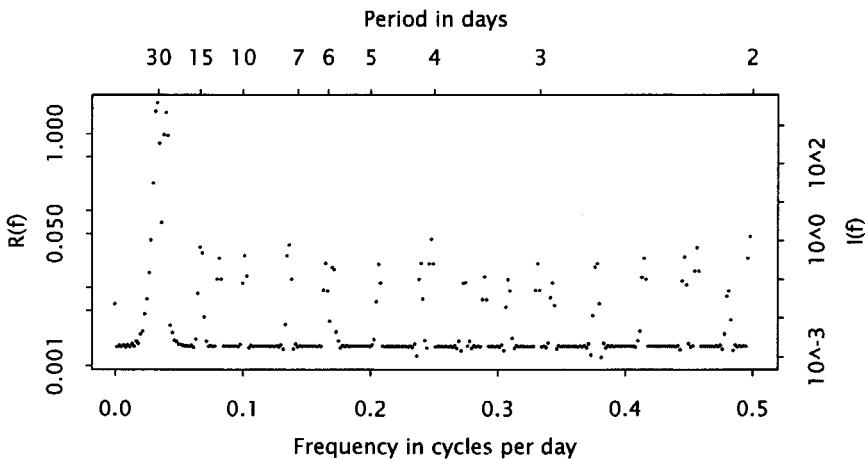


Fig. 6.7 Periodogram of the variable star data, tapered 100%.

Figures 6.6 and 6.7 show the results of 50% and 100% tapering, respectively, the latter being the periodogram derived from the hanned transform (6.6). The progressive reduction of leakage is clear, especially around the main peaks. The accompanying broadening of the peaks may also be seen.

The near constancy of the periodogram away from the peaks is highly unusual. More typically, periodogram values fluctuate randomly in any

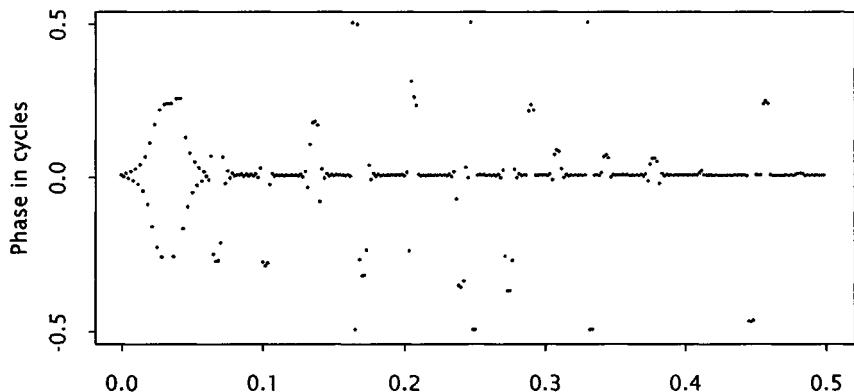


Fig. 6.8 Phase of the hanned transform of the variable star data, centered at $t_0 = 300$.

interval of frequencies where there is no peak (see Section 6.8). This suggests that tapering has made the transforms of the periodic components small enough to reveal the presence of yet another component, one with a constant periodogram. The simplest series with a constant periodogram is one in which only a single value is nonzero (see Example (ii) of Chapter 4, p. 47), and often such a component arises because of an error in a single observation. In this case the phase of the transform is a linear function of frequency (mod 1), and the slope of the phase plot is the negative of the time offset of the nonzero term from the start of the record. The magnitude of the nonzero term is n times the magnitude of the transform, $nR(f) = n|d(f)|$. The value in this case is therefore around $600 \times 0.0017 \approx 1$. Since the data are in integers, a likely value for an error is 1, but it would have to be near the middle of the data to keep this value after tapering.

Figure 6.8 is a graph of the phase of the hanned transform, corrected for a slope of $-t_0 = -300$ and reduced (mod 1); that is, the j th ordinate is the residue (mod 1) (defined to lie in the interval from $-1/2$ to $1/2$) of $\phi_j + 300f_j$, where ϕ_j is the phase (measured in fractions of a cycle) of the j th term in the transform. Away from the peaks in the periodogram, this centered phase is almost identically 0 and is certainly a linear function of frequency. This confirms that the remaining term is of the "single error" form (although not finally; the behavior of its phase at the peak frequencies is still not known). Furthermore, centering the phase at $t_0 = 300$ has reduced the slope to 0, and therefore the error must be in x_{300} ,

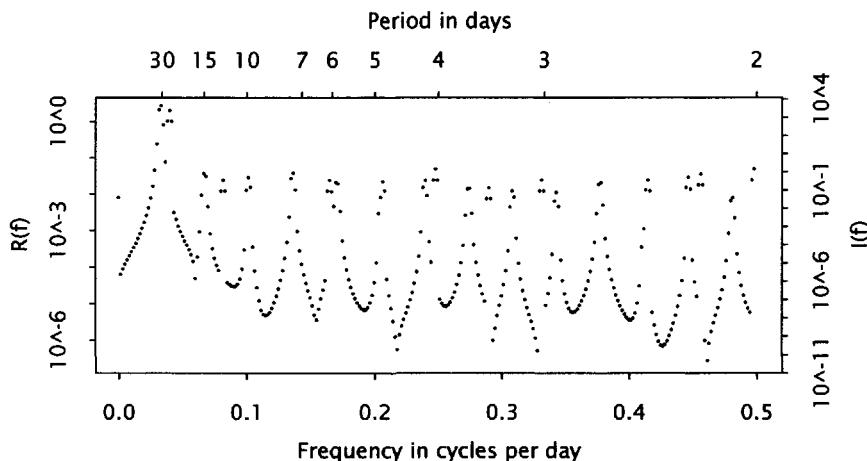


Fig. 6.9 Periodogram of the variable star data with x_{300} modified, tapered 100%.

the 301st observation. Finally, since the phases are all equal to 0, the “error” is positive. The only other possible value for the phase at $f = 0$ (and hence for $f \neq 0$ also) is $\pm 1/2$, which would correspond to a negative “error” (see Example (ii) of Chapter 4, p. 47).

Had the initial guess of $t_0 = 300$ been less lucky, there would have been a remaining slope identifying the necessary adjustment. Thus had $t_0 = 295$ been the choice, the resulting slope of -5 would have made the phase drift down from 0 at $f = 0$ to $-1/2$ at $f = 0.1$, and then show two complete cycles from $1/2$ to $-1/2$ as f increases to $f = 1/2$.

Figure 6.9 shows the periodogram of the data with $x_{300} = 19$ replaced by $x_{300} - 1 = 18$. The troughs now fall to values of around 10^{-8} to 10^{-10} , where previously they were supported by the “floor” at roughly 10^{-5} . The difference between this graph and Figure 6.7 confirms that x_{300} was in some sense perturbed.

It is interesting that the smallest possible perturbation in a single observation (a change of 1 in the least significant digit) should have such a visible effect on the analysis. This is possible only because the rest of the data have such a strongly periodic character; however, a larger perturbation could similarly affect the transform of less highly structured data. This sensitivity to a few large “errors” may be traced to the least squares interpretation of the periodogram developed in Chapter 2, or equivalently to the fact that the discrete Fourier transform is a linear function of the data. Any analysis that was less sensitive to errors would necessarily be nonlinear in the data.

Exercise 6.6 Effect of Tapering on Periodogram Peaks

Consider a sinusoid

$$x_t = e^{2\pi i f_0 t}$$

tapered by w_t , $0 \leq t < n$, to give $y_t = w_t x_t$. In the light of Exercise 6.2, all that must be assumed about the taper is that its transform is nonzero at $f = 0$ and small elsewhere.

- (i) Find the ratio of the heights of the peaks at $f = f_0$ in the periodograms of $\{x_t\}$ and $\{y_t\}$. In what way should data windows be normalized if unbiased estimates of amplitude are desired?
- (ii) The curvature of the peak can be measured by the second derivative of the periodogram at $f = f_0$. Show that for $\{y_t\}$ this is

$$-\left(\frac{8}{n^2}\right)\left\{\sum(t - \bar{t})^2 w_t\right\} \sum w_t,$$

where

$$\bar{t} = \frac{\sum t w_t}{\sum w_t}$$

is the average value of t with respect to the weights w_t [usually because of symmetry this is $(n - 1)/2$]. Show that this curvature is less than the corresponding curvature for the untapered $\{x_t\}$, when the normalization in (i) has been used (i.e., when the peaks are of the same height).

6.4 WOLF'S SUNSPOT NUMBERS

Wolf's sunspot numbers are an index of surface activity of the sun. They are a much analyzed set of data for which no completely satisfactory explanation exists (see, e.g., Bray and Loughhead, 1964; Newton, 1958). It has been shown that solar activity has an impact on many terrestrial phenomena, especially the earth's magnetic field, and its climate. Figure 1.3 (p. 4) shows the annual averages of the sunspot index for the years 1700 to the present. The succession of peaks and troughs shows that a definitely periodic phenomenon is at work in the data. For a variety of reasons from scientific curiosity to the need for forecasting future peaks of activity, it is desirable to describe this periodicity as accurately as possible.

The periodogram of these data is shown in Figure 6.10. The mean of the data was subtracted out, and the data were then tapered 5% at each end. The tapering covers 13 years at each end of the data, or a

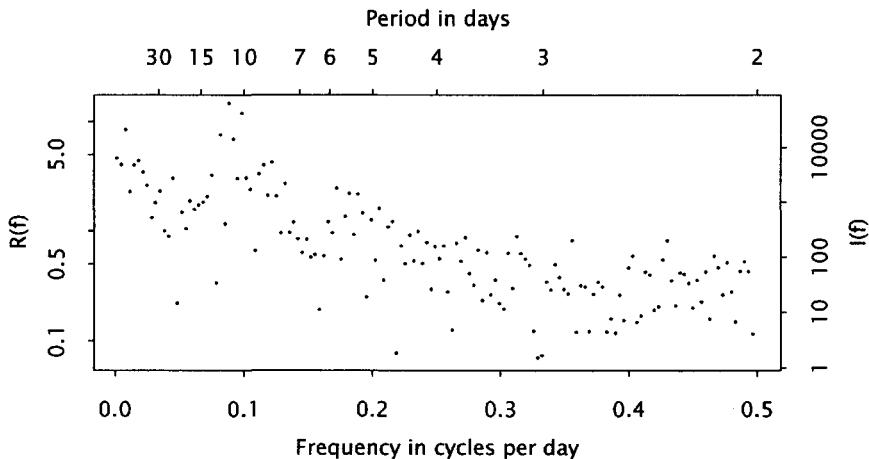


Fig. 6.10 Periodogram of the annual sunspot numbers for $0 < f \leq 1/2$.

little more than one cycle. This may seem rather small in the light of the improvement made by much heavier tapering in Section 6.3. However, the graph shows that there is, in fact, no strong peak in the periodogram, and relatively little tapering should suffice. As usual, the periodogram is displayed logarithmically for frequencies $0 < f \leq 1/2$.

This periodogram shows a much less strongly defined periodicity than did any of the periodograms of the variable star data in Sections 6.1 and 6.3. The largest ordinate is at $f_{21} = 21/235 = 0.08936$ cycles per year, corresponding to a period of $235/21 \approx 11.2$ years, the commonly quoted "period" of sunspots. However, this ordinate is not dramatically larger than its neighbors, whose magnitude cannot be explained by leakage. There is instead a broad peak of indeterminate width, extending roughly from $f = 0.06$ to $f = 0.12$ cycles per year. Several very low frequencies show almost the same power. Thus it would be misleading to name any single frequency as dominant.

Is there anything in the original data that could warn that harmonic analysis will not give the desired clear picture? Recall that a harmonic analysis decomposes the data into sinusoidal terms. The oscillations in these data do, in fact, show a number of departures from purely sinusoidal behavior. The first is that the sequence of peaks and troughs is not completely regular. For instance, there is no peak between 1787 and 1802, a gap of 15 years, whereas peaks in 1761 and 1769 are separated by only 8 years. The second visible departure involves the amplitude of the oscillations. All the troughs occur at around the same level, but the heights of

the peaks vary widely. A third feature which may also be seen under close examination is that the individual oscillations are not purely sinusoidal. Typically there is a sharp, well-defined peak and a broad trough, and it often appears that the rise is steeper than the drop (Newton, 1958, pp. 529–554).

The next two sections will describe the general effect on a harmonic analysis of such departures from sinusoidal behavior. Section 6.7 examines the extent to which the oscillations may be made more sinusoidal by a suitable transformation.

6.5 NONSINUSOIDAL OSCILLATIONS

The theory of Fourier series shows that any periodic function may be represented as the sum of a (usually infinite) series of sine and cosine functions. For the discrete Fourier transform described in Chapter 4 there is a corresponding result, given in Example (vi) of Chapter 4 (p. 51). It states that, if a series $\{x_t\}$ is periodic with period h , its transform is nonzero only at the frequency $1/h$ and its multiples.² The frequency $1/h$ is the *fundamental* frequency of the oscillation, and its multiples are *harmonics*.

Some aspects of nonsinusoidal behavior in the sunspot series (Figure 1.3, p. 4) were noted in Section 6.4. The first is that the troughs tend to be flatter than the peaks, and the second is that the rises tend to be slightly steeper than the falls. Figure 6.11 shows two periodic functions, each displaying one of these phenomena. The upper curve is $\cos 2\pi u + 1/4 \cos 4\pi u$ and the lower is $\cos 2\pi u - 1/8 \sin 4\pi u$. Thus each contains just its fundamental and second harmonic frequencies.

This suggests that the periodogram of the sunspot data (Figure 6.10) should show some power at the second harmonic of the fundamental frequency. There is indeed some evidence of a second, ill-defined peak at around $f = 0.2$ cycles per year, but it is very slight. Periodogram smoothing techniques, described in Chapter 8, give a clearer picture of these broad peaks.

The periodogram provides information only about the *magnitude* of the discrete Fourier transform. In the present case, the *direction* of the

²To be precise, for discrete time series, periodicity may be defined only for *integer* periods h , and the result holds only if the series length n is a multiple of the period. However, when the data are obtained by sampling a continuous time phenomenon, a noninteger period has an obvious interpretation and usefulness. If n is not a multiple of the period, a corresponding approximate result holds.

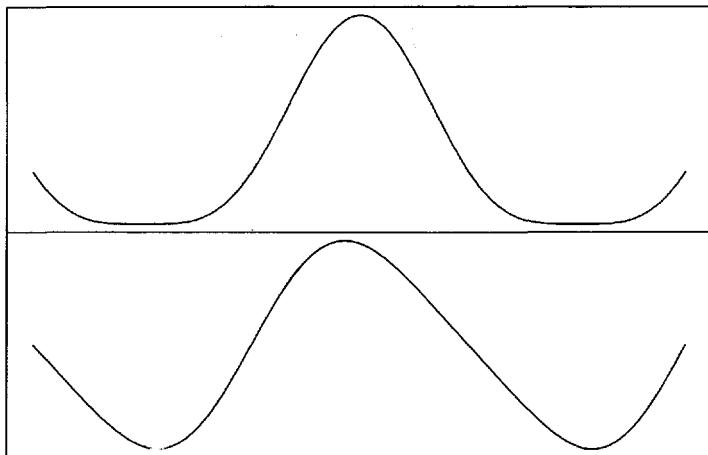


Fig. 6.11 Two periodic functions: $\cos 2\pi u + 1/4 \cos 4\pi u$ (upper curve) and $\cos 2\pi u - 1/8 \sin 4\pi u$ (lower curve).

departure from sinusoidal behavior should be evident in the phase of the transform. Both of the functions in Figure 6.11 may be written as $\cos 2\pi u + R \cos 2\pi(2u + \psi)$, where the phase ψ has the values 0 and $1/4$ cycles, respectively. An intermediate phase gives a function showing both types of behavior (see Exercise 6.7). If the scale of the u axis is changed by a factor f_0 and the u -origin is changed to $u = u_0$ from $u = 0$, the common form $\cos 2\pi u + R \cos 2\pi(2u + \psi)$ becomes

$$\begin{aligned} \cos(2\pi f_0(u - u_0) + R \cos 2\pi\{2f_0(u - u_0) + \psi\}) \\ = \cos 2\pi(f_0u + \phi_1) + R \cos 2\pi(2f_0u + \phi_2), \end{aligned}$$

say, where $\phi_1 = -f_0u_0$ and $\phi_2 = -2f_0u_0 + \psi$ are the phases of the fundamental and the second harmonic, respectively. The intrinsic shape of the function is, of course, not affected by such changes, and is still characterized by ψ , which may be obtained from ϕ_1 and ϕ_2 as $\psi = \phi_2 - 2\phi_1$, the *relative phase* of the second harmonic to the fundamental.

Since the sunspot periodogram shows a broad peak rather than a single frequency, it seems appropriate to examine this relative phase for each frequency in a band covering the peak. The frequencies chosen were the Fourier frequencies $f_j = j/n$ satisfying $0.06 < f_j < 0.12$ cycles per year. If the transform at frequency f_j is $d(f_j) = R(f_j)e^{2\pi i\phi_j}$, then the relative phase is $\psi_j = \phi_{2j} - 2\phi_j$.

A natural way to present these phases graphically is as points on a circle. This corresponds to plotting, in the complex plane, the points $e^{2\pi i\psi_j}$

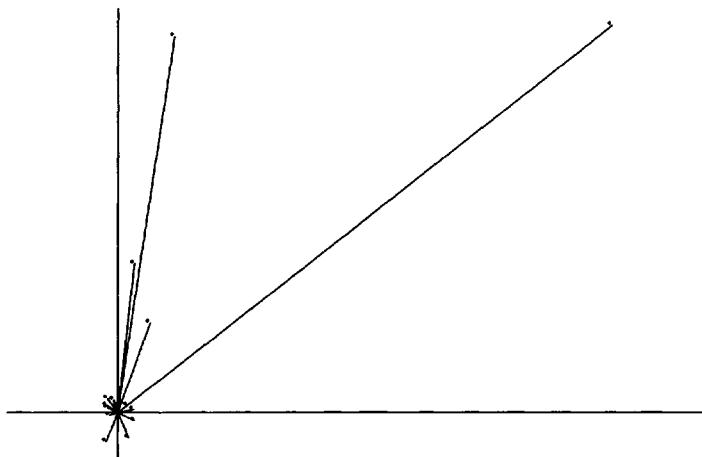


Fig. 6.12 Some values of the third order periodogram of the sunspot data, showing the relative phase of the second harmonic.

for which $0.06 < f_j < 0.12$. Since interest is focused on frequencies for which the fundamental and second harmonic are strong, it is desirable for the plot also to contain some information about the magnitudes of the transform at these frequencies. Displacing each point radially by a suitable amount achieves this goal. A convenient amount is $R(f_j)^2 R(f_{2j})$, since

$$\begin{aligned} R(f_j)^2 R(2f_{2j}) e^{2\pi i \psi_j} &= R(f_j)^2 e^{-4\pi i \phi_j} R(f_{2j}) e^{2\pi i \phi_{2j}} \\ &= \overline{d(f_j)}^2 d(2f_j), \end{aligned} \quad (6.10)$$

permitting a simple calculation of the complex values to be graphed. Figure 6.12 shows the resulting graph. The cluster of points near the origin corresponds to frequencies where either the fundamental or second harmonic is weak. The only point at any distance from the origin is in the quadrant suggested by Exercise 6.7.

Function 6.10 is a special case of the *third order periodogram* $d(f_1, f_2)$, which is proportional to $d(f_1)d(f_2)\overline{d(f_1 + f_2)}$. Higher order periodograms may, in fact, be defined to an arbitrarily high degree in a similar way, although they are rarely used. Brillinger and Rosenblatt (1967b) describe a similar but more extensive analysis of monthly sunspot data. The quantities they compute are smoothed third and fourth order periodograms, computed by techniques similar to those described in Chapter 8 for smoothing the conventional (or, in this terminology, *second order*) periodogram.

Exercise 6.7 Relative Phase of the Second Harmonic

Consider the function

$$\cos 2\pi u + \frac{a}{4} \cos 4\pi u - \frac{b}{8} \sin 4\pi u,$$

where $a, b > 0$ and $a + b = 1$.

- (i) Show that this function may be written in the form

$$\cos 2\pi u + R \cos 2\pi(2u + \psi),$$

where $0 < \psi < 1/4$.

- (ii) Show that this function combines the features of the two curves in Figure 6.11. (*Hint:* Evaluate the first derivative at $u = \pm 1/4$ and the second derivative at $u = 0$ and $u = \pm 1/2$.)

6.6 AMPLITUDE AND PHASE FLUCTUATIONS

The simplest example of a cosine wave with a fluctuating amplitude is the phenomenon of *beats*. If two cosine waves with nearly equal frequencies $f \pm \delta f$ are superimposed, the result is

$$\cos 2\pi(f + \delta f)t + \cos 2\pi(f - \delta f)t = 2 \cos 2\pi ft \cos 2\pi \delta f t.$$

This oscillates at the average frequency frequency f , but the amplitude changes slowly according to the *modulating function* $\cos 2\pi \delta f t$. The period of the modulating function is $1/\delta f$, which is large if δf is small. Conversely, if the modulated cosine wave is decomposed by Fourier analysis, the apparent frequency f splits into the pair of original frequencies $f \pm \delta f$. In this extreme case the transform is, in fact, *zero* at the apparent frequency f .

In the case of the sunspot data (Figure 1.3, p. 4), the modulation is not as simple as this. In particular, the amplitude seems to vary about a positive value, rather than about zero, as in the case of beats. In addition, it appears that the frequency of the oscillations is not constant, but similarly changes slowly with time. Thus the data might be represented approximately as

$$x_t = R_t \cos 2\pi(f_t t + \phi), \quad (6.11)$$

where R_t and f_t are slowly varying sequences.³ If f_t varies about some

³The term *slowly varying* is used here with no precise mathematical definition in mind.

typical value f_0 , the last equation may also be written

$$\begin{aligned} x_t &= R_t \cos 2\pi \{f_0 t + \phi + (f_t - f_0)t\} \\ &= R_t \cos 2\pi(f_0 t + \phi_t), \end{aligned} \quad (6.12)$$

say, in which the frequency is constant and the phase $\phi_t = \phi + (f_t - f_0)t$ varies. The frequency f_0 need not be a typical value of f_t to carry out this rearrangement, but it makes more sense if this is the case, since otherwise the phase ϕ_t would show a systematic drift.

The complex analog of (6.12) is

$$\begin{aligned} x_t &= R_t e^{2\pi i(f_t + \phi_t)} \\ &= z_t e^{2\pi i f_t}, \end{aligned} \quad (6.13)$$

say, where $z_t = R_t e^{2\pi i \phi_t}$ is a complex-valued modulating function. Expression (6.13), a *modulated sinusoid*, is formally identical to the tapered transform of Exercise 6.2, and therefore its transform consists of that of $\{z_t\}$ centered about frequency f_0 . Specifically,

$$\begin{aligned} d_x(f) &= \frac{1}{n} \sum x_t e^{-2\pi i f t} \\ &= \frac{1}{n} \sum z_t e^{-2\pi i(f-f_0)t} \\ &= d_z(f-f_0). \end{aligned}$$

Now $\{z_t\}$ varies slowly and smoothly, and hence its transform is large only for low frequencies (see Section 4.5). Therefore the transform of $\{x_t\}$ is large only for frequencies close to f_0 . However, the single spike that would be seen for a pure (unmodulated) sinusoid is *split* or *spread* into a displaced image of the transform of $\{z_t\}$. Typically, the transform of a modulating function $\{z_t\}$ would show an irregular peak surrounding $f = 0$, and in particular would rarely vanish at $f = 0$. Correspondingly, the transform of $\{x_t\}$ is usually nonzero at $f = f_0$, although this may not be the largest value.

In the sunspot data, four or five intervals may be identified where peaks tend to be large, separated by intervals where they are smaller. The phase variation is harder to evaluate by eye. If the phase is assumed to be changing no more rapidly than the amplitude, the transform of the (complex) modulating function might be expected to be large only for frequencies in the interval $-5/n < f < 5/n$ cycles per year, or, since $n = 261$ years, $-0.02 < f < 0.02$ cycles per year. Thus the periodogram of the sunspot numbers might be expected to show a peak with a width of around 0.04 cycles per year. Figure 6.10 does, in fact, show a peak of roughly this

width, suggesting that the poor definition of that peak may well be caused by amplitude and phase variations. *Complex demodulation*, described in Chapter 7, is a useful tool for analyzing data containing amplitude or phase variations.

6.7 TRANSFORMATIONS

Sections 6.5 and 6.6 described two common ways in which a periodic series may fail to be purely sinusoidal. Both have the effect that the transform of the series has power at frequencies other than the fundamental frequency corresponding to the period of the series. To this extent, there is information about the periodicity of the series in its discrete Fourier transform at frequencies other than the fundamental. In particular, no single ordinate of the transform contains all of the available information.

A series may sometimes be transformed in such a way that it becomes more closely sinusoidal. This means that a greater proportion of the power associated with the periodicity appears at the fundamental frequency in the transform, and hence the fundamental becomes more pronounced. For example, the upper curve in Figure 6.11 is

$$\begin{aligned}\cos u + \frac{1}{4} \cos 2u &= \cos u + \frac{1}{4} \{2(\cos u)^2 - 1\} \\ &= \frac{1}{2}(1 + \cos u)^2 - \frac{3}{4}.\end{aligned}$$

Thus if the periodic signal were

$$\frac{3}{4} + \cos u + \frac{1}{4} \cos 2u,$$

which, like the sunspot data, is nonnegative and has zero as its lowest value, it could be transformed into a pure sinusoid by the square root transformation.

As a second example, suppose that a series is $x_t = y_t^2$, say, where

$$y_t = R \cos 2\pi(ft + \phi) + z_t$$

and $\{z_t\}$ is a slowly varying series. Then

$$\begin{aligned}x_t &= R^2 \cos 2\pi(ft + \phi)^2 + 2z_t R \cos 2\pi(ft + \phi) + z_t^2 \\ &= \frac{1}{2}R^2 \{\cos 4\pi(ft + \phi) + 1\} + 2z_t R \cos 2\pi(ft + \phi) + z_t^2,\end{aligned}\quad (6.14)$$

which consists of a term with frequency $2f$, a slowly varying term $R^2/2 + z_t^2$, and an amplitude-modulated term

$$2z_t R \cos 2\pi(ft + \phi).$$

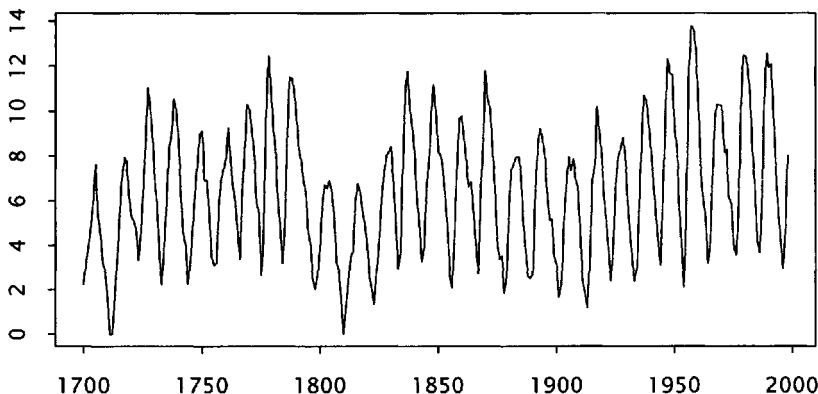


Fig. 6.13 Square roots of the annual sunspot numbers, 1700 to 1960.

Thus, in this case, the square root transformation would eliminate both the term with frequency $2f$ (the second harmonic) and the amplitude modulation. Of course, the original data must have the special structure (6.14) for this to be the case.

Notice that a transformation of the *values* of a series can do nothing to change *time* asymmetry in a waveform, such as that shown by the lower curve of Figure 6.11. Similarly, variations in the phase of an otherwise periodic signal cannot be changed. These features could be removed only by a transformation of the time axis, which would, in general, be difficult both to find and to interpret.

Figure 6.13 shows the square roots of the sunspot numbers. The tendency for the troughs to be flatter than the peaks has been largely removed and possibly partly reversed. The values at the troughs show greater variation than before, but still not as great as that of the heights of the peaks. It also appears that the amplitude modulation has been reduced and a low-frequency component introduced, as in the example above. These observations suggest that the power at the second harmonic should be reduced, and that the main peak should be more clearly defined.

Figure 6.14 shows the periodogram of square roots, which is, in fact, very similar to the original periodogram, Figure 6.10 (p. 77). Close inspection suggests that the power at the second harmonic has been reduced slightly, but the main peak is no clearer. The phase fluctuations, which are not affected by this transformation, and the remaining amplitude fluctuations are sufficiently strong for the peak still to be diffuse rather than

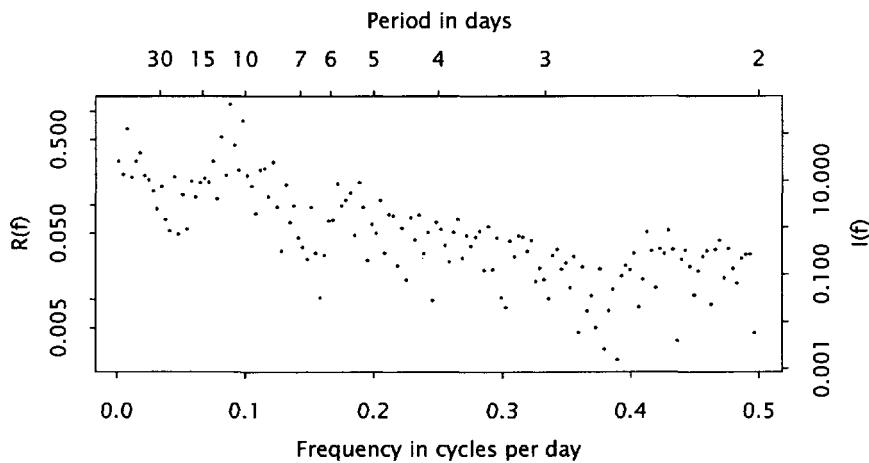


Fig. 6.14 Periodogram of the square roots of the sunspot numbers.

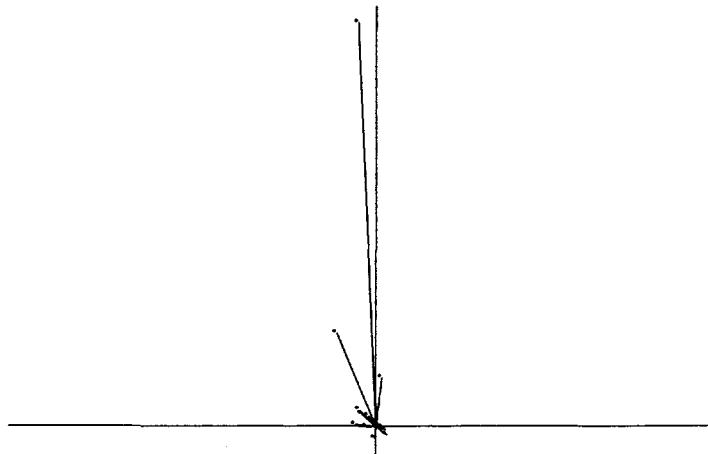


Fig. 6.15 Some values of the third-order periodogram of the square roots of the sunspot numbers, showing the relative phase of the second harmonic.

sharp.

Figure 6.15 is the relative phase plot for the second harmonic of the frequencies f_j between 0.06 and 0.12 cycles per year, computed as described in Section 6.5. Only one point falls at a large radius, corresponding to $f = 0.08936$ cycles per year, and this point falls much closer to the imaginary axis than before. This confirms that the square root trans-

formation has substantially reduced the difference between the nature of peaks and troughs, while leaving the time asymmetry relatively unchanged.

The most generally useful transformations are the power transformations $y = x^\alpha$ and the logarithmic transformation $y = \log x$, which arises as the limit as α approaches 0 of $(x^\alpha - 1)/\alpha$. These may be applied only to nonnegative data (positive data if $\alpha \leq 0$). However, the data most in need of transformation are usually nonnegative, so this is not a serious restriction. In certain other cases, other transformations may be suitable. In the case of the sunspot data, it is doubtful whether another transformation would bring about a substantial improvement. The amplitude modulation was not entirely eliminated, and it appears that a stronger transformation, perhaps cube roots, is needed for this. On the other hand, the difference between the natures of the peaks and troughs has been largely eliminated, and a stronger transformation would presumably reintroduce a difference in the opposite direction.

Exercise 6.8 Effect of a Transformation

Suppose that $\{y_0, y_1, \dots, y_{n-1}\}$ is obtained from $\{x_t\}$ by a transformation $y_t = f(x_t)$, with inverse $x_t = g(y_t)$. Suppose also that

$$y_t = \mu + R \cos 2\pi(ft + \phi),$$

where $f = j/n$ is the j th Fourier frequency.

- (i) Suppose, in addition, that j divides n . Show that the transform of $\{x_t\}$ is nonzero only at multiples of f .
- (ii) Suppose, more generally, that the greatest common divisor of j and n is l . Show that the transform of $\{x_t\}$ is nonzero only at multiples of $f_l = l/n$. (*Hint:* Find the period of $\{y_t\}$ and hence of $\{x_t\}$, and note that the period of a discrete time series must be an integer.)

Exercise 6.9 Continuation

Suppose that $\{x_t\}$ and $\{y_t\}$ are related as in Exercise 6.8, but that

$$y_t = \mu + R \cos 2\pi(ft + \phi) + R' \cos 2\pi(f't + \phi'),$$

where $f' = j'/n$ is another Fourier frequency. Show that the transform of $\{x_t\}$ is nonzero only at multiples of f_l , where l is the greatest common divisor of j , j' , and n .

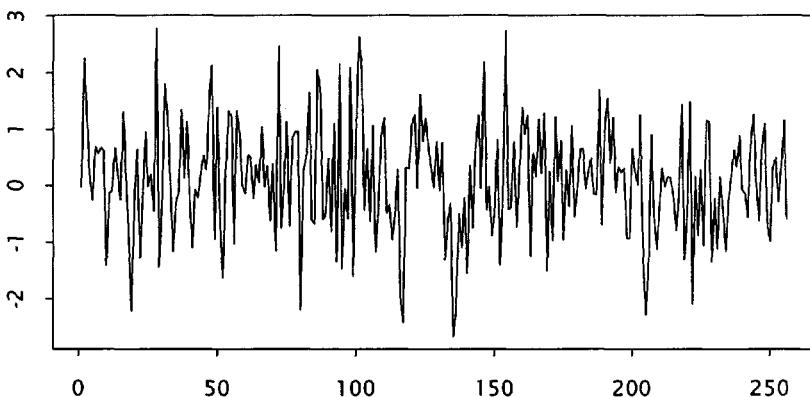


Fig. 6.16 A series of 256 pseudorandom numbers drawn from the standard normal distribution.

Exercise 6.10 An Approximation

Suppose that $\{x_t\}$ and $\{y_t\}$ are as in Exercise 6.9, and that R and R' are small compared with μ . Use a Taylor series expansion about μ to show that in the transform of $\{x_t\}$ there are first-order terms at f and f' , and second-order terms at $2f$, $2f'$, and $|f \pm f'|$.

6.8 PERIODOGRAM OF A NOISE SERIES

One feature common to most of the periodograms shown in this chapter is the roughness of the graphs in any interval in which there is no peak. The exceptions were the periodograms of the heavily tapered variable star data, which showed an unusual constancy (Figures 6.6, p. 73, and 6.7, p. 73). Even in that case, however, the characteristic roughness reappeared when the exceptional value was corrected (Figure 6.9). This behavior is typical of periodograms of empirical data and is usually caused by the presence of a random component in the data.

The simplest such component is a series of independent random numbers. Figure 6.16 shows 256 pseudorandom numbers drawn from the normal (or Gaussian) distribution. Their periodogram, shown in Figure 6.17, shows the typical roughness, but otherwise has no interesting structure or overall shape. By analogy with the spectrum of white light, which is also flat and featureless, a series of independent errors is also known as

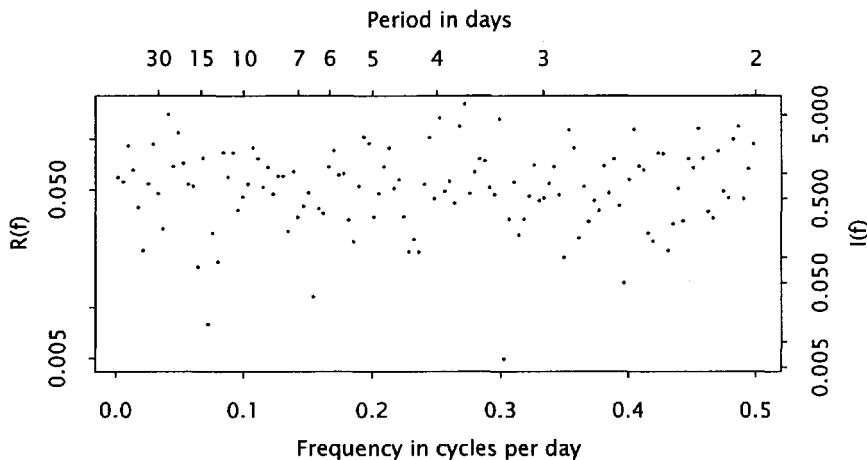


Fig. 6.17 Periodogram of the pseudorandom data.

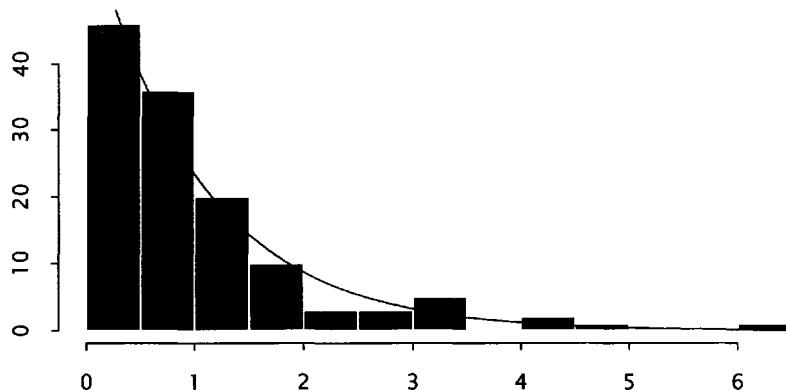


Fig. 6.18 Histogram of the periodogram of the pseudorandom data.

white noise.

Figure 6.18 is the histogram of this periodogram, with an exponential curve superimposed. The similarity is no coincidence. When the data are independent random numbers drawn from a Gaussian distribution, it may be shown that the periodogram ordinates at the Fourier frequencies (other than 0 and $1/2$ cycles per unit time) are independently exponentially distributed (see Exercise 6.11). It is this independence that makes the

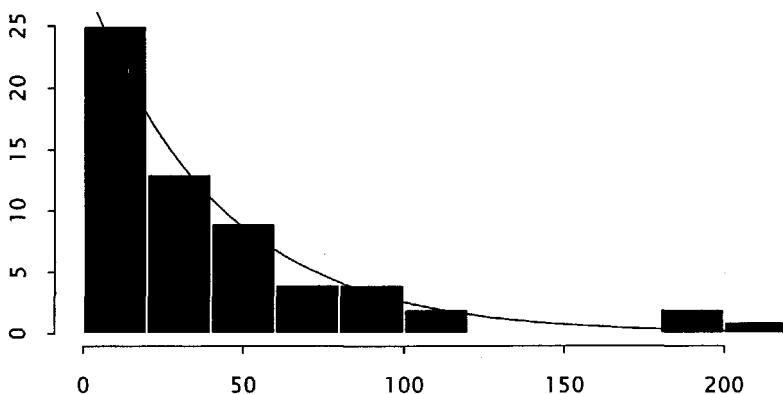


Fig. 6.19 Histogram of the periodogram of the annual sunspot numbers (shown in Figure 6.10) for $0.3 < f < 0.5$ cycles per year.

graph of the periodogram rough, since it means that there is no tendency for adjacent ordinates to have similar values.

When the periodogram is calculated at frequencies other than the Fourier frequencies (e.g., when it is calculated after extending the data by zeros), each ordinate is still approximately exponentially distributed, but neighboring ordinates are by no means independent (see Exercise 6.12). Tapering a series adds a further complication, since then the distribution of the periodogram ordinates is only approximately exponential even at the Fourier frequencies (see Exercise 6.13).

Figure 6.19 is the histogram of the sunspot periodogram (Figure 6.10, p. 77) for $0.3 < f < 0.5$ cycles per year, again with an exponential curve superimposed. Despite the fact that the data were tapered, no gross departure from the exponential distribution is evident. The number of ordinates used, 47, is too small to detect slight departures. No more ordinates could be used, however, as the periodogram begins to increase as f is reduced below $f = 0.3$ cycles per year. Inclusion of these ordinates would result in mixing the graphed distribution with exponential distributions having larger scale parameters.

The exponential distribution of periodogram ordinates has been derived here on the assumption of independent, normally distributed errors (or *noise*). The key point, however, is that the cosine and sine sums are normally distributed. If it is assumed only that the errors are independent and have finite variance, the central limit theorem (see, e.g., Feller,

1968, pp. 244, 254) assures that these sums are approximately normally distributed for large n . It follows that periodogram ordinates are still approximately exponentially distributed under this weaker assumption, and indeed in any situation where a central limit theorem holds.

Exercise 6.11 Periodogram of a Noise Series

Suppose that $\{x_0, x_1, \dots, x_{n-1}\}$ are independent and identically distributed with the standard normal distribution.

- (i) If f is a Fourier frequency and $0 < f < 1/2$, show that both

$$\sum_{t=0}^{n-1} x_t \cos 2\pi f t \quad \text{and} \quad \sum_{t=0}^{n-1} x_t \sin 2\pi f t$$

have zero expectation, that their variances are both $n/2$, and that their covariance is zero. It follows that the periodogram has a χ^2 distribution with 2 degrees of freedom, which is the exponential distribution.

- (ii) If f' is another Fourier frequency in the same interval, show that the cosine and sine sums involving f' are uncorrelated with those involving f . This lack of correlation, together with the normal distribution, implies independence of the sums, and hence of the periodogram values.

Recall that when the periodogram is calculated at the Fourier frequencies, prior subtraction of the series mean has no effect; thus the result of this exercise holds, regardless of whether the data are centered at their mean before computation of the periodogram.

Exercise 6.12 Periodogram at a Non-Fourier Frequency

Suppose the $\{x_t\}$ is as in Exercise 6.11, but that f is not a Fourier frequency. Show that the values found in Exercise 6.11 for the moments of the cosine and sine sums are still approximately valid, and deduce that the periodogram at f is still approximately exponentially distributed. (See also Exercise 2.6, p. 24.)

Exercise 6.13 Periodogram of a Tapered Noise Series

Suppose the $\{x_t\}$ is as in Exercise 6.11, and consider the cosine and sine sums of the series tapered by $\{w_t\}$,

$$\sum_{t=0}^{n-1} w_t x_t \cos 2\pi f t \quad \text{and} \quad \sum_{t=0}^{n-1} w_t x_t \sin 2\pi f t.$$

Suppose further that $w_t = w\{(t + 1/2)/n\}$ for some smooth function $w(x)$ defined on $0 \leq x \leq 1$. Find the variances and covariance of these sums, and show that they may be approximated by integrals involving $w(x)^2$. Use the fact that

$$\int_0^1 w(x)^2 \cos Ax dx \quad \text{and} \quad \int_0^1 w(x)^2 \sin Ax dx$$

converge to zero as $A \rightarrow \infty$ to deduce, as in Exercise 6.12, that the periodogram is approximately exponentially distributed.

6.9 FISHER'S TEST FOR PERIODICITY

The distributional results of the preceding section mean that it may be misleading to look for peaks in a periodogram. Since the ordinates at the Fourier frequencies are approximately independent, they are bound to fluctuate and show many small peaks and troughs. Furthermore, since the distribution is exponential, the largest ordinate, even for a noise series, tends to be large compared with its neighbors (see Exercise 6.14) and may appear to indicate a reasonably strong periodicity.

Fisher (1929) proposed a test of the significance of the largest peak in a periodogram and gave a table of critical values for various series lengths. The test statistic is the ratio of the largest of the periodogram ordinates at the Fourier frequencies to the sum of the ordinates. Fuller (1995, Table 7.1.2) gives a more extensive table for an equivalent statistic, the ratio of the largest ordinate to the average.

The null hypothesis in each case is that the data consist of independent errors (*white noise*). For an exact theory it is assumed that the errors are normally distributed (*Gaussian white noise*), but the discussion of Section 6.8 suggests that, if the Gaussian assumption is not satisfied, the theory should continue to provide a useful approximation.

This null hypothesis is often inappropriate, in that it may be clear from the data that even if no periodicity is present, the data do not consist of

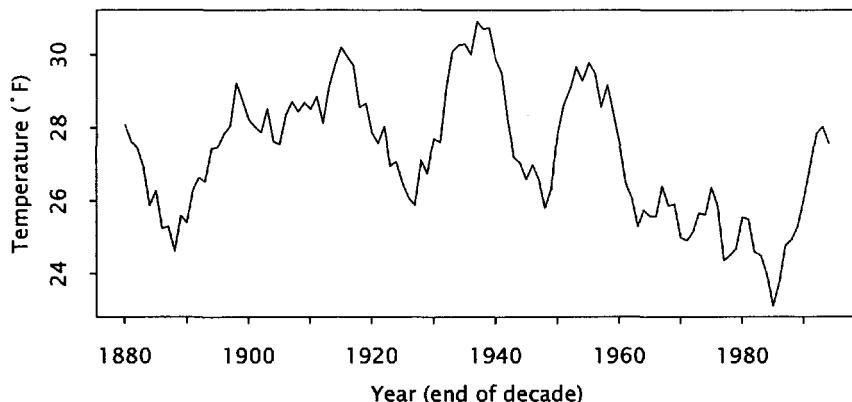


Fig. 6.20 Ten-year running means of mean January temperatures at Central Park Observatory, New York City, for 1870 to 1994 (Karl et al., 1990; Easterling et al., 1996).

independent errors. Whittle (1952) describes a modified procedure that is appropriate for testing a more general null hypothesis.

Fisher's test may be used to assess the significance of an apparent periodicity in the data shown in Figure 6.20, which are 10-year running means of January temperatures for 1870 to 1994 at Central Park Observatory, New York City. The data from 1869 to 1973 were discussed by Spar and Mayer (1973); the data shown here were obtained from the U.S. Historical Climatology Network (Revision 3) at the National Climatic Data Center (Karl et al., 1990; Easterling et al., 1996).⁴ There appears to be a strong 20-year cycle, especially in parts of the record.

Figure 6.21 shows the periodogram of the *original*, unsmoothed, January temperatures. The running means shown in Figure 6.20, a smoothed series, would have a periodogram with less power at high frequencies. The largest value in the periodogram occurs at the lowest frequency in the transform, and is associated with the overall trend. With this ordinate set aside, the next highest is at 0.048 cycles per year, corresponding

⁴The web site for the Historical Climate Network (HCN) is <http://cdiac.esd.ornl.gov/epubs/ndp019/ndp019.html>. The monthly mean temperatures for all HCN stations in New York state are maintained in a single file: <http://cdiac.esd.ornl.gov/r3d/us/hcn/state/NY/NY94mea.html>, in which each line beginning with the station code for the Central Park station, 305801, contains a single year's data. Each year's record contains the monthly means, quarterly means, and the annual mean, where the quarters are Winter (Dec, Jan, Feb), Spring (Mar, Apr, May), Summer (Jun, Jul, Aug), and Fall (Sep, Oct, Nov).

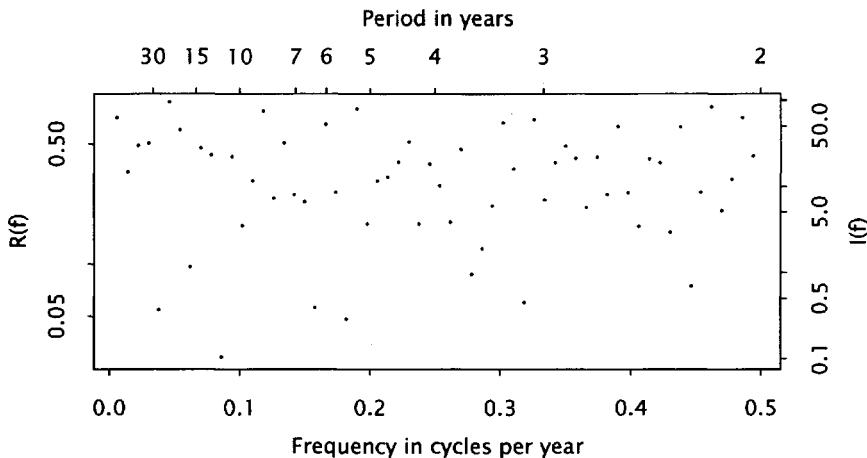


Fig. 6.21 Periodogram of unsmoothed mean January temperatures.

to a period of 20.8 years. The ordinate is 4.420 times the average of these $m = 61$ ordinates. The value of the test statistic for which Fuller's table is relevant is therefore 4.420. Linear interpolation between the entries for $m = 60$ and $m = 70$ yields an upper 10% point of 6.161. Thus the peak is far short of statistical significance. It follows that these data alone provide no statistical evidence for the presence of a 20-year cycle.

Exercise 6.14 Largest Periodogram Ordinate

Suppose that I_1, I_2, \dots, I_m are independent and exponentially distributed with mean σ^2 , that is,

$$\text{pr}(I_j/\sigma^2 \leq x) = 1 - e^{-x}, \quad x \geq 0, j = 1, 2, \dots, m.$$

Let $X_m = \max\{I_1, I_2, \dots, I_m\}$.

- (i) Show that

$$\text{pr}(X_m/\sigma^2 \leq x) = (1 - e^{-x})^m,$$

and hence that for large m ,

$$\text{pr}(X_m/\sigma^2 \leq x + \log m) \approx \exp(-e^{-x}).$$

This implies that X_m is probably close to $\sigma^2 \log m$.

- (ii) Show that for large m ,

$$Y_m = I_1 + I_2 + \dots + I_m \approx \sigma^2 m,$$

in the sense that the distribution of $Y_m/(\sigma^2 m)$ becomes concentrated arbitrarily close to the value 1.

- (iii) Deduce that if $\xi_m = X_m/(Y_m/m)$ is Fuller's version of the test statistic, then

$$\text{pr}(\xi_m \leq x + \log m) \approx \exp(-e^{-x})$$

or

$$\text{pr}(\xi_m \leq \xi) \approx \exp(-me^{-\xi}).$$

For the temperature data of Figure 6.20, $n = 125$ and there are 62 Fourier frequencies. Omitting the first leaves $m = 61$ ordinates to be tested. The observed value $\xi_{\text{obs}} = 4.420$ would be exceeded with a probability of approximately

$$1 - \exp(-61e^{-4.420}) = .520,$$

or 52% of the time, under the null hypothesis that the data are a pure noise series.

The approximation derived in this exercise may be used to obtain approximate percentage points:

$$\text{pr}\{\xi_m \leq \log m - \log(-\log q)\} \approx q,$$

which gives the approximate upper 10% point for $m = 61$ as 6.361 (cf. the interpolated value 6.161).

Exercise 6.15 A Better Approximation

Chiu (1989) has studied the properties of a family of tests related to Fisher's, and has shown that the following argument provides a better approximation to the distribution of ξ_m . Write

$$Y'_m = Y_m - X_m,$$

the sum of all the ordinates except the largest. Chiu shows that

$$Y'_m \approx \sigma^2(m - 1 - \log m)$$

and hence that

$$\xi'_m = \frac{X_m(m - 1 - \log m)}{Y'_m}$$

has the same approximate distribution as ξ_m :

$$\text{pr}(\xi'_m \leq \xi') \approx \exp(-me^{-\xi'}).$$

Since

$$\xi'_m = \xi_m \frac{(m - 1 - \log m)}{m - \xi_m}$$

it follows that

$$\text{pr}(\xi_m \leq \xi) \approx \exp(-me^{-\xi'}),$$

where

$$\xi' = \xi \frac{(m - 1 - \log m)}{m - \xi}.$$

Chiu compared this approximation with the exact distribution obtained by Fisher and tabulated by Fuller, and found that it provides a surprisingly accurate approximation. For these data the value $\xi_{\text{obs}} = 4.420$ corresponds to $\xi'_{\text{obs}} = 4.366$, which would be exceeded approximately 53.9% of the time for pure noise, leaving the conclusion unchanged. This more accurate approximation gives the upper 10% point of the distribution of ξ as 6.233, which is indeed closer to the (interpolated) exact value of 6.161 than the value 6.361 found in the previous exercise.

Appendix

The S-Plus function `spec.pgram()` is a general purpose tool for estimating spectra (Chapter 8) based on smoothing the periodogram, and will return the periodogram of tapered data as a default. The function `fft()` is designed for the more specific purpose of calculating the discrete Fourier transform $d(f)$, and was used for the calculations in this chapter. This function returns the *unnormalized* transform, that is, without the n^{-1} multiplier used here. Thus a typical usage would be

```
star.dft <- fft(star)/length(star)
```

Split cosine bell tapers were computed from the function

```
taper <- function(n, p)
{
  p <- max(0, min(1, p))
  x <- (2 * (1:n) - 1)/(2 * n)
  (1 - cos(2 * pi * pmin(x/p, 1/2, (1 - x)/p)))/2
}
```

which could be used as in

```
star.dft50 <- fft((star - mean(star)) * taper(n =
  length(star), p = 0.5))/length(star)
```

96 EXAMPLES OF HARMONIC ANALYSIS

The amplitude function $R(f) = |d(f)|$ was obtained using the S-PLUS function Mod():

```
star.amp50 <- Mod(star.dft50)
```

and the periodogram $I(f) = n|d(f)|^2 = nR(f)^2$ was found similarly:

```
star.pgm50 <- length(star) * Mod(star.dft50)^2
```

The third-order periodograms shown in Figures 6.12 and 6.15 can be found using, for instance,

```
n <- length(sunspots)
spot.dft10 <- fft((sunspots - mean(sunspots))*taper(n,
  0.1))/n
spot.dft10 <- spot.dft10[-1] # drop the zero-frequency term
f <- (1:(n - 1))/n
i <- (1:(n - 1))[f > 0.06 & f < 0.12]
spot.3rd <- d[2 * i] * Conj(d[i])^2
```

7

Complex Demodulation

Harmonic analysis of the sunspot series in Chapter 6 showed that not all “periodic” phenomena have simple representations in terms of cosine functions, even when much has been done to improve the analysis. *Complex demodulation* is a more flexible approach to the analysis of such data (Bingham et al., 1967). By trading off some *frequency resolution* for *time resolution*, complex demodulation can describe features of data that would be missed by harmonic analysis, and also to verify in some cases that no such features exist. The price of this flexibility is a loss of precision in describing pure frequencies, for which harmonic analysis is most exact.

7.1 INTRODUCTION

Suppose that a set of data contains a perturbed periodic component

$$x_t = R_t \cos(2\pi(f_0 t + \phi_t)). \quad (7.1)$$

Here $\{R_t\}$ is a slowly changing amplitude, and $\{\phi_t\}$ is a slowly changing phase. It was shown in Section 6.4 that the oscillations in the sunspot series could reasonably be regarded as having the structure (7.1). The aim of complex demodulation is to extract approximations to the series $\{R_t\}$ and $\{\phi_t\}$. It may be regarded as a *local* version of harmonic analysis; it is analogous to harmonic analysis in that it seeks to describe the amplitude

and phase of an oscillation, but it is local in that these are allowed to change slowly over time.

How rapidly the amplitude and phase are allowed to change is the issue of time *resolution*, and is one of the key choices in carrying out complex demodulation. It is a matter of compromise: allowing more rapid changes, that is, increasing the time resolution, inevitably reduces the frequency resolution.

Consider first the complex analog of (7.1),

$$x_t = R_t e^{2\pi i(f_0 t + \phi_t)}. \quad (7.2)$$

The extraction of $\{R_t\}$ and $\{\phi_t\}$ is trivial if f_0 is known, as is assumed, since then it is possible to construct

$$\begin{aligned} y_t &= x_t e^{-2\pi i f_0 t} \\ &= R_t e^{2\pi i \phi_t}. \end{aligned}$$

Then

$$R_t = |y_t| \quad \text{and} \quad e^{2\pi i \phi_t} = \frac{y_t}{|y_t|}.$$

The new series $\{y_t\}$ is said to be obtained from $\{x_t\}$ by *complex demodulation*. Now the real form (7.1) may be written as

$$x_t = \frac{1}{2} R_t \left\{ e^{2\pi i(f_0 t + \phi_t)} + e^{-2\pi i(f_0 t + \phi_t)} \right\}$$

and is thus the sum of two complex terms, the first similar to (7.2) and the second its complex conjugate. This real series may be analyzed in one of two ways. The first is to use complex demodulation initially ignoring the second term, and then to remove it using *linear filtering* techniques. The second way is to separate the two terms and then to use complex demodulation on either. The separation is also based on filtering, so the two ways are only superficially different. However, they suggest different directions of development, and will be presented separately. The former approach is described below, and the latter in Section 7.6.

Complex demodulation of the real series (7.1) yields

$$y_t = \frac{1}{2} R_t e^{2\pi i \phi_t} + \frac{1}{2} e^{-2\pi i(2f_0 t + \phi_t)}.$$

The first term is the desired one, since as before $\{R_t\}$ and $\{\phi_t\}$ are easily extracted. The second term, which is a perturbed complex sinusoid with frequency $-2f_0$, has to be removed.

In general, the series being analyzed does not consist solely of the perturbed sinusoid (7.1). For instance, it was shown in Sections 6.4 and 6.7

that the sunspot series contains some low frequency terms, especially after transformation, and in Section 6.5 that it contains some terms that may be identified as second harmonics. It is also clear from the graph of the series, Figure 1.3 (p. 4), that there is a noise component. Thus, more generally, the data series would be

$$x_t = R_t e^{2\pi i(f_0 t + \phi_t)} + z_t$$

where z_t contains the additional terms, and hence

$$\begin{aligned} y_t &= x_t e^{-2\pi i f_0 t} \\ &= \frac{1}{2} R_t e^{2\pi i \phi_t} + \frac{1}{2} e^{-2\pi i(2f_0 t + \phi_t)} + z_t e^{-2\pi i f_0 t}. \end{aligned} \quad (7.3)$$

The basic problem in complex demodulation is to separate the first term in (7.3) from the others. The feature that makes this possible is that, since both $\{R_t\}$ and $\{\phi_t\}$ are assumed to be smooth, the first term likewise is smooth. The second term oscillates at a frequency around $-2f_0$. All frequencies in the final term are shifted by $-f_0$ by the demodulation (see Exercise 7.1). Now $\{z_t\}$ may be assumed to have no component at frequency f_0 , since any such component would be indistinguishable from $R_t \cos 2\pi(f_0 t + \phi_t)$. Thus $\{z_t e^{-2\pi i f_0 t}\}$ contains no component around zero frequency, and hence is not smooth (see Section 4.5). The problem, therefore, is to extract the smooth component of $\{y_t\}$. This is usually accomplished by *linear filtering*, which is described briefly in the next section.

Exercise 7.1 Transform of a Demodulated Series

For any series $\{x_t\}$, the demodulated series

$$y_t = x_t e^{-2\pi i f_0 t}, \quad t = 0, 1, \dots, n - 1$$

is formally the same as the tapered sinusoid of Exercise 6.2. Verify that the transform of $\{y_t\}$ consists of the transform of $\{x_t\}$ centered at f_0 , that is,

$$\begin{aligned} d_y(f) &= \frac{1}{n} \sum y_t e^{-2\pi i f t} \\ &= \frac{1}{n} \sum x_t e^{-2\pi i(f + f_0)t} \\ &= d_x(f + f_0). \end{aligned}$$

In this sense, complex demodulation just shifts all the frequencies by $-f_0$.

7.2 SMOOTHING: LINEAR FILTERING

The problem just encountered arises in many contexts: given a series that consists of a smooth function, the *signal*, plus disturbance or *noise*, how can the two components be separated? The question is essentially how to smooth the series. *Linear filters* are often used for this purpose. Since these filters have been discussed extensively elsewhere (see, e.g., Otnes and Enochson, 1972; Hamming, 1998), only a brief account is given here.

Suppose that the series $\{y_t\}$ may be written as

$$y_t = a_t + e_t,$$

where $\{a_t\}$ is smooth and $\{e_t\}$ represents errors or disturbances. Since $\{a_t\}$ is smooth, a_{t-1} and a_{t+1} are approximately the same as a_t . Thus the average of y_{t-1} , y_t , and y_{t+1} is approximately a_t plus the average of e_{t-1} , e_t , and e_{t+1} . However, these errors tend to cancel out, so that the average error tends to be smaller than the individual errors. If this averaging is carried out for each t , a new series is obtained, say $\{z_t\}$, which consists approximately of $\{a_t\}$ plus errors that tend to be smaller than before. Hence some progress has been made toward extracting the smooth series $\{a_t\}$.¹

The clearest way to describe the effect of this procedure, which is known as *simple moving averaging*, is through a frequency approach. Suppose first that the series $\{y_t\}$ is exactly sinusoidal,

$$y_t = R \cos 2\pi(ft + \phi).$$

Then

$$\begin{aligned} z_t &= \frac{1}{3}(y_{t-1} + y_t + y_{t+1}) \\ &= \frac{1}{3}R\{\cos 2\pi(ft - f + \phi) + \cos 2\pi(ft + \phi) + \cos 2\pi(ft + f + \phi)\}, \end{aligned}$$

which is most easily evaluated as the real part of

$$\begin{aligned} \frac{1}{3}R\{e^{2\pi i(ft-f+\phi)} + e^{2\pi i(ft+\phi)} + e^{2\pi i(ft+f+\phi)} \\ &= \frac{1}{3}Re^{2\pi i(ft+\phi)}\{e^{-2\pi if} + 1 + e^{2\pi if}\} \\ &= \frac{1}{3}Re^{2\pi i(ft+\phi)}(1 + 2 \cos 2\pi f). \end{aligned} \tag{7.4}$$

¹One point at each end of the series cannot be calculated without some modification of the procedure.

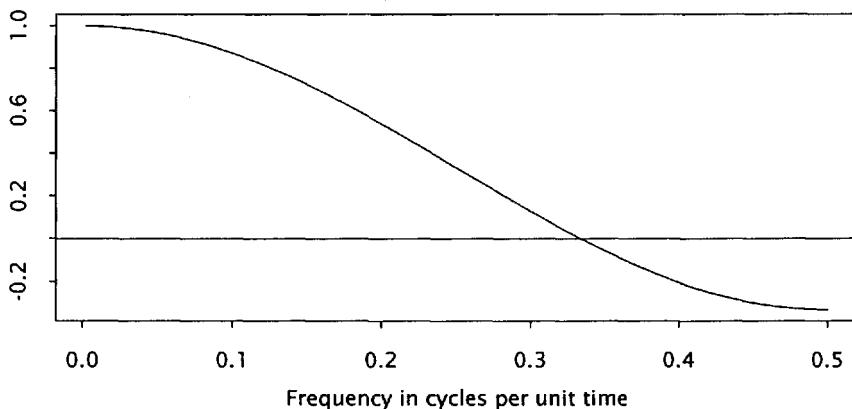


Fig. 7.1 The function $(1 + 2 \cos 2\pi f)/3$.

The real part is thus $R/3 \cos 2\pi(f t + \phi)(1 + 2 \cos 2\pi f)$. Thus the output of this procedure $\{z_t\}$ is obtained from the input $\{y_t\}$ by multiplying by $(1 + 2 \cos 2\pi f)/3$. A graph of this as a function of f is given in Figure 7.1.

Because in this smoothing procedure the output is a linear function of the input, it is also clear what happens when the input is the sum of a number of cosine terms. The output then contains cosine terms with the same frequencies, but with each amplitude changed by the factor $(1 + 2 \cos 2\pi f)/3$, f being the corresponding frequency. Thus terms with frequency near zero pass through almost unchanged, whereas a term with $f = 1/3$ is removed completely.

However, it was shown in Chapter 4 that any series may be written as a sum of cosine terms, and hence the action of this filter on an arbitrary series may be described in these frequency terms. This provides an alternative description of the effect of simple averaging in the original problem, where $\{y_t\}$ is smooth "signal" plus error, or "noise": it was shown in Chapter 4 that for a function to be smooth its transform must be concentrated at low frequencies, whereas at least for *white noise* (see Section 6.8) the magnitude of the transform is relatively constant. Thus the averaging procedure described here passes most of the signal, but cuts down the power of the noise, at least over certain frequency bands.

A general linear filter consists of a set of weights $\{g_r, g_{r+1}, \dots, g_s\}$, such that if the input to the filter is $\{y_t\}$ the output is

$$z_t = \sum_{u=r}^s g_u y_{t-u}. \quad (7.5)$$

The three-term simple moving average above has $r = -1$, $s = 1$, and $g_u = 1/3$, $u = -1, 0, 1$. Note that if the input series is available for $t = 0, 1, \dots, n-1$, the output may be computed only for $t = s, s+1, \dots, n-1+r$. When the input series is the sinusoid $R \cos 2\pi(f t + \phi)$, the output, for t in the latter range, is the real part of

$$\sum_{u=r}^s g_u R e^{2\pi i(f t - f u + \phi)} = R e^{2\pi i(f t + \phi)} \sum_{u=r}^s g_u e^{-2\pi i f u}.$$

The second factor,

$$G(f) = \sum_{u=r}^s g_u e^{-2\pi i f u},$$

is called the *transfer function* of the filter, since it describes the way in which a sinusoid with frequency f is transferred from the input to the output. Its squared magnitude $|G(f)|^2$ is the *power transfer function*.

In the case of a symmetric filter, one for which $r = -s$ and $g_{-u} = g_u$, the transfer function is real, and the output is $R G(f) \cos 2\pi(f t + \phi)$. Notice that in this case

$$\begin{aligned} G(f) &= \sum g_u \cos 2\pi f u \\ &= g_0 + 2 \sum_{u>0} g_u \cos 2\pi f u, \end{aligned}$$

and since $\cos ux$ is a polynomial in $\cos x$, $G(f)$ may be expanded as a polynomial in $\cos 2\pi f$.

More generally, $G(f)$ may be complex, say $G(f) = \Gamma(f) e^{2\pi i \gamma(f)}$, where $\Gamma(f)$ is real and nonnegative and $\gamma(f)$ is real. The output is then the real part of

$$R e^{2\pi i(f t + \phi)} \Gamma(f) e^{2\pi i \gamma(f)},$$

which is

$$R \Gamma(f) \cos 2\pi \{f t + \phi + \gamma(f)\}.$$

In this case, the amplitude is changed by $\Gamma(f)$ as before, but in addition the phase is changed by $\gamma(f)$. It should be noted that when the input series is real, the filter weights are usually also real-valued, and consequently $G(-f) = \overline{G(f)}$. Thus $\Gamma(-f) = \Gamma(f)$ and $\gamma(-f) = -\gamma(f)$. Thus, as might be expected, it is necessary to study the behavior of transfer functions only for positive frequencies.

Relation 7.5 defining the output of a filter in terms of its inputs and a set of filter weights is an example of *convolution*. If the values at the ends of the output are computed as if the input were part of a periodic series with period n , the operation is the *circular convolution* defined in

Exercise 6.4. It follows from the results of that exercise that if f is a Fourier frequency, then

$$\begin{aligned} d_z(f) &= n d_y(f) d_g(f) \\ &= G(f) d_y(f), \end{aligned} \tag{7.6}$$

since the transform of the weights is just $d_g(f) = n^{-1}G(f)$. Typically, some other rule is used to compute the end values of the output series. Two simple rules are to treat unavailable values in the input series as equal to the corresponding end value, and to treat the input series as if it were symmetric about each end value. It may be shown that for any reasonable rule (7.6) is approximately true (see Exercise 7.3).

Since the transfer function is proportional to the transform of the weights, it may be inverted to yield the weights. Thus a general linear filter may be defined in terms of either the weights $\{g_u\}$ or the transfer function $G(f)$. The weights are also known as the *impulse response function*, since if the input series is an *impulse* (i.e., consists of a single nonzero value in a string of zeros), the output consists precisely of the weights (see Exercise 7.6). The filters described above are known as *finite impulse response* (FIR) filters. All filters used below are in this class. The other main class consists of *recursive* filters, in which the output at a given time depends on a finite number of values of the input series and also on values of the output at a finite number of other times (usually in the past). Recursive filters do not have finite impulse response. The transfer function of a filter is also known as its *frequency response function*, since in a dual way it describes the output when the input contains a single frequency. Recursive and nonrecursive filters are discussed in more detail by Hamming (1998).

The simplest filters are those which *shift* the series; that is, if the input is $\{y_t\}$, the output is $\{z_t\}$ where $z_t = y_{t+h}$ for some h . Since $z_0 = y_h$ this may be viewed as a change of time origin to $t = h$. The transfer function is $e^{2\pi i fh}$, whence $\Gamma(f) = 1$ and $y(f) = fh$, a linear function of frequency f . Whenever a transfer function has a linear phase it may be viewed as including a shift, even if $\Gamma(f)$ is not constant.

A convenient way to construct filters is by repeated application; if both are *low-pass filters* (see Section 7.3), this is called *resmoothing* (Tukey, 1977). When filters with transfer functions $G_1(f)$ and $G_2(f)$ are applied in succession, they are equivalent to a composite filter with transfer function $G(f) = G_1(f)G_2(f)$: if the input to the first is $e^{2\pi i ft}$ then the output, which is the input to the second, is $G_1(f)e^{2\pi i ft}$, whence the final output is $G_1(f)G_2(f)e^{2\pi i ft}$. The order in which they are applied does not matter except for possible effects at the ends of the series: the transfer functions

$G_1(f)G_2(f)$ and $G_2(f)G_1(f)$ are identical and by the above argument the transfer function determines the filter.

Clearly, the result of resmoothing is an output series that is more smooth than that of either filter applied on its own. Sometimes it is desired to combine filters so that the output is *less* smooth, for instance, if it is judged that an otherwise satisfactory filter has eliminated part of the signal. One way to achieve this is by *reroughing* (Tukey, 1977). The second filter is applied not to the *output* of the first but to the corresponding *rough*:

$$\text{rough} = \text{input} - \text{output}.$$

The output of the second filter is then *added to* the output of the first, to restore the eliminated part of the signal. The transfer function is easily seen to be $G(f) = G_1(f) + G_2(f) - G_1(f)G_2(f)$, whence $1 - G(f) = \{1 - G_1(f)\}\{1 - G_2(f)\}$.

Exercise 7.2 Circular Filters

Suppose that a linear filter with weights $\{g_u : r \leq u \leq s\}$ is defined as a circular convolution. Show that (7.6) holds approximately when f is not a Fourier frequency. (*Hint:* Write out (7.6) explicitly for the end terms and note that only finitely many terms are affected.)

Exercise 7.3 (Continuation) Noncircular Filters

A possible definition of a *reasonable* rule for extending an input series in order to compute the output of a noncircular filter would be that any substitute value falls in the *range* of the data. In other words, if \hat{y} is any value substituted for y_t , $t < 0$ or $t \geq n$, then

$$\min\{y_t : 0 \leq t < n\} \leq \hat{y} \leq \max\{y_t : 0 \leq t < n\}.$$

With this definition, show that (7.6) holds approximately for any f .

Exercise 7.4 Impulse Response Function

Suppose that a linear filter has weights $\{g_u : r \leq u \leq s\}$, where $r \leq 0 \leq s$, and that its input is

$$y_t = \begin{cases} 1 & \text{if } t = v \\ 0 & \text{otherwise} \end{cases}$$

for some v satisfying $s \leq v < n + r$. Show that the output is

$$z_t = \begin{cases} g_{t-v} & \text{if } v + r \leq t \leq v + s \\ 0 & \text{otherwise.} \end{cases}$$

7.3 DESIGNING A FILTER

Two essentially distinct approaches may be taken to designing a filter for any given application. The first is to take a convenient set of primitive filters, such as simple moving averages of various lengths, and use them as building blocks in assembling a filter with the desired characteristics. The second is to specify the requirements precisely, and then to construct a filter directly to satisfy them. Examples of both approaches are given in this section and the next, and applied to complex demodulation later in the chapter and to other filtering problems in other chapters. Hamming (1998) discusses filter design at length.

The requirements for a filter are usually specified in terms of both the impulse response function and the frequency response function.

- Other things being equal, it is usually desirable for the filter to have a short *span*, where the span is

$$\max\{u : g_u \neq 0\} - \min\{u : g_u \neq 0\}.$$

The span is one more than the total number of points that are lost (or have to be computed specially) at the ends of the data, and the computation⁻¹ effort required to apply the filter is usually roughly proportional to the span.

- It is usually undesirable for the impulse response function to show eccentric behavior such as large ripples.

The desired frequency response function, or transfer function, may usually be specified more precisely. The problem that arose in Section 7.1, that of separating a low-frequency component from other terms, is typical. A filter that does this is called a *low-pass filter*. Its transfer function should be close to 1 for frequencies in the band associated with the low-frequency signal (the *pass band*, extending from $f = 0$ to the *pass frequency*) and close to 0 elsewhere (the *stop band*, extending from the *stop frequency* to $f = 1/2$ cycles per unit time). Since transfer functions are continuous, there is necessarily an intermediate band (the *transition band*) where the transfer function lies between 0 and 1 in magnitude.

The transfer function of a symmetric filter (only symmetric filters will be used in the present application) is a polynomial in $\cos 2\pi f$, and hence these requirements cannot be met exactly for all frequencies. For complex demodulation, additional information shows which frequencies are most important. Equation (7.3) showed that one unwanted component contains a sinusoid with frequency² around $2f_0$. Thus the transfer function should be 0 at $2f_0$, and small in a band surrounding $2f_0$. The other unwanted frequencies are those in the final term. Since in the present case $\{z_t\}$ is known to contain a low-frequency term and the second harmonic of f_0 , this term contains power at around f_0 and $3f_0$. Thus the transfer function should similarly vanish at f_0 and $3f_0$.

The simplest way to meet these minimal requirements is to find a polynomial $P(x)$ such that

$$P(1) = 1, \quad P(\cos 2\pi f_0) = P(\cos 4\pi f_0) = P(\cos 6\pi f_0) = 0. \quad (7.7)$$

The transfer function $P(\cos 2\pi f)$ is then necessarily close to 1 for low frequencies and close to 0 around f_0 , $2f_0$, and $3f_0$. The lowest degree polynomial satisfying (7.7) is

$$P(x) = \frac{(x - \cos 2\pi f_0)(x - \cos 4\pi f_0)(x - \cos 6\pi f_0)}{(1 - \cos 2\pi f_0)(1 - \cos 4\pi f_0)(1 - \cos 6\pi f_0)}.$$

The graph of $P(\cos 2\pi f)$ is shown in Figure 7.2 for $f_0 = 0.08936$ cycles per year, the approximate frequency of the sunspot cycle. The graphed behavior is unacceptable because

- the values between the zeros are too large, and
- the function becomes large in magnitude as f approaches 1/2 cycles per year.

A sinusoid with frequency between 1/4 and 1/2 cycles per year would be strongly amplified (and have its phase changed by one half of a cycle). If further zeros are added at $4f_0$ and $5f_0$, the transfer function becomes that shown in Figure 7.3, which is more acceptable.

The filter corresponding to any transfer function may be found in several ways.

- Evaluate the transfer function on a sufficiently fine grid of frequencies, and apply the inverse discrete Fourier transform.

²Strictly the frequency is $-2f_0$, but since the transfer function is symmetric, the sign may be ignored.

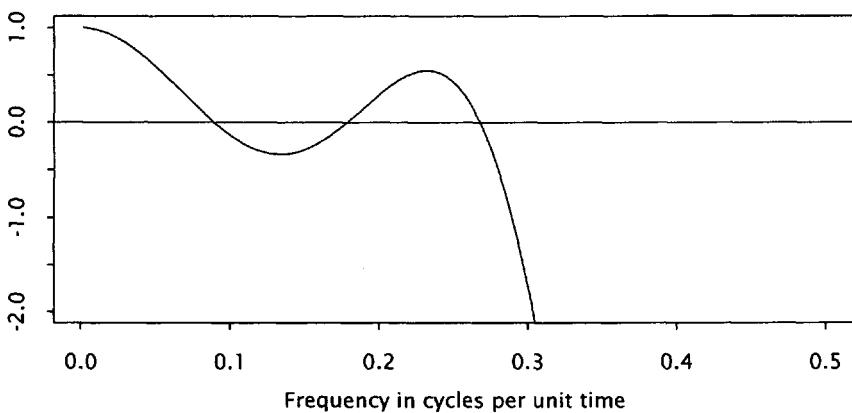


Fig. 7.2 A transfer function with three zeros.

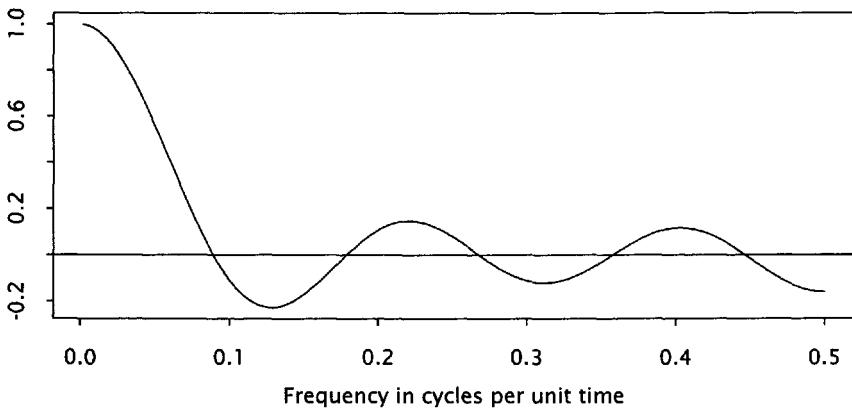


Fig. 7.3 A transfer function with five zeros.

- Replace $\cos 2\pi f$ by $(e^{2\pi i f} + e^{-2\pi i f})/2$ and expand in powers of $e^{2\pi i f}$.
- If the transfer function is specified in a factorized form

$$P(\cos 2\pi f) = \prod_j (a_j + 2b_j \cos 2\pi f)$$

it may be applied by successive application of three-term filters with

weights b_j , a_j , b_j , since the transfer function of the j th of these is exactly the j th factor.

The last is an example of building a filter using three-term filters as primitives.

The simplest way to fill in the two end values when using a three-term filter is to treat y_{-1} and y_n as if they vanish. One modification is to scale the resulting values by $(a_j + 2b_j)/(a_j + b_j)$, thus making the sum of the weights for the end values the same as the sum of the basic weights. Both these methods destroy the symmetry of the filter, however, and it may be preferable to use y_0 or y_{n-1} multiplied by $a_j + 2b_j$.

When the frequency f_0 corresponds to an integer period p , the p -term simple moving average provides a useful filter. If p is odd, say $p = 2q+1$, this is

$$z_t = \frac{1}{p} \sum_{u=-q}^q y_{t-u},$$

with transfer function $D_p(f)$, the Dirichlet kernel (see Section 2.2). Since $D_p(0) = 1$ and $D_p(kf_0) = 0$, $k = 1, 2, \dots, q$, this transfer function has zeros at the required frequencies, and its oscillations become smaller as f increases. The period of the sunspot series is close to 11 years, and, in fact, $D_{11}(f)$ is very similar to the transfer function shown in Figure 7.3 (see also Exercise 7.6).

When p is even, say $p = 2q$, there is no symmetric simple moving average of length p . The modified average

$$z_t = \frac{1}{p} \left(\frac{1}{2} y_{t-q} + \sum_{u=-q+1}^{q-1} y_{t-u} + \frac{1}{2} y_{t+q} \right) \quad (7.8)$$

is typically used. It is the combination of an asymmetric simple moving average of length p and an asymmetric simple moving average of length 2, whence its transfer function is $D_p(f)D_2(f) = D_p(f) \cos \pi f$ (see Exercise 7.7). The main effect of the factor $\cos \pi f$ is to add a second zero at $f = 1/2$ cycles per unit time.

For simple moving averages a natural rule for filling in end values is to take the average of the available data. For example, if $p = 2q+1$ the values of z_0, z_1, \dots would be

$$\frac{1}{q+1} \sum_{u=-q}^0 y_{-u}, \quad \frac{1}{q+2} \sum_{u=-q}^1 y_{1-u}, \quad \dots$$

However, these averages are asymmetric, and an alternative is to use centered averages of the available data, $y_0, (y_0 + y_1 + y_3)/3, \dots$

Since simple moving averages are easy to compute, they could be added to the set of primitive filters. A filter could then be built by starting with a suitable simple moving average (or combination of simple moving averages), and using three-term filters to incorporate any further zeros that might be needed in the transfer function. End values would be filled in by an appropriate rule at each step.

The transfer function may be made arbitrarily small in the *stop* band by these resmoothing techniques. Reroughing may be used to improve its behavior in the *pass* band. If, for instance, the transfer function $G_1(f)$ decreases too rapidly as f increases from 0, the following simple modification provides an improvement. Suppose that $G_2(f)$ is the transfer function of a second filter with a similar pass band. When these are combined by reroughing, the transfer function $G(f)$ of the combination satisfies $1 - G(f) = \{1 - G_1(f)\}\{1 - G_2(f)\}$. Since both $G_1(f)$ and $G_2(f)$ are close to 1 in the pass band, $G(f)$ is closer than either. Thus the pass band behavior has been improved without seriously compromising³ the stop band. One cost is that the span of the filter is increased (see Exercise 7.5).

Exercise 7.5 Combination of Filters

Suppose that a filter is obtained by using a filter with weights $\{g_u : r \leq u \leq s\}$ on the output of another with weights $\{g'_u : r' \leq u \leq s'\}$. Show that the span of the combined filter is the sum of the individual spans less 1, and find the impulse response function of the combined filter.

Exercise 7.6 Impulse Response Function

Find the impulse response function of the 11-term filter whose transfer function has zeros at $jf_0, j = 1, 2, \dots, 5$ for $f_0 = 0.08936$ and takes the value 1 at $f = 0$. Verify that each weight is approximately 1/11.

Exercise 7.7 Simple Moving Averages of Even Length

Verify that the modified simple moving average of length $p = 2q$ defined

³In the stop band $|G_1(f)| \ll 1$ and $|G_2(f)| \ll 1$, whence $G(f) \approx G_1(f) + G_2(f)$.

in (7.8) gives the same output as the combined asymmetric filters

$$\begin{aligned} z'_t &= \frac{1}{p} \sum_{u=-q+1}^q y_{t-u}, \\ z_t &= \frac{1}{2}(z'_t + z'_{t+1}). \end{aligned}$$

7.4 LEAST SQUARES FILTER DESIGN

A more systematic way of designing a filter is as follows. Suppose that the context suggests an ideal transfer function $H(f)$ and that we decide to approximate it by a finite impulse response filter. For instance, to design a low-pass filter we might decide on a *pass frequency* f_p , a *stop frequency* f_s , and *tolerances* δ_p and δ_s , and then look for the filter with the shortest span whose transfer function $G(f)$ satisfies

$$\begin{aligned} |1 - G(f)| &\leq \delta_p, \quad 0 \leq |f| \leq f_p, \\ |G(f)| &\leq \delta_s, \quad f_s \leq |f| \leq \frac{1}{2}. \end{aligned}$$

Such problems are usually complex, and can be solved only by numerical optimization.

An easier approach is to fix the span of the impulse response function and then find the filter whose transfer function best approximates $H(f)$ in the least squares sense. In other words, for given r and s we find $\{g_u : r \leq u \leq s\}$ to minimize

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \left| H(f) - \sum_{u=r}^s g_u e^{-2\pi i f u} \right|^2 df. \quad (7.9)$$

The optimal value of g_u may be verified to be

$$h_u = \int_{-\frac{1}{2}}^{\frac{1}{2}} H(f) e^{2\pi i f u} df, \quad r \leq u \leq s, \quad (7.10)$$

the u th *Fourier coefficient* of $H(f)$ (see Exercise 7.8). It is interesting and convenient that the optimal value of g_u does not depend on the values of r and s , a consequence of the orthogonality of these complex sinusoids.

The transfer function of the approximating filter is

$$\begin{aligned} H_{r,s}(f) &= \sum_{u=r}^s h_u e^{-2\pi i f u} \\ &= (s - r + 1) \int_{-\frac{1}{2}}^{\frac{1}{2}} H(f - f') D_{s-r+1}(f') e^{-\pi i (r+s)f'} df' \quad (7.11) \end{aligned}$$

where $D_{s-r+1}(f)$ is the Dirichlet kernel (see Exercise 7.9). The sum is just the Fourier series of $H(f)$ truncated at r and s . If $H(f)$ is real and symmetric and $r = -s$, then

$$\begin{aligned} h_u &= \int_{-\frac{1}{2}}^{\frac{1}{2}} H(f) \cos 2\pi f u \, df \\ &= 2 \int_0^{\frac{1}{2}} H(f) \cos 2\pi f u \, df \\ &= h_{-u} \end{aligned}$$

and

$$\begin{aligned} H_{r,s}(f) &= \sum_{u=-s}^s h_u e^{-2\pi i f u} \\ &= h_0 + 2 \sum_{u=1}^s h_u \cos 2\pi f u \\ &= (2s+1) \int_{-\frac{1}{2}}^{\frac{1}{2}} H(f-f') D_{2s+1}(f') \, df' \\ &= H_s(f), \end{aligned} \tag{7.12}$$

say.

A Low-pass filter

Suppose that this method is used to approximate the ideal low-pass transfer function

$$H(f) = \begin{cases} 1 & 0 \leq |f| \leq f_c, \\ 0 & f_c < |f| \leq \frac{1}{2}, \end{cases} \tag{7.13}$$

where f_c is the *cutoff frequency*. The Fourier coefficients are

$$h_u = 2 \int_0^{f_c} \cos 2\pi f u \, df = \begin{cases} 2f_c & u = 0, \\ \frac{\sin 2\pi f_c u}{\pi u} & u \neq 0. \end{cases}$$

For example, in the sunspot data a reasonable cutoff frequency would be $f_c = f_0/2 = 0.04968$ cycles per year. Figure 7.4 shows the approximating transfer functions for $s = 5$ (an 11-term filter) and $s = 20$ (a 41-term filter), together with the ideal transfer function. The curve for $s = 20$ shows pronounced *overshoot* on either side of the cutoff frequency. This is known as *Gibbs's phenomenon* (Lanczos, 1961, p. 225) and is characteristic of the truncated Fourier series of a discontinuous function.

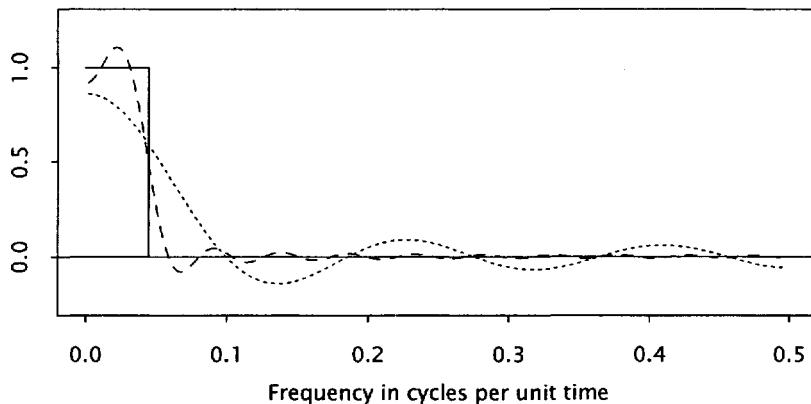


Fig. 7.4 Transfer functions of least squares low-pass filters, $s = 5$ (short dashed line) and $s = 20$ (long dashed line).

Convergence Factors

Overshoot and the accompanying sidelobes may be greatly reduced as follows: The wavelength of the sidelobes may be shown to be approximately $\delta = 2/(2s + 1)$ (see Exercise 7.10). Thus, the smoothed function

$$\tilde{H}_s(f) = \frac{1}{\delta} \int_{-\frac{\delta}{2}}^{\frac{\delta}{2}} H_s(f - f') df'$$

has smaller sidelobes, since the integration is over one complete cycle. But

$$\tilde{H}_s(f) = h_0 + 2 \sum_{u=1}^s h_u \frac{\sin \pi \delta u}{\pi \delta u} \cos 2\pi f u \quad (7.14)$$

and this corresponds to replacing the Fourier coefficient h_u by

$$h_u \frac{\sin \pi \delta u}{\pi \delta u} = h_u \frac{\sin 2\pi u / (2s + 1)}{2\pi u / (2s + 1)}. \quad (7.15)$$

The multipliers

$$\sigma_{s,u} = \frac{\sin 2\pi u / (2s + 1)}{2\pi u / (2s + 1)}.$$

are an example of *convergence factors* and are essentially the same as the σ -factors introduced by Lanczos (1961, pp. 225-229) to accelerate

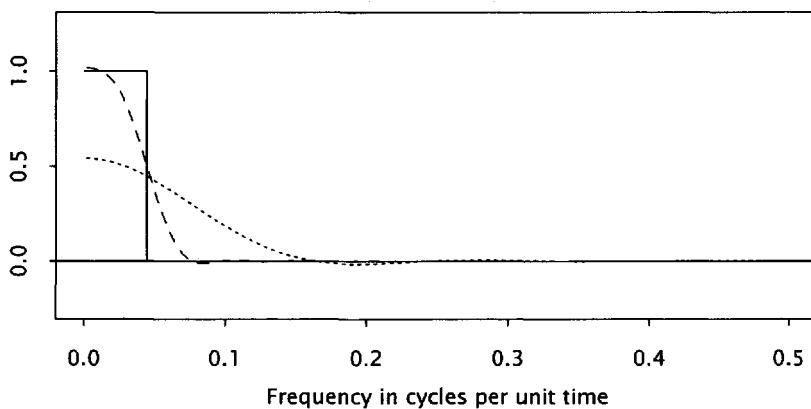


Fig. 7.5 Transfer functions of least squares low-pass filters with convergence factors applied, $s = 5$ (short dashed line) and $s = 20$ (long dashed line).

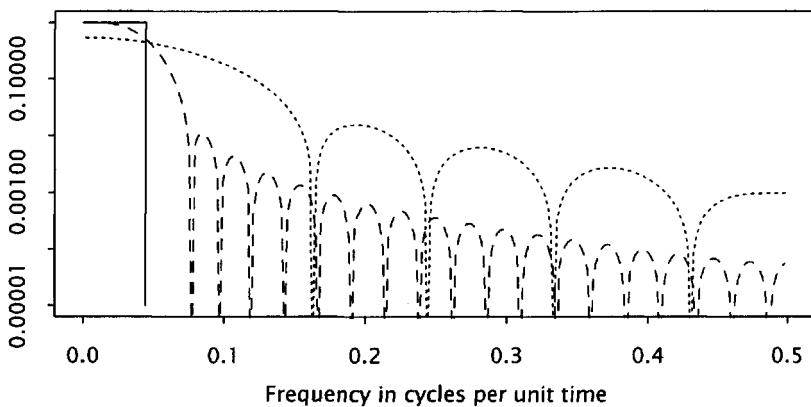


Fig. 7.6 Absolute values of transfer functions in Figure 7.5, logarithmic vertical scale.

the convergence of Fourier series. Figure 7.5 shows the functions $\tilde{H}_5(f)$ and $\tilde{H}_{20}(f)$ corresponding to the functions in Figure 7.4. The sidelobes in each have been substantially reduced, as has the overshoot in $H_{20}(f)$. The absolute values of these functions are plotted on a logarithmic scale in Figure 7.6. The plot shows that the sidelobes, especially those of $\tilde{H}_{20}(f)$, are uniformly small.

The use of convergence factors is analogous to the use of a data window (see Section 6.2). A truncated Fourier series may be regarded as the infinite series with the coefficients multiplied by a boxcar function. This multiplication is equivalent to convolving the original function with the transform of the boxcar, the Dirichlet kernel. Such equivalence has, in fact, been shown in (7.11). The convergence factors, which are initially 1 and decay smoothly to 0, are a smooth approximation to the boxcar. The modified truncated sum can be represented as the convolution of the original function with the transform of the convergence factors (see Exercise 7.11), which has smaller sidelobes than the Dirichlet kernel. A split cosine bell taper (see Section 6.2) would have achieved much the same effect. Note that the cosine bell, in fact, approaches 0 more smoothly than the convergence factors, which behave like $(\sin x)/x$ (see Figure 4.1 on p. 47). It is, therefore, likely that cosine bell tapering applied only to the higher coefficients would be roughly equivalent to the use of the factors derived in this section.

Transition Band

Another interpretation of the application of convergence factors may be derived as follows. The convergence factors are the Fourier coefficients of the boxcar function

$$C(f) = \begin{cases} \frac{1}{\delta} & |f| \leq \frac{\delta}{2} \\ 0 & \text{otherwise,} \end{cases}$$

and therefore the products (7.15) are the Fourier coefficients of the convolution of the ideal transfer function $H(f)$ with this boxcar (see Exercise 7.12). Thus the modified partial sum (7.14) may also be regarded as the least squares approximation to this convolution, which is a smoothed version of the ideal transfer function.

The modified transfer function thus has two interpretations, as:

- a smoothed version of the least squares approximation to the ideal transfer function (the original derivation), or
- the least squares approximation to a smoothed version of the ideal transfer function.

For the ideal low-pass filter (7.13), when $f_c \geq \delta/2$ the effect of this smoothing is to replace the ideal low-pass transfer function, with its sharp cutoff at f_c cycles per unit time, by a modified function that decays linearly from the value 1 at $f_c - \delta/2$ to 0 at $f_c + \delta/2$ cycles per unit time.

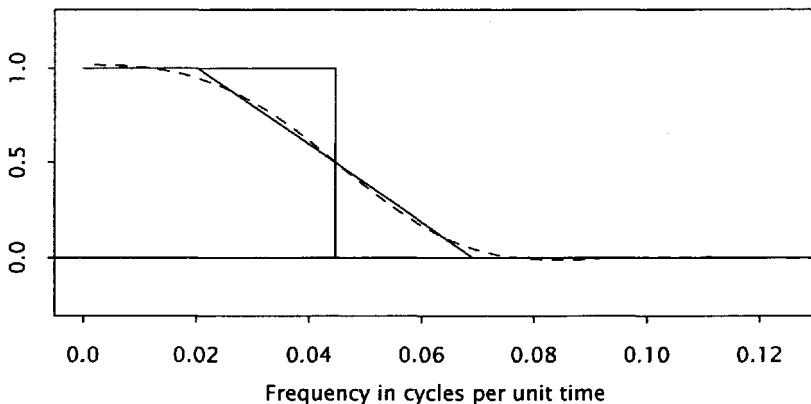


Fig. 7.7 Ideal low-pass transfer functions with and without transition bands, and transfer function of least squares low-pass filter with convergence factors applied, $s = 20$, graphed for $0 \leq f \leq 0.125$.

This creates a *transition band* of width $\delta = 2/(2s + 1)$. The transition band is defined by its lower limit, the *pass frequency* $f_c - \delta/2$, and its upper limit, the *stop frequency* $f_c + \delta/2$.

Convergence factors have a less desirable effect when the condition $f_c \geq \delta/2$ is *not* met. In the case of $\tilde{H}_5(f)$, the half-width is $\delta/2 = 1/11 = 0.0909$, while $f_c = 0.04968$ cycles per year. When $f_c < \delta/2$, the transition band is centered at $\delta/2$ instead of at f_c , and has width $2f_c$ instead of δ . Thus convergence factors have the intended effect only if $2s+1 \geq 1/f_c$. This explains the poor approximation given by $\tilde{H}_5(f)$, since in the present case $1/f_c \approx 22$.

The ideal low-pass transfer function with its sharp cutoff, the modification with a transition band of width $\delta = 2/41$, and $\tilde{H}_{20}(f)$ are shown in greater detail in Figure 7.7 for $0 \leq f \leq 0.125$ cycles per year.

Passing a Constant Term

A filter is sometimes required to pass a zero-frequency component (that is, a constant term) without change. This requires that the transfer function have the value 1 at zero frequency. For the filters constructed in Section 7.3 this requirement is easily imposed, but Figures 7.4 to 7.7 show that least squares approximations to an ideal filter with this property do not, in general, share it, whether or not convergence factors are

used. The least squares argument is easily modified to include this constraint, the only effect on the solution being that h_u is replaced by $h_u + (1 - \sum_{u'=r}^s h_{u'}) / (2s + 1)$. When convergence factors $\{\sigma_{s,u}, -s \leq u \leq s\}$ are used, an appropriate modification is to replace h_u by

$$h_u + \frac{1 - \sum_{u'=-s}^s h_{u'} \sigma_{s,u'}}{\sum_{u'=-s}^s \sigma_{s,u'}}.$$

In either case, a simple alternative is just to rescale the coefficients so that they sum to 1.

End Values

Filling in end values is not as easy with the filters described in this section as with the three-term and simple moving average filters of the preceding section. The least squares approach suggests the constraint that the weights attached to unavailable data to be 0, with the remaining coefficients chosen optimally. However, these optimal values are not affected by the constraints, and thus the overall effect is the same as replacing unavailable values by 0. The resulting filter becomes quite asymmetric at the extreme ends of the series and hence may be unacceptable in some cases.

If convergence factors are used, they should be computed separately for the two sides of the filter. Thus when $0 \leq t < s$, for instance, the output calculation, normally

$$\begin{aligned} z_t &= \sum_{u=-s}^s h_u \sigma_{s,u} y_{t-u} \\ &= \sum_{u=-s}^0 h_u \sigma_{s,u} y_{t-u} + \sum_{u=1}^s h_u \sigma_{s,u} y_{t-u} \end{aligned}$$

would be computed instead as

$$z_t = \sum_{u=-s}^0 h_u \sigma_{s,u} y_{t-u} + \sum_{u=1}^t h_u \sigma_{t,u} y_{t-u},$$

with a similar modification at the other end of the series.

Exercise 7.8 The Fourier Coefficients

Verify that the values of g_u , $u = r, r+1, \dots, s$, that minimize (7.9) are the Fourier coefficients (7.10). (*Hint:* Write $H(f)$ as $X(f) + iY(f)$ and g_u as

$x_u + iy_u$, expand (7.9) in terms of these real quantities, and differentiate with respect to x_u and y_u .)

Exercise 7.9 The Truncated Fourier Series

Verify (7.11). (*Hint:* Substitute the integral formula (7.10) for h_u , and sum.)

Exercise 7.10 Gibbs's Phenomenon

Specialize (7.12) to the case of the ideal low-pass filter with transfer function (7.13). Show that, if s is large and f_c is small, the oscillations in $H_s(f)$ in any small interval (f_1, f_2) not containing 0 are approximately sinusoidal, with period $2/(2s + 1)$. (*Hint:* The denominator of $D_n(f)$ is approximately constant in any such interval.)

Exercise 7.11 General Convergence Factors

Suppose that $\{h_u\}$ are the Fourier coefficients of $H(f)$, defined as in (7.10), and that $\{c_u, r \leq u \leq s\}$, are a set of numbers which, in the present context, are interpreted as convergence factors. Let

$$H_c(f) = \sum_{u=r}^s c_u h_u e^{-2\pi i f u}$$

be the corresponding modified partial sum of the Fourier series for $H(f)$. Show that

$$H_c(f) = \int_{-\frac{1}{2}}^{\frac{1}{2}} H(f - f') C(f') df',$$

where

$$C(f) = \sum_{u=r}^s c_u e^{-2\pi i f u}.$$

(*Hint:* Substitute the integral formula for h_u .)

Notes:

- No assumption about the convergence of the infinite Fourier series of $H(f)$ is needed.
- This identity is analogous to the discrete result derived in Exercise 6.4. In words, the Fourier coefficients of the convolution of two functions are the products of the respective Fourier coefficients,

provided that one of the functions has a finite Fourier series [in this instance, $C(f)$].

Exercise 7.12 (Continuation) A Generalization

Suppose that $H_c(f)$ is the convolution of $H(f)$ and $C(f)$, as in Exercise 7.11, but make no assumption about the finiteness or convergence of the Fourier series of either $H(f)$ or $C(f)$. Show that the Fourier coefficients of $H_c(f)$ are the products of those of $H(f)$ and $C(f)$.

7.5 DEMODULATING THE SUNSPOT SERIES

The first application of complex demodulation is to the sunspot series analyzed in Section 6.4. It was verified in that section that the oscillations in this series have a period of around 11 years, but that the amplitude and phase of the oscillations vary. This section shows how complex demodulation can be used to describe those variations, first using moving average filters and then using least squares filters.

Moving Average Filters

Figure 7.8 shows the *instantaneous amplitude* and *instantaneous phase* of the oscillations in the sunspot series as functions of time. They were calculated by forming the demodulated series

$$y_t = x_t e^{-2\pi i f_0 t}, \quad t = 0, 1, \dots, n - 1,$$

where x_t is the annual sunspot number for year $1700 + t$ and $f_0 = 1/11$ cycles per year is the base frequency. The demodulated series was then smoothed by taking a simple moving average of 11-year blocks. As explained in Section 7.3, this removes most of the unwanted components of $\{y_t\}$ (see equation 7.3). The integer period of 11 years was used in preference to the period of 11.2 years, which corresponds to the largest peak in the periodogram, because the integer value allows the use of a simple moving average filter. With this combination of frequency and filter, the interpretation of complex demodulation as a local version of harmonic analysis is, in fact, exact, since the t th term in the smoothed demodulated series is just the discrete Fourier transform of $x_{t-5}, x_{t-4}, \dots, x_{t+5}$ evaluated at the frequency $1/11$ cycles per year, the fundamental frequency of a series (or subseries) of length 11. In the same way, if a more

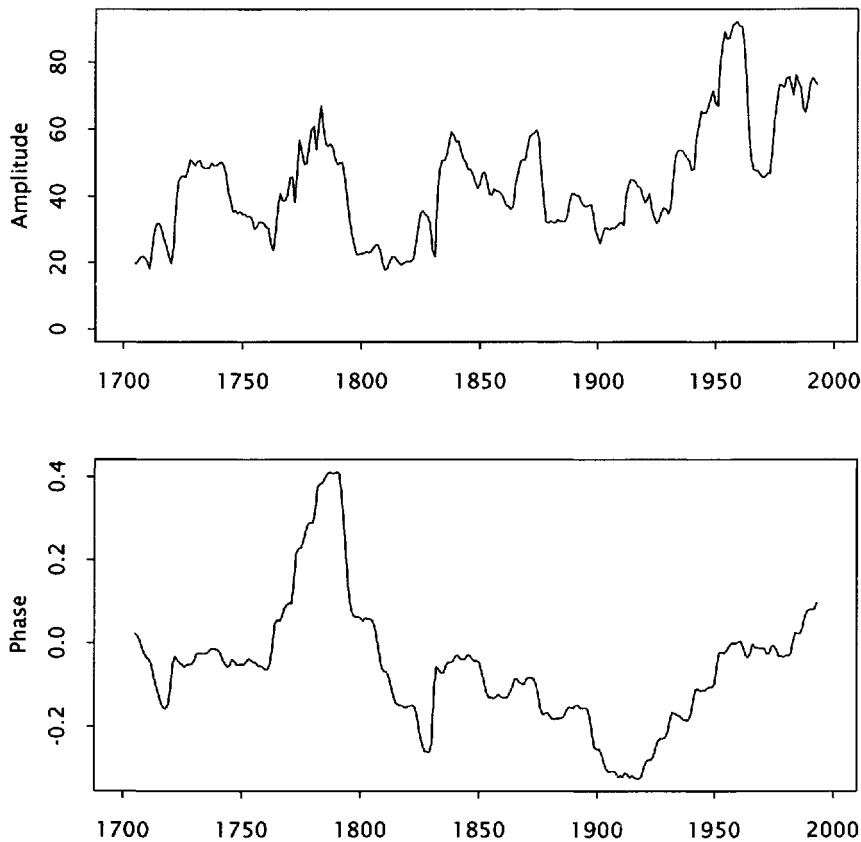


Fig. 7.8 Instantaneous amplitude and phase of sunspot series at $f_0 = 1/11$ cycles per year (11-term simple moving average filter).

sophisticated filter is used, the t th term in the smoothed demodulated series is the transform of a *tapered* stretch of data centered at time t , the data window weights being (proportional to) the weights of the filter.

The smoothed series is

$$z_t = u_t + i v_t \approx \frac{1}{2} R_t e^{2\pi i \phi_t},$$

and R_t and ϕ_t are approximated by solving the equations $R \cos 2\pi \phi = 2u_t$, $R \sin 2\pi \phi = 2v_t$. The solution for R is $2\sqrt{u_t^2 + v_t^2}$, while for ϕ it is the modification of $\arctan v_t/u_t$ described in Section 2.2.

The amplitude plot (Figure 7.8, upper panel) shows that there are indeed substantial variations in the amplitude of the oscillations, with a range of about 3 : 1. The raggedness of this graph shows that the 11-year moving average did not remove all of the unwanted components (or perhaps allows a better identification of what is wanted and what is not). The short-term fluctuations in the graph are difficult to interpret as fluctuations in the amplitude of a sinusoid with an 11-year period.

The phase plot also displays substantial variations. Note that in any time interval in which the phase is a linear function $\phi_t \approx a + bt$ the oscillations are approximated by

$$R_t \cos 2\pi(f_0 t + a + bt) = R_t \cos 2\pi\{(f_0 + b)t + a\},$$

a modulated sinusoid with frequency $f_0 + b$. That is, linear behavior of the phase indicates a shift in frequency. Overall the phase graph shows a slight downward slope, indicating that the basic frequency is slightly below 1/11 cycles per year, or, in other words, that the basic period is slightly greater than 11 years. However, estimating the slope from different stretches of data would give very different answers, meaning that the period is not well determined by these data.

Figure 7.9 shows the instantaneous amplitude and phase calculated using two consecutive 11-term simple moving averages. The transfer function of this filter is the square of the Dirichlet kernel and, therefore, has much smaller sidelobes. Thus more of the unwanted components are removed, and the two graphs are smoother, especially the amplitude plot. The broad features of these graphs are very similar to those of Figure 7.8, suggesting that the extra smoothness has not been gained at the cost of loss of accuracy.

The end values were not filled in for either of the filters used to obtain Figures 7.8 and 7.9. The possibilities described in Section 7.3 are either

- asymmetric, thus introducing phase shifts into the smoothed values,
or
- based on only a few data points, leaving them insufficiently smoothed.

Both of these characteristics are undesirable, and unless the end values are important it may be preferable simply to omit them. In the case of the 11-term simple moving average this amounts to losing 5 points at each end of the series, which often will not be critical. With the second pass of this filter an additional 5 points are lost at each end, which may be more important, especially if the amplitude and phase of future cycles must be forecast.

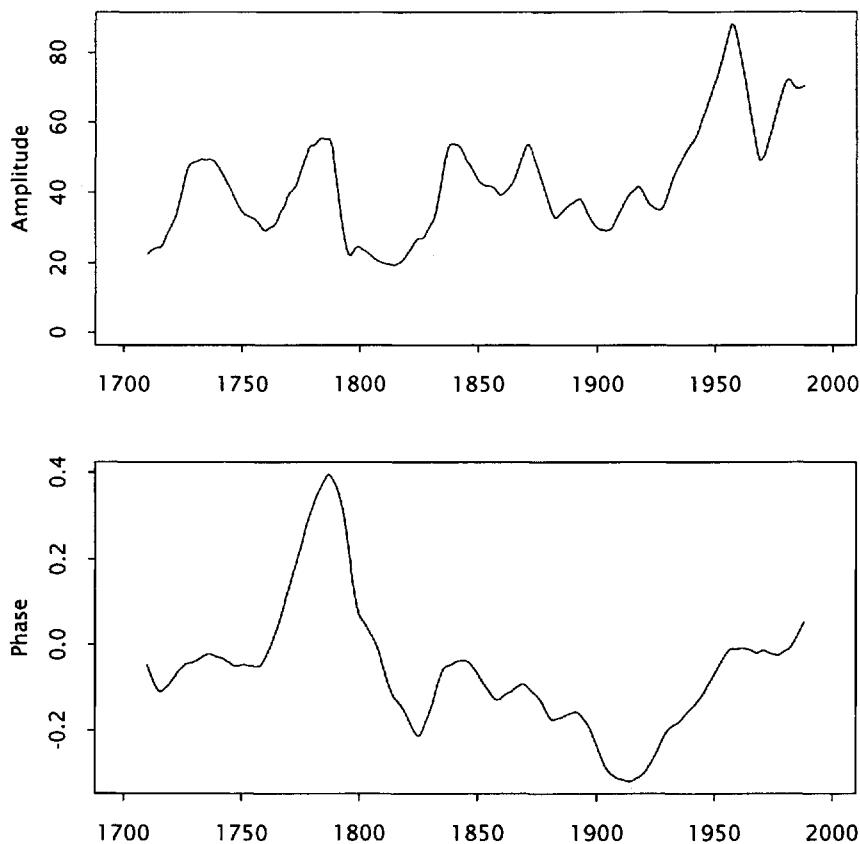


Fig. 7.9 Instantaneous amplitude and phase of sunspot series at $f_0 = 1/11$ cycles per year (two passes of 11-term simple moving average filter).

Figures 7.8 and 7.9 give information about variations in the fundamental frequency of the sunspot oscillations. The analysis of Section 6.5 suggests that the *instantaneous relative phase* of the second harmonic might be useful in exploring the nonsinusoidal nature of the oscillations. This is the instantaneous analog of the relative phase discussed in Section 6.5 and is constructed by demodulating at frequency $2f_0$ in addition to f_0 . Figure 7.10 shows the results for the two filters used previously. In each panel the quantity graphed is $\phi_t(2f_0) - 2\phi_t(f_0)$, where $\phi_t(f_0)$ and $\phi_t(2f_0)$ denote instantaneous phase obtained from demodulation at frequencies f_0 and $2f_0$ cycles per year, respectively. Where the principal

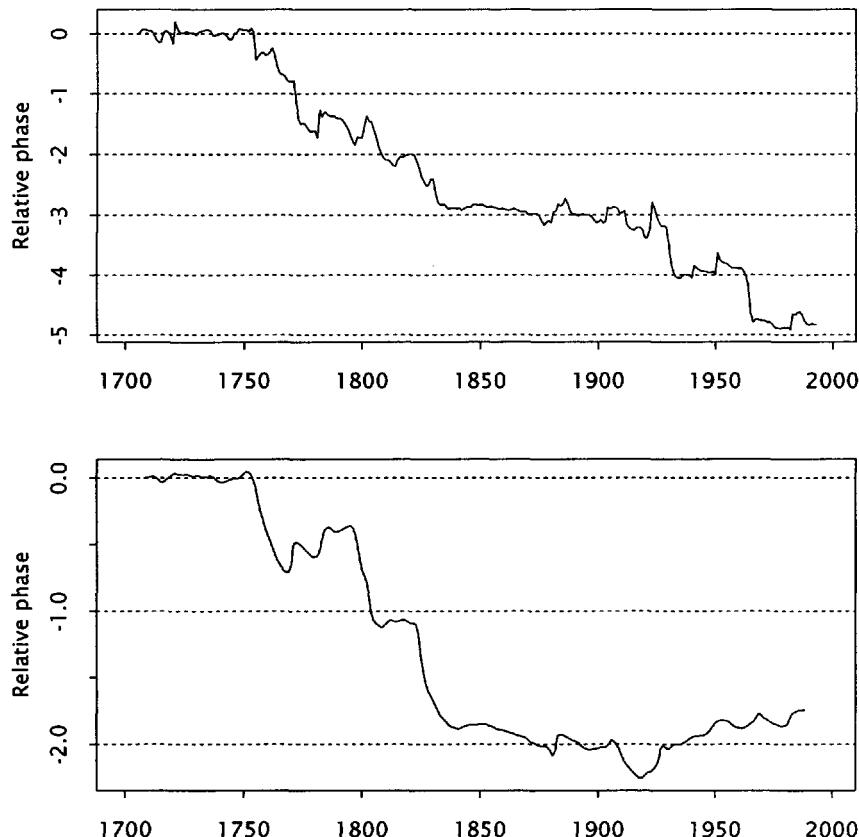


Fig. 7.10 Instantaneous relative phase of second harmonic of the annual sunspot numbers, at $f_0 = 1/11$ cycles per year; upper panel based on 11-year simple moving average, lower panel based on two passes of 11-year simple moving average.

value of phase makes a jump between $1/2$ and $-1/2$ cycles, the graph switches to the alias that keeps the curve continuous. Both curves tend to stay in the principal range 0 to $1/2$ or one of its aliases, as is to be expected from the nature of the oscillations (see Section 6.5). Transitions from one alias of this range to another occur where the amplitude of the second harmonic is weak, where its phase is not well determined.

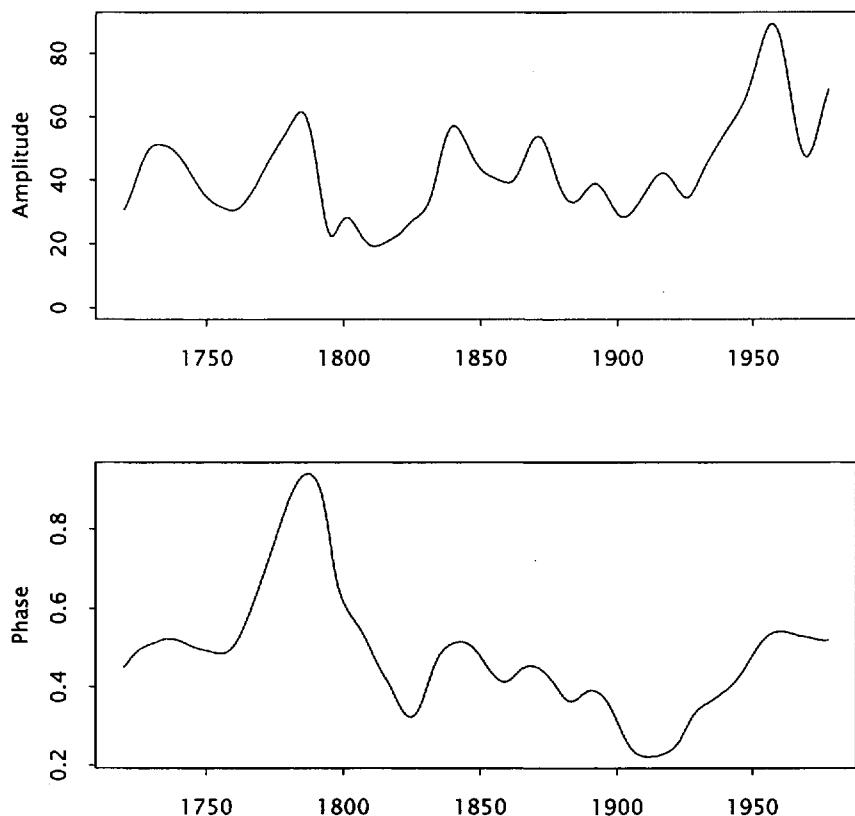


Fig. 7.11 Instantaneous amplitude and phase of sunspot series at $f_0 = 1/11$ cycles per year (41-term least squares filter with cutoff frequency $f_c = 0.04968$ and convergence factors).

Least Squares Filters

Figure 7.11 shows the instantaneous amplitude and phase that result from using a least squares filter with convergence factors, as described in Section 7.4. For comparison with the earlier figures, the demodulation frequency was $f_0 = 1/11$ cycles per year. The filter has 41 terms ($s = 20$), giving a transition band of width $\delta = 2/(2s + 1) = 2/41$ cycles per year. The transition band extends from the *pass frequency* of 0.02529 cycles per year to the *stop frequency* of 0.07407 cycles per year. The transfer

function of this filter is shown in Figures 7.5 to 7.7. The instantaneous amplitude and phase obtained in this way are both very similar to those resulting from two passes of an 11-term simple moving average, shown in Figure 7.9.

7.6 COMPLEX TIME SERIES

The essential step in complex demodulation is the separation of the two parts of the perturbed sinusoid (7.1):

$$\begin{aligned}x_t &= R_t \cos 2\pi(f_0 t + \phi_t) \\&= \frac{1}{2} R_t \left\{ e^{2\pi i(f_0 t + \phi_t)} + e^{-2\pi i(f_0 t + \phi_t)} \right\}.\end{aligned}$$

The discrete Fourier transform (Section 4.2) provides a way to achieve this that is superficially quite different from the steps described above. Recall that any stretch of a time series of length n may be written as the inverse of its discrete Fourier transform:

$$x_t = \sum_j d(f_j) e^{2\pi i f_j t}, \quad t = 0, 1, \dots, n - 1,$$

where

$$d(f) = \frac{1}{n} \sum_{t=0}^{n-1} x_t e^{-2\pi i f t}$$

and $f_j = j/n$. The sum over j may be viewed as covering frequencies $0 \leq f_j < 1$ or, alternatively, $-1/2 < f_j \leq 1/2$. The latter is the more interesting in the present context, as it displays the time series as a sum of terms with positive and negative frequencies together with the central term with $f = 0$. When n is even, there is also a term with frequency $f_{n/2} = 1/2$, which may equally be regarded as representing $f = -1/2$. Like the term at $f = 0$, this term is purely real. The series may thus be written

$$x_t = x_t^+ + x_t^- + x_t^{\text{real}}, \quad t = 0, 1, \dots, n - 1, \quad (7.16)$$

where

$$\begin{aligned}x_t^+ &= \sum_{0 < f_j < \frac{1}{2}} d(f_j) e^{2\pi i f_j t}, \\x_t^- &= \sum_{-\frac{1}{2} < f_j < 0} d(f_j) e^{2\pi i f_j t},\end{aligned}$$

and x_t^{real} consists of the remaining one or two terms. The two parts $\{x_t^+\}$ and $\{x_t^-\}$ are inherently complex, and are complex conjugates of each other. Each contains the information to recover the original series except for $\{x_t^{\text{real}}\}$:

$$x_t - x_t^{\text{real}} = 2\Re x_t^+ = 2\Re x_t^-,$$

where $\Re z$ denotes the real part of the complex number z . The part containing the positive frequencies, $\{x_t^+\}$, is the *complex time series* derived from $\{x_t\}$.

Equation (7.16) suggests a way to separate the two parts of (7.16). Suppose first that $\{R_t\}$ and $\{\phi_t\}$ vary slowly enough that $\{R_t e^{2\pi i \phi_t}\}$ may be written as

$$R_t e^{2\pi i \phi_t} = \sum_{|f_j| < \delta} \rho(f_j) e^{2\pi i f_j t} \quad \text{for some } \delta < f_0. \quad (7.17)$$

Notice that this, in fact, requires $\{R_t e^{2\pi i \phi_t}\}$ to vary slowly for all t ; since the representation is periodic, there must also be no sharp breaks between $t = -1$ and $t = 0$, $t = n - 1$ and $t = n$, and so on.

In this case (7.16) may be written

$$\begin{aligned} 2x_t &= R_t \left\{ e^{2\pi i (f_0 t + \phi_t)} + e^{-2\pi i (f_0 t + \phi_t)} \right\} \\ &= e^{2\pi i f_0 t} R_t e^{2\pi i \phi_t} + e^{-2\pi i f_0 t} \overline{R_t e^{2\pi i \phi_t}} \\ &= e^{2\pi i f_0 t} \sum_{|f_j| < \delta} \rho(f_j) e^{2\pi i f_j t} + e^{-2\pi i f_0 t} \sum_{|f_j| < \delta} \overline{\rho(f_j)} e^{-2\pi i f_j t} \\ &= \sum_{|f_j| < \delta} \rho(f_j) e^{2\pi i (f_0 + f_j)t} + \sum_{|f_j| < \delta} \overline{\rho(f_j)} e^{-2\pi i (f_0 + f_j)t}. \end{aligned}$$

Because $|f_j| < \delta < f_0$, all of the frequencies $f_0 + f_j$ in the first sum are positive, and all of those in the second are negative. Thus, if the steps outlined in (7.16) are followed, the result will be

$$\begin{aligned} x_t^+ &= \frac{1}{2} \sum_{|f_j| < \delta} \rho(f_j) e^{2\pi i (f_0 + f_j)t} \\ &= \frac{1}{2} R_t e^{2\pi i (f_0 t + \phi_t)} \end{aligned}$$

and

$$\begin{aligned} x_t^- &= \frac{1}{2} \sum_{|f_j| < \delta} \overline{\rho(f_j)} e^{-2\pi i (f_0 + f_j)t} \\ &= \frac{1}{2} R_t e^{-2\pi i (f_0 t + \phi_t)}. \end{aligned}$$

That is, $\{x_t^+\}$ and $\{x_t^-\}$ are exactly the two parts of (7.16), respectively. From either part, $\{R_t\}$ and $\{\phi_t\}$ are easily extracted.

In practice, it is unlikely that $\{R_t e^{2\pi i \phi_t}\}$ would have the band-limited representation (7.17). In particular, there is in general no reason for it to be periodic with period n , meaning that there *would* be breaks at the ends of the record when it is extended periodically. These can be smoothed out by tapering the data before carrying out this analysis. It may still be reasonable to assume that the transform $\rho(f)$ is large only for frequencies in a band $|f| < \delta < f_0$. To the extent that other values of $\rho(f)$ may be ignored, $\{x_t^+\}$ and $\{x_t^-\}$ continue to provide approximations to the two parts of (7.16).

Another problem that often arises in practice is that there are other components to a series that may obscure the identification of $\{R_t\}$ and $\{\phi_t\}$ (Section 7.1). In the earlier analysis of the sunspot series (Section 7.5), these terms were eliminated by filtering. The natural way to accomplish much the same goal in the present approach is by further limiting the set of frequencies in constructing the complex time series. This is illustrated in the next section.

7.7 SUNSPOTS: THE COMPLEX SERIES

The complex time series for the sunspot data, computed as in (7.16), is shown in Figure 7.12. The upper panel shows the amplitude, and the lower panel shows the phase as deviations from the base frequency $f_0 = 1/11$ cycles per year (i.e., the phase of $\{e^{-2\pi i f_0 t} x_t^+\}$). The data were centered at their mean and tapered 10% (5% at each end) to reduce the impact of their lack of circularity.

Evidently, the simple construction of the complex time series has not produced useful approximations to $\{R_t\}$ and $\{\phi_t\}$, especially to the former. The amplitude plot shows substantial oscillations at roughly the period of the underlying phenomenon (compare with Figure 1.3 on p. 4); the interpretation of $\{R_t\}$ as the instantaneous amplitude of the phenomenon, however, requires it to be nearly constant on the time scale of one period.

It was noted in Section 7.5 (p. 118) that the sunspot data contain components at other frequencies than the roughly 11-year period of the basic phenomenon. In that section, these other components were removed by filtering, revealing a clearer picture of the amplitude and phase variations in that phenomenon. In the spirit of the current approach, the same effect may be found by restricting the set of frequencies included in the

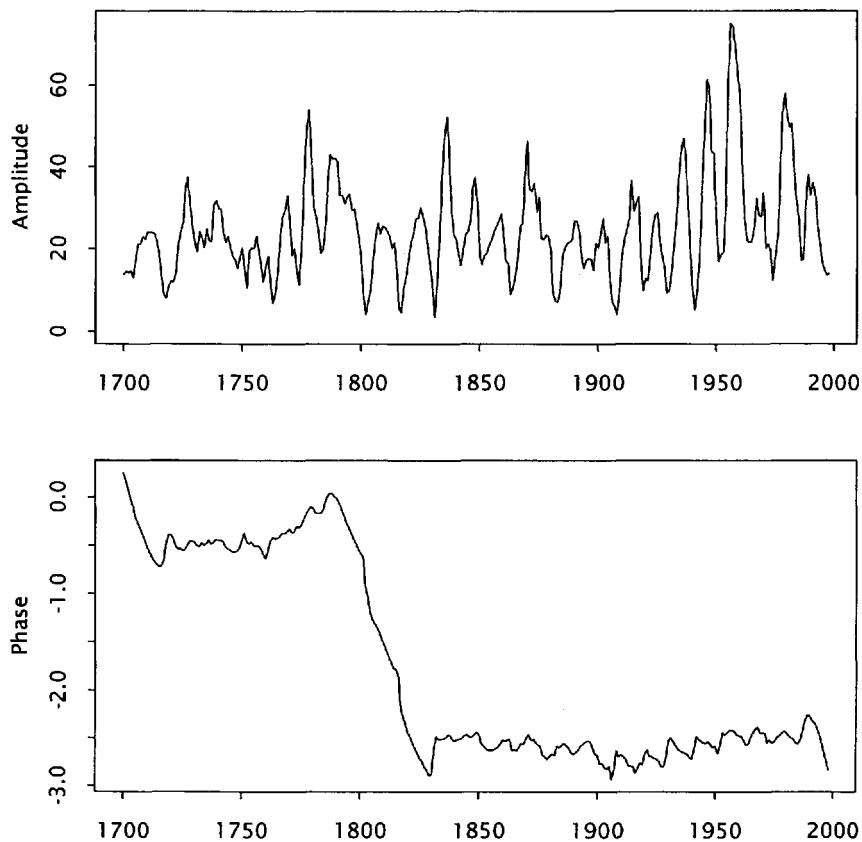


Fig. 7.12 Complex time series for the sunspot data. Amplitude (upper panel) and phase deviations from $f_0 = 1/11$ cycles per year (lower panel).

complex series:

$$x_t^+ = \sum_{f < f_j < \bar{f}} d(f_j) e^{2\pi i f_j t},$$

where the default values of \underline{f} and \bar{f} are 0 and $1/2$ cycles per year, respectively.

Figure 7.13 shows the result when $\underline{f} = 1/22$ cycles per year and \bar{f} is left at its default value of $1/2$ cycles per year. While the amplitude and phase are still far from smooth, the strong oscillation in amplitude seen in Figure 7.12 has been eliminated. The graphs are similar in quality to those in Figure 7.8.

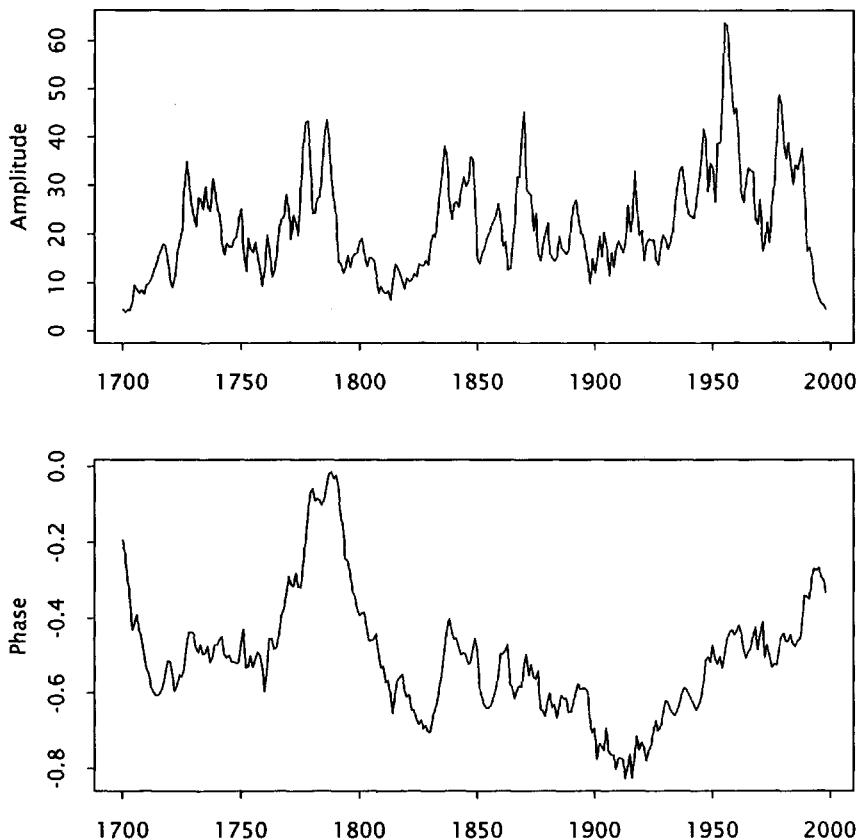


Fig. 7.13 Complex time series for the sunspot data, with $f = 1/22$ cycles per year. Amplitude (upper panel) and phase deviations from $f_0 = 1/11$ cycles per year (lower panel).

Figure 7.14 shows the result when in addition $\bar{f} = 3/22$ cycles per year. Excising the higher frequencies gives greater smoothness without changing the broad picture; the results are now more comparable with those in Figure 7.9.

Omitting a term from the sum that defines $\{x_t^+\}$ is equivalent to multiplying it by zero. Thus the series shown in Figures 7.13 and 7.14 may also be found by multiplying the discrete Fourier transform by an appropriate set of weights (all zeros and ones) before inverting it. They may, therefore, also be viewed as the result of applying circular filters to the original

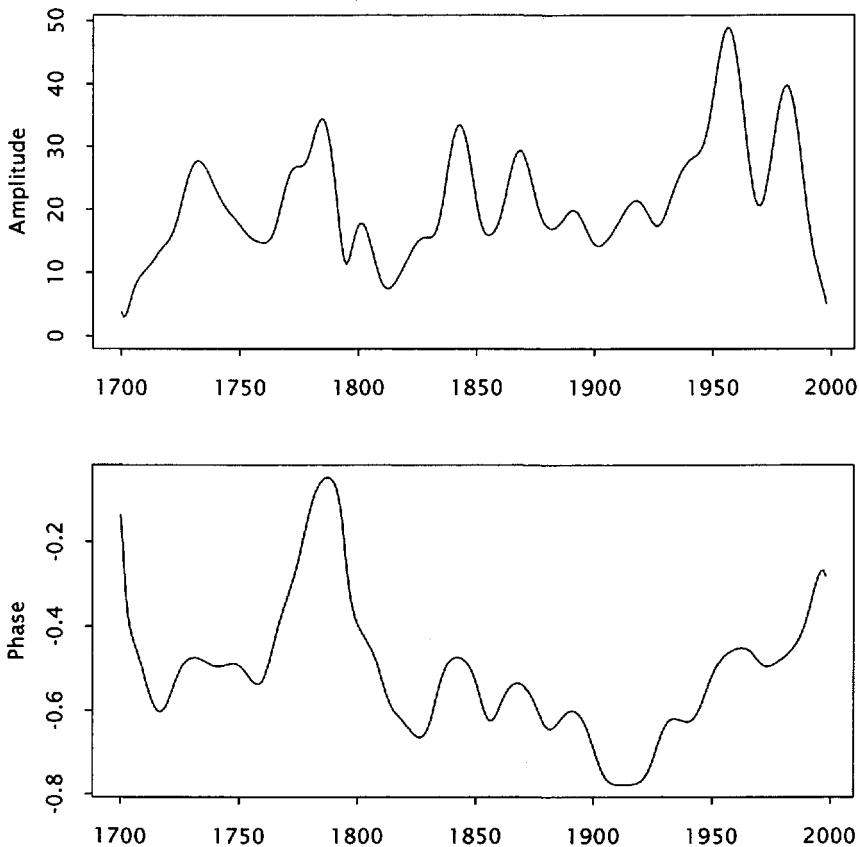


Fig. 7.14 Complex time series for the sunspot data, with $f = 1/22$ cycles per year and $\bar{f} = 3/22$ cycles per year. Amplitude (upper panel) and phase deviations from $f_0 = 1/11$ cycles per year (lower panel).

data. The coefficients in the filters, as the transforms themselves of these weights, have a complicated structure and, in particular, have substantial sidelobes. They are, therefore, quite different from the filters described in Section 7.3. However, this equivalence makes the similarity between the results not unexpected.

Appendix

The basic steps in obtaining the results shown in this chapter are:

- demodulation, followed by
- filtering.

If f_0 is the frequency of interest, the demodulated form of x may be obtained in S-PLUS by

```
y <- x * complex(real = cos(2 * pi * f0 * time(x)),
                   imag = - sin(2 * pi * f0 * time(x)))
```

where the standard S-PLUS function `complex()` is used to construct the vector $e^{-2\pi if_0 t}$ from its real and imaginary parts. The simple moving average filters are easily calculated in an explicit loop.

For graphical purposes, the instantaneous amplitude and phase may be obtained using the S-PLUS functions `Mod(z)` and `Arg(z)`, respectively. The latter is in radians, and is more useful when converted into cycles, as $\text{Arg}(z)/(2\pi)$. To avoid distracting and irrelevant discontinuities where the phase jumps from $-1/2$ to $1/2$ or back, it may be rendered continuous using the function

```
unwrap <- function(p)
{
  pd <- diff(p)
  p[] <- cumsum(c(p[1], pd - round(pd)))
  p
}
```

The S-PLUS function `demod()` provides a one-step implementation of complex demodulation using least squares filters, and was used to generate Figure 7.11. Its arguments, in addition to the time series to be demodulated and the demodulation frequency f_0 , are the pass frequency and stop frequency of the filter. It returns the instantaneous amplitude and phase, the latter unwrapped as above by default (the option exists to suppress the unwrapping). The phase is returned in radians, and should be divided by 2π to convert it to the more useful measure of cycles.

The number of terms in the filter constructed by `demod()` is controlled by the width δ of the transition band delimited by the specified pass frequency and stop frequency. If $\delta = 2/(2s + 1)$ cycles per unit time for some integer s , the result is precisely the least squares approximation to the ideal low-pass transfer function with convergence factors described in

Section 7.4. Regardless of the form of δ it has the second interpretation as the least squares approximation to the transition band version of the ideal transfer function.

The calculation in S-PLUS of the complex time series for a given time series x , discussed in Sections 7.6 and 7.7, is straightforward:

```
n <- length(x)
y <- (x - mean(x)) * taper(n, p)
y <- fft(y)/n
f <- (0:(n - 1))/n
y[f <= f1o | f >= fhi] <- 0.
y <- fft(y, inv = T)
```

Here p is the fraction of the data to be tapered, and f_{lo} and f_{hi} are the parameters f and \bar{f} used in Section 7.7 to control the character of the complex time series.

8

The Spectrum

The results of harmonic analysis can be difficult to interpret, even when the data show definite periodicity in the form of successive, fairly regular, peaks and troughs. The sunspot series analyzed in Chapters 6 and 7 is ample evidence of this fact.

What, then, can be achieved by harmonic analysis of a series with less well defined oscillations, such as Beveridge's wheat price series (Figure 1.4, p. 5)? These data were collected and published by Beveridge (1921) as part of a study of the impact of meteorological variables on economic conditions. They are a price index for wheat in Europe, normalized to make the average price for 1700 to 1745 equal to 100. Historical records from 48 separate locations were combined to produce the index. Although the graph of the data shows a succession of peaks and troughs, these are by no means as regular as those of the sunspots. Nevertheless, harmonic analysis (and its close relative, *spectrum analysis*) of such economic series is widely used. The kind of information it can yield will be explored in this chapter.

8.1 PERIODOGRAM ANALYSIS OF WHEAT PRICES

The first problem in analyzing these data is the change in scale. No sinusoid can match oscillations that grow in amplitude. Beveridge produced

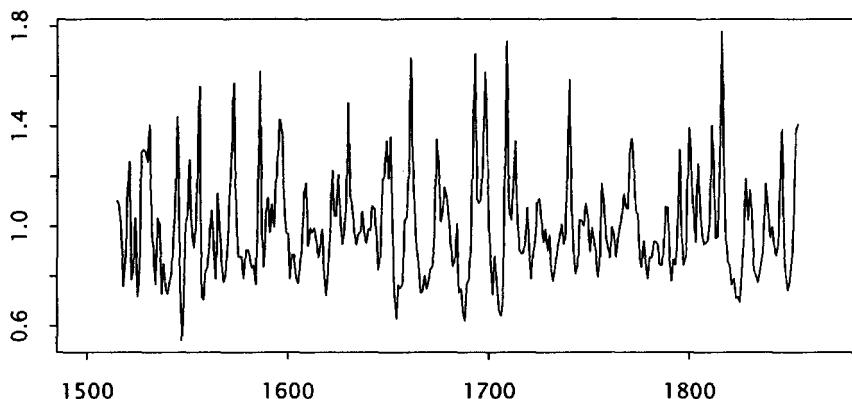


Fig. 8.1 Index of fluctuation of wheat prices in Western Europe, 1515 to 1854.

an “index of fluctuation,” shown in Figure 8.1, by dividing each value in the series by the average of 31 centered adjacent values. The oscillations in this index are more uniform and, in particular, show no tendency for their amplitude to change over time. Beveridge (1921) gave various values of the periodogram of the index of fluctuations, and in a later paper (Beveridge, 1922) presented a more extensive analysis, including some corrections of the earlier values.

The construction of the index of fluctuations may be motivated by modeling the data as $x_t = T_t I_t$, where T_t is the *trend* at time t and I_t is an *irregular* or oscillating term. The interpretation of these terms is that the trend reflects long-term economic forces such as inflation, whereas the irregular terms are caused by short-term effects such as fluctuations in supply from year to year. The wheat-price series was constructed to allow examination of these short-term fluctuations, and the trend term is an unwanted complication. The 31-year moving average may be regarded as an approximation to the trend term, and thus the index of fluctuations is the corresponding estimate of the irregular component.

However, operations such as these introduce their own effects into the data, as is most easily seen in the case of the additive model $x_t = T_t + I_t$. A natural way to estimate T_t is by applying a linear filter (see Section 7.2) to the data $\{x_t\}$. Suppose that the filter has weights $\{g_u : -s \leq u \leq s\}$ and transfer function $G(f)$. Then T_t is approximated by

$$y_t = \sum_{u=-s}^s g_u x_{t-u}. \quad (8.1)$$

Now $\{T_t\}$ is assumed to be smooth, and thus $T_{t-u} \approx T_t$ for $-s \leq u \leq s$, provided that s is reasonably small. Hence

$$\begin{aligned}\gamma_t &= \sum_{u=-s}^s g_u T_{t-u} + \sum_{u=-s}^s g_u I_{t-u} \\ &\approx T_t + \sum_{u=-s}^s g_u I_{t-u},\end{aligned}$$

provided $\sum_u g_u = 1$, which is the natural normalization. Then I_t is approximated by

$$\begin{aligned}z_t &= x_t - \gamma_t \\ &\approx I_t - \sum_{u=-s}^s g_u I_{t-u},\end{aligned}$$

which is the result of applying a linear filter to $\{I_t\}$. The transfer function of the filter is

$$1 - \sum_u g_u e^{-2\pi i f u} = 1 - G(f).$$

Thus the series $\{z_t\}$, constructed as an approximation to $\{I_t\}$, is really an approximation to a *filtered version* of $\{I_t\}$.

In the case of the multiplicative model $x_t = T_t I_t$, the trend term is approximated in the same way by applying a linear filter to $\{x_t\}$, but it is then removed by division instead of subtraction. The resulting series $z_t = x_t / \gamma_t$, the general analog of Beveridge's index of fluctuations, is also approximately a filtered version of $\{I_t\}$, with the same transfer function $1 - G(f)$, as is shown by Granger and Hughes (1971) (see also Exercise 8.1). For the index of fluctuations, $G(f) = D_{31}(f)$, a Dirichlet kernel. From the results of Section 7.2, it follows that the transforms of $\{z_t\}$ and $\{I_t\}$, $d_z(f)$ and $d_I(f)$, respectively, are related by

$$\begin{aligned}d_z(f) &\approx \{1 - G(f)\} d_I(f) \\ &= \{1 - D_{31}(f)\} d_I(f).\end{aligned}$$

Thus the periodogram of $\{z_t\}$ consists approximately of that of $\{I_t\}$ multiplied by $|1 - D_{31}(f)|^2$, the corresponding power transfer function.

A graph of the function $|1 - D_{31}(f)|^2$ appears in Figure 8.2. The function has the value 1 whenever $D_{31}(f)$ vanishes, that is, at $f = k/31$ cycles per year, $k = 1, 2, \dots$. There are alternate peaks and troughs between these points, occurring roughly at $(k + 1/2)/31$ cycles per year, $k = 1, 2, \dots$, with the first peak almost reaching the value 1.5. The corresponding period is $62/3 = 20.67$ years, and it follows that the periodogram values around this period are considerably amplified. The

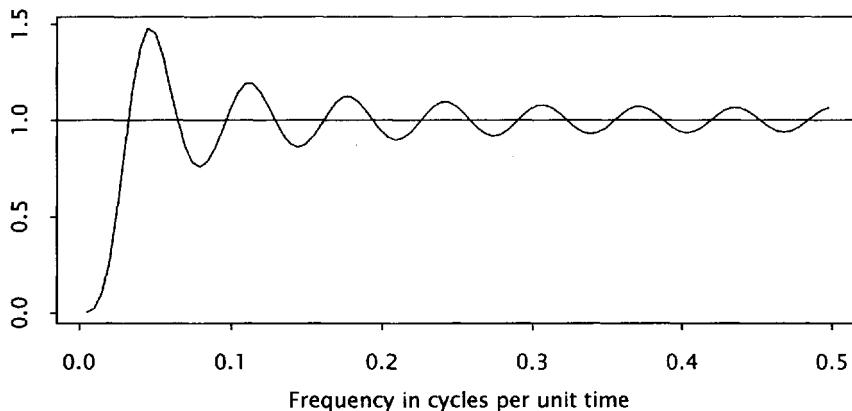


Fig. 8.2 The power transfer function $|1 - D_{31}(f)|^2$.

first trough, by contrast, falls almost to the value 0.75 and occurs near $f = 2.5/31$ cycles per year, corresponding to a period of $62/5 = 12.4$ years. The periodogram is correspondingly attenuated around this period.

The enhancement of the periodogram at periods near 21 years and its attenuation near 12 years could easily lead one to infer the existence of a peak in the periodogram at around 21 years. The possibility that spurious periodicities may be introduced into data by operations involving linear filters was raised by Slutsky in 1927 (see, e.g., Slutsky, 1937), and the phenomenon is known by his name. Schuster (1898) was also aware that linear filtering may cause such distortions in the transform of a series.

Table 8.1 shows the effect on the periodogram of the wheat-price series. The first two columns contain the integer periods from 2 to 36 years and the periodogram ordinates given by Beveridge (1921), ordered by the magnitudes of the ordinates.¹ The last three columns show the periods, the periodogram ordinates corrected by dividing by $\{1 - D_{31}(f)\}^2$, and the rank of the ordinate before correction. As might be expected, the 20-year period moves down several places, while the 12- and 13-year periods move up. Granger and Hughes (1971) describe the effect of a similar correction and find that the largest peak is moved from a 15.2-year period, close to Beveridge's figure of 15.3 years, to 13.3 years.

A different procedure is suggested by the model $x_t = T_t I_t$. All three

¹Some of these values were revised by Beveridge (1922), but not substantially.

Table 8.1 Beveridge's periodogram for integer periods,
before and after corrections for the power transfer function.

Original Analysis		After Correction		
Period (years)	Periodogram Ordinate	Period (years)	Periodogram Ordinate	Previous Rank
15	47.28	15	50.48	1
11	40.93	11	46.51	2
20	32.44	36	36.92	7
17	29.35	13	36.41	5
13	27.81	12	25.99	9
24	26.48	17	24.59	4
36	26.27	35	36.92	10
16	20.14	20	22.39	3
12	20.11	16	18.90	8
35	17.52	24	18.51	6
18	17.26	34	13.50	15
25	14.95	18	13.23	11
6	12.29	7	12.12	16
8	12.05	6	11.53	13
34	11.04	8	11.31	14
7	10.43	25	10.81	12
30	7.86	30	7.38	17
23	7.54	23	5.15	18
22	7.50	22	5.06	19
21	6.33	10	5.06	21
10	5.39	31	5.04	23
29	5.05	5	4.73	24
31	5.04	29	4.46	22
5	4.43	21	4.28	20
4	3.44	4	3.23	25
26	3.35	33	3.22	27
33	2.82	26	2.52	26
9	2.39	9	2.00	28
28	2.00	28	1.67	29
19	1.50	32	1.12	32
27	1.25	19	1.08	30
32	1.05	27	0.99	31
3	0.23	3	0.25	33
14	0.02	14	0.02	34
2	0.01	2	0.01	35

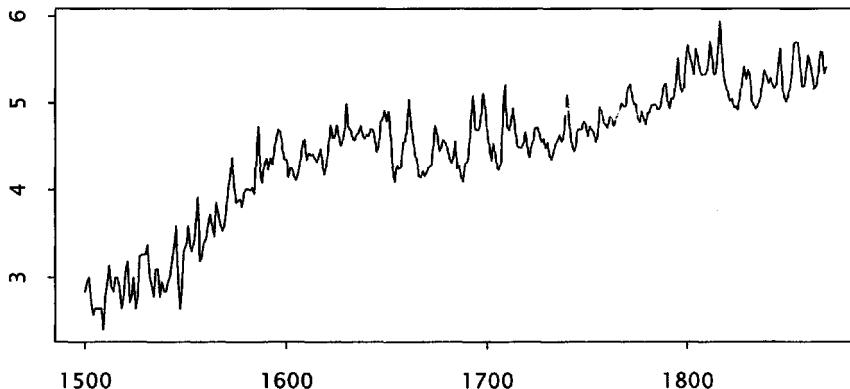


Fig. 8.3 Logarithms of the index of wheat prices.

terms are strictly positive, and the model may be written

$$\log x_t = \log T_t + \log I_t.$$

Thus the logarithms of the data should consist of a smooth term, $\log T_t$, with an *added* irregular component, $\log I_t$. The trend term T_t is naturally thought of as the typical value of x_t in the neighborhood of t , and then I_t is a dimensionless quantity close to 1. Hence $\log I_t$ fluctuates around 0. Figure 8.3, which shows the logarithms of the original data, is a considerable improvement over the original data (Figure 1.4, p. 5), in that the fluctuations show no definite tendency either to increase or to decrease in magnitude over time. This is, in fact, to be expected, given the similarly constant amplitude of the oscillations in the index of fluctuations shown in Figure 8.1 (see Exercise 8.2).

The logarithms also show some advantages over the index of fluctuations, however, in that the spikiness of the peaks has been reduced. Section 6.5 showed that nonsinusoidal behavior such as a marked disparity between the natures of the peaks and those of the troughs of a series introduces structure into its Fourier transform that is not revealed by the periodogram. Thus transformation of a series to remove or reduce such disparities is desirable. The logarithmic transformation goes some way toward achieving this for the wheat-price series.

Beveridge (1921) argued that the early part of the series is unreliable, as it is based on data from relatively few sources, and that the later part of the series is of a different nature because of economic changes in the

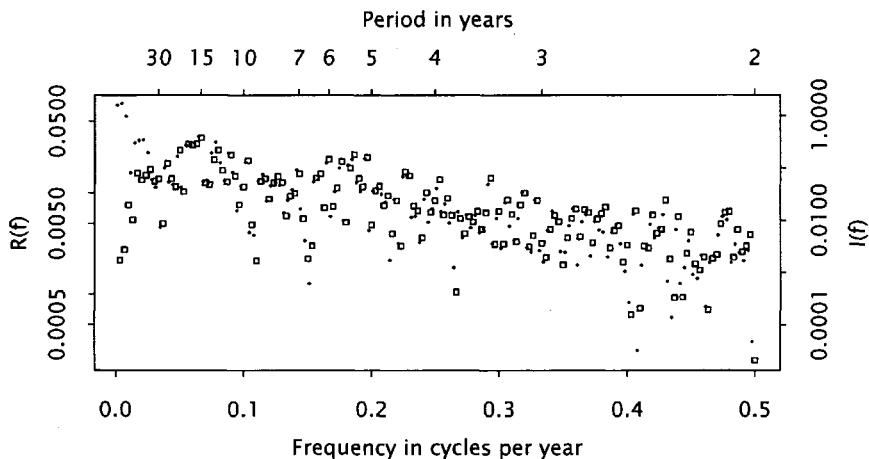


Fig. 8.4 Periodograms of the index of fluctuations (squares) and logarithms of the index of wheat prices (dots).

nineteenth century. He chose the years 1545–1844 for his periodogram analysis, giving a series of 300 terms. Figure 8.4 shows the periodograms of the index of fluctuations and of the natural logarithms (see Exercise 8.2) of the original data, for the same time window. The index of fluctuations was first corrected for its typical value of 1 and then tapered 20%. The logarithms of the index were detrended by subtracting the least squares straight line and then similarly tapered 20%.

The periodogram of the logarithms is considerably larger than that of the index of fluctuations at low frequencies, since even after a trend-line has been removed, obvious low frequency terms remain. Near $f = 0.05$ cycles per year the ordering is reversed. This is the range of frequencies where the transfer function in Figure 8.2 reaches its first and highest peak, amplifying the periodogram of the index of fluctuations. Over the remaining frequencies the ordering varies, although the values are generally similar.

Exercise 8.1 Removing a Multiplicative Trend

Suppose that a series $\{x_t\}$ may be represented as $x_t = T_t I_t$, where $\{T_t\}$ is a smooth trend series, and $\{I_t\}$ is an irregular series consisting of small fluctuations around 1. Suppose that T_t is approximated by y_t as in (8.1). Show that $z_t = x_t / y_t$ satisfies

$$z_t \approx 1 + I_t - \sum_u g_u I_{t-u}.$$

(Hint: Since T_t is smooth,

$$y_t \approx T_t \sum_u g_u I_{t-u},$$

and hence

$$z_t \approx \frac{I_t}{\sum_u g_u I_{t-u}}.$$

Write $I_t = 1 + \epsilon_t$, where $|\epsilon_t| \ll 1$, and expand using a Taylor series.)

Exercise 8.2 (Continuation) Logarithmic Transformation

Suppose that the series $\{x_t\}$ is as in Exercise 8.1. Let

$$y_t = \sum_u g_u \log x_{t-u}$$

and $z_t = \log x_t - y_t$. Show that

$$z_t \approx I_t - \sum_u g_u I_{t-u}.$$

Note that this implies that the fluctuations in the logarithms of a series should be approximately the same as the quotients defined in Exercise 8.1—for example, Beveridge's index of fluctuations.

8.2 ANALYSIS OF SEGMENTS OF A SERIES

To investigate the consistency of the periodicities in the wheat-price series, Beveridge (1921) also gives some terms from the periodograms of the two halves of the series (as suggested by Schuster, 1898), 1545–1694 and 1695–1844, respectively. Corresponding periodograms of the logarithms of the series are shown in Figure 8.5. Each half of the series was detrended and tapered as in Section 8.1.

The two periodograms have the same general shape. They are large at the lowest frequencies and show a broad peak between 0.06 and 0.10 cycles per year, corresponding to periods of around 10 to 17 years. They then show a gentle decline over the rest of the periodogram, with small fluctuations superimposed. However, the fine structures of the two periodograms are quite unrelated: a local peak in one is just as likely to be matched by a local trough as a local peak in the other. In other words, the

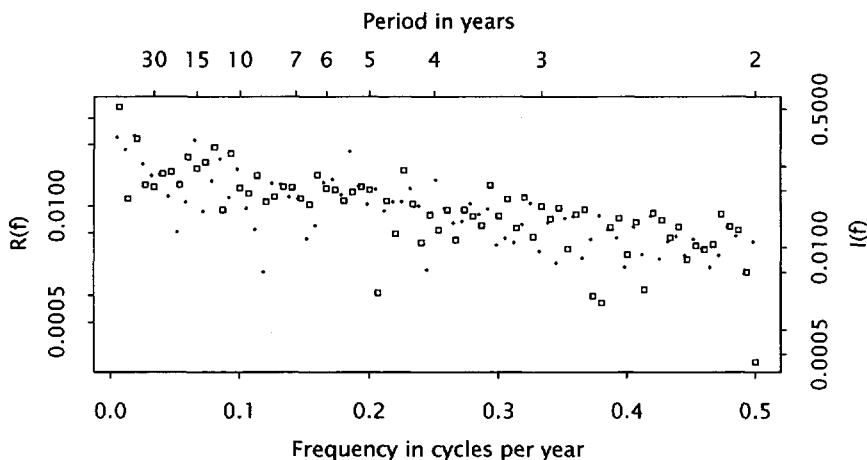


Fig. 8.5 Periodograms of logarithms of two halves of the index of wheat prices (1545–1694 = squares, 1695–1944 = dots).

fine structure is not *repeated* from one segment to the next, but the broad features show a *statistical regularity* or *consistency* across segments.

Thus the fine structure of the periodogram of these data is not characteristic of the series as a whole, but depends on the particular segment being analyzed. On the other hand, the broad features of the periodogram do not appear to vary in this way, and may be characteristic of the series as a whole. Hence if interest is focused on the series as a whole, the fine structure should be ignored in favor of the broad features of the periodogram.

By way of contrast, Figure 8.6 shows the periodograms of the first and second halves of the variable star data. Each half was centered at its mean and tapered 20%. Here the peaks are perfectly aligned with each other, and clearly the fine structure *is* characteristic of the whole series.

Thus different series may be characterized by different aspects of the periodogram. For series similar to the variable star data, the periodogram itself is a useful tool. For series like Beveridge's, the fine details need to be suppressed, so that the broad behavior that is common to different segments and characteristic of the series as a whole can be seen. In this case, the periodograms of different segments may be regarded as the same underlying smooth curve, with different segment-specific fluctuations superimposed. As with the complex demodulation problem discussed in Chapter 7, approximations to the underlying smooth curve are often constructed by *linearly filtering* or *smoothing* the periodogram.

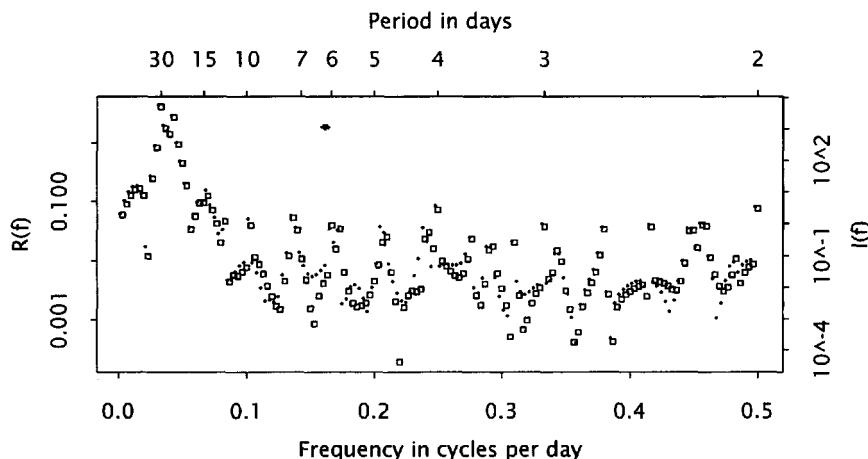


Fig. 8.6 Periodograms of two halves of the variable star series.

The underlying smooth curve is called the *spectrum* of the series. The theory to be sketched in Chapter 9 shows that such a curve exists for many time series models. Constructing an estimate of the spectrum is one of the possible uses of the periodogram for this kind of time series data.

8.3 SMOOTHING THE PERIODOGRAM

To see the common features of the two periodograms of Figure 8.5, a simple procedure would be to graph the average of the two ordinates at each frequency. Since the periodogram is typically graphed on a logarithmic scale, either the average of the logarithms or the logarithm of the averages could be used. The latter is preferable, since it puts more weight on larger values than on small ones, and the small values are the most susceptible to perturbations of all kinds, particularly leakage from other frequencies.

More generally, the data could be divided into a number of segments, the periodograms computed and averaged, and graphed on a logarithmic scale. This procedure was first suggested by Bartlett (1948) (see also Bartlett, 1950; Kendall, 1948). Figure 8.7 shows three such averages, with 2, 15, and 60 segments, respectively. Each segment was tapered 20%, and to partially compensate for the resulting loss of data, the segments were arranged so that each overlapped its two neighbors by 10%. As a result, each data point received either full weight in a single segment or partial

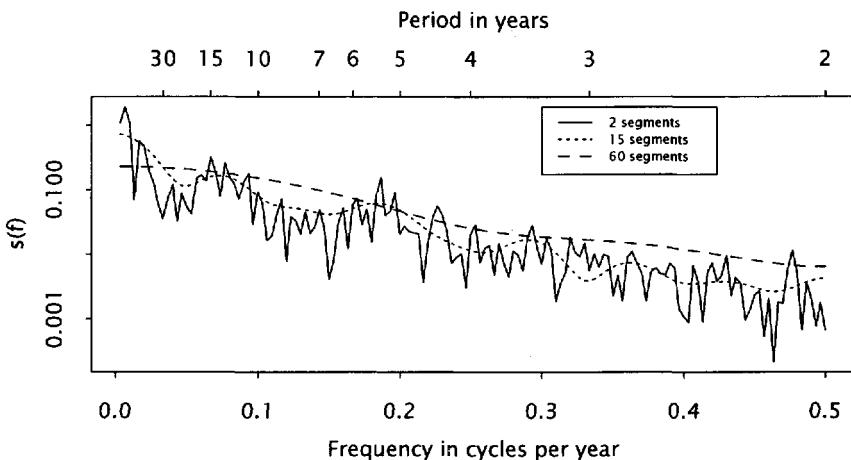


Fig. 8.7 Averaged segment periodograms for the logarithms of the wheat prices, 1545–1844

weight in two segments.

The three graphs show that varying degrees of smoothness may be attained in this way, ranging from essentially none to quite severe. To understand the process better, an alternative representation of the periodogram will help.

The complex form of the periodogram will be used, for convenience. Also for convenience, the data will be treated as deviations around zero, so that no centering is needed. The data are x_0, x_1, \dots, x_{n-1} , their transform is

$$d(f) = \frac{1}{n} \sum_t x_t e^{-2\pi i f t},$$

and the periodogram is $I(f) = n|d(f)|^2$. Since $|d(f)|^2$ can be found as $d(f)$ times its complex conjugate $\overline{d(f)}$, the periodogram may be rearranged as

$$\begin{aligned} I(f) &= n \times d(f) \times \overline{d(f)} \\ &= \frac{1}{n} \sum_t x_t e^{-2\pi i f t} \sum_{t'} x_{t'} e^{2\pi i f t'} \\ &= \frac{1}{n} \sum_t \sum_{t'} x_t x_{t'} e^{-2\pi i f(t-t')}. \end{aligned}$$

There are n^2 terms in this double sum, indexed by t and t' . However, since the exponential factor in each depends only on $t - t'$, rather

than on t and t' separately, and $t - t'$ takes on only the $2n - 1$ values $-n + 1, -n + 2, \dots, -1, 0, 1, \dots, n - 2, n - 1$, it follows that there are only $2n - 1$ corresponding values of the exponential factor. The expression may therefore be rewritten

$$\begin{aligned} I(f) &= \frac{1}{n} \sum_{r=-n+1}^{n-1} \sum_{t-t'=r} x_t x_{t'} e^{-2\pi i f(t-t')} \\ &= \frac{1}{n} \sum_{r=-n+1}^{n-1} \sum_{t-t'=r} x_t x_{t'} e^{-2\pi i f r} \\ &= \frac{1}{n} \sum_{r=-n+1}^{n-1} e^{-2\pi i f r} \sum_{t-t'=r} x_t x_{t'}. \end{aligned}$$

The limits on the inner sum are not explicit. It is read simply as the sum over all pairs (t, t') with $t - t' = r$, and listing these pairs is elementary if a little tedious.

The periodogram is thus itself a Fourier series, in which the coefficient of $e^{-2\pi i f r}$ is $n^{-1} \sum_{t-t'=r} x_t x_{t'}$. Some manipulation yields

$$I(f) = \sum_{|r| < n} c_r e^{-2\pi i f r}, \quad (8.2)$$

where

$$c_r = \begin{cases} \frac{1}{n} \sum_{t=r}^{n-1} x_t x_{t-r} & r \geq 0, \\ c_{-r} & r < 0 \end{cases} \quad (8.3)$$

(see Exercise 8.3). The quantity c_r is the *autocovariance* of $\{x_t\}$ at lag r . Because of the symmetry of the autocovariances, (8.2) may also be written as

$$I(f) = c_0 + 2 \sum_{r=1}^{n-1} c_r \cos 2\pi f r.$$

Suppose that the data are divided into k segments, for convenience not overlapping and not tapered, each of length $m = n/k$. Write $I_j(f)$ for the periodogram of the j th segment, and $\{c_{j,r}\}$ for the corresponding

autocovariances. The average of these periodograms is

$$\begin{aligned}\hat{s}(f) &= \frac{1}{k} \sum_{j=1}^k I_j(f) \\ &= \sum_{|r| < m} \left(\frac{1}{k} \sum_{j=1}^k c_{j,r} \right) e^{-2\pi i f r} \\ &= \sum_{|r| < m} \left(1 - \frac{|r|}{m} \right) \frac{1}{k(m-|r|)} \sum_{j=1}^k m c_{j,r} e^{-2\pi i f r}.\end{aligned}\quad (8.4)$$

Now $\sum_j m c_{j,r}$ is, like nc_r , a sum of products of the form $x_t x_{t+r}$, but not of all such products that are available: the term $x_t x_{t+r}$ is included only if x_t and x_{t+r} fall in the same segment. There are $m - |r|$ terms in each $c_{j,r}$, and thus

$$\frac{1}{k(m-|r|)} \sum_{j=1}^k m c_{j,r}$$

is the average of these products. Replacing this part of (8.4) by

$$\frac{nc_r}{n-|r|} = \frac{c_r}{1-|r|/n}$$

seems reasonable, since this is the average of *all* available products of this form. This replacement yields the modified function

$$\hat{s}_B(f) = \sum_{|r| < m} \frac{1-|r|/m}{1-|r|/n} c_r e^{-2\pi i f r} \quad (8.5)$$

$$= \sum_{|r| < m} w_r c_r e^{-2\pi i f r}, \quad (8.6)$$

say, where

$$w_r = \frac{1-|r|/m}{1-|r|/n}. \quad (8.7)$$

The function $\hat{s}_B(f)$ defined in (8.5) is known as the *Bartlett spectrum estimate*, and the quantities $\{w_r\}$ are the corresponding *lag weights*.

If, more generally, the segments are tapered and overlapped, as in the computations underlying Figure 8.7, a similar analysis is possible. The coefficient of $e^{-2\pi i f r}$ is now a more general weighted sum of the products $x_t x_{t+r}$. In the extreme case where consecutive segments differ by only one time point, the calculation is essentially the same as complex demodulation, and the weights on these products become equal, except for some end terms (see Exercise 8.8, p. 154).

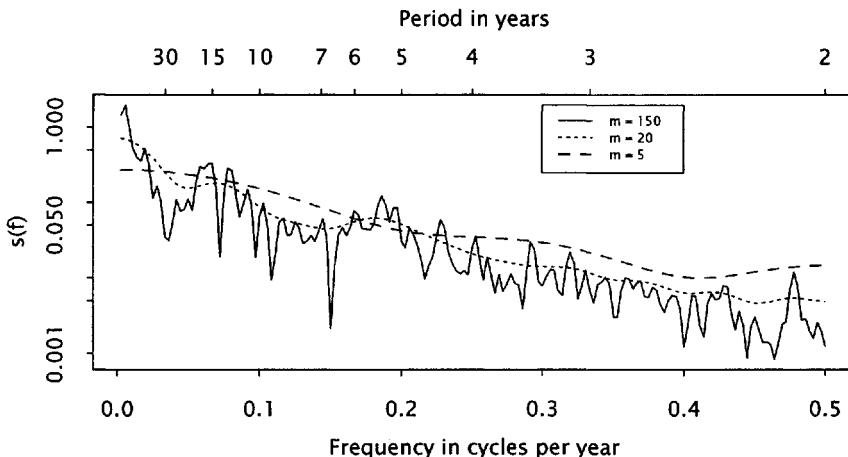


Fig. 8.8 Bartlett spectrum estimates for the logarithms of the wheat prices, 1545–1844

The Bartlett spectrum estimate differs from the periodogram in two ways:

- All terms with $|r| \geq m$ have been omitted.
- The remaining terms are progressively reduced in magnitude by the factor w_r .

Both effects make the function smoother than the periodogram. Furthermore, varying the *truncation point* m (which no longer needs to be required to divide n) provides control over the degree of smoothness.

One characteristic of the Bartlett spectrum estimate is that, unlike the averaged periodograms to which it is an approximation, it is not guaranteed to be positive. Figure 8.8 shows the three estimates corresponding to those in Figure 8.7; these are, in fact, positive at all frequencies. The calculation was carried out along the lines described in Section 8.4. The data were detrended and tapered 20% before calculation of the autocovariances.

Modifying the periodogram to make it smoother is analogous to the use of a *data window* (see Section 6.2) on a data series. In each case a series of quantities is tapered at its end to make its Fourier transform smoother. It is due partly to historical accident and partly to a difference in interpretation that the classes of tapers used are different in each case.

As in the case of a data window, the significant property of the lag weights is that they decay smoothly from the value 1 as $r = 0$ to 0 at

$r = \pm m$, and many such sets of weights have been used to construct spectrum estimates. Anderson (1994) lists the most commonly used sets of weights and their properties. Note that, in contrast with the tapers used as data windows, the Bartlett weights begin their decay linearly, since for small r , $w_r \approx 1 - |r|(m^{-1} - n^{-1})$. All the other sets of weights used to construct spectrum estimates have the property that the weights stay closer to 1 for small r , and typically behave like $1 - \alpha r^2$ for some $\alpha > 0$. For this and related reasons (see Section 8.5) the Bartlett estimate is rarely used in practice.

The most obvious way to compute any spectrum estimate of the form (8.6) is by

- calculating the autocovariances from (8.3), and
- summing the terms in (8.6).

This strategy is especially attractive when the truncation point m is small, since relatively few autocovariances need to be calculated and the sum (8.6) has relatively few terms. The next section shows that *all* of the autocovariances $\{c_r\}$ may be computed efficiently using the fast Fourier transform and its inverse. The calculation (8.6), which is itself a Fourier transform, may then be carried out using a third application of the fast Fourier transform. This method involves less computation when the truncation point m is large. A simpler way to construct spectrum estimates, involving even less computation, is described in Section 8.5

Exercise 8.3 Alternative Expression for the Periodogram

Verify that the periodogram may be written as in (8.2), where the autocovariances $\{c_r\}$ are defined by (8.3).

8.4 COMPUTING AUTOCOVARIANCES AND LAG-WEIGHTS SPECTRUM ESTIMATES

Computing the r th autocovariance of $\{x_t\}$ directly from (8.3) requires $n - r$ multiplications and $n - r - 1$ additions. Computing c_0, c_1, \dots, c_m therefore requires $m\{n - (m - 1)/2\}$ multiplications and $m\{n - (m + 1)/2\}$ additions, or roughly mn of each. However, (8.3) displays the periodogram as the Fourier transform of the autocovariances, raising the possibility of using the inverse fast Fourier transform to obtain the autocovariances (see, e.g., Gentleman and Sande, 1966).

When the periodogram is evaluated at a Fourier frequency $f_j = j/n$, it may be rewritten

$$I(f_j) = \sum_{r=0}^{n-1} (c_r + c_{n-r}) e^{-2\pi i f_j r} \quad (8.8)$$

(see Exercise 8.4), provided c_r is defined to be 0 for $|r| \geq n$. Therefore

$$c_r + c_{n-r} = \frac{1}{n} \sum_{j=0}^{n-1} I(f_j) e^{2\pi i f_j r},$$

a computation which may be carried out using the inverse fast Fourier transform. However, the only part of the autocovariance sequence that this yields is c_0 , which emerges in the case $r = 0$ because $c_n = 0$. All other autocovariances appear in pairs. If r is small relative to n , the term c_{n-r} contains many fewer products than c_r , whence, in general, $c_r + c_{n-r} \approx c_r$, and the computation may be viewed as giving approximations to the autocovariances at small lags.

The discrete Fourier transform and the periodogram are, however, easily obtained on a finer grid $f'_j = j/n'$ for some $n' > n$ (see Section 5.3), by *padding* the data with a block of $n' - n$ zeros. At these frequencies, (8.8) becomes

$$I(f'_j) = \sum_{r=0}^{n-1} (c_r + c_{n'-r}) e^{-2\pi i f'_j r} \quad (8.9)$$

and inversion now gives

$$c_r + c_{n'-r} = \frac{1}{n'} \sum_{j=0}^{n'-1} I(f'_j) e^{2\pi i f'_j r},$$

yielding $c_0, c_1, \dots, c_{n'-n}$ exactly. In particular, if $n' = 2n - 1$, all autocovariances are obtained.

Since the computational cost of the fast Fourier transform of a series of length n' is roughly $2n' \log_2 n'$, the first m autocovariances may be found with a cost of roughly $2(m+n) \log_2(m+n)$, and they may all be found with a cost of roughly $4n \log_2 2n$. If m and n are both large, the fast Fourier transform approach offers substantial computational savings.

Once the autocovariances have been computed, the fast Fourier transform may be used one more time to construct the spectrum estimate for any set of lag weights $\{w_r\}$. Suppose that the spectrum estimate is needed on a grid of frequencies $f'_j = j/n'$, $0 \leq f'_j \leq 1/2$, where n' is no

longer assumed to be larger than n . In fact, since a spectrum estimate is typically a smooth function of frequency, often relatively few frequencies are required. To use the fast Fourier transform, the weighted series of autocovariances is first extended to length n' :

$$d_r = \begin{cases} w_r c_r, & 0 \leq r < n \\ 0, & m \leq r < n' \end{cases}$$

and $e_r = d_r + d_{n'-r}$, $r = 0, 1, \dots, n' - 1$. An argument similar to that used to obtain (8.9) shows that the transform of $\{e_r\}$ is the required spectrum estimate.

Exercise 8.4 Inverting the Periodogram

Show that the periodogram, evaluated at a Fourier frequency, may be written as in (8.8). (*Hint:* Use (8.2), the symmetry of the autocovariances $\{c_r\}$, and the special properties of the Fourier frequencies.)

Note that (8.9) follows by an identical argument.

8.5 ALTERNATIVE REPRESENTATIONS OF A SPECTRUM ESTIMATE

Suppose that a spectrum estimate $\hat{s}(f)$ is given by

$$\hat{s}(f) = \sum_{|r| < n} w_r c_r e^{-2\pi i f r}, \quad (8.10)$$

where it is no longer assumed that all the lag weights vanish beyond some truncation point. By the integral inversion formula (see Exercise 4.3, p. 40) applied to the periodogram,

$$c_r = \int_0^1 I(f) e^{2\pi i f r} df.$$

Then

$$\hat{s}(f) = \int_0^1 W_n(f - f') I(f') df', \quad (8.11)$$

where

$$W_n(f) = \sum_{|r| < n} w_r e^{-2\pi i f r}$$

(see Exercise 8.5). Hence any spectrum estimate of the form (8.10) may be written as an integral average of the periodogram. Equation (8.11) is the integral analog of the discrete linear filters described in Section 7.2.

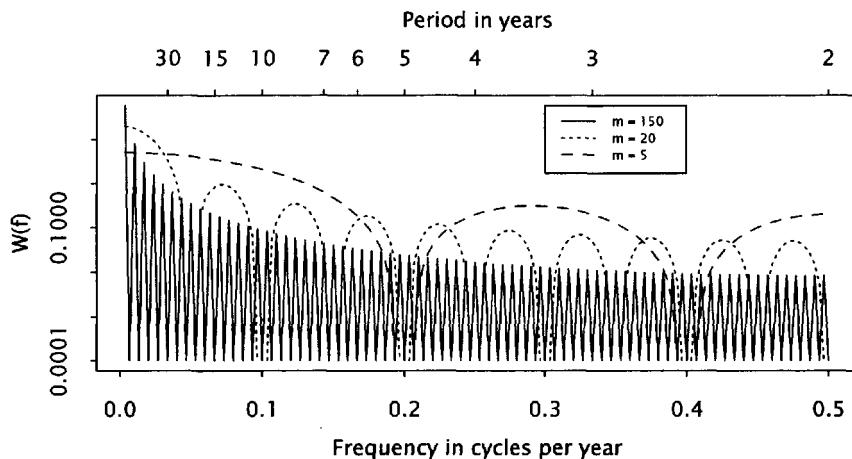


Fig. 8.9 Spectral windows $W_n(f)$ for the Bartlett spectrum estimates of Figure 8.8.

The function $W_n(f)$ is called the *spectral window* associated with the spectrum estimate. The spectral window of the Bartlett estimate (8.5) cannot be written in closed form, but is easily computed. Figure 8.9 shows the spectral windows for the estimates in Figure 8.8. Each window is negative near the frequencies $f = j/m$, $j = 1, 2, \dots$; nonpositive values are omitted from this logarithmic plot.

For a modified Bartlett estimate with lag weights

$$w_r = \begin{cases} 1 - |r|/m, & |r| < m \\ 0, & |r| \geq m \end{cases} \quad (8.12)$$

the spectral window is

$$\begin{aligned} W_n(f) &= \sum_{|r| < m} (1 - |r|/m) e^{-2\pi i f r} \\ &= m D_m(f)^2, \end{aligned}$$

(see Exercise 8.6), where $D_m(f)$ is the Dirichlet kernel (see Section 2.2). Since this spectral window is nonnegative, the integrand in (8.11) is also nonnegative, and consequently the spectrum estimate is guaranteed to be nonnegative. Figure 8.10 shows the modified Bartlett spectrum estimates for the logarithms of the wheat price series corresponding to the Bartlett estimates of Figure 8.8. The spectral windows are shown in Figure 8.11. These windows vanish at the frequencies $f = j/m$, $j = 1, 2, \dots$, and these values are omitted from the logarithmic plot.

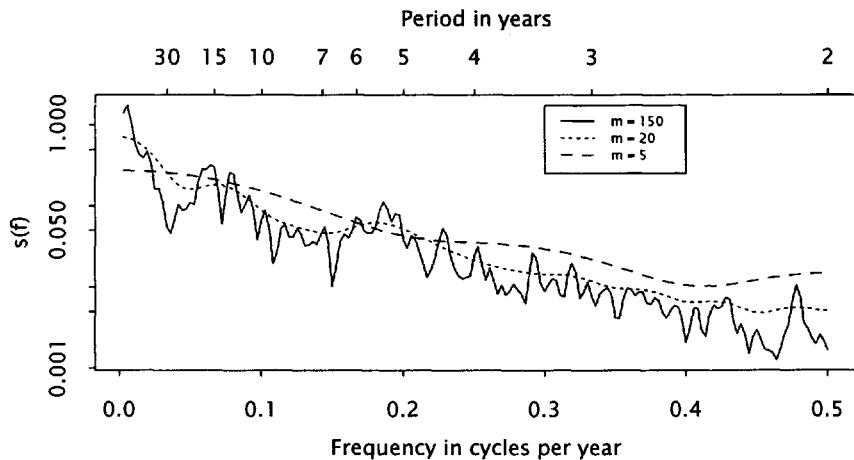


Fig. 8.10 Modified Bartlett spectrum estimates for the logarithms of the wheat prices, 1545-1844.

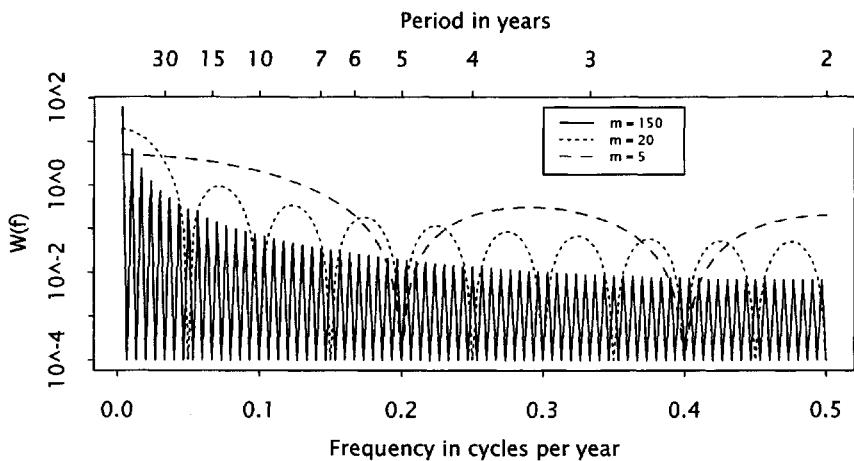


Fig. 8.11 Spectral windows $W_n(f)$ for the modified Bartlett spectrum estimates of Figure 8.10.

The central peak in the spectral window of the modified Bartlett estimate is of height m and the first zeros on either side are at $f = \pm 1/m$ cycles per unit time. However, a sizable proportion of the mass of $W_n(f)$ is contained not in the main peak but in the sidelobes. These decay slowly, and hence periodogram values at some distance from f may contribute substantially to the integral. Previous examples have demonstrated that

periodogram ordinates may vary by several orders of magnitude, in which case the estimated spectrum in one frequency band where there is little power may be swamped by such *leakage* from another band with high power, even when these bands are not adjacent.

This source of leakage is quite distinct from the ones mentioned in previous chapters. It was shown in Chapter 6 that if no data window is used, there may be substantial leakage in the periodogram itself. However, even if leakage in the periodogram is controlled by the use of a data window, leakage may appear in a smoothed version of the periodogram because of sidelobes in the spectral window.

The sidelobes of the modified Bartlett window are larger and decay more slowly than those of the most commonly used estimates (see Anderson, 1994, Chapter 9), and for this reason it is rarely used. However, sidelobes of some magnitude are bound to exist for any spectrum estimate for which $w_r = 0$ for $|r| \geq m$ (that is, with a *truncation point* m), for $W_n(f)e^{2\pi i(m-1)f}$ is a polynomial of degree $2(m-1)$ in $e^{2\pi if}$ and hence may vanish for at most $2(m-1)$ frequencies. For a given m these frequencies must be placed so that $W_n(f)$ remains small for f not close to 0. It is desirable for $W_n(f)$ to be nonnegative, since otherwise negative values of the estimate $\hat{s}(f)$ may arise. This requires that the zeros of the polynomial occur in pairs.

Equation (8.11) suggests a different way of constructing spectrum estimates. If $W(f)$ is any suitable function, then

$$\hat{s}(f) = \int_0^1 W(f')I(f - f')df' \quad (8.13)$$

is a smoothed version of the periodogram. Furthermore, it may be written in the form (8.10) with

$$w_r = \int_0^1 W(f)e^{2\pi ifr}df \quad (8.14)$$

(see Exercise 8.5). This implies that $\hat{s}(f)$ also satisfies (8.11) for some spectral window $W_n(f)$, but this need not be the same function as $W(f)$. From a mathematical perspective, the lag weights w_r are the Fourier coefficients of $W(f)$, and therefore $W_n(f)$ is a partial sum of the Fourier series for $W(f)$. The spectrum estimate based on a series of length n depends on only the first n Fourier coefficients, whence different functions can yield identical spectrum estimates (for a given series length n).

Daniell (1946) suggested an estimate of the form (8.13), with

$$W(f) = \begin{cases} \frac{1}{2\delta}, & |f| < \delta, \\ 0, & |f| \geq \delta. \end{cases}$$

The resulting estimate $\hat{s}_D(f)$ is simply the integral average of the periodogram over an interval of length 2δ surrounding f . This is the integral analog of the simple moving average filters discussed in Section 7.2 and may, in the same way, be applied successively to build up more complex filters (see Section 8.7, p. 157).

A spectrum estimate defined as in (8.13) cannot be computed directly from the integral (8.13). It may be computed as a lag-weights estimate through (8.10); alternatively, the integral (8.13) could be approximated by a numerical quadrature formula. If the truncated Fourier series $W_n(f)$ is available, $\hat{s}(f)$ may also be calculated as

$$\hat{s}(f) = \frac{1}{n'} \sum_j W_n(f - f'_j) I(f'_j), \quad (8.15)$$

where $f'_j = n'/j$ are the Fourier frequencies associated with any series length $n' \geq 2n-1$. Equation 8.15 may be regarded as an exact quadrature formula for (8.11) or (8.13). The result of Exercise 8.7 shows that, when the estimate is in fact a lag-weights estimate with truncation point m , then $n' \geq n + m - 1$ is sufficient for (8.15) to hold.

Equation (8.15) shows that the estimate $\hat{s}(f)$ may be obtained from the periodogram, evaluated on a sufficiently fine grid, by a *linear filtering* operation (see Section 7.2). This is the simplest way to describe such an estimate and is also a useful computational route, for the fast Fourier transform will compute the required periodogram ordinates and the linear filter may be applied directly. If the filter has weights $\{\tilde{w}_j\}$, which in this context are called *spectral weights*, the formula is

$$\hat{s}(f) = \sum_j \tilde{w}_j I(f - f'_j). \quad (8.16)$$

This type of estimate is called a *discrete spectral average*. It may also be written in the lag-weights form (8.10) with $w_r = \tilde{W}(-f'_r)$, where $\tilde{W}(f)$ is the *transfer function* of the filter $\{\tilde{w}_j\}$,

$$\tilde{W}(f) = \sum_j \tilde{w}_j e^{-2\pi i f j}, \quad (8.17)$$

and in the integral form (8.13). However, (8.16) provides the simplest representation, and the spectral weights may be chosen directly to make the filtering operation have the required smoothing effect while remaining simple in nature. The choice of spectral weights is discussed in the next section.

Exercise 8.5 Alternative Representations

- (i) Verify (8.11). Note that $W_n(f)$ is a periodic function of f with period 1.
- (ii) Verify that the spectrum estimate $\hat{s}(f)$ defined by (8.13) satisfies (8.10) with lag weights given by (8.14).

Exercise 8.6 Modified Bartlett Window

Verify that

$$\sum_{|r| < m} (1 - |r|/m) e^{-2\pi i f r} = m D_m(f)^2.$$

(Hint: Exercise 2.2 shows that the Dirichlet kernel $D_m(f)$ satisfies

$$\sum_{t=0}^{n-1} e^{2\pi i f t} = e^{\pi i f(n-1)} m D_m(f).$$

Use $D_m(f)^2 = |D_m(f)|^2 = D_m(f)\overline{D_m(f)}$ and rearrange the latter product as a double sum, in the same way that the periodogram was rearranged in Section 8.3.)

Exercise 8.7 Discrete Spectral Averages

Suppose that $\hat{s}(f)$ is defined by (8.11), and that it has truncation point m (i.e., that $w_r = 0$ for $r \geq m$). Note that $m \leq n$. Show that (8.15) holds for any $n' \geq n + m - 1$. (Hint: Substitute the definition of $W_n(f)$ below (8.11) into (8.11) and use the orthogonality relations to show that the result simplifies to (8.10).)

Note that (8.15) does *not* hold in general if $n' < n + m - 1$.

Exercise 8.8 Spectrum Estimation by Complex Demodulation

Suppose that a series $\{x_0, x_1, \dots, x_{n-1}\}$ is demodulated at frequency f and then filtered using weights $\{g_u\}$ (see Sections 7.1 and 7.2). The instantaneous amplitude $\hat{R}_t(f)$ satisfies

$$\hat{R}_t(f)^2 = \left| \sum_u g_u y_{t-u} \right|^2,$$

where

$$y_t = e^{-2\pi ift} x_t.$$

- (i) Show that

$$\hat{s}(f) = \frac{\sum_t \hat{R}_t(f)^2}{n \sum_u g_u^2}$$

is approximately the lag-weights spectrum estimate

$$\sum_{|r| < n} w_r c_r e^{-2\pi i f r},$$

where

$$w_r = \frac{\sum_u g_u g_{u-r}}{\sum_u g_u^2}.$$

(Hint: The sum $\sum \hat{R}_t(f)^2$ may be written as a triple sum, in which every term is some multiple of a product of the form $x_t x_{t'} e^{-2\pi i f(t-t')}$. Collect terms with $t - t' = r$, and note that the multiplier in each term is $\sum_u g_u g_{u-r}$, except for a fixed number of terms at the ends of the sum.)

- (ii) Show that the corresponding spectral window is

$$W_n(f) = \frac{|G(f)|^2}{\sum_u g_u^2},$$

where $G(f)$ is the transfer function of the filter weights $\{g_u\}$.

Note that for any lag-weights spectrum estimate with truncation point m , if $W_n(f)$ is nonnegative then it may be factorized as $|G(f)|^2$, where $G(f)$ is the transfer function of a filter of span m . The class of spectrum estimates obtainable by complex demodulation is therefore essentially the same as the class of nonnegative lag-weights estimates. The connection between these estimates and the segment-average estimates was mentioned in Section 8.2. For more details see Bingham et al. (1967).

8.6 CHOICE OF A SPECTRAL WINDOW

Four factors need to be taken into account when choosing a spectral window:

- Resolution, or bandwidth;
- Stability;

- Leakage; and
- Smoothness.

Resolution is the ability of a spectrum estimate to represent fine structure in the frequency properties of the series, such as narrow peaks in the spectrum. Because of the averaging involved in computing a spectrum estimate, a spike or narrow peak in the periodogram is spread out into a broader peak. This peak is roughly an image of the spectral window of the estimate, and its “width,” suitably defined, is the *bandwidth* of the estimate (see Section 9.6). If the spectrum of a series contains two narrow peaks closer together than the bandwidth of the estimate used, the resulting broad peaks in the spectrum estimate overlap and form a single peak. In this case the estimate has failed to *resolve* the peaks.

The *stability* is the extent to which estimates computed from different segments of a series agree, or, in other words, the extent to which irrelevant fine structure in the periodogram is eliminated. Resolution and stability are conflicting requirements, because high stability requires averaging over many periodogram ordinates, while this increases bandwidth and consequently reduces resolution. Section 9.5 gives a statistical treatment of the stability of spectrum estimates.

Leakage has been discussed in the context of the Bartlett estimate in Section 8.5. It is caused by sidelobes in the spectral window, which are always present in a lag-weights estimate with a nontrivial truncation point $m < n$. However, the computationally simpler discrete spectral averages described in Section 8.5 can avoid leakage problems entirely. The first step is to use a data window to control leakage in the periodogram, and the second is to use a spectral window of a desired compact form.

The *smoothness* of a spectrum estimate is a less tangible property, but an important issue in graphical presentation and visual interpretation. The need for smoothness can introduce further conflict in the choice of a window. For instance, a simple definition of the bandwidth of an estimate is the span of frequencies over which the periodogram is averaged. Results of the next chapter show that the most stable estimate for a given bandwidth is the corresponding Daniell estimate. However, if the periodogram contains one or more large spikes, as is often the case, each of these is flattened out into a rectangle (or *boxcar*). The result can hardly be described as smooth, and the angular appearance of the spectrum estimate makes visual assessment difficult. Repeated smoothing using successive Daniell windows can yield a satisfactory estimate, as is shown in the next section, but necessarily gives a less stable estimate than is achievable for the resulting bandwidth.

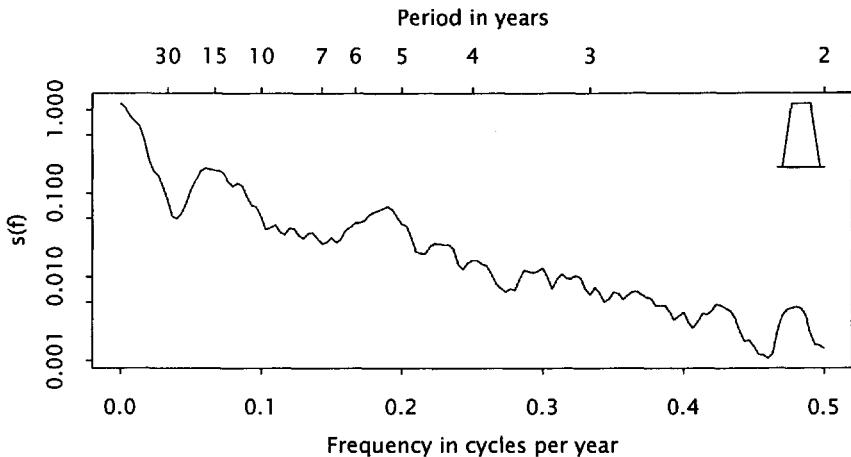


Fig. 8.12 Smoothed periodogram of logarithms of wheat price index, with spectral window inset. (Modified Daniell filter, $m = 6$.)

8.7 EXAMPLES OF SMOOTHING THE PERIODOGRAM

Wheat prices

Figure 8.12 shows the result of smoothing the periodogram of the logarithms of the wheat price series, plotted as usual on a logarithmic scale. The periodogram itself is shown in Figure 8.4 (p. 139), and its computation is described in Section 8.1. The frequencies at which it was calculated are its Fourier frequencies $f_j = j/300$, $j = 0, 1, \dots, 150$. The smoothing was carried out using a modified discrete Daniell procedure,

$$\hat{s}_6(f) = \frac{1}{6} \left\{ \frac{1}{2} I(f - f_3) + \sum_{j=-2}^2 I(f - f_j) + \frac{1}{2} I(f + f_3) \right\}, \quad (8.18)$$

which is the simple moving average of length 6 (see Section 7.3). The spectral weights are shown inset in Figure 8.12.

The graph in Figure 8.12 is still by no means smooth, and in a number of places an image of the spectral window is visible. This occurs where a relatively large periodogram ordinate has not been smoothed sufficiently. Figure 8.13 shows the result of applying a second modified Daniell filter to Figure 8.12, this time a simple moving average of length 12:

$$\hat{s}_{6,12}(f) = \frac{1}{12} \left\{ \frac{1}{2} \hat{s}_6(f - f_6) + \sum_{j=-5}^5 \hat{s}_6(f - f_j) + \frac{1}{2} \hat{s}_6(f + f_6) \right\}.$$

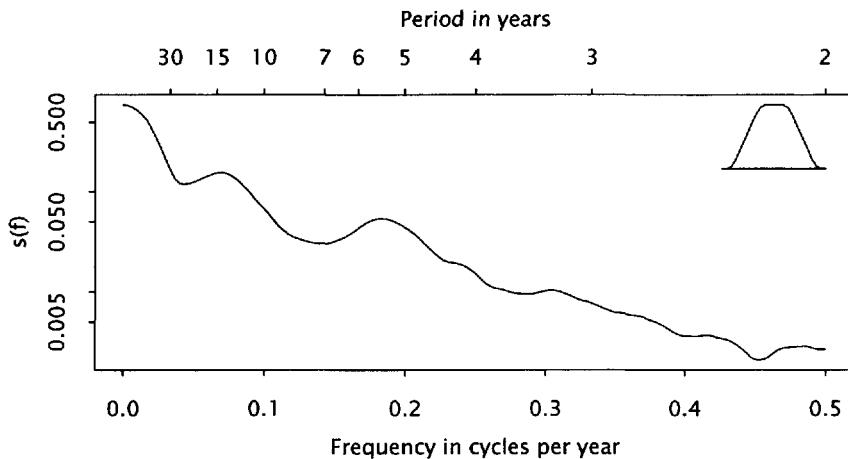


Fig. 8.13 Smoothed periodogram of logarithms of wheat price index, with spectral window inset. (Modified Daniell filter, $m = 6, 12$.)

As was shown in Section 7.2, the result is equivalent to applying a combined filter directly to the periodogram, and the corresponding spectral weights are shown inset. The many local fluctuations in Figure 8.12 have been smoothed out, but the broad features remain. The two peaks in the spectrum at around $f = 0.07$ and $f = 0.18$ cycles per year are quite clear. The only remaining structure in the spectrum appears to be a decrease in spectral power at higher frequencies.

Sunspots

Figure 8.14 shows estimates of the spectra of the yearly sunspot numbers (Figure 1.3, p. 4) and the square roots of these numbers (Figure 6.13, p. 84). They were obtained by smoothing the periodograms of the two series (Figures 6.10, p. 77 and 6.14, p. 85, respectively), using three applications of the modified Daniell procedure (8.18).²

The spectral weights are again shown inset. They have the same *span* as those in Figure 8.13 (i.e., they cover the same number of periodogram ordinates) but they are smoother and have a narrower peak. The spectrum estimates correspondingly show more rounded but slightly larger fluctuations. This argument is made more precise in Chapter 9.

²The square root series was rescaled to make its variance match that of the original sunspot numbers.

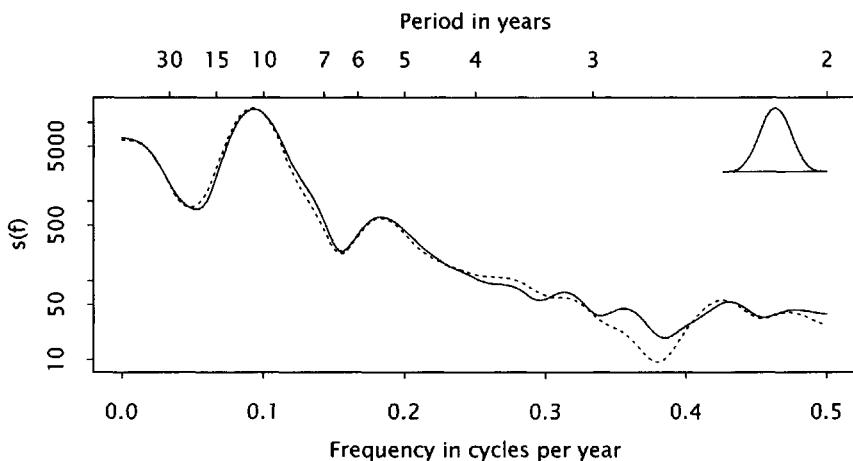


Fig. 8.14 Smoothed periodograms of yearly sunspot numbers (solid line) and their square roots (broken line), with spectral window inset. (Modified Daniell filter, $m = 6, 6, 6$.)

The presence of the second harmonic of the basic sunspot cycle is clear, although no higher harmonic is visible. Also visible is that the square root transformation makes the main peak slightly more prominent. Thus the transformation sought in Section 6.7 has the desired effect of making the main peak more distinct, but not to any great extent. Recall that the change is not visible in the periodograms themselves (Figures 6.10, p. 77 and 6.14, p. 85); one of the benefits of smoothing the periodogram is revealing such small effects.

Further discussion of the choice of a spectral window or lag weights for spectrum estimates may be found in Jenkins (1961) and Parzen (1961) (see also Section 9.6).

Exercise 8.9 Combination of Filters

Find the weights in the combined filters that were used to obtain the spectrum estimates in Figures 8.13 and 8.14. You should find that the first set is trapezoidal with some modifications in the angles, and that the second set is quadratic in three sections, again with some modifications at the boundaries. (*Hint:* The moving averages of lengths 6 and 12 can be written as moving averages of lengths 5 and 11, respectively, followed in each case by a moving average of length 2. Since filtering operations may be applied in any order, the moving averages of length 2 may be applied last. Combining the longer filters should give the piecewise linear and

piecewise quadratic shapes described.)

In this case the longer averages are themselves fairly short, and the "modifications" introduced by the averages of length 2 leave little of the exact behavior. The exercise may be repeated with longer windows to see the structure more clearly.

8.8 REROUGHING THE SPECTRUM

In the wheat price spectrum shown in Figure 8.13 (p. 158), the trough between the peak at zero frequency and the peak near $f = 0.07$ cycles per year is not as well defined as it is in Figure 8.12 (p. 157), because the bandwidth in the estimate shown in Figure 8.13 is large enough for the two peaks to begin to overlap. The corresponding slight loss of resolution suggests that the spectrum estimate may be somewhat oversmoothed. Similarly, in the sunspot spectra shown in Figure 8.14 (p. 159), the principal features are similar in width to the spectral weights, again raising the possibility of oversmoothing. As in filter design (Section 7.2), when a smoothing operation has possibly smoothed away part of the signal, some of the details may be restored by *reroughing* or *twicing* (Tukey, 1977).

In the context of linear filters, the *rough* is defined by

$$\text{rough} = \text{input} - \text{output},$$

and if the output has been oversmoothed, then part of the signal finds its way into the rough. Reroughing consists of smoothing the rough, and adding the smoothed rough back into the original output. Since periodograms and spectra are inherently nonnegative, the context of spectrum estimation calls for the rough to be defined by division instead of subtraction:

$$\text{rough} = \frac{\text{input}}{\text{output}},$$

and for the smoothed rough to be incorporated by multiplication instead of addition (Bloomfield, 1991). To be specific, if $I(f)$ is the periodogram and $\hat{s}(f)$ is the current spectrum estimate, the rough is

$$r(f) = \frac{I(f)}{\hat{s}(f)}.$$

If $\hat{s}(f)$ suffers from oversmoothing, there are narrow-band features in the periodogram that were not fully transferred to $\hat{s}(f)$, and these therefore appear partially in $r(f)$. They can be extracted with another round of smoothing, say as

$$\tilde{r}(f) = \sum_u g_u r(f - f_u),$$

and the reroughed spectrum estimate is

$$\hat{s}_r(f) = \tilde{r}(f)\hat{s}(f).$$

If the same filter is used in the second round as in the first, the process is called *twicing*.

Wheat Prices

The results of twicing the spectrum estimate of Figure 8.13 are shown in Figure 8.15. The upper panel shows the rough $r(f)$ and the smoothed rough $\tilde{r}(f)$, with the same modified Daniell smoothing as that used for Figure 8.13 (modified Daniell filters of lengths 6 and 12). The graph of $\tilde{r}(f)$ deviates from 1 principally in the dip at low frequencies, suggesting that the trough between the peak at zero frequency and the peak near $f = 0.07$ cycles per year should indeed be more sharply defined than in Figure 8.13. The lower panel shows the twiced spectrum estimate $\hat{s}_r(f)$ (solid line) and the original estimate $\hat{s}(f)$ (broken line). As expected, twicing has deepened the trough, adding emphasis to the separation of the two peaks.

Because of the way it is constructed, the reroughed spectrum estimate is not a discrete spectral average for any set of spectral weights. The spectral window shown inset in Figure 8.15 is the one that would result from *additive* reroughing of the spectrum estimate. The negative sidelobes confirm that additive reroughing could, undesirably, give negative estimates. This window is shown here to give a rough guide to the resolution of the estimate. Careful comparison shows that the main lobe of this window is slightly narrower than that shown in Figure 8.13, and this partly explains the deepening of the trough. When the center of the window is located in the trough, the negative sidelobes fall on the adjacent peaks, and also contribute to the deepening.

Sunspots

Figure 8.16 shows the result of twicing the sunspot spectrum estimate of Figure 8.14. Again, the main effect is the deepening of the troughs and a resulting enhancement of the peaks, consistent with the suggestion that the original spectrum estimate may be oversmoothed.

The corresponding window again has a somewhat narrower main lobe than that shown in Figure 8.14. In fact, the effect of twicing, and of reroughing in general, is to make the smooth somewhat *less* smooth, which will generally result in a narrower main lobe. The window has a

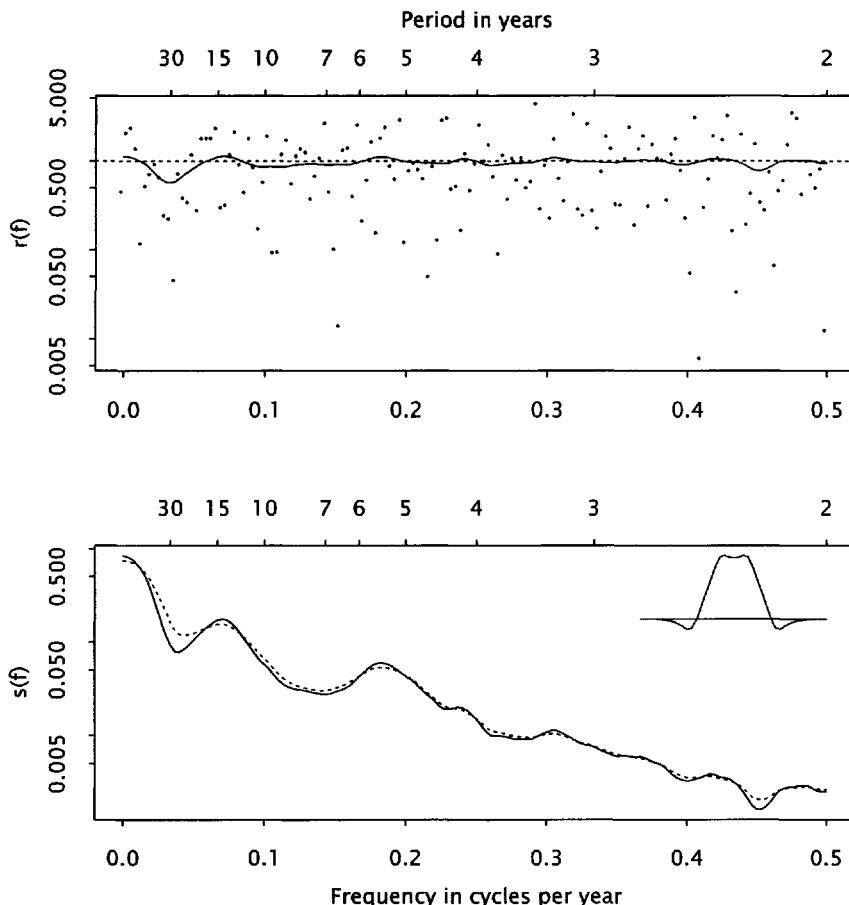


Fig. 8.15 Twicing the spectrum estimate of logarithms of wheat price index. Upper panel: rough $r(f)$ and smoothed rough $\tilde{r}(f)$. Lower panel: twiced estimate (solid line) and original estimate (broken line), with corresponding spectral window inset. (Modified Daniell filter, $m = 6, 12$.)

more satisfactory shape than that in Figure 8.15 as a result of the smoother profile of the original window (compare Figures 8.13 and 8.14). When modified Daniell filters are used, a minimum of three passes is required to achieve such a result.

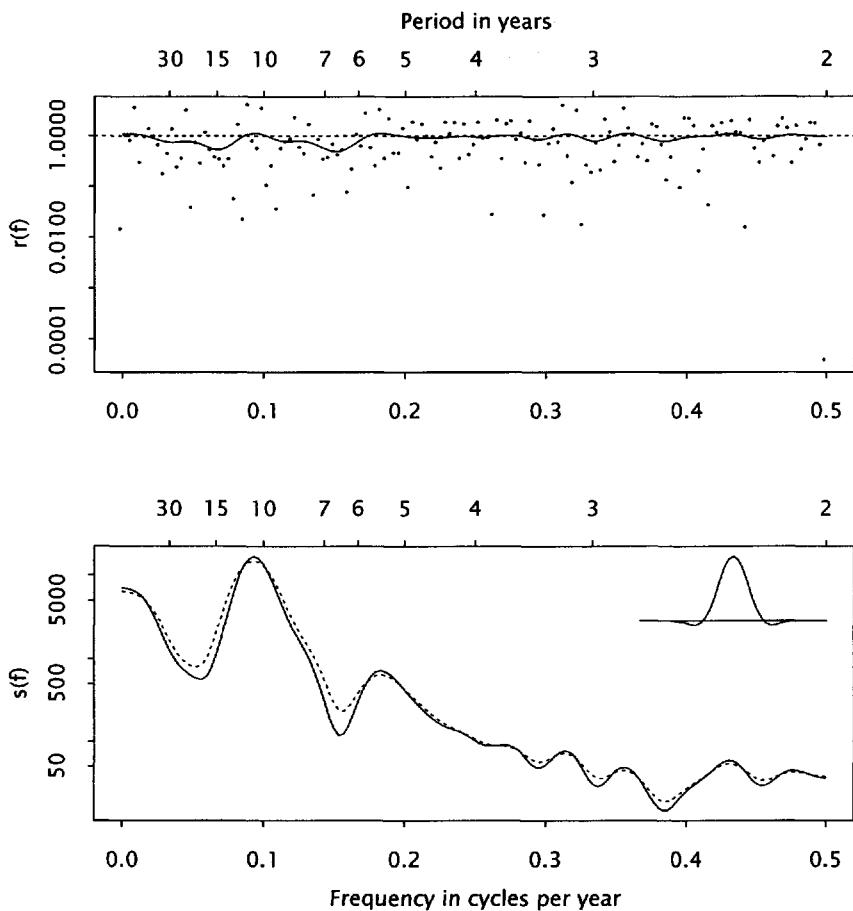


Fig. 8.16 Twicing the spectrum estimate of yearly sunspot numbers. Upper panel: rough $r(f)$ and smoothed rough $\tilde{r}(f)$. Lower panel: twiced estimate (solid line) and original estimate (broken line), with corresponding spectral window inset. (Modified Daniell filter, $m = 6, 6, 6$.)

Prewhitenning

Reroughing is closely related to *prewhitening* (Blackman and Tukey, 1959). Oversmoothing the periodogram leads to leakage from frequency bands with high power to adjacent bands with lower power. From this perspective, the oversmoothing problem is caused by a large *dynamic range* in the spectrum. Prewhitenning is a technique for reducing dynamic range prior to forming the periodogram and estimating the spectrum. It reduces the

problem of leakage, and allows the use of a more stable estimate with lower resolution.

The simplest form of prewhitening is replacing the data $\{x_t\}$ by their first differences $\{y_t = x_t - x_{t-1}\}$. This is a linear filtering operation with transfer function $G(f) = 1 - e^{2\pi if}$ and power transfer function $|G(f)|^2 = |1 - e^{2\pi if}|^2 = 4(\sin \pi f)^2$. Differencing often reduces dynamic range in the common situation where the highest power occurs at low frequencies, and other high-pass filters have a similar effect. If the high power occurs in some other frequency band, a filter may be designed to reduce the imbalance.

The periodogram of the prewhitened (i.e., filtered) series is approximately the same as the periodogram of the original series multiplied by $|G(f)|^2$, the power transfer function of the prewhitening filter. This distorts the underlying spectrum, but the distortion may be reversed by dividing the resulting spectrum estimate by the same function. The parallel with reroughing is clear, with *multiplication* by the power transfer function playing the role of *division* by an initial spectrum estimate. However, the parallel is not exact, precisely because the periodogram of the prewhitened series is only *approximately* the same as the periodogram of the original series multiplied by the power transfer function. Reroughing may be thought of as an enhancement to *spectrum estimation*, while prewhitening is a form of *preprocessing*.

Appendix

The Bartlett estimates shown in this chapter were computed using S-PLUS commands similar to the following:

```
n <- length(x)
y <- c(taper(n, p) * x, rep(0, m))
acov <- fft(Mod(fft(y))^2, inv = T)/n^2
acov <- acov * bartweights(0:(length(acov) - 1), m, n)
acov[-1] <- acov[-1] + rev(acov[-1])
spec <- fft(acov)
```

Here x is the time series, prepared for Fourier transformation by centering or removing a trend line, p is the fraction of the series length to be tapered (e.g., $p = 0.2$ for 10% tapering at each end of the series), and m is the parameter of the Bartlett estimate. The function *bartweights* provides the required lag weights:

```
bartweights <- function(r, m, n)
```

```

{
  w <- pmax(0, 1 - abs(r)/m)
  w[w > 0] <- w[w > 0]/(1 - abs(r)/n)[w > 0]
  w
}

```

The modified Daniell spectrum estimates were obtained using the S-PLUS function `spec.pgram()`, which may also be used to calculate the periodogram. In addition to the time series, `spec.pgram()` has arguments that control

- the number of passes and the spans,
- tapering,
- padding with zeros,
- adjusting the series by centering at the mean or subtracting the least squares straight line, and
- plotting.

The spans may be specified as even integers, as above, or as odd integers, the actual spans of the filters. For instance, the component described above as a 6-term moving average actually has a span of 7 because of the modification at the ends of the window, and may be specified as either `span = 6` or `span = 7`. The argument controlling tapering specifies the fraction of data to be tapered *at each end* of the series. Similarly, the argument controlling padding specifies the number of zeros to be appended to the data as a fraction of the series length, with a default of zero. A small number of additional zeros may be appended if the length of the padded series has large prime factors, which would otherwise make the fast Fourier transform relatively inefficient.

The function returns a list with many components, including the spectrum estimate (in decibels, $10 \log_{10} \hat{s}(f)$) and the weights of the composite filter. The figures in this chapter were made by graphing the actual spectrum estimate on a logarithmic scale, rather than by using the `plot` option of `spec.pgram()`.

The spectrum estimates were reroughed (to be specific, twiced), using code similar to the following:

```

a <- spec.pgram(x, spans = c(6, 12))
a$spec <- 10^(a$spec/10)
b <- spec.pgram(x)
b$spec <- 10^(b$spec/10)

```

```
r <- b$spec/a$spec
lenpad <- round(1/diff(a$freq[1:2]))
r <- r[pmin(1:lenpad, lenpad + 2 - 1:lenpad)]
rtilde <- filter(r, filter = a$filter, circular = T)
twice <- rtilde[1:length(a$spec)] * a$spec
```

The default options of tapering 10% of each end of the series and detrending the series were adopted. In the second use of `spec.pgram()` the `span` defaults to 1, and the result `b$spec` is thus the unsmoothed periodogram. The function `filter()` was used carry out the smoothing of the rough, `r`, using the filter weights `a$filter` in a circular convolution. Since `spec.pgram()` returns the spectrum only for $0 \leq f \leq 1/2$, `r` must first be extended to frequencies $0 \leq f < 1$, in the steps involving the length of the padded series `lenpad`.

9

Some Stationary Time Series Theory

The *spectrum* of a time series was introduced in Chapter 8, as the aspects of its periodogram that show statistical regularity and are characteristic of the series as a whole. The concept is statistical in nature, and this chapter discusses a class of theoretical models that show such statistical regularity, namely, the class of *stationary time series*. A *theoretical spectrum* may be defined for a stationary time series, and the spectrum estimates described in Chapter 8 may be thought of as providing estimates of it. Some properties of the sampling distributions of spectrum estimates may be found and, in particular, approximate confidence intervals for the theoretical spectrum may be constructed.

This chapter gives a brief nonrigorous description of a small part of stationary time series theory. More extensive and more rigorous mathematical discussion may be found in Anderson (1994), Brillinger (1981), Hannan (1970), Koopmans (1995), or Priestley (1981a).

9.1 STATIONARY TIME SERIES

A (weakly) stationary time series is a collection of random variables $\{X_t\}$ defined for all real t or all integers t , as the case may be, with the following properties:

- (i) $E(X_t)$ is constant, and

(ii) $E(X_t X_{t'})$ depends on t and t' only through $t - t'$.

Since $E(X_t) = \mu$ is constant, it may be estimated and subtracted from the series with little effect on what follows. Henceforth, $E(X_t)$ will be assumed to be zero, and modifications that must be made when the series mean is subtracted from the data will be described where necessary. Property (ii) implies that $\text{var}(X_t)$ is constant and that

$$\gamma_r = \text{cov}(X_t, X_{t+r}) = \gamma_{-r}$$

does not depend on t . The quantity γ_r is the *theoretical autocovariance* at lag r of $\{X_t\}$ and may be estimated by

$$c_r^* = \frac{1}{n} \sum_{t=|r|}^{n-1} (X_t - \mu)(X_{t-|r|} - \mu), \quad (9.1)$$

which for $r = 0$ reduces to $(1/n) \sum (X_t - \mu)^2$. If μ is replaced by its estimate \bar{X} in (9.1), the sample autocovariance c_r used in Chapter 8 emerges. The expected value of c_r^* is $(1 - |r|/n)\gamma_r$, whence it is a biased estimate of γ_r , although the bias is small for large n (see Exercise 9.1).

The sequence $\{\gamma_r\}$ is *nonnegative definite*, meaning that for any $k > 0$ and constants a_1, a_2, \dots, a_k

$$\sum_{r,s} a_r \gamma_{r-s} a_s \geq 0,$$

since the double sum is the variance of a linear combination of random variables in the series and hence is certainly nonnegative (see Exercise 9.2). A theorem due to Herglotz states that for any such sequence there exists a nondecreasing function $S(f)$ such that

$$\gamma_r = \int_0^1 e^{2\pi i f r} dS(f)$$

(for a proof, see Doob, 1953, p. 474). If $S(f)$ has a derivative $s(f)$, then

$$\gamma_r = \int_0^1 e^{2\pi i f r} s(f) df. \quad (9.2)$$

The functions $S(f)$ and $s(f)$ are the *spectral distribution function* and *spectral density function*, respectively. The spectral density function is the theoretical counterpart of the (empirical) spectrum defined in Chapter 8. Equation (9.2) shows that the theoretical autocovariances $\{\gamma_r\}$ are the Fourier coefficients of the spectral density function $s(f)$. Hence under

mild conditions on $s(f)$, the latter may be represented as the Fourier series

$$\begin{aligned}s(f) &= \sum_{r=-\infty}^{\infty} y_r e^{-2\pi i f r} \\&= y_0 + 2 \sum_{r=1}^{\infty} y_r \cos 2\pi f r.\end{aligned}\tag{9.3}$$

The equation

$$\text{var}(X_t) = y_0 = S(1) - S(0) = \int_0^1 s(f) df$$

shows that the variance of the series may be regarded as the sum of components associated with each frequency in the interval $[0, 1]$, and also that $s(f)$ has a finite integral.

Since the autocovariances $\{y_r\}$ are real, it follows that the spectral density function is symmetric about zero, and the spectral distribution function is also symmetric in the sense that

$$S(f) - S(f') = -\{S(-f) - S(-f')\}.$$

The Fourier series (9.3) represents a periodic function with period 1, and it is natural to view the spectral density function as periodic. Any integral with limits 0 and 1 may then also be written with limits $-1/2$ and $1/2$. Because of the symmetry of $s(f)$, the autocovariance y_r may also be written

$$\begin{aligned}y_r &= \int_0^1 e^{2\pi i f r} s(f) df \\&= \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i f r} s(f) df \\&= \int_0^{\frac{1}{2}} \{e^{-2\pi i f r} s(-f) + e^{2\pi i f r} s(f)\} df \\&= \int_0^{\frac{1}{2}} \{e^{-2\pi i f r} + e^{2\pi i f r}\} s(f) df \\&= \int_0^{\frac{1}{2}} 2 \cos(2\pi f r) s(f) df,\end{aligned}$$

where the integrand is now entirely real.

The following sections give some simple but central examples of stationary time series and their spectra.

Sinusoid

If

$$X_t = A \cos 2\pi f t + B \sin 2\pi f t, \quad -\infty < t < \infty,$$

where

$$E(A) = E(B) = E(AB) = 0$$

and

$$E(A^2) = E(B^2) = \frac{a^2}{2},$$

say, then $\{X_t\}$ is weakly stationary with

$$E(X_t) = 0$$

and

$$y_r = \frac{a^2}{2} \cos 2\pi f r.$$

Thus $S(f)$ is a step function with jumps of $a^2/4$ at $-f$ and f (see Exercise 9.4). The sum of several series of the same kind, with various frequencies and amplitudes, is itself stationary if all the coefficients are uncorrelated (see Exercise 9.5). The spectral distribution function then has a jump at each of the frequencies involved.

Notice that Fourier analysis of a set of observed values from such a series would provide estimates of the coefficients A and B , and hence of the squared amplitude $A^2 + B^2$. However, the parameter a^2 is the *expected value* of this quantity, and even a long record yields only a single observed value of each of A and B . Thus even a long record does not provide a good estimate of $S(f)$, unless $A^2 + B^2$ is nonrandom (has a degenerate distribution). It is possible for $A^2 + B^2$ to be nonrandom, for instance if

$$A = a \cos 2\pi \Phi$$

and

$$B = a \sin 2\pi \Phi,$$

where Φ is uniformly distributed on $[0, 1]$. The series may then be written

$$X_t = a \cos 2\pi(f t - \Phi),$$

a *random phase* model. Since $S(f)$ is not differentiable, there is no spectral density function except in the sense of generalized functions (δ -functions).

White Noise

If

$$\begin{aligned} E(X_t) &= 0, \\ E(X_t^2) &= \sigma^2, \quad \text{and} \\ X(X_t X_{t'}) &= 0 \quad \text{for } t \neq t', \end{aligned}$$

then $\{X_t\}$ is stationary, and

$$\gamma_r = \begin{cases} \sigma^2 & r = 0, \\ 0 & r \neq 0. \end{cases}$$

Therefore,

$$S(f) = \sigma^2 f.$$

(see Exercise 9.6). In this case, $S(f)$ has no discontinuities, and the series contains no pure sinusoids as in the previous example. The spectral density function is $s(f) = \sigma^2$, a constant.

Output of a Filter

Suppose that $\{X_t\}$ is stationary with mean zero and spectral distribution function $S_X(f)$, and that

$$Y_t = \sum_u g_u X_{t-u}, \quad -\infty < t < \infty.$$

Then

$$E(Y_t) = 0,$$

and

$$\begin{aligned} E(Y_t Y_{t+r}) &= \sum_{u,u'} g_u g_{u'} E(X_{t-u} X_{t+r-u'}) \\ &= \sum_{r,s} \int_0^1 e^{2\pi i f(r+u-u')} dS_X(f) \\ &= \int_0^1 |G(f)|^2 e^{2\pi i f r} dS_X(f), \end{aligned}$$

where

$$G(f) = \sum g_u e^{-2\pi i f u}$$

is the *transfer function* of the filter (see Section 7.2). Thus $\{Y_t\}$ is also stationary, and its spectral distribution function is

$$S_Y(f) = \int_0^f |G(f')|^2 dS_X(f').$$

The function $|G(f)|^2$ is called the *power transfer function* of the filter.

The relationship between the spectral density functions is simpler:

$$s_Y(f) = |G(f)|^2 s_X(f).$$

Exercise 9.1 Mean of the Sample Autocovariance

If c_r^* is defined as in (9.1), verify that

$$E(c_r^*) = \left(1 - \frac{|r|}{n}\right) \gamma_r.$$

Note that $c_r^*/(1 - |r|/n)$ provides an unbiased estimate of γ_r . However, in the context of spectrum estimation the sample autocovariances are heavily damped at large lags r , whence this modification would have little effect. The lag weights of the Bartlett spectrum estimate (8.5) (p. 145) contain such a divisor, derived from heuristic reasoning, while the modified Bartlett lag weights (8.12) (p. 150) do not. Compare the spectrum estimates for the wheat price series, shown in Figure 8.8 (p. 146) and Figure 8.10 (p. 151), respectively.

Exercise 9.2 Autocovariances Are Nonnegative Definite

Find the variance of $\sum_r a_r X_{t-r}$, and hence deduce that $\{\gamma_r\}$ is nonnegative definite.

Exercise 9.3 (Continuation) Sample Autocovariances

Show that the sample autocovariance sequence $\{c_r\}$, where c_r is defined to be zero for lags greater than or equal to the series length, is nonnegative definite. (*Hint:* Use the integral inversion formula

$$c_r = \int_0^1 I(f) e^{2\pi i f r} df, \quad -\infty < r < \infty,$$

where $I(f)$ is the periodogram, and the property $I(f) \geq 0.$)

Exercise 9.4 Random Sinusoid

Verify that the sinusoidal series $\{X_t\}$ defined in Section 9.1 is weakly stationary.

Exercise 9.5 Superposition of Stationary Series

Suppose that $\{X_t\}$ and $\{Y_t\}$ are weakly stationary, and that

$$E(X_t Y_{t'}) = E(X_t)E(Y_{t'}), \quad \text{all } t, t'.$$

Show that $\{Z_t = X_t + Y_t\}$ is also weakly stationary.

Exercise 9.6 White Noise

Show that the spectral distribution function $S(f) = \sigma^2 f$ corresponds to the autocovariance sequence

$$\gamma_r = \begin{cases} \sigma^2 & r = 0 \\ 0 & r \neq 0 \end{cases}$$

and hence to a series of uncorrelated random variables.

9.2 CONTINUOUS SPECTRA

The examples in the previous section showed that if a time series contains a periodic component then its spectral distribution function has a discontinuity at the corresponding frequency, and hence the spectral density function does not exist. When the spectral distribution function has no discontinuities and the spectral density function exists, the series is said to have a *continuous spectrum*. The autocovariances $\{\gamma_r\}$ of such a series are given by (9.2). It was noted above that $s(f)$ has the properties:

- Nonnegativity: $s(f) \geq 0$
- Symmetry: $s(-f) = s(f)$
- Periodicity: $s(f + 1) = s(f)$
- Integrability: $\int_0^1 s(f) df < \infty.$

In fact, any function $s(f)$ with these properties may arise as a spectral density function. Write $G_1(f) = \sqrt{s(f)}$. Then there are filters with transfer functions arbitrarily close to $G_1(f)$. For instance, suppose that $s(f)$ is continuous. Then so is $Q_1(x)$, where $Q_1(\cos 2\pi f) = G_1(f)$. But in that case there is a polynomial that is arbitrarily close to $Q_1(x)$, and hence a polynomial in $\cos 2\pi f$ that is arbitrarily close to $G_1(f)$. But it was shown in Section 7.2 that any cosine polynomial is the transfer function of some filter. Taking limits along a sequence of such filters leads to a limit filter with transfer function $G_1(f)$. The limiting filter may involve an infinite number of data points and may, therefore, be somewhat hypothetical in nature.

Now suppose that $\{U_t\}$ is a white noise process with variance 1. If the weights of the limiting filter are $\{g_r\}$, let

$$X_t = \sum_r g_r U_{t-r}. \quad (9.4)$$

From Section 9.1, the spectral density function of $\{U_t\}$ is the constant 1. From results in the same section, therefore, the spectral density function of $\{X_t\}$ is $|G_1(f)|^2 = s(f)$. A process of the form (9.4), that is, the output of a filter to which the input is white noise, is called a *moving average process*. It follows from the above argument that *for any nonnegative function $s(f)$ with a finite integral there exists a moving average process whose spectral density function is $s(f)$* .

There is another result which states that *every time series with a continuous spectrum is a moving average process* (Doob, 1953, p. 498). The argument is along the following lines: Suppose that the series is $\{X_t\}$ and that its spectral density function is $s(f)$. It is convenient, although not necessary, to assume that $s(f)$ is continuous and strictly positive. Then $G_2(f) = 1/\sqrt{s(f)}$ is bounded and continuous, and as before a limiting filter may be found with transfer function $G_2(f)$. Suppose that $\{U_t\}$ is the result of applying this filter to $\{X_t\}$. Then the spectral density function of $\{U_t\}$ is

$$\begin{aligned} s_U(f) &= |G_2(f)|^2 s(f) \\ &= 1, \quad 0 \leq f < 1, \end{aligned}$$

and hence $\{U_t\}$ is white noise. Now suppose that the filter constructed above, with transfer function $G_1(f) = \sqrt{s(f)}$, is applied to $\{U_t\}$. The result is the same as applying to $\{X_t\}$ a combined filter with transfer function $G_1(f)G_2(f) = 1$, $0 \leq f < 1$. In other words, the combined filter is the identity filter, which leaves $\{X_t\}$ unchanged. That is, the white noise series $\{U_t\}$ has been constructed so that (9.4) holds, thus proving the result.

It must be remembered that the white noise process has been defined only in terms of its first and second moments, and that lack of correlation does not imply independence. To see how far from independent the terms in a white noise process may be, consider the following example: Suppose that Y is drawn from a symmetric distribution F on $[0, 1]$, and that Φ is drawn from the uniform distribution on the same interval, independently of Y . Let

$$X_t = a \cos 2\pi(Yt + \Phi).$$

Conditionally on $Y = f$, this becomes the random phase model (see Section 9.1), with mean zero and autocovariances

$$E(X_t X_{t+r} | Y = f) = \frac{a^2}{2} \cos 2\pi f r.$$

The unconditional moments are thus

$$\begin{aligned} E(X_t X_{t+r}) &= \frac{a^2}{2} \int_0^1 \cos(2\pi f r) dF(f) \\ &= \frac{a^2}{2} \int_0^1 e^{2\pi i f r} dF(f), \end{aligned}$$

and hence the spectral distribution function is $a^2 F(f)/2$. In particular, if F is the uniform distribution on $[0, 1]$, $\{X_t\}$ is a white noise process. However, any *realization* of the series is a pure sinusoid!

9.3 TIME AVERAGING AND ENSEMBLE AVERAGING

The spectrum of an observed series was defined as the aspects of its periodogram that show statistical regularity from one segment of the series to another. The first spectrum estimate described in Chapter 8 was derived by averaging periodograms of a number of segments. Exercise 8.8 (p. 154) shows that most spectrum estimates may similarly be regarded as the average over time of a local measure of spectral power.

However, the probabilistic model for a stationary time series described in Section 9.1 is one in which an *ensemble* of possible *realizations* of the series is envisaged, and the expected value operator $E(\dots)$ refers to averaging across the ensemble, rather than along time. A series is said to be *ergodic* if the time average of a quantity is (with probability 1) equal to its ensemble average (see Brillinger, 1981, Section 2.11). Statistical inferences about the structure of such a series, such as estimation of its spectrum, should therefore be based on the probabilistic properties of *ergodic* series.

The random phase model (Section 9.1) and the related white noise example constructed in Section 9.2 show that nonergodic series are easy to construct. In the random phase model, the periodogram $I(f; t)$ of the stretch of data $\{X_t, X_{t+1}, \dots, X_{t+n-1}\}$ is always equal to $(A^2 + B^2)/a^2$ times its expected value, and hence so is the time average of $I(f; t)$. Thus the series can be ergodic only if $A^2 + B^2$ has a degenerate distribution (i.e., if $A^2 + B^2 = a^2$ with probability 1). Similarly, any series with discontinuities in its spectral distribution function can be ergodic only if all the amplitudes are (with probability 1) constants. The white noise series obtained in Section 9.2 by randomizing the frequency illustrates the problem more clearly. From any one realization of the series, its spectrum would be inferred to be purely discontinuous, whereas it is in fact continuous. However, a white noise series in which the observations are *independent* and *identically distributed* is ergodic, and so too is any moving average series constructed from it (such a moving average process is called a *linear process*). In the remainder of this chapter, we examine the statistical properties of the periodogram of a moving average series, and of spectrum estimates constructed from it.

9.4 PERIODOGRAM OF A TIME SERIES WITH A CONTINUOUS SPECTRUM

It was shown in Chapter 6 that, if $\{U_t\}$ is a Gaussian white noise series with variance σ^2 , the periodogram ordinates at the Fourier frequencies (other than 0 and $1/2$ cycles per unit time) are independently exponentially distributed, with expected value σ^2 . At $f = 0$, and at $f = 1/2$ if it is a Fourier frequency, the expected value is the same, but the distribution is χ^2 with 1 degree of freedom instead of 2. If the mean is subtracted from the series, the periodogram vanishes identically at zero frequency. If the $\{U_t\}$ are independent with a finite variance σ^2 but a non-Gaussian distribution, the central limit theorem assures that for large n the joint distribution of any finite number of periodogram ordinates approaches the same distribution. Central limit theorems for dependent processes (see, e.g., Ibragimov and Linnik, 1971) show that the same result holds for dependent white noise processes, provided the dependence is limited in time. The rather extreme example of Section 9.2 shows that the result cannot be true for all white noise processes.

Now suppose that $\{X_t\}$ has a continuous spectrum, with spectral density function $s(f)$. Suppose further that

$$X_t = \sum_r g_r U_{t-r},$$

where $\{U_t\}$ is a white noise process with variance 1, either

- Gaussian, and hence independent, or
- Non-Gaussian but independent, or
- Non-Gaussian and dependent, but close enough to independence for the limiting distribution of the periodogram is exponential.

Then the transfer function $G(f)$ of the filter weights $\{g_r\}$ satisfies $|G(f)|^2 = s(f)$, and the argument of Section 7.2 shows that the Fourier transform of a stretch $\{X_0, X_1, \dots, X_{n-1}\}$ satisfies

$$d_X(f) \approx G(f)d_U(f).$$

Hence the periodogram satisfies

$$\begin{aligned} I_X(f) &\approx |G(f)|^2 I_U(f) \\ &= s(f) I_U(f). \end{aligned} \tag{9.5}$$

It is easily shown that $E\{I_U(f)\} = 1$, and hence that $E\{I_X(f)\} \approx s(f)$. Thus $I_X(f)$ is an approximately unbiased estimate of $s(f)$. However, since its distribution is χ^2 with 2 degrees of freedom, it is a poor estimate and does not improve with increasing series length.

The approximation in (9.5) is caused by end effects of the filter and improves with increasing series length n . Thus for large n the periodogram of $\{X_t\}$ consists of the spectral density function of $\{X_t\}$ multiplied by the periodogram of a white noise process. This accounts for the spikiness of the periodograms shown in earlier chapters. Olshen (1967) gives a rigorous study of the asymptotic behavior of periodograms.

9.5 APPROXIMATE MEAN AND VARIANCE OF SPECTRUM ESTIMATES

The background material is now in place to derive approximations to the statistical behavior of the spectrum estimates introduced in Chapter 8 (see also Priestley, 1981a, Chapter 6).

Discrete Averages

The simplest spectrum estimates are discrete averages of periodogram ordinates evaluated at the Fourier frequencies,

$$\hat{s}(f) = \sum_u g_u I(f - f_u). \tag{9.6}$$

By the arguments of Section 9.4, such an estimate is a sum of approximately independent and exponentially distributed terms, and thus its approximate distribution may be derived analytically. The following cruder approximation is usually satisfactory.

It was shown in Section 9.4 that

$$E\{I(f)\} \approx s(f).$$

If the averaging in (9.6) is assumed to cover a small interval of frequencies and that $s(f)$ is continuous in that interval, then $s(f)$ is approximately constant over the same interval. Hence

$$E\{I(f - f_u)\} \approx s(f)$$

and

$$E\{\hat{s}(f)\} \approx s(f) \sum_u g_u.$$

Thus $\hat{s}(f)$ is an approximately unbiased estimate of $s(f)$, provided that $\sum g_u = 1$. In Section 9.6, a more refined argument yields an approximation to the bias (see also Parzen, 1957b).

The variance of an exponentially distributed random variable is the square of its expected value, and hence

$$\text{var}\{I(f)\} \approx s(f)^2.$$

It was shown in Section 9.4 that periodogram ordinates at different Fourier frequencies are approximately independent, and hence if $f \neq 0$ or $1/2$,

$$\text{var}\{\hat{s}(f)\} \approx \sum_u g_u^2 s(f - f_u)^2.$$

If $s(f)$ is assumed as before to be approximately constant over the interval of averaging, then

$$\text{var}\{\hat{s}(f)\} \approx s(f)^2 \sum_u g_u^2. \quad (9.7)$$

Equation (9.7) shows that the coefficient of variation of $\hat{s}(f)$ does not depend on f , unless the spectral weights $\{g_u\}$ themselves are frequency-dependent. In the examples discussed in Chapter 8 and later, the spectral weights are always chosen to be the same at all frequencies. However, there is still an *implicit* frequency-dependence at frequencies close to 0 and $1/2$ cycles per unit time. Near these frequencies, the convolution defining $\hat{s}(f)$ extends to frequencies outside the principal interval $[0, 1/2]$. These are aliased with frequencies *inside* this interval, and

the spectral weights effectively accumulate on the corresponding periodogram ordinates. This phenomenon leads to a progressive increase in sampling variability as f approaches 0 or $1/2$ cycles per unit time, doubling the variance at these limiting frequencies. For instance, confidence intervals are widened by a factor that increases to $\sqrt{2}$.

Lag Weights Estimates

A lag-weights estimate may be written in the integral form

$$\hat{s}(f) = \int_0^1 W_n(f - f') I(f') df', \quad (9.8)$$

where

$$W_n(f) = \sum_{|r| < m} w_r e^{-2\pi i f r}$$

is the corresponding spectral window. The integral (9.8) may be approximated by the sum

$$\hat{s}_d(f) = \frac{1}{n} \sum_u W_n(f - f_u) I(f_u), \quad (9.9)$$

which is of the form (9.6), and hence

$$\begin{aligned} \text{var}\{\hat{s}(f)\} &\approx \text{var}\{\hat{s}_d(f)\} \\ &\approx \left(\frac{1}{n}\right)^2 s(f)^2 \sum_u W_n(f - f_u)^2 \\ &\approx \frac{1}{n} s(f)^2 \int_0^1 W_n(f')^2 df'. \end{aligned}$$

Now

$$\int_0^1 W_n(f')^2 df' = \sum_{|r| < m} w_r^2,$$

and thus an alternative form is

$$\text{var}\{\hat{s}(f)\} \approx \frac{1}{n} s(f)^2 \sum_{|r| < m} w_r^2.$$

The lag weights are usually given by a formula such as $w_r = w(r/m)$, where $w(x) = 0$ for $|x| \geq 1$. Then

$$\begin{aligned}\sum_{|r| < m} w_r^2 &= \sum_{|r| < m} w\left(\frac{r}{m}\right)^2 \\ &\approx m \int_{-1}^1 w(x)^2 dx.\end{aligned}$$

The function $w(x)$ is the *lag window* of the estimate. This gives another approximation,

$$\text{var}\{\hat{s}(f)\} \approx \frac{m}{n} s(f)^2 \int_{-1}^1 w(x)^2 dx. \quad (9.10)$$

This shows that for a given lag window the variance of the spectrum estimate is determined by the ratio m/n . This ratio is just the proportion of sample autocovariances for which the lag weights do not vanish.

Discrete Averages on a Refined Grid

Lastly consider the discrete spectral average computed from the periodogram ordinates at a set of frequencies $f'_j = j/n'$ more finely spaced than the Fourier frequencies,

$$\hat{s}(f) = \sum_u g_u I(f - f'_u).$$

If the spectral weights $\{g_u\}$ approximate a smooth window $W(f)$, in the sense that $g_u = (1/n')W(f'_u)$, then

$$\hat{s}(f) \approx \int W(f') I(f - f') df'.$$

This approximation may be justified by an argument similar to that used in Exercise 9.8. As for a lag weights estimate, it follows that

$$\text{var}\{\hat{s}(f)\} \approx \frac{1}{n} s(f)^2 \int_0^1 W(f')^2 df'.$$

Now

$$\begin{aligned}\sum_u g_u^2 &= \left(\frac{1}{n'}\right)^2 \sum_u W(f'_u)^2 \\ &\approx \frac{1}{n'} \int_0^1 W(f')^2 df',\end{aligned}$$

and thus

$$\text{var}\{\hat{s}(f)\} \approx \frac{n'}{n} s(f)^2 \sum_u g_u^2.$$

The factor n'/n may be regarded as a correction for the finer spacing of the frequencies.

Often the smooth window $W(f)$ is given by

$$W(f) = mV(mf),$$

where $V(x) = 0$ for $|x| \geq 1$, and

$$\int_{-1}^1 V(x)dx = 1.$$

The function $V(x)$ is referred to as a standardized spectral window. By varying m a family of windows is obtained with different widths. Now

$$\begin{aligned} \int_{-1}^1 W(f)^2 df &= m^2 \int_{-1}^1 V(mf)^2 df \\ &= m \int_{-1}^1 V(x)^2 dx, \end{aligned}$$

and hence

$$\text{var}\{\hat{s}(f)\} = \frac{m}{n} s(f)^2 \int_{-1}^1 V(x)^2 dx.$$

Comparison with (9.10) shows that the parameter m has the same effect as the truncation point of a lag weights estimate. Note that the span of the resulting filter $\{g_u\}$ is roughly n'/m , and hence its bandwidth (see Section 8.6) is at most $1/m$.

Tapering

The effect of tapering a set of data before computing the periodogram from which a spectrum estimate is constructed also needs to be taken into account. If the data window is $\{u_t : 0 \leq t < n\}$, then

$$E\{I(f)\} \approx U_2 s(f) \tag{9.11}$$

(see Exercise 9.9), where

$$U_2 = \frac{1}{n} \sum_t u_t^2.$$

Thus for the periodogram to continue to provide an approximately unbiased estimate of the spectral density function $s(f)$, the data window should be normalized so that $U_2 \approx 1$.¹ With this normalization the effect on the variance of any of the spectrum estimates is to multiply it by a factor that is approximately

$$U_4 = \frac{1}{n} \sum_t u_t^4$$

(see, e.g., Hannan, 1970, Section V.4).

In terms of the logarithm of a spectrum estimate, the possible biasing factor U_2 becomes an added constant and is relatively unimportant in many situations. The variance of the logarithm is multiplied by U_4/U_2^2 , and this, of course, does not depend on whether or not the added bias is removed.

For the split cosine bell taper described in Section 6.2,

$$u_t = u\left(\frac{2t+1}{2n}\right),$$

where

$$u(x) = \begin{cases} \frac{1}{2}\{1 - \cos(2\pi x/p)\} & 0 \leq x \leq p/2, \\ 1, & p/2 < x \leq 1 - p/2, \\ \frac{1}{2}[1 - \cos\{2\pi(1-x)/p\}], & 1 - p/2 < x \leq 1, \end{cases}$$

and p is the total proportion of the series tapered ($p = 2m/n$, where m is as in equation (6.9), p. 69). It follows that

$$\begin{aligned} U_2 &\approx 1 - p + p \int_0^1 \left\{ \frac{1}{2}(1 - \cos 2\pi x) \right\}^2 dx \\ &= 1 - p + \frac{3p}{8} \\ &= 1 - \frac{5p}{8}, \end{aligned}$$

¹Note that this normalization differs from that which is appropriate when the discrete Fourier transform of the tapered data is used to estimate the parameters of a sinusoidal component of the data. (See Exercise 6.6, p. 76.)

and

$$\begin{aligned} U_4 &\approx 1 - p + p \int_0^1 \left\{ \frac{1}{2}(1 - \cos 2\pi x) \right\}^4 dx \\ &= 1 - p + \frac{35p}{128} \\ &= 1 - \frac{93p}{128} \end{aligned}$$

and hence

$$\begin{aligned} \frac{U_4}{U_2^2} &\approx \frac{1 - \frac{93p}{128}}{\left(1 - \frac{5p}{8}\right)^2} \\ &= \frac{128 - 93p}{2(8 - 5p)^2}. \end{aligned}$$

For $p = 0.1$ and 0.2 , the levels of tapering used on the sunspot numbers and the wheat-price index, respectively, U_4/U_2^2 has the values 1.055 and 1.116 , respectively. These modest increases in variances are a relatively small price to pay for the protection against possible leakage given by tapering in each case.

Combining these factors, we find that the variance of the discrete spectral average

$$\hat{s}(f) = \sum_u g_u I(f - f'_u)$$

is

$$\begin{aligned} \text{var}\{\hat{s}(f)\} &\approx \frac{U_4}{U_2^2} \frac{n'}{n} s(f)^2 \sum_u g_u^2 \\ &= g^2 s(f)^2, \end{aligned} \tag{9.12}$$

say. These arguments have to be modified for estimates at frequencies $f = 0$ and $f = 1/2$, for the periodogram is symmetric about both of these frequencies, and thus a central average at either frequency is the same as a one-sided average of one-half the length. The result is that the variance of $\hat{s}(0)$ and $\hat{s}(1/2)$ is twice the value given by (9.12). This is true whether or not the series mean is subtracted before the analysis.

Confidence Intervals

To construct confidence intervals for $\hat{s}(f)$ or $\ln \hat{s}(f)$ we need an approximation to the distribution of $\hat{s}(f)$. The simplest one is found by noting that, since $\hat{s}(f)$ is approximately a sum of independent variables, its

distribution is approximately normal. Since its mean is positive and its variance is small, $\ln \hat{s}(f)$ is also approximately normally distributed. A Taylor series expansion of $\ln \hat{s}(f)$ about $s(f)$ gives

$$\ln \hat{s}(f) \approx \ln s(f) + \frac{\hat{s}(f) - s(f)}{s(f)},$$

and hence

$$\begin{aligned} E\{\ln \hat{s}(f)\} &\approx \ln s(f), \\ \text{var}\{\ln \hat{s}(f)\} &\approx \frac{\text{var}\{\hat{s}(f)\}}{s(f)^2} \\ &\approx g^2. \end{aligned}$$

Thus, for example, a 95% approximate confidence interval for $\ln \hat{s}(f)$ is

$$\ln \hat{s}(f) \pm 1.96g$$

and the corresponding interval for $s(f)$ is

$$e^{-1.96g} \hat{s}(f) \leq s(f) \leq e^{1.96g} \hat{s}(f).$$

When g^2 is not small, the normal distribution may give a poor approximation to the distribution of $\ln \hat{s}(f)$. Since $\hat{s}(f)$ is approximately a sum of independent exponentially distributed quantities, another possibility is to use the χ^2 -distribution as an approximation to that of $\hat{s}(f)$ (see, e.g., Jenkins and Watts, 1968, pp. 87, 252-255). The mean μ , variance σ^2 , and degrees of freedom v of the χ^2 -distribution are related by $v\sigma^2 = 2\mu^2$ or $v = 2\mu^2/\sigma^2$. Thus, we approximate the distribution of $\hat{s}(f)$ by the χ^2 -distribution with

$$v = \frac{2s(f)^2}{g^2 s(f)^2} = \frac{2}{g^2}$$

degrees of freedom. The quantity $2/g^2$ is called the *equivalent number of degrees of freedom* of the estimate $\hat{s}(f)$. To be exact,

$$\chi^2 = \frac{v\hat{s}(f)}{s(f)}$$

is approximately χ^2 -distribution with v degrees of freedom. Thus, for instance, a 95% confidence interval for $\hat{s}(f)$ is given by

$$\frac{v\hat{s}(f)}{\chi_v^2(0.975)} \leq s(f) \leq \frac{v\hat{s}(f)}{\chi_v^2(0.025)},$$

where $\chi_v^2(0.025)$ and $\chi_v^2(0.975)$ are the 2.5% and 97.5% points of the χ^2 -distribution with v degrees of freedom.

In constructing confidence intervals from the χ^2 -distribution it is conventional to place equal masses of probability in two tails. This does not give the shortest interval for a given level of confidence, and the interval is not symmetric about the estimated value on either a linear or a logarithmic scale. However, it does provide the convenient property that, for instance, the upper limit of a 95% confidence interval is also a 97.5% *upper confidence bound*.

The distribution of $\hat{s}(f)$ is in fact exactly χ^2 for the Daniell estimate (to be exact, for the discrete Daniell estimate with $n' = n$ and no tapering, applied to a Gaussian white noise series). However, for other estimates it is more skewed than the approximating χ^2 -distribution. The distribution of $\hat{s}(f)^\alpha$ is less skewed than that of $\hat{s}(f)$ if $\alpha < 1$, and thus a better approximation can be found by matching the first three approximate moments of $\hat{s}(f)$ to those of $\beta \times (\chi^2)^{1/\alpha}$, where χ^2 has the χ^2 -distribution with v degrees of freedom. However, this is not an elementary problem, and we have not used intervals based on its solution.

Examples

The estimate of the spectrum of the wheat-price index shown in Figure 8.12 (p. 157) is a discrete average of periodogram ordinates at the Fourier frequencies, for a data set of length 300, and the factor n'/n is therefore 1. The value of $\sum g_u^2$ is 0.153. The correction factor for tapering (of 20%, or $p = 0.2$) is 1.116. Thus, the combined factor $g^2 = (U_4/U_2^2)(n'/n) \sum g_u^2$ has the value 0.171. As noted in Section 9.5, this value is appropriate for all frequencies not close to 0 or 1/2 cycles per unit time; g^2 progressively increases to twice this value as f approaches either of these limiting frequencies.

The approximate 95% confidence limits for $\hat{s}(f)$ based on the normal distribution are $\hat{s}(f)$ multiplied by $e^{\pm 1.96g} = 2.247$ and 0.245. The degrees of freedom in the approximating χ^2 -distribution are $2/0.171 = 11.7$, and the factors providing the 95% confidence interval are 2.765 and 0.511. Figure 9.1 shows the same spectrum estimate as Figure 8.12, with these confidence limits added.² The normal theory limits are symmetric around the spectrum estimate because the graph has a logarithmic axis, and are shown by short dashed lines. The χ^2 theory limits are asymmetric, and are shown by the long dashed lines.

In the same way, Figure 9.2 shows the spectrum estimate of Figure 8.13

²The confidence intervals are wider for frequencies close to 0 and 1/2 cycles per year because of the adjustment of g^2 .

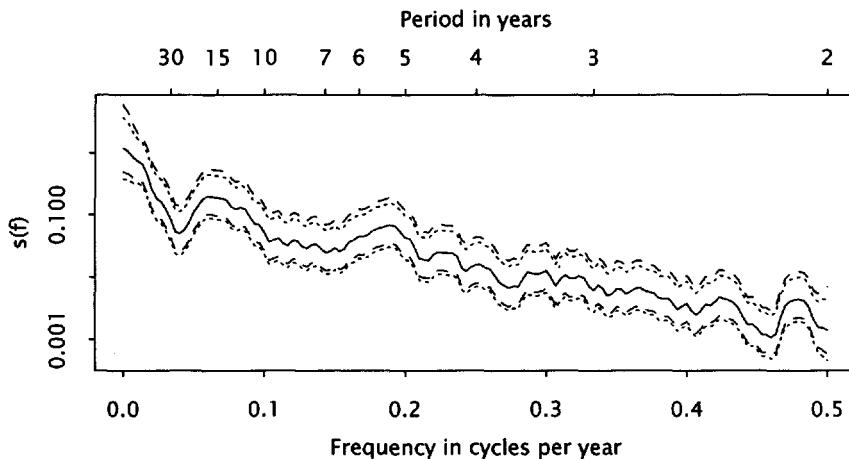


Fig. 9.1 Smoothed periodogram of logarithms of wheat price index, with approximate 95% confidence limits. (Modified Daniell filter, $m = 6$.)

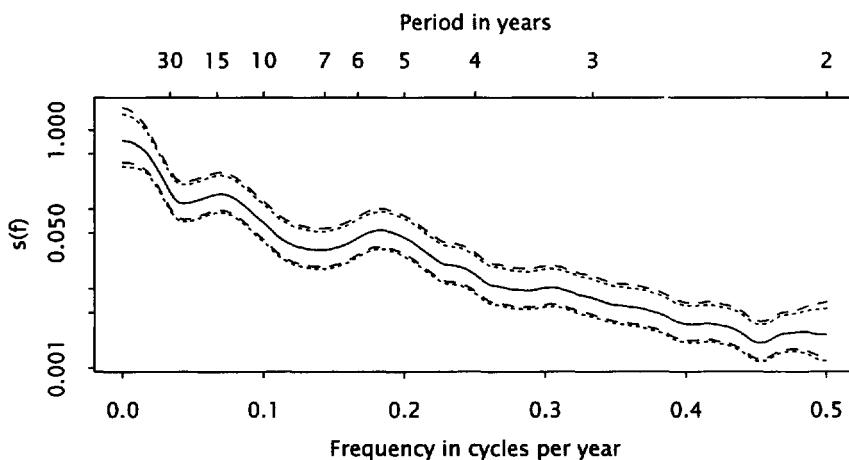


Fig. 9.2 Smoothed periodogram of logarithms of wheat price index, with approximate 95% confidence limits. (Modified Daniell filter, $m = 6, 12$.)

(p. 158) with approximate confidence limits added. In this case the value of $\sum g_u^2$ is 0.0687, and with the same correction for tapering the combined factor g^2 has the value 0.0767. The factors for the normal theory limits are 1.721 and 0.581, and those for the χ^2 theory limits, based on $2/0.0767 = 26.1$ degrees of freedom, are 1.876 and 0.621.

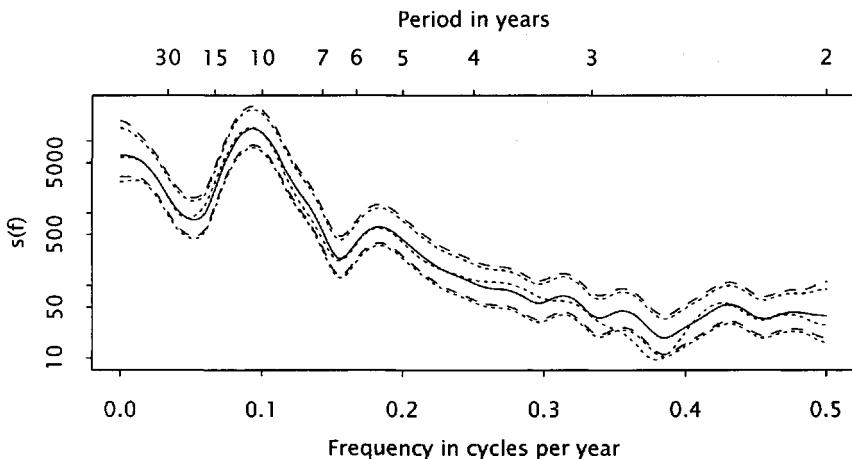


Fig. 9.3 Smoothed periodograms of yearly sunspot numbers (solid line) and their square roots (broken line), with approximate 95% confidence limits. (Modified Daniell filter, $m = 6, 6, 6$.)

The limits shown in Figure 9.1 are quite wide, and clearly none of the local fluctuations in that curve represents a statistically significant feature. The interval based on the χ^2 -distribution is quite asymmetric, and is 7% wider than the more approximate interval based on normal distribution theory. The limits shown in Figure 9.2 are narrower. The χ^2 limits are more nearly symmetric and in this case are only 2% wider than the normal theory limits. In both graphs the normal theory limits provide very reasonable approximations, at least as a visual guide to the statistical variability in the spectrum estimates.

For the spectrum estimates of the sunspot series shown in Figure 8.14 on p. 159, we have $n = 299$, $n' = 300$, $U_4/U_2^2 = 1.055$, and $\sum g_u^2 = 0.0943$. The approximate 95% confidence limits based on the normal distribution are therefore 1.826 and 0.548, and the limits based on the appropriate χ^2 -distribution (with 21.1 degrees of freedom) are 2.034 and 0.593. Figure 9.3 shows the same pair of spectrum estimates as Figure 8.14, with Gaussian-based and χ^2 -based confidence intervals for the spectrum of the original untransformed values. Again, the normal theory interval is accurate enough for graphical presentation. Note that the difference between the estimated spectra of the original series and of the square-rooted series is mostly much smaller than the width of the confidence interval.

Either of the confidence intervals we have described may be plotted as a

band surrounding the estimated spectrum, as in the figures shown above. However, neither interval provides a *simultaneous confidence band*, which would have the property that the true spectrum is everywhere covered by the band with a given probability, say 0.95. Instead, we should expect the true spectrum to fall outside the band over about 5% of the interval $[0, 1/2]$. The results of Woodroffe and Van Ness (1967) would allow the construction of a simultaneous confidence band for the whole of the estimated spectrum. However, this simultaneous confidence band would be wider in each case by a factor of around $\log m$, where m is $1/(\text{bandwidth of estimate})$. For relatively short series such as the wheat-price index and the annual sunspot numbers, these wide simultaneous confidence bands are not particularly useful. It also seems likely that the approximation is good only for very long series (see Hannan, 1970, p. 294).

Exercise 9.7 Approximate Variances

Verify the steps leading to the approximation (9.7) for the variance of an average of periodogram ordinates at Fourier frequencies.

Exercise 9.8 Approximation

Show that the difference between (9.8) and (9.9) is

$$2 \sum_{r=1}^m w_r c_{n-r} \cos 2\pi r f. \quad (9.13)$$

Note that for $r > 0$

$$c_{n-r} = \frac{1}{n} \sum_{t=0}^{r-1} x_t x_{t+n-r},$$

and hence

$$\begin{aligned} E(c_{n-r}^2) &= \frac{1}{n^2} \sum_{t,t'=0}^{r-1} E(x_t x_{t+n-r} x_{t'} x_{t'+n-r}) \\ &\leq \frac{1}{n^2} \sum_{t,t'=0}^{r-1} \{E(x_t^4) E(x_{t+n-r}^4) E(x_{t'}^4) E(x_{t'+n-r}^4)\}^{1/4} \\ &= \frac{k^2 r^2}{n^2}, \end{aligned}$$

where $k^2 = E(x_t^4)$ is assumed to be finite. Thus $|c_{n-r}|$ is at most of order kr/n in probability, and hence the difference (9.13) is at most of the order

$$\frac{2k}{n} \sum_{r=1}^m r|w_r|$$

in probability. If $w_r = w(r/m)$, then

$$\begin{aligned} \sum_{r=1}^m r|w_r| &= \sum_{r=1}^m r \left| w\left(\frac{r}{m}\right) \right| \\ &\approx \frac{m}{2} \int_{-1}^1 |xw(x)|dx, \end{aligned}$$

and so the difference (9.13) is at most of the order

$$\frac{km}{n} \int_{-1}^1 |xw(x)|dx$$

in probability. Now the variance of the discrete average spectrum estimate (9.9) is of the order m/n , whence its standard deviation is of the order $\sqrt{m/n}$, and thus the difference between (9.8) and (9.9) is negligible compared with the standard deviation of either.

Exercise 9.9 Effect of Tapering

Suppose that $\{X_t\}$ is a Gaussian white noise series, that is, that the random variables $\{X_t\}$ are independent, each with the standard normal distribution. Suppose that the stretch $\{X_0, X_1, \dots, X_{n-1}\}$ is tapered by $\{u_0, u_1, \dots, u_{n-1}\}$, and that the periodogram $I(f)$ and a spectrum estimate $\hat{s}(f)$ with lag weights $\{w_r\}$ are computed from the tapered data.

(i) Verify (9.11).

(ii) Show that

$$\hat{s}(f) = \frac{1}{n} \sum_{t,t'=0}^{n-1} X_t X_{t'} u_t u_{t'} w_{t-t'} e^{-2\pi i f(t-t')}.$$

(iii) Find the expected value and variance of $\hat{s}(f)$. (Hint: Note that

$$E(X_t X_{t'}) = \delta_{t,t'} = \begin{cases} 1, & t = t' \\ 0, & t \neq t' \end{cases}$$

and

$$E(X_t X_{t'} X_u X_{u'}) = \delta_{t,t'} \delta_{u,u'} + \delta_{t,u} \delta_{t',u'} + \delta_{t,u'} \delta_{t',u},$$

from the properties of the normal distribution.)

- (iv) If $u_t = u\{(2t + 1)/2n\}$ for some smooth data window $u(x)$, show that for large n

$$E\{\hat{s}(f)\} \approx \int_0^1 u(x)^2 dx,$$

and if $f \neq 0$ or $1/2$ cycles per unit time,

$$\text{var}\{\hat{s}(f)\} \approx \frac{1}{n} \left(\sum w_r^2 \right) \int_0^1 u(x)^4 dx.$$

9.6 PROPERTIES OF SPECTRAL WINDOWS

It was shown in Section 9.5 that the variance of the discrete spectral average

$$\hat{s}(f) = \frac{m}{n'} \sum_u V(mf'_u) I(f - f'_u) \quad (9.14)$$

is approximately

$$\frac{mU_4}{n} s(f)^2 \int V(x)^2 dx,$$

where U_4 is the correction for tapering. It was also shown that $\hat{s}(f)$ is approximately unbiased. To obtain a more exact description of the bias, $E\{\hat{s}(f)\} - s(f)$ may be approximated by

$$b(f) = \left\{ \frac{m}{n'} \sum_u V(mf'_u) s(f - f'_u) \right\} - s(f).$$

Now if $s(f)$ is smooth enough,

$$s(f - f'_u) \approx s(f) - f'_u s'(f) + \frac{f'^2_u}{2} s''(f),$$

and hence

$$\begin{aligned} b(f) &\approx \frac{m}{n'} s(f) \sum_u V(mf'_u) - s(f) \\ &\quad - \frac{m}{n'} s'(f) \sum_u f'_u V(mf'_u) \\ &\quad + \frac{m}{2n'} s''(f) \sum_u f'^2_u V(mf'_u). \end{aligned}$$

Now

$$\frac{m}{n'} \sum_u V(mf'_u) \approx \frac{1}{m} \int V(x) dx = 1,$$

and

$$\frac{m}{n'} \sum f'_u V(mf'_u) \approx \frac{1}{m} \int xV(x)dx,$$

which vanishes if $V(x)$ is symmetric, the usual case. A similar approximation for the last term³ gives

$$b(f) \approx \frac{1}{2} \frac{s''(f)}{2m^2} \int x^2V(x)dx. \quad (9.15)$$

Since this approximation to the bias depends on the spectral density function only through $s''(f)$, it clearly represents only the contribution from local "smudging" of the spectrum, and not leakage. As might be expected, the bias is negative at peaks and positive at troughs. Failure to represent the local features of a spectrum was referred to in Section 8.6 as loss of *resolution*. Since loss of resolution is caused by a large bandwidth, it seems reasonable to define the bandwidth of the window $W(f) = mV(mf)$ to be

$$\frac{1}{m} \sqrt{\int x^2V(x)dx}.$$

To the present order of approximation, the statistical properties of the spectrum estimate $\hat{s}(f)$ depend only on n , m , $s(f)$, $s''(f)$, $\int V(x)^2dx$, and $\int x^2V(x)dx$. The approximations to bias and variance show that making both of the latter two quantities small is desirable. If $V_c(x) = cV(cx)$, then $V_c(x)$ may also be regarded as a standardized spectral window, since

$$\int V_c(x)dx = \int V(x)dx = 1.$$

Now

$$\int V_c(x)^2dx = c \int V(x)^2dx,$$

and

$$\int x^2V_c(x)dx = \frac{1}{c^2} \int x^2V(x)dx,$$

³This approximation depends on the integrability of $x^2V(x)$, which fails for the Bartlett window, since its sidelobes decay as x^{-2} (see, e.g., Jenkins and Watts, 1968, p. 247, for an account of the bias of the Bartlett estimate).

and hence

$$\left\{ \int V_c(x)^2 dx \right\}^2 \int x^2 V_c(x) dx = \left\{ \int V(x)^2 dx \right\}^2 \int x^2 V(x) dx.$$

Since $V(x)$ and $V_c(x)$ generate the same family of windows, the invariant quantity

$$\left\{ \int V(x)^2 dx \right\}^2 \int x^2 V(x) dx \quad (9.16)$$

is a suitable figure of merit of the standardized window $V(x)$, small values being desirable. Exercise 9.10 shows that choosing the bandwidth to minimize the approximate mean squared error leads to the same functional of $V(x)$.

For nonnegative windows the figure of merit is always at least $9/125 = 0.072$.⁴ For the rectangular window that corresponds most closely to the modified Daniell procedure used in Section 8.7, the value is $1/12 = 0.083$. For the combined window that results from many repeated applications of the Daniell procedure, the value is approximately $1/4\pi = 0.080$ (see Exercise 9.13). Neither of these values is unacceptably high compared with the minimum. The results of Exercise 9.12 show that with two applications of the Daniell procedure with one averaging length being twice the other, the minimum is almost achieved. However, three applications give the spectrum estimate a smoother appearance (see Section 8.7).

Smaller values of the figure of merit may be obtained when the window is not constrained to be nonnegative, and there is, in fact, no positive lower bound in this case. However, negative sidelobes in a spectral window are generally undesirable, as they may easily lead to negative values of the spectrum estimate.

The values of the figure of merit for many commonly used windows may be computed from values given by Anderson (1994, Table 9.3.4); $\int x^2 V(x) dx$ is $2k$ in Anderson's notation, and similarly $\int V(x)^2 dx$ is $(1/2\pi) \int k(x)^2 dx$. For example, the value for the widely used Parzen window is 0.088.

The bandwidth of a spectral window is sometimes defined as the width of the rectangular window that gives the same variance. For $W(f) = mV(mf)$ this is therefore $\{m \int V(x)^2 dx\}^{-1}$. However, this definition has the disadvantage that both bandwidth and approximate variance depend on the shape of the window through the same functional of $V(x)$. Thus it provides no tool for discriminating between windows of different shapes.

⁴The window that attains the minimum value is not particularly desirable (see Exercise 9.11).

Further discussion of the properties of spectral windows may be found in Parzen (1957a, 1961), Jenkins (1961), Marple (1987), Priestley (1981a, Chapter 7) and Rosenblatt (1971). The figure of merit described above was also used by Parzen (1957a). The lower bound and the window that attains it are given by Epanechnikov (1969). The same inequality was derived in a different statistical context by Hodges and Lehmann (1956).

Exercise 9.10 Minimum Mean Squared Error

Combine the expressions (9.15) and (9.14) to find an approximation to the mean squared error (bias² + variance). Find the value of the parameter m that minimizes the mean squared error, and the minimum value.

Notes:

- (i) The minimum mean squared error depends on the standardized window $V(x)$ only through the invariant figure of merit (9.16).
- (ii) The bandwidth that achieves the minimum depends on $s(f)$ and $s''(f)$ and, therefore, itself depends on f .
- (iii) An optimal bandwidth that does not depend on f may be defined by minimizing the integral, possibly weighted, of the mean squared error. It still depends on the standardized window $V(x)$ only through the figure of merit.
- (iv) All minimum mean squared error considerations lead to optimal bandwidths that depend on the true spectrum. Since the latter is typically unknown, finding an optimal m is not straightforward.

Exercise 9.11 Optimal Window

Let

$$V_0(x) = \begin{cases} \frac{3}{4}(1 - x^2), & |x| \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

- (i) Verify that

$$\int v_0(x)dx = 1,$$

and evaluate $\int x^2 V_0(x)dx$ and $\int (V_0(x))^2 dx$. Show that the figure of merit

$$\left\{ \int V_0(x)^2 dx \right\}^2 \int x^2 V_0(x)dx$$

has the value 9/125.

- (ii) Let $V(x)$ be any nonnegative window for which

$$\int V(x)dx = 1$$

and

$$\int x^2V(x)dx = \int x^2V_0(x)dx.$$

Show that

$$\int V(x)^2dx \geq \int V_0(x)^2dx,$$

with equality only if $V(x) \equiv V_0(x)$.

(Hint: Write $V(x) = V_0(x) + \delta(x)$, and note that $\int \delta(x)dx = 0$, $\int x^2\delta(x)dx = 0$, and $\delta(x) \geq 0$ for $|x| \geq 1$.)

Note that because this window does not approach 0 smoothly, it would not result in particularly smooth spectrum estimates.

Exercise 9.12 Daniell Window and Combinations

- (i) Evaluate the figure of merit for the standardized Daniell window

$$V(x) = \begin{cases} \frac{1}{2}, & |x| \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

- (ii) Two applications of Daniell smoothing with the same length of averaging in each step result in the triangle window

$$V(x) = \begin{cases} 1 - |x|, & |x| \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Show that the value of the figure of merit is $2/27 = 0.074$.

- (iii) Two applications of Daniell smoothing with different lengths of averaging result in the trapezium window

$$V(x) = \begin{cases} \frac{1}{1+p}, & |x| \leq p, \\ \frac{1-|x|}{1-p^2}, & p < x \leq 1, \\ 0, & \text{otherwise,} \end{cases}$$

for some $0 \leq p \leq 1$, where the ratio of the two averaging lengths is $(1+p):(1-p)$. Show that the value of the figure of merit is

$$\frac{2(1+2p)^2(1+p^2)}{27(1+p)^4}$$

and that the values for $p = 1/2$ and $p = 1/3$ are 0.0732 and 0.0723, respectively. Note that these values of p correspond to ratios of smoothing lengths of 3 : 1 and 2 : 1, respectively.

Exercise 9.13 Gaussian Window

Any smooth probability density function may be used as a spectral window. The Gaussian window is

$$V(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

- (i) Show that $\int x^2 V(x) dx = 1$. (*Hint:* Integrate by parts, and use the fact that $\int V(x) dx = 1$.)
- (ii) Show that $\int V(x)^2 dx = 1/4\pi = 0.0796$.

Note that repeated application of filters corresponds to convolution of the windows, and hence by the central limit theorem the combined window is approximately Gaussian.

9.7 ALIASING AND THE SPECTRUM

The time series under analysis often consists of measurements made at a discrete, equally spaced, set of times on some phenomenon that is actually evolving continuously, or at least on a much finer time scale. Thus there may well be oscillations in the behavior of the phenomenon with frequencies higher than the Nyquist frequency associated with the sampling interval; see Section 2.5.

The simplest situation is as follows. Suppose that $\{X(t)\}$ is some random process defined for all real values of t . The two most common ways of deriving a discrete set of measurements of such a process are as follows:

- (i) *Sampling:* $Y_t = X(t)$; and
- (ii) *Averaging:* $Z_t = \int_{t-1}^t X(u) du$.

It is, in fact, sufficient to consider sampling.

The theory of spectra of random processes (stochastic processes) defined for a continuous variable t is essentially similar to that described in Section 9.1. The analog of the autocovariance sequence is the *autocovariance function*

$$\gamma(\tau) = \text{cov}\{X(t), X(t + \tau)\},$$

which has the representation

$$\gamma(\tau) = \int_{-\infty}^{\infty} e^{2\pi if\tau} dS(f).$$

Note that the limits of integration are now infinite, since there is no Nyquist frequency to limit the frequency of oscillations. The finiteness of the variance $\gamma(0)$ requires that the total increase of the spectral distribution function, $S(\infty) - S(-\infty)$, be finite. It is easily seen that this implies the continuity of $\gamma()$.

Now if $\{X(t)\}$ is sampled according to (i),

$$\begin{aligned}\gamma_r &= \text{cov}(Y_t, Y_{t+r}) \\ &= \text{cov}\{X(t), X(t+r)\} \\ &= \gamma(r) \\ &= \int_{-\infty}^{\infty} e^{2\pi ifr} dS(f).\end{aligned}$$

For simplicity, suppose that $S(f)$ has a derivative $s(f)$. Then for any integer r ,

$$\begin{aligned}\gamma_r &= \int_{-\infty}^{\infty} e^{2\pi ifr} s(f) df \\ &= \sum_{k=-\infty}^{\infty} \int_k^{k+1} e^{2\pi ifr} s(f) df \\ &= \sum_{k=-\infty}^{\infty} \int_0^1 e^{2\pi i(f-k)r} s(f-k) df \\ &= \sum_{k=-\infty}^{\infty} \int_0^1 e^{2\pi ifr} s(f-k) df,\end{aligned}$$

since $e^{-2\pi ikr} = 1$. Interchanging the order of summation and integration yields

$$\gamma_r = \int_0^1 e^{2\pi ifr} \left\{ \sum_{k=-\infty}^{\infty} s(f-k) \right\} df,$$

and consequently the spectral density function of $\{Y_t\}$ is

$$s_Y(f) = \sum_{k=-\infty}^{\infty} s(f-k), \quad 0 \leq f < 1. \quad (9.17)$$

The right-hand side of (9.17) is the *aliased spectral density function*. Its value at frequency f cycles per unit time consists of the sum of contributions from all frequencies of the form $f - k$ cycles per time for integer k .

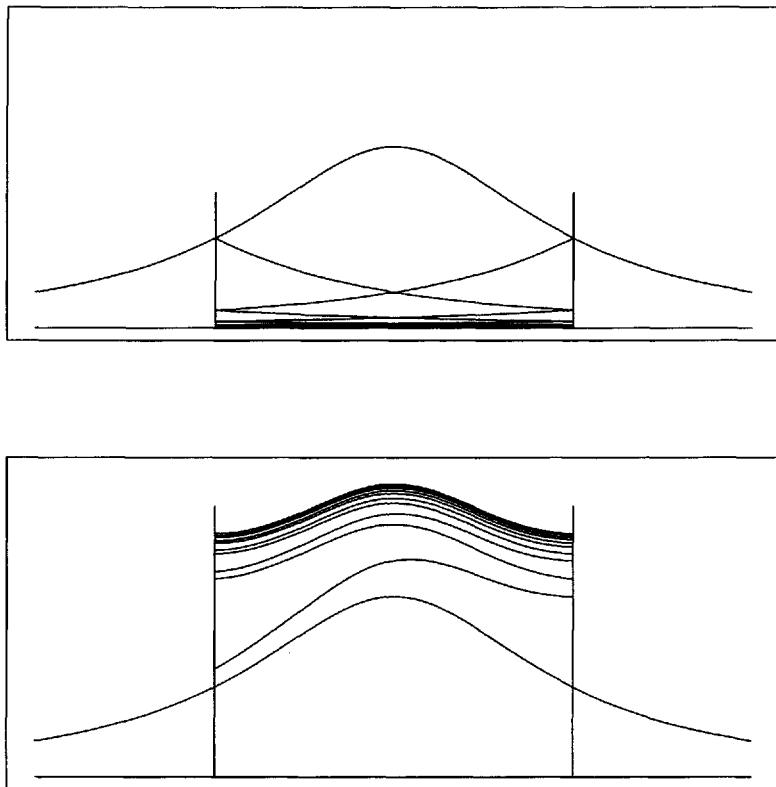


Fig. 9.4 An aliased spectrum

Since f is the *principal alias* of all of these frequencies (see Section 2.5), this is an intuitively natural result.

Aliasing of the spectral density function is illustrated for the function $s(f) = 1/(1 + 4f^2)$ in Figure 9.4. The upper graph shows this function “folded in” onto the interval $(-1/2, 1/2)$, and the lower graph shows the sum of the various contributions. In this case, the series may be summed analytically to yield $s_Y(f) = \pi \sinh \pi/2(\cosh \pi - \cos 2\pi f)$.

For the series $\{Z_t\}$ obtained by averaging the input series as in (ii), a similar result holds. If the continuous time process $\{W(t)\}$ is defined by

$$W(t) = \int_{t-1}^t X(t') dt', \quad (9.18)$$

then $\{Z_t\}$ is the result of sampling $\{W(t)\}$. The operation (9.18) is an example of continuous time filtering and is exactly analogous to the discrete time filtering described in Section 7.2. The transfer function is again defined to be the factor by which a complex sinusoid of frequency f cycles per unit time is changed, and for the operation (9.18) is

$$e^{-\pi i f} \frac{\sin \pi f}{\pi f}.$$

The spectral density function of $\{W(t)\}$ is

$$s_W(f) = |G(f)|^2 s(f) = \left(\frac{\sin \pi f}{\pi f} \right)^2 s(f).$$

Since $\{Z_t\}$ is found by sampling $\{W(t)\}$, its spectral density function is given (9.17) with $s(f)$ replaced by $s_W(f)$. Thus

$$\begin{aligned} s_Z(f) &= \sum_{k=-\infty}^{\infty} s(f-k) \left\{ \frac{\sin \pi(f-k)}{\pi(f-k)} \right\}^2 \\ &= \left(\frac{\sin \pi f}{\pi f} \right)^2 \left\{ s(f) + \sum_{k \neq 0} s(f-k) \left(\frac{f}{f-k} \right)^2 \right\}. \end{aligned} \quad (9.19)$$

The series in (9.19) converges more rapidly than that in (9.17) because of the additional factor, which decays as $1/|k|^2$. Thus aliasing is reduced, but at the cost of some distortion of the leading term. The factor $(\sin \pi f)^2 / (\pi f)^2$ has the value 1 at $f = 0$, but falls to $4/\pi^2 = 0.405$ at $f = 1/2$ cycles per unit time. Thus power at frequencies close to the Nyquist frequency is attenuated to around 40% of its true value.

Aliasing can be a confusing phenomenon, since a peak in the true spectrum at a frequency beyond the Nyquist frequency may be strong enough to be seen in the aliased spectrum. This may give the impression that a frequency is significant when it is not, or the peak may partly obscure another frequency of interest. When the sampling interval (the time interval between consecutive observations) is at the disposal of the experimenter, it may be chosen small enough that no substantial amount of power falls beyond the Nyquist frequency, avoiding aliasing problems completely.

However, the cost of collecting the data may become prohibitive. One compromise is to collect the data initially with a small sampling interval, filter these high sampling rate data to remove unwanted power at high frequencies, and then *decimate* the output (i.e., take every k th observation, where k is chosen to increase the sampling interval to a desired level). The simplest such filter would be the discrete analog of (ii), namely, an unweighted average of k successive values.

More sophisticated filters could be designed at in Sections 7.2 and 7.3. A desirable filter would distort the spectrum as little as possible up to the reduced Nyquist frequency, while passing as little power as possible from higher frequencies. These characteristics are essentially those of a low-pass filter, with its pass band ending at the new Nyquist frequency.

For many common data collection tasks, specialized data collection systems are available that incorporate high frequency sampling, filtering, and decimation. The occasional user of such equipment or of the resulting data may be quite unaware of the steps that are taken between the actual measurement and the presentation of results to the user.

10

Analysis of Multiple Series

By a multiple series we mean a number of time series observed simultaneously. This should not be confused with situations in which "time" itself is multidimensional, as when data are collected on a grid in the plane. Fourier-analytic methods may be extended in a fairly obvious way to the latter type of data, but the subject will not be discussed here. For most of this chapter we examine the situation in which a pair of series is observed; that is, at each time point t we have available a pair of numbers, (x_t, y_t) . The separate series $\{x_t\}$ and $\{y_t\}$ will be referred to as *components*.

When only a single series is observed, the questions that we usually wish to answer are of the form "What is the *internal* structure of this set of data?" If the data have an oscillatory appearance of the loose form encountered in the examples of preceding chapters, one answer will usually be a description of the data by one of the methods already discussed, namely, harmonic analysis (Chapter 6), complex demodulation (Chapter 7), and spectrum analysis (Chapter 8). When two series are observed, we shall usually be interested in the internal structure of each, but in addition we shall be concerned with their *joint* structure, or the dependence of either series on the other.

Harmonic analysis and its local form, complex demodulation, may be used without modification on a multiple series, by analyzing the series component by component. Suppose, for instance, that we observe a pair of series $\{x_t\}$ and $\{y_t\}$, each containing oscillations at a frequency around

f cycles per unit time, but with amplitude and phase fluctuations. Thus we may write

$$x_t = R_{x,t} \cos(2\pi ft + \phi_{x,t}) + x'_t,$$

and

$$y_t = R_{y,t} \cos(2\pi ft + \phi_{y,t}) + y'_t,$$

where $\{x'_t\}$ and $\{y'_t\}$ denote other components in the two series. By complex demodulation of the two component series, we may find approximations to the instantaneous amplitudes $R_{x,t}$ and $R_{y,t}$, and the instantaneous phases $\phi_{x,t}$ and $\phi_{y,t}$. From the phases we may also compute the *instantaneous relative phase* of $\{x_t\}$ relative to $\{y_t\}$, $\phi_{x,t} - \phi_{y,t}$. The instantaneous phase of either series is interpreted as the amount by which its oscillations have drifted from the oscillations in a pure sinusoid with frequency f cycles per unit time, and thus the relative phase measures the extent to which oscillations in both series drift together. In this way it gives information about the joint behavior of the two series.

Estimated spectra, on the other hand, give information only about the oscillations in individual series. Similarities between the spectra of two series, such as peaks at similar frequencies, may suggest that the series are related. To investigate such suggestions we may compute estimates of the *cross spectrum* of the two series. This is an extension of the definition of the spectrum, and is conveniently estimated by smoothing the *cross periodogram*.

10.1 CROSS PERIODOGRAM

The periodogram has a natural extension to the multiple series context. Let $d_x(f)$ and $d_y(f)$ denote the Fourier transforms of the component series; that is,

$$d_x(f) = \frac{1}{n} \sum_{t=0}^{n-1} x_t e^{-2\pi i f t},$$

with $d_y(f)$ defined similarly. The periodograms of the component series are $n|d_x(f)|^2$ and $n|d_y(f)|^2$, respectively. These will now be denoted $I_{x,x}(f)$ and $I_{y,y}(f)$ and referred to as the *autoperiodograms* of the component series to distinguish them from the *cross periodogram*,

$$I_{x,y}(f) = n d_x(f) \overline{d_y(f)}.$$

Notice that, in contrast with autoperiodograms, the cross periodogram $I_{x,y}(f)$ is, in general, not positive, nor even real, and not symmetric. Instead it satisfies the identities

$$I_{x,y}(-f) = I_{y,x}(f) = \overline{I_{x,y}(f)}.$$

These identities do imply that $I_{x,y}(f)$ is real at $f = 0$ and $1/2$ cycles per unit time, but still not necessarily positive.

Another identity is

$$\left| I_{x,y}(f) \right|^2 = I_{x,x}(f)I_{y,y}(f), \quad (10.1)$$

whence the magnitude of the cross periodogram contains no information that is not in the autoperiodograms. If we write the transform $d_x(f)$ as $|d_x(f)|e^{2\pi i\phi_x(f)}$, the phase of the cross periodogram is $\phi_x(f) - \phi_y(f)$, the difference of the phases of the two component series, or the *relative phase*. Note that the relative phase does not depend on the choice of origin for the time scale. It would be the same for any stretch of n consecutive observations on these series, unlike the phases of the two transforms $d_x(f)$ and $d_y(f)$ separately.

Thus the cross periodogram contains no information that is not contained in the two Fourier transforms. However, the same is true of the autoperiodogram, and yet it is a useful tool, especially in a smoothed form. Recall the observation made in Chapter 8 that when the series being analyzed contains oscillations with no strong periodicity, the periodogram may still show statistical regularity, in that its overall shape, although not its fine structure, is consistent from one stretch of the series to another. For the same kind of data the cross periodogram shows a similar statistical regularity. The smooth function that results from averaging cross periodograms from different stretches of data is called the *cross spectrum*. It is in general complex-valued, like the cross periodogram.

Estimates of the cross spectrum will be constructed by smoothing the cross periodogram, using much the same procedures as for smoothing autoperiodograms. The only new problem that arises with smoothing cross periodograms involves the *alignment* of the series. This is discussed below in Section 10.6.

Exercise 10.1 Properties of the Cross Periodogram

- (i) Verify that $I_{x,y}(-f) = I_{y,x}(f) = \overline{I_{x,y}(f)}$.
- (ii) Suppose that two stretches of data $\{x_0, x_1, \dots, x_{n-1}\}$ and $\{y_0, y_1, \dots, y_{n-1}\}$ are shifted cyclically by amounts h and k , respectively. Show that the phase of the cross periodogram $I_{x,y}(f)$ changes by a linear function of frequency, which vanishes if $h = k$.

10.2 ESTIMATING THE CROSS SPECTRUM

The formula for the cross periodogram may be rearranged to give

$$I_{x,y}(f) = \sum_{|r| < n} c_{x,y,r} e^{-2\pi i f r}, \quad (10.2)$$

where

$$c_{x,y,r} = \frac{1}{n} \sum x_t y_{t-r}, \quad |r| < n,$$

the latter sum being over all t for which both t and $t - r$ lie in the range $0, 1, \dots, n - 1$. The quantity $c_{x,y,r}$ is the *cross covariance* of $\{x_t\}$ and $\{y_t\}$ at lag r . Notice that, in general, $c_{x,y,-r} \neq c_{x,y,r}$, but instead $c_{x,y,-r} = c_{y,x,r}$.

The expansion (10.2) is analogous to the expression given in Section 8.3 for the autoperiodogram. As a result, the discussion of smoothing auto-periodograms in Chapter 8 is immediately applicable to smoothing cross periodograms. In particular, all of the various equivalences derived in Section 8.5 between the different types of estimates extend to the multiple series situation. In the examples that follow, the *discrete spectral averages* described in Section 8.5 will be used for smoothing the cross periodogram.

For an example of estimating cross spectra we use the one-year and ten-year interest rates¹ shown in Figure 10.1 and the industrial production index² shown in Figure 10.2. Industrial production was converted into the monthly percentage changes shown in Figure 10.3.

Spectrum estimates of the interest rates and monthly changes in industrial production are shown in Figures 10.4 and 10.5, respectively. Each series was corrected for its mean and tapered 20%. The periodogram of each was then smoothed by three applications of modified Daniell smoothing of length 16 (see Section 8.7; the resulting spectral window is shown inset).

The cross periodogram $I_{x,y}(f)$ was smoothed in the same way to pro-

¹These time series are retrievable from FRED, the Federal Reserve Economic Data archive <http://www.stls.frb.org/fred/index.html> maintained by the Federal Reserve Bank of Saint Louis, <http://www.stls.frb.org>. The one-year rates are in the file <http://www.stls.frb.org/fred/data/irates/gs1> and the ten-year rates are in the file <http://www.stls.frb.org/fred/data/irates/gs10>.

²The industrial production index is also retrievable from the FRED data archive, in the file <http://www.stls.frb.org/fred/data/business/indpro>.

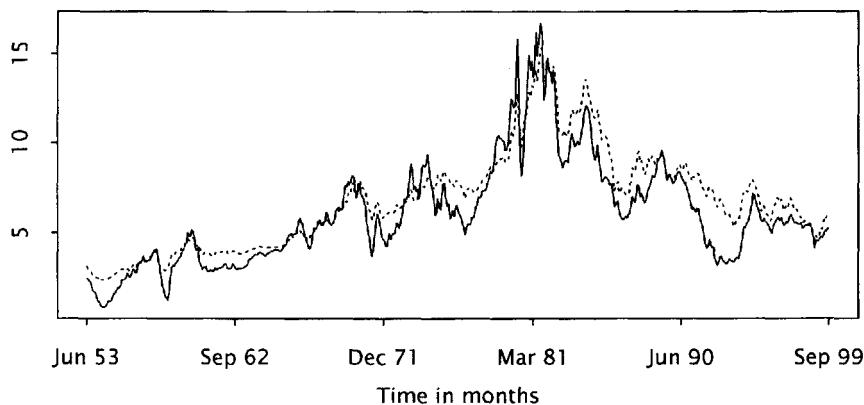


Fig. 10.1 One-Year (solid line) and Ten-Year (broken line) Treasury Constant Maturity Rates. (Source: H.15 Release — Federal Reserve Board of Governors.)

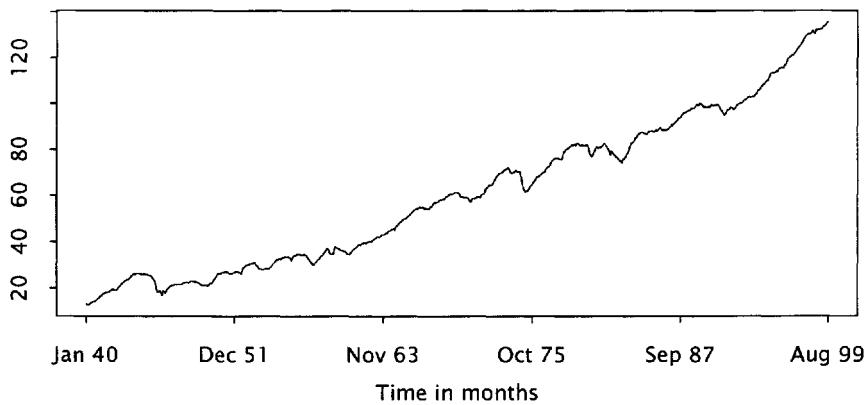


Fig. 10.2 Industrial Production Index. 1992=100, Seasonally Adjusted. (Source: Federal Reserve Board, Washington, DC.)

duce an estimate of the cross spectrum:

$$\hat{s}_{x,y}(f) = \sum_u g_u I_{x,y}(f - f_u). \quad (10.3)$$

This function has the same symmetries as $I_{x,y}(f)$, namely,

$$\hat{s}_{x,y}(-f) = \hat{s}_{y,x}(f) = \overline{\hat{s}_{x,y}(f)}.$$

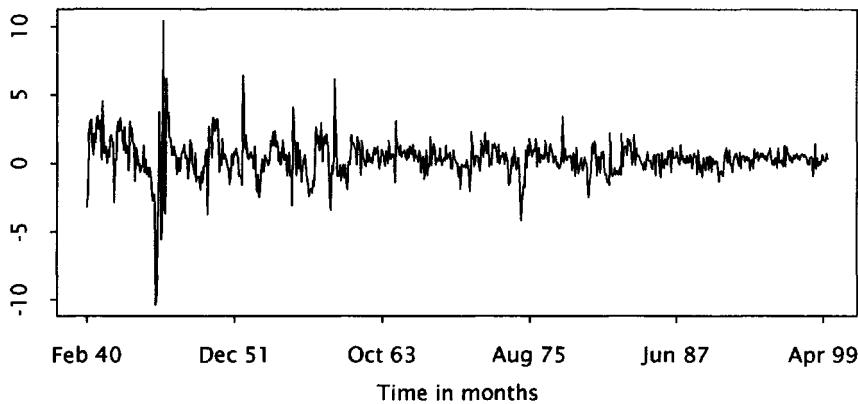


Fig. 10.3 Monthly percentage change in Industrial Production Index.

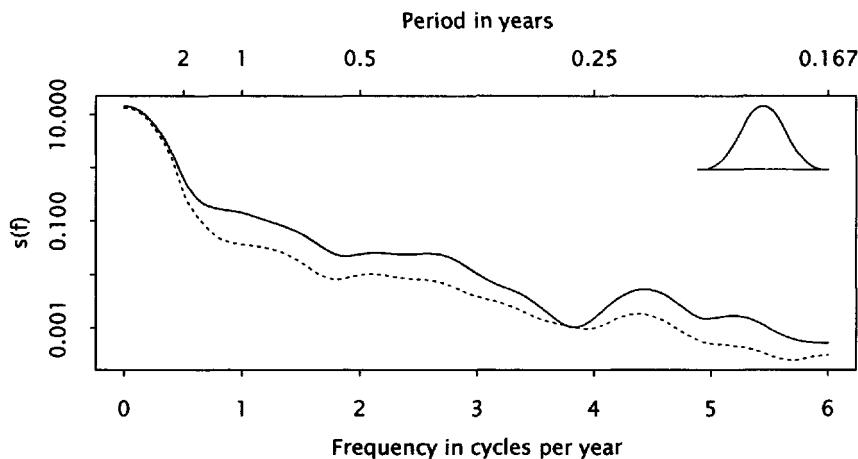


Fig. 10.4 Smoothed periodograms of one-year and ten-year interest rates, with spectral window inset. (Modified Daniell filter, $m = 16, 16, 16$.)

Like the cross periodogram $I_{x,y}(f)$, the cross spectrum $\hat{s}_{y,x}(f)$ is not, in general, real except at $f = 0$ and $1/2$ cycles per unit time. In place of the identity (10.1), $\hat{s}_{y,x}(f)$ satisfies the *inequality*

$$\left| \hat{s}_{x,y}(f) \right|^2 \leq \hat{s}_{x,x}(f) \hat{s}_{y,y}(f) \quad (10.4)$$

(see Exercise 10.4).

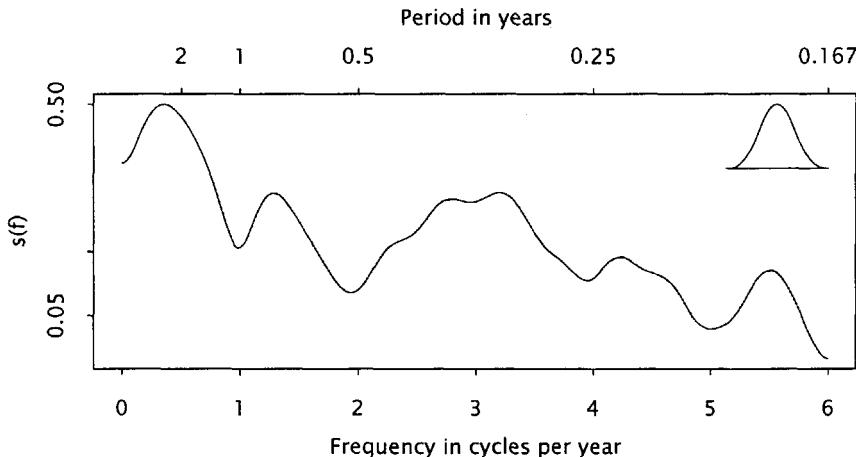


Fig. 10.5 Smoothed periodogram of monthly changes in industrial production, with spectral window inset. (Modified Daniell filter, $m = 16, 16, 16$.)

The real and imaginary parts of the cross spectrum are called the *cospectrum* and the *quadrature spectrum*, respectively. The cospectrum measures the extent to which the two series oscillate with the same phase (or with opposite sign, that is with a phase shift of a half cycle), while the quadrature spectrum measures the extent to which they tend to oscillate with a phase difference of a quarter cycle in either direction.

Typically the estimated cross spectrum $\hat{s}_{x,y}(f)$ is more usefully presented in terms of its magnitude $|\hat{s}_{x,y}(f)|$ and phase $\hat{\phi}_{x,y}(f)$, the latter defined by

$$\hat{s}_{x,y}(f) = |\hat{s}_{x,y}(f)| e^{2\pi i \hat{\phi}_{x,y}(f)},$$

than by the cospectrum and quadrature spectrum. The difference is essentially that between using polar coordinates and using Cartesian coordinates to represent a complex number. The phase, of course, is undefined where the cross spectrum vanishes. As in other statistical contexts, this means that it is also essentially meaningless where the cross spectrum is small.

Since the autospectra $\hat{s}_{x,x}(f)$ and $\hat{s}_{y,y}(f)$ are also available, the information in the magnitude $|\hat{s}_{x,y}(f)|$ is the same as that in the dimensionless ratio

$$\hat{r}_{x,y}^2(f) = \frac{|\hat{s}_{x,y}(f)|^2}{\hat{s}_{x,x}(f)\hat{s}_{y,y}(f)},$$

the estimated *squared coherency* of $\{x_t\}$ and $\{y_t\}$. The inequality (10.4)

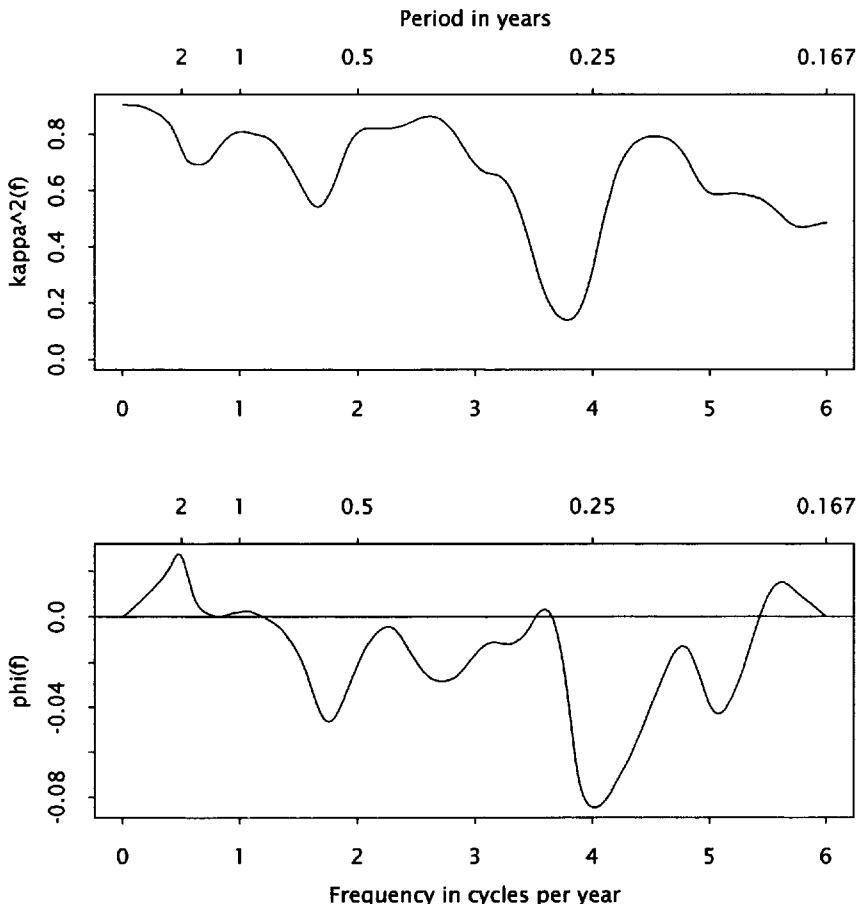


Fig. 10.6 Estimated squared coherency (upper panel) and phase (lower panel) of one-year and ten-year interest rates.

shows that the squared coherency lies between 0 and 1, with 0 corresponding to no dependence and 1 corresponding to exact linear dependence at frequency f cycles per unit time.

The estimated squared coherency and phase for the one-year and ten-year interest rates are shown in Figure 10.6. The estimated squared coherency is strong, as might be expected from the way the two interest rate series track each other in Figure 10.1. The estimated phase is close to zero, although negative at most frequencies. Phase spectra are difficult to interpret, except in the simple case of a linear phase, which indicates a lead/lag relationship (see Exercise 10.7). In the present case the phase

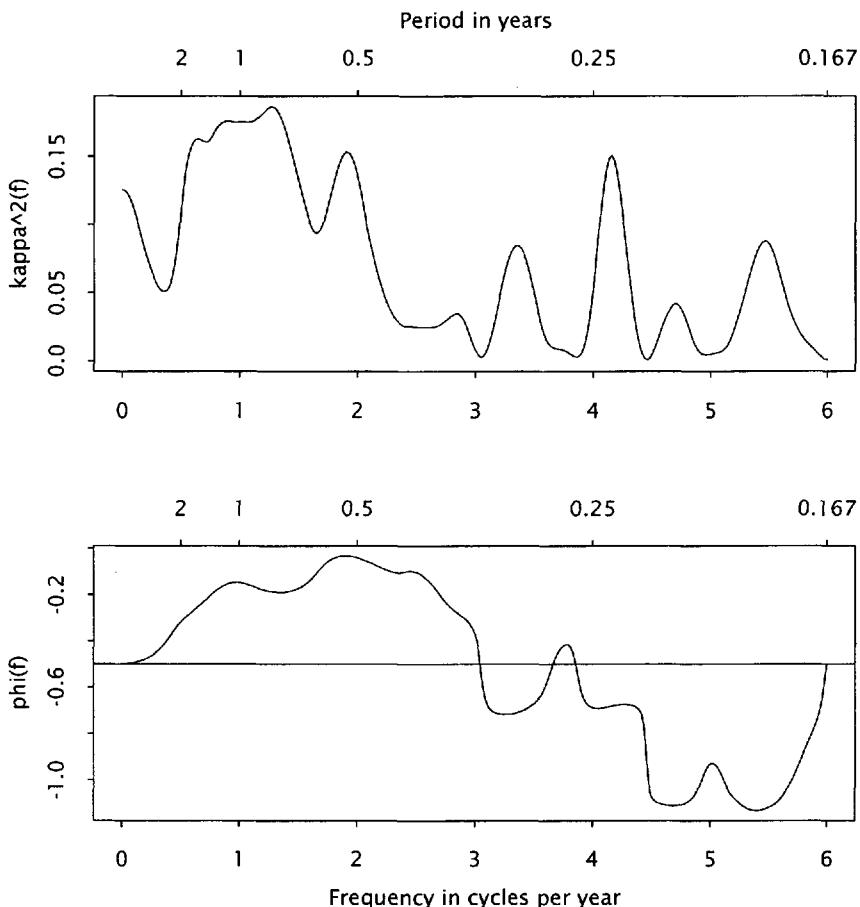


Fig. 10.7 Estimated squared coherency (upper panel) and phase (lower panel) of ten-year interest rates and monthly changes in industrial production.

does not show linear dependence on frequency; however, if it were approximated by a downward-sloping line through the origin and with the value -0.02 cycles at $f = 3$ cycles per year, the negative slope would suggest that the $\{y_t\}$ series, the ten-year interest rates, leads the $\{x_t\}$ series, one-year interest rates. The absolute value of the slope, $-0.02/3 = 0.007$ years or around 2–3 days, would suggest that the lead is of around this magnitude.

Figure 10.7 shows similar estimates of squared coherency and phase for ten-year interest rates and industrial production. The squared coherency is much lower in this case, and the phase is therefore less well determined.

The phase is $-1/2$ of a cycle at zero frequency, meaning that at the lowest frequencies, interest rates and changes in industrial production are negatively related. The positive slope in the part of the phase spectrum corresponding to the larger squared coherencies suggests that at these frequencies, say up to two cycles per year, interest rates lead changes in industrial production. Since the phase has progressed to around zero cycles at $f = 2$ cycles per year, the slope is around 0.25 years. That is, high interest rates in a given month are associated with a decline in industrial production around one quarter later, and conversely.

The interpretation of these graphs of estimated squared coherencies and phases is complicated by the need to distinguish frequencies where the squared coherency is large, where the phase is meaningful, from those where it is small, and the phase is correspondingly meaningless. The distinction is usually guided by statistical ideas: assessing the statistical significance of the squared coherency, and attaching confidence limits to the phase. Tools developed below in Section 10.5 embody these techniques.

Exercise 10.2 Rearranging the Cross Periodogram

Verify (10.2), and find explicit limits for the sum defining $c_{x,y,r}$. Note that they take different forms depending on the sign of r . Verify also that $c_{x,y,r} = c_{y,x,-r}$.

Exercise 10.3 Properties of Estimated Cross Spectra

Show that the estimated cross spectrum $\hat{s}_{x,y}(f)$ has the properties of the cross periodogram derived in Exercise 10.1.

Exercise 10.4 Coherency Inequality

Suppose that in the calculation of the estimated cross spectrum $\hat{s}_{x,y}(f)$ from equation (10.3), the weights $\{g_u\}$ are nonnegative. Show that in this case, $|\hat{s}_{x,y}(f)|^2 \leq \hat{s}_{x,x}(f)\hat{s}_{y,y}(f)$. (*Hint:* Use the Cauchy-Schwartz inequality.)

Note that this result remains true even if the smoothing procedures used to compute the estimated autospectra $\hat{s}_{x,x}(f)$ and $\hat{s}_{y,y}(f)$ differ from that used to compute $\hat{s}_{x,y}(f)$, provided the weights satisfy $g_{x,x,r} \geq 0$, $g_{y,y,r} \geq 0$, and $g_{x,y,r}^2 = g_{x,x,r}g_{y,y,r}$. Here $\{g_{x,x,r}\}$ are the spectral weights used to compute $\hat{s}_{x,x}(f)$, etc.

Exercise 10.5 Invariance of Estimated Squared Coherency

Suppose that $\{z_t\}$ is obtained from $\{y_t\}$ by linear filtering. Show that the estimated coherencies $\hat{s}_{x,y}(f)$ and $\hat{s}_{x,z}(f)$ are approximately the same. (*Hint:* Assume that the transfer function of the filter is smooth and, in particular, may be treated as constant over the span of the spectral window. Use the result of Exercise 7.3.)

Exercise 10.6 (Continuation) Coherency as a Measure of Dependence

With $\{y_t\}$ and $\{z_t\}$ as in the previous exercise, show that the estimated squared coherency of $\{y_t\}$ and $\{z_t\}$ is approximately 1 for all f . Note that $\{y_t\}$ and $\{z_t\}$ are completely linearly dependent, since each term in $\{z_t\}$ may be computed linearly from the terms of $\{y_t\}$.

10.3 THEORETICAL CROSS SPECTRUM

The theoretical cross spectrum of a pair of series may be defined by making the appropriate extension to the theory described in Chapter 9. The basic extension is to the definition of stationarity.

The stochastic processes $\{X_t\}$ and $\{Y_t\}$ are *jointly weakly stationary* if

- (i) $E(X_t)$ and $E(Y_t)$ are constant, and
- (ii) $E(X_t X_{t'})$, $E(X_t Y_{t'})$, and $E(Y_t Y_{t'})$ depend on t and t' only through $t - t'$.

This implies that both of the component processes $\{X_t\}$ and $\{Y_t\}$ separately are weakly stationary under the definition of Section 9.1. The extra strength of the present definition lies in the requirement that $E(X_t Y_{t'})$ depend only on $t - t'$, since this concerns the *joint* probabilistic structure of $\{X_t\}$ and $\{Y_t\}$.

For jointly weakly stationary processes $\{X_t\}$ and $\{Y_t\}$, the covariances

$$\gamma_{X,X,r} = \text{cov}(X_t, X_{t-r}),$$

$$\gamma_{X,Y,r} = \text{cov}(X_t, Y_{t-r}),$$

$$\gamma_{Y,Y,r} = \text{cov}(Y_t, Y_{t-r}),$$

do not depend on t . The first and third are the *autocovariances* of $\{X_t\}$ and $\{Y_t\}$, respectively, while the second, $\gamma_{X,Y,r}$, is the theoretical *cross covariance* of $\{X_t\}$ and $\{Y_t\}$ at lag r . Like the empirical cross covariances

defined in Section 10.2, the theoretical cross covariances are not generally symmetric, but satisfy $\gamma_{X,Y,r} = \gamma_{Y,X,-r}$. These sequences are *jointly non-negative definite*, in the sense that for any k and constants a_1, a_2, \dots, a_k and b_1, b_2, \dots, b_k ,

$$\sum_{r,s} (a_r \gamma_{X,X,r-s} a_s + a_r \gamma_{X,Y,-s} b_s + b_r \gamma_{Y,X,-s} a_s + b_r \gamma_{Y,Y,-s} b_s) \geq 0.$$

This implies that, in addition to the spectral distribution functions $S_{X,X}(f)$ and $S_{Y,Y}(f)$, which satisfy

$$\gamma_{X,X}(f) = \int_0^1 e^{2\pi i f r} dS_{X,X}(f)$$

and

$$\gamma_{Y,Y}(f) = \int_0^1 e^{2\pi i f r} dS_{Y,Y}(f),$$

there exists a function $S_{X,Y}(f)$ for which

$$\gamma_{X,Y}(f) = \int_0^1 e^{2\pi i f r} dS_{X,Y}(f).$$

From the properties of cross covariances it follows that the *cross spectral distribution function* $S_{X,Y}(f)$ satisfies

$$\begin{aligned} S_{X,Y}(f) - S_{X,Y}(f') &= \overline{S_{X,Y}(-f') - S_{X,Y}(-f)} \\ &= \overline{S_{Y,X}(f) - S_{Y,X}(f')}. \end{aligned}$$

The spectral distribution functions $S_{X,X}(f)$ and $S_{Y,Y}(f)$ are nondecreasing (see Section 9.1); this is equivalent to stating that for $f > f'$,

$$S_{X,X}(f) - S_{X,X}(f') \geq 0$$

and

$$S_{Y,Y}(f) - S_{Y,Y}(f') \geq 0.$$

The cross spectral distribution function satisfies the additional condition

$$\begin{aligned} \{S_{X,X}(f) - S_{X,X}(f')\} \{S_{Y,Y}(f) - S_{Y,Y}(f')\} \\ - |S_{X,Y}(f) - S_{X,Y}(f')|^2 \geq 0. \quad (10.5) \end{aligned}$$

These three conditions are equivalent to the matrix-valued function

$$\mathbf{S}(f) = \begin{bmatrix} S_{X,X}(f) & S_{X,Y}(f) \\ S_{X,Y}(f) & S_{Y,Y}(f) \end{bmatrix} = \begin{bmatrix} S_{X,X}(f) & S_{X,Y}(f) \\ S_{Y,X}(f) & S_{Y,Y}(f) \end{bmatrix}$$

having *nonnegative definite increments*.

If any of these functions is discontinuous at a frequency f , then a pure sinusoid of frequency f is present in one or both of the component processes. The other important case is where $s_{X,X}(f)$, $s_{X,Y}(f)$ and $s_{Y,Y}(f)$ all have derivatives, $s_{X,X}'(f)$, $s_{X,Y}'(f)$ and $s_{Y,Y}'(f)$, respectively. Condition (10.5) is then equivalent to

$$|s_{X,Y}(f)|^2 \leq s_{X,X}(f)s_{Y,Y}(f), \quad 0 \leq f < 1, \quad (10.6)$$

or, in other words, to stating that the matrix-valued function

$$\mathbf{s}(f) = \begin{bmatrix} s_{X,X}(f) & s_{X,Y}(f) \\ s_{X,Y}(f) & s_{Y,Y}(f) \end{bmatrix} = \begin{bmatrix} s_{X,X}(f) & s_{X,Y}(f) \\ s_{Y,X}(f) & s_{Y,Y}(f) \end{bmatrix}$$

is *nonnegative definite*. The functions $s_{X,X}(f)$ and $s_{Y,Y}(f)$ are the spectral density functions of $\{X_t\}$ and $\{Y_t\}$, respectively (see Section 9.1), while $s_{X,Y}(f)$ is the *cross spectral density function* of $\{X_t\}$ and $\{Y_t\}$, the theoretical counterpart of the empirical cross spectrum estimate $\hat{s}_{X,Y}(f)$ described in Section 10.2. Like its empirical counterpart, $s_{X,Y}(f)$ is not in general real except for $f = 0$ and $1/2$ cycles per unit time, but satisfies

$$s_{X,Y}(-f) = s_{Y,X}(f) = \overline{s_{X,Y}(f)}.$$

The counterparts of the empirical coherency and phase are $\kappa_{X,Y}^2(f)$ and $\phi_{X,Y}(f)$, where

$$\kappa_{X,Y}^2(f) = \frac{|s_{X,Y}(f)|^2}{s_{X,X}(f)s_{Y,Y}(f)}$$

and $\phi_{X,Y}(f)$ is defined by

$$\begin{aligned} s_{X,Y}(f) &= |s_{X,Y}(f)| e^{2\pi i \phi_{X,Y}(f)} \\ &= \kappa_{X,Y}(f) \sqrt{s_{X,X}(f)s_{Y,Y}(f)} e^{2\pi i \phi_{X,Y}(f)}. \end{aligned}$$

From (10.6) it follows that the theoretical squared coherency, like its empirical counterpart, satisfies $0 \leq \kappa_{X,Y}^2(f) \leq 1$. The statistical properties of the autoperiodogram and cross periodogram, described in the following section, depend on these various theoretical spectra.

Exercise 10.7 Operations on Jointly Weakly Stationary Processes

Suppose that $\{X_t\}$ and $\{Y_t\}$ are jointly weakly stationary, with spectral densities $s_{X,X}(f)$, $s_{X,Y}(f)$, and $s_{Y,Y}(f)$.

- (i) Let $Z_t = X_t + Y_t$. Show that $\{Y_t\}$ and $\{Z_t\}$ are jointly weakly stationary, and find the corresponding spectral density functions. (Hint:

Find the autocovariances and cross covariances, and use the Fourier series representation

$$s_{X,Y}(f) = \sum_r y_{X,Y,r} e^{-2\pi i f r}$$

for $s_{X,Y}(f)$.)

- (ii) Let $Z_t = \sum_u g_u X_{t-u}$. Show again that $\{Y_t\}$ and $\{Z_t\}$ are jointly weakly stationary, and find the spectral density functions.
- (iii) Suppose that the weights $\{g_u\}$ above are symmetric about $u = h$. Thus $\{Z_t\}$ is the result of applying a symmetric filter and a shift of h time units to $\{X_t\}$. Show that the squared coherency of $\{X_t\}$ and $\{Z_t\}$ is identically 1, and that the phase spectrum is $\phi_{X,Z}(f) = h f$.

10.4 DISTRIBUTION OF THE CROSS PERIODOGRAM

In this section we derive approximations to the sampling distribution of the autoperiodograms and cross periodograms of a pair of jointly weakly stationary processes $\{X_t\}$ and $\{Y_t\}$, in terms of their spectral densities $s_{X,X}(f)$, $s_{X,Y}(f)$, and $s_{Y,Y}(f)$. For rigorous derivations of the results of this section and the next, see Anderson (1994, Section 8.2), Brillinger (1981, Section 7.2), Hannan (1970, Section V.2), or Priestley (1981b, Chapter 9). We assume first that the series are Gaussian (i.e., that all their joint distributions are multivariate normal), and discuss later how this assumption may be relaxed.

The joint distribution of the transforms of two series is simplest if the two series are not cross correlated (i.e., if their cross covariances vanish at all lags, or equivalently if their cross spectrum vanishes at all frequencies). Under the Gaussian assumption, this implies independence; hence the two transforms are independent, and their joint distribution is the product of their marginal distributions, found in Section 9.4.

We show first that the joint distribution of the two transforms in the general, cross correlated, case may be found from this special case. Suppose then that the cross spectral density function $s_{X,Y}(f)$ is not identically zero, and let $G(f) = s_{X,Y}(f)/s_{X,X}(f)$. Then by an argument similar to that used in Section 9.4 we may find a filter, perhaps only in a limiting sense, with transfer function $G(f)$. Suppose that $\{Z'_t\}$ is the result of applying this filter to $\{X_t\}$, and let $Z_t = Y_t - Z'_t$. Then the cross spectrum of $\{X_t\}$ and $\{Z_t\}$ vanishes at all frequencies. The variable Z'_t is in fact the regression of Y_t on $\{X_u : -\infty < u < \infty\}$, and Z_t is the residual from

this regression. The autospectral density function of $\{Z_t\}$ is

$$s_{Z,Z}(f) = s_{Y,Y}(f) - \frac{|s_{X,Y}(f)|^2}{s_{X,X}(f)},$$

the *residual spectral density function* of $\{Y_t\}$ after regression on $\{X_t\}$.

Since $\{X_t\}$ and $\{Z_t\}$ are not cross correlated, the joint distribution of their transforms is the product of the marginal distributions. The real and imaginary parts have a joint four-dimensional distribution with all first moments zero, all covariances either exactly or approximately zero, and

$$\begin{aligned} \text{var}\{\Re d_X(f)\} &\approx \text{var}\{\Im d_X(f)\} \approx \frac{s_{X,X}(f)}{2n}, \\ \text{var}\{\Re d_Z(f)\} &\approx \text{var}\{\Im d_Z(f)\} \approx \frac{s_{Z,Z}(f)}{2n}. \end{aligned} \quad (10.7)$$

Here $\Re d_X(f)$ and $\Im d_X(f)$ denote the real and imaginary parts, respectively, of $d_X(f)$.

Now $d_Y(f)$, the transform of $\{Y_0, Y_1, \dots, Y_{n-1}\}$, must satisfy $d_Y(f) = d_Z(f) + d_{Z'}(f)$, and by the argument of Section 7.2 $d_{Z'}(f) \approx G(f)d_X(f)$, whence

$$d_Y(f) \approx d_Z(f) + G(f)d_X(f). \quad (10.8)$$

As linear combinations of Gaussian random variables, the real and imaginary parts of $d_X(f)$ and $d_Y(f)$ have a joint four-dimensional Gaussian distribution, with moments that follow from (10.7) and (10.8). First moments are all zero, and the second moments are:

$$\begin{aligned} \text{var}\{\Re d_X(f)\} &\approx \text{var}\{\Im d_X(f)\} \approx \frac{s_{X,X}(f)}{2n}, \\ \text{var}\{\Re d_Y(f)\} &\approx \text{var}\{\Im d_Y(f)\} \approx \frac{s_{Y,Y}(f)}{2n}, \\ \text{cov}\{\Re d_X(f), \Im d_X(f)\} &\approx \text{cov}\{\Re d_Y(f), \Im d_Y(f)\} \approx 0, \\ \text{cov}\{\Re d_X(f), \Re d_Y(f)\} &\approx \text{cov}\{\Im d_X(f), \Im d_Y(f)\} \approx \frac{c_{X,Y}(f)}{2n}, \\ \text{cov}\{\Re d_X(f), \Im d_Y(f)\} &\approx -\text{cov}\{\Im d_X(f), \Re d_Y(f)\} \approx \frac{-q_{X,Y}(f)}{2n}. \end{aligned} \quad (10.9)$$

Here $c_{X,Y}(f)$ and $q_{X,Y}(f)$ are the cospectrum $\Re s_{X,Y}(f)$ and quadrature spectrum $\Im s_{X,Y}(f)$, respectively.

The results of Section 9.4 show that

$$E\{I_{X,X}(f)\} \approx s_{X,X}(f), \quad \text{var}\{I_{X,X}(f)\} \approx s_{X,X}(f)^2,$$

and

$$E \{I_{Y,Y}(f)\} \approx s_{Y,Y}(f), \quad \text{var } \{I_{Y,Y}(f)\} \approx s_{Y,Y}(f)^2.$$

From (10.9) it follows additionally that

$$\begin{aligned} E \{I_{X,Y}(f)\} &\approx s_{X,Y}(f), \\ \text{var } \{\Re I_{X,Y}(f)\} &\approx \frac{1}{2} \left\{ s_{X,X}(f)s_{Y,Y}(f) + c_{X,Y}(f)^2 - q_{X,Y}(f)^2 \right\}, \\ \text{var } \{\Im I_{X,Y}(f)\} &\approx \frac{1}{2} \left\{ s_{X,X}(f)s_{Y,Y}(f) - c_{X,Y}(f)^2 + q_{X,Y}(f)^2 \right\}, \\ \text{cov } \{\Re I_{X,Y}(f), \Im I_{X,Y}(f)\} &\approx c_{X,Y}(f)q_{X,Y}(f), \\ \text{cov } \{\Re I_{X,Y}(f), I_{X,X}(f)\} &\approx c_{X,Y}(f)s_{X,X}(f), \\ \text{cov } \{\Im I_{X,Y}(f), I_{X,X}(f)\} &\approx q_{X,Y}(f)s_{X,X}(f), \\ \text{cov } \{\Re I_{X,Y}(f), I_{Y,Y}(f)\} &\approx c_{X,Y}(f)s_{Y,Y}(f), \\ \text{cov } \{\Im I_{X,Y}(f), I_{Y,Y}(f)\} &\approx q_{X,Y}(f)s_{Y,Y}(f), \\ \text{cov } \{I_{X,X}(f), I_{Y,Y}(f)\} &\approx |s_{X,Y}(f)|^2 = c_{X,Y}(f)^2 + q_{X,Y}(f)^2. \end{aligned} \tag{10.10}$$

Equation (10.10) gives the means, variances, and covariances of the various periodograms at a single frequency f . As in Section 9.4, it may be shown that these quantities evaluated at a Fourier frequency are approximately independent of the same quantities evaluated at any other Fourier frequency. This is all the information needed to derive approximate sampling distributions of the smoothed periodograms described in Section 10.2.

Since the discrete Fourier transform is a linear function of the data, and the periodograms, auto- and cross, are quadratic functions of the data, the moment formulas given above depend on the joint distributions of the data only through moments up to the fourth. That is, they are valid for any non-Gaussian series with the same moments, up to the fourth, as a Gaussian process. In other cases there would be additional contributions from the fourth-order cumulants.

An approximation to the full joint distribution of the periodograms may also be found, but by contrast this does depend on the Gaussian distribution of the Fourier transforms. In this approximation, the marginal distributions of $\Re I_{X,Z}(f)$ and $\Im I_{X,Z}(f)$, the real and imaginary parts of the cross periodogram of the *independent* series $\{X_t\}$ and $\{Z_t\}$, are both double exponential (or Laplacian). The real and imaginary parts are not independent, however, but instead have a circularly symmetric distribution. The approximate joint distribution of the real and imaginary parts

of the cross periodogram of the *dependent* series $\{X_t\}$ and $\{Y_t\}$ is not simple if $s_{X,Y}(f) \neq 0$.

The approximations discussed in this section depend only on the joint distribution of the discrete Fourier transforms $d_X(f)$ and $d_Y(f)$ being approximately Gaussian; normality of the series $\{X_t\}$ and $\{Y_t\}$ is only the most convenient assumption that ensures this. The assumption may be weakened in various ways—for instance, by assuming instead that all the terms in the white noise series that generate $\{X_t\}$ and $\{Z_t\}$ are independent of each other; see Section 9.4 for further discussion.

Exercise 10.8 Regression of Y_t on $\{X_t\}$

The *regression* of Y_t on $\{X_t : -\infty < t < \infty\}$ is the linear combination of the variables $\{X_t\}$ that is closest to Y_t in the sense of mean squared error.

- (i) Show that the weights $\{g_u\}$ that minimize

$$E \left(Y_t - \sum_u g_u X_{t-u} \right)^2$$

satisfy

$$G(f) = \sum_u g_u e^{-2\pi i f u} = \frac{\overline{s_{X,Y}(f)}}{s_{X,X}(f)}.$$

- (ii) Let $Z_t = Y_t - \sum_u g_u X_{t-u}$, where the weights $\{g_u\}$ are these optimal weights. Use the results of Exercise 10.7 to find the spectral densities of $\{X_t\}$ and $\{Z_t\}$. Note that

$$s_{Z,Z}(f) = s_{Y,Y}(f) - \frac{|s_{X,Y}(f)|^2}{s_{X,X}(f)}$$

is nonnegative by the coherency inequality (10.6).

Exercise 10.9 Moments of Transforms

Verify that the moments of $d_X(f)$ and $d_Y(f)$ are as stated in (10.9). (*Hint:* Use (10.7) and (10.8) and the definition $G(f) = \overline{s_{X,Y}(f)}/s_{X,X}(f) = \{c_{X,Y}(f) - iq_{X,Y}(f)\}/s_{X,X}(f)$.)

Exercise 10.10 Moments of Periodograms

Verify that the moments of the periodograms $I_{X,X}(f)$, $I_{X,Y}(f)$, and $I_{Y,Y}(f)$ are as stated in (10.10). (*Hint:* If A , B , C , and D have a joint Gaussian dis-

tribution with zero means, then $E(ABCD) = E(AB)E(CD)+E(AC)E(BD)+E(AD)E(BC).$

Exercise 10.11 Distribution of the Cross Periodogram

Suppose that A , B , C , and D are independent, each with the standard normal distribution.

- (i) Show that $AB + CD$ has the double exponential distribution. (*Hint:* $AB = 1/4\{(A+B)^2 - (A-B)^2\}.$)
- (ii) Show that $P = AB + CD$ and $Q = AD - BC$ have a circularly symmetric joint distribution. That is, if $c^2 + s^2 = 1$, then the joint distribution of $cP + sQ$ and $cQ - sP$ is the same as that of P and Q .

10.5 DISTRIBUTION OF ESTIMATED CROSS SPECTRA

As with autoperiodograms, cross periodogram ordinates calculated at different Fourier frequencies are uncorrelated. Furthermore, a cross periodogram ordinate at one Fourier frequency is uncorrelated with either autoperiodogram ordinate at a different Fourier frequency, and the same is true for the two autoperiodograms. This may be summarized by saying that any periodogram ordinate (cross or auto-) at a given Fourier frequency is uncorrelated with any periodogram ordinate at any other frequency.

It follows, therefore, that if a periodogram is smoothed using weights $\{g_u\}$, and if the theoretical spectra are approximately constant over the span of these weights, then the resulting spectrum estimates $\hat{s}_{X,X}(f)$, $\hat{s}_{X,Y}(f)$, and $\hat{s}_{Y,Y}(f)$ have the same variances and covariances as $I_{X,X}(f)$, $I_{X,Y}(f)$, and $I_{Y,Y}(f)$ (given at the end of Section 10.4) but multiplied by the factor $\sum_u g_u^2$. If the data were tapered before the periodogram was calculated, this factor is multiplied by the correction factor U_4/U_2^2 derived in Section 9.5. Lastly, if the periodogram was computed on a finer grid than the Fourier frequencies, we must also multiply by the factor $n'/n > 1$, the ratio of the spacings. In the most general case, the variances and covariances of $\hat{s}_{X,X}(f)$, $\hat{s}_{X,Y}(f)$, and $\hat{s}_{Y,Y}(f)$ are the same as those of $I_{X,X}(f)$, $I_{X,Y}(f)$, and $I_{Y,Y}(f)$ in (10.10), but multiplied by the factor

$$g^2 = \frac{n'}{n} \frac{U_4}{U_2^2} \sum_u g_u^2.$$

The phase $\hat{\phi}_{X,Y}(f)$ satisfies

$$\tan \{2\pi \hat{\phi}_{X,Y}(f)\} = \frac{\Im \hat{s}_{X,Y}(f)}{\Re \hat{s}_{X,Y}(f)},$$

and is calculated using the branch of the arctangent that gives the correct sign to, say, $\Re \hat{s}_{X,Y}(f)$, as described in Section 2.2. Approximations to the moments of $\hat{\phi}_{X,Y}(f)$ may be obtained only by assuming that $\hat{s}_{X,Y}(f)$ has been smoothed enough to ensure that it is close to $s_{X,Y}(f)$. If $\hat{s}_{X,Y}(f) = s_{X,Y}(f) + a + ib$ for small a and b , we may expand $\hat{\phi}_{X,Y}(f)$ in a Taylor series in a and b . The zero- and first-order terms are

$$\hat{\phi}_{X,Y}(f) \approx \phi_{X,Y}(f) - \frac{aq_{X,Y}(f) - bc_{X,Y}(f)}{2\pi |s_{X,Y}(f)|^2},$$

and thus

$$E \{\hat{\phi}_{X,Y}(f)\} \approx \phi_{X,Y}(f)$$

and

$$\text{var} \{\hat{\phi}_{X,Y}(f)\} \approx \frac{1}{(2\pi)^2} \frac{g^2}{2} \left\{ \frac{1}{\kappa_{X,Y}^2(f)} - 1 \right\},$$

where $\phi_{X,Y}(f)$ and $\kappa_{X,Y}^2(f)$, defined in Section 10.3, are the theoretical phase and squared coherency, respectively. Notice that the approximate variance is large if the coherency is small. The approximation is, in fact, valid only when it is small, that is, when $\kappa_{X,Y}^2(f)$ is not small.

The estimated phase is approximately normally distributed when this approximation is good, and thus, for instance, an approximate 95% confidence interval for $\phi_{X,Y}(f)$ is

$$\hat{\phi}_{X,Y}(f) \pm \frac{1.96}{2\pi} \sqrt{\frac{g^2}{2} \left\{ \frac{1}{\hat{\kappa}_{X,Y}^2(f)} - 1 \right\}}, \quad (10.11)$$

where the theoretical coherency $\kappa_{X,Y}^2(f)$ has been replaced by its estimate $\hat{\kappa}_{X,Y}^2(f)$. The arguments used by Hannan (1970, Section V.2) suggest that

$$\hat{\phi}_{X,Y}(f) \pm \frac{1}{2\pi} \arcsin \left[t_v(.05) \sqrt{\frac{g^2}{2(1-g^2)} \left\{ \frac{1}{\hat{\kappa}_{X,Y}^2(f)} - 1 \right\}} \right] \quad (10.12)$$

would be a better approximation, where $t_v(\alpha)$ is the $100\alpha\%$ point of the t -distribution with v degrees of freedom, and $v = 2/g^2 - 2$.

The approximate variance of $\hat{\kappa}_{X,Y}^2(f)$ and the covariance of $\hat{\phi}_{X,Y}(f)$ and $\hat{\kappa}_{X,Y}^2(f)$ may be found similarly. If $\hat{s}_{X,X}(f) = s_{X,X}(f) + c$ and $\hat{s}_{Y,Y}(f) = s_{Y,Y}(f) + d$, then, provided $\kappa_{X,Y}^2(f) \neq 0$,

$$\begin{aligned}\hat{\kappa}_{X,Y}(f) &= \frac{|\hat{s}_{X,Y}(f)|}{\sqrt{\hat{s}_{X,X}(f)\hat{s}_{Y,Y}(f)}} \\ &\approx \kappa_{X,Y}(f) + \frac{ac_{X,Y}(f) + b\hat{q}_{X,Y}(f)}{\kappa_{X,Y}(f)s_{X,X}(f)s_{Y,Y}(f)} \\ &\quad - \frac{1}{2}\kappa_{X,Y}(f) \left\{ \frac{c}{s_{X,X}(f)} + \frac{d}{s_{Y,Y}(f)} \right\},\end{aligned}\tag{10.13}$$

and hence

$$\begin{aligned}E\{\hat{\kappa}_{X,Y}(f)\} &\approx \kappa_{X,Y}(f), \\ \text{var}\{\hat{\kappa}_{X,Y}(f)\} &\approx \frac{g^2}{2} \left\{ 1 - \kappa_{X,Y}^2(f) \right\}^2,\end{aligned}$$

and

$$\text{cov}\{\hat{\phi}_{X,Y}(f), \hat{\kappa}_{X,Y}(f)\} \approx 0.$$

The approximate bias in $\hat{\kappa}_{X,Y}(f)$ may be found by taking the expectation of the next term in the Taylor series expansion. The result is

$$\frac{g^2}{4} \frac{\left\{ 1 - \kappa_{X,Y}^2(f) \right\}^2}{\kappa_{X,Y}(f)},$$

which is always positive. Recall that, if the coherency is computed from the periodograms without smoothing, its value is identically 1. We see that some positive bias remains even after smoothing. There is another source of bias when the spectra are not effectively constant over the bandwidth of the spectral window. Unlike the present source, this increases with the bandwidth. Thus a trade-off is called for in controlling these two sources of bias.

The variance of $\hat{\kappa}_{X,Y}(f)$ depends on $\kappa_{X,Y}(f)$ in the same way as that of a correlation coefficient depends on the theoretical correlation. Thus the arctanh transformation makes the variance constant to this order of approximation (see Jenkins and Watts, 1968, p. 379, or Brillinger, 1981, Section 8.9). In fact,

$$E\{\text{arctanh } \hat{\kappa}_{X,Y}(f)\} \approx \text{arctanh } \kappa_{X,Y}(f)$$

and

$$\text{var}\{\text{arctanh } \hat{\kappa}_{X,Y}(f)\} \approx \frac{g^2}{2}.$$

Thus an approximate 95% confidence interval for $\kappa_{X,Y}(f)$ may be found as $\tanh z_1 \leq \kappa_{X,Y}(f) \leq \tanh z_2$, where z_1 and z_2 are

$$\operatorname{arctanh} \hat{\kappa}_{X,Y}(f) \pm 1.96 \sqrt{\frac{g^2}{2}} = \frac{1}{2} \ln \frac{1 + \hat{\kappa}_{X,Y}(f)}{1 - \hat{\kappa}_{X,Y}(f)} \pm 1.96 \sqrt{\frac{g^2}{2}}. \quad (10.14)$$

Expansion (10.13) is invalid for $\kappa_{X,Y}(f) = 0$, and therefore gives poor approximations for small $\kappa_{X,Y}(f)$. The distribution of $\hat{\kappa}_{X,Y}^2(f)$ when $\kappa_{X,Y}(f) = 0$ is given by

$$\operatorname{pr} \left\{ \hat{\kappa}_{X,Y}^2(f) \leq \sigma(p)^2 \right\} \approx p,$$

where

$$\sigma(p)^2 = 1 - (1 - p)^{g^2/(1-g^2)}$$

(see, e.g., Brillinger, 1981, p. 317). Thus, for example, the 95% point of the distribution is

$$\sigma(0.95)^2 = 1 - 20^{-g^2/(1-g^2)}. \quad (10.15)$$

Observed values of $\hat{\kappa}_{X,Y}^2(f)$ less than $\sigma(0.95)^2$ should therefore be regarded as not significantly different from 0, and the confidence interval (10.14) should be used only if $\hat{\kappa}_{X,Y}^2(f)$ exceeds this value. Further discussion of the construction of confidence intervals for the estimated phase and coherency may be found in Hannan (1970, Section V.2) and Brillinger (1981, Sections 6.9, 8.9).

Examples

Figure 10.8 shows the estimated squared coherency of one-year and ten-year interest rates as in Figure 10.6 (p. 208), but with level lines showing the significance level of the observed coherency under the null hypothesis that the coherency is zero. Note that all probability statements are made on a frequency-by-frequency basis. A test of the hypothesis that coherency is zero at *all* frequencies would be quite different.

The observed coherency is evidently strong. Figure 10.9 is another version of the graph, with 95% confidence intervals computed from (10.14) added. These intervals show that the coherency is estimated relatively precisely.

Figure 10.10 shows the estimated phase of one-year and ten-year interest rates with 95% confidence intervals computed from (10.12). The intervals are relatively tight, with a length of less than 0.05 cycles at frequencies where the coherency is strongest, and mostly including the zero

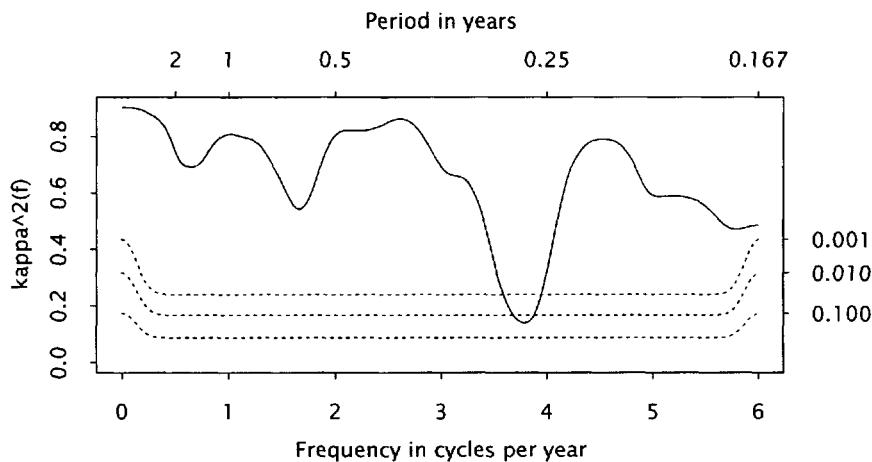


Fig. 10.8 Estimated squared coherency of one-year and ten-year interest rates, with cutoff levels for 10%, 1%, and 0.1% significance levels under the null hypothesis of incoherency.

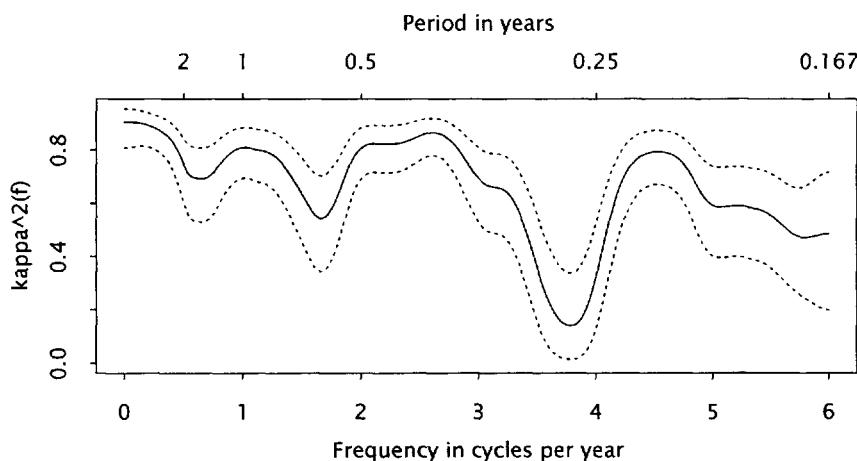


Fig. 10.9 Estimated squared coherency of one-year and ten-year interest rates, with approximate 95% confidence interval.

line. If the true phase were zero at all frequencies, the interval would be expected to include zero at around 95% of frequencies. It appears from the graph that the percentage of frequencies where the interval includes zero is somewhat less than 95%, suggesting that the lead of ten-year rates over one-year rates is real. However, since the indicated lead is of only

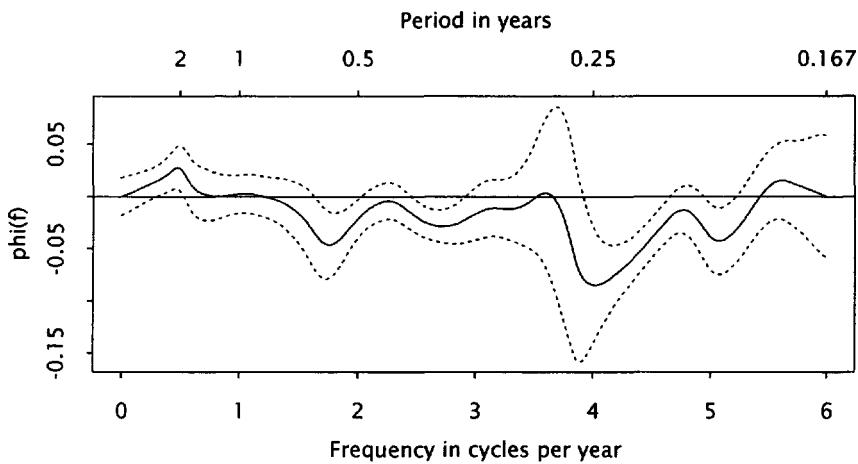


Fig. 10.10 Estimated phase of one-year and ten-year interest rates, with approximate 95% confidence interval.

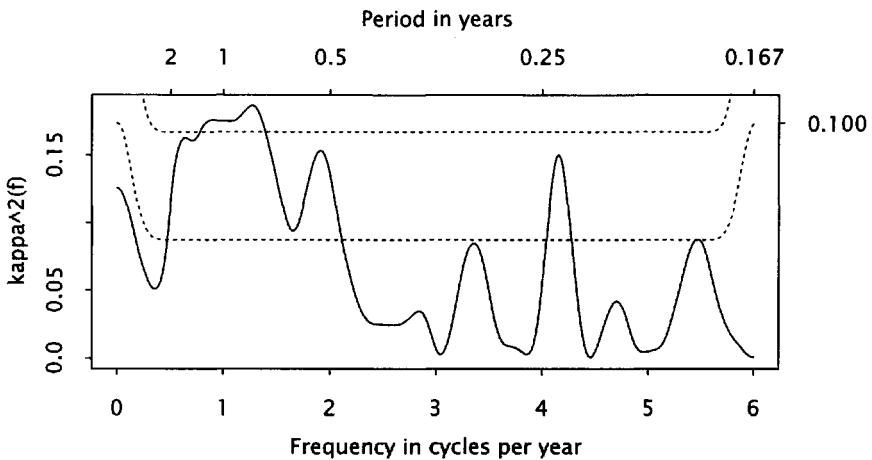


Fig. 10.11 Estimated squared coherency of ten-year interest rates and monthly changes in industrial production, with cutoff levels for 10% and 1% significance levels under the null hypothesis of incoherency.

a few days, these monthly data are less useful for exploring it than daily data would be.

Figures 10.11 and 10.12 display the coherency of ten-year interest rates and industrial production from Figure 10.7 (p. 209), similarly enhanced by the addition of significance levels and 95% confidence intervals, re-

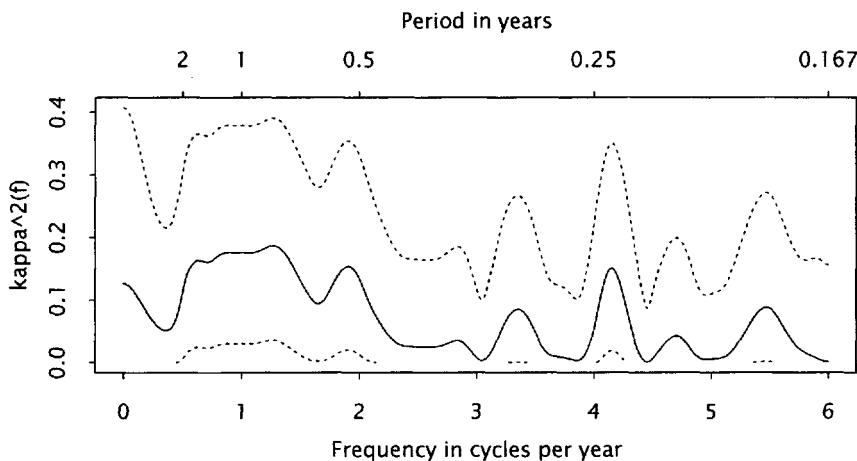


Fig. 10.12 Estimated squared coherency of ten-year interest rates and monthly changes in industrial production, with approximate 95% confidence interval.

spectively. The coherency shown in these figures is much lower than that between the interest rates. However, it appears that there is some real coherency, in that the observed coherency lies above the 10% and 1% points of the null distribution for somewhat more than 10% and 1% of frequencies, respectively.

Figure 10.13 shows the corresponding phase with 95% confidence intervals, again obtained from (10.12). The intervals are considerably wider than those in Figure 10.10 because of the weaker coherency, and are, in fact, undefined at some frequencies, where the argument of the inverse sine function is outside the interval $[-1, 1]$. However, where the coherency is strongest, at frequencies up to around two cycles per year, the confidence intervals are consistent with the linear phase behavior discussed in Section 10.2.

Comparison of Spectra

It is sometimes of interest to compare the autospectra of a number of series. For instance, the two interest rate spectra shown in Figure 10.4 (p. 206) are commensurate, and a frequency-by-frequency comparison of the power in each could be illuminating. Because the series are not independent, the autospectra are all positively correlated, and this fact has to

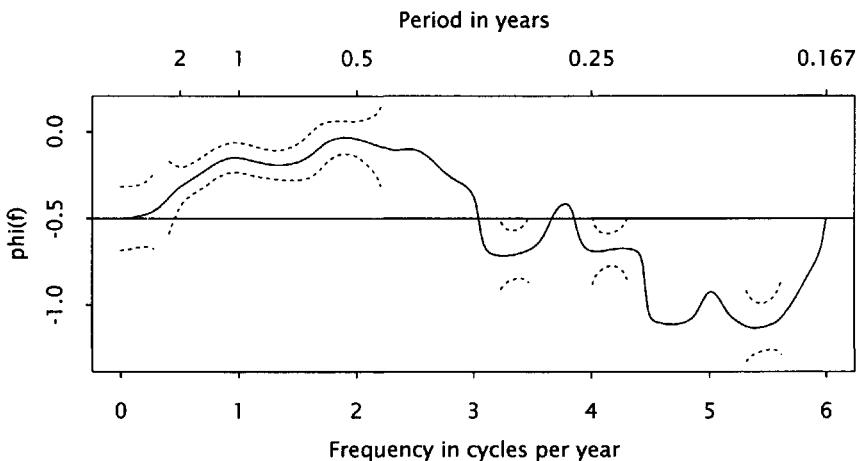


Fig. 10.13 Estimated phase of ten-year interest rates and monthly changes in industrial production, with approximate 95% confidence interval.

be taken into account in setting up confidence intervals for differences.³ The resulting intervals decrease in width as the coherency increases. Thus very precise comparisons may be made between the autospectra of two highly coherent series, even though the autospectrum of each may not be estimated very precisely.

The variance of

$$\ln \frac{\hat{S}_{X,X}(f)}{\hat{S}_{Y,Y}(f)}$$

is approximately $2g^2 \{1 - \kappa_{X,Y}^2(f)\}$; thus, for example, an approximate 95% confidence interval for

$$\ln \frac{S_{X,X}(f)}{S_{Y,Y}(f)}$$

is

$$\frac{\hat{S}_{X,X}(f)}{\hat{S}_{Y,Y}(f)} \pm 1.96 \sqrt{2g^2 \{1 - \hat{\kappa}_{X,Y}^2(f)\}},$$

where $\kappa_{X,Y}(f)$ has been replaced by its estimate $\hat{\kappa}_{X,Y}(f)$. Figure 10.14 shows the ratio of the estimated spectrum of ten-year interest rates to that of one-year interest rates, with the corresponding approximate 95% confidence interval. The varying width of the interval is apparent, as is the conclusion that the ten-year interest rates have significantly higher

³These differences are most naturally calculated as differences of logarithms of spectra or equivalently as logarithms of ratios of spectra.

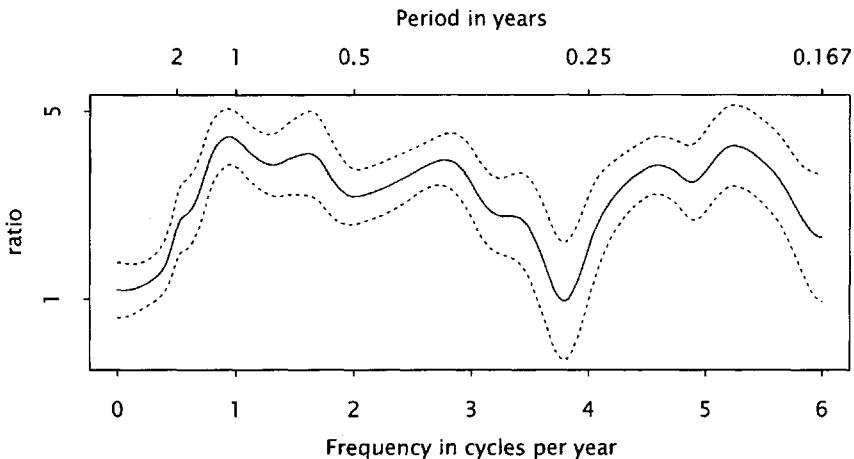


Fig. 10.14 Ratio of smoothed periodograms of one-year and ten-year interest rates, with 95% confidence interval.

spectral power at most frequencies, though notably not at the lowest frequencies, where both series have highest power.

10.6 ALIGNMENT

It was stated in Section 10.1 that smoothing a cross periodogram is essentially no more difficult than smoothing an autoperiodogram, except for the question of *alignment* of the series. In this section we illustrate this problem and discuss ways to avoid it.

Suppose that a pair of series $\{x_t\}$ and $\{y_t\}$ is observed, and that in fact $\{y_t\}$ consists of $\{x_t\}$ shifted by h ; that is, $y_t = x_{t-h}$. We say that $\{y_t\}$ lags $\{x_t\}$ and $\{x_t\}$ leads $\{y_t\}$ by h time units. Then the transforms of $\{x_0, x_1, \dots, x_{n-1}\}$ and $\{y_0, y_1, \dots, y_{n-1}\}$ satisfy

$$d_y(f) \approx d_x(f)e^{-2\pi ifh},$$

this being an approximation rather than an identity because of h terms at each end of the corresponding sums that are not equal. Thus

$$I_{x,y}(f) = nd_x(f)\overline{d_y(f)} \approx I_{x,x}(f)e^{2\pi ifh}.$$

Similarly the cross spectrum of a pair of jointly weakly stationary processes $\{X_t\}$ and $\{Y_t\}$ satisfying the same lag- h relationship satisfies

$$s_{X,Y}(f) = s_{X,X}(f)e^{2\pi ifh}.$$

Then the spectrum estimates should satisfy

$$\hat{s}_{x,y}(f) \approx \hat{s}_{x,x}(f)e^{2\pi ifh}.$$

However, if h is large enough this cannot be the case.

To see this, consider a lag-weights spectrum estimate with truncation point m ,

$$\hat{s}_{x,y}(f) = \sum_{|r| < m} w_r c_{x,y,r} e^{-2\pi ifr}.$$

Now

$$\begin{aligned} c_{x,y,r} &= \frac{1}{n} \sum x_t y_{t-r} \\ &\approx \frac{1}{n} \sum x_t x_{t-h-r} \\ &= c_{x,x,r+h}, \end{aligned}$$

whence

$$\hat{s}_{x,y}(f) \approx \sum_{|r| < m} w_r c_{x,x,r+h} e^{-2\pi ifr}. \quad (10.16)$$

Now $\{c_{x,x,r} : r = 0, \pm 1, \dots, \pm(n-1)\}$ is a symmetric sequence with its maximum at $r = 0$. Thus, if $|h| \geq m$, the largest term in the sequence $\{c_{x,x,r}\}$ is omitted from the sum (10.16), and therefore $\hat{s}_{x,y}(f)$ will tend to be smaller than $\hat{s}_{x,x}(f)e^{2\pi ifh}$. If h is large enough, all of the large autocovariances are omitted from (10.16), and thus

$$|\hat{s}_{x,y}(f)| \ll \hat{s}_{x,x}(f) = \hat{s}_{y,y}(f).$$

Hence the estimated squared coherency between $\{x_t\}$ and $\{y_t\}$,

$$\hat{R}_{x,y}^2(f) = \frac{|\hat{s}_{x,y}(f)|^2}{\hat{s}_{x,x}(f)\hat{s}_{y,y}(f)},$$

is close to zero. We would therefore conclude that the two series are unrelated, whereas they are in fact completely dependent.

The phenomenon does not appear only for lag-weights estimates. Any estimate may also be written as a spectral average (see Section 8.5), with an appropriate window. Now $I_{x,x}(f)$ is real and nonnegative and hence averages to a positive value. However,

$$I_{x,y}(f) \approx I_{x,x}(f)e^{2\pi ifh}$$

and if the complex factor $e^{2\pi ifh}$ goes through one or more complete cycles within the bandwidth of the window, the values of $I_{x,y}(f)$ tend to cancel out, and thus their average $\hat{s}_{x,y}(f)$ again may satisfy

$$|\hat{s}_{x,y}(f)| \ll \hat{s}_{x,x}(f).$$

This happens if the bandwidth is at least $1/h$, which is equivalent to $h \geq m$ if we define the bandwidth of a lag-weights estimate to be $1/m$.

The simplest way out of this predicament is to *realign* the data. If we define $\{z_t\}$ by $z_t = y_{t+h}$, there are no alignment problems for $\{x_t, z_t\}$. We have to know h to carry out this realignment, at least to within an error that is small compared with m . The lag of the largest entry in $\{c_{x,y,r}\}$ is the most obvious candidate. A similar correction could be implemented by analyzing $I_{x,y}(f)e^{-2\pi ifh}$.

In this example, realignment of the two series removes the problem. However, in general, it is not possible to find a single realignment that removes the problem at all frequencies. Consider the following example. The series $\{x_t\}$ is the sum of two components $\{u_t\}$ and $\{v_t\}$, where the first has more power at low frequencies, and the second at high frequencies. Now, if $y_t = u_t + v_{t-h}$, then $I_{x,y}(f) \approx I_{u,u}(f)$ at low frequencies, and $I_{x,y}(f) \approx I_{v,v}(f)e^{2\pi ifh}$ at high frequencies. Thus, whereas no realignment is called for at low frequencies, a realignment of h time units is needed at high frequencies. More generally, there could be several frequency bands, each requiring its own realignment.

Another way of describing the problem is as follows: Write

$$s_{X,Y}(f) = |s_{X,Y}(f)| e^{2\pi i \phi_{X,Y}(f)}$$

and suppose that both $|s_{X,Y}(f)|$ and $\phi'_{X,Y}(f)$ are smooth, but the latter is large. Then for f' close to f we have

$$s_{X,Y}(f') \approx |s_{X,Y}(f)| e^{2\pi i \{\phi_{X,Y}(f) + (f' - f)\phi'_{X,Y}(f)\}}.$$

Hence

$$\begin{aligned} E\{\hat{s}_{X,Y}(f)\} &\approx \sum_u g_u s_{X,Y}(f - f_u) \\ &\approx |s_{X,Y}(f)| e^{2\pi i \phi_{X,Y}(f)} \sum_u g_u e^{2\pi i f_u \phi'_{X,Y}(f)} \\ &= s_{X,Y}(f) \sum_u g_u e^{2\pi i \frac{u \phi'_{X,Y}(f)}{n}}. \end{aligned}$$

If $\phi'_{X,Y}(f)$ is an integer, then by equation (8.17) (p. 153), the sum is, in fact, the “lag weight” of this lag. If $\phi'_{X,Y}(f)$ is not an integer, it may be

regarded as an interpolated value between adjacent lag weights. Thus the problems we have described are associated with rapidly varying phase, in that the sum may be small or even vanish if $\phi'_{X,Y}(f)$ is large enough.

One way to remove the problem therefore would be to modify the periodogram so that its phase cannot vary rapidly. Since

$$s_{X,Y}(f) = |s_{X,Y}(f)| e^{2\pi i \phi_{X,Y}(f)}$$

it follows that $I_{X,Y}(f)e^{-2\pi i \phi_{X,Y}(f)}$ is an estimate of the real, positive quantity $|s_{X,Y}(f)|$. Thus there are no more problems in smoothing this function than in smoothing the autoperiodogram. One procedure would be to follow these steps:

- Obtain an initial estimate of the phase spectrum $\phi_{X,Y}(f)$;
- Correct $I_{X,Y}(f)$ for phase using this estimate;
- Smooth the result to obtain an estimate of $|s_{X,Y}(f)|$; and
- Combine this with the phase information to obtain an estimate of $s_{X,Y}(f)$.

The resulting estimate of $s_{X,Y}(f)$ could then be used to obtain a new estimate of the phase spectrum, thus setting up an iteration. A few steps, say two or three, should be sufficient to achieve reasonable convergence.

The problem of alignment is also discussed by Hannan (1970, Section V.7) and Brillinger (1981, p. 266). Hannan and Thompson (1973) suggest that estimation of the phase spectrum $\phi_{X,Y}(f)$ be replaced by estimation of the *group delay* $\phi'_{X,Y}(f)$, and discuss the problems involved.

Example

Figures 10.7 (p. 209) and 10.13 (p. 225) showed that there is around a one quarter misalignment between ten-year interest rates and monthly changes in industrial production. Figure 10.15 displays the coherency between these series when the interest rates are lagged (*realigned*) by one quarter. Comparison with Figure 10.11 (p. 223) shows that this relatively small realignment changes the estimated squared coherency only slightly, and results in no change in the general conclusion of a weak association limited to low frequencies.

Figure 10.16 displays the corresponding estimated phase spectrum. Comparison with Figure 10.13 (p. 225) confirms that the realignment has essentially eliminated the slope in the phase spectrum over the frequency band where the coherency is largest.

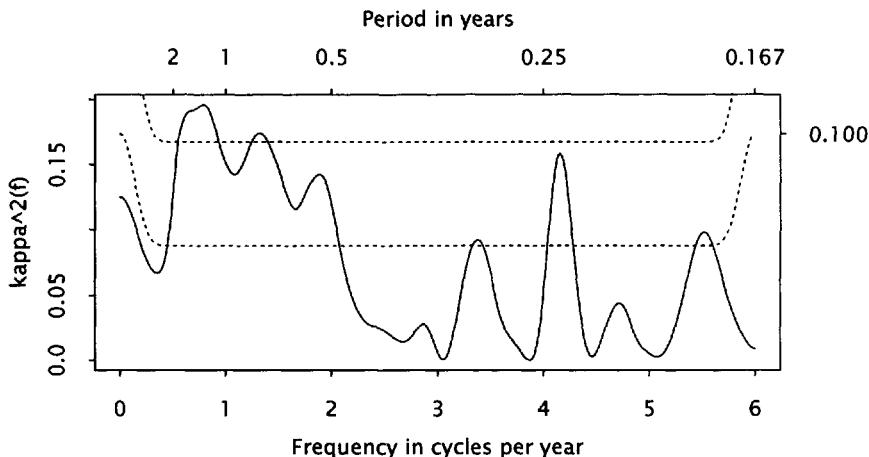


Fig. 10.15 Estimated squared coherency of lagged ten-year interest rates and monthly changes in industrial production, with cutoff levels for 10% and 1% significance levels under the null hypothesis of incoherency. Interest rates lagged one quarter.

The realignment has not simply adjusted the phase by a linear function of frequency, although this is what was suggested by Exercise 10.7. However, the new phase differs from the result of such an adjustment chiefly by (irrelevant) whole cycles, making transitions where the coherency is small. Interestingly, at all of the frequency bands where the coherency is large enough for the phase confidence interval to be defined, the phase spectrum is close either to $1/2$ or to an alias of $1/2$, suggesting that a single realignment is appropriate at all frequencies.

Appendix

The S-Plus function `spec.pgram()` described in the Appendix to Chapter 8 (p. 164) may be given a multiple time series as its first argument, in which case the list that it returns includes a matrix of estimated squared coherencies and phases in addition to the spectra. It was used in this way to produce the spectra graphed in this chapter.

In constructing the multiple time series (the functions `tsmatrix()` and `ts.intersect()` may be used for this), the natural alignment is used, all elements in a given row corresponding to the same time. If a different alignment is needed, the function `lag()` may be used to change the time scale of one of the components.

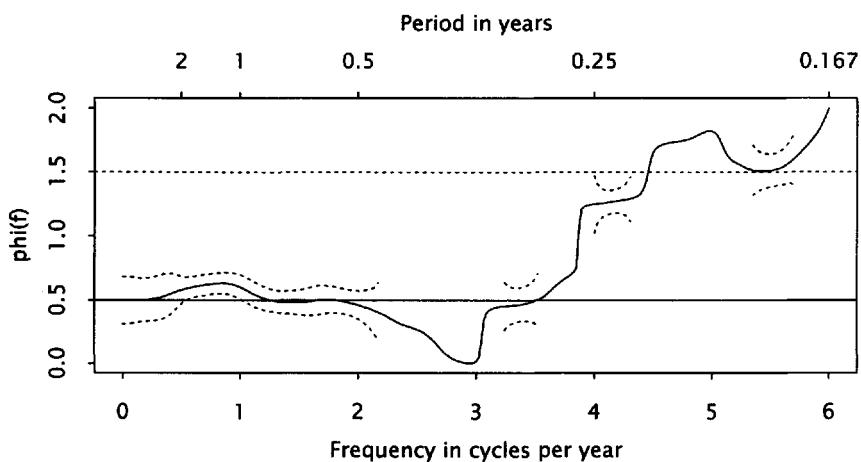


Fig. 10.16 Estimated phase of lagged ten-year interest rates and monthly changes in industrial production, with approximate 95% confidence interval. Interest rates lagged one quarter.

11

Further Topics

Many aspects of time series analysis have not been covered in the preceding chapters. Some of these have been omitted because they are not Fourier analysis methods, while others are more advanced and cannot be treated without a more thorough development of the theories of statistics and stochastic processes. This chapter contains brief descriptions of some of these topics, with references to more extensive discussion.

11.1 TIME DOMAIN ANALYSIS

All the methods of analysis described in this book may be termed *frequency domain* methods, in that they seek to describe the fluctuations in one or more series in terms of sinusoidal behavior at various frequencies. The other main type of analysis is *time domain* analysis, in which the behavior of a series is described in terms of the way in which observations at different times are related statistically.

The basic tools in time domain analysis are the sample autocovariances $\{c_r\}$ defined in Section 8.3 and the related sequence of *autocorrelations*, usually defined as

$$r_r = \frac{c_r}{c_0}, \quad |r| < n.$$

These are a normalized version of the autocovariances and satisfy $r_0 = 1$, $|r_r| \leq 1$. The autocorrelation at lag r , r_r , measures the extent to which

the observation at time t is related to the observation at time $t - r$.

The autocorrelation sequence and the periodogram $I(f)$ [strictly, the normalized periodogram $I(f)/c_0$] are equivalent in the sense that each is the Fourier transform of the other. However, the periodogram, either unsmoothed or in its smoothed form as a spectrum estimate, is the simpler of the two to interpret. This is so because periodogram ordinates of a stretch of n observations at frequencies separated by more than $1/n$ are approximately independent, whereas estimated autocorrelations in general have a complicated covariance structure.

Time domain methods have been applied with great success to some specific problems in time series analysis, including *forecasting* and *control*. In the forecasting problem one observes a series $\{x_t\}$ (or, more generally, a number of series, $\{x_t\}$, $\{y_t\}$, ...) up to time n and wishes to predict or forecast some future value x_{n+r} , $r > 0$. In the control problem one observes a series $\{x_t\}$ that depends on another series $\{y_t\}$ whose values may be determined by the observer (more generally, $\{x_t\}$ depends on a number of observed series $\{y_t\}$, $\{z_t\}$, ...), some of whose values are under the control of the observer). The problem here is to manipulate the controllable series in such a way that future values of $\{x_t\}$ lie as close as possible to desired values. The theory of these problems is described by Whittle (1963), and their solution by time domain methods is discussed extensively by Box et al. (1994).

11.2 SPATIAL SERIES

In the examples used in earlier chapters the variable t has always represented time in one unit or another. However, the methods described are immediately applicable to any set of observations associated with a single variable with constant increments, such as the thickness of a lamina at equally spaced points along a line. More generally, we might have observations at the points of a rectangular grid in the plane (or a higher-dimensional space), say,

$$x_{t_1, t_2}, \quad t_1 = 0, 1, \dots, n_1 - 1, \quad t_2 = 0, 1, \dots, n_2 - 1.$$

Such a collection of data is called a *spatial series*. The basic tool in the Fourier analysis of such data is the multidimensional discrete Fourier transform

$$\begin{aligned} d_{j_1, j_2} &= d(f_{1, j_1}, f_{2, j_2}) \\ &= \frac{1}{n_1 n_2} \sum_{t_1=0}^{n_1-1} \sum_{t_2=0}^{n_2-1} x_{t_1, t_2} e^{-2\pi i (f_{1, j_1} t_1 + f_{2, j_2} t_2)}, \end{aligned}$$

where $f_{1,j} = j/n_1$ and $f_{2,j} = j/n_2$ are Fourier frequencies associated with n_1 and n_2 , respectively. The inverse transform is

$$x_{t_1, t_2} = \sum_{j_1=0}^{n_1-1} \sum_{j_2=0}^{n_2-1} d_{j_1, j_2} e^{2\pi i(f_{1,j_1} t_1 + f_{2,j_2} t_2)}, \quad (11.1)$$

which represents the data as linear combinations of elementary sinusoids. As a function of t_1 and t_2 , the (j_1, j_2) term

$$d_{j_1, j_2} e^{2\pi i(f_{1,j_1} t_1 + f_{2,j_2} t_2)}, \quad (11.2)$$

depends only on $f_{1,j_1} t_1 + f_{2,j_2} t_2$. If we make an orthogonal change of variables to

$$\begin{aligned} t'_1 &= \frac{f_{1,j_1} t_1 + f_{2,j_2} t_2}{\sqrt{f_{1,j_1}^2 + f_{2,j_2}^2}}, \\ t'_2 &= \frac{f_{2,j_2} t_1 - f_{1,j_1} t_2}{\sqrt{f_{1,j_1}^2 + f_{2,j_2}^2}}, \end{aligned}$$

the term (11.2) becomes

$$d_{j_1, j_2} e^{2\pi i F t'_1}$$

where $F = \sqrt{f_{1,j_1}^2 + f_{2,j_2}^2}$, and does not depend on t'_2 . Thus the (j_1, j_2) term is a sinusoidal surface, constant along the lines $t'_1 = \text{constant}$, which are parallel to the t'_2 -axis, and with wavelength $1/F = 1/\sqrt{f_{1,j_1}^2 + f_{2,j_2}^2}$ (measured orthogonally to these lines). The inverse transform (11.1), therefore, represents the data as a sum of sinusoidal surfaces with different orientations and wavelengths.

If we were looking for such purely sinusoidal components in the data, we would compute the periodogram

$$I(f_1, f_2) = \frac{1}{n_1 n_2} |d(f_1, f_2)|^2$$

from a suitably tapered set of data. More commonly, however, we will be interested in a smoothed version

$$\hat{s}(f_1, f_2) = \int \int W(f_1 - f'_1, f_2 - f'_2) I(f'_1, f'_2) df'_1 df'_2$$

or a discrete analog. The simplest way to carry out this smoothing is to divide the (f_1, f_2) plane into possibly overlapping sets, average each set of periodogram ordinates, and associate the result with the average of the frequencies. An alternative is to divide the data into possible overlapping rectangles, compute periodograms for each, and average them. In terms

of the arithmetic operations required, these two approaches are similar, but the latter requires less storage, since the data are processed in segments. Since both n_1 and n_2 may be large, computation time and storage requirements may be important considerations in the choice of method.

If more sophisticated smoothing is required, the discrete averaging method may be improved by introducing nonconstant spectral weights and by increasing the number of points at which averages are computed. The equivalent spectral window for the second method is

$$W(f, f') = |w(f, f')|^2,$$

where

$$w(f, f') = \sum_{t, t'} w_{t, t'} e^{-2\pi i (ft + f't')}$$

and $w_{t, t'}$ is the data window used on each rectangular segment. Thus any spectral window may be used implicitly by a suitable choice of data window.

Often the orientation of the grid on which the data are collected is arbitrary. In this case it is desirable that the smoothed periodogram be approximately invariant under rotation of the grid (or, rather, equivariant, since the orientation of the sinusoidal surfaces is referred to the orientation of the grid). This may be achieved by using a spectral window with circular symmetry. In the second case the spectral window is circularly symmetric if the data window is likewise circularly symmetric.

The statistical properties of the spectrum estimates obtained in this way may be derived by making the appropriate extensions to the definition of a stationary process. Bartlett (1955) discusses processes defined in this way, and Rayner (1971) describes the spectrum analysis of spatial series. Unwin and Hepple (1974) review the general analysis of spatial processes and give an extensive bibliography, including several applications of spectrum analysis. General issues in the analysis of spatial data are discussed by Cressie (1991) and Ripley (1981, 1988).

11.3 MULTIPLE SERIES

In Chapter 10 we discussed estimating the cross spectrum, coherency, and phase of a pair of series. These are covariance- and correlation-like quantities, in that they are symmetric (or Hermitian) functions of the two series, and describe the way in which the two series are *related*. Often, however, one series *depends* (or is thought to) on one or more other series,

and this dependence is the real object of the investigation. In this case an analog of regression analysis is needed.

Suppose that the dependent series is $\{X_{1,t}\}$, and that it depends on $\{X_{2,t}\}, \{X_{3,t}\}, \dots, \{X_{p,t}\}$, through

$$X_{1,t} = \sum_{j=2}^p \sum_u g_{j,u} X_{j,t-u} + Y_t, \quad (11.3)$$

where Y_t is uncorrelated with all the terms in the independent series. Then the sum in (11.3) is the linear combination of the terms in the independent series that is closest to $X_{1,t}$ in the sense of mean squared error and is called the *linear regression* of $X_{1,t}$ on the independent series. If $\{X_{2,t}\}, \{X_{3,t}\}, \dots, \{X_{p,t}\}$ are jointly weakly stationary (i.e., all pairs satisfy the definition in Section 10.3), and $\{Y_t\}$ is weakly stationary, then $X_{1,t}, \{X_{2,t}\}, \{X_{3,t}\}, \dots, \{X_{p,t}\}$ are jointly weakly stationary, and we may define the spectra $\{s_{j,k}(f), j, k = 1, 2, \dots, p\}$. If $G_j(f)$ is the transfer function of the j th filter $\{g_{j,u}\}$, it may be shown that

$$\begin{bmatrix} \overline{G_2(f)} \\ \vdots \\ \overline{G_p(f)} \end{bmatrix} = \begin{bmatrix} s_{2,2}(f) & \dots & s_{2,p}(f) \\ \vdots & & \vdots \\ s_{p,2}(f) & \dots & s_{p,p}(f) \end{bmatrix}^{-1} \begin{bmatrix} s_{2,1}(f)(f) \\ \vdots \\ s_{2,p}(f) \end{bmatrix}. \quad (11.4)$$

The *residual spectrum* $s_{Y,Y}(f)$ is given by

$$\begin{aligned} s_{Y,Y}(f) &= s_{1,1}(f) - \left[s_{1,2}(f)(f) \quad \dots \quad s_{1,p}(f) \right] \begin{bmatrix} \overline{G_2(f)} \\ \vdots \\ \overline{G_p(f)} \end{bmatrix} \\ &= s_{1,1}(f) \{1 - R(f)\}, \end{aligned}$$

say. The quantity $R(f)$ is the *multiple coherency* of $X_{1,t}$ with $\{X_{2,t}\}, \{X_{3,t}\}, \dots, \{X_{p,t}\}$ and is the analog of the multiple correlation coefficient. Note that if $p = 2$, (11.4) simplifies to

$$\overline{G_2(f)} = \frac{s_{2,1}(f)}{s_{2,2}(f)}$$

and

$$\begin{aligned} R(f) &= \frac{s_{1,2}(f)s_{2,1}(f)}{s_{1,1}(f)s_{2,2}(f)} \\ &= \frac{|s_{1,2}(f)|^2}{s_{1,1}(f)s_{2,2}(f)}, \end{aligned}$$

the squared coherency discussed in Chapter 10.

If it is known that two of the series, say $X_{1,t}$ and $\{X_{2,t}\}$, depend on the others, we may wish to investigate whether there is any other dependence between them, or whether they are related only by the fact that both depend on $\{X_{3,t}\}, \dots, \{X_{p,t}\}$. To do this we compute the *partial spectra*

$$\begin{bmatrix} s_{1,1 \cdot 3, \dots, p}(f) & s_{1,2 \cdot 3, \dots, p}(f) \\ s_{2,1 \cdot 3, \dots, p}(f) & s_{2,2 \cdot 3, \dots, p}(f) \end{bmatrix} = \begin{bmatrix} s_{1,1}(f) & s_{1,2}(f) \\ s_{2,1}(f) & s_{2,2}(f) \end{bmatrix}^{-1} \begin{bmatrix} s_{3,3}(f) & \dots & s_{3,p}(f) \\ \vdots & \ddots & \vdots \\ s_{p,3}(f) & \dots & s_{p,p}(f) \end{bmatrix} \begin{bmatrix} s_{3,1}(f) & s_{3,2}(f) \\ \vdots & \vdots \\ s_{p,1}(f) & s_{p,2}(f) \end{bmatrix}$$

and hence we may find the *partial coherency* and *partial phase* (see Priestley, 1981b, Chapter 9). If the partial coherency is not significantly different from 0, the two series $\{X_{1,t}\}$ and $\{X_{2,t}\}$ have no further dependence.

In practice, the theoretical spectra are replaced by estimates. Hannan (1970, Chapter V) describes the statistical properties of the resulting estimates of transfer functions, multiple coherencies, and partial coherencies. Brillinger (1981, Chapter 8) gives extensive discussion of the computation and use of these quantities.

Other methods of multivariate analysis such as principal components and canonical correlations may also be extended to apply to time series data (see, e.g., Brillinger, 1981, Chapters 9 and 10).

11.4 HIGHER ORDER SPECTRA

In Section 6.5, it was shown that the third-order periodogram can give us information about nonsinusoidal behavior of a series (in that case, the sunspot series). In Section 7.5 a similar quantity was calculated in a local way using complex demodulation. Brillinger and Rosenblatt (1967a,b) define the k th order autoperiodogram as

$$I_k(f_1, f_2, \dots, f_k) = n^{k-1} \prod_{j=1}^k d(f_j), \quad (11.5)$$

where $\sum f_j \equiv 0 \pmod{1}$. We have

$$\begin{aligned} I_2(f, -f) &= n d(f) \overline{d(f)} \\ &= n |d(f)|^2 \\ &= I(f) \end{aligned}$$

and thus (11.5) contains the familiar periodogram as a special case. The definition may be extended in an obvious way to give k th order cross periodograms of a multiple series.

The expectation of a k th order periodogram is (approximately) the Fourier transform of the cumulants of the process being analyzed (Brillinger and Rosenblatt, 1967a). Since all cumulants of orders higher than 2 of a Gaussian process vanish, these higher order spectra provide information about nonnormality of a process. They also give information about nonlinearity in the structure of a process and may be used to indicate whether some transformation such as those discussed in Section 6.7 will lead to a series with simpler structure (see Brillinger, 1965; Godfrey, 1965).

It is easily seen that nonsinusoidal oscillations in a series are evidence of nonnormality in its probability structure. For a zero-mean Gaussian process $\{X_t\}$ has the symmetry properties that $\{Y_t\}$ and $\{Z_t\}$ have the same distribution as $\{X_t\}$, where $Y_t = -X_t$ and $Z_t = X_{-t}$. However, in the sunspot series (Figure 1.3, p. 4) we see oscillations that do not have these symmetry properties. In the square roots of the sunspot series (Figure 6.13, p. 84) the spatial asymmetry has been largely eliminated, but the time asymmetry remains (see also Figure 6.15, p. 85). Since no transformation of the data can remove this time asymmetry, it represents a fundamental nonnormality in the distribution of the sunspot numbers.

Like spectrum estimates for series defined for multidimensional "time," higher order spectrum estimates may be calculated either by frequency domain smoothing of the corresponding higher order periodogram of the whole series, or by time domain averaging of periodograms of segments of the data. Brillinger and Rosenblatt (1967b) describe an estimate of the first type, while Godfrey (1965) uses the second type. The second approach is preferable from a computational point of view. It requires less storage because the data are processed one segment at a time and the periodograms of successive segments may be accumulated as they are computed. Also, for $k > 2$, fewer arithmetic operations are required to compute the periodograms from the transforms.

11.5 NONQUADRATIC SPECTRUM ESTIMATES

All the spectrum estimates discussed in Chapters 8 and 9 are quadratic forms of the data $\{x_t\}$; that is, they may be written as

$$\hat{s}(f) = \sum_{t,t'=0}^{n-1} x_t a_{t,t'}(f) x_{t'}.$$

Similarly, the cross spectrum estimates of Chapter 9 are bilinear forms of the two series and may be written as

$$\hat{s}_{x,y}(f) = \sum_{t,t'=0}^{n-1} x_t a_{t,t'}(f) y_{t'}.$$

In recent years there has been some interest in other kinds of spectrum estimates. One approach is to assume that the spectrum belongs to some parametric family $s(f; \theta_0, \theta_1, \dots, \theta_p)$, where $\theta_0, \theta_1, \dots, \theta_p$ are unknown parameters. Estimates $\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p$ are computed from the data, and the spectrum is then estimated by

$$\hat{s}(f) = s(f; \hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p).$$

The family used most widely is that of *autoregressive spectra*

$$s(f; \theta_0, \theta_1, \dots, \theta_p) = \frac{\theta_0}{\left| 1 - \theta_1 e^{2\pi i f} - \dots - \theta_p e^{2\pi i f p} \right|^2}$$

(see, e.g. Parzen, 1969; Marple, 1987, Chapters 8–10). This is the spectrum of the autoregressive process $\{X_t\}$, defined by

$$X_t = \theta_1 X_{t-1} + \dots + \theta_p X_{t-p} + U_t, \quad (11.6)$$

where $\{U_t\}$ is a zero-mean white noise process with variance θ_0 , and

$$E(U_t X_{t-r}) = 0, \quad r = 1, 2, \dots, p. \quad (11.7)$$

The parameters $\theta_0, \theta_1, \dots, \theta_p$ are constrained to satisfy

- $\theta_0 > 0$, and
- $1 - \theta_1 z - \dots - \theta_p z^p \neq 0$ for any $|z| \leq 1$,

the second condition being required to ensure the existence of a stationary solution to the difference equation (11.6) satisfying (11.7). Relations (11.6) and (11.7) show that the autoregressive process has some of the properties of the conventional multiple linear regression model, and its parameters may be estimated in a similar way (see, e.g., Box et al., 1994, Sections 7.3.1 and A7.5). The exponential model of Bloomfield (1973) may be used in the same way to produce estimated spectra.

Another approach was used by Burg (see, e.g., Lacoss, 1971). Suppose that we know the first m autocovariances of a weakly stationary process (see Section 9.1, p.167), y_0, y_1, \dots, y_{m-1} or, alternatively, that we have

good estimates of them. These are consistent with any spectrum $s(f)$ for which

$$\int_0^1 s(f) e^{2\pi i f r} df = y_r, \quad r = 0, 1, \dots, m-1 \quad (11.8)$$

and Burg's procedure is to use the spectrum satisfying (11.8) that maximizes the *entropy*

$$\int_0^1 \log s(f) df.$$

In this sense it is the smoothest spectrum satisfying (11.8). It may be shown (Lacoss, 1971) that the resulting *maximum entropy spectrum* is of autoregressive form; in fact, if the m autocovariances are estimated from a stretch of data (the most common case), the resulting estimate is the same as the autoregressive spectrum estimated from that stretch of data (with $p = m - 1$).

A different estimate, called the *maximum likelihood spectrum*, is suggested by Capon (1969). A relationship between this and the maximum entropy spectrum (and hence also the autoregressive spectrum) is given by Burg (1972). Pisarenko (1972) describes a class of generally nonquadratic estimates that includes as special cases the conventional quadratic spectrum estimate (in the form of the Bartlett estimate) and the maximum likelihood spectrum (see also Marple, 1987, Chapter 13).

Various arguments have been made in support of the use of these spectrum estimates. The general procedure of assuming a parametric model for a spectrum and then estimating the parameters is efficient (in a statistical sense) when the true spectrum belongs to the family, and is a flexible method if the family is diverse enough. However, the resulting spectrum estimate is a more complex function of the data than a conventional quadratic estimate, and on these grounds is harder to interpret. It is also computationally more complex, at least in the case of a large number of parameters.

It has also been shown that nonquadratic spectra may have greater *resolution* than a conventional quadratic spectrum estimate using the same number of autocovariances (though we note that some of the comparisons that have been made use the Bartlett estimate as the conventional estimate, and that the Bartlett estimate is undesirable on at least two counts; see Sections 8.3, p. 142, and 8.5, p. 149). Nonquadratic spectra would therefore be useful in situations where only a fixed number of *autocovariances* are known or estimated, and a spectrum estimate of the highest possible resolution is needed. One such situation would be where we observe many short segments of time series with the same structure;

autocovariances pooled across segments could provide good estimates at short lags, but no estimates would be available at longer lags.

However, in the more common case where a single stretch of the original *data* is available for analysis, conventional estimates may be constructed with any desired resolution. In the extreme case where it is desired to estimate the frequency of a sinusoid from a few cycles or a fraction of a cycle, the exact least squares methods of Chapter 3 give the maximum resolution.

11.6 INCOMPLETE DATA, IRREGULARLY SPACED DATA, AND POINT PROCESSES

Often a series of observations from which it is desired to compute a spectrum estimate is incomplete (or contains bad values, which should be omitted). There are a number of ways in which this situation may be handled.

Gaps Large and Few

If the gaps are relatively few and far apart, and especially if each gap is fairly large, the simplest procedure is to treat each uninterrupted stretch of data as a separate series, compute a periodogram for each, and average them. The segments should each be tapered and then extended by zeros to a common length, to ensure that each periodogram is relatively free of leakage and that all are computed at the same frequency.

Suppose that there are m stretches of data, of lengths n_1, n_2, \dots, n_m , and that we compute the weighted average

$$\hat{s}_1(f) = \sum_{j=1}^m a_j I_j(f),$$

where $a_j > 0, j = 1, \dots, m$, and $\sum_j a_j = 1$. This is computed at the frequencies $f'_j = j/n'$ for some $n' \geq \max\{n_1, n_2, \dots, n_m\}$ and then smoothed to form

$$\begin{aligned}\hat{s}_2(f) &= \sum_u g_u \hat{s}_1(f - f'_u) \\ &= \sum_{j=1}^m \sum_u g_u I_j(f - f'_u).\end{aligned}$$

Now the periodograms of disjoint stretches of data are approximately

uncorrelated, whence

$$\text{var} \{ \hat{s}_2(f) \} \approx \sum_{j=1}^m a_j^2 \text{var} \left\{ \sum_u g_u I_j(f - f'_u) \right\} \quad (11.9)$$

and

$$\text{var} \left\{ \sum_u g_u I_j(f - f'_u) \right\} \approx s(f)^2 \min \left\{ 1, \frac{n'}{n} U_j \sum_u g_u^2 \right\}, \quad (11.10)$$

where U_j is the variance inflation factor due to tapering of the j th stretch of data (see Section 9.5, p. 177). Note that the variance inflation factors may be different for different stretches, since we may wish to taper the same *number* of observations, rather than the same *proportion*, at the ends of each stretch.

Expression (11.10) incorporates a modification that prevents it from exceeding $s(f)^2$; if n_j were so small that

$$\frac{n'}{n_j} U_j \sum_u g_u^2 > 1, \quad (11.11)$$

then filtering with weights $\{g_u\}$ would have little effect on $I_j(f)$. Thus

$$\sum_u g_u I_j(f - f'_u) \approx I_j(f)$$

and hence

$$\begin{aligned} \text{var} \left\{ \sum_u g_u I_j(f - f'_u) \right\} &\approx \text{var} \{ I_j(f) \} \\ &\approx s(f)^2. \end{aligned}$$

From (11.9) and (11.10) it follows that $\text{var} \{ \hat{s}_2(f) \}$ is minimized by taking

$$a_j = \frac{a}{\min \left\{ 1, \frac{n'}{n} U_j \sum_u g_u^2 \right\}},$$

where a is chosen so that $\sum_j a_j = 1$; that is,

$$\frac{1}{a} = \sum_{j=1}^m \frac{1}{\min \left\{ 1, \frac{n'}{n} U_j \sum_u g_u^2 \right\}}.$$

Thus the optimal weights in the averaging of $I_j(f)$ depend on the amount of subsequent smoothing. Often (11.11) does not hold for any j , and then

$$a_j = \frac{n_j/U_j}{\sum_{k=1}^m n_k/U_k}. \quad (11.12)$$

Thus, if $\hat{s}_2(f)$ is a sufficiently heavily smoothed form of $\hat{s}_1(f)$, the optimal weights are given by (11.12) and do not depend on the amount of smoothing. If the weights (11.12) are used when (11.11) does hold for some j , the effect is to place less weight on the j th segment. Since only the shortest segments are affected in this way, it seems reasonable that weights (11.12) should be used in any case. The resolution of these short segments is poor, and such down-weighting therefore reduces the bias in $\hat{s}_2(f)$ because of this poor resolution.

Gaps Small and Few

On the other hand, if missing or bad observations tend to occur in isolated ones or twos, it is simplest to replace any such observation by a linear combination of its neighbors. It may be shown that the effect of this is to introduce a small bias into the spectrum, proportional to the fraction of data that are missing. The effect on the variance of the estimate is to replace $s(f)$ by the biased form, and thus the variance of $\log \hat{s}(f)$ is unaffected.

A combination of these two approaches may also be used. The data are divided into segments at any long gaps, and shorter gaps within each segment are then filled in by linear combinations of their neighbors.

Gaps Small and Many

When a moderate proportion of the data is missing, say more than 10% or 20%, it may not be possible to divide the data into stretches with relatively few data points missing in each stretch. In this case even the combined approach may lead to an unacceptable bias in the estimated spectrum. However, in the special case where missing data are replaced by zeros (a trivial linear combination), the bias may be estimated and removed. Estimates constructed in this way have been considered by several authors. Jones (1962) and Parzen (1963) assume that the data are missed in some periodic way. Scheinok (1965) assumes that each observation is observed with a given probability, independently of the others. Bloomfield (1970) generalizes this to a general random mechanism. Jones (1971) makes no assumption about the mechanism that causes observations to be missed, and gives a formula for the approximate variance that differs from the expressions of Scheinok and Bloomfield.

Any of these procedures may result in negative spectrum estimates, which are generally undesirable. This is less likely to happen if the spectrum is close to white than if it has a large *dynamic range*. Thus the data

should usually be *prewhitened* (see Section 8.8, p. 160) to some extent, even though this increases the number of missing observations.

Irregular Data

It has been assumed throughout this book that the data are collected at equally spaced times. The missing data problems considered above represent a small departure from this assumption. More generally, however, we may wish to estimate the spectrum of a continuous time series $\{x(t)\}$, observed at times t_1, t_2, \dots, t_n which do not fall on a grid.

If these epochs of observation are generated by some random mechanism, they constitute a realization of a *point process*. The estimation of the spectrum of $\{x(t)\}$ is related to the estimation of the spectrum of the point process $\{t_i\}$, which is defined by Bartlett (1963). These estimation problems are discussed by Bartlett and by Brillinger (1972). Several issues in the analysis of irregular data are discussed by contributors to Parzen (1984). It is interesting that the first theoretical study of the statistical properties of the periodogram was presented by Schuster (1897) in the context of a point process, the occurrence of earthquakes in Japan.

References

- Anderson, T. W. (1994). *The Statistical Analysis of Time Series* (Reprint ed.). New York: Wiley.
- Bartlett, M. S. (1948). Smoothing periodograms from time series with continuous spectra. *Nature* 161, 686-687.
- Bartlett, M. S. (1950). Periodogram analysis and continuous spectra. *Biometrika* 37, 1-16.
- Bartlett, M. S. (1955). *An Introduction to Stochastic Processes, with Special Reference to Methods and Applications*. Cambridge: Cambridge University Press.
- Bartlett, M. S. (1963). The spectral analysis of point processes. *J. Roy. Statist. Soc., Ser. B* 25, 264-280.
- Bergland, G. D. (1968). A fast Fourier transform algorithm using base 8 iterations. *Math. Comput.* 22, 275-279.
- Beveridge, W. H. (1921). Weather and harvest cycles. *Econ. J.* 31, 429-452.
- Beveridge, W. H. (1922). Wheat prices and rainfall in Western Europe. *J. Roy. Statist. Soc.* 85, 412-459.
- Bingham, C., M. D. Godfrey, and J. W. Tukey (1967). Modern techniques of power spectrum estimation. *IEEE Trans. Audio Electroacoust.* AU-15, 56-66.
- Blackman, R. B. and J. W. Tukey (1959). *The Measurement of Power Spectra, from the Point of View of Communications Engineering*. New York: Dover.

- Bloomfield, P. (1970). Spectral analysis with randomly missing observations. *J. Roy. Statist. Soc., Ser. B* 32, 369–380.
- Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika* 60, 217–226.
- Bloomfield, P. (1991). Time series methods. In D. V. Hinkley, N. Reid, and E. J. Snell (Eds.), *Statistical Theory and Modelling*, pp. 152–176. London: Chapman and Hall.
- Box, G. E. P., G. M. Jenkins, and G. C. Reinsel (1994). *Time Series Analysis: Forecasting and Control* (3rd ed.). San Francisco: Holden-Day.
- Bray, R. J. and R. E. Loughhead (1964). *Sunspots*. New York: Wiley.
- Brigham, E. O. (1988). *The Fast Fourier Transform* (Facsimile ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Brillinger, D. R. (1965). An introduction to polyspectra. *Ann. Math. Statist.* 36, 1351–1374.
- Brillinger, D. R. (1972). The spectral analysis of stationary interval functions. In L. M. LeCam, J. Neyman, and E. L. Scott (Eds.), *Proceedings Sixth Berkeley Symposium*, pp. 483–513. Berkeley: University of California Press.
- Brillinger, D. R. (1973). An empirical investigation of the Chandler wobble and two proposed excitation processes. *Bull. Int. Statist. Inst.* 45, Book 3, 413–434.
- Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory* (Expanded ed.). New York: Holt, Rinehart, and Winston.
- Brillinger, D. R. (1993). The digital rainbow: Some history and applications of numerical spectrum analysis. *Can. J. Statist.* 21, 1–19.
- Brillinger, D. R. and M. Rosenblatt (1967a). Asymptotic theory of estimates of k -th order spectra. In B. Harris (Ed.), *Spectral Analysis of Time Series*, pp. 153–188. New York: Wiley.
- Brillinger, D. R. and M. Rosenblatt (1967b). Computation and interpretation of k -th order spectra. In B. Harris (Ed.), *Spectral Analysis of Time Series*, pp. 189–232. New York: Wiley.
- Burg, J. P. (1972). The relationship between maximum entropy spectra and maximum likelihood spectra. *Geophysics* 37, 375–376.
- Capon, J. (1969). High-resolution frequency-wavenumber spectral analysis. *Proc. IEEE* 57, 1408–1418.
- Chiu, S.-T. (1989). Detecting periodic components in a white Gaussian time series. *J. Roy. Statist. Soc. B* 51, 249–259.

- Cooley, J. W., P. A. W. Lewis, and P. D. Welch (1967). Historical notes on the fast Fourier transform. *IEEE Trans. Audio Electroacoust.* AU-15, 76-79.
- Cooley, J. W. and J. W. Tukey (1965). An algorithm for the machine computation of complex Fourier series. *Math. Comput.* 19, 297-301.
- Cressie, N. A. C. (1991). *Statistics for Spatial Data*. New York: Wiley.
- Dale, J. B. (1914a). Note on the number of components of a compound periodic function. *Mon. Notic. Roy. Astron. Soc.* 74, 664.
- Dale, J. B. (1914b). The resolution of a compound periodic function into simple periodic functions. *Mon. Notic. Roy. Astron. Soc.* 74, 628-648.
- Daniell, P. J. (1946). Discussion on the Symposium on Autocorrelation in Time Series. *J. Roy. Statist. Soc. (Suppl.)* 8, 88-90.
- Doob, J. L. (1953). *Stochastic Processes*. New York: Wiley.
- Durbin, J. and G. S. Watson (1950). Testing for serial correlation in least squares regression. I. *Biometrika* 37, 408-428.
- Durbin, J. and G. S. Watson (1951). Testing for serial correlation in least squares regression. II. *Biometrika* 38, 159-178.
- Durbin, J. and G. S. Watson (1971). Testing for serial correlation in least squares regression. III. *Biometrika* 58, 1-19.
- Easterling, D. R., T. R. Karl, E. H. Mason, P. Y. Hughes, D. P. Bowman, R. C. Daniels, and T. A. Boden (Eds.) (1996). *United States Historical Climatology Network (U.S. HCN) Monthly Temperature and Precipitation Data*. Oak Ridge, Tennessee: Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory.
- Einstein, A. (1914). Méthode pour la détermination de valeurs statistiques d'observations concernant des grandeurs soumises à des fluctuations irrégulières. *Arch. Sci. Phys. Natur., Ser. 4* 37, 254-256.
- Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory Probab. Appl.* 14, 153-158.
- Feller, W. (1968). *An Introduction to the Theory of Probability and Its Applications* (3rd ed.), Volume I. New York: Wiley.
- Fisher, R. A. (1929). Tests of significance in harmonic analysis. *Proc. Roy. Soc., Ser. A* 125, 54-59.
- Fuller, W. A. (1995). *Introduction to Statistical Time Series* (2nd ed.). New York: Wiley.

- Gentleman, W. M. and G. Sande (1966). Fast Fourier transforms—for fun and profit. In *1966 Fall Joint Comput. Conf., AFIPS Conf. Proc.*, Volume 29, pp. 563–578.
- Godfrey, M. D. (1965). An exploratory study of the bispectrum of an economic time series. *Appl. Statist.* 14, 48–69.
- Good, I. J. (1958). The interaction algorithm and practical Fourier analysis. *J. Roy. Statist. Soc. Ser. B* 20, 361–372.
- Good, I. J. (1971). The relationship between two fast Fourier transforms. *IEEE Trans. Comput. C-20*, 310–317.
- Granger, C. W. J. and A. O. Hughes (1971). A new look at some old data: The Beveridge wheat price series. *J. Roy. Statist. Soc. Ser. A* 134, 413–428.
- Grenander, U. and M. Rosenblatt (1953). Statistical spectral analysis of time series arising from stationary stochastic processes. *Ann. Math. Statist.* 24, 537–558.
- Grenander, U. and M. Rosenblatt (1957). *Statistical Analysis of Stationary Time Series*. New York: Wiley.
- Hamming, R. W. (1987). *Numerical Methods for Scientists and Engineers* (2nd ed.). New York: Dover.
- Hamming, R. W. (1998). *Digital Filters*. New York: Dover.
- Hamming, R. W. and J. W. Tukey (1949). Measuring noise color. Bell Telephone Laboratories Memorandum.
- Hannan, E. J. (1970). *Multiple Time Series*. New York: Wiley.
- Hannan, E. J. and P. J. Thompson (1973). Estimating group delay. *Biometrika* 60, 241–253.
- Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proc. IEEE* 66, 51–88.
- Hart, B. I. and J. von Neumann (1942). Tabulation of the probabilities for the ratio of the mean square successive difference to the variance. *Ann. Math. Statist.* 13, 207–214.
- Hodges, J. L. and E. L. Lehmann (1956). The efficiency of some nonparametric competitors of the *t*-test. *Ann. Math. Statist.* 27, 324–335.
- Ibragimov, I. A. and Y. V. Linnik (1971). *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff.
- Jenkins, G. M. (1961). General considerations in the analysis of spectra. *Techometrics* 3, 133–166.

- Jenkins, G. M. and D. G. Watts (1968). *Spectral Analysis and Its Applications*. San Francisco: Holden-Day.
- Jones, R. H. (1962). Spectral analysis with regularly missed observations. *Ann. Math. Statist.* 31, 568–573.
- Jones, R. H. (1971). Spectrum estimation with missing observations. *Ann. Inst. Statist. Math.* 23, 387–398.
- Kaiser, J. F. and R. W. Schafer (1980). On the use of the I_0 -sinh windows for spectrum analysis. *IEEE Trans. Acoust. Speech Signal Process.* ASSP-28, 105–107.
- Karl, T. R., C. N. Williams, F. T. Quinlan, and T. A. Boden (1990). United States Historical Climate Network (HCN) serial temperature and precipitation data. Technical Report 3404, Carbon Dioxide Information and Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN.
- Kendall, D. G. (1948). Oscillatory time series. Review of *Contributions to the Study of Oscillatory Time Series*, by M. G. Kendall. *Nature* 161, 187.
- Kendall, M. G. (1971). Studies in the history of probability and statistics. XXVI. The work of Ernst Abbe. *Biometrika* 58, 369–373.
- Knott, C. G. (1897). On lunar periodicities in earthquake frequency. *Proc. Roy. Soc.* 60, 457–466.
- Koopmans, L. H. (1995). *The Spectral Analysis of Time Series* (Reprint ed.). New York: Academic.
- Lacoss, R. T. (1971). Data adaptive spectral analysis methods. *Geophysics* 36, 661–675.
- Lagrange (1873). Recherches sur la manière de former des tables des planètes d'après les seules observations. In *Oeuvres de Lagrange*, Volume VI, pp. 507–627.
- Lanczos, C. (1961). *Applied Analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- Marple, Jr., S. L. (1987). *Digital Spectral Analysis, with Applications*. Englewood Cliffs, NJ: Prentice-Hall.
- Newton, H. W. (1958). *The Face of the Sun*. Harmondsworth: Penguin.
- Olshen, R. A. (1967). Asymptotic properties of the periodogram of a discrete stationary process. *J. Appl. Probab.* 4, 508–528.
- Otnes, R. K. and L. Enochson (1972). *Digital Time Series Analysis*. New York: Wiley.
- Parzen, E. (1957a). On choosing an estimate of the spectral density function of a stationary time series. *Ann. Math. Statist.* 28, 921–932.

- Parzen, E. (1957b). On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Statist.* 28, 329–348.
- Parzen, E. (1961). Mathematical considerations in the estimation of spectra. *Technometrics* 3, 167–190.
- Parzen, E. (1963). On spectral analysis with missing observations and amplitude modulation. *Sankhya, Ser. A* 25, 383–392.
- Parzen, E. (1969). Multiple time series modelling. In P. R. Krishnaiah (Ed.), *Multivariate Analysis*, Volume I, pp. 389–409. New York: Academic.
- Parzen, E. (Ed.) (1984). *Time Series Analysis of Irregularly Observed Data*. New York: Springer-Verlag.
- Pisarenko, V. F. (1972). On the estimation of spectra by means of nonlinear functions of the covariance matrix. *Geophys. J. Roy. Astron. Soc.* 28, 511–531.
- Pisarenko, V. F. (1973). Parameter estimation for 2 harmonics with closing frequencies on noise background. *Theory Probab. Appl.* 18, 826.
- Priestley, M. B. (1981a). *Spectral Analysis and Time Series*, Volume 1. London: Academic.
- Priestley, M. B. (1981b). *Spectral Analysis and Time Series*, Volume 2. London: Academic.
- Rayner, J. N. (1971). *An Introduction to Spectral Analysis*. London: Pion.
- Richards, E. G. (1998). *Mapping Time: The Calendar and its History*. Oxford: Oxford University Press.
- Ripley, B. D. (1981). *Spatial Statistics*. New York: Wiley.
- Ripley, B. D. (1988). *Statistical Inference for Spatial Process*. Cambridge: Cambridge University Press.
- Rosenblatt, M. (1971). Curve estimates. *Ann. Math. Statist.* 42, 1815–1842.
- Scheinok, P. A. (1965). Spectral analysis with randomly missed observations: The binomial case. *Ann. Math. Statist.* 36, 971–977.
- Schuster, A. (1897). On lunar and solar periodicities of earthquakes. *Proc. Roy. Soc.* 61, 455–465.
- Schuster, A. (1898). On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terr. Magn.* 3, 13–41.
- Schuster, A. (1900). The periodogram of magnetic declination as obtained from the records of the Greenwich Observatory during the years 1871–1895. *Cambridge Phil. Trans.* 18, 107–135.

- Schuster, A. (1906). On the periodicities of sunspots. *Phil. Trans. Roy. Soc., Ser. A* 206, 69-100.
- Slutsky, E. (1937). The summation of random causes as the source of cyclic processes. *Econometrica* 5, 105-146.
- Spar, J. and J. A. Mayer (1973). Temperature trends in New York City: A postscript. *Weatherwise* 26, 128-130.
- Spector, P. (1994). *An Introduction to S and S-Plus*. Boston: Duxbury.
- Stewart, B. and W. Dodgson (1879). Preliminary report to the Committee on Solar Physics on a method of detecting the unknown inequalities of a series of observations. *Proc. Roy. Soc.* 29, 106-122.
- Stokes, G. G. (1879). Note on the paper by Stewart and Dodgson. *Proc. Roy. Soc.* 29, 122-123.
- Stolarski, R. S., P. Bloomfield, R. D. McPeters, and J. R. Herman (1991). Total ozone trends deduced from Nimbus 7 TOMS data. *Geophys. Res. Lett.* 18, 1015-1018.
- Thomson, W. (1876). On an instrument for calculating ($\int \phi(x)\psi(x)dx$), the integral of the product of two given functions. *Proc. Roy. Soc.* 24, 266-268.
- Thomson, W. (1878). Harmonic analyzer. *Proc. Roy. Soc.* 27, 371-373.
- Titchmarsh, E. C. (1939). *Theory of Functions*. London: Oxford University Press.
- Tukey, J. W. (1967). An introduction to the calculations of numerical spectrum analysis. In B. Harris (Ed.), *Spectral Analysis of Time Series*, pp. 25-46. New York: Wiley.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
- Unwin, D. J. and L. W. Hepple (1974). The statistical analysis of spatial series. *Statistician* 23, 211-227.
- von Neumann, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Statist.* 12, 367-395.
- von Neumann, J. (1942). A further remark concerning the distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Statist.* 13, 86-88.
- von Neumann, J., R. H. Kent, H. R. Bellinson, and B. I. Hart (1941). The mean square successive difference. *Ann. Math. Statist.* 12, 153-162.
- Walker, A. M. (1971). On the estimation of a harmonic component in a time series with stationary independent residuals. *Biometrika* 58, 21-36.

- Whittaker, E. T. and G. Robinson (1944). *The Calculus of Observations* (4th ed.). London: Blackie and Son.
- Whittle, P. (1952). The simultaneous estimation of a time series harmonic components and covariance structure. *Trab. Estadist.* 3, 43–57.
- Whittle, P. (1963). *Prediction and Regulation by Linear Least-Squares Methods*. London: The English Universities Press.
- Wold, H. O. A. (1954). *A Study in the Analysis of Stationary Time Series* (2nd ed.). Stockholm: Almqvist and Wiksell.
- Woodroofe, M. B. and J. W. Van Ness (1967). The maximum deviation of sample spectral densities. *Ann. Math. Statist.* 38, 1558–1569.
- Yaglom, A. M. (1987). Einstein's 1914 paper on the theory of irregularly fluctuating series of observations. *IEEE Acoust. Speech Signal Process. Mag.* 4, 7–11.

Author Index

- Abbe, Ernst, 53
Anderson, T. W., 6, 147, 152, 167, 192,
214
Bartlett, M. S., 7, 142, 236, 245
Bellinson, H. R., 53
Bergland, G. D., 61
Beveridge, W. H., 4, 6, 133–136, 138, 140,
141
Bingham, C., 97, 155
Blackman, R. B., 7, 163
Bloomfield, P., 2, 160, 240, 244
Bluestein, L. I., 58, 61
Boden, T. A., 92
Bowman, D. P., 92
Box, G. E. P., 234, 240
Bray, R. J., 76
Brigham, E. O., 7, 58, 61
Brillinger, D. R., 41, 80, 167, 175, 214,
220, 221, 229, 238, 239, 245
Burg, J. P., 240, 241
Buys-Ballot, C. D. H., 6
Capon, J., 241
Chiu, S.-T., 94
Cooley, J. W., 7, 58
Cressie, N. A. C., 236
Dale, J. B., 5
Daniell, P. J., 7, 152
Daniels, R. C., 92
Dodgson, W., 6
Doob, J. L., 168, 174
Durbin, J., 53
Easterling, D. R., 92
Einstein, A., 6, 7
Enochson, L., 100
Epanechnikov, V. A., 193
Feller, W., 23, 90
Fisher, R. A., 91
Fuller, W. A., 91
Gentleman, W. M., 58, 147
Godfrey, M. D., 97, 155, 239
Good, I. J., 58
Granger, C. W. J., 135, 136
Grenander, U., 7
Hamming, R. W., 5, 7, 100, 103, 105
Hannan, E. J., 167, 182, 188, 214, 219,
221, 229, 238
Harris, F. J., 70
Hart, B. I., 53
Hepple, L. W., 236
Herglotz, G., 168
Herman, J. R., 2

- Hodges, J. L., 193
 Hughes, A. O., 135, 136
 Hughes, P. Y., 92
Ibragimov, I. A., 176
 Jenkins, G. M., 159, 184, 191, 193, 220,
 234, 240
 Jones, R. H., 244
 Kaiser, J. F., 70
 Karl, T. R., 92
 Kendall, D. G., 7, 142
 Kendall, M. G., 53
 Kent, R. H., 53
 Knott, C. G., 6
 Koopmans, L. H., 167
 Lacoss, R. T., 240, 241
 Lagrange, J. L., 5
 Lanczos, C., 111, 112
 Lehmann, E. L., 193
 Lewis, P. A. W., 58
 Linnik, Yu. V., 176
 Loughhead, R. E., 76
 Marple, S. L., Jr., 70, 193, 240, 241
 Mason, E. H., 92
 Mayer, J. A., 92
 McPeters, R. D., 2
 Newton, H. W., 76, 78
 Olshen, R. A., 177
 Otnes, R.K., 100
 Parzen, E., 159, 178, 193, 240, 244, 245
 Pisarenko, V. F., 33, 241
 Priestley, M. B., 167, 177, 193, 214, 238
 Prony, R. de, 5
 Quinlan, F. T., 92
 Rayner, J. N., 236
 Reinsel, G. C., 234, 240
 Ripley, B. D., 236
 Robinson, G., 3, 6
 Rosenblatt, M., 7, 80, 193, 238, 239
 Sande, G., 58, 147
 Schafer, R. W., 70
 Scheinok, P. A., 244
 Schuster, A., 6, 136, 140, 245
 Slutsky, E., 136
 Spar, J., 92
 Spector, P., 24
 Stewart, B., 6
 Stokes, G. G., 6
 Stolarski, R. S., 2
 Thompson, P. J., 229
 Thomson, W., 6
 Titchmarsh, E. C., 12
 Tukey, J. W., 7, 58, 69, 97, 103, 104, 155,
 160, 163
 Unwin, D. J., 236
 Van Ness, J. W., 188
 von Neumann, J., 53
 Walker, A. M., 31, 33
 Watson, G. S., 53
 Watts, D. G., 184, 191, 220
 Welch, P. D., 58
 Whittaker, E. T., 3, 6
 Whittle, P., 30, 92, 234
 Williams, C. N., 92
 Wold, H. O. A., 6
 Woodroffe, M. B., 188

Subject Index

- Alias, 8, 21–23
 - principal, 21, 197
- Aliased spectral density function, 196
- Aliased spectrum, 195–199
- Alignment, 203, 226–230
 - of interest rates and industrial production, 229–230
- Amplitude
 - instantaneous, 118–124, 126, 130, 154
 - least squares estimate of, 12
 - of a sinusoid, 8
- Analysis of variance, 45
- Autocorrelation, 233
- Autocovariance, 144, 233
 - theoretical, 168, 211
- Autoregressive spectrum, 240
- Band
 - pass, 105
 - stop, 105
 - transition, 105, 114–115, 123
- Bandwidth, 156, 191
- Bartlett spectrum estimate, 145–154, 172, 241
- Boxcar data window, 68, 69, 71, 72, 114
- Boxcar function, 114
- Circular convolution, 102
- Coherency spectrum, 207–230, 236
 - confidence interval for, 221, 223
 - multiple, 237
 - of interest rate series, 208, 221
 - of interest rates and industrial production, 209, 223
 - realigned, 229
 - partial, 238
- Complex demodulation, 97–131
 - of sunspot series, 118–124
 - spectrum estimate, 154
- Complex sinusoid, 41
 - orthogonality of, 41, 110, 154
- Complex time series, 124–129
 - for sunspot series, 126–129
- Confidence interval
 - for coherency spectrum, 221, 223
 - for phase spectrum, 219, 221, 223, 224
 - for ratio of spectra, 225
 - for spectrum, 179, 183–188
- Continuous spectrum, 173
- Control problem, 234
- Convergence factor, 112
- Convolution, 71, 102
 - circular, 102
- Correlation, *see* Autocorrelation, Autocovariance, Cross covariance

- Cosine bell data window, 69, 71, 72, 76, 95
 Cosine wave, *see* Sinusoid
 Cosepectrum, 207, 215
 Covariance, *see* Autocorrelation, Auto-covariance, Cross covariance
 Cross covariance, 204
 theoretical, 211-214
 Cross periodogram, 202
 Cross spectral density function, 213, 214
 Cross spectrum, 203-230, 236
 theoretical, 211-214
 Cutoff frequency, 111
- Daniell spectrum estimate, 152, 156-166
 Data series (examples), *see* Industrial production, Interest rate, Ozone, Sunspot, Temperature, Variable star, Wheat price
 Data window, 66-76, 114, 119, 145-147, 152, 156, 181-183, 189, 236
 boxcar, 68, 69, 71, 72, 114
 cosine bell, 69, 71, 72, 76, 95
 Decimation, 198
 Degrees of freedom
 equivalent, 184
 Demodulation, complex, *see* Complex demodulation
 Density function, *see* Spectral density function
 Diffuseness, 47
 Digital filter, *see* Filter
 Dirichlet kernel, 12, 46, 69, 111, 114, 150
 Discrete Fourier transform, 40-43
 see also Fourier transform, Harmonic analysis
 Discrete spectral average, 153, 154
 Distribution function, *see* Spectral distribution function
 Durbin-Watson statistic, 53
 Dynamic range, 163, 244
- Ensemble, 175
 Entropy, 241
 Equivalent degrees of freedom, 184
 Ergodicity, 175
 Estimate, *see* Least squares estimate, Spectrum estimate
 Euler relation, 14, 41
- Fader, 66-76
 see also Data window, Taper
 Fast Fourier transform, 57-62, 147, 148, 153
 Filter, 100-118, 134, 135, 153, 155
 composite, 103
 construction, 103
 design, 105-118
 least squares, 110-118
 finite impulse response, 103, 110
 high-pass, 164
 low-pass, 105, 110-115
 recursive, 103
 span of, 105, 109
 spectrum of output, 171
 symmetric, 102
 transfer function of, 106
 weights, 101, 104, 134
 Fisher's test for periodicity, 91-95
 Folding frequency, 21
 see also Alias, Nyquist frequency
 Forecasting problem, 234
 Fourier analysis, *see* Harmonic analysis
 Fourier coefficient, 110, 116
 Fourier frequency, 20, 23, 37-40, 46, 51, 52, 57, 148
 Fourier transform
 discrete, 40-43
 effect of symmetry, 50
 fast, 57-62, 147, 148, 153
 linearity of, 43
 of impulse, 47
 of periodic series, 51
 of sinusoid, 45
 of smooth series, 53
 of step function, 48
 of straight line, 48
- Frequency
 cutoff, 111
 folding, 21
 see also Alias, Nyquist frequency
 Fourier, 23, 37-40, 46, 51, 52, 57
 fundamental, 52
 harmonic, 20, 52
 least squares estimate of, 26, 30, 31
 Nyquist, 21-23, 195-199
 of a sinusoid, 7
 pass, 105, 110
 sampling, 22
 stop, 105, 110
- Frequency resolution, 97
 Frequency response function, 103, 105
 Fundamental frequency, 52

- Gaussian noise, 87-89, 91, 176, 177, 185, 189
 Gibbs's phenomenon, 111, 117
 Group delay, 229
 Harmonic analysis, 2, 37-55
 examples, 63-96
 local, 97
 Harmonic frequency, 20, 52
 High-pass filter, 164
 Higher order periodogram, 238
 Higher order spectrum, 239
- Impulse**
 Fourier transform of, 47
Impulse response function, 103-105
Industrial production series, 204-226, 229-230
 alignment with interest rates, 229-230
 coherency spectrum with interest rates, 209, 223
 realigned, 229
 phase spectrum with interest rates, 209, 224
 realigned, 229
 spectrum estimate of, 204
Instantaneous amplitude, 118-124, 126, 130, 154
Instantaneous phase, 118-124, 130
 relative, 121, 202
Interest rate series, 204-226, 229-230
 alignment with industrial production, 229-230
 coherency spectrum of, 208, 221
 coherency spectrum with industrial production, 209, 223
 realigned, 229
 phase spectrum of, 208, 221
 phase spectrum with industrial production, 209, 224
 realigned, 229
 ratio of spectra, 225
 spectrum estimate of, 204
Interval, *see Confidence interval, Sampling interval*
Irregular component, 134, 138
Irregular observations, 245
 Joint nonnegative definiteness, 212
 Joint stationarity, 211-230
 Lag weights, 145, 146, 148-150, 152, 154, 164, 172
 Lag window, 180
 Leakage, 46, 62, 152, 156, 163
 Least squares, 10, 26
 estimate, 10-17
 of amplitude, 12
 of frequency, 26, 30, 31
 of phase, 12
 filter design, 110-118
 principle, 11
 Linear filter, *see Filter*
 Linear regression, 237
 Linearity of Fourier transform, 43
 Local harmonic analysis, 97
 Localization, 47
 Low-pass filter, 105, 110-115
 Maximum entropy spectrum, 241
 Maximum likelihood spectrum, 241
 Missing data, 242, 244
 Moving average
 simple, 100
 Moving average process, 174
 Multiple coherency, 237
 Multiple series, 236
 Noise, 100
 see also Gaussian noise, White noise
 Nonnegative definiteness, 168, 172
 joint, 212
 Nonquadratic spectrum estimate, 240
 Nyquist frequency, 21-23, 195-199
 Orthogonality
 of complex sinusoids, 41, 110, 154
 of sinusoids, 19, 37, 38
 Oscillation
 perturbed, 97
 Ozone series, 2, 9, 13-17
 seasonal cycle in, 9, 13-17
 Padding, 61, 148
 Partial coherency, 238
 Partial phase, 238
 Partial spectrum, 238
 Pass band, 105
 Pass frequency, 105, 110, 115, 123
 Period of a sinusoid, 7
 Periodic series
 Fourier transform of, 51
Periodogram, 42, 165, 167
 cross, 202
 higher order, 238
 of sunspot series, 76-86
 histogram, 89

- Periodogram (cont.)
 of temperature series, 92
 of variable star series, 63-65
 of wheat price series, 133-142
 of white noise, 87-91
 histogram, 88
- Perturbed oscillation, 97
- Phase
 instantaneous, 118-124, 130
 least squares estimate of, 12
 of a sinusoid, 8
 random, 170
 relative, 79-80, 203
 instantaneous, 121, 202
- Phase spectrum, 207-230, 236
 confidence interval for, 219, 221, 223, 224
 of interest rate series, 208, 221
 of interest rates and industrial production, 209, 224
 realigned, 229
 partial, 238
- Point process, 245
 spectrum of, 245
- Power transfer function, 102, 172
- Prewhitenning, 163-164, 245
- Principal alias, 21, 197
- Quadrature spectrum, 207, 215
- Random phase model, 170
- Random sinusoid model, 170, 173
- Realization, 175, 176
- Recursive filter, 103
- Regression
 linear, 237
- Relative phase, 79-80, 203
 instantaneous, 121, 202
- Reroughing, 104, 109
 multiplicative, 160
 spectrum estimate, 160-164
- Residual, 9-17, 23, 26-30
- Residual spectrum, 215, 237
- Resmoothing, 103, 109
- Resolution, 156, 160, 164, 191, 241
 frequency, 97
 time, 97
- Response function, *see* Frequency response function, Impulse response function
- Sampling
 irregular, 245
- Sampling frequency, 22
- Sampling interval, 195, 198
- Seasonal cycle
 in ozone series, 9, 13-17
- Segment, 140-145
- Sidelobe, 47-49, 69, 112-114, 151, 152, 156, 158, 192
- Signal, 100
- Simple moving average, 100
- Sine wave, *see* Sinusoid
- Sinusoid, 7, 8
 amplitude of, 8, 12
 complex, 41
 orthogonality of, 41, 110, 154
- Fourier transform of, 45
 frequency of, 7
 orthogonality of, 19, 37, 38
 period of, 7
 phase of, 8, 12
 random, 170, 173
 tapered, 99
- Smoothing, 100-105
- Smoothness
 effect on Fourier transform, 53
- Span
 of filter, 105, 109
 of spectral weights, 158
- Spatial series, 234-236
- Spectral density function, 168-177, 182, 191, 196-198, 213-215
 cross, 213, 214
 residual, 215, 237
- Spectral distribution function, 168, 169
- Spectral weights, 153, 157, 158
- Spectral window, 150, 155, 156, 179, 236
- Spectrum, 142, 167
 aliased, 195-199
 autoregressive, 240
 co-, 207, 215
 confidence interval
 for ratio, 225
 confidence interval for, 179, 183-188
 continuous, 173
 cross, 203-230, 236
 higher order, 239
 maximum entropy, 241
 maximum likelihood, 241
 partial, 238
 quadrature, 207, 215
 ratio, 224-226
 confidence interval for, 225
 for interest rate series, 225
 residual, 215, 237

- Spectrum (cont.)
 theoretical, 167, 168
see also Coherency spectrum, Phase spectrum
- Spectrum estimate, 147
 bandwidth, 156
 Bartlett, 145-154, 172, 241
 complex demodulation, 154
 Daniell, 152, 156-166
 discrete average, 153, 154
 leakage, 156
 nonquadratic, 240
 of industrial production series, 204
 of interest rate series, 204
 of sunspot series, 158-159, 187
 reroughed, 161-162
 of wheat price series, 142, 146, 150,
 157-158, 172, 185
 reroughed, 161
 reroughed, 160-164
 of sunspot series, 161-162
 of wheat price series, 161
 resolution, 156
 smoothness, 156
 stability, 156
 Stability of a spectrum estimate, 156
 Stationary time series, 167-173
 joint, 211-230
 Step function
 Fourier transform of, 48
 Stop band, 105
 Stop frequency, 105, 110, 115, 123
 Straight line
 Fourier transform of, 48
 Sunspot series, 4, 6, 97, 99, 108, 133,
 158, 159, 183, 187, 188
 complex demodulation of, 118-124
 complex time series for, 126-129
 periodogram of, 76-86
 histogram, 89
 spectrum estimate of, 158-159, 187
 reroughed, 161-162
 Superposability of Fourier transform, 43
 Symmetric filter, 102
 transfer function of, 106
 Symmetry
 effect on Fourier transform, 50
- Taper, 66-76, 90, 91, 95, 119, 126, 145,
 181-183, 189
 Tapered sinusoid, 99
 Temperature series, 92-95
 periodogram of, 92
 Theoretical autocovariance, 168
 Theoretical cross covariance, 211-214
 Theoretical cross spectrum, 211-214
 Theoretical spectrum, 167, 168
 Time domain analysis, 233, 234
 Time origin, 8
 Time resolution, 97
 Time series
 examples, *see* Industrial production,
 Interest rate, Ozone, Sunspot,
 Temperature, Variable star, Wheat
 price
 stationary, 167-173
 joint, 211-230
 Transfer function, 102, 153, 155, 172
 ideal, 110
 inversion, 103, 106
 of symmetric filter, 106
 power, 102, 172
 Transform, *see* Fourier transform
 Transition band, 105, 114-115, 123
 Trend component, 134, 138
 Truncation point, 146, 147, 152, 153
- Variable star series, 3, 25-32, 141
 periodogram of, 63-65
 tapering, 72-76
 Von Neumann ratio, 53
- Wheat price series, 4, 6, 133-139, 183,
 186, 188
 periodogram of, 133-142
 spectrum estimate of, 142, 146, 150,
 157-158, 172, 185
 reroughed, 161
 White noise, 88, 89, 91, 171, 173-177,
 185, 189
 periodogram of, 87-91
 histogram, 88
 Window, *see* Data window, Lag window,
 Spectral window

WILEY SERIES IN PROBABILITY AND STATISTICS
ESTABLISHED BY WALTER A. SHEWHART AND SAMUEL S. WILKS

Editors

*Vic Barnett, Noel A. C. Cressie, Nicholas I. Fisher,
Iain M. Johnstone, J. B. Kadane, David G. Kendall, David W. Scott,
Bernard W. Silverman, Adrian F. M. Smith, Jozef L. Teugels;
Ralph A. Bradley, Emeritus, J. Stuart Hunter, Emeritus*

Probability and Statistics Section

- *ANDERSON · The Statistical Analysis of Time Series
ARNOLD, BALAKRISHNAN, and NAGARAJA · A First Course in Order Statistics
ARNOLD, BALAKRISHNAN, and NAGARAJA · Records
BACCELLI, COHEN, OLSDER, and QUADRAT · Synchronization and Linearity:
An Algebra for Discrete Event Systems
BASILEVSKY · Statistical Factor Analysis and Related Methods: Theory and
Applications
BERNARDO and SMITH · Bayesian Statistical Concepts and Theory
BILLINGSLEY · Convergence of Probability Measures, *Second Edition*
BOROVKOV · Asymptotic Methods in Queuing Theory
BOROVKOV · Ergodicity and Stability of Stochastic Processes
BRANDT, FRANKEN, and LISEK · Stationary Stochastic Models
CAINES · Linear Stochastic Systems
CAIROLI and DALANG · Sequential Stochastic Optimization
CONSTANTINE · Combinatorial Theory and Statistical Design
COOK · Regression Graphics
COVER and THOMAS · Elements of Information Theory
CSÖRGÖ and HORVÁTH · Weighted Approximations in Probability Statistics
CSÖRGÖ and HORVÁTH · Limit Theorems in Change Point Analysis
DETTE and STUDDEN · The Theory of Canonical Moments with Applications in
Statistics, Probability, and Analysis
DEY and MUKERJEE · Fractional Factorial Plans
*DOOB · Stochastic Processes
DRYDEN and MARDIA · Statistical Analysis of Shape
DUPUIS and ELLIS · A Weak Convergence Approach to the Theory of Large Deviations
ETHIER and KURTZ · Markov Processes: Characterization and Convergence
FELLER · An Introduction to Probability Theory and Its Applications, Volume 1,
Third Edition, Revised; Volume II, Second Edition
FULLER · Introduction to Statistical Time Series, *Second Edition*
FULLER · Measurement Error Models
GHOSH, MUKHOPADHYAY, and SEN · Sequential Estimation
GIFI · Nonlinear Multivariate Analysis
GUTTORP · Statistical Inference for Branching Processes
HALL · Introduction to the Theory of Coverage Processes
HAMPEL · Robust Statistics: The Approach Based on Influence Functions
HANNAN and DEISTLER · The Statistical Theory of Linear Systems
HUBER · Robust Statistics
IMAN and CONOVER · A Modern Approach to Statistics
JUREK and MASON · Operator-Limit Distributions in Probability Theory
KASS and VOS · Geometrical Foundations of Asymptotic Inference

*Now available in a lower priced paperback edition in the Wiley Classics Library.

Probability and Statistics (Continued)

- KAUFMAN and ROUSSEEUW · Finding Groups in Data: An Introduction to Cluster Analysis
KELLY · Probability, Statistics, and Optimization
LINDVALL · Lectures on the Coupling Method
McFADDEN · Management of Data in Clinical Trials
MANTON, WOODBURY, and TOLLEY · Statistical Applications Using Fuzzy Sets
MORGENTHALER and TUKEY · Configural Polysampling: A Route to Practical Robustness
MUIRHEAD · Aspects of Multivariate Statistical Theory
OLIVER and SMITH · Influence Diagrams, Belief Nets and Decision Analysis
*PARZEN · Modern Probability Theory and Its Applications
PRESS · Bayesian Statistics: Principles, Models, and Applications
PUKELSHEIM · Optimal Experimental Design
RAO · Asymptotic Theory of Statistical Inference
RAO · Linear Statistical Inference and Its Applications, *Second Edition*
RAO and SHANBHAG · Choquet-Deny Type Functional Equations with Applications to Stochastic Models
ROBERTSON, WRIGHT, and DYKSTRA · Order Restricted Statistical Inference
ROGERS and WILLIAMS · Diffusions, Markov Processes, and Martingales, Volume I: Foundations, *Second Edition*; Volume II: Itô Calculus
RUBINSTEIN and SHAPIRO · Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method
RUZSA and SZEKELY · Algebraic Probability Theory
SCHEFFE · The Analysis of Variance
SEBER · Linear Regression Analysis
SEBER · Multivariate Observations
SEBER and WILD · Nonlinear Regression
SERFLING · Approximation Theorems of Mathematical Statistics
SHORACK and WELLNER · Empirical Processes with Applications to Statistics
SMALL and McLEISH · Hilbert Space Methods in Probability and Statistical Inference
STAPLETON · Linear Statistical Models
STAUDTE and SHEATHER · Robust Estimation and Testing
STOYANOV · Counterexamples in Probability
TANAKA · Time Series Analysis: Nonstationary and Noninvertible Distribution Theory
THOMPSON and SEBER · Adaptive Sampling
WELSH · Aspects of Statistical Inference
WHITTAKER · Graphical Models in Applied Multivariate Statistics
YANG · The Construction Theory of Denumerable Markov Processes

Applied Probability and Statistics Section

- ABRAHAM and LEDOLTER · Statistical Methods for Forecasting
AGRESTI · Analysis of Ordinal Categorical Data
AGRESTI · Categorical Data Analysis
ANDERSSON, AUQUIER, HAUCK, OAKES, VANDAELE, and WEISBERG · Statistical Methods for Comparative Studies
ARMITAGE and DAVID (editors) · Advances in Biometry
*ARTHANARI and DODGE · Mathematical Programming in Statistics
ASMUSSEN · Applied Probability and Queues
*BAILEY · The Elements of Stochastic Processes with Applications to the Natural Sciences
BARNETT and LEWIS · Outliers in Statistical Data, *Third Edition*

*Now available in a lower priced paperback edition in the Wiley Classics Library.

Applied Probability and Statistics (Continued)

- BARTHOLOMEW, FORBES, and MCLEAN · Statistical Techniques for Manpower Planning, *Second Edition*
- BATES and WATTS · Nonlinear Regression Analysis and Its Applications
- BECHHOFER, SANTNER, and GOLDSMAN · Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons
- BELSLEY · Conditioning Diagnostics: Collinearity and Weak Data in Regression
- BELSLEY, KUH, and WELSCH · Regression Diagnostics: Identifying Influential Data and Sources of Collinearity
- BHAT · Elements of Applied Stochastic Processes, *Second Edition*
- BHATTACHARYA and WAYMIRE · Stochastic Processes with Applications
- BIRKES and DODGE · Alternative Methods of Regression
- BLOOMFIELD · Fourier Analysis of Time Series: An Introduction, *Second Edition*
- BOLLEN · Structural Equations with Latent Variables
- BOULEAU · Numerical Methods for Stochastic Processes
- BOX · Bayesian Inference in Statistical Analysis
- BOX and DRAPER · Empirical Model-Building and Response Surfaces
- BOX and DRAPER · Evolutionary Operation: A Statistical Method for Process Improvement
- BUCKLEW · Large Deviation Techniques in Decision, Simulation, and Estimation
- BUNKE and BUNKE · Nonlinear Regression, Functional Relations and Robust Methods: Statistical Methods of Model Building
- CHATTERJEE and HADI · Sensitivity Analysis in Linear Regression
- CHERNICK · Bootstrap Methods: A Practitioner's Guide
- CHILÈS and DELFINER · Geostatistics: Modeling Spatial Uncertainty
- CHOW and LIU · Design and Analysis of Clinical Trials: Concepts and Methodologies
- CLARKE and DISNEY · Probability and Random Processes: A First Course with Applications, *Second Edition*
- *COCHRAN and COX · Experimental Designs, *Second Edition*
- CONOVER · Practical Nonparametric Statistics, *Second Edition*
- CORNELL · Experiments with Mixtures, Designs, Models, and the Analysis of Mixture Data, *Second Edition*
- *COX · Planning of Experiments
- CRESSIE · Statistics for Spatial Data, *Revised Edition*
- DANIEL · Applications of Statistics to Industrial Experimentation
- DANIEL · Biostatistics: A Foundation for Analysis in the Health Sciences, *Sixth Edition*
- DAVID · Order Statistics, *Second Edition*
- *DEGROOT, FIENBERG, and KADANE · Statistics and the Law
- DODGE · Alternative Methods of Regression
- DOWDY and WEARDEN · Statistics for Research, *Second Edition*
- DRYDEN and MARDIA · Statistical Shape Analysis
- DUNN and CLARK · Applied Statistics: Analysis of Variance and Regression, *Second Edition*
- ELANDT-JOHNSON and JOHNSON · Survival Models and Data Analysis
- EVANS, PEACOCK, and HASTINGS · Statistical Distributions, *Second Edition*
- FLEISS · The Design and Analysis of Clinical Experiments
- FLEISS · Statistical Methods for Rates and Proportions, *Second Edition*
- FLEMING and HARRINGTON · Counting Processes and Survival Analysis
- GALLANT · Nonlinear Statistical Models
- GLASSERMAN and YAO · Monotone Structure in Discrete-Event Systems
- GNANADESIKAN · Methods for Statistical Data Analysis of Multivariate Observations, *Second Edition*
- GOLDSTEIN and LEWIS · Assessment: Problems, Development, and Statistical Issues
- GREENWOOD and NIKULIN · A Guide to Chi-Squared Testing

*Now available in a lower priced paperback edition in the Wiley Classics Library.

Applied Probability and Statistics (Continued)

*HAHN · Statistical Models in Engineering

HAHN and MEEKER · Statistical Intervals: A Guide for Practitioners

HAND · Construction and Assessment of Classification Rules

HAND · Discrimination and Classification

HEIBERGER · Computation for the Analysis of Designed Experiments

HEDAYAT and SINHA · Design and Inference in Finite Population Sampling

HINKELMAN and KEMPTHORNE · Design and Analysis of Experiments, Volume 1:
Introduction to Experimental Design

HOAGLIN, MOSTELLER, and TUKEY · Exploratory Approach to Analysis of Variance

HOAGLIN, MOSTELLER, and TUKEY · Exploring Data Tables, Trends and Shapes

HOAGLIN, MOSTELLER, and TUKEY · Understanding Robust and Exploratory
Data Analysis

HOCHBERG and TAMHANE · Multiple Comparison Procedures

HOCKING · Methods and Applications of Linear Models: Regression and the Analysis
of Variables

HOGG and KLUGMAN · Loss Distributions

HOSMER and LEMESHOW · Applied Logistic Regression

HØYLAND and RAUSAND · System Reliability Theory: Models and Statistical Methods

HUBERTY · Applied Discriminant Analysis

JACKSON · A User's Guide to Principle Components

JOHN · Statistical Methods in Engineering and Quality Assurance

JOHNSON · Multivariate Statistical Simulation

JOHNSON and KOTZ · Distributions in Statistics

Continuous Multivariate Distributions

JOHNSON, KOTZ, and BALAKRISHNAN · Continuous Univariate Distributions,
Volume 1, Second Edition

JOHNSON, KOTZ, and BALAKRISHNAN · Continuous Univariate Distributions,
Volume 2, Second Edition

JOHNSON, KOTZ, and BALAKRISHNAN · Discrete Multivariate Distributions

JOHNSON, KOTZ, and KEMP · Univariate Discrete Distributions, *Second Edition*

JUREČKOVÁ and SEN · Robust Statistical Procedures: Asymptotics and Interrelations

KADANE · Bayesian Methods and Ethics in a Clinical Trial Design

KADANE AND SCHUM · A Probabilistic Analysis of the Sacco and Vanzetti Evidence

KALBFLEISCH and PRENTICE · The Statistical Analysis of Failure Time Data

KELLY · Reversability and Stochastic Networks

KHURI, MATHEW, and SINHA · Statistical Tests for Mixed Linear Models

KLUGMAN, PANJER, and WILLMOT · Loss Models: From Data to Decisions

KLUGMAN, PANJER, and WILLMOT · Solutions Manual to Accompany Loss Models:
From Data to Decisions

KOVALENKO, KUZNETZOV, and PEGG · Mathematical Theory of Reliability of
Time-Dependent Systems with Practical Applications

LAD · Operational Subjective Statistical Methods: A Mathematical, Philosophical, and
Historical Introduction

LANGE, RYAN, BILLARD, BRILLINGER, CONQUEST, and GREENHOUSE ·
Case Studies in Biometry

LAWLESS · Statistical Models and Methods for Lifetime Data

LEE · Statistical Methods for Survival Data Analysis, *Second Edition*

LEPAGE and BILLARD · Exploring the Limits of Bootstrap

LINHART and ZUCCHINI · Model Selection

LITTLE and RUBIN · Statistical Analysis with Missing Data

LLOYD · The Statistical Analysis of Categorical Data

MAGNUS and NEUDECKER · Matrix Differential Calculus with Applications in
Statistics and Econometrics

*Now available in a lower priced paperback edition in the Wiley Classics Library.

Applied Probability and Statistics (Continued)

- MALLER and ZHOU · Survival Analysis with Long Term Survivors
MANN, SCHAFER, and SINGPURWALLA · Methods for Statistical Analysis of Reliability and Life Data
MC LACHLAN and KRISHNAN · The EM Algorithm and Extensions
MC LACHLAN · Discriminant Analysis and Statistical Pattern Recognition
MC NEIL · Epidemiological Research Methods
MEEKER and ESCOBAR · Statistical Methods for Reliability Data
MILLER · Survival Analysis
MONTGOMERY and PECK · Introduction to Linear Regression Analysis, *Second Edition*
MYERS and MONTGOMERY · Response Surface Methodology: Process and Product in Optimization Using Designed Experiments
NELSON · Accelerated Testing, Statistical Models, Test Plans, and Data Analyses
NELSON · Applied Life Data Analysis
OCHI · Applied Probability and Stochastic Processes in Engineering and Physical Sciences
OKABE, BOOTS, and SUGIHARA · Spatial Tesselations: Concepts and Applications of Voronoi Diagrams
PANKRATZ · Forecasting with Dynamic Regression Models
PANKRATZ · Forecasting with Univariate Box-Jenkins Models: Concepts and Cases
PIANTADOSI · Clinical Trials: A Methodologic Perspective
PORT · Theoretical Probability for Applications
PUTERMAN · Markov Decision Processes: Discrete Stochastic Dynamic Programming
RACHEV · Probability Metrics and the Stability of Stochastic Models
RÉNYI · A Diary on Information Theory
RIPLEY · Spatial Statistics
RIPLEY · Stochastic Simulation
ROUSSEEUW and LEROY · Robust Regression and Outlier Detection
RUBIN · Multiple Imputation for Nonresponse in Surveys
RUBINSTEIN · Simulation and the Monte Carlo Method
RUBINSTEIN and MELAMED · Modern Simulation and Modeling
RYAN · Statistical Methods for Quality Improvement
SCHUSS · Theory and Applications of Stochastic Differential Equations
SCOTT · Multivariate Density Estimation: Theory, Practice, and Visualization
*SEARLE · Linear Models
SEARLE · Linear Models for Unbalanced Data
SEARLE, CASELLA, and McCULLOCH · Variance Components
SENNOTT · Stochastic Dynamic Programming and the Control of Queueing Systems
STOYAN, KENDALL, and MECKE · Stochastic Geometry and Its Applications, *Second Edition*
STOYAN and STOYAN · Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics
THOMPSON · Empirical Model Building
THOMPSON · Sampling
THOMPSON · Simulation: A Modeler's Approach
TIJMS · Stochastic Modeling and Analysis: A Computational Approach
TIJMS · Stochastic Models: An Algorithmic Approach
TITTERINGTON, SMITH, and MAKOV · Statistical Analysis of Finite Mixture Distributions
UPTON and FINGLETON · Spatial Data Analysis by Example, Volume I: Point Pattern and Quantitative Data
UPTON and FINGLETON · Spatial Data Analysis by Example, Volume II: Categorical and Directional Data
VAN RIJCKEVORSEL and DE LEEUW · Component and Correspondence Analysis

*Now available in a lower priced paperback edition in the Wiley Classics Library.

Applied Probability and Statistics (Continued)

- VIDAKOVIC · Statistical Modeling by Wavelets
WEISBERG · Applied Linear Regression, *Second Edition*
WESTFALL and YOUNG · Resampling-Based Multiple Testing: Examples and Methods for *p*-Value Adjustment
WHITTLE · Systems in Stochastic Equilibrium
WOODING · Planning Pharmaceutical Clinical Trials: Basic Statistical Principles
WOOLSON · Statistical Methods for the Analysis of Biomedical Data
*ZELLNER · An Introduction to Bayesian Inference in Econometrics

Texts and References Section

- AGRESTI · An Introduction to Categorical Data Analysis
ANDERSON · An Introduction to Multivariate Statistical Analysis, *Second Edition*
ANDERSON and LOYNES · The Teaching of Practical Statistics
ARMITAGE and COLTON · Encyclopedia of Biostatistics: Volumes 1 to 6 with Index
BARTOSZYNSKI and NIEWIADOMSKA-BUGAJ · Probability and Statistical Inference
BERRY, CHALONER, and GEWEKE · Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner
BHATTACHARYA and JOHNSON · Statistical Concepts and Methods
BILLINGSLEY · Probability and Measure, *Second Edition*
BOX · R. A. Fisher, the Life of a Scientist
BOX, HUNTER, and HUNTER · Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building
BOX and LUCEÑO · Statistical Control by Monitoring and Feedback Adjustment
BROWN and HOLLANDER · Statistics: A Biomedical Introduction
CHATTERJEE and PRICE · Regression Analysis by Example, *Third Edition*
COOK and WEISBERG · Applied Regression Including Computing and Graphics
COOK and WEISBERG · An Introduction to Regression Graphics
COX · A Handbook of Introductory Statistical Methods
DILLON and GOLDSTEIN · Multivariate Analysis: Methods and Applications
DODGE and ROMIG · Sampling Inspection Tables, *Second Edition*
DRAPER and SMITH · Applied Regression Analysis, *Third Edition*
DUDEWICZ and MISHRA · Modern Mathematical Statistics
DUNN · Basic Statistics: A Primer for the Biomedical Sciences, *Second Edition*
FISHER and VAN BELLE · Biostatistics: A Methodology for the Health Sciences
FREEMAN and SMITH · Aspects of Uncertainty: A Tribute to D. V. Lindley
GROSS and HARRIS · Fundamentals of Queueing Theory, *Third Edition*
HALD · A History of Probability and Statistics and their Applications Before 1750
HALD · A History of Mathematical Statistics from 1750 to 1930
HELLER · MACSYMA for Statisticians
HOEL · Introduction to Mathematical Statistics, *Fifth Edition*
HOLLANDER and WOLFE · Nonparametric Statistical Methods, *Second Edition*
HOSMER and LEMESHOW · Applied Survival Analysis: Regression Modeling of Time to Event Data
JOHNSON and BALAKRISHNAN · Advances in the Theory and Practice of Statistics: A Volume in Honor of Samuel Kotz
JOHNSON and KOTZ (editors) · Leading Personalities in Statistical Sciences: From the Seventeenth Century to the Present
JUDGE, GRIFFITHS, HILL, LÜTKEPOHL, and LEE · The Theory and Practice of Econometrics, *Second Edition*
KHURI · Advanced Calculus with Applications in Statistics
KOTZ and JOHNSON (editors) · Encyclopedia of Statistical Sciences: Volumes 1 to 9 with Index

*Now available in a lower priced paperback edition in the Wiley Classics Library.

Texts and References (Continued)

- KOTZ and JOHNSON (editors) · Encyclopedia of Statistical Sciences: Supplement
Volume
- KOTZ, REED, and BANKS (editors) · Encyclopedia of Statistical Sciences: Update
Volume 1
- KOTZ, REED, and BANKS (editors) · Encyclopedia of Statistical Sciences: Update
Volume 2
- LAMPERTI · Probability: A Survey of the Mathematical Theory, *Second Edition*
- LARSON · Introduction to Probability Theory and Statistical Inference, *Third Edition*
- LE · Applied Categorical Data Analysis
- LE · Applied Survival Analysis
- MALLOWS · Design, Data, and Analysis by Some Friends of Cuthbert Daniel
- MARDIA · The Art of Statistical Science: A Tribute to G. S. Watson
- MASON, GUNST, and HESS · Statistical Design and Analysis of Experiments with
Applications to Engineering and Science
- MURRAY · X-STAT 2.0 Statistical Experimentation, Design Data Analysis, and
Nonlinear Optimization
- PURI, VILAPLANA, and WERTZ · New Perspectives in Theoretical and Applied
Statistics
- RENCHER · Methods of Multivariate Analysis
- RENCHER · Multivariate Statistical Inference with Applications
- RENCHER · Linear Models in Statistics
- ROSS · Introduction to Probability and Statistics for Engineers and Scientists
- ROHATGI · An Introduction to Probability Theory and Mathematical Statistics
- RYAN · Modern Regression Methods
- SCHOTT · Matrix Analysis for Statistics
- SEARLE · Matrix Algebra Useful for Statistics
- STYAN · The Collected Papers of T. W. Anderson: 1943–1985
- TIERNEY · LISP-STAT: An Object-Oriented Environment for Statistical Computing
and Dynamic Graphics
- WONNACOTT and WONNACOTT · Econometrics, *Second Edition*

WILEY SERIES IN PROBABILITY AND STATISTICS
ESTABLISHED BY WALTER A. SHEWHART AND SAMUEL S. WILKS

Editors

*Robert M. Groves, Graham Kalton, J. N. K. Rao, Norbert Schwarz,
Christopher Skinner*

Survey Methodology Section

- BIEMER, GROVES, LYBERG, MATHIOWETZ, and SUDMAN · Measurement
Errors in Surveys
- COCHRAN · Sampling Techniques, *Third Edition*
- COUPER, BAKER, BETHLEHEM, CLARK, MARTIN, NICHOLLS, and O'REILLY
(editors) · Computer Assisted Survey Information Collection
- COX, BINDER, CHINNAPPA, CHRISTIANSON, COLLEDGE, and KOTT (editors) ·
Business Survey Methods
- *DEMING · Sample Design in Business Research

*Now available in a lower priced paperback edition in the Wiley Classics Library.

Survey Methodology (Continued)

DILLMAN · Mail and Telephone Surveys: The Total Design Method

GROVES and COUPER · Nonresponse in Household Interview Surveys

GROVES · Survey Errors and Survey Costs

GROVES, BIEMER, LYBERG, MASSEY, NICHOLLS, and WAKSBERG ·

Telephone Survey Methodology

*HANSEN, HURWITZ, and MADOW · Sample Survey Methods and Theory,
Volume 1: Methods and Applications

*HANSEN, HURWITZ, and MADOW · Sample Survey Methods and Theory,
Volume II: Theory

KISH · Statistical Design for Research

*KISH · Survey Sampling

KORN and GRAUBARD · Analysis of Health Surveys

LESSLER and KALSBEEK · Nonsampling Error in Surveys

LEVY and LEMESHOW · Sampling of Populations: Methods and Applications,

Third Edition

LYBERG, BIEMER, COLLINS, de LEEUW, DIPPO, SCHWARZ, TREWIN (editors) ·

Survey Measurement and Process Quality

SIRKEN, HERRMANN, SCHECHTER, SCHWARZ, TANUR, and TOURANGEAU

(editors) · Cognition and Survey Research

*Now available in a lower priced paperback edition in the Wiley Classics Library.