ON EMPIRICAL MULTIPLE
TIME SERIES ANALYSIS

EMANUEL PARZEN
STANFORD UNIVERSITY

1. Introduction

Like probability theory, modern time series analysis has the feature that many of its most elementary theorems are based on rather deep mathematics, while many of its most advanced theorems are known and understood by research workers who do not have the mathematical background to understand the proofs. It is natural to think of the theory of time series analysis as composed of two parts, *foundations* (a probabilistic part involving deep mathematics and based on the unrealistic assumption that one knows the probability law of the time series) and *empirical* (in which one considers statistical and computational procedures). While the probabilistic theory of time series can be pursued for the sake of its great beauty, it would be a mistake if the statistical theory were to be developed only for its elegance. The ultimate aim of the statistical theory of time series analysis must be to provide *data-handling procedures* for achieving the aim of time series analysis, *synthesis of stochastic models* which can be used to describe and perhaps to control the mechanisms generating each time series and relating various time series. For this reason, one may define a field which may be called "empirical time series analysis" with aims such as the following:

(1) to develop the statistical theory in such a way that it provides a philosophy for judging and interpreting the statistical data reduction which can be provided by computers;

(2) to develop efficient computer programs for the statistical analysis of empirical time series;

(3) to obtain experience in the small sample applicability and robustness of statistical procedures derived from asymptotic theory;

(4) to focus attention on theoretical questions requiring further investigation.

One of my concerns in recent years has been to develop a computer program for empirical time series analysis. There were several reasons motivating this concern:

(1) I discovered that when a researcher came to me for advice on time series analysis, I could do him the most good by (in addition to telling him which formulas to use) making available to him a computer program for carrying out the analysis.

(2) I was curious to see if there were any truth to the proposition that a statistician interested in data analysis need not be interested in theorems, since experience with computer output would provide all the insights he needs; my experience leads me to conclude that a knowledge of relevant theorems is indispensable if one desires to be able to interpret as many features of the computer output as possible.

(3) I desired to develop an approach to empirical time series analysis. Before describing this approach, let me quote some recent remarks of John Tukey (who in my view is "The Father of Empirical Time Series Analysis"). Tukey notes ([30], p. 1284) that, "It is a commonplace of science that where one can, one learns faster by deliberately reaching in and changing something, by seeing what happens when something is varied in a controlled way." Unfortunately, time series often arise in the field sciences rather than in the laboratory sciences; and it is nature rather than the observer who determines the conditions under which the data will be observed. Nevertheless, quoting Tukey again, "How can at least some of the advantages of reaching in be had when one can only sit and look?" One important answer is given by Tukey, "The answer is simple and well known: look in two [or more] places and try to assess the relationship of the things observed." He sums up this point of view in the maxim, "Look here, look there, compare, and interrelate." I believe there is another way in which to compare and interrelate; this is by varying the way in which one analyzes the data. One should consider a variety of models for the observed time series. For each model one should estimate the parameters which represent the incompletely specified characteristics of the probability law. Comparing the analyses often provides rough tests of hypotheses concerning which model provides a better fit to the data.

As in statistics, so in time series analysis, one may distinguish three main problems:

(1) *estimation* of the parameters of a given model for the observed time series (in particular, in any field where the properties of the phenomenon being studied can be characterized in terms of its behavior in the frequency domain, one needs to estimate spectral density functions and other spectral characteristics associated with stationary multiple time series);

(2) *hypothesis testing* and hypothesis suggesting (testing the fit of various models and suggesting possible models to fit); and

(3) *description* (to provide measurements about a phenomenon, which together with other kinds of measurements, represent the observational regularities which it is a purpose of any theory of the phenomenon to explain).

Some techniques for fitting models to single time series have been discussed in previous papers (see Parzen [22], [24]). Fitting models to multiple time series seems a much harder problem. While cross-spectral analysis is clearly one of the main tools for fitting models to multiple time series, it is not yet clear what are the sample cross-spectral functions which should be routinely computed. To an observed sample of a multiple time series one can associate a bewildering array

of cross-spectral density quantities involving such adjectives as "co-spectral," "quadrature-spectral," "partial cross-spectral," and such nouns as "amplitude," "phase," "coherence," and "gain."

The aim of this paper is to sketch a unified exposition of cross-spectral analysis. The exposition is not entirely rigorous, but rather attempts to indicate the theoretical questions which require further investigation. Among the results believed to be new are those in the section on the asymptotic sampling theory of partial cross-spectra. Rigorous proofs of all results stated are given in the Stanford Ph.D. thesis of Grace Wahba, June 1966.

Applied statisticians, actually computing sample spectra, often complain that papers written on spectral analysis are highly mathematical and offer no guide on how to proceed in practice. I am willing to grant some merit to this complaint in general. While this paper is not by itself a guide to how to proceed in practice, I hope that it will be of value as a discussion of some of the main mathematical considerations which need to be borne in mind in order to interpret sample cross-spectra. Excellent introductions to general considerations in empirical time series analysis are given by Jenkins [12] and Tukey [29].

2. Sample cross-spectra

Observed time series come in a variety of shapes. Economic and social time series often have the typical shapes shown in figure 1. In analyzing observed time series, I have found it valuable to distinguish two consecutive stages: (i) time series transformation and detrending, and (ii) correlation and spectral computations. In forming sample correlations and covariances, one should not automatically subtract out sample means (or fitted straight lines, and so on); any such subtractions should be done in the time series transformation and detrending stage. Consequently, given finite samples of r real time series

$$(2.1) \qquad \{X_1(t), t = 1, 2, \cdots, T\}, \cdots, \{X_r(t), t = 1, 2, \cdots, T\},$$

we make the following definitions.

The sample cross-covariance $R_{hj;T}(v)$ of lag v between $X_h(\cdot)$ and $X_j(\cdot)$ is defined to be

$$R_{hj;T}(v) = \frac{1}{T} \sum_{t=1}^{T-v} X_h(t) X_j(t+v) \qquad \text{for} \quad v = 0, 1, \cdots, T-1,$$

$$(2.2) \qquad R_{hj;T}(v) = \frac{1}{T} \sum_{t=-v+1}^{T} X_h(t) X_j(t+v) \quad \text{for} \quad v = -1, -2, \cdots, -(T-1),$$

$$R_{hj;T}(v) = 0 \qquad\qquad\qquad\qquad\qquad \text{otherwise.}$$

Since

$$(2.3) \qquad R_{jh;T}(v) = \frac{1}{T} \sum_{t=1}^{T-v} X_j(t) X_h(t+v) = \frac{1}{T} \sum_{s=v+1}^{T} X_h(t) X_j(s-v) = R_{hj;T}(-v),$$

it suffices to compute all the cross-covariances for positive lags v to know them for all v.

The sample cross-spectral density function between $X_h(\cdot)$ and $X_j(\cdot)$ is defined by (writing i for $\sqrt{-1}$)

$$(2.4) \qquad f_{hj;T}(\omega) = \frac{1}{2\pi T} \sum_{s=1}^{T} e^{i\omega s} X_h(s) \sum_{t=1}^{T} e^{-i\omega t} X_j(t).$$

It may be verified that these quantities form a pair of Fourier transforms:

$$(2.5) \qquad R_{hj;T}(v) = \int_{-\pi}^{\pi} e^{i\omega v} f_{hj;T}(\omega) \, d\omega,$$

$$(2.6) \qquad f_{hj;T}(\omega) = \frac{1}{2\pi} \sum_{|v|<T} e^{-iv\omega} R_{hj;T}(v).$$

The sample cross-correlation function $\rho_{hj;T}(v)$ is defined by

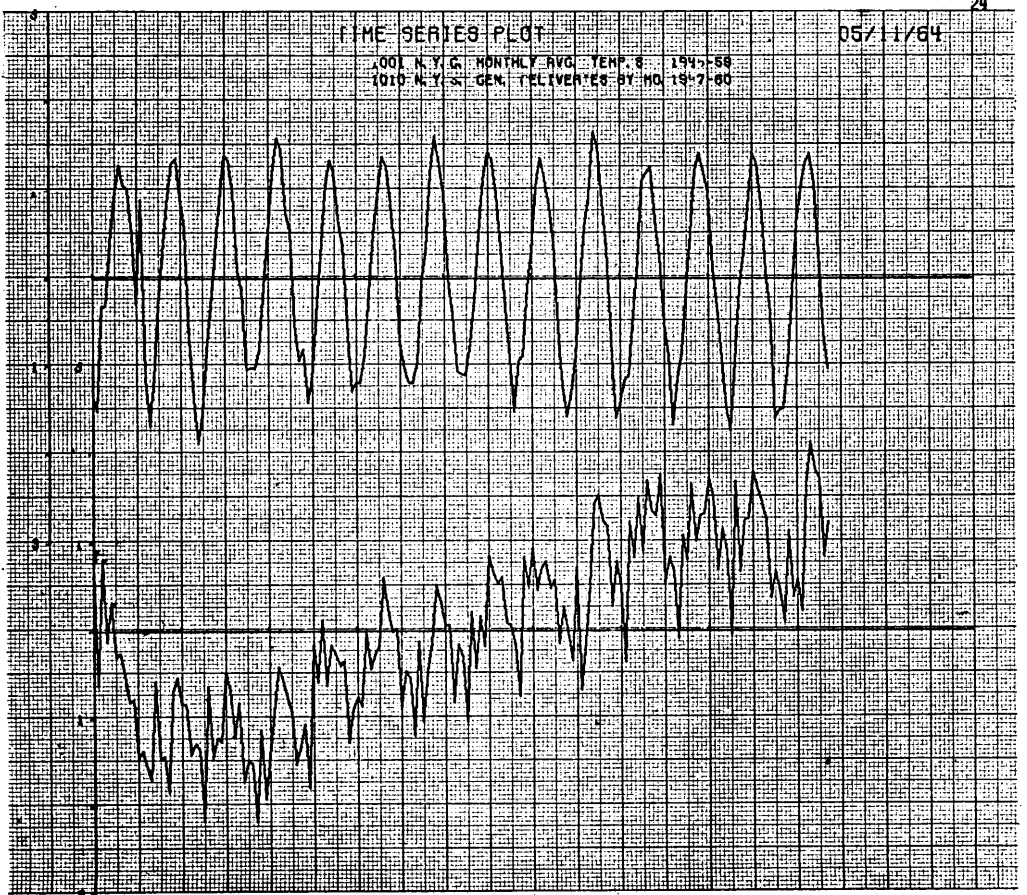$$(2.7) \qquad \rho_{hj;T}(v) = R_{hj;T}(v) \div \{R_{hh;T}(0) R_{jj;T}(0)\}^{1/2}.$$

FIGURE 1

The normalized sample spectral density function $\bar{f}_{hj;T}(\omega)$ is defined by

$$(2.8) \qquad \bar{f}_{hj;T}(\omega) = f_{hj;T}(\omega) \div \{R_{hh;T}(0)R_{jj;T}(0)\}^{1/2}.$$

The sample cross-spectral density function is generally complex-valued. The following notation and terminology, due to Tukey, is used to describe the real and negative imaginary parts of $f_{hj;T}(\omega)$:

$$(2.9) \qquad c_{hj;T}(\omega) = \mathrm{Re} f_{hj;T}(\omega), \quad \textit{sample co-spectral density},$$

$$q_{hj;T}(\omega) = -\mathrm{Im} f_{hj;T}(\omega), \quad \textit{sample quadrature spectral density}.$$

In the foregoing definitions, we are guided by the idea that when the observed time series are *zero mean* covariance stationary time series, the sample cross-spectral quantities should provide sample versions of corresponding population cross-spectra. However, these quantities can be defined for *any* sample, and their statistical characteristics can be investigated for *any* model that one may want to consider for an observed set of time series. Consequently, one can interpret sample cross-spectra without necessarily making the assumption of zero mean covariance stationarity.

One important class of models for time series for which one desires to understand the properties of sample cross-spectra is the following: for $j = 1, 2, \cdots, r$

$$(2.10) \qquad X_j(t) = m_j(t) + Z_j(t), \qquad\qquad t = 1, 2, \cdots,$$

where $m_j(\cdot)$ is the mean value function of $X_j(\cdot)$,

$$(2.11) \qquad m_j(t) = E[X_j(t)],$$

and $\mathbf{Z}(\cdot) = (Z_1(\cdot), \cdots, Z_r(\cdot))$ has zero means, is jointly normal, and is covariance stationary with covariance functions

$$(2.12) \qquad R_{hj}(v) = E[Z_h(t)Z_j(t + v)]$$

(for $h, j = 1, \cdots, r; t = 1, 2, \cdots$; and $v = 0, \pm 1, \pm 2, \cdots$), and *spectral density matrix*

$$(2.13) \qquad \mathbf{f}(\omega) = \begin{bmatrix} f_{11}(\omega) & \cdots & f_{1r}(\omega) \\ \vdots & & \vdots \\ f_{r1}(\omega) & \cdots & f_{rr}(\omega) \end{bmatrix}$$

satisfying

$$(2.14) \qquad R_{hj}(v) = \int_{-\pi}^{\pi} e^{iv\omega} f_{hj}(\omega)\, d\omega.$$

The diagonal element $f_{jj}(\omega)$ is called the spectral density function of the series $X_j(\cdot)$; the (h, j)-th element $f_{hj}(\omega)$ of the spectral density matrix is called the *cross-spectral density* of the series $X_h(\cdot)$ and $X_j(\cdot)$. Following the terminology introduced by Tukey, the real and negative imaginary parts of $f_{hj}(\omega)$ are called, respectively, the *co-spectral* density, denoted $c_{hj}(\omega)$, and *quadrature spectral* density, denoted $q_{hj}(\omega)$.

We do not demand that time series submitted for spectral analysis have vanishing mean value functions. Therefore, in studying the behavior of sample

spectra, we must distinguish two general cases: the observed time series are jointly covariance stationary with absolutely continuous spectrum and have:

(1) zero means,
(2) possibly nonzero means.

We call the second case the *mixed spectrum* case.

The problem of mixed spectra has been extensively discussed for univariate time series (see Hext [10] for a history of the problem). The theory of mixed cross-spectral analysis of multiple time series is not discussed in this paper (which is already too long) but will be discussed in a separate paper. (In her Ph.D. thesis, Grace Wahba gives a rigorous derivation of the small sample distribution theory of sample cross-spectral estimates for jointly stationary normal multiple time series with bounded nonzero mean value functions.)

In order to study the properties of the sample cross-spectral density function in the mixed spectrum case, one would introduce the sample cross-spectral density function of $m_h(\cdot)$ and $m_j(\cdot)$, defined by

$$(2.15) \qquad f_{hj;m,T}(\omega) = \frac{1}{2\pi T} \sum_{t=1}^{T} e^{i\omega t} m_h(t) \sum_{t=1}^{T} e^{-i\omega t} m_j(t).$$

The sample cross-spectral density function of the time series $Z_h(\cdot)$ is defined similarly:

$$(2.16) \qquad f_{hj;Z,T}(\omega) = \frac{1}{2\pi T} \sum_{t=1}^{T} e^{i\omega t} Z_h(t) \sum_{t=1}^{T} e^{-i\omega t} Z_j(t).$$

One use of these expressions is in writing the mean of a sample cross-spectral density:

$$(2.17) \qquad E[f_{hj;X,T}(\lambda)] = E[f_{hj;Z,T}(\lambda)] + f_{hj;m,T}(\lambda).$$

It is important to note that in order to study the properties of sample spectra it is *not* necessary to assume that the sample cross-spectral density function of the mean value functions, defined by (2.15), possesses a limit as T tends to ∞.

3. Windowed sample cross-spectra

As is well known, if one is seeking to estimate the spectral density functions of covariance stationary time series, one cannot use the sample spectral density functions but must use *windowed* sample spectra.

Given a kernel $k(v)$ and truncation point M, the *windowed cross-spectral density function*, denoted $f_{hj;T,M}(\omega)$, is defined by

$$(3.1) \qquad f_{hj;T,M}(\omega) = \frac{1}{2\pi} \sum_{|v|<M} e^{-iv\omega} k\left(\frac{v}{M}\right) R_{hj;T}(v).$$

Its real and negative imaginary parts, denoted $c_{hj;T,M}(\omega)$ and $q_{hj;T,M}(\omega)$, are called respectively the windowed sample co-spectral density function and the windowed sample quadrature spectral density function. The *windowed normalized cross-*

spectral density function, denoted $\bar{f}_{hj;T,M}(\omega)$, is defined similarly in terms of the sample cross-correlation function:

$$(3.2) \qquad \bar{f}_{hj;T,M}(\omega) = \frac{1}{2\pi} \sum_{|v|<M} e^{-iv\omega} k\left(\frac{v}{M}\right) \rho_{hj;T}(v).$$

For ease of comparing sample spectra arising from different time series, I believe it is wisest to *compute and plot normalized versions of these functions*. Indeed, I believe that normalization is vital for interpretation and that it facilitates the exchange of ideas among research workers concerned with time series arising in quite different fields. It should be noted that the theory of normalized spectra is more difficult than the unnormalized theory.

There is an extensive literature (in particular, see *Technometrics* [28] and Jenkins [12]) concerning the choice of the function $k(\cdot)$, called the lag window, and the integer $M(<T)$, called the *truncation* point (since it represents the number of sample correlations actually used in computing the spectrum). It should be noted that most methods of computing sample spectra can be essentially represented in the form (3.1) even if a formula of this kind is not explicitly employed. An extensive comparison of the effects of different choices of $k(\cdot)$ and M is beyond the scope of this paper (although an empirical comparison of a few windows is given in the next section).

At this point, let us merely note the choices of $k(\cdot)$ and M we normally make. In our work we use mainly the following lag window:

$$(3.3) \qquad \begin{aligned} k(u) &= 1 - 6u^2 + 6|u|^3, & |u| &\leq 0.5, \\ &= 2(1 - |u|)^3 & 0.5 &\leq |u| \leq 1.0, \\ &= 0, & |u| &\geq 1. \end{aligned}$$

A kernel widely used in existing spectral analysis programs is one suggested by Tukey (see Blackman and Tukey [5], p. 14):

$$(3.4) \qquad \begin{aligned} k(u) &= \tfrac{1}{2}(1 + \cos \pi u), & |u| &< 1, \\ &= 0, & &\text{otherwise.} \end{aligned}$$

This lag window is not used in our work because the corresponding windowed spectrum is not necessarily nonnegative (and the corresponding estimates of coherence are not necessarily between 0 and 1).

Two other kernels which might be considered are one generally known as the Bartlett kernel,

$$(3.5) \qquad \begin{aligned} k(u) &= 1 - |u|, & |u| &\leq 1, \\ &= 0, & &\text{otherwise,} \end{aligned}$$

and one which we call the Bohman kernel (after Bohman [6] who introduced it in connection with the numerical inversion of characteristic functions to compute distribution functions),

$$(3.6) \qquad k(u) = (1 - u) \cos \pi u + \frac{1}{\pi} \sin \pi u, \qquad 0 < u < 1,$$

$$= 0, \qquad u > 1,$$

$$= k(-u), \qquad u < 0.$$

The *spectral window* of a windowed sample spectrum of the form of (3.2) is defined to be the function

$$(3.7) \qquad K_M(\omega) = \frac{1}{2\pi} \sum_{|v| \leq M} e^{-iv\omega} k\left(\frac{v}{M}\right);$$

the *spectral window* generator is defined to be the Fourier transform

$$(3.8) \qquad K(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iu\omega} k(u) \, du.$$

For the lag window (3.4), it may be shown that

$$(3.9) \qquad K_M(\omega) = \frac{3}{8\pi M^3} \left\{ \frac{\sin (M\omega/4)}{\frac{1}{2} \sin \frac{\omega}{2}} \right\}^4 \left\{ 1 - \frac{2}{3} \left(\sin \frac{\omega}{2} \right)^2 \right\},$$

$$(3.10) \qquad K(\omega) = \frac{3}{8\pi} \left\{ \frac{\sin (\omega/4)}{\omega/4} \right\}^4.$$

It may be shown that a windowed sample spectral density function $f_{hj;T,M}(\omega)$ is the convolution of the sample cross-spectral density function $f_{hj;T}(\omega)$, and the spectral window $K_M(\omega)$,

$$(3.11) \qquad f_{hj;T,M}(\omega) = \int_{-\pi}^{\pi} K_M(\omega - \lambda) f_{hj;T}(\lambda) \, d\lambda.$$

Therefore, its mean is also a convolution,

$$(3.12) \qquad E[f_{hj;T,M}(\omega)] = \int_{-\pi}^{\pi} K_M(\omega - \lambda) E[f_{hj;T}(\lambda)] \, d\lambda.$$

While it is more difficult to justify universal advice on the choice of the truncation point, my experience leads me to believe that it is necessary and sufficient to use *three* truncation points, M_1, M_2, M_3, satisfying a condition of the following kind:

$$(3.13) \qquad 5\% \leq \frac{M_1}{T} \leq 10\%, \qquad 10\% \leq \frac{M_2}{T} \leq 25\%, \qquad 25\% \leq \frac{M_3}{T} \leq 75\%.$$

I have several justifications for this advice: (1) in general, if one is in doubt as to which of two ways to perform an analysis, one should do it both ways and decide by a comparison of results which way was right; (2) the three truncation points given in (3.13) span the range of possible truncation points, and not too much additional information can be obtained by using additional truncation points; and (3) the presence of peaks and the smoothness of spectra can be determined by comparing spectra corresponding to different truncation points.

Given finite samples of r time series, for brevity we often denote by $\hat{f}_{hj}(\omega)$ the windowed normalized cross-spectral density function defined by (3.2). We let

$$(3.14) \qquad \hat{\mathbf{f}}(\omega) = \begin{bmatrix} \hat{f}_{11}(\omega) & \cdots & \hat{f}_{1r}(\omega) \\ \vdots & & \vdots \\ \hat{f}_{r1}(\omega) & \cdots & \hat{f}_{rr}(\omega) \end{bmatrix}$$

denote the windowed sample cross-spectral density matrix. A variety of derived spectral quantities may be computed which hopefully will provide insight into the relations among the observed time series. For two series $X_h(\cdot)$ and $X_j(\cdot)$, one can form the following derived sample spectral quantities: the *sample regression transfer function*

$$(3.15) \qquad \hat{B}_{h;j}(\omega) = \frac{\hat{f}_{hj}(\omega)}{\hat{f}_{jj}(\omega)} = \frac{\hat{c}_{hj}(\omega)}{\hat{f}_{jj}(\omega)} - i\,\frac{\hat{q}_{hj}(\omega)}{\hat{f}_{jj}(\omega)} = \hat{\alpha}_{h;j}(\omega) + i\hat{\beta}_{h;j}(\omega),$$

the *sample residual spectral density function*

$$(3.16) \qquad \hat{f}_{h;j}(\omega) = \hat{f}_{hh}(\omega)\{1 - \hat{W}_{hj}(\omega)\}$$

where $\hat{W}_{hj}(\omega)$, called the *sample coherence* between series h and series j, is defined by

$$(3.17) \qquad \hat{W}_{hj}(\omega) = \frac{|\hat{f}_{hj}(\omega)|^2}{\hat{f}_{hh}(\omega)\hat{f}_{jj}(\omega)} = \frac{\hat{c}_{hj}^2(\omega) + \hat{q}_{hj}^2(\omega)}{\hat{f}_{hh}(\omega)\hat{f}_{jj}(\omega)}.$$

The regression transfer function may be written as

$$(3.18) \qquad \hat{B}_{h;j}(\omega) = \hat{G}_{h;j}(\omega)e^{-i\hat{\varphi}_{h;j}(\omega)}$$

where

$$(3.19) \qquad \hat{G}_{h;j}(\omega) = \frac{|\hat{f}_{hj}(\omega)|}{\hat{f}_{jj}(\omega)},$$

called the *sample gain* at frequency ω of the predictor of $X_h(\cdot)$ given $X_j(\cdot)$, and

$$(3.20) \quad \begin{aligned} \hat{\varphi}_{h;j}(\omega) &= \tan^{-1}\left(\frac{\hat{q}_{hj}(\omega)}{\hat{c}_{hj}(\omega)}\right) & \text{if} \quad \hat{c}_{hj}(\omega) \geq 0, \\[2mm] \hat{\varphi}_{h;j}(\omega) &= \left\{\tan^{-1}\left(\frac{\hat{q}_{hj}(\omega)}{\hat{c}_{hj}(\omega)}\right) + \pi\,\text{sign}\,[\hat{q}_{hj}(\omega)]\right\} & \text{if} \quad \hat{c}_{hj}(\omega) < 0, \end{aligned}$$

called the *sample phase* difference between the two series at frequency ω. Interpretations and generalizations of these quantities are given in section 7; their sampling theory is discussed in section 8.

4. A comparison of spectral windows

In interpreting windowed sample cross-spectra of observed time series, it is valuable to compare them with similarly computed windowed sample cross-spectra of artificially generated time series. We present here examples of sample cross-spectra for a few simple artificial time series. Our aim is first to gain some idea of what sample cross-spectra look like, and second, to see some of the ways in which the choice of lag window affects the results.

A time series identically equal to 1,

(4.1)
$$X(t) = 1 \qquad \text{for all} \quad t,$$

has as its sample covariance function

(4.2)
$$R_T(v) = \frac{1}{T} \sum_{t=1}^{T} X(t)X(t+v) = \left(1 - \frac{v}{T}\right).$$

The corresponding windowed sample spectral density is

(4.3)
$$K_{M,T}(\omega) = \frac{1}{2\pi} \sum_{|v|<T} e^{-iv\omega} \left(1 - \frac{|v|}{T}\right) k\left(\frac{v}{M}\right).$$

For many purposes it can be verified that approximately

(4.4)
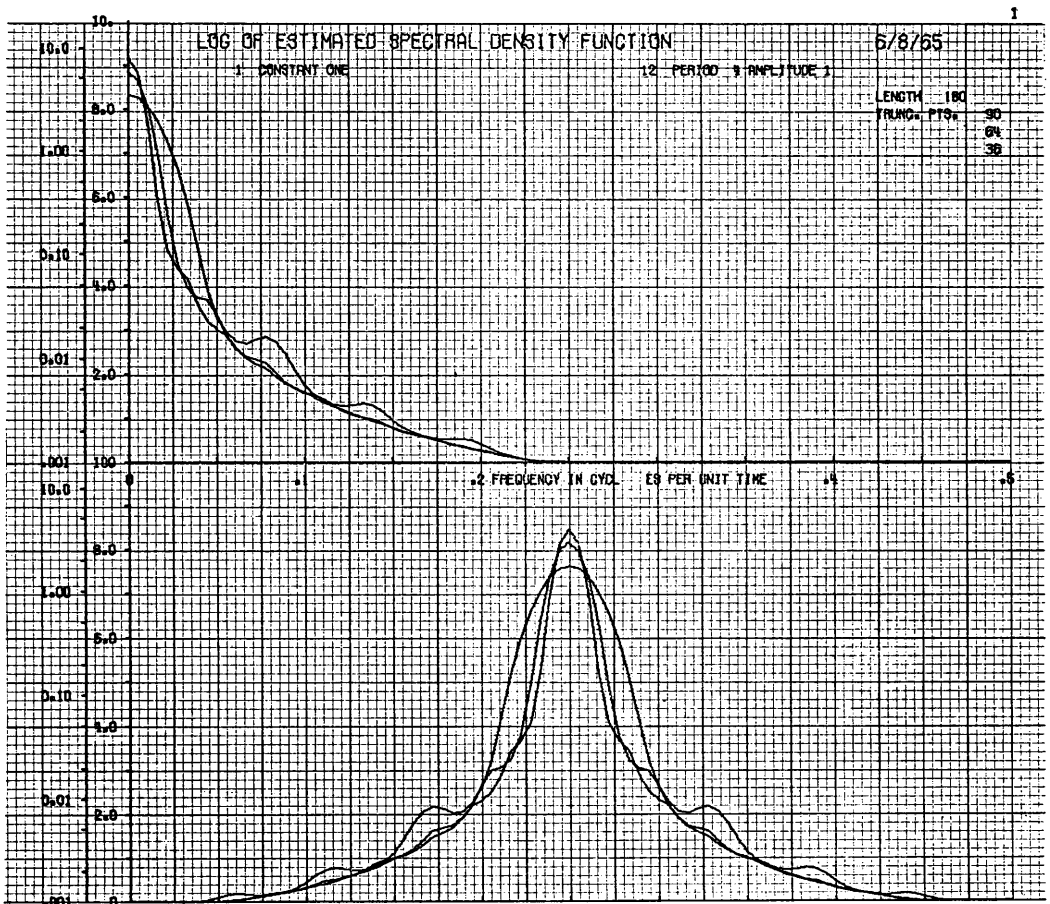$$K_{M,T}(\omega) \doteq K_M(\omega).$$

FIGURE 2

Figures 2–5 plot (in the top half) the function $K_{M,T}(\omega)$ for $T = 180$, $M = 90$, 64, 36, and for the four lag windows we have mentioned. The bottom half of each of these figures plots the windowed sample spectral density function of the time series

$$(4.5) \qquad\qquad X(t) = \cos \frac{2\pi}{4} t.$$

The horizontal axis of these figures measures frequency on an axis from 0 to 0.5, representing ν (cycles per unit time) rather than ω (radians per unit time). The function plotted is not the windowed sample spectral density $\hat{f}(\omega)$, but rather

$$(4.6) \qquad\qquad 1000 \log_e \hat{f}\left(\frac{\omega}{2\pi}\right).$$
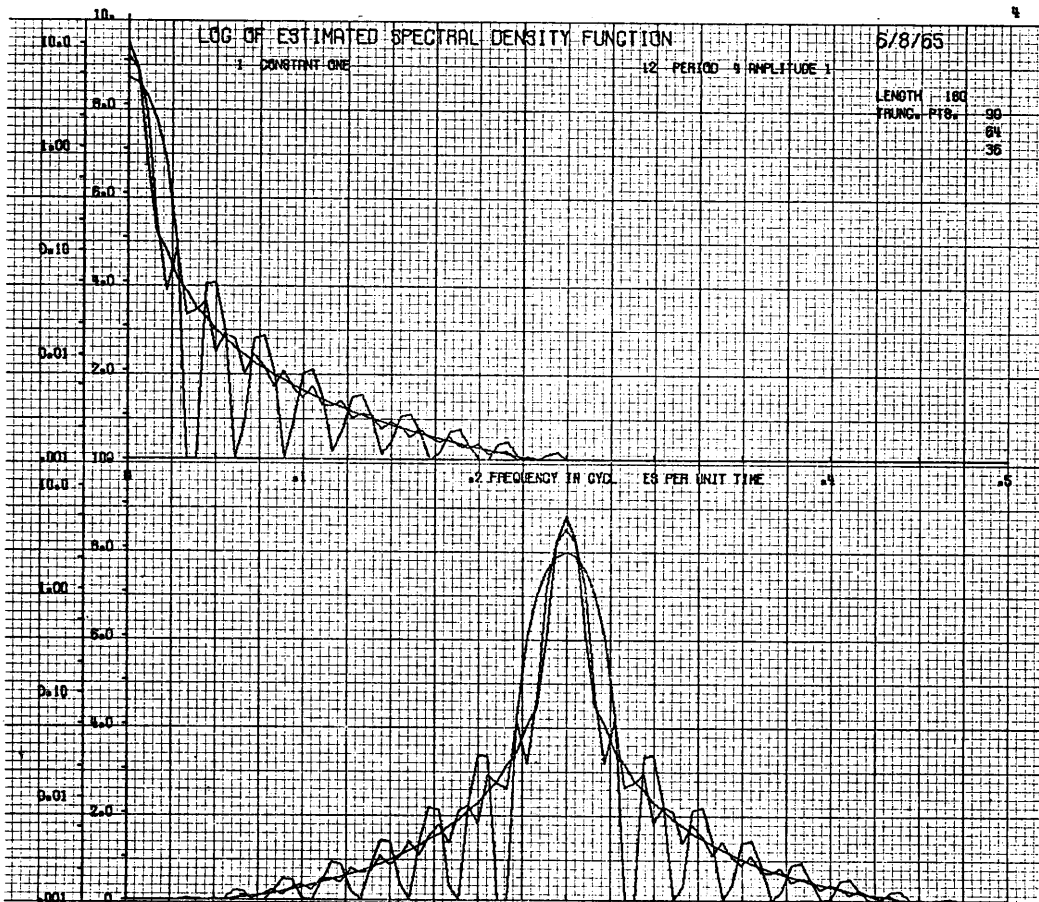

FIGURE 3

By white noise we mean any time series of uncorrelated random variables with zero means; it is covariance stationary with covariance function

(4.7)
$$R(v) = 1 \qquad \text{if} \quad v = 0,$$
$$R(v) = 0 \qquad \text{if} \quad v \neq 0,$$

when normalized to have unit variance. The corresponding spectral density function is

(4.8)
$$f(\omega) = \frac{1}{2\pi}, \qquad -\pi \leq \omega \leq \pi.$$

Now

(4.9)
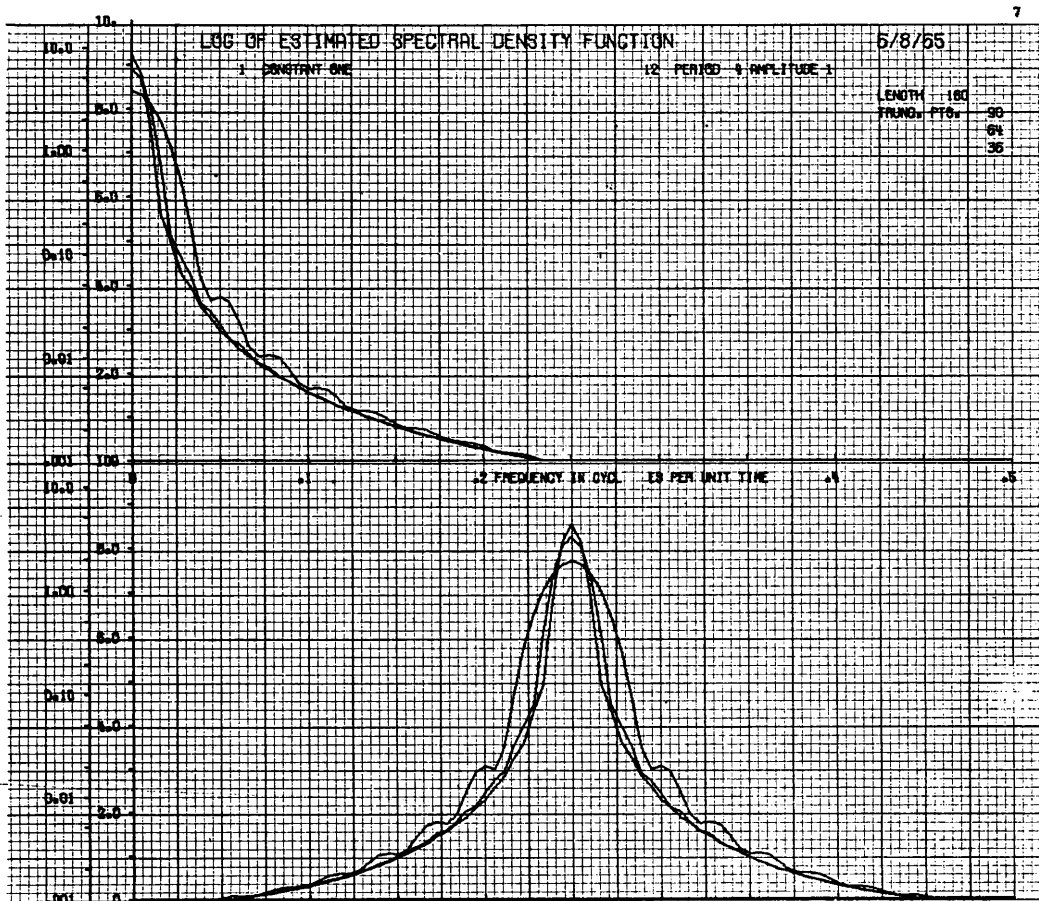$$1000 \log_e \frac{1}{2\pi} = 5.07.$$

FIGURE 4

Therefore, in figures 2–5 the normalized windowed sample spectral density functions are plotted on a scale that goes from 1 to 10. A sample of white noise would be expected to oscillate about the middle of the graph.

Various numerical measures for comparing properties of various spectral windows have been introduced in previous papers [see Parzen [21], [23]). Here, taking an empirical attitude, we study the computer output one obtains from an empirical time series analysis of various series. Comparing in figures 2–5 the window sample spectra of a pure sine wave, one sees that figure 3 (the Tukey kernel) is more oscillatory than figure 4 (the Bohman kernel), which in turn is slightly more oscillatory than figure 2 (the Parzen kernel). Figure 5 (the Bartlett kernel) shows strong oscillations as well as a very unsatisfactory failure to damp down. These differences in behavior hold more for small truncation points than for large truncation points.
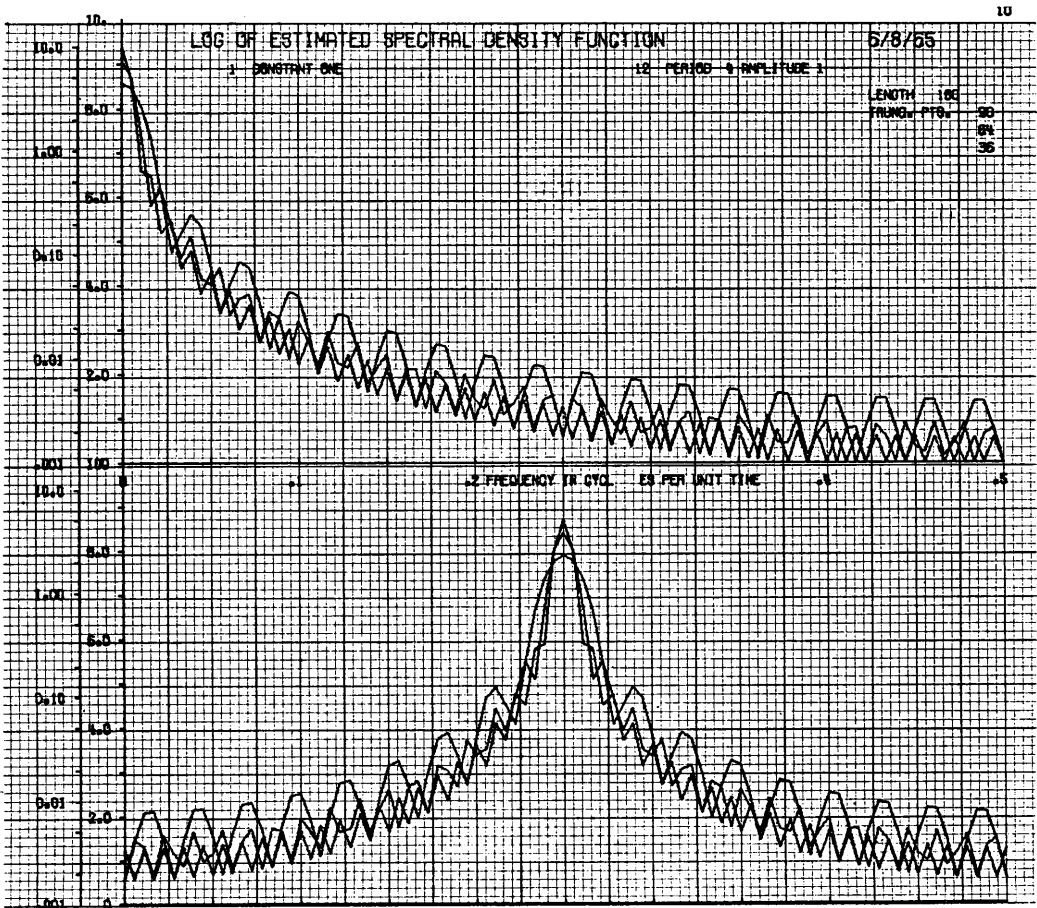

FIGURE 5

Figures 6–9 present windowed sample spectra of time series $N(t) + \cos(2\pi/4)t$ and $N(t)$, where $N(t)$ is a "sample of white noise" internally generated by the computer. For these time series the differences between the various windows is much less pronounced than in figures 2–5.

However, there seems to be much difference between the graphs of the sample coherence, plotted in figures 10–13 for the various windows. Figure 10 (the Parzen window) seems to have the smoothest behavior.

Although spectral distribution functions are not discussed in this paper, plots of windowed sample distribution functions are given in figure 14.

5. Sampling theory of sample cross-spectra

In this section we outline the properties of the windowed sample cross-spectra when the observed time series are jointly covariance stationary with absolutely continuous spectrum and have zero means.
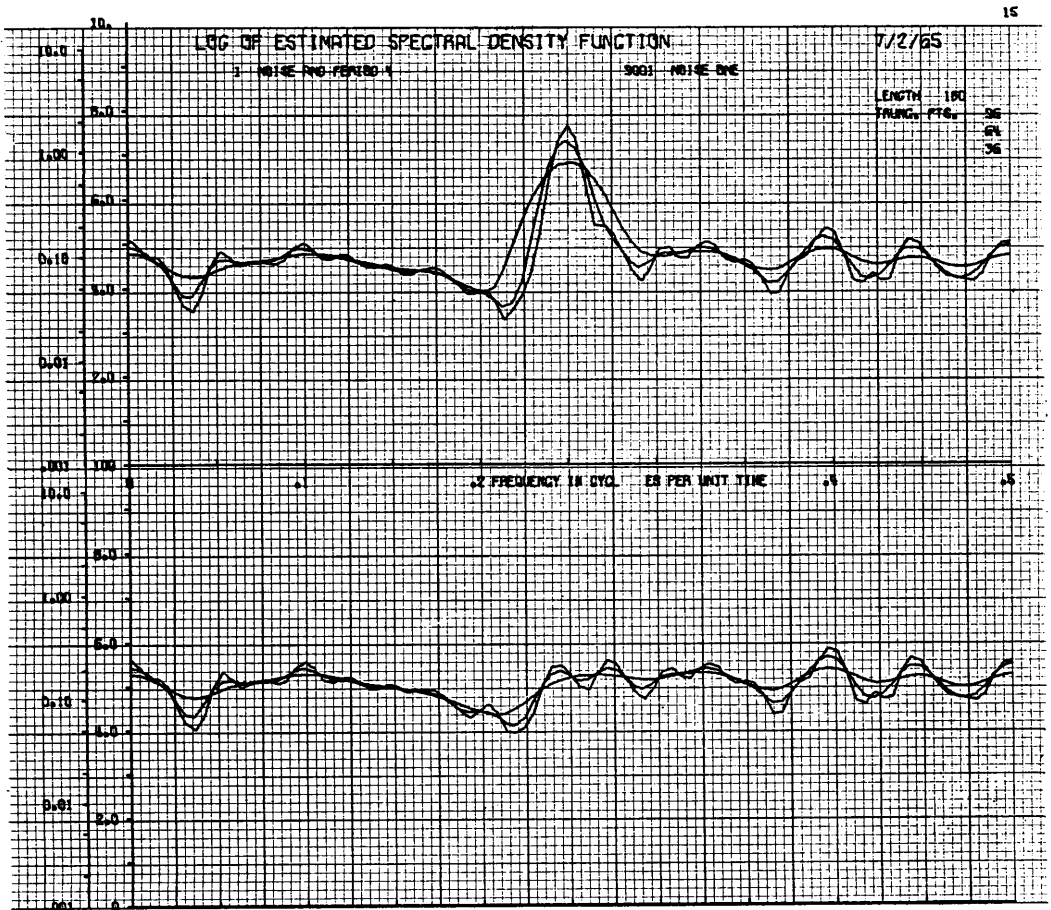

FIGURE 6

We first consider the mean of a windowed sample cross-spectral density

$$(5.1) \qquad E[\hat{f}_{hj}(\omega)] = \frac{1}{2\pi} \sum_{|v| < M} e^{-iv\omega} k\left(\frac{v}{M}\right)\left(1 - \frac{|v|}{T}\right) R_{hj}(v)$$

$$= \int_{-\pi}^{\pi} K_{M,T}(\omega - \lambda) f_{hj}(\lambda) \, d\lambda,$$

where $K_{M,T}(\omega)$ is defined by (4.3). Assuming that (4.4) holds, we obtain the following approximation for the mean of a sample cross-spectral density

$$(5.2) \qquad E[\hat{f}_{hj}(\omega)] \doteq \int_{-\pi}^{\pi} K_M(\omega - \lambda) f_{hj}(\lambda) \, d\lambda.$$

To evaluate this integral it is often assumed that, in the neighborhood of ω, the real and imaginary parts $f_{hj}(\lambda)$ are both varying slowly compared to $K_M(\omega - \lambda)$; then approximately
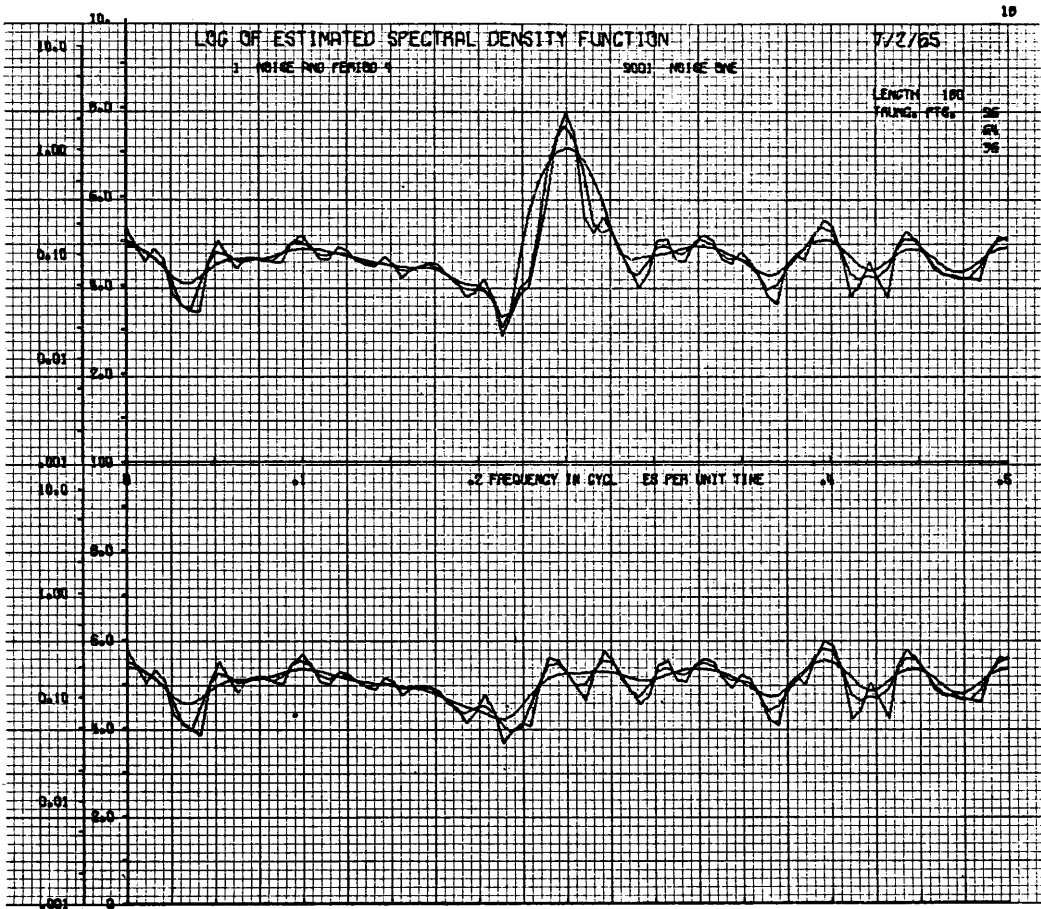

FIGURE 7

$$(5.3) \qquad \int_{-\pi}^{\pi} K_M(\omega - \lambda) f_{hj}(\lambda) \, d\lambda \doteq f_{hj}(\omega) \int_{-\pi}^{\pi} K_M(\omega - \lambda) \, d\lambda = f_{hj}(\omega).$$

It is thus implied that to a first-order approximation (as $M \to \infty$), the windowed sample cross-spectral density is an *unbiased* estimate of the true cross-spectral density (when the observed time series is zero mean covariance stationary). While this is a correct statement from the asymptotic point of view, for finite samples there is a bias in cross-spectral estimates not present in auto-spectral estimates; this bias is discussed in section 9 since in order to discuss it, we need to first introduce the notions of gain and phase. A comprehensive and rigorous discussion of bias in cross-spectral estimates is given by Nigel Nettheim in his 1966 Stanford Ph.D. thesis.

Much of the mathematical literature on cross-spectral analysis has been concerned with variability rather than bias. One can investigate the sampling theory
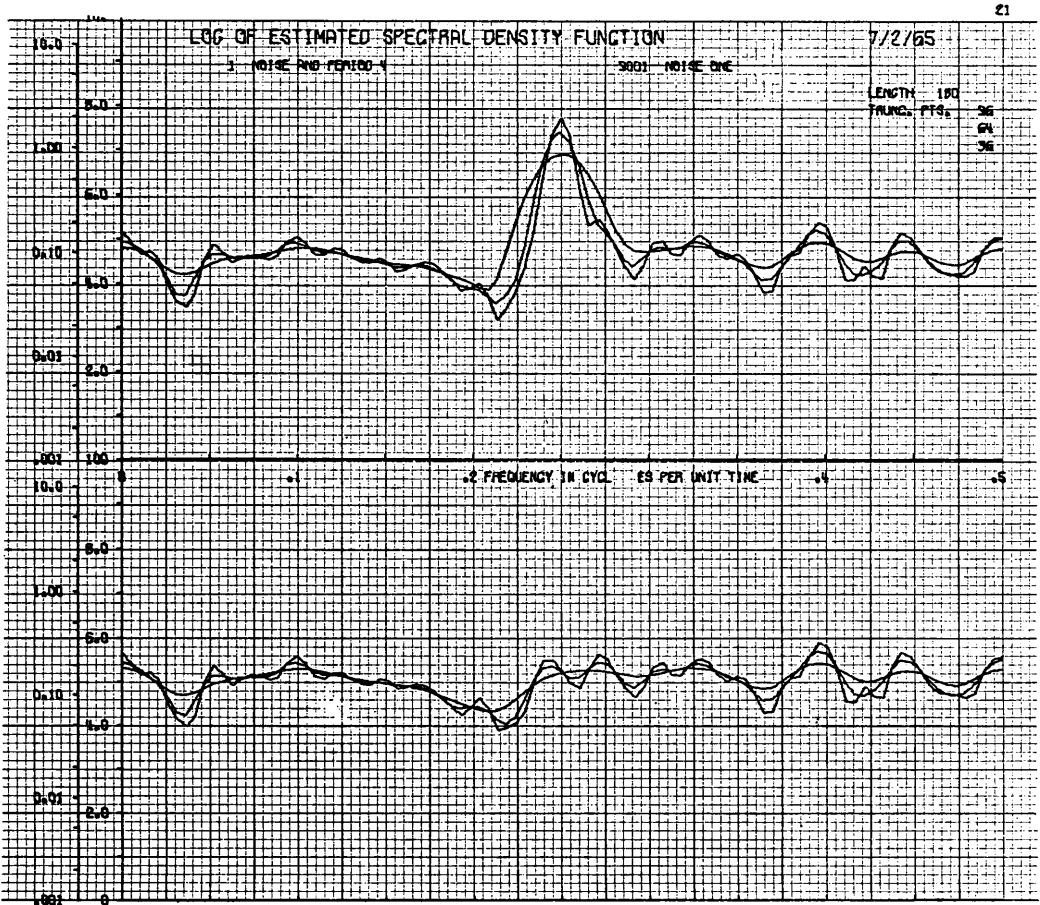

FIGURE 8

of sample cross-spectra from an asymptotic point of view or from a small sample point of view in the case that the observed time series are assumed to be zero mean, normal, jointly covariance stationary, and possessing spectral density functions.

The basic formula of the asymptotic point of view is (under suitable conditions on the kernel $k(u)$ and assuming that the same kernel and truncation are used in every estimate)

$$(5.4) \qquad \operatorname{cov}\left[\hat{f}_{hj}(\omega), \hat{f}_{kn}(\omega)\right] \doteq C f_{hk}(\omega)\bar{f}_{jn}(\omega), \qquad 0 < \omega < \pi,$$

writing $\bar{z}$ to denote the complex conjugate of a complex number z and defining

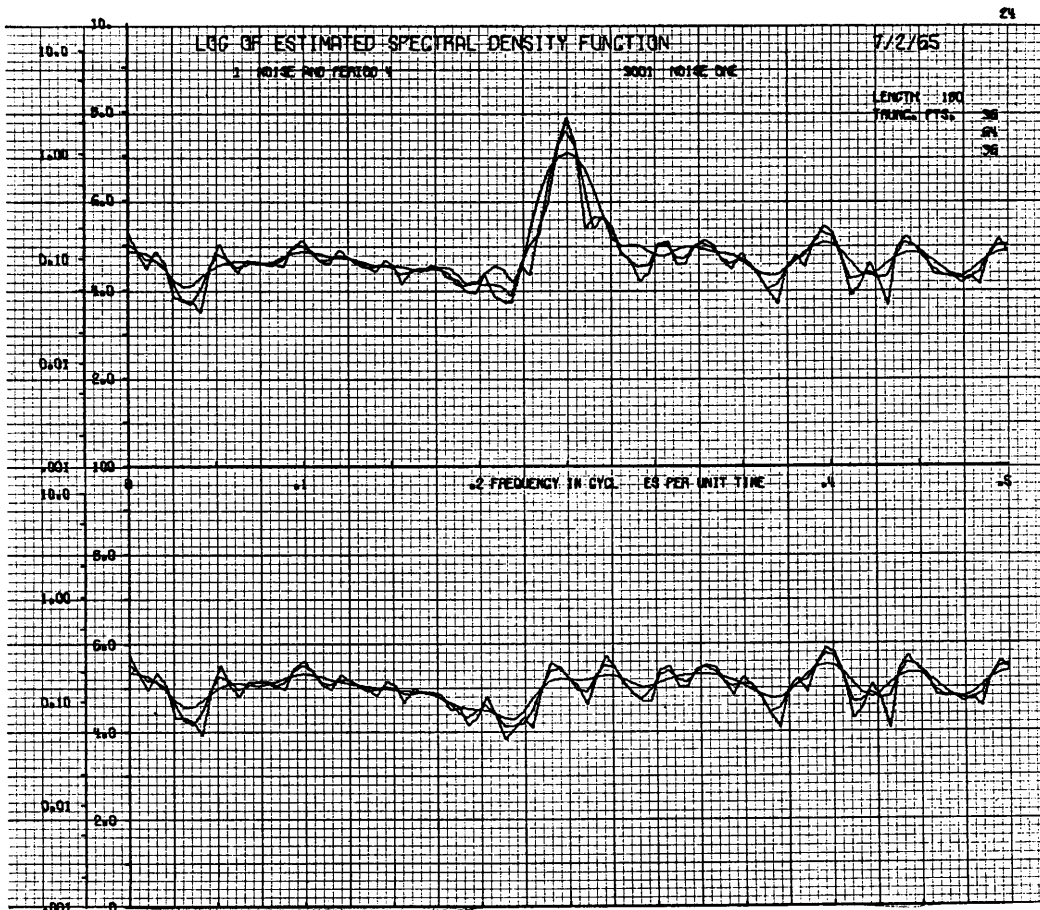$$(5.5) \qquad C = \frac{M}{T} \int_{-\infty}^{\infty} k^2(u) \, du.$$

FIGURE 9

To prove (5.4) let us first note, without proof, that (compare Rosenblatt [27])

$$(5.6) \qquad \qquad \text{cov } [\hat{f}_{11}(\omega), \hat{f}_{22}(\omega)] \doteq C|f_{12}(\omega)|^2, \qquad \qquad 0 < \omega < \pi.$$

From (5.6) one derives (5.4) as follows: consider arbitrary linear combinations of the observed time series $Z_j(t)$,

$$(5.7) \qquad \qquad Y_1(t) = \sum_j a_j Z_j(t), \qquad Y_2(t) = \sum_k b_k Z_k(t).$$

Their sample spectra can be written (using $\hat{f}_{11}(\omega)$ with two meanings, as the windowed sample spectral density of both $Y_1(t)$ and $Z_1(t)$; similarly for $\hat{f}_{22}$ and $\hat{f}_{12}$)

$$(5.8) \qquad \qquad \hat{f}_{11} = \sum_{h,j} a_h \hat{f}_{hj} \bar{a}_j, \qquad \hat{f}_{22} = \sum_{k,n} b_k \hat{f}_{kn} \bar{b}_n;$$
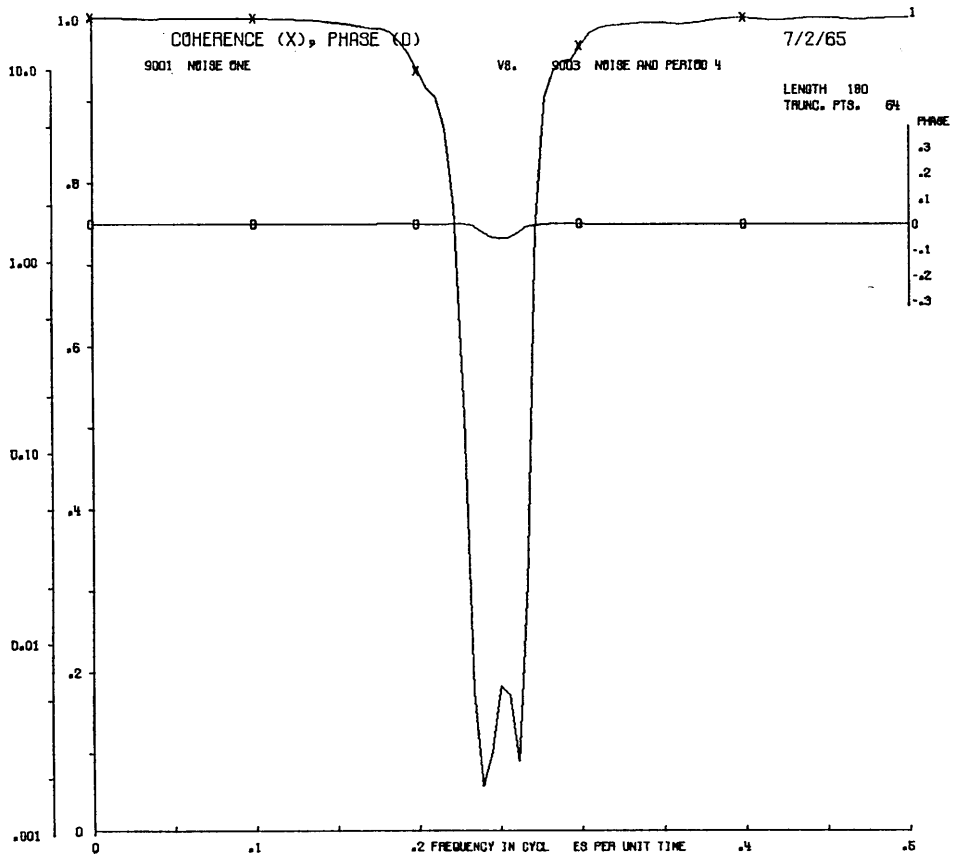
FIGURE 10

consequently,

$$(5.9) \qquad \text{cov} \, [\hat{f}_{11}(\omega), \hat{f}_{22}(\omega)] = \sum_{h,j,k,n} a_h \bar{a}_j \, \text{cov} \, [\hat{f}_{hj}(\omega), \hat{f}_{kn}(\omega)] \bar{b}_k b_n,$$

$$\doteq C|f_{12}(\omega)|^2,$$

$$= C|\sum_{h,k} a_h f_{hk}(\omega) \bar{b}_k|^2,$$

$$= C \sum_{h,j,k,n} a_h \bar{a}_j f_{hk}(\omega) \bar{f}_{jn}(\omega) \bar{b}_k b_n.$$

One may now infer (5.4).

The meaning of (5.4) is best understood by writing out the variance-covariance matrix of the estimates $\hat{f}_{11}(\omega)$, $\hat{f}_{22}(\omega)$, $\hat{f}_{12}(\omega)$, $\hat{f}_{21}(\omega)$.

FIGURE 11

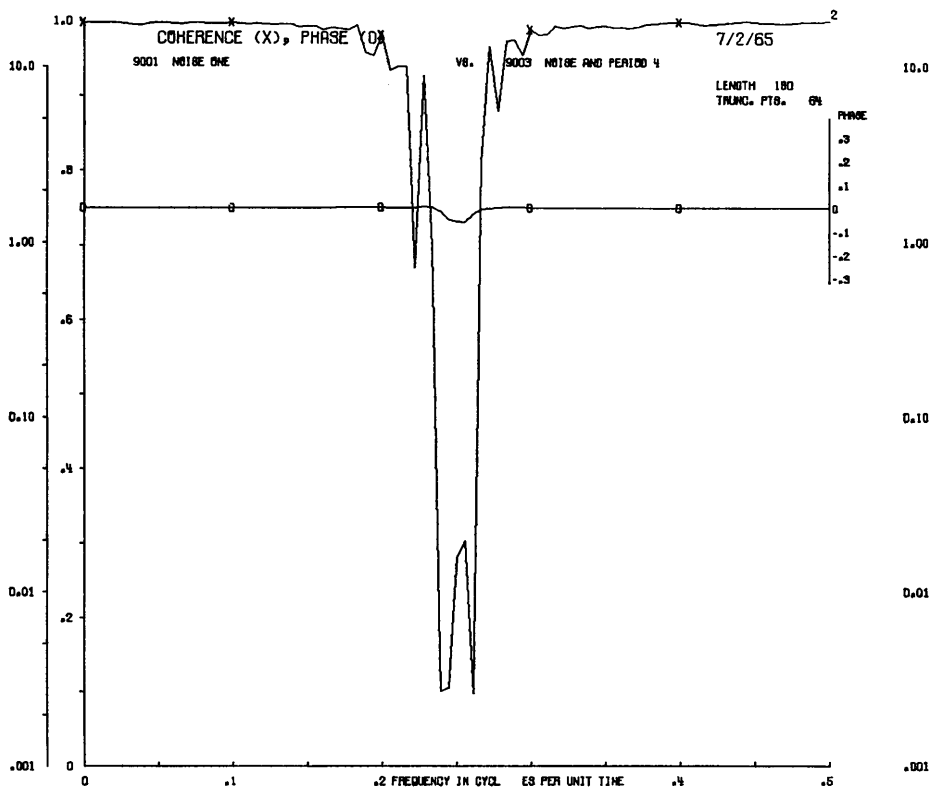|	$\hat{f}_{11}(\omega)$	$\hat{f}_{22}(\omega)$	$\hat{f}_{12}(\omega)$	$\hat{f}_{21}(\omega)$
$\hat{f}_{11}(\omega)$	$C\|f_{11}(\omega)\|^2$	$C\|f_{12}(\omega)\|^2$	$Cf_{11}(\omega)\bar{f}_{12}(\omega)$	$Cf_{11}(\omega)\bar{f}_{21}(\omega)$
$\hat{f}_{22}(\omega)$		$C\|f_{22}(\omega)\|^2$	$Cf_{22}(\omega)\bar{f}_{12}(\omega)$	$Cf_{22}(\omega)\bar{f}_{21}(\omega)$
$\hat{f}_{12}(\omega)$			$Cf_{11}(\omega)f_{22}(\omega)$	$C\{f_{12}(\omega)\}^2$
$\hat{f}_{21}(\omega)$				$Cf_{11}(\omega)f_{22}(\omega)$

(5.10)

From (5.10) one obtains the covariances of $\hat{f}_{11}(\omega)$, $\hat{f}_{22}(\omega)$, $\hat{c}_{12}(\omega) = \text{Re}\,\hat{f}_{12}(\omega)$, and $\hat{q}_{12}(\omega) = -\text{Im}\,\hat{f}_{12}(\omega)$. In writing the following table we have *omitted* from every entry the factor C defined by (5.5).

FIGURE 12

(5.11)

	$\hat{f}_{11}(\omega)$	$\hat{f}_{22}(\omega)$	$\hat{c}_{12}(\omega)$	$\hat{q}_{12}(\omega)$
$\hat{f}_{11}(\omega)$	$\lvert f_{11}(\omega)\rvert^2$	$c_{12}^2(\omega) + q_{12}^2(\omega)$	$f_{11}(\omega)c_{12}(\omega)$	$-f_{11}(\omega)q_{12}(\omega)$
$\hat{f}_{22}(\omega)$		$\lvert f_{22}(\omega)\rvert^2$	$f_{22}(\omega)c_{12}(\omega)$	$-f_{22}(\omega)q_{12}(\omega)$
$\hat{c}_{12}(\omega)$			$\frac{1}{2}\{f_{11}(\omega)f_{22}(\omega) + c_{12}^2(\omega) - q_{12}^2(\omega)\}$	$c_{12}(\omega)q_{12}(\omega)$
$\hat{q}_{12}(\omega)$				$\frac{1}{2}\{f_{11}(\omega)f_{22}(\omega) + q_{12}^2(\omega) - c_{12}^2(\omega)\}$
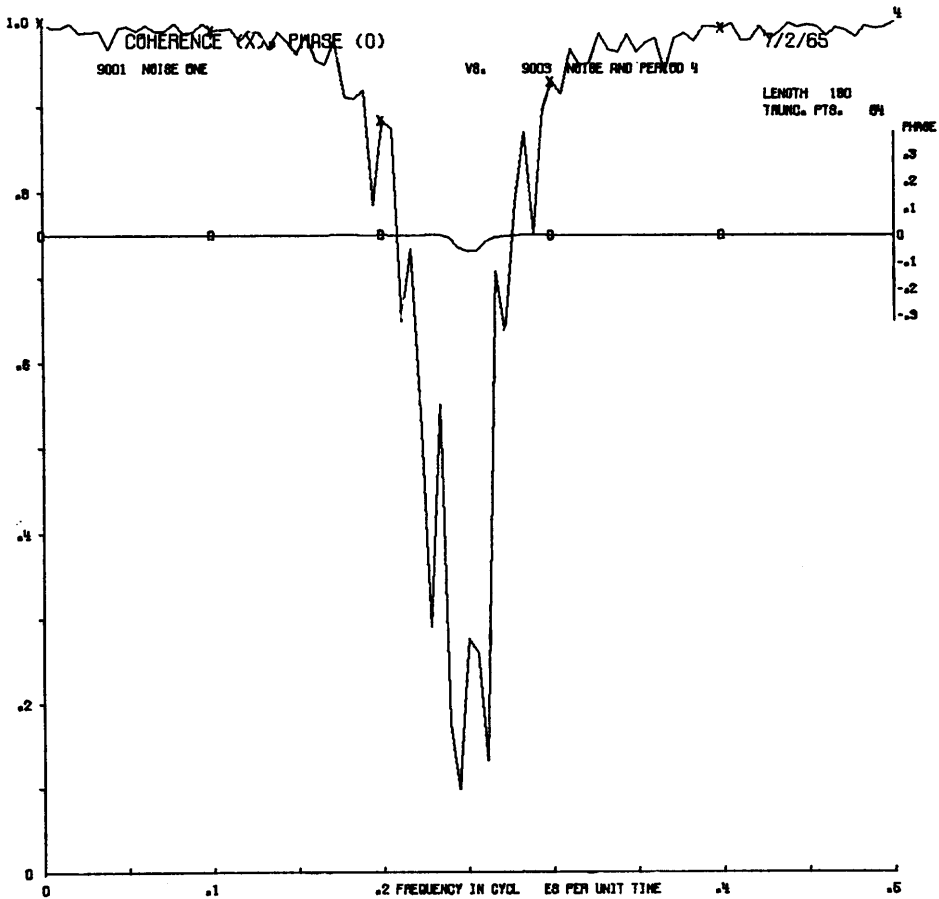
FIGURE 13

One approach to the "small sample" distribution theory of sample cross-spectra is due to Goodman [9] who uses an analogy between windowed sample cross-spectra and a sample covariance matrix

$$(5.12) \qquad \hat{K}_{jk} = \frac{1}{n} \sum_{t=1}^{n} X_j(t)\overline{X}_k(t), \qquad\qquad j, k = 1, \cdots, r,$$

where for $t = 1, 2, \cdots, n$ $\{X_j(t), j = 1, \cdots, r\}$ are independent *complex* random vectors identically distributed as the vector $\{X_j, j = 1, \cdots, r\}$ which is assumed to (i) be normally distributed, (ii) have zero means, and (iii) satisfy the conditions for any indices j and k,

$$(5.13) \qquad\qquad E[X_j X_k] = 0, \qquad E[X_j \overline{X}_k] = K_{jk}.$$

One important case in which the first equation in (5.13) holds is when

$$(5.14) \qquad\qquad X_j = U_j + iV_j, \qquad X_k = U_k + iV_k,$$

FIGURE 14

where U_j, V_j, U_k, V_k are jointly normal random variables with zero means and covariances

(5.15)

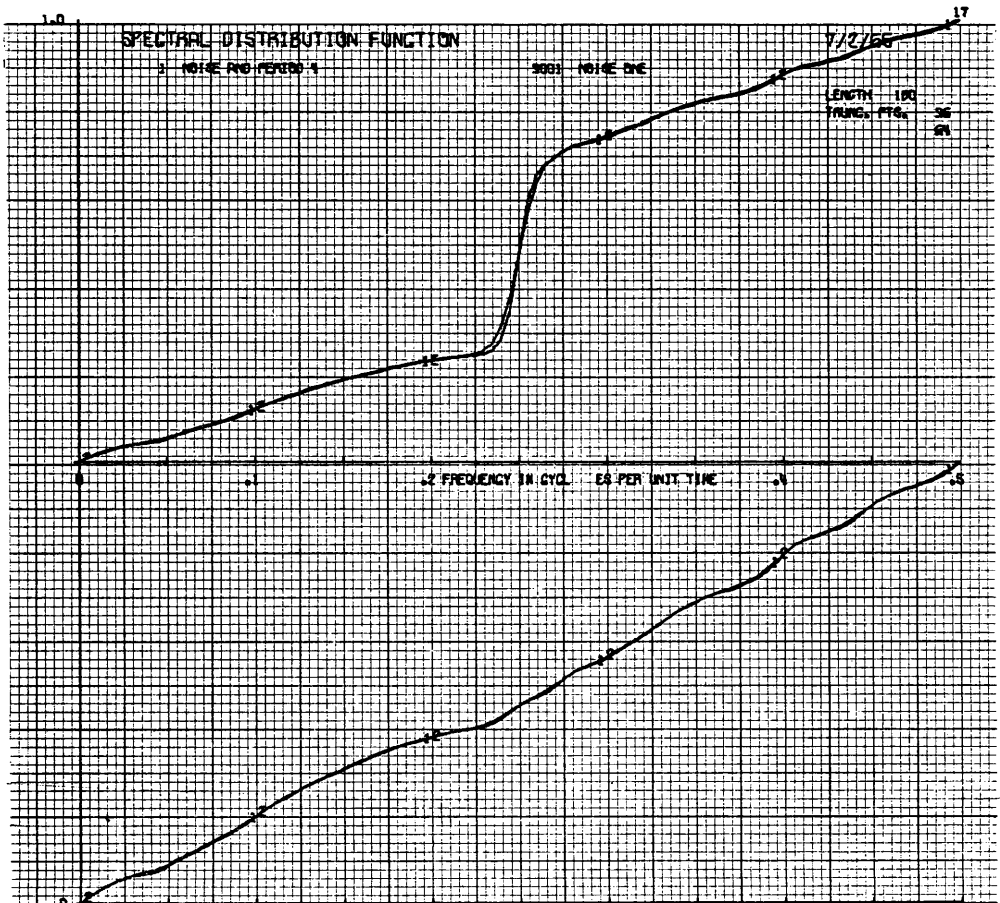$$\begin{array}{c} \quad\; U_j \quad\; V_j \quad\; U_k \quad\; V_k \\ \begin{array}{c} U_j \\ V_j \\ U_k \\ V_k \end{array} \begin{bmatrix} \sigma_j^2 & 0 & \alpha_{jk} & \beta_{jk} \\ 0 & \sigma_j^2 & -\beta_{jk} & \alpha_{jk} \\ \alpha_{jk} & -\beta_{jk} & \sigma_k^2 & 0 \\ \beta_{jk} & \alpha_{jk} & 0 & \sigma_k^2 \end{bmatrix} \end{array}.$$

From (5.15) it follows that

$$\begin{aligned} (5.16) \quad \operatorname{cov}\,[X_h\overline{X}_j, X_k\overline{X}_m] &= E[X_h\overline{X}_j\overline{X}_kX_m] - E[X_h\overline{X}_j]E[\overline{X}_kX_m], \\ &= E[X_h\overline{X}_k]E[\overline{X}_jX_m] + E[X_hX_m]E[\overline{X}_j\overline{X}_k], \\ &= K_{hk}K_{jm}. \end{aligned}$$

We thus obtain a basic formula for the covariances of the sample covariance function (true for every n, and not only asymptotically):

$$(5.17) \qquad\qquad \operatorname{cov}\,[\hat{K}_{hj}, \hat{K}_{km}] = \frac{1}{n} K_{hk}\overline{K}_{jm}.$$

This formula is reminiscent of (5.4), identifying $1/n$ with C.

To illustrate the application of this result, let us consider two time series of length $T = 180$ whose windowed sample cross-spectra are computed for a truncation point $M = 64$. Then

$$(5.18) \qquad\qquad n = 1 \div C = T \div M \int_{-\infty}^{\infty} k^2(u)\,du$$

$$= \begin{cases} 5.21 & \text{for Parzen window,} \\ 3.75 & \text{for Tukey window,} \end{cases}$$

since

$$(5.19) \qquad \int_{-\infty}^{\infty} k^2(u)\,du = \begin{cases} 0.54 & \text{for Parzen window,} \\ 0.75 & \text{for Tukey window.} \end{cases}$$

From Goodman's small sample approximation to the sampling theory of windowed sample cross-spectra, one can derive significance levels for the sample coherence. For T/M approximately 3, the 95% significance levels to test the hypothesis that true coherence at a given frequency is *zero* are, respectively, 0.464 (for Parzen window) and 0.632 (for the Tukey window); these values are obtained from Amos and Koopmans [4]).

6. Prediction filters and partial cross-spectra

Increasingly, techniques of regression analysis, correlation analysis, and multivariate analysis are being applied by research workers in various disciplines as a means of studying the relations between various variables. A lucid discussion of the basic methodology is given by Kendall and Stuart [13]. The aim of this section is to indicate how these ideas generalize to time series. Rigorous developments of some of these ideas have been given by Koopmans [14], [15].

The notions introduced in multivariate analysis to describe the relations between a family of random variables may be generalized in two ways to time series, depending on whether one uses two-sided or one-sided prediction filters. The generalization using two-sided prediction filters (which for technical reasons is the one most often considered) is discussed in detail. The generalization using one-sided prediction filters is briefly mentioned.

Let $X_1(\cdot), \cdots, X_r(\cdot)$ be r time series. Let P be a subset of the set $D = \{1, 2, \cdots, r\}$ of indices, and let j be any index in D. We define a new time series, denoted

$$(6.1) \qquad\qquad X_{j;P}(t), \qquad\qquad t = 0, \pm 1, \cdots,$$

and called the minimum mean square error linear predictor of $X_j(t)$, given $\{X_k(s), s = 0, \pm 1, \cdots, k \in \mathrm{P}\}$, as follows: $X_{j,P}(t)$ is a linear combination of the predictor random variables, which we write

$$(6.2) \qquad\qquad X_{j;P}(t) = \sum_{k \in P} \sum_{s=-\infty}^{\infty} b_{jk;P}(t - s) X_k(s)$$

whose mean square prediction error as an estimate of $X_j(t)$ is a minimum (that is, does not exceed the mean square prediction error of any other predictor of $X_j(t)$ which is a linear combination of $\{X_k(t), t = 0, \pm 1, \cdots, k \in P\}$). It should be noted that in general $X_{j;P}(t)$ cannot be written as an infinite series; this assumption is made only for ease of exposition. As shown by Koopmans [15], the conclusions given may be shown to hold under somewhat more general conditions using the Hilbert space theory of time series.

The coefficient $b_{jk;P}(t - s)$ is a function only of the time difference $t - s$ because of the joint stationarity; similarly, the mean square prediction error is independent of t. We call $b_{jk;P}(t - s)$ the *partial regression coefficient* of $X_j(t)$ on $X_k(s)$ given $\{X_k(\cdot), k \in P\}$. These coefficients are determined by the conditions

$$(6.3) \quad E[X_{j;P}(t)X_h(u)] = E[X_j(t)X_h(u)], \qquad \text{for} \quad u = 0, \pm 1, \cdots, h \in P,$$

which lead to the normal equations

$$(6.4) \qquad \sum_{k \in P} \sum_{s=-\infty}^{\infty} b_{jk;P}(t - s) E[X_k(s)X_h(u)] = E[X_j(t)X_h(u)],$$

which may be written in terms of covariance functions

$$(6.5) \qquad \sum_{k \in P} \sum_{v=-\infty}^{\infty} b_{jk;P}(v) R_{kh}(u + v - t) = R_{jh}(u - t),$$

and in terms of spectral density functions

$$(6.6) \qquad \sum_{k \in P} \sum_{s=-\infty}^{\infty} b_{jk;P}(v) \int_{-\pi}^{\pi} e^{i\omega(u+v-t)} f_{kh}(\omega) \, d\omega = \int_{-\pi}^{\pi} e^{i\omega(u-t)} f_{jh}(\omega) \, d\omega.$$

The Fourier transform of the partial regression coefficients is called the *partial regression transfer function* and denoted

(6.7) $$B_{jk;P}(\omega) = \sum_{v=-\infty}^{\infty} b_{jk;P}(v)e^{-iv\omega};$$

its interpretation is discussed in section 7. Writing (6.6) in the form

(6.8) $$\int_{-\pi}^{\pi} d\omega \, e^{i\omega(u-t)} \left\{ \sum_{k \in P} \overline{B}_{jk;P}(\omega)f_{kh}(\omega) - f_{jh}(\omega) \right\} = 0,$$

we obtain a system of *normal equations* for the regression transfer functions. For each ω in $-\pi \le \omega \le \pi$ and h in P,

(6.9) $$\sum_{k \in P} \overline{B}_{jk;P}(\omega)f_{kh}(\omega) = f_{jh}(\omega).$$

The *partial covariance function* between two time series $X_h(\cdot)$ and $X_j(\cdot)$, given predictors $\{X_k(t), t = 0, \pm 1, \cdots, k \in P\}$, is denoted by $K_{hj;P}(v)$ and is defined by

(6.10) $$K_{hj;P}(t_2 - t_1) = E[\epsilon_{h;P}(t_1)\epsilon_{j;P}(t_2)]$$

where

(6.11) $$\epsilon_{h;P}(t) = X_h(t) - X_{h;P}(t)$$

is the *residual* series of $X_h(\cdot)$ given the predictors. We next show that the partial covariance function depends on t_1 and t_2 only through the time difference $t_2 - t_1$ by a method which also obtains a *spectral representation* for the partial covariance function $E[\epsilon_{h;P}(t_1)\epsilon_{j;P}(t_2)] = E[\epsilon_{h;P}(t_1)X_j(t_2)]$. This is equal to

(6.12) $$E[X_h(t_1)X_j(t_2)] - E\left[\left\{ \sum_{k \in P} \sum_{s=-\infty}^{\infty} b_{hk;P}(s)X_h(t_1 - s) \right\} X_j(t_2) \right]$$

$$= R_{hj}(t_2 - t_1) - \sum_{k \in P} \sum_{s=-\infty}^{\infty} b_{hk;P}(s)R_{kj}(t_2 - t_1 + s)$$

$$= \int_{-\pi}^{\pi} d\omega \, e^{i\omega(t_2-t_1)} \left\{ f_{hj}(\omega) - \sum_{k \in P} \sum_{s=-\infty}^{\infty} b_{hk;P}(s)f_{kj}(\omega)e^{i\omega s} \right\}$$

$$= \int_{-\pi}^{\pi} d\omega \, e^{i\omega(t_2-t_1)} \left\{ f_{hj}(\omega) - \sum_{k \in P} \overline{B}_{hk;P}(\omega)f_{kj}(\omega) \right\}.$$

The *partial spectral density function* of two series $X_h(\cdot)$ and $X_j(\cdot)$ given predictors $\{X_k(t), t = 0, \pm 1, \cdots, k \in P\}$, denoted $f_{hj;P}(\omega)$, may be defined by the condition that it provides a spectral representation for the partial covariance function

(6.13) $$K_{hj;P}(t_2 - t_1) = \int_{-\pi}^{\pi} e^{i\omega(t_2-t_1)}f_{hj;P}(\omega) \, d\omega.$$

From (6.12) we obtain the *basic* formula

(6.14) $$f_{hj;P}(\omega) = f_{hj}(\omega) - \sum_{k \in P} \overline{B}_{hk;P}(\omega)f_{kj}(\omega).$$

To interpret the partial spectral density function, let us first consider the properties of the residual series $\epsilon_{j;P}(t)$. Its spectral density function, called the

residual spectral density function of the series $X_j(\cdot)$ given the predictors $\{X_k(\cdot), \, k \in P\}$, is given by

$$(6.15) \qquad f_{jj;P}(\omega) = f_{jj}(\omega) - \sum_{k \in P} \overline{B}_{jk;P}(\omega) f_{kj}(\omega),$$

$$= f_{jj}(\omega) \{1 - W_{j;P}(\omega)\},$$

defining

$$(6.16) \qquad W_{j;P}(\omega) = 1 - \frac{f_{jj;P}(\omega)}{f_{jj}(\omega)},$$

$$= \sum_{k \in P} \overline{B}_{jk;P}(\omega) f_{kj}(\omega) \div f_{jj}(\omega).$$

One calls $W_{j;P}(\omega)$ the *multiple coherence* function of the series $X_j(\cdot)$ given the predictors $\{X_k(t), \, t = 0, \pm 1, \cdots, k \in P\}$. It is analogous to the squared multiple correlation coefficient and is a measure of the predictability of the component of $X_j(\cdot)$ at frequency ω from the components of $\{X_k(\cdot), \, k \in P\}$ at frequency ω. The analogue of the square of the partial correlation coefficient is called the *partial coherence* between series $X_h(\cdot)$ and $X_j(\cdot)$, given $\{X_k(\cdot), \, k \in P\}$; it is denoted, and given, by

$$(6.17) \qquad W_{hj;P}(\omega) = |f_{hj;P}(\omega)|^2 \div f_{hh;P}(\omega) f_{jj;P}(\omega).$$

It is instructive to consider the case where only a single series $\{X_k(t), \, t = 0, \pm 1, \cdots\}$ is used as the predictor. The regression transfer function $B_{jk;\{k\}}(\omega)$ will be denoted $B_{j;k}(\omega)$, the partial spectral density function $f_{hj;\{k\}}(\omega)$ will be denoted $f_{hj;k}(\omega)$, and the multiple spectral density function $W_{j;\{k\}}(\omega)$ will be denoted $W_{j;k}(\omega)$. We obtain the following formulas: $\overline{B}_{j;k}(\omega) f_{kk}(\omega) = f_{jk}(\omega)$, so that the regression transfer function is given by (assuming $f_{kk}(\omega)$ never vanishes)

$$(6.18) \qquad \overline{B}_{j;k}(\omega) = \frac{f_{jk}(\omega)}{f_{kk}(\omega)}.$$

The partial cross-spectral density function is given by

$$(6.19) \qquad f_{hj;k}(\omega) = f_{hj}(\omega) - \overline{B}_{h;k}(\omega) f_{kj}(\omega),$$

$$= f_{hj}(\omega) - \frac{f_{hk}(\omega) f_{kj}(\omega)}{f_{kk}(\omega)}.$$

In particular, the residual spectral density function is given by

$$(6.20) \qquad f_{jj;k}(\omega) = f_{jj}(\omega) - \frac{|f_{jk}(\omega)|^2}{f_{kk}(\omega)} = f_{jj}(\omega) \{1 - W_{j;k}(\omega)\}$$

where

$$(6.21) \qquad W_{j;k}(\omega) = \frac{|f_{jk}(\omega)|^2}{f_{jj}(\omega) f_{kk}(\omega)}.$$

If one examines the formula for the multiple coherence function of $X_j(\cdot)$ given $X_k(\cdot)$, one sees that the indices j and k play a symmetrical role. We therefore define the symbol

$$(6.22) \qquad W_{jk}(\omega) = \frac{|f_{jk}(\omega)|^2}{f_{jj}(\omega) f_{kk}(\omega)},$$

which is called the *coherence* between the series $X_j(\cdot)$ and $X_k(\cdot)$. It is denoted by

the letter W in commemoration of Norbert Wiener who first introduced the notion of coherence. The coherence is related to a frequency decomposition of the residual series when one uses either of the series $X_j(\cdot)$ and $X_k(\cdot)$ to predict the other. It remains an open question whether it is more informative to plot the coherence $W_{jk}(\omega)$ or the residual spectral density functions $f_{jj;k}(\omega)$ and $f_{kk;j}(\omega)$.

Inductive formulas. Partial spectral densities and regression transfer functions are best computed by adding a variable at a time. Let P be an index set. By $P + m$ we mean the index set $\{j, j \in P \text{ or } j = m\}$; it is understood in this case that m does not belong to P. By $P - m$ we mean the index set $\{j, j \in P \text{ and } j \neq m\}$; it is understood in this case that m belongs to P.

From innovation theory one obtains the basic formula

$$(6.23) \qquad X_{j;P+m}(t) = X_{j;P}(t) + \sum_{s=-\infty}^{\infty} b_{jm;P+m}(t - s)\{X_m(s) - X_{m;P}(s)\}.$$

The regression coefficients $b_{jm;P+m}(t - s)$ are determined by the conditions (for $u = 0, \pm 1, \cdots$)

$$(6.24) \qquad E[X_j(t)\{X_m(u) - X_{m;P}(u)\}] = E[X_{j;P+m}(t)\{X_m(u) - X_{m;P}(u)\}]$$

which lead to the formulas

$$(6.25) \qquad K_{jm;P}(u - t) = \sum_{s=-\infty}^{\infty} b_{jm;P+m}(t - s)K_{mm;P}(u - s),$$

$$f_{jm;P}(\omega) = \overline{B}_{jm;P+m}(\omega)f_{mm;P}(\omega).$$

Thus

$$(6.26) \qquad \overline{B}_{jm;P+m}(\omega) = \frac{f_{jm;P}(\omega)}{f_{mm;P}(\omega)}.$$

Similarly, one derives other inductive formulas from (6.23):

$$(6.27) \qquad \begin{aligned} B_{jk;P+m}(\omega) &= B_{jk;P}(\omega) - B_{jm;P+m}(\omega)B_{mk;P}(\omega), \\ f_{hj;P+m}(\omega) &= f_{hj;P}(\omega) - \frac{f_{hm;P}(\omega)f_{jm;P}(-\omega)}{f_{mm;P}(\omega)}. \end{aligned}$$

More generally, one can conveniently compute a matrix A_P defined as follows. Fix a subset P of indices, and let Q denote the set of indices in D but not in P. Define A_P by

(6.28)

k \ j	P	Q
P	inverse matrix of $\{f_{jk}(\omega), j, k \in P\}$, denoted $\{g_{jk;P}(\omega)\}$	$\overline{B}_{kj;P}(\omega)$, conjugate of regression transfer function at ω
Q	$-B_{jk;P}(\omega)$, negative of regression transfer function	$f_{jk;P}(\omega)$, partial spectral density at ω

Given A_P, and an index m not in P, it can be shown that one forms $A_{P+m} = \{a_{jk;P+m}\}$ by the formulas

$$(6.29) \qquad a_{mm;P+m} = 1 \div a_{mm;P},$$

$$a_{mj;P+m} = a_{mj;P} \div a_{mm;P} \qquad \text{for} \quad j \neq m,$$

$$a_{jm;P+m} = -a_{jm;P} \div a_{mm;P} \qquad \text{for} \quad j \neq m,$$

$$a_{jk;P+m} = a_{jk;P} - a_{jm;P+m}a_{mk;P} \qquad \text{for} \quad j \neq m \quad \text{and} \quad k \neq m.$$

Similarly, given A_P and an index m in P, the same formulas yield a matrix whose entries contain the regression transfer functions and partial cross-spectra for the set of predictors $\{X_k, k \in P, \text{ but } k \neq m\}$.

To prove (6.29), one needs the following formulas for the inverse matrix $\{g_{hj;P+m}(\omega)\} = \{f_{hj}(\omega), h, j \in P + m\}^{-1}$ when one adds an index m to a predictor set P (the argument ω is omitted for ease of writing):

$$(6.30) \qquad g_{mm;P+m} = 1 \div f_{mm;P},$$

$$g_{mh;P+m} = \bar{g}_{hm;P+m} = -B_{mh;P} \div f_{mm;P} \qquad \text{for} \quad h \in P,$$

$$g_{hj;P+m} = g_{hj;P} + \frac{\bar{B}_{mh;P}B_{mj;P}}{f_{mm;P}},$$

$$= g_{hj;P} - g_{hm;P+m}B_{mj;P} \qquad \text{for} \quad h, j \in P.$$

One-sided prediction filters. The predictors considered in the foregoing are two-sided. One often desires to examine one-sided prediction filters.

Let $X_1(\cdot), \cdots, X_q(\cdot)$ be q time series. Let P be a subset of the set $D = \{1, 2, \cdots, q\}$ of indices, and let j be any index in D. We define a new time series, denoted $X_{j;P}^{(r)}(t)$, $t = 0, \pm 1, \cdots$ and called the minimum mean square error linear predictor of $X_j(t)$ given $\{X_k(s), s = t - r, t - r - 1, \cdots, k \in P\}$ as follows: $X_{j;P}^{(r)}(t)$ is a linear combination of the predictor random variables up to time $t - r$, which we write

$$(6.31) \qquad X_{j;P}^{(r)}(t) = \sum_{k \in P} \sum_{s=-\infty}^{t-r} b_{jk;P}^{(r)}(t - s)X_k(s) = \sum_{k \in P} \sum_{v=r}^{\infty} b_{jk;P}^{(r)}(v)X_k(t - v)$$

whose mean square prediction error as an estimate of $X_j(t)$, denoted

$$(6.32) \qquad K_{jj;P}^{(r)} = E[|X_j(t) - X_{j;P}^{(r)}(t)|^2],$$

is minimized. The regression coefficients $b_{jk;P}^{(r)}(t - s)$ are now determined by the conditions

$$(6.33) \qquad E[X_{j;P}^{(r)}(t)X_h(u)] = E[X_j(t)X_h(u)] \qquad \text{for} \quad u \leq t - r \quad \text{and} \quad h \text{ in } P,$$

which lead to the normal equations

$$(6.34) \qquad \sum_{k \in P} \sum_{v=r}^{\infty} b_{jk;P}^{(r)}(v)E[X_k(t - v)X_h(u)] = E[X_j(t)X_h(u)]$$

$$\text{for} \quad u \leq t - r,$$

$$\sum_{k \in P} \sum_{v=r}^{\infty} b_{jk;P}^{(r)}(v)R_{kh}(u - t - v) = R_{jh}(u - t) \qquad \text{for} \quad u - t \leq -r,$$

$$\sum_{k \in P} \sum_{v=r}^{\infty} b_{jk;P}^{(r)}(v)R_{hk}(s + v) = R_{hj}(s) \qquad \text{for} \quad s \geq r.$$

Solving this system of equations for $b_{jk;P}^{(r)}(v)$ is the well-known Wiener-Hopf problem. The regression transfer functions

$$(6.35) \qquad B_{jk;P}^{(r)}(\omega) = \sum_{v=r}^{\infty} b_{jk;P}^{(r)}(v)e^{-iv\omega}$$

can be obtained by a method involving factorization of the spectral density functions which is difficult to carry out (see Whittle [31]). If one is content with a numerical solution on a computer, rather than an analytical solution, one can find the regression coefficients $b_{jk}^{(r)}(v)$ directly and then compute the regression transfer function.

7. Gain and phase

Let $X_1, \cdots, X_r$ be jointly normal random variables. For any subset P of $D = \{1, \cdots, r\}$ and index j not in P, one can form (i) the regression coefficients $\{b_{jk;P}, k \in P\}$, (ii) the partial covariances $K_{hj;P}$ and partial correlation coefficients, and (iii) multiple correlation coefficient.

Similarly, for jointly covariance stationary time series $X_1(\cdot), \cdots, X_r(\cdot)$ one can form (i) the regression transfer functions $\{B_{jk;P}(\omega), k \in P\}$, (ii) the partial spectral density functions $f_{hj;P}(\omega)$ and partial coherence functions $W_{hj;P}(\omega)$, and (iii) the multiple coherence function $W_{j;P}(\omega)$.

The regression transfer function $B_{jk;P}(\omega)$ is best interpreted by regarding it as the frequency transfer function of a filter and introducing its gain and phase.

A discrete time invariant filter is described by its pulse response sequence $\{b_s, s = 0, \pm1, \cdots\}$ or its frequency transfer function

$$(7.1) \qquad B(\omega) = \sum_{s=-\infty}^{\infty} b_s e^{-is\omega}.$$

In terms of pulse response function, the output $\hat{X}(t)$ of the filter corresponding to an input $X(t)$ is given by

$$(7.2) \qquad \hat{X}(t) = \sum_{s=-\infty}^{\infty} b_s X(t-s), \qquad s = 0, \pm1, \cdots.$$

For a sinusoidal input $X(t) = e^{i\omega t}$, the output is $\hat{X}(t) = B(\omega)e^{i\omega t}$. Therefore, for an input which is a superposition of harmonics,

$$(7.3) \qquad X(t) = \int_{-\pi}^{\pi} e^{it\omega} \, dZ(\omega),$$

the output is

$$(7.4) \qquad \hat{X}(t) = \int_{-\pi}^{\pi} e^{it\omega} B(\omega) \, dZ(\omega).$$

The frequency response function $B(\omega)$ of a filter is a complex number which we can write

$$(7.5) \qquad B(\omega) = \alpha(\omega) + i\beta(\omega) = G(\omega)e^{-i\varphi(\omega)}$$

where

$$(7.6) \qquad \alpha(\omega) = \mathrm{Re}B(\omega), \qquad \beta(\omega) = \mathrm{Im}B(\omega).$$

The *gain* $G(\omega)$ and *phase* $\varphi(\omega)$ of a filter are defined by

(7.7)
$$G(\omega) = \sqrt{\alpha^2(\omega) + \beta^2(\omega)},$$

(7.8)
$$\varphi(\omega) = \arctan\{-\beta(\omega)/\alpha(\omega)\} \qquad \text{if} \quad \alpha(\omega) \geq 0,$$
$$= \arctan\{-\beta(\omega)/\alpha(\omega)\} + \pi \operatorname{sign}\{-\beta(\omega)\} \qquad \text{if} \quad \alpha(\omega) < 0.$$

To interpret the gain and phase of a filter, consider an input signal

(7.9)
$$X(t) = f(t)e^{i\omega_0 t}$$

whose frequency spectrum is nonvanishing only in a neighborhood of the frequency ω_0. Further, assume that in this region the gain of the filter is essentially constant and the phase is essentially a linear function of ω. Then the output signal will be a delayed but undistorted replica of the original (see Mason and Zimmerman [16], p. 367):

(7.10)
$$\hat{X}(t) = G(\omega_0)f(t - t_1)e^{i\omega_0(t-t_0)}$$

where

(7.11)
$$t_0 = \left.\frac{\varphi(\omega)}{\omega}\right|_{\omega=\omega_0}, \qquad \text{carrier delay or phase delay};$$

(7.12)
$$t_1 = \varphi'(\omega)|_{\omega=\omega_0}, \qquad \text{envelope delay or group delay}.$$

The terminology "carrier delay" and "envelope delay" is used in the communication theory literature (for example, Mason and Zimmerman [16]). The terminology "phase delay" and "group delay" is used by Robinson ([26], p. 31), who extensively discusses these concepts.

In summary, one way to describe the relations between time series is by describing the characteristics of various regression transfer functions. There are a number of characteristics which need to be looked at: gain, logarithm of gain (or attenuation), phase, phase delay, and group delay. One of the problems of empirical multiple time series analysis is to determine which of these characteristics is most wisely used in routine statistical data reduction of multiple time series.

8. Sampling theory of sample partial and derived cross-spectra

Given a windowed sample spectral density matrix, one can form estimated partial regression transfer functions

(8.1)
$$\hat{B}_{jk;P}(\omega) = \hat{\alpha}_{jk;P}(\omega) + i\hat{\beta}_{jk;P}(\omega),$$
$$= \hat{G}_{jk;P}(\omega) \exp[-i\hat{\varphi}_{jk;P}(\omega)].$$

This estimate is computed by the methods of section 6; by analogy with (6.25) the estimates of $B_{jk;P}(\omega)$ can be explicitly written as

(8.2)
$$\hat{B}_{jk;P}(\omega) = \bar{\hat{f}}_{jk;P-k}(\omega) \div \hat{f}_{kk;P-k}(\omega).$$

By analogy with results of the usual theory of partial correlation (see Kendall and Stuart ([13], p. 333)), one might conjecture that for normal stationary time series with zero means

$$(8.3) \qquad (\hat{B}_{jk;P}(\omega) - B_{jk;P}(\omega)) \left\{ \frac{\hat{f}_{kk;P-k}(\omega)}{C\hat{f}_{jj;P}(\omega)} \right\}^{1/2}$$

is asymptotically complex normal with mean 0 and variance 1.

To establish the plausibility of (8.3), let us relate it to certain established results for ordinary cross-spectral analysis (compare Jenkins [11]). In the case that the prediction set P contains only the predictor k, we write

$$(8.4) \qquad \hat{B}_{j;k}(\omega) = \hat{\alpha}_{j;k}(\omega) + i\hat{\beta}_{j;k}(\omega)$$

for the sample regression transfer function. The estimates are formed by

$$(8.5) \qquad \hat{\alpha}_{j;k}(\omega) = \frac{\hat{c}_{jk}(\omega)}{\hat{f}_{kk}(\omega)}, \qquad \hat{\beta}_{j;k}(\omega) = \frac{\hat{q}_{jk}(\omega)}{\hat{f}_{kk}(\omega)}.$$

The variance of $\hat{\alpha}$ can be derived by the well-known delta method (compare Kendall and Stuart ([13], vol. I, p. 231)); writing $\hat{c}$ and $\hat{f}$, respectively, for the numerator and denominator of $\hat{\alpha}$,

$$(8.6) \qquad \text{var} \, [\hat{\alpha}] = \frac{\text{var} \, [\hat{c}]}{E^2[\hat{f}]} - \frac{2 \, \text{cov} \, [\hat{c}, \hat{f}] E[\hat{c}]}{E^3[\hat{f}]} + \frac{\text{var} \, [\hat{f}] E^2[\hat{c}]}{E^4[\hat{f}]},$$

one obtains the asymptotic covariances

$$(8.7) \qquad \text{var} \, [\hat{\alpha}_{j;k}(\omega)] = \text{var} \, [\hat{\beta}_{j;k}(\omega)] = \tfrac{1}{2} C \frac{f_{jj}(\omega)}{f_{kk}(\omega)} \{1 - W_{jk}(\omega)\},$$

$$(8.8) \qquad \text{cov} \, [\hat{\alpha}_{j;k}(\omega), \hat{\beta}_{j;k}(\omega)] = 0.$$

Therefore,

$$
\begin{aligned}
(8.9) \qquad \text{var} \, [\hat{B}_{j;k}(\omega)] &= C \frac{f_{jj}(\omega)}{f_{kk}(\omega)} \{1 - W_{jk}(\omega)\}, \\
&= C \frac{f_{jj;P}(\omega)}{f_{kk;P-k}(\omega)},
\end{aligned}
$$

which agrees with (8.3).

Under the assumptions $\text{var} \, [\hat{\alpha}] = \text{var} \, [\hat{\beta}]$ and $\text{cov} \, [\hat{\alpha}, \hat{\beta}] = 0$, the gain $\hat{G}$ and phase $\hat{\varphi}$ defined by

$$(8.10) \qquad \hat{G} e^{-i\varphi} = \hat{\alpha} + i\hat{\beta}$$

have asymptotic variances (by the delta method)

$$(8.11) \qquad \text{var} \, [\hat{G}] = \text{var} \, [\hat{\alpha}], \qquad \text{var} \, [\hat{\varphi}] = \frac{1}{G^2} \text{var} \, [\hat{\alpha}]$$

where $G e^{-i\varphi} = \alpha + i\beta$. In view of (8.3) and (8.7), we conjecture that

$$\text{var } [\hat{\alpha}_{jk;P}(\omega)] = \text{var } [\hat{\beta}_{jk;P}(\omega)],$$

$$= \tfrac{1}{2}C \frac{f_{jj;P}(\omega)}{f_{kk;P-k}(\omega)},$$

(8.12)
$$= \tfrac{1}{2}C \left|\frac{f_{jk;P-k}(\omega)}{f_{kk;P-k}(\omega)}\right|^2 \frac{f_{jj;P}(\omega)f_{kk;P-k}(\omega)}{|f_{jk;P-k}(\omega)|^2},$$

$$= \tfrac{1}{2}C|G_{jk;P}(\omega)|^2 \frac{f_{jj;P}(\omega) \div f_{jj;P-k}(\omega)}{W_{jk;P-k}(\omega)},$$

$$= \tfrac{1}{2}C|G_{jk;P}(\omega)|^2 \left\{\frac{1}{W_{jk;P-k}(\omega)} - 1\right\}.$$

From (8.11) and (8.12) one obtains expressions for the asymptotic variances of the partial gain and phase. In particular,

(8.13)
$$\text{var } [\hat{\varphi}_{jk;P}(\omega)] = \text{var } [\log_e \hat{G}_{jk;P}(\omega)],$$

$$= \tfrac{1}{2}C \left\{\frac{1}{W_{jk;P-k}(\omega)} - 1\right\}.$$

One may interpret (8.13) in words as follows: the variability of the estimated partial attenuation (log gain) and phase is determined by the partial coherency $W_{jk;P-k}(\omega)$; in particular, the variance tends to 0 as the partial coherence tends to 1. These results provide one interpretation of partial coherency.

To actually compute partial regression functions and their sampling error, one should use the algorithm (6.21), since using (6.30) one can rewrite (8.3): for any index k in P, asymptotic variance of $\hat{B}_{jk;P}(\omega)$ is $Cf_{jj;P}(\omega) \, g_{kk;P}(\omega)$. Stopping rules for selecting a significant set P of indices remain to be investigated.

9. Mean and bias of cross-spectral estimates

The behavior and interpretation of windowed sample cross-spectral density functions cannot be understood on the basis of their variability theory alone. Their means must be investigated.

To study the means of windowed sample spectra, one needs to consider two possible assumptions for the observed time series: (i) they are jointly covariance stationary with zero means, (ii) they are the sum of mean value functions and jointly covariance stationary zero mean fluctuations. Only case (i) is discussed in this paper.

We consider separately auto-spectra and cross-spectra. Asymptotic expressions for the means of windowed sample auto-spectral density functions have been studied by many writers, especially Parzen [18] and Hext [10]. We consider only the case that the spectral window satisfies the assumptions $K_{M,1}(\omega) = 0$ and $K_{M,2}(\omega) > 0$, defining

(9.1)
$$K_{M,\nu}(\omega) = \int_{-\pi}^{\pi} (\lambda - \omega)^\nu K_M(\lambda - \omega) \, d\lambda.$$

Then the mean of a windowed sample auto-spectral density function may be approximated

$$E[\hat{f}_{jj}(\omega)] \doteq \int_{-\pi}^{\pi} K_M(\lambda - \omega) f_{jj}(\lambda) \, d\lambda,$$

(9.2)

$$\doteq f_{jj}(\omega) - \frac{1}{2M^2} k''(0) f''(\omega),$$

where

(9.3)
$$k''(0) = - \int_{-\infty}^{\infty} \omega^2 K(\omega) \, d\omega$$

is the value at 0 of the second derivative of the covariance kernel $k(u) = \int_{-\infty}^{\infty} e^{iu\omega} K(\omega) \, d\omega$. We digress for a moment to note that some authors (Daniels [7], Akaike [1]) have suggested that the spectral window K_M be chosen so that $K_{M,\nu}(\omega) = 0$ for as many values of ν as possible. While this reduces the bias, it necessarily leads to possibly negative estimates which may lead to difficulties of interpretation of spectral estimates.

In evaluating the mean of a windowed sample cross-spectral density function,

(9.4)
$$E[\hat{f}_{jk}(\omega)] \doteq \int_{-\pi}^{\pi} K_M(\lambda - \omega) f_{jk}(\lambda) \, d\lambda,$$

it is most convenient to express $f_{jk}(\lambda)$ in terms of the true regression transfer function

(9.5)
$$\overline{B}_{j;k}(\lambda) = \frac{f_{jk}(\lambda)}{f_{kk}(\lambda)} = G_{j;k}(\lambda) \exp \left[i\varphi_{j;k}(\lambda) \right]$$

by

(9.6)
$$f_{jk}(\lambda) = f_{kk}(\lambda) G_{j;k}(\lambda) \exp \left[i\varphi_{j;k}(\lambda) \right].$$

To understand the special sources of bias in cross-spectral estimation, let us first find the leading term of the mean $E[\hat{f}_{jk}(\omega)]$ by assuming that in the region $|\omega - \lambda| \leq B$ where $K_M(\omega - \lambda)$ is appreciably nonzero, both the auto-spectral density $f_{kk}(\cdot)$ and the gain are practically constant while the phase is linear; then approximately

(9.7)
$$f_{jk}(\lambda) = f_{kk}(\omega) G_{j;k}(\omega) \exp \left[i\{\varphi_{j;k}(\omega) + (\lambda - \omega)\varphi'_{j;k}(\omega)\} \right],$$
$$= f_{jk}(\omega) \exp \left[i(\lambda - \omega)\varphi'_{j;k}(\omega) \right];$$

recall that the phase derivative $\varphi'_{j;k}(\omega)$ may be interpreted as a group delay or carrier delay. From (9.7) it follows that

$$E[\hat{f}_{jk}(\omega)] = f_{jk}(\omega) \int_{-\pi}^{\pi} K_M(\lambda - \omega) \exp \left[i(\lambda - \omega)\varphi'_{j;k}(\omega) \right] d\lambda,$$

(9.8)
$$\doteq f_{jk}(\omega) \int_{-\infty}^{\infty} M K(M\mu) \exp \left[i\mu\varphi'_{j;k}(\omega) \right] d\mu,$$

$$= f_{jk}(\omega) k(\varphi'_{j;k}(\omega)/M).$$

In words, if the truncation point M is not chosen large compared to the group

delay, there will be an appreciable bias in estimating the cross-spectral density function.

A possible method of avoiding this source of bias in cross-spectral density estimation is to use *shifted* cross-spectral estimates, which we now define (this method is due to Akaike [3]).

Let L be an integer (positive or negative). Define the shifted windowed sample cross-spectral density function with shift L by

$$(9.9) \qquad f_{jk;T,M,L}(\omega) = \frac{1}{2\pi} \sum_{|v|<M} e^{-iv\omega} k\left(\frac{v}{M}\right) R_{jk;T}(v + L).$$

One may verify that its mean is approximately given by

$$(9.10) \qquad \begin{aligned} E[f_{jk;T,L,M}(\omega)] &\doteq \int_{-\pi}^{\pi} d\lambda\, f_{jk}(\lambda) e^{i\lambda L} K_M(\omega - \lambda), \\ &\doteq \int_{-\pi}^{\pi} d\lambda\, K_M(\omega - \lambda) e^{i\lambda L} f_{kk}(\lambda) G_{j;k}(\lambda) e^{i\varphi_{j;k}(\lambda)}\, d\lambda. \end{aligned}$$

Using the same approximations as before, one may show that the mean is approximately equal to

$$(9.11) \qquad f_{jk}(\omega) e^{i\omega L} \int_{-\pi}^{\pi} d\lambda\, K_M(\lambda - \omega) e^{i(\lambda-\omega)L} e^{i(\lambda-\omega)\varphi'_{j;k}(\omega)}.$$

Finally, one obtains the following approximation:

$$(9.12) \qquad E[f_{jk;T,L,M}(\omega)] = f_{jk}(\omega) e^{i\omega L} k\left(\frac{L + \varphi'(\omega)}{M}\right) + \text{terms in } \frac{1}{M^2}.$$

If L is so chosen that

$$(9.13) \qquad L + \varphi'_{j;k}(\omega) \ll M,$$

then an approximately unbiased estimate of $f_{jk}(\omega)$ is given by

$$(9.14) \qquad e^{-i\omega L} f_{jk;T,M,L}(\omega).$$

The question of how to choose L remains; it may vary with ω and may have to be estimated from the sample phase. As a first guess, it could be taken to be the lag at which the sample cross-covariance function $R_{jk;T}(v)$ achieves its maximum absolute value.

We do not discuss here the terms in the bias of cross-spectral estimates which are of the order of $1/M^2$; they are analogous to the bias of auto-spectral estimates. It should be noted that the foregoing derivations are very heuristic; a complete and rigorous discussion is given by Nigel Nettheim in his Stanford Ph.D. thesis.

If one investigates (using the delta method) how the bias in cross-spectral density estimates propagates into the estimates of derived cross-spectral quantities, one finds that the bias is present in the estimated coherence but is absent in the estimated phase. It would seem that corrections for bias could be introduced using the estimated phase derivative. It remains to be investigated whether it would not be wise to directly estimate the phase derivative (group delay)

$$\varphi'_{j:k}(\omega) = \frac{d}{d\omega} \arctan \frac{q_{jk}(\omega)}{c_{jk}(\omega)},$$

(9.15)

$$= \frac{c_{jk}(\omega)q'_{jk}(\omega) - q_{jk}(\omega)c'_{jk}(\omega)}{c^2_{jk}(\omega) + q^2_{jk}(\omega)},$$

by directly estimating the derivatives of the co-spectral and quadrature-spectral density functions.

The group delay (or phase derivative) should be routinely estimated in cross-spectral analysis since it seems easier to interpret than the phase. Further, the phase may be estimated without ambiguities modulo 2π by integrating (by Simpson's rule) the phase derivative.

REFERENCES

[1] H. Akaike, "On the design of lag window for the estimation of spectra," *Ann. Inst. Statist. Math.*, Vol. 14 (1962), pp. 1–21.

[2] H. Akaike and Y. Yamanouchi, "On the statistical estimation of frequency response function," *Ann. Inst. Statist. Math*, Vol. 14 (1962), pp. 23–56.

[3] H. Akaike, "Statistical measurement of frequency response function," *Ann. Inst. Statist. Math.*, Supplement III, Vol. 15 (1964), pp. 5–17.

[4] D. E. Amos and L. H. Koopmans, "Tables of the distribution of the coefficient of coherence for stationary bivariate Gaussian processes," (1963), Albuquerque, New Mexico, Sandia Corporation (available from the Office of Technical Services, Department of Commerce, Washington, D. C.).

[5] R. B. Blackman and J. Tukey, *The Measurement of Power Spectra from the Point of View of Communication Engineering*, New York, Dover, 1959.

[6] H. Bohman, "Approximate Fourier analysis of distribution functions," *Ark. Mat.*, Vol. 4 (1960), pp. 99–157.

[7] H. E. Daniels, "The estimation of spectral densities," *J. Roy. Statist. Soc. Ser. B*, Vol. 24 (1962), pp. 185–198.

[8] W. F. Freiberger, "Approximate distributions of cross-spectral estimates for Gaussian processes," *Time Series Analysis Symposium Proceedings*, edited by M. Rosenblatt, New York, Wiley, 1963, pp. 244–259.

[9] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution," *Ann. Math. Statist.*, Vol. 34 (1963), pp. 152–177.

[10] G. Hext, "A new approach to time series with mixed spectra," Ph.D. thesis, Statistics Department, Stanford University, 1966.

[11] G. M. Jenkins, "Cross-spectral analysis and the estimation of linear open loop transfer functions," *Time Series Analysis Symposium Proceedings*, edited by M. Rosenblatt, New York, Wiley, 1963, pp. 267–278.

[12] ———, "Some examples of and comments on spectral analysis," *Proceedings IBM Scientific Computing Symposium on Statistics*, White Plains, IBM, 1965, pp. 205–246.

[13] M. G. Kendall and A. Stuart, *Advanced Theory of Statistics*, Vols. 1 and 2, London, Griffin, 1958 and 1961.

[14] L. H. Koopmans, "On the coefficient of coherence for weakly stationary stochastic processes," *Ann. Math. Statist.*, Vol. 35 (1964), pp. 532–549.

[15] ———, "On the multivariate analysis of weakly stationary stochastic processes," *Ann. Math. Statist.*, Vol. 35 (1964), pp. 1765–1780.

[16] S. J. Mason and H. J. Zimmerman, *Electronic Circuits, Signals, and Systems*, New York, Wiley, 1960.

[17] V. K. MURTHY, "Estimation of the cross spectrum," *Ann. Math. Statist.*, Vol. 34 (1963), pp. 1012–1021.
[18] E. PARZEN, "On consistent estimates of the spectrum of a stationary time series," *Ann. Math. Statist.*, Vol. 28 (1957), pp. 329–348.
[19] ———, "On choosing an estimate of the spectral density function of a stationary time series," *Ann. Math. Statist.*, Vol. 28 (1957), pp. 921–932.
[20] ———, "On asymptotically efficient consistent estimates of the spectral density function of a stationary time series," *J. Roy. Statist. Soc. Ser. B*, Vol. 20 (1958), pp. 303–322.
[21] ———, "Mathematical considerations in the estimation of spectra," *Technometrics*, Vol. 3 (1961), pp. 167–190.
[22] ———, "An approach to empirical time series analysis," *Radio Science*, Vol. 68D (1964), pp. 937–957.
[23] ———, "On statistical spectral analysis," *Proceedings of Symposia in Applied Mathematics*, Vol. XVI, Providence, American Mathematical Society, 1964, pp. 221–246.
[24] ———, "The role of spectral analysis in time series analysis," International Statistical Institute (Belgrade meeting, September, 1965).
[25] ———, *Empirical Time Series Analysis*, San Francisco, Holden-Day, 1967.
[26] E. A. ROBINSON, *Random Wavelets and Cybernetic Systems*, London, Griffin, 1962.
[27] M. ROSENBLATT, "Statistical analysis of stochastic processes with stationary residuals," *Probability and Statistics* (Cramér volume), edited by U. Grenander, New York, Wiley, 1959, pp. 300–330.
[28] *Technometrics*, Papers on Spectral Analysis of Time Series, Vol. 3 (1961), pp. 133–268.
[29] J. W. TUKEY, "An introduction to the measurement of spectra," *Probability and Statistics* (Cramér volume), edited by U. Grenander, New York, Wiley, 1959, pp. 1283–1289.
[30] ———, "Data analysis and the frontiers of geophysics," *Science*, Vol. 148 (1965), pp. 1283–1289.
[31] P. WHITTLE, "On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix," *Biometrika*, Vol. 50 (1963), p. 129.