Invited Paper, Royal Statistical Society Conference, Cardiff, Wales Sept 29 - Oct 1, 1964

A Frequency Approach to the Techniques of Principle Components, Factor Analysis and Canonical Variates in the Case of Stationary Time Series.

By DAVID R. BRILLINGER
Princeton University

SUMMARY

In the subject of multivariate time series analysis, there arise the direct analogs of problems problems that are studied in classical multivariate analysis. If the series involved are stationary one is able to estimate relevant statistics from a single realization. The purpose of this paper is to derive analogs of principle components, factors and canonical variates that may be estimated from a single realization. The distribution of the proposed estimates is approximated. In addition an adaptive approach is presented that may be used in the case of a class non-stationary series. to investigate stationarity

1. INTRODUCTION

Consider a multivariate, stationary, Gaussian time series $\{x_1(t), \ldots, x_k(t)\}$ with zero mean and power spectral density matrix $f(\omega) = ||f_{ij}(\omega)||$, where t is either discrete or continuous. The matrix $f(\omega)$ may be thought of as the variance-covariance matrix of a particular set of Gaussian variates. Namely if,

$$x_i(t) = \int e^{i\omega t} \, d\, Z_i(\omega) \;,$$

is the Cramer representation of the process $x_i(t)$ then,

$$f_{ij}(\omega)\, d\omega = E\, \{d\, Z_i^*(\omega)\; d\, Z_j(\omega)\} \;.$$

This result leads one to consider analyzing the series by applying techniques developed for the analysis of a sample from a multidimensional Gaussian distribution.

Specifically, suppose that an estimate $\hat{f}_{ij}(\omega)$ of $f_{ij}(\omega)$ is available where $\hat{f}_{ij}(\omega)$ is a weighted average of the cross-periodogram. If the weighted average corresponds to a positive kernel the analogy with the analysis of a multivariate Gaussian sample continues for in this case $||\hat{f}_{ij}(\omega)||$ may be written (in the discrete case) in the form,

$$\frac{1}{T} \sum_{j=1}^{T} \xi_j \bar{\xi}_j' \quad ,$$

where ξ_j is a complex-valued vector Gaussian random variable with mean 0 and variance-covariance a weighted average of $f(\omega)$. (There is a similar expression in the continuous case.) The ξ_j above are dependent $j = 1, \ldots, T$; however it will be seen later that they may be approximated by ΩT independent variables where Ω is the bandwidth of the kernel. That is one is in possession of an approximate sample from a multivariate Gaussian distribution.

The particular classical multivariate techniques of analysis of such a sample to be considered in this paper include, principle components, factor analysis and canonical variates. The results obtained from such an analysis have applications to problems such as, the construction of economic indices, the analysis of signals received by an array of seismic stations, and the enquiry into the relationship between the components of pressure and velocity in a wind field.

The distribution of the statistics proposed in this paper may be approximated by the distributions of similar statistics derived from a complex Gaussian sample. These latter distributions were provided recently by James (1964).

2. PRINCIPLE COMPONENTS

Bowley (1920) makes the following statement, "Index numbers are used to measure the change in some quantity which we cannot observe directly, which we know to have a definite influence on many other quantities which we can so observe, tending to increase all, or diminish all, while this influence is concealed by the action of many causes affecting the separate quantities in various ways."

Admittedly the concept of a quantity which can not be observed directly has many philosophical pitfalls connected with it, but still the intent of Bowley's definition does seem reasonable. Let us not commence with a definition involving unobservable quantities, but instead let us think of an index number as a time series describing in a simple manner the complex concomitant behaviour present in a set of stationary time series. Specifically suppose that one is dealing with time series $x_1(t), \ldots, x_k(t)$ defined for $t = 0, \pm 1, \pm 2, \ldots$ and one wishes to find functions $h_i(t)$ such that subject to a suitable normalization the series,

$$\zeta(t) = \sum_i \sum_\tau h_i(t - \tau)\, x_i(\tau) ,$$

explains, in a variance sense, as much of the variation present in the series $\{x_i(t)\}$ as is possible.

An example of a use to which such a series $\zeta(t)$ might be put is to describe the comovement of the prices of a number of securities.

The following theorem presents an explicit method of calculating index numbers as defined above.

<u>Theorem:</u> Consider a multivariate, wide-sense stationary time series $\{x_1(t), \ldots, x_k(t)\}$. Consider time series $\zeta(t)$ of the form,

$$\sum_i \sum_\tau h_i(t - \tau)\ x_i(\tau) \ .$$

Among all time series of this form, subject to the condition,

$$\sum_{i,j} H_i(\omega)\ K_{ij}(\omega)\ H_j^*(\omega) = 1 \ , \qquad *$$

where $H_i(\omega)$ is the Fourier transform of $h_i(t)$, and the matrix $||K_{ij}(\omega)||$ is Hermitian, the series with maximum variance is the one generated by the functions $H_i(\omega)$ such that,

$$\sum_j \{f_{ij}(\omega) - \lambda(\omega)\ K_{ij}(\omega)\}\ H_j(\omega) = 0 \qquad (2.1)$$

where $\lambda(\omega)$ is the maximum latent root satisfying (2.1).

<u>Proof:</u> The variance of $\zeta(t)$ may be written,

$$\int_{-\pi}^{\pi} \sum_{i,j} H_i(\omega)\ f_{ij}(\omega)\ H_j^*(\omega)\ d\omega \ .$$

A classic result now applies to yield the result that the expression,

$$\sum_{i,j} H_i(\omega)\ f_{ij}(\omega)\ H_j^*(\omega)$$

is maximised subject to the condition,

$$\sum_{i,j} H_i(\omega)\ K_{ij}(\omega)\ H_j^*(\omega) = 1 \ ,$$

by selecting the $H_i(\omega)$ as described in the theorem.

The variance of the $\zeta(t)$ actually achieved by the above choice of $H_i(\omega)$ at each frequency is,

$$\int_{-\pi}^{\pi} \lambda_{max}(\omega)\ d\omega$$

The series $\zeta(t)$ may now be determined by constructing filters $h_i(t)$ having the desired transfer functions $H_i(\omega)$.

In practice $f_{ij}(\omega)$ in (2.1) will be replaced by an estimate $\hat{f}_{ij}(\omega)$. Reasonable choices for $K_{ij}(\omega)$ to use in practice would appear to be,

$$K_{ij}(\omega) = \delta_{ij} \tag{2.2}$$

$$K_{ij}(\omega) = \delta_{ij}\ f_{ii}(\omega) \tag{2.3}$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

(2.2) appears to be applicable if the series may reasonably be thought of as being in the same scale, while (2.3) appears useful if it is desired to eliminate some of the effects of scale and is equivalent to carrying out the analysis on the matrix of coherencies.

The equation (2.1) will have k roots. At this point we have only made use of the largest of these at each frequency; however as in the classical situation the remaining roots may well be most useful. They explain in a [variance] correlation sense the variation remaining in the series after the component $\zeta(t)$ has been removed, the second largest root explaining the largest possible amount of the remaining variation and so on.

One may consequently construct a sequence of series $\zeta_1(t), \ldots, \zeta_k(t)$ uncorrelated with one another, explaining the movement of the complex of time series $\{x_1(t), \ldots, x_k(t)\}$ to various degrees.

The existence of the series $\zeta_2(t), \ldots, \zeta_k(t)$ leads one to a position that many constructors of indices do not appear to have adopted, that is, in many situations it may well be meaningful to quote a number of indices.

The first index typically providing a form of average of the series involved, the second a form of average of the deviations from the first index and so on.

The index arrived at in this section if evaluated for a number of economic time series differs from the sort of index arrived at by Rhodes (1937) and by Stone (1947) in that it makes use of all the observed values of the series as more than weights. To make this contrast more apparent, Rhodes and Stone would arrive at the same index no matter what the time order of the basic data they use.

If one is considering any form of projection, making use of the order of the observations does appear essential.

3. REDUCTION

One way of viewing the construction of indices is to view it as the reduction of a number of series to fewer series with as little a loss of information as is possible. Let us quantize this viewpoint as follows; consider the time series $\{x_1(t), \ldots, x_k(t)\}$. By means of linear time invariant operations reduce these k series to ℓ series, that is form

$$y_j(t) = \sum_i \sum_\tau h_{ij}(t - \tau)\, x_i(\tau) \qquad j = 1, \ldots, \ell\ .$$

Now by means of linear operations on the ℓ series $y_j(t)$ return to k series

$$x_i'(t) = \sum_j \sum_\tau g_{ij}(t - \tau)\, y_j(\tau) \qquad i = 1, \ldots, k\ .$$

What are the functions $h_{ij}(t)$, $g_{ij}(t)$ such that as little information as possible is lost by these operations? Interpreting the loss of information as a mean-squared error, this question may be answered as follows; the mean-squared error of $x'(t) - x(t)$ is,

$$\text{trace } \{\text{var-covar matrix of } x_i'(t) - x_i(t)\} \quad .$$

This may be written as,

$$\int \text{trace } (I - G(\omega)\ H(\omega))\ f(\omega)\ (I - G(\omega)\ H(\omega))^*\ d\omega \qquad (3.1)$$

where $f(\omega)$ denotes the spectral matrix of the series and $G(\omega)$, $H(\omega)$ denote the matrices of the Fourier transforms of $g_{ij}(t)$, $h_{ij}(t)$. I is a $k \times k$ identity matrix.

The essential point in the problem of minimizing (3.1) is to note that the matrix product $G(\omega)\ H(\omega)$ is of rank ℓ .

Consider the problem of minimizing

$$\text{trace } (I - M)^T R(I - M)$$

where M is a matrix of rank ℓ and size $k \times k$, R is $k \times k$ and positive definite. This problem is equivalent to the problem of minimizing,

$$\text{trace } (R^{\frac{1}{2}} - R^{\frac{1}{2}}M)^T (R^{\frac{1}{2}} - R^{\frac{1}{2}}M),$$

or of minimizing

$$\text{trace } (R^{\frac{1}{2}} - N)^T (R^{\frac{1}{2}} - N),$$

where N is of rank ℓ.

The matrix N providing the required minimum is given by,

$$\hat{R} = W_e^T W_e$$

(see Eckart and Young (1936)) where W_e is an $\ell \times k$ matrix having as its ℓ rows, the ℓ rows of the eigenvector matrix of R corresponding to the largest ℓ eigenvectors.

This result remains true if R is a complex hermitian matrix and in terms of (3.1), one is led to the calculation of the eigenvalues and eigenvectors of the spectral matrix once again.

Elementary manipulations indicate $H = W_e$ and $G = R^{\frac{1}{2}} W_e^T$, $R \equiv f(\omega)$, as possible choices of $H(\omega)$ and $G(\omega)$.

Making use of $H(\omega)$ one can consequently construct ℓ series $y_j(t)$, linearly related to the original series $x_i(t)$, that in the sense of being able to rederive the original series $x_i(t)$ back from them, contain as much of the information as is possible. The series $y_j(t)$ are seen to correspond to the principle components of Section 2.

However something more has been added at this stage. One has been led to think of $x_i(t)$ as being of the structure

$$x_i(t) \sim \sum_j \sum_\tau g_{ij}(t - \tau)\, y_j(\tau) \quad .$$

This structure leads one into the very specific factor analysis model of the next section.

4. FACTOR ANALYSIS

Suppose that k seismometers have been set up in an array of k seismic stations. Let $x_i(t)$ denote the output of the seismometer at the i^{th} station at time t. Suppose that some sort of seismic event has taken place. The output $x_i(t)$ then has the form,

$$x_i(t) = s_i(t) + n_i(t) \qquad i = 1, \ldots, k$$

where $s_i(t)$ is the component of the output attributable to the seismic event and $n_i(t)$ is the undesired, but ever present noise. If it is assumed that the earth in the region containing the k stations behaves like a linear filter. The signals $s_i(t)$ are then related to the "true" signal s(t) by,

$$s_i(t) = \sum_\tau h_i(t - \tau)\, s(\tau) \ .$$

Consequently an imagined model for the k records $x_i(t)$ is

$$x_i(t) = \sum_\tau h_i(t - \tau)\, s(\tau) + n_i(t) \qquad (4.1)$$

and one is interested in estimating $h_i(t)$ and s(t).

One has consequently been led to a very specific structural model. This example is a particular case of the factor analysis model, namely imagining that there exist factor series $y_j(t)$, loading series $g_{ij}(t)$, noise series $n_i(t)$ such that

$$x_i(t) = \sum_j \sum_\tau g_{ij}(t - \tau)\, y_j(\tau) + n_i(t) \qquad (4.2)$$

Assuming that the $y_j(t)$ are independent unit white noise series uncorrelated with the $n_i(t)$ and that the latter are uncorrelated with one another and have power spectral matrix $\Sigma(\omega)$, one obtains from (4.2)

$$f(\omega) = G^T(\omega)\, G(\omega) + \Sigma(\omega) \ .$$

This is the basic equation of factor analysis, see Harmon (1960). One may now proceed to the estimation of $G(\omega)$ by means of the classical techniques of factor analysis, e.g. inserting comunalities, evaluating latent roots etc.

The reader may wonder at the apparent dissimilarity of the models (4.1) and (4.2). The model (4.2), specifically the white noise assumption for the factors, finds meaning in Ricker's (1940) theory of wavelets as developed by Robinson (1957). The $g_{ij}(t)$ represent the wavelets received at location i from the impulse excitation $y_j(t)$ at some location. The $g_{ij}(t)$ may be thought of as representing the effect that the earth has upon the various impulses.

The advantage of the model presented herein is that it allows the separation of a number of disturbances. Also one can estimate the direction of the various disturbances and the velocities with which they are propagating.

5. CANONICAL VARIATES AND CANONICAL CORRELATIONS

Suppose that one is considering two sets of time series $\{x_1(t), \ldots, x_k(t)\}$ and $\{y_1(t), \ldots, y_m(t)\}$ and that one is interested in representing a possible relationship between the two sets of series. More specifically suppose ~~that one wishes to determine in some sense~~ one wishes a technique that in some sense determines the extent to which the two sets of series reflect the same underlying traits. A quantitative means of answering this question would be to find that linear combination of the first set of series which is most highly correlated with a linear combination of the second set of series. One could continue by finding the pair of next most highly correlated series orthogonal to the first pair and so on. In a certain sense what are being found are factors accounting for the correlations between the two sets of time series.

As an example of the sort of problem being considered here, suppose that at a particular point in the atmosphere one is able to measure the three components of wind velocity $(u(t), v(t), w(t))$ and also one is able to measure the three components of the pressure gradient $(p_u(t), p_v(t), p_w(t))$ at time t. Theory leads one to imagine that there is a linear relationship between the two sets of measurements, specifically one can imagine the existence of functions $a(t)$, $b(t)$, $c(t)$; $\alpha(t)$, $\beta(t)$, $\gamma(t)$

such that,

$$\int a(t-\tau)\, u(\tau)\, d\tau + \int b(t-\tau)\, v(\tau)\, d\tau + \int c(t-\tau)\, w(\tau)\, d\tau$$
$$= \int \alpha(t-\tau)\, p_u(\tau)\, d\tau + \int \beta(t-\tau)\, p_v(\tau)\, d\tau + \int \gamma(t-\tau)\, p_w(\tau)\, d\tau \quad .$$

In the investigation of the existence of such a relationship one may perhaps take advantage of the following result.

<u>Theorem</u> Let there be given two sets of zero mean wide sense stationary time series $\{x_1(t), \ldots, x_k(t)\}$ and $\{y_1(t), \ldots, y_m(t)\}$. The series

$$\alpha(t) = \sum_i \sum_\tau g_i(t-\tau)\; x_i(\tau),$$

$$\beta(t) = \sum_j \sum_\tau h_i(t-\tau)\; y_j(\tau),$$

which have the extrema co-spectra $\mathrm{Re}\{f_{\alpha\beta}(\omega)\}$ at each ω subject to

$$G^T(\omega)\; f_{xx}(\omega)\; G^*(\omega) = 1, \tag{5.1}$$

$$H^T(\omega)\; f_{yy}(\omega)\; H^*(\omega) = 1, \tag{5.2}$$

(where G and H denote the column vectors of Fourier transforms of $g_i(t)$ and $h_j(t)$) are provided by first solving the determinental equation;

$$|\lambda(\omega)\; f_{xx}(\omega) - f_{xy}(\omega)\; f_{yy}^{-1}(\omega)\; f_{yx}(\omega)| = 0 \; ,$$

where the spectral density matrix of the process $\{x_1(t), \ldots, x_k(t);$ $y_1(t), \ldots, y_m(t)\}$ has been partitioned as follows;

$$\left\| \begin{matrix} f_{xx}(\omega) & f_{xy}(\omega) \\ f_{yx}(\omega) & f_{yy}(\omega) \end{matrix} \right\|$$

f_{xx} being k x k, f_{yy} being m x m.

Then solving

$$\left\| \begin{matrix} \rho(\omega)\ f_{xx}(\omega) & -\ f_{xy}(\omega) \\ -f_{yx}(\omega) & \rho(\omega)\ f_{yy}(\omega) \end{matrix} \right\| \cdot \begin{bmatrix} G(\omega) \\ H(\omega) \end{bmatrix} = 0 \qquad (5.3)$$

where $\rho(\omega) = \sqrt{\lambda(\omega)}$.

The two solutions $\alpha(t)$, $\beta(t)$ having the greatest co-spectrum at each ω correspond to the largest $\lambda(\omega)$ at each frequency ω.

Proof: The co-spectrum of $\alpha(t)$ and $\beta(t)$ at frequency ω is,

$$\text{Re}\ \{G^T(\omega)\ f_{xy}(\omega)\ H^*(\omega)\}$$

$$= \frac{1}{2}\ \{G^T(\omega)\ f_{xy}(\omega)\ H^*(\omega) + G^{*T}(\omega)\ f_{yx}(\omega)\ H(\omega)\ \} \qquad (5.4)$$

The extrema of this expression, subject to the conditions (5.1), (5.2), may be found by the differentiation of,

$$(5.4) + \rho(\omega)\ G^T(\omega)\ f_{xx}(\omega)\ G^*(\omega) + \sigma(\omega)\ H^T(\omega)\ f_{yy}(\omega)\ H^*(\omega)$$

with respect to the real and imaginary parts of the components of $G(\omega)$ and $H(\omega)$, $\rho(\omega)$, $\sigma(\omega)$ being undetermined Lagrange multipliers.

Carrying out this differentiation one is quickly led to the equation (5.3).

The remainder of the theorem results from simple algebraic manipulations.

It seems relevant to point out that if one is concerned over possible differences in scale of the series involved he may carry out the procedure proposed above on the matrix of coherencies.

6. AN ADAPTIVE APPROACH

There is another quite different approach that may be made to the problems discussed in this paper. This different approach brings out the essential linear nature of the problems considered and has the advantage that it may provide a useable, adaptive, time varying, solution if the basic relationships and structures of the series under consideration are changing slowly in time.

When one considers the problems investigated in this paper, he notes that the proposed solution of each of them was based upon the spectral density matrix. Consequently if one is in possession of some observed stretches of series, he will require an estimate of the spectral density matrix in order to implement one of the proposed solutions.

The spectral density matrix may be estimated by first estimating relevant covariances and then Fourier transforming the covariances with the use of convergence factors. This approach is discussed in Jenkins (1963) for example.

Another approach is to be found in Brillinger (1964a), this latter approach is based upon the technique of complex demodulation (Tukey (1961).)

The process of the complex demodulation of a time series proceeds as follows; let there be given an observed segment $\{x(t);\ t = 1, \ldots, T\}$ of a series. Select a frequency ω_0 . Form the series $\{x(t) \cos \omega_0 t; t = 1, \ldots, T\}$ and $\{x(t) \sin \omega_0 t;\ t = 1, \ldots, T\}$. Smooth these two series, forming the series $\{u(t,\omega_0);\ t = 1, \ldots, T'\}$ and $\{u^H(t, \omega_0);\ t = 1, \ldots, T'\}$ where $T' < T$. These series are the complex demodulates at frequency ω_0.

It is pointed out in Brillinger (1964a) that the cross-spectrum $f_{12}(\omega_0)$ of the two series $\{x_1(t), x_2(t)\}$ may be estimated by time averaging the product,

$$\{u_1(t, \omega_0) + i\, u_1^H(t, \omega_0)\}\ \{u_2(t, \omega_0) - i\, u_2^H(t, \omega_0)\} .$$

The advantage of estimating $f_{12}(\omega_0)$ by this technique is that it is very easily implimented to provide a running average in time of $f_{12}(\omega_0)$ This allows one to look for certain classes of non-stationarities.

Let us consider this technique in respect of the desire to factor analyze the spectral matrix. If the various cross-spectra are estimated as above, the estimated cross-spectral matrix is essentially an estimated variance-covariance matrix for the random variables $\{u_j(t, \omega_0) + i\, u_j^H(t, \omega_0)\}$ where j runs across the series involved. In other words then, the problem of principle components for example, may be looked upon as the problem of finding constants a_j such that,

$$\Sigma\, a_j \left\{ u_j(t, \omega_0) + i\, u_j^H(t, \omega_0) \right\},$$

has maximum variance, and the solution to this problem consists of carrying out the same procedure as developed for the case of independent observations. This leads to the solution of this paper.

7. DISTRIBUTIONS OF THE PROPOSED STATISTICS

In the Appendix of this paper it will be shown that if $\| \hat{f}_{ij}(\omega) \|$ is an estimate of the spectral density matrix $\| f_{ij}(\omega) \|$ where $\hat{f}_{ij}(\omega)$ is a weighted average, corresponding to a positive kernel, of the cross-periodogram of $x_i(t)$ and $x_j(t)$, then $\| \hat{f}_{ij}(\omega) \|$ may be written

$$\frac{1}{T} \sum_{j=1}^{T} \xi_j\, \xi_j^{-1}, \tag{7.1}$$

where ξ_j is a complex valued vector Gaussian random variable with mean 0 and variance-covariance matrix,

$$\| \int K(\omega - \omega')\, f_{ij}(\omega')\, d\omega' \| \tag{7.2}$$

K corresponding to the kernel. The ξ_j, $j = 1, \ldots, T$ are dependent;

however it is argued in the Appendix that (7.1) may be approximated by,

$$\frac{1}{u} \sum_{j=1}^{u} n_j \bar{n}_j' , \tag{7.3}$$

where u equals the integral part of ΩT (Ω denoting the bandwidth of K), and the n_j are now independent with mean 0 and variance covariance matrix (7.2).

One may consequently imagine himself in possession of a sample from the complex Gaussian distribution (Goodman (1963),) and may note that (7.3) is an observation from the corresponding complex Wishart distribution.

James (1964) has considered the distributions of various statistics calculated from an observation from a complex Wishart distribution. The distributions he considers that have specific relevance to this paper are the distributions of eigenvalues and canonical correlations. These distributions are given by expressions (95) and (112) of his paper and from what was said above would appear to provide suitable approximations to the distributions of the eigenvalues and canonical correlations of this paper.

8. COMPUTATIONAL PROCEDURES

It has been seen that the solution of a number of problems relating to the joint analysis of a number of stationary time series comes down to the determination of various roots and related vectors of a complex Hermitian matrix. As the standard computational techniques developed for the corresponding problems in multivariate analysis apply only to real symmetric matrices a comment seems in order.

The determination of the roots of a k x k complex Hermitian matrix is equivalent to the determination of the roots of a 2k x 2k real matrix, this results from the isomorphism of the complex numbers and 2 x 2 matrices,

$$\alpha + i\beta \longleftrightarrow \begin{Vmatrix} \alpha & \beta \\ -\beta & \alpha \end{Vmatrix} .$$

One simply replaces each complex entry of the spectral density matrix by the corresponding 2 x 2 real matrix and evaluates the desired roots for the larger matrix. Each root will however be duplicated.

This technique is discussed in Brenner (1961) and Goodman (1963).

REFERENCES

BOWLEY, A. L. (1920). Elements of Statistics. P. S. King and Son, London.

BRENNER, J. L. (1961). "Expanded matrices from matrices with complex elements". SIAM Review, 3, 165-166.

BRILLINGER, D. R. (1964a). "A technique for measuring the spectral density matrix of two signals". Proc. I. E. E. E., 52, 103-104.

(1964b). "A moving average representation for a finite collection of covariance stationary random variables". Submitted to Biometrika.

ECKART, C. and YOUNG, G. (1936). "The approximation of one matrix by another of lower rank". Psychometrika, 1, 211-218.

FREIBERGER, W. and GRENANDER, U. (1959). "Approximate distributions of noise power measurements". Quart. Appl. Math., 17, 271-283.

GOODMAN, N. R. (1963). "Statistical analysis based upon a certain multivariate complex Gaussian distribution (an introduction)". Ann. Math. Statist., 34, 152-177.

GRENANDER, U., POLLAK, H. O. and SLEPIAN, D. (1959). "The distribution theory of quadratic forms in normal variates: a small sample theory with applications to spectral analysis". J. Soc. Indus. Appl. Math., 7, 374-401.

GRENANDER, U. and Szego, G. (1958). Toeplitz Forms and Their Applications. University of California Press, California.

HARMON, H. H. (1960). Modern Factor Analysis. University of Chicago Press, Chicago.

JAMES, A. T. (1964). "Distributions of matrix variates and latent roots derived from normal samples". Ann. Math. Statist., 35, 475-501.

JENKINS, G. M. (1963). "Cross-spectral analysis and the estimation of linear open loop transfer functions". Time Series Analysis, edited by Murray Rosenblatt. Wiley, New York.

PARZEN, E. (1957). "On consistent estimates of the spectrum of a stationary time series". Ann. Math. Statist., 28, 329-348.

RICKER, N. (1940). "The form and nature of seismic waves and the structure of seismograms". Geophysics, 5, 348-366.

ROBINSON, E. A. (1957). "Predictive decomposition of seismic traces". Geophysics, 22, 767.

STONE, R. (1947). "On the interdependence of blocks of transactions". J. R. Statist. Soc. B, 9, 1-32.

TUKEY, J. W. (1961). "Discussion, emphasizing the connection between analysis of variance and spectrum". Technometrics, 3, 191-219.

APPENDIX

Let $\hat{f}_{ij}(\omega)$ be a weighted average of the cross-periodogram of $\{x_i(t)\}$ and $\{x_j(t)\}$ corresponding to the positive kernel $K(\omega)$. $\hat{f}_{ij}(\omega)$ may consequently be written,

$$\frac{1}{T} \sum_{s,t} k(s - t)\, x_i(s)\, x_j(t), \tag{A.1}$$

where $k(s)$ is an even function of s such that,

$$k(s) = \int_{-\pi}^{\pi} e^{is\omega}\, K(\omega)\, d\omega \quad ,$$

where $K(\omega)$ has the approximate form,

$$K(\omega) = 1 \text{ for } |\omega + \omega_0| < \tfrac{1}{4}\Omega \text{ and } |\omega - \omega_0| < \tfrac{1}{4}\Omega$$

$$= 0 \text{ otherwise.}$$

As stated above it is assumed that $k(\omega) \geq 0$. This implies that the matrix $k = ||\ k_{st}|| = ||\ k(s - t)||$ is non-negative definite. Assume that this matrix is m X m. $||k(s - t)||$ has a $2m - 1$ X m matrix square root $b = ||b_{st}|| = ||b(s - t)||$ such that $k = b^T b$. (See Brillinger (1964b)).

The expression (A.1) may now be written,

$$\frac{1}{T}(b\ x_i)^T\ (b\ x_j)\ ,$$

where x_i^T denotes the row vector $||x_i(1), \ldots, x_i(T)||$. In other words (A.1) may be written,

$$\frac{1}{T}(y_i)^T\ (y_j)\ , \tag{A.2}$$

where y_i, y_j are zero mean, multivariate Gaussian random variables such that,

$$E\, y_i(\alpha)\, y_j(\beta) = E \sum_{\gamma} b(\alpha - \gamma)\, x_i(\gamma) \sum_{\rho} b(\beta - \rho)\, x_j(\rho)$$

$$= \int e^{i\omega(\alpha-\beta)}\, K(\omega)\, f_{ij}(\omega)\, d\omega \quad .$$

At this point let us assume that $f_{ij}(\omega)$ is essentially a constant, $f_{ij}(\omega_0)$, on the support of $K(\omega)$. Consequently,

$$E\, y_i(\alpha)\, y_j(\beta) \sim f_{ij}(\omega_0)\cdot\|k(\alpha - \beta)\| \; .$$

Now suppose that q is an orthogonal matrix that diagonalizes k that is $q^T kq = \Lambda$ where Λ is diagonal. Define the $2m - 1 \times m$ matrix p to have the form,

$$\left\| \begin{matrix} q \\ - - - \\ 0 \end{matrix} \right\|$$

Let $y = pz$ define m variates z. The expression (A.2) becomes

$$\frac{1}{T}\{z_i\}^T \{z_j\} \; . \tag{A.3}$$

The z_i are Gaussian, mean 0 and

$$E\, z_i\, z_j^T = p\, E\, y_i\, y_j^T\, p^T$$

$$\sim f_{ij}(\omega_0)\, p\, k\, p^T$$

$$= f_{ij}(\omega_0)\, \Lambda$$

Summing up the calculations to this point it has been seen that the spectral estimate A.1 is approximately of the form,

$$\frac{1}{T}\sqrt{f_{ii}(\omega_0)\, f_{jj}(\omega_0)}\; \Sigma\, \lambda_\alpha\, u_{i\alpha}\, v_{j\alpha}$$

where the $(u_{i\alpha}, v_{j\alpha})$ are independent for different α and where $(u_{i\alpha}, v_{j\alpha})$ have a joint Gaussian distribution with means 0, variances 1 and correlation $R_{ij}(\omega_0)$.

The intention of the remainder of this section is to argue that the λ_α if ordered such that $\lambda_1 \geq \lambda_2 \geq \cdots$ are very nearly 1 for $\alpha \leq \Omega T$ and 0 thereafter. If this can be argued convincingly then it is immediately apparent that A.1 has the approximate form of a multiple of a simple covariance calculated from a sample of ΩT independent variates.

Because $k = ||k(\alpha - \beta)||$ is a Toeplitz matrix, results are known concerning its eigenvalues. In fact a theorem of Grenander and Szegö (1958) states that s $T \longrightarrow \infty$, $\lim \{\frac{\text{number } \lambda_\alpha \leq x}{T}\} = \text{meas } [\lambda | K(\lambda) \leq x]$.

The reader will remember that $K(\omega)$ was selected to have the approximate shape,

$$K(\omega) = 1 \text{ for } |\omega + \omega_0| < \frac{1}{4}\Omega,\ |\omega - \omega_0| < \frac{1}{4}\Omega$$

$$= 0 \text{ otherwise.}$$

Consequently, the number of $\lambda_\alpha \geq 1 - \epsilon$ is approximately ΩT for any ϵ, $0 < \epsilon < 1$, while the rest of the λ_α are $< \epsilon$.

In other words the first ΩT eigenvalues are near 1 and the remainder near 0.

References to the use of Toeplitz approximations in the consideration of the distribution of spectral estimates include Freiberger and Grenander (1959) and Grenander, Pollak and Slepian (1959).

I have learned from Professor E. Parzen that in a preliminary version of Parzen (1957) he obtained the representation (A.2) for the case of a single time series.