The Analysis of Multiple Stationary Time Series

By P. Whittle

University Institute of Statistics, Upsala.

[Received December, 1952]

SUMMARY

After some preparatory work, the least square estimation equations are derived for a purely nondeterministic stationary multiple process (Th. 6). The asymptotic covariances of the parameter estimates are calculated for a normal process (Th. 9) and a test of fit derived (Th. 10). The testing of a sunspot model provides an illustration of the methods developed.

(1) *Introductory*

In a series of recent publications (Whittle, 1951–52*c*) the author has considered the application of the classical least square principle to the analysis of a single stationary time series, in an attempt to obtain a uniform treatment of the estimation and test problems which arise. (By "least square estimation equations" is meant here those equations which would have been obtained on the maximum likelihood principle if the variates had been assumed normally distributed.) The method depends essentially upon the natural complement which least square theory forms to the spectral and autocovariance theory of stationary processes, concerning as they both do only the second moments of the observations. Thus, quantities such as the residual sum of squares may be expressed in terms of the theoretical spectral function and the observed autocovariances or the periodogram.

From the theoretical point of view, this method should lead in most cases to estimates and tests with the restricted optimum properties enjoyed by least square statistics in general, at least asymptotically; from the practical point of view, experience has shown it to be as simple and direct as the subject would seem to allow.

However, it is seldom that one has occasion to analyse a single series. The vast majority of practical problems require the analysis of a multiple series—meteorological and economic data provide perhaps the most obvious examples. In this paper an attempt is therefore made to extend the least square theory to this more general case.

Now, the difficulty of analysing a q-tuple series may be said to increase roughly as q^2 (the number of auto- and cross-correlograms which must be calculated, and the order of the number of parameters to be estimated), while the number of observations increases only as q. These facts are undoubtedly discouraging, and offset only in part by the fact that the stochastic relations of a multiple process should in general be fairly simple in nature. Indeed, intuition would lead us to believe that if account were taken of all the relevant variates of a self-contained system, then the process would be Markovian (see Moyal, 1949, p. 200.) Further, if the variates fall into a well-defined causal chain, there are good reasons for believing that the set of stochastic relations reduces to a so-called recursive system (see Wold, 1952, p. 64.)

Another complication introduced by the multiplicity of the series is that of *nonidentifiability* (Frisch, 1938). That is, the number of unknown parameters exceeds the number of distinct estimation equations available, so that the solution for the parameter estimates is indeterminate. The indeterminacy generally takes the form that the separate equations of the fitted scheme may be arbitrarily added and subtracted, and still be consistent with the estimation equations. (Indeterminacies of another type can arise in the single series case, cf. Wold, 1938, p. 126, but these are not so serious.)

To attain identifiability it is necessary to reduce the number of parameters by requiring them to obey extra relations—relations which may often be quite arbitrary. However, since it is certain that statistics alone cannot provide a solution of the identifiability problem, we shall assume

that identifiability has been achieved, so that the estimation equations always yield determinate solutions.

We note a last difficulty which will become more apparent in the course of the paper: that those parameters referring to essential features of the process and those referring simply to the variances of the residual variates are in general so intermixed as to be inseparable (separation is possible in the single series case). This not only complicates the estimation equations, but also affects the asymptotic properties of the parameter estimates.

The aim of the paper is to provide a practical solution of the problem rather than a theoretical treatment of the many points which arise. It has nevertheless been impossible to avoid rather much mathematics, so much that it has been considered best to give the main results in theorem form. This is done solely in the interests of clarity, however, and not to provide a pretence of rigour where no particular degree of rigour has been attempted.

Notation. The complex conjugate of z is denoted by $\bar{z}$. A matrix $\mathbf{A}$ with elements a_{jk} will sometimes be written (a_{jk}). Its determinant, transpose and trace will be denoted $|\mathbf{A}|$, $\mathbf{A}'$, and $tr\mathbf{A}$ respectively. The complex conjugate of the transpose, $\mathbf{A}'$, will be written $\mathbf{A}^*$.

(2) *Specification of the Process*

Suppose that we consider a q-variate process $\{x_{1t}, x_{2t} \ldots x_{qt}\}$, $(t = \ldots -2, -1, 0, 1, 2 \ldots)$, the variates for time t forming a $q \times 1$ vector $\mathbf{X}_t$. We shall restrict ourselves to a purely nondeterministic process, so that $E(\mathbf{X}_t) = 0$. (The effect of a non-zero mean will be considered later.) We shall now define the theoretical autocovariances and spectral functions together with their empirical equivalents:

$$\Gamma_{jk}(s) = E(x_{j,\,t+s}x_{kt}). \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.1)$$

$$F_{jk}(\omega) = \sum_{s=-\infty}^{\infty} \Gamma_{jk}(s)e^{i\omega s}. \quad . \quad . \quad . \quad . \quad . \quad (2.2)$$

$$\mathbf{F}(\omega) = (F_{jk}(\omega)). \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.3)$$

$$C_{jk}(s) = (1/N) \sum_{t=1}^{N-s} x_{j,\,t+s}x_{kt}. \quad . \quad . \quad . \quad . \quad . \quad (2.4)$$

$$f_{jk}(\omega) = \sum_{s=-N}^{N} C_{jk}(s)e^{i\omega s}. \quad . \quad . \quad . \quad . \quad . \quad (2.5)$$

$$\mathbf{f}(\omega) = (f_{jk}(\omega)). \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.6)$$

$\mathbf{F}(\omega)$ is thus the matrix of theoretical spectral functions $F_{jk}(\omega)$. $C_{jk}(s)$ is the empirical cross-covariance of x_{jt} and x_{kt}, lagged s steps, and is based upon a sample series of N terms, $(\mathbf{X}_1, \mathbf{X}_2 \ldots \mathbf{X}_N)$. Strictly, the sum in (2.4) should be divided by $N - s$ instead of N, but we shall consistently assume N so large that end effects may be neglected. (It will later be apparent that we need only consider the earlier covariances, for which s is relatively small.) The $f_{jk}(\omega)$ are the empirical spectral functions, or periodograms, and $\mathbf{f}(\omega)$ is their matrix.

The first facts to be noted are that the matrices $\mathbf{F}(\omega)$, $\mathbf{f}(\omega)$ are Hermitian and Hermitian positive semi-definite (see Cramér, 1940). That $\mathbf{F}(\omega)$ is Hermitian follows from the fact that $F_{jk}(\omega) = F_{kj}(-\omega) = \overline{F_{kj}(\omega)}$. That it is Hermitian positive definite follows from the fact that $\boldsymbol{\zeta}'\mathbf{F}(\omega)\bar{\boldsymbol{\zeta}}$ is the spectral function of the variate $\boldsymbol{\zeta}'\mathbf{X}_t$, ($\boldsymbol{\zeta}$ an arbitrary vector), and is thus never negative. Similarly for $\mathbf{f}(\omega)$.

Now, Wold (1938) has given a canonical representation of the univariate stationary process, which in the case of a purely nondeterministic variate reduces to a one-sided moving average

$$x_t = b_0\,\varepsilon_t + b_1\,\varepsilon_{t-1} + \ldots \quad . \quad . \quad . \quad . \quad . \quad (2.7)$$

over a series of uncorrelated variates ε_t. (The interpretation of (2.7) is well known: if a series of uncorrelated disturbances $\ldots \varepsilon_{t-1}, \varepsilon_t, \varepsilon_{t+1} \ldots$ is injected into a system, and ε_t has a lingering effect $b_s\varepsilon_t$ at time $t + s$, and the effects of the different disturbances are additive, then the total effect is given by (2.7). The series $b_0, b_1, b_2 \ldots$ is known as the *transient response* of the system.)

Zasuhin (1941) has generalized Wold's representation to the multivariate case, so that corresponding to (2.7) we have

$$x_{jt} = \sum_{m=0}^{\infty} b_{jkm}\varepsilon_{k,\,t-m} \qquad . \quad . \quad . \quad . \quad . \quad . \quad (2.8)$$

($j = 1, 2 \ \ldots \ q$; $t = \ \ldots \ -1, 0, 1 \ \ldots$) where all the ε's are uncorrelated, have zero mean and unit variance. A feature of Zasuhin's representation is that the b_{jkm} coefficients are not uniquely determined by $\mathbf{F}(\omega)$. This is just the nonidentifiability phenomenon mentioned before, however, and will not trouble us, since the constraints to which we have assumed the parameters subject will also serve to uniquely determine the b_{jkm} coefficients.

Let us now define a shift operator U for which

$$Ux_t = x_{t-1} \qquad . \quad . \quad . \quad . \quad . \quad . \quad (2.9)$$

and define a matrix $\mathbf{B}(U)$.

$$\mathbf{B}(U) = [B_{jk}(U)]$$

$$B_{jk}(U) = \sum_{m=0}^{\infty} b_{jkm}U^m, \qquad . \quad . \quad . \quad . \quad . \quad . \quad (2.10)$$

so that the elements of $\mathbf{B}(U)$ may be expanded in positive powers of U. Then (2.8) may be written more neatly

$$\mathbf{X}_t = \mathbf{B}(U)\,\boldsymbol{\epsilon}_t, \qquad . \quad . \quad . \quad . \quad . \quad . \quad (2.11)$$

where $\boldsymbol{\epsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t} \ \ldots \ \varepsilon_{qt})$. We shall term (2.11) the *moving average representation* of the process. If the equation in z, $|\mathbf{B}(z)| = 0$, has no roots on or within $|z| = 1$, then (2.11) may be inverted into an *autoregressive representation* of the process

$$\boldsymbol{\epsilon}_t = [\mathbf{B}(U)]^{-1}\mathbf{X}_t = \mathbf{A}(U)\mathbf{X}_t \qquad . \quad . \quad . \quad . \quad . \quad (2.12)$$

say. The reciprocal of $\mathbf{B}(U)$, $\mathbf{A}(U)$, is to be obtained by first inverting $\mathbf{B}(U)$ as a matrix, and then expanding each element of the inverted matrix in powers of U. Only positive powers of U will occur (due to the condition on the roots of $|\mathbf{B}(z)|$) so that future values, $\mathbf{X}_{t+1}, \mathbf{X}_{t+2} \ \ldots$, do not occur in the representation. The elements of $\mathbf{A}(U)$ will be denoted

$$\mathbf{A}(U) = [A_{jk}(U)]$$

$$A_{jk}(U) = \sum_{m=0}^{\infty} a_{jkm}U^m. \qquad . \quad . \quad . \quad . \quad . \quad (2.13)$$

Given the autoregressive representation (2.12), a necessary and sufficient condition that it be invertable to the moving average representation (2.11) is that $|\mathbf{A}(z)| = 0$ have no roots on or within the unit circle (cf. Mann and Wald, 1943; Moran, 1949). This result corresponds to the usual condition that a simple autoregression generate a stationary process (Wold, 1938, p. 99).

There is a theorem connecting the spectral matrix with the coefficients of the two representations (cf. Bartlett, 1946, p. 92).

Theorem 1

$$\mathbf{F}(\omega) = \mathbf{B}(e^{i\omega})\ \mathbf{B}'\ (e^{-i\omega}) = [\mathbf{A}'(e^{-i\omega})\ \mathbf{A}(e^{i\omega})]^{-1}.$$

Writing $z = e^{i\omega}$, we have

$$F_{jk}(\omega) = E(\sum_s x_{j,\,t+s}\, x_{kt}\, z^s)$$

$$= E(\sum_s \sum_m \sum_n \sum_u \sum_v b_{jmu}\, b_{knv}\, \varepsilon_{m,\,t+s-u}\, \varepsilon_{n,\,t-v}\, z^s$$

$$= \sum_s \sum_m \sum_u b_{jmu}\, b_{km,\,u-s}\, z^s = \sum_m \sum_u \sum_v b_{jmu}\, z^u b_{kmv} z^{-v}$$

$$= \sum_m B_{jm}(z)\, B_{km}(z^{-1}),$$

which is equivalent to the desired result.

Example—Consider the double process

$$x_{1t} + ax_{1,\,t-1} + bx_{1,\,t-2} = c\,\varepsilon_{1t}$$
$$x_{2t} + dx_{1,\,t-1} = e(\varepsilon_{2t} - f\varepsilon_{2,\,t-1})$$

or

$$\begin{bmatrix} 1 + aU + bU^2 & . \\ dU & 1 \end{bmatrix} \mathbf{X}_t = \begin{bmatrix} c & . \\ . & e(1 - fU) \end{bmatrix} \boldsymbol{\epsilon}_t$$

If we write the last equation as $\mathbf{P}\mathbf{X}_t = \mathbf{Q}\boldsymbol{\epsilon}_t$, then

$$\mathbf{A}(U) = \mathbf{Q}^{-1}\mathbf{P} = \begin{bmatrix} \dfrac{1 + aU + bU^2}{c} & . \\ \dfrac{dU}{e(1 - fU)} & \dfrac{1}{e(1 - fU)} \end{bmatrix}$$

and

$$\mathbf{B}(U) = \mathbf{P}^{-1}\,\mathbf{Q} = \frac{1}{1 + aU + bU^2} \begin{bmatrix} c & \\ -cdU & e(1 - fU)\,(1 + aU + bU^2) \end{bmatrix}$$

We have

$$\mathbf{F}(\omega) = \mathbf{B}\mathbf{B}^* = \frac{1}{|\,1 + az + bz^2\,|^{\,2}} \begin{bmatrix} c^2 & -c^2\,dz^{-1} \\ -c^2 dz & c^2d^2 + e\,|\,1 - fz\,|^{\,2} \end{bmatrix}$$

where $z = e^{i\omega}$.

(3) *The Prediction Variance*

Consider the autoregressive representation of the process (2.12). Let us denote the absolute term in the expansion of $\mathbf{A}(U)$ in powers of U by $\mathbf{A}_0$, so that (2.12) may be written

$$\mathbf{A}_0\,\mathbf{X}_t + (\text{past } \mathbf{X}_t \text{ values}) = \boldsymbol{\epsilon}_t,$$

or

$$\mathbf{X}_t = \mathbf{A}_0^{-1}\,\boldsymbol{\epsilon}_t + (\text{past } \mathbf{X}_t \text{ values})$$
$$= \mathbf{B}_0\,\boldsymbol{\epsilon}_t + (\text{past } \mathbf{X}_t \text{ values}), \qquad . \quad . \quad . \quad . \quad . \quad (3.1)$$

where $\mathbf{B}_0$ is correspondingly the absolute term in the expansion of $\mathbf{B}(U)$. Now, the bracketed term in (3.1) provides the best linear prediction of $\mathbf{X}_t$ in terms of $\mathbf{X}_{t-1}, \mathbf{X}_{t-2}$ . . . , so that $\mathbf{B}_0\boldsymbol{\epsilon}_t$ is the deviation of prediction from reality, that part of $\mathbf{X}_t$ which cannot be deduced from past history, consisting as it does of the random variation which has been injected into the process since the last instant of time.

Corresponding to the variance of a single variate is the *total variance* of a vector variate, equal to the determinant of the covariance matrix (Wilks, 1932). In the present case

$$E[(\mathbf{B}_0\,\boldsymbol{\epsilon}_t)\,(\mathbf{B}_0\,\boldsymbol{\epsilon}_t)'] = \mathbf{B}_0\,\mathbf{B}'_0, \qquad . \quad . \quad . \quad . \quad . \quad (3.2)$$

so that the total variance of $\mathbf{B}_0\,\boldsymbol{\epsilon}_t$, a quantity which we may term the *total prediction variance*, is

$$V = |\,\mathbf{B}_0\,\mathbf{B}'_0\,| = |\,\mathbf{B}_0\,|^{\,2} = |\,\mathbf{A}_0\,|^{\,-2}. \qquad . \quad . \quad . \quad . \quad (3.3)$$

V gives a measure of the total random variance entering the process at every step. It may be derived in another way: if the likelihood of the sample $(\mathbf{X}_1, \mathbf{X}_2 \ . \ . \ . \ \mathbf{X}_N)$ is expressed in terms of the likelihood of $(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2 \ . \ . \ . \ \boldsymbol{\epsilon}_N)$, then the Jacobian of the transformation is

$$\left|\frac{\partial(\mathbf{X}_1, \mathbf{X}_2 \ . \ . \ . \ \mathbf{X}_N)}{\partial(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2 \ . \ . \ . \ \boldsymbol{\epsilon}_N)}\right| = |\,\mathbf{B}_0\,|^{\,N}, \qquad . \quad . \quad . \quad . \quad (3.4)$$

and we should expect this quantity to be equal to just $V^{N/2}$.

The total prediction variance may be expressed in terms of the spectral matrix of the process:

Theorem 2

$$V = \exp\left[\frac{1}{2\pi}\int_0^{2\pi} \log \mid \mathbf{F}(\omega) \mid d\omega\right].$$

In the case of a univariate process this is Kolmogoroff's result (1941). To prove it, we make use of the fact that, since $\mid \mathbf{B}_0 \mid$ is the absolute term in the expansion of $\mid \mathbf{B}(z) \mid$, ($\mid z \mid \leqslant 1$), then $\log \mid \mathbf{B}_0 \mid$ is the absolute term in the expansion of $\log \mid \mathbf{B}(z) \mid$. We have thus

$$\frac{1}{2\pi}\int_0^{2\pi} \log \mid \mathbf{F}(\omega) \mid d\omega = \frac{1}{2\pi}\int_0^{2\pi} \log \mid \mathbf{B}(z) \mid d\omega + \frac{1}{2\pi}\int_0^{2\pi} \log \mid \mathbf{B}(z^{-1}) \mid d\omega$$

$$= \log \mid \mathbf{B}_0 \mid + \log \mid \mathbf{B}_0 \mid = \log V,$$

which is the required result.

As is the case in so many problems of this type, the multivariate process may be reduced in a standard fashion to q mutually uncorrelated univariate processes (cf. Wiener, 1949, p. 44). For, suppose that the matrix $\mathbf{F}(\omega)$ has eigenvalues $\lambda_1(\omega)$, $\lambda_2(\omega)$ . . . $\lambda_q(\omega)$, with corresponding normalized eigenvectors $\mathbf{C}_1(z)$, $\mathbf{C}_2(z)$ . . . $\mathbf{C}_q(z)$, so that

$$\mathbf{F}(\omega) = \sum_j \lambda_j \mathbf{C}_j \mathbf{C}_j^* \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.5)$$

$$\mathbf{C}_j^* \mathbf{C}_k \equiv \delta_{jk} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.6)$$

Theorem 3.—*The Unitary Transformation*

$$y_{jt} = \mathbf{C}'_j(U^{-1})\mathbf{X}_t$$

produces q *mutually uncorrelated processes* $\{y_{jt}\}$ *with respective spectral functions* $\lambda_j(\omega)$, (j = 1, 2 . . . q).

For, if $\mathbf{Y}'_t = (y_{1t}, y_{2t} \ldots y_{qt})$ and $\mathbf{C}^*(z) = (\mathbf{C}_1, \mathbf{C}_2 \ldots \mathbf{C}_q)$, then

$$\mathbf{Y}_t = \mathbf{C}(U)\mathbf{X}_t \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.7)$$

and, by the same argument as that used to establish Th. 1 from (2.11), we see that the matrix of spectral functions for the $\mathbf{Y}_t$ process is

$$\mathbf{F}^{(y)}(\omega) = \mathbf{C}(z)\,\mathbf{F}(\omega)\,\mathbf{C}^*(z)$$

$$= [\mathbf{C}'_j(z^{-1})\,\mathbf{F}(\omega)\,\mathbf{C}_k(z)] = (\delta_{jk}\lambda_j(\omega)) \quad . \quad . \quad . \quad . \quad (3.8)$$

by (3.5), (3.6). This is equivalent to the result of the theorem.

The q $\{y_{jt}\}$ processes, which may be termed the *canonical components* of the multivariate process, will have prediction variances

$$v_j = \exp\left[\frac{1}{2\pi}\int_0^{2\pi} \log \lambda_j(\omega)\, d\omega\right] \quad . \quad . \quad . \quad . \quad . \quad (3.9)$$

and these may be termed the *canonical prediction variances* of the process. Obviously, $\prod_1^q v_j = V$.

While the canonical transformation of Th. 3 has a meaning, and can be useful in theoretical work, it can very seldom be realised in practice, and we shall not find great use for it. There are, however, certain exceptional cases where the eigenvalues and eigenvectors of $\mathbf{F}(\omega)$ may be readily calculated, and where the eigenvectors do not contain any of the process parameters. These cases may be treated very elegantly, for then the q y series may be considered completely separately one from the other.

Example.—Consider the symmetric double autoregression

$$\alpha(U)x_{1t} + \beta(U)x_{2t} = \varepsilon_{1t}$$

$$\beta(U)x_{1t} + \alpha(U)x_{2t} = \varepsilon_{2t},$$

for which

$$\mathbf{F}(\omega)^{-1} = \begin{bmatrix} \alpha(z)\,\alpha(z^{-1}) + \beta(z)\,\beta(z^{-1}) & \alpha(z)\,\beta(z^{-1}) + \alpha(z^{-1})\,\beta(z) \\ \alpha(z)\,\beta(z^{-1}) + \alpha(z^{-1})\,\beta(z) & \alpha(z)\,\alpha(z^{-1}) + \beta(z)\,\beta(z^{-1}) \end{bmatrix}.$$

We find the eigenvalues of $\mathbf{F}(\omega)$ to be $\lambda(\omega) = |\,\alpha(z) \pm \beta(z)\,|^2$ with corresponding eigenvectors $\sqrt{\frac{1}{2}}(1, \pm 1)$. The canonical components of the process are thus $\{(x_{1t} \pm x_{2t})/\sqrt{2}\}$. If $\alpha(z) = \alpha_0 + \alpha_1 z + \ldots$ and $\beta(z) = \beta_0 + \beta_1 z + \ldots$ then the canonical prediction variances are $(\alpha_0 \pm \beta_0)^{-2}$.

(4) *Moments of Sample Functions*

In this section we shall consider the distribution of linear functions (real coefficients) of the auto- and cross-covariances, on the assumption that the residual variates ε are normally distributed. First, we shall show that any linear function ξ of these covariances may be written)

$$\xi = N\mathscr{A}\, tr[\mathbf{Q}(z)\,\mathbf{f}(\omega)] = (N/2\pi)\int_0^{2\pi} tr[\mathbf{Q}(z)\,\mathbf{f}(\omega)]\,d\omega, \quad . \quad . \quad . \quad (4.1)$$

where $\mathbf{Q}(z)$ is a $q \times q$ Hermitian matrix whose elements are functions of $z = e^{i\omega}$, and $\mathscr{A}$ denotes the operation "absolute term in z in". For

$$\begin{aligned} \xi &= N\,\mathscr{A}\, tr[\mathbf{Q}(z)\,\mathbf{f}(\omega)] \\ &= N\,\mathscr{A} \sum_j \sum_k Q_{kj}(z)\,f_{jk}(\omega) = N \sum_j \sum_k \sum_s q_{kj,\,-s}C_{jk}(s). \quad . \quad . \quad . \quad (4.2) \end{aligned}$$

Now, to require that $\mathbf{Q}$ be Hermitian is to require that $q_{kj,\,-s} = q_{jks}$ for all appropriate j, k and s, since the q's are assumed real. But $q_{kj,\,-s}$ is the coefficient of $C_{jk}(s)$, and q_{jks} is the coefficient of $C_{kj}(-s) = C_{jk}(s)$ also. The restriction is thus only an apparent one, and (4.2) represents the general linear function of the covariances. We shall now prove our main result of this section:

Theorem 4.—*If* $Q(z)$ *possesses a Laurent expansion in* z *on* $|\,z\,| = 1$, *then the* r^{th} *cumulant of* ξ *is given asymptotically by*

$$\begin{aligned} k_r(\xi) &= 2^{r-1}(r-1)!N\,\mathscr{A}\, tr[\mathbf{Q}(z)\,\mathbf{F}(\omega)]^r \\ &= \frac{2^{r-1}(r-1)!N}{2\pi}\int_0^{2\pi} tr[\mathbf{Q}(z)\,\mathbf{F}(\omega)]^r\,d\omega. \end{aligned}$$

We shall first prove the result for the case $\mathbf{F} = \mathbf{I}$, when the x's are all independent standard variates.

We shall introduce the auxiliary $N \times N$ circulant matrix

$$\mathbf{W} = \begin{bmatrix} \cdot & 1 & \cdot & \ldots & \cdot \\ \cdot & \cdot & 1 & \ldots & \cdot \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \cdot & \cdot & \cdot & \ldots & 1 \\ 1 & \cdot & \cdot & \ldots & \cdot \end{bmatrix}. \quad . \quad . \quad . \quad . \quad . \quad (4.3)$$

Denoting the vector $(x_{j1}, x_{j2} \ \ldots \ x_{jN})$ by $\mathbf{x}'_j$, then

$$NC_{jk}(s) \doteq \mathbf{x}'_j \mathbf{W}^s \mathbf{x}_k \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.4)$$

if s is small compared with N, so that we can write

$$\begin{aligned} \xi &= N \mathscr{A} \sum_j \sum_k Q_{kj}(z) f_{jk}(\omega) \\ &\doteq \sum_j \sum_k x'_j \, Q_{kj}(W) x_k. \quad . \quad . \quad . \quad . \quad . \quad (4.5) \end{aligned}$$

By ordinary normal theory, the characteristic function of ξ is then

$$\Phi_\xi(\theta) \doteq \left| \mathbf{I} - 2i\theta \, \mathbf{Q}(\mathbf{W}) \right|^{-\frac{1}{2}} \quad . \quad . \quad . \quad . \quad . \quad (4.6)$$

and

$$k_r(\xi) \doteq 2^{r-1}(r-1)! \, tr \, \mathbf{Q}(\mathbf{W})^n \quad . \quad . \quad . \quad . \quad . \quad (4.7)$$

where $\mathbf{Q}(\mathbf{W})$ is the matrix of matrices

$$\mathbf{Q}(\mathbf{W}) = [Q_{jk}(\mathbf{W})]. \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.8)$$

Since $\mathbf{Q}(z)$ is Hermitian

$$Q_{jk}(\mathbf{W}) = Q_{kj}(\mathbf{W}^{-1}) = Q_{kj}(\mathbf{W})' \quad . \quad . \quad . \quad . \quad (4.9)$$

so that $\mathbf{Q}(\mathbf{W})$ is symmetric, as it must be for (4.6) and (4.7) to hold. (This is the reason why $Q(z)$ was required to be Hermitian). Now, if $P_1(z)$, $P_2(z)$ . . . are functions with Laurent expansions in z on $|z| = 1$, then it may be shown (Whittle, 1951, pp. 35–49) that

$$tr[P_1(\mathbf{W}) \, P_2(\mathbf{W}) \quad \ldots \quad] \doteq (N/2\pi) \int_0^{2\pi} [P_1(z) \, P_2(z) \quad \ldots \quad] d\omega. \quad . \quad . \quad (4.10)$$

Setting this result in (4.7) we have

$$\begin{aligned} k_r(\xi) &= 2^{r-1}(r-1)! \, tr \sum_{\nu_1} \sum_{\nu_2} \quad \ldots \quad \sum_{\nu_r} Q_{\nu_1 \nu_2}(\mathbf{W}) \, Q_{\nu_2 \nu_3}(\mathbf{W}) \quad \ldots \quad Q_{\nu_r \nu_1}(\mathbf{W}) \\ &= \frac{2^{r-1}(r-1)!N}{2\pi} \int_0^{2\pi} \sum_{\nu_1} \sum_{\nu_2} \quad \ldots \quad \sum_{\nu_r} Q_{\nu_1 \nu_2}(z) \, Q_{\nu_2 \nu_3}(z) \quad \ldots \quad Q_{\nu_r \nu_1}(z) d\omega \\ &= \frac{2^{r-1}(r-1)!N}{2\pi} \int_0^{2\pi} tr \, \mathbf{Q}(z)^r \, d\omega \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.11) \end{aligned}$$

which is the result of the theorem when $\mathbf{F} = \mathbf{I}$. This result may be readily extended to the general case. We have

$$\begin{aligned} Nf_{jk}(\omega) &= N \sum_s C_{jk}(s) z^s = \sum_s \sum_t x_{j, t+s} x_{kt} z^s \\ &= \sum_s \sum_t \sum_m \sum_n \sum_u \sum_v b_{jmu} \, b_{knv} \, \varepsilon_{m, t+s-u} \, \varepsilon_{n, t-v} \, z^s \end{aligned}$$

Setting now $t - v = t'$, $t + s - u = t' + s'$ (and immediately dropping the primes!), the expression becomes

$$\begin{aligned} Nf_{jk}(\omega) &\doteq \sum_s \sum_t \sum_m \sum_n \sum_u \sum_v b_{jmu} \, z^u \, b_{knv} \, z^{-v} \, \varepsilon_{m, t+s} \, \varepsilon_{nt} \, z^s \\ &= N \sum_m \sum_n \sum_u \sum_v b_{jmu} \, z^u \, b_{knv} \, z^{-v} f_{mn}^{(\varepsilon)}(\omega) \\ &= N \sum_m \sum_n B_{jm}(z) \, B_{kn}(z^{-1}) f_{mn}^{(\varepsilon)}(\omega). \quad . \quad . \quad . \quad . \quad (4.12) \end{aligned}$$

Here $f_{mn}^{(\varepsilon)}(\omega)$ is the cross-periodogram of the two series $(\varepsilon_{m1}, \varepsilon_{m2} \ \ldots \ \varepsilon_{mN}), (\varepsilon_{n1}, \varepsilon_{n2} \ \ldots \ \varepsilon_{nN})$. Equation (4.12) may be written

$$\mathbf{f}(\omega) \doteq \mathbf{B}(z) \, \mathbf{f}^{(\varepsilon)}(\omega) \, \mathbf{B}(z^{-1})' = \mathbf{B} \mathbf{f}^{(\varepsilon)} \mathbf{B}^*. \quad . \quad . \quad . \quad . \quad (4.13)$$

Thus

$$\xi = N\mathscr{A}\ tr(\mathbf{Qf}) \simeq N\mathscr{A}\ tr(\mathbf{QBf}^{(\varepsilon)}\mathbf{B}^*) = N\mathscr{A}\ tr(\mathbf{B}^*\mathbf{QBf}^{(\varepsilon)}), \quad . \quad . \quad (4.14)$$

so that by (4.11)

$$k_r(\xi) \simeq 2^{r-1}(r-1)!N\mathscr{A}\ tr(\mathbf{B}^*\mathbf{QB})^r = 2^{r-1}(r-1)!N\mathscr{A}\ tr(\mathbf{QBB}^*)^r$$
$$= 2^{r-1}(r-1)!N\mathscr{A}\ tr(\mathbf{QF})^r$$

which is the required result.

(5) *The Estimation Equations*

We are now prepared to turn our attention to the sampling side of the question. The least square method of estimation consists in that one expresses the sum of squares of the residual variates (which will in this case not be the ε's, but a constant times the ε's) in terms of the observations $\mathbf{X}_t$, and then minimizes this expression with respect to the unknown parameter values) subject to the condition that the Jacobian $|\mathbf{J}|$ of the transformation from the residuals to the $\mathbf{X}$'s be constant. This last condition, which fixes the relative scales of the observed and residual variates, is necessary if the minimization is to be non-trivial. It is not mentioned in classical theory, because for uncorrelated variates $\mathbf{J} = \mathbf{I}$, and the condition is automatically fulfilled, but as soon as correlation is introduced explicit regard must be taken to it.

Now, the estimation equations obtained upon this principle are equivalent to those obtained upon the maximum likelihood principle if the variates are assumed normally distributed. We shall use this fact to simplify our presentation; that is, we shall derive our estimation equations by maximizing a Gaussian likelihood. This is a purely formal device—we do not necessarily assume a normal distribution.

We shall now deduce the Gaussian likelihood of the $\mathbf{X}$'s from that of the ε's. In Th. 2 we have already expressed the Jacobian $\left|\partial(X)/\partial(\varepsilon)\right|$ in terms of the process parameters. All that remains then is to express $\sum_{j=1}^{q}\sum_{t=1}^{N}\varepsilon_{jt}^2$ in terms of these parameters and the observations X_t. (That we can consider the ε's for the same time range as the $\mathbf{X}$'s, i.e., for $t = 1, 2 \ldots N$, is a result of the facts that the dependence of $\mathbf{X}_t$ upon $\boldsymbol{\epsilon}_{t-s}$ tends to zero as $s \to \infty$, and that we are neglecting end effects.)

Theorum 5.—*If* $\mathbf{F}(\omega)^{-1}$ *is analytic in* $e^{i\omega}$ *on* $(0, 2\pi)$ *then*

$$\sum_{j=1}^{q}\sum_{t=1}^{N}\varepsilon_{jt}^2 \simeq N\mathscr{A}\ tr(\mathbf{fF}^{-1}) = (N/2\pi)\int_0^{2\pi} tr[\mathbf{f}(\omega)\,\mathbf{F}(\omega)^{-1}]\,d\omega.$$

The proof is direct; we have

$$\begin{aligned}
\sum_{j=1}^{q}\sum_{t=1}^{N}\varepsilon_{jt}^2 &= \sum_{j}\sum_{t}\sum_{m}\sum_{n}\sum_{u}\sum_{v} a_{jmu}\,a_{jnv}\,x_{m,\,t-u}\,x_{n,\,t-v} \\
&\simeq N\sum_{j}\sum_{m}\sum_{n}\sum_{u}\sum_{v} a_{jmu}\,a_{jnv}\,C_{mn}(v-u) \\
&= N\mathscr{A}\sum_{s}\sum_{j}\sum_{m}\sum_{n}\sum_{u}\sum_{v} a_{jmu}\,z^{u}\,a_{jnv}\,z^{-v}\,C_{mn}(s)z^{s} \\
&= N\mathscr{A}\sum_{j}\sum_{m}\sum_{n} A_{jm}(z)\,A_{jn}(z^{-1})\,f_{mn}(\omega) \\
&= N\mathscr{A}\ tr(\mathbf{fA}^*\mathbf{A}) = N\mathscr{A}\ tr(\mathbf{fF}^{-1}),
\end{aligned}$$

which is the required result. If $\mathbf{F}^{-1}$ has singularities on $(0, 2\pi)$, then the autoregressive representation of the series does not exist (cf. §2), so the proof breaks down.

The Gaussian likelihood is thus given, apart from end-effects, by

$$p(x) = (2\pi)^{-Nq/2}\left|\frac{\partial(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2 \ldots \boldsymbol{\epsilon}_N)}{\partial(\mathbf{X}_1\,\mathbf{X}_2, \ldots \mathbf{X}_N)}\right| \exp\left[-\frac{1}{2}\sum_{j=1}^{q}\sum_{t=1}^{N}\boldsymbol{\epsilon}_{jt}^2\right]$$
$$= (2\pi)^{-Nq/2}\,V^{-N/2}\exp\left[-\frac{N}{4\pi}\int_0^{2\pi} tr(\mathbf{fF}^{-1})d\omega\right] \quad . \quad . \quad (5.1)$$

where V is given by Th. 2.

Theorem 6.—The Least Square Parameter Estimates are Obtained by Minimizing

$$L = (N/2\pi) \int_0^{2\pi} [\log | \mathbf{F}(\omega) | + tr[\mathbf{f}(\omega)\, \mathbf{F}(\omega)^{-1}]]\, d\omega.$$

This is immediately seen from the fact that $L = -2 \log p$.

Th. 6 yields the estimation equations in the form of various weighted integrals of the periodograms, not a very applicable form. If, however

$$\mathbf{F}(\omega)^{-1} = (\sum_s c_{jks}\, z^s), \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.2)$$

then

$$L = (N/2\pi) \int_0^{2\pi} \log | \mathbf{F}(\omega) |\, d\omega \quad + N \sum_{j=1}^{q} \sum_{k=1}^{q} \sum_{s=-\infty}^{\infty} c_{jks}\, C_{jk}(s), \quad . \quad (5.3)$$

so that the same values enter the estimation equations as linear functions of the covariances (hence the importance of §4). The sum in (5.3) is strictly infinite with respect to s, but the c's converge to zero as $s \to \pm \infty$ (ultimately on account of the nondeterminacy of the process), so that the weight falls on the earlier covariances in (5.3), and the infinite sum can be approximated by a finite one. The practical estimation equations are obtained by minimizing L as given by equation (5.3).

Example.—Consider the example of §2. We have

$$A_0 = \begin{bmatrix} 1/c & . \\ . & 1/e \end{bmatrix},$$

so that $V = (ce)^2$. Further

$$\mathbf{F}(\omega)^{-1} = \begin{bmatrix} \dfrac{|\, 1 + az + bz^2 \,|^2}{c^2} + \dfrac{d^2}{e^2\, |\, 1 - fz \,|^2} & \dfrac{dz^{-1}}{e^2\, |\, 1 - fz \,|^2} \\ \dfrac{dz}{e^2\, |\, 1 - fz \,|^2} & \dfrac{1}{e^2\, |\, 1 - fz \,|^2} \end{bmatrix}$$

so that by (5.3)

$$L/N = 2 \log (ce) + c^{-2}[(1 + a^2 + b^2)\, C_{11}(0) + 2(a + ab)\, C_{11}(1) + 2bC_{11}(2)]$$
$$+ \frac{1}{e^2(1 - f^2)} \sum_s f^{|s|} [d^2\, C_{11}(s) + dC_{12}(s - 1) + dC_{21}(s + 1) + C_{22}(s)].$$

Upon minimizing with respect to a, b and c we obtain the usual estimation equations for a simple autoregressive scheme, as might have been expected, since x_{1t} has no dependence upon x_{2t}. The estimates of d, e and f are more complicated, but iterative methods give the solution quite rapidly.

The next logical step would now be to specify the parameters somewhat more exactly, but at this stage a new difficulty arises. In the single series case the author defined a *normalized spectral function* $G(\omega) = F(\omega)/V$ (Whittle, 1951, p. 80), the aim of this procedure being to separate V from the parameters of $G(\omega)$. For in this case it is the G parameters which are the essential parameters of the process, defining as they do the linear operation to which the process is equivalent, while V is really only a scale factor. Further, the least square estimates of the G parameters have second moments which are asymptotically independent of the form of the distribution function of the residual variate (at least, if the residuals are statistically independent) but this is not true of V. In fact, all that can be said of V is that its least square estimate, $\hat{V}$, is asymptotically uncorrelated with the others. However, V provides the least square criterion of the fit of a given hypothesis (the maximized Gaussian likelihood $\hat{p} = \text{const.}\ \hat{V}^{-N/2}$).

Now, it is by no means obvious in what manner this step should be generalized to the present

case. One possibility would be to define $\mathbf{G}(\omega) = V^{-1/q}\mathbf{F}(\omega)$, so that the estimation equations of Th. 6 become

$$\frac{1}{2\pi q}\int_0^{2\pi} tr[\mathbf{f}\hat{\mathbf{G}}^{-1}]d\omega = \text{minimum} = \hat{V}^{1/q}, \quad . \quad . \quad . \quad . \quad (5.4)$$

$$\int_0^{2\pi} \log|\hat{\mathbf{G}}|\, d\omega = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.5)$$

where the circumflex denotes "least square estimate". Equation (5.5) is a consequence of the normalization of $\mathbf{G}$, and is an identity in the parameter values.

However, the choice of V as a parameter may not necessarily be a natural one, as examination of a few special cases will show. Another possible choice would be to take the canonical prediction variances as q of the parameters of the process, but this is generally even less satisfactory from the computational point of view. The conclusion seems to be, then, that the choice of parameters is best governed by the particular case, and that general rules cannot be laid down.

Inserting equation (5.4) in

$$\hat{p}(x) = (2\pi)^{-Nq/2}\,\hat{V}^{-N/2}\exp\left[-\frac{N}{4\pi\hat{V}^{1/q}}\int_0^{2\pi} tr[\mathbf{f}\hat{\mathbf{G}}^{-1}]d\omega\right] \quad . \quad . \quad . \quad (5.6)$$

we obtain

Theorem 7.—*The maximized value of the Gaussian likelihood may be expressed*

$$\hat{p}(x) = (2\pi e)^{-Nq/2}\,\hat{V}^{-N/2}.$$

Thus the only sample function entering $\hat{p}$ is $\hat{V}$. This has its importance in test theory.

We have hitherto assumed that all variates have zero mean, but this is obviously unrealistic. Substituting $x_{jt} - \mu_j$ for x_{jt} in the expression L of Th. 6 ($j = 1, 2 \ldots q$), and minimizing w.r.t. μ_j, we find that the least square estimate of the mean of x_{jt} is asymptotically

$$\hat{\mu}_j = \sum_{t=1}^{N} x_{jt}/N, \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.7)$$

and is thus independent of $\mathbf{F}(\omega)$.

Lastly, it may be of interest to see what value of $\mathbf{F}(\omega)$ maximizes $p(x)$ absolutely. That is, we no longer require that $\mathbf{F}$ be undetermined only to within a finite number of parameters, but let it vary freely. This is of course a problem in the calculus of variations. Differentiating L of Th. 6 with respect to the *function* $F_{jk}(\omega)$ we obtain

$$\{F^{(jk)}/|F|\} - tr(\mathbf{f}\mathbf{F}^{-1}(\delta_j\,\delta_k)\,\mathbf{F}^{-1}) = 0 \quad . \quad . \quad . \quad . \quad (5.8)$$

where $F^{(jk)}$ is the co-factor of $F_{jk}(\omega)$ in $\mathbf{F}(\omega)$. Now, the first term in (5.8) is the kj^{th} element of the matrix $\mathbf{F}^{-1}$, and the second may be directly verified as being the jk^{th} element of the matrix $(\mathbf{f}\mathbf{F}^{-1})'\,(\mathbf{F}^{-1})'$. Combining the q^2 relations (5.8) we have then

$$\mathbf{F}^{-1} = \mathbf{F}^{-1}\mathbf{f}\mathbf{F}^{-1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.9)$$

with solution $\mathbf{F}(\omega) = \mathbf{f}(\omega)$, as could well have been expected. Setting $\mathbf{F} = \mathbf{f}$ in the expression for V, we see that the least possible value of the estimated total prediction variance which can be obtained by fitting a stationary nondeterministic model is

$$V_{min} = \exp\left[\frac{1}{2\pi}\int_0^{2\pi} \log|\mathbf{f}(\omega)|\, d\omega\right] \quad . \quad . \quad . \quad . \quad (5.10)$$

(6) *Second Moments of the Parameter Estimates*

A careful treatment of the properties of the least square estimates (i.e., the establishment of conditions for their consistency, limited optimality, etc.) would require a paper to itself. For the moment, therefore, we shall content ourselves with assuming the consistency of the estimates. It may be noted, however, that a very complete treatment already exists (Mann and Wald, 1943) for one particular case: that of the estimation of the coefficients of a finite autoregression, with results also for the non-normal case. This treatment does not lend itself to immediate generalization, however; not so much on account of the restriction of autoregressivity, as of that implied by the particular choice of parameters to be estimated. We must in general consider the case where estimation of individual coefficients is unnecessary, the whole body of coefficients being determined by a relatively small number of parameters.

If the parameters of the process are $(\theta_1, \theta_2 \ \ldots \ \theta_p)$, let us denote the $p \times 1$ vector of derivatives of L with respect to these parameters by $\mathbf{L}_1$ (the derivatives being evaluated at the true parameter values), and the $p \times p$ matrix of corresponding second derivatives by $\mathbf{L}_{11}$.

Theorem 8.—*If the sample variates are normally distributed, then*

$$E(\mathbf{L}_1) = 0, \ E(\mathbf{L}_1 \mathbf{L}'_1) = 2\mathbf{M}, \text{ and } E(\mathbf{L}_{11}) = \mathbf{M},$$

where $\mathbf{M}$ *is the matrix*

$$\mathbf{M} = \left[\frac{N}{2\pi}\int_0^{2\pi} tr\left[\frac{\partial \mathbf{F}}{\partial \theta_j}\mathbf{F}^{-1}\frac{\partial \mathbf{F}}{\partial \theta_k}\mathbf{F}^{-1}\right]d\omega\right].$$

We have

$$\frac{\partial L}{\partial \theta_j} = \frac{N}{2\pi}\int_0^{2\pi} tr\left[\frac{\partial \mathbf{F}}{\partial \theta_j}\mathbf{F}^{-1} - \mathbf{f}\mathbf{F}^{-1}\frac{\partial \mathbf{F}}{\partial \theta_j}\mathbf{F}^{-1}\right]d\omega \quad . \quad . \quad . \quad . \quad (6.1)$$

$$\frac{\partial^2 L}{\partial \theta_j \partial \theta_k} = \frac{N}{2\pi}\int_0^{2\pi} tr\left[\frac{\partial^2 \mathbf{F}}{\partial \theta_j \partial \theta_k}\mathbf{F}^{-1} - \frac{\partial \mathbf{F}}{\partial \theta_j}\mathbf{F}^{-1}\frac{\partial \mathbf{F}}{\partial \theta_k}\mathbf{F}^{-1} + \mathbf{f}\mathbf{F}^{-1}\frac{\partial \mathbf{F}}{\partial \theta_j}\mathbf{F}^{-1}\frac{\partial \mathbf{F}}{\partial \theta_k}\mathbf{F}^{-1} + \mathbf{f}\mathbf{F}^{-1}\frac{\partial \mathbf{F}}{\partial \theta_k}\mathbf{F}^{-1}\frac{\partial \mathbf{F}}{\partial \theta_j}\mathbf{F}^{-1}\right]d\omega. \quad . \quad (6.2)$$

Regarded as linear functions of the covariances, these quantities are already in the form (4.1), and their expectations and covariances may be obtained by a direct application of Th. 4. The reader will readily verify the results of the theorem. (Note that $\mathrm{cov}(\xi_1, \xi_2)$ is the coefficient of 2λ in $k_2(\xi_1 + \lambda\xi_2)$.)

It may be remarked that Th. 8 only needs to be supplemented with a few regularity conditions, and we have sufficient to prove that the root $(\hat{\theta}_1, \hat{\theta}_2 \ \ldots \ \hat{\theta}_p)$ of $\hat{\mathbf{L}}_1 = 0$ falls with asymptotic probability 1 in any arbitrarily small region enclosing $(\theta_1, \theta_2 \ \ldots \ \theta_p)$, i.e., the least square estimates are consistent at least for a normal process.

Theorem 9.—*For a normal process the least square estimates* $\hat{\theta}_1, \hat{\theta}_2 \ \ldots \ \hat{\theta}_p$ *have asymptotic covariance matrix*

$$2\mathbf{M}^{-1} = \left[\frac{N}{4\pi}\int_0^{2\pi} tr\left[\frac{\partial \mathbf{F}}{\partial \theta_j}\mathbf{F}^{-1}\frac{\partial \mathbf{F}}{\partial \theta_k}\mathbf{F}^{-1}\right]d\omega\right]^{-1} \quad . \quad . \quad . \quad . \quad (6.3)$$

Let us denote the minimized value of L by $\hat{L}$, and the corresponding value of $\mathbf{L}_1$ (i.e., $\mathbf{L}_1$ with the θ's replaced by $\hat{\theta}$'s) by $\hat{\mathbf{L}}_1$. Further, let the vectors of parameters and parameter estimates be denoted by $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$ respectively. Then

$$\hat{L} = L + \mathbf{L}_1'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + \tfrac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbf{L}_{11}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + \ \ldots \quad . \quad . \quad (6.4)$$

$$\hat{\mathbf{L}}_1 = 0 = \mathbf{L}_1 + \mathbf{L}_{11}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + \ \ldots \ . \quad . \quad . \quad . \quad . \quad (6.5)$$

Since the estimates are consistent, $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ will with asymptotic certainty be of small order in N, usually $0(N^{-\frac{1}{2}})$, and we may as an approximation truncate (6.4) and (6.5) at the points indicated. From (6.5) we have then that

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \doteq -\mathbf{L}_{11}^{-1}\mathbf{L}_1 \doteq (E\mathbf{L}_{11})^{-1}\mathbf{L}_1, \quad . \quad . \quad . \quad . \quad (6.6)$$

since $\mathbf{L}_{11}$ has a coefficient of variation of small order in N. Thus the covariance matrix of $\hat{\boldsymbol{\theta}}_1$, $\hat{\boldsymbol{\theta}}_2 \ \ldots \ \hat{\boldsymbol{\theta}}_p$ is

$$\begin{aligned} E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'] &\doteq (E\mathbf{L}_{11})^{-1} E(\mathbf{L}_1\mathbf{L}_1')(E\mathbf{L}_{11})^{-1} \\ &\doteq \mathbf{M}^{-1}(2\mathbf{M})\mathbf{M}^{-1} = 2\mathbf{M}^{-1}, \quad . \quad . \quad . \quad . \quad . \quad (6.7) \end{aligned}$$

which is the result of the theorem.

It is interesting to note that $\mathbf{M}$ is invariant if $\mathbf{F}$ is replaced by $\mathbf{F}^{-1}$. That is, if for a given process the roles of $\mathbf{X}_t$ and $\boldsymbol{\epsilon}_t$ are interchanged, then the parameter estimates of the new process thus formed will have the same covariance matrix as before, despite the fact that the parameters have radically changed in interpretation.

The form of the covariance matrix may be further specified for special choices of the parameters. Thus, if we take V as one of the parameters, we find that $\hat{V}$ has asymptotic variance $2qV^2/N$, and that it is uncorrelated with the remaining parameters, which have a covariance matrix identical with that of Th. 8, except that $\mathbf{F}$ is replaced by $\mathbf{G}$. If we choose the canonical prediction variances v_j as q of the parameters, we find that the estimates $\hat{v}_j$ have variances $2v_j^2/N$, and that they are correlated neither with one another nor with remaining parameter estimates. The estimates of the means, $\hat{\mu}_j$, are uncorrelated with remaining estimates, and have covariance matrix $\mathbf{F}(0)/N$.

The question of whether the above expressions still hold if the residuals are other than normally distributed is as yet an unsolved one. We can give an example of the type of result. Suppose that apart from the parameters which enter into the expression for V we have only a single parameter, θ, so that

$$\hat{\theta} - \theta \doteq -\frac{\partial L}{\partial\theta}\Big/\frac{\partial^2 L}{\partial\theta^2} = \frac{\boldsymbol{\varepsilon}'\mathbf{P}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'\mathbf{Q}\boldsymbol{\varepsilon}} \quad . \quad . \quad . \quad . \quad . \quad (6.8)$$

say, where we have written $\hat{\theta} - \theta$ as a quotient of two quadratic forms in the ε_{jt}'s, $(j = 1, 2 \ \ldots \ q;\ t = 1, 2 \ \ldots \ N)$. If the estimate is to be unbiased we must have $tr\mathbf{P} \equiv \sum_i p_{ii} = 0$, which it is. We readily find that the condition that var $(\hat{\theta})$ be asymptotically independent of $k_4(\varepsilon)$ is that $\sum_i p_{ii}^2 = 0$, or that all p_{ii} be zero. This is found to be equivalent to the condition that all the diagonal elements of

$$\int_0^{2\pi} [\mathbf{B}(z)]^{-1}\frac{\partial\mathbf{F}}{\partial\theta}[\mathbf{B}(z^{-1})']^{-1}d\omega \quad . \quad . \quad . \quad . \quad . \quad (6.9)$$

be zero, which might be interpreted as requiring that all the diagonal elements of

$$\int_0^{2\pi} \log \mathbf{F}(\omega)d\omega \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.10)$$

be independent of θ. Condition (6.9) reduces in the single series case to the condition for unbiasedness, and so is then always fulfilled.

(7) *Tests of Fit*

Substituting $(\hat{\theta} - \theta)$ from (6.6) into (6.4) we obtain

$$L - \hat{L} \doteq \tfrac{1}{2}\mathbf{L}_1'\mathbf{L}_{11}^{-1}\mathbf{L}_1 \doteq \tfrac{1}{2}\mathbf{L}_1'(E\mathbf{L}_{11})^{-1}\mathbf{L}_1 \quad . \quad . \quad . \quad . \quad (7.1)$$

and since $2E(L_{11}) = E(L_1 L_1')$ we have

$$L - \hat{L} \doteq \mathbf{L}_1' E(\mathbf{L}_1\mathbf{L}_1')^{-1}\mathbf{L}_1, \quad . \quad . \quad . \quad . \quad . \quad (7.2)$$

so that $L - \hat{L}$ is approximately equal to the sum of squares of p orthogonalized standard statistical variates, where p is the number of parameters fitted. Now, since the cumulants of $\mathbf{L}_1$ are all $0(N)$, we may expect that $\mathbf{L}_1$ will be asymptotically normally distributed, so that the same may be said of the standardized variates, and $L - \hat{L}$ is asymptotically distributed as χ^2 with p degrees of freedom.

This simple fact may be made the basis of a theory of fit testing. For, suppose that two hypotheses H_1 and H_2 involve respectively p_1 and p_2 parameters ($p_2 > p_1$), and that p_1 of H_2's parameters are identical with H_1's parameters, a fact which we may express by saying that H_2 *includes* H_1. If now the minimized values of L are respectively $\hat{L}_1$ and $\hat{L}_2$, then $L - \hat{L}_j$ is distributed as χ^2 with p_j degrees of freedom, and by the partition theorem

$$(L - \hat{L}_2) - (L - \hat{L}_1) = \hat{L}_1 - \hat{L}_2 \quad . \quad . \quad . \quad . \quad . \quad (7.3)$$

is asymptotically distributed as χ^2 with $p_2 - p_1$ degrees of freedom, since the variates contributing to $L - \hat{L}_1$ ($\partial L/\partial\theta_j$; $j = 1, 2 \ \ldots \ p_1$) also contribute to $L - \hat{L}_2$.

Theorem 10.—*If* $\hat{\mathrm{V}}_1$, $\hat{\mathrm{V}}_2$ *are the least square estimates of the total prediction variance on Gaussian hypotheses* H_1, H_2, *then*

$$\psi^2 = (N - p_2/q) \log_e(\hat{V}_1/\hat{V}_2)$$

is asymptotically distributed as χ^2 *with* $\mathrm{p}_2 - \mathrm{p}_1$ *degrees of freedom.*

To see this, we note that $\log \hat{p} = \text{const.} - \frac{1}{2}\hat{L} = \text{const.} - (N/2) \log \hat{V}$ (cf. Th. 7), so that $\hat{L} = \text{const.} + N \log \hat{V}$, and $\hat{L}_1 - \hat{L}_2 = N \log (\hat{V}_1/\hat{V}_2)$. The second order term, $-p_2/q$, while not directly indicated in the derivation, has been added to allow for lost "degrees of freedom" in the estimation of the residual variance V_2 (cf. Whittle, 1952*a* and *c*; Walker, 1952).

Now, the ratio $\hat{V}_1/\hat{V}_2$ measures the improvement in fit obtained by introducing the $p_2 - p_1$ extra parameters which distinguish H_2 from H_1. Alternatively, we may say that it provides a measure of the fit of H_1 relative to the wider alternatives permitted by H_2. From this point of view we have a test of fit, and Th. 9 gives the approximate distribution of the test statistic.

For an interesting alternative approach to the test problem, see Bartlett and Rajalakshman (1953).

Example.—Consider the hypotheses: H_1, that the variates $\mathbf{X}_t$ are uncorrelated with past values, and H_2, that $\mathbf{X}_t$ is linearly dependent upon $\mathbf{X}_{t-1}$. That is

$$H_1 \qquad \mathbf{A}_0 \mathbf{X}_t = \boldsymbol{\epsilon}_t$$

$$H_2 \qquad \mathbf{A}_0 \mathbf{X}_t + \mathbf{A}_1 \mathbf{X}_{t-1} = \boldsymbol{\epsilon}_t.$$

Estimating the elements of $\mathbf{A}_0$, and $\mathbf{A}_0$ and $\mathbf{A}_1$ by least squares (note that A_0 has only $q(q + 1)/2$ statistically distinguishable elements) we obtain

$$\hat{V}_1 = \left| (C_{jk}(0)) \right|$$

and

$$\hat{V}_2 = \begin{vmatrix} (C_{jk}(0)) & (C_{jk}(1)) \\ (C_{kj}(1)) & (C_{jk}(0)) \end{vmatrix} \Big/ \left| C_{jk}(0) \right| .$$

The test function is

$$\psi^2 = (N - (3q + 1)/2) \log_e(\hat{V}_1/\hat{V}_2),$$

and is approximately distributed as χ^2 with q^2 degrees of freedom. If means have been fitted, the factor multiplying the logarithm is best modified to $(N - (5q + 1)/2)$.

(8) *An Application to Sunspot Data*

To illustrate the application of the methods developed we shall describe part of an analysis of sunspot observations. An account of the complete investigation is to appear elsewhere (Whittle, 1953) so we shall restrict ourselves to that which is relevant in the present connection.

The observations give the total sunspot area for a series of 120 six-monthly periods, and for two belts of solar latitude (16°–21° N., and 16°–21° S.), so that we have two time series, whose terms we shall denote N_t and S_t ($t = 1, 2 \ldots 120$). A reasonable first assumption is that the spot generating mechanism is symmetric about the equator, so that it is plausible that the two variates follow a symmetric double autoregression

$$\begin{aligned} \alpha(U)\, N_t + \beta(U)\, S_t &= \varepsilon_t \\ \beta(U)\, N_t + \alpha(U)\, S_t &= \varepsilon'_t, \end{aligned} \qquad . \quad . \quad . \quad . \quad . \quad (8.1)$$

where ε_t and ε_t' may without loss of generality be assumed uncorrelated. Adding and subtracting these two equations (cf. the example of §3) we find that the two series, $x_t = N_t + S_t$, and $y_t = N_t - S_s$ develop independently one of the other.

It is most plausible that x_t and y_t depend largely upon immediately preceding values and upon values in the previous sunspot cycle (i.e., x_{t-22} and y_{t-22}). Fitting then simple autoregressions with lags 1, 2 and 22 to the two series, we find by the test of fit of Th. 9 that lags 2 are superfluous in the presence of lags 22, which improve the fit vastly. The fitted autoregressions become then

$$\begin{aligned} x_t &= 0{\cdot}503\, x_{t-1} + 0{\cdot}388\, x_{t-22} \\ y_t &= 0{\cdot}114\, y_{t-1} - 0{\cdot}107\, y_{t-22}. \end{aligned} \qquad . \quad . \quad . \quad . \quad . \quad (8.2)$$

Adding these two equations we obtain

$$N_t = 0{\cdot}309\, N_{t-1} + 0{\cdot}141\, N_{t-22} + 0{\cdot}200\, S_{t-1} + 0{\cdot}248\, S_{t-22}. \qquad . \quad . \quad (8.3)$$

Equation (8.3) corresponds to neither of the equations of (8.1) directly, but is one of the equations of the so-called *reduced form*, expressing N_t and S_t solely in terms of *past* values. It is just this reduced form which we shall require, however.

Alfvén (1950) has proposed a theory of sunspot generation which leads to the conclusion that the sunspot intensity in the one hemisphere should directly depend upon that in the opposite hemisphere in the previous cycle. Consequently, the coefficient of S_{t-22} in (8.3) should be relatively large, and this is the point we intend to test.

To test the significance of the S_{t-22} coefficient, we set up a model in which it is assumed zero, i.e., in which the coefficients of x_{t-22} and y_{t-22} in (8.2) are required to be equal, so that

$$\begin{aligned} x_t &= ax_{t-1} + bx_{t-22} \\ y_t &= cy_{t-1} + by_{t-22} \end{aligned} \qquad . \quad . \quad . \quad . \quad . \quad . \quad (8.4)$$

say. By (5.3) the estimation equations for such a scheme are derived from the condition that

$$\begin{aligned} &\log(vv') + [(1 + a^2 + b^2)\, C_0 - 2aC_1 - 2bC_{22} + 2abC_{21}]/v \\ &\quad + [(1 + c^2 + b^2)\, C'_0 - 2cC'_1 - 2bC'_{22} + 2cbC'_{21}]/v' \end{aligned} \qquad . \quad . \quad (8.5)$$

be a minimum, where v, v' are the prediction variances, and C_s, C'_s the s^{th} observed autocovariances of the x_t and y_t series respectively. The total estimated prediction variance is then $\hat{v}\,\hat{v}'$. We find that when we take the more general model corresponding to (8.2) then $\hat{v} = 421{\cdot}969$ and $\hat{v}' = 234{\cdot}587$, while for the more restricted model (8.4) $\hat{v} = 529{\cdot}969$ and $\hat{v}' = 236{\cdot}791$. Thus (see Th. 9)

$$\psi^2 = \left(120 - \frac{4}{2}\right) \log \frac{(529{\cdot}969)\ (236{\cdot}791)}{(421{\cdot}969)\ (234{\cdot}587)} = 28{\cdot}76.$$

On hypothesis (8.4) ψ^2 should be distributed as χ^2 with one degree of freedom. The probability of obtaining so large a χ^2 as 28·76 is vanishingly small, however, and we conclude that the S_{t-22} coefficients is extremely significant.

The coefficient of N_{t-22} is also significant, so that N_t has a significant direct dependence upon N_{t-22}. This would not have been expected upon Alfvén's original theory, and a certain reflection region must be introduced to account for it, a region whose existence Alfvén has independently deduced by purely physical considerations.

While it may be expected that model (8.3) reproduces the more important features of the mechanism, it is certainly a gross oversimplification of reality. In an attempt to investigate the "fine structure" of the mechanism the author has therefore fitted a double autoregression involving all lags from 1 to 26. The simple model (8.3) is confirmed, but evidence is also found for several new phenomena. (For a full description of these results see Whittle, 1953.)

The author would wish to express his indebtedness to Professor H. Wold, Upsala, for a number of stimulating conversations on the subject of multiple series.

References

ALFVÉN, H. (1950), *Cosmical Electrodynamics.* Oxford.
BARTLETT, M. S. (1946), *Stochastic Processes. Mimeographed N. Carolina lecture notes.*
BARTLETT, M. S., and RAJALAKSHMAN, D. V. (1953), *J. R. Statist. Soc.*, B, **15**, 107–124.
CRAMÉR, H. (1940), *Ann. Math.*, **41**, 215–230.
FRISCH, R. (1938), *Statistical versus Theoretical Relationships in Economic Macrodynamics. Mimeographed memorandum.*
KOLMOGOROFF, A. N. (1941), *Bull. Univ. Moscou*, **5**, 3–14.
MANN, H., and WALD, A. (1943), *Econometrica*, **11**, 173–220.
MORAN, P. (1949), *Biometrika*, **36**, 63–70.
MOYAL, J. (1949), *J. R. Statist. Soc.*, B, **11**, 150–210.
WALKER, A. (1952), *J. R. Statist. Soc.*, B, **14**, 117–134.
WHITTLE, P. (1951), *Hypothesis Testing in Time Series Analysis.* Upsala, Almquist and Wicksell.
—— (1952*a*), *Skand. AktuarieTidskr.*, 48–60.
—— (1952*b*), *Trabajos de Estadistica*, **3**, 43–57.
—— (1952*c*), *Biometrika*, **39**, 309–318.
—— (1953), *Astrophysical Journal.* (To appear.)
WIENER, N. (1949), *The Extrapolation, Interpolation and Smoothing of Stationary Time Series.* New York: Wiley.
WILKS, S. (1932), *Biometrika*, **24**, 471–494.
WOLD, H. (1938), *A Study in the Analysis of Stationary Time-Series.* Upsala, Almquist and Wicksell.
—— (1952), *Demand Analysis.* New York: Wiley.
ZASUHIN, V. (1941), *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS*, **23**, 435–437.