ESD-TDR-65-58

Technical Note 1965-4

L. A. Gardner, Jr.

Computationally Feasible Frequency Estimates in the Presence of Unknown Phase and Amplitude

11 March 1965

Prepared under Electronic Systems Division Contract AF 19(628)-500 by

Lincoln Laboratory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Lexington, Massachusetts

The work reported in this document was performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology, with the support of the U.S. Air Force under Contract AF 19(628)-500.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

COMPUTATIONALLY FEASIBLE FREQUENCY ESTIMATES IN THE PRESENCE OF UNKNOWN PHASE AND AMPLITUDE

L. A. GARDNER, JR.

Group 28

TECHNICAL NOTE 1965-4

11 MARCH 1965

LEXINGTON MASSACHUSETTS

ABSTRACT

The problem of estimating the angular frequency in a single trigonometric regression function, observed in the presence of correlated noise, can be approached by the method of Least Squares or procedures based on the theory of empirical spectral analysis. For real time applications, such methods generally are prohibitively time consuming.

We introduce two very simple and consistent frequency estimates which use at most the first three sample covariances, and derive and compare their large sample distributions. One interesting by-product of our calculations is a precise analysis of the asymptotic behavior of the (inconsistent) technique known in numerical analysis as Prony's method.

Accepted for the Air Force
Stanley J. Wisniewski
Lt Colonel, USAF
Chief, Lincoln Laboratory Office

TABLE OF CONTENTS

1. INTRODUCTION AND SUMMARY

Many statistical problems connected with re-entry physics require that we estimate the angular frequency in a single trigonometric regression function when the phase and amplitude are unknown. When confronted with the task of designing estimation procedures which must keep up with data arriving in real time, the computational aspect obviously becomes all important. In many problems of current physical interest, an optimum method from the point of view of statistical efficiency (a classical goal of theoretical statistics) leads to calculations which are prohibitively time consuming.

Let us suppose we have observations

$$z_t = \rho \cos(\theta t - \varphi) + v_t \qquad (t=1, 2, \ldots, n), \qquad (1.1)$$

where $\{v_t\}$ is some zero mean stationary noise process and $0 < \theta < \pi$ is to be estimated. Consider the simplest (unrealistic) situation in which $\varphi = 0$ and $\rho \neq 0$ is known. When the errors are independent normal variates with variance σ^2, the Maximum Likelihood estimate of the parameter value $\xi = \cos\theta$ is a solution of the equation

$$8 \sum_{t=1}^{n} z_t T_t'(\xi) = \rho U_{2n}'(\xi) \qquad (1.2)$$

where prime denotes differentiation with respect to ξ, and $T_n(\xi) = \cos n\theta$ and $U_{n-1}(\xi) = \sin n\theta / \sin\theta$ are the first and second kind Tchebichev polynomials in ξ of degree n and $n-1$ respectively. If there is a solution $\hat{\xi} = \hat{\xi}(z_1, \cdots, z_n)$ of Eq. (1.2) which consistently estimates ξ as $n \to \infty$, then $\hat{\theta} = \cos^{-1} \hat{\xi}$ will attain the Cramér-Rao lower bound:

$$\mathcal{E}(\hat{\theta} - \theta)^2 \cong \frac{6\sigma^2}{\rho^2} \frac{1}{n^3}. \qquad (1.3)$$

The price paid for this rate of convergence would be a difficult and lengthy computational job. Indeed, $\hat{\xi}$ is a solution of a polynomial of degree $2n-1$ with the first n coefficients depending on the data. In many applications it is sheer folly to consider solving such a numerical problem in real time. The difficulty is of course compounded when φ and ρ are unknown, since then we would have to locate the maximum of a highly nonlinear function of three variables. (The right side of Eq. (1.3) would be increased by a factor of 4, independent of whether or not ρ is known.)

A computationally more feasible method for estimating θ can be based on the theory of empirical spectral analysis (see Parzen, Ref. [9]). Let the noise in Eq. (1.1) now be a weakly stationary process with summable covariances $\sigma(k) = \mathcal{E} v_t v_{t+|k|}$. For indices $k = 0, \pm 1, \pm 2, \ldots, \pm(n-1)$, define sample covariances

$$C_n(k) = \frac{1}{n} \sum_{t=1}^{n-|k|} z_t z_{t+|k|} . \tag{1.4}$$

The time series $\{z_t\}$ is called asymptotically stationary because

$$\lim_n \mathcal{E} C_n(k) = \frac{1}{2} \rho^2 \cos k\theta + \sigma(k) \tag{1.5}$$

for each fixed k. The right side is the Stieltjes cosine transform of a bounded nondecreasing function having a discontinuity of magnitude $\frac{1}{2}\rho^2$ at the frequency θ. By transforming $m = o(\sqrt{n})$ of the covariances in Eq. (1.4) with suitable weights λ_k, which depend on the truncation point m, one obtains a sample function of frequency $S^*(\omega)$. In large samples this will have a mean value approximately equal to $f(\omega)$ if $\omega \neq \theta$, but at θ equal to $A^2 \rho^2 m + f(\theta)$ where A^2 is a known constant. One expects, therefore, the graph of $S^*(\cdot)$ to have a dominant peak at $\omega = \theta$ for large truncation points. The frequency, say θ^*, at which $S^*(\cdot)$ achieves its largest value is a consistent estimate of the unknown θ.

One can approximate θ^* without going through the time consuming operations involved in calculating $S^*(\omega)$ over a grid of ω values. (One would want to do so, of course, when there is interest in the overall shape of the spectrum.) This is accomplished by self-convolving the even sequence $\lambda_k C_n(k) = \psi_0(k)$ over $|k| \leq m$ to obtain a new sequence $\psi_1(k)$, normalized so that $\psi_1(0) = 1$, which is nonzero for $|k| \leq 2m$. Replacing $\psi_0(\cdot)$ by $\psi_1(\cdot)$, we generate a sequence $\psi_2(k)$ for $|k| \leq 4m$. In a recursive fashion, therefore, we arrive at $\psi_J(0) = 1$, $\psi_J(1), \psi_J(2), \cdots, \psi_J(K)$ after J iterations. By a proper choice of $K = o(2^J)$, the number of axis crossings of these K+1 numbers (excluding a certain small band around the zero value of the ordinate), after division by K, yields a ratio which converges to θ^*/π as $J \to \infty$. (Cf. Gardner, Ref. [4].)

Although this method is considerably faster than Least Squares, and correspondingly statistically less accurate, it is still in essence a (large) fixed sample size procedure. Such methods are generally not appropriate for real time estimation. What is needed are techniques which are recursive in the sample size n, so that the estimate can be rapidly updated as each new observation arrives using only small finite-memory computations. Albert and Gardner Ref. [1] introduce and analyze a class of estimates of a parameter in certain nonlinear regression problems with this point of view in mind. It is appropriate that we briefly discuss these procedures here. Suppose at time t we observe

$$z_t = F_t(\theta) + v_t \qquad (t = 1, 2, \ldots)$$

where the sequence $F_1(\cdot), F_2(\cdot), \ldots$ is prescribed. A recursively computed θ-estimate (suggested by the recursive formula for the Least Squares estimate in the linear case) is

$$\theta_n = \theta_{n-1} + a_n [z_n - F_n(\theta_{n-1})] \tag{1.6}$$

$$(n=1, 2, \ldots ; \quad \theta_0 \text{ arbitrary}).$$

The procedure (called "successive relinearization") is specified by a choice of the so-called gain sequence $a_1, a_2, \ldots$, where a_n may or may not depend on the available iterates $\theta_0, \theta_1, \ldots, \theta_{n-1}$. Under certain restrictions on the sequence of regression function derivatives, and the assumption that $\{v_t\}$ has uniformly bounded second moments, θ_n is a strongly consistent estimate of θ as $n \to \infty$ for a wide variety of recursively computable gains. For the case of independent errors with common variance, a transformed version of Eq. (1.6) with particular iterate-dependent gains is asymptotically efficient when (and only when) these errors are Gaussian. However, the mean square error of the estimate has the proper n dependence in all cases, i.e. it goes to 0 like (an assumption) $1/\sum_{t=1}^{n} F_t'^2(\theta)$ as $n \to \infty$. Those familiar with the behavior of stochastic approximation schemes (to which the above bears a very close formal similarity) will not be surprised that $F_t(\cdot)$ is required to be monotone over the parameter space Θ for all sufficiently large time indices:

$$\operatorname{sgn} F_t'(x) = s_t = \pm 1$$

independent of the argument $x \epsilon \Theta$. Such obviously fails to hold for the regression function in Eq. (1.1), because as time progresses it has an increasing number of zeroes with respect to θ. Even if the nuisance parameters φ and ρ were known, then, we could not directly apply this existing technique for nonlinear regression problems. (Can a scheme of the form of Eq. (1.6) be exhibited which consistently estimates θ for $F_t(\theta) = \cos \theta t$?)

We are going to treat the frequency estimation problem in its own right, and introduce two θ-estimates which are simple functions of at most the first three sample covariances in Eq. (1.4). Since the lags are fixed and do not increase with sample size, these estimates are easily generated in a recursive fashion. They are, of course, highly inefficient relative to estimates requiring an unlimited number of covariances, but this is the price paid for computational simplicity. Actually, we will deal with a model which is more general than Eq. (1.1); viz.,

$$z_t = \mathcal{L}[\rho \cos(\theta t - \varphi)] + u_t \tag{1.7}$$

where $\mathfrak{L}[\cdot]$ is a realizable linear operator with summable impulse response coefficients. We will assume $\{u_t\}$ is a one-sided zero mean linear process (and ultimately Gaussian for simplicity) with a summable fourth order moment sequence. In the special case

$$u_t = \mathfrak{L} v_t , \tag{1.8}$$

we obtain Eq. (1.1) passed through a linear filter. Although Eq. (1.8) is the case in most applications, there is no reason to so specialize Eq. (1.7) because the mathematics is essentially the same.

In Sec. 2 we derive the limiting statistical behavior of the sample covariances in Eq. (1.4) when $\{z_t\}$ is given by Eq. (1.7). In generalization of Eq. (1.5) we have, as a limit as $n \to \infty$ both with probability one and in mean square,

$$C_n(k) \to C(k) = \frac{1}{2}\rho^2 |A(\theta)|^2 \cos k\theta + \sigma(k)$$

for every fixed k, where $A(\cdot)$ is the complex-valued transfer function of the operator $\mathfrak{L}$ and $\sigma(\cdot)$ now denotes the covariance sequence of $\{u_t\}$. Any finite collection of random variables

$$D_n(k) = \sqrt{n}\,(C_n(k) - C(k)),$$

corresponding to distinct choices of the lag variable k, tend to joint normality. When $\{u_t\}$ is Gaussian, the covariance between $D_n(k_1)$ and $D_n(k_2)$ of the limiting distribution, written in terms of frequency, is

$$\psi_{k_1 k_2} = 4\pi\rho^2 |A(\theta)|^2 f(\theta) \cos k_1\theta \cos k_2\theta + 4\pi \int_{-\pi}^{\pi} \cos k_1\omega \cos k_2\omega\, f^2(\omega) d\omega$$

where $f(\cdot)$ is the spectral density function of $\{u_t\}$.

Using these results we introduce in Secs. 3 and 4 two different methods for estimating $\xi = \cos\theta$ $(-1 < \xi < 1)$. We present them here only for the special case when $\{u_t\}$ is a white Gaussian process with variance σ^2. The respective estimates are then

$$\xi_n^{I} = \frac{C_n(1)}{C_n(0) - \sigma^2}$$

$$\xi_n^{II} = \frac{C_n(2) + \sqrt{C_n^2(2) + 8\,C_n^2(1)}}{4\,C_n(1)} \quad .$$

We estimate θ by $\cos^{-1}\xi_n$. In each case, $\sqrt{n}(\xi_n - \xi)$ tends to be normally distributed about 0 as $n \to \infty$ with a variance $V^2(\xi)$. These functions are respectively given by

$$R^2 V_I^2(\xi) = \frac{1}{|A(\theta)|^4}(2\xi^2 + 1)$$

$$R^2 V_{II}^2(\xi) = \frac{1}{|A(\theta)|^4}\,\frac{4\xi^4 - 3\xi^2 + 1}{(2\xi^2 + 1)^2}$$

where $R = \frac{1}{2}\rho^2/\sigma^2$ is the "signal to noise" ratio. We have $V_{II}^2(\xi)/V_I^2(\xi) \leq 1$ with equality only at $\theta = \pi/2$, so that the additional computation in Method II offsets the knowledge of σ^2 presumed in Method I. In the general nonwhite case, ξ_n^I assumes knowledge of $\sigma(0)$ and $\sigma(1)$ and ξ_n^{II} of $\sigma(1)/\sigma(0)$ and $\sigma(2)/\sigma(0)$. In addition, the latter involves the sample variance $C_n(0)$.

Figure 1 is a computational flow chart (without initialization) for Method II in the white noise case. The input to a "box" is to be multiplied by the value of the enclosed symbol, while "circled" symbols indicate operations to be performed on the input (I). The top segment of the diagram is the recursive covariance calculation for lags 1 and 2. This portion is run continuously, i.e. the time index n is stepped by unity. When an estimate of θ is desired it is only necessary to connect the bottom segment to the top via the indicated terminals 1 and 2.

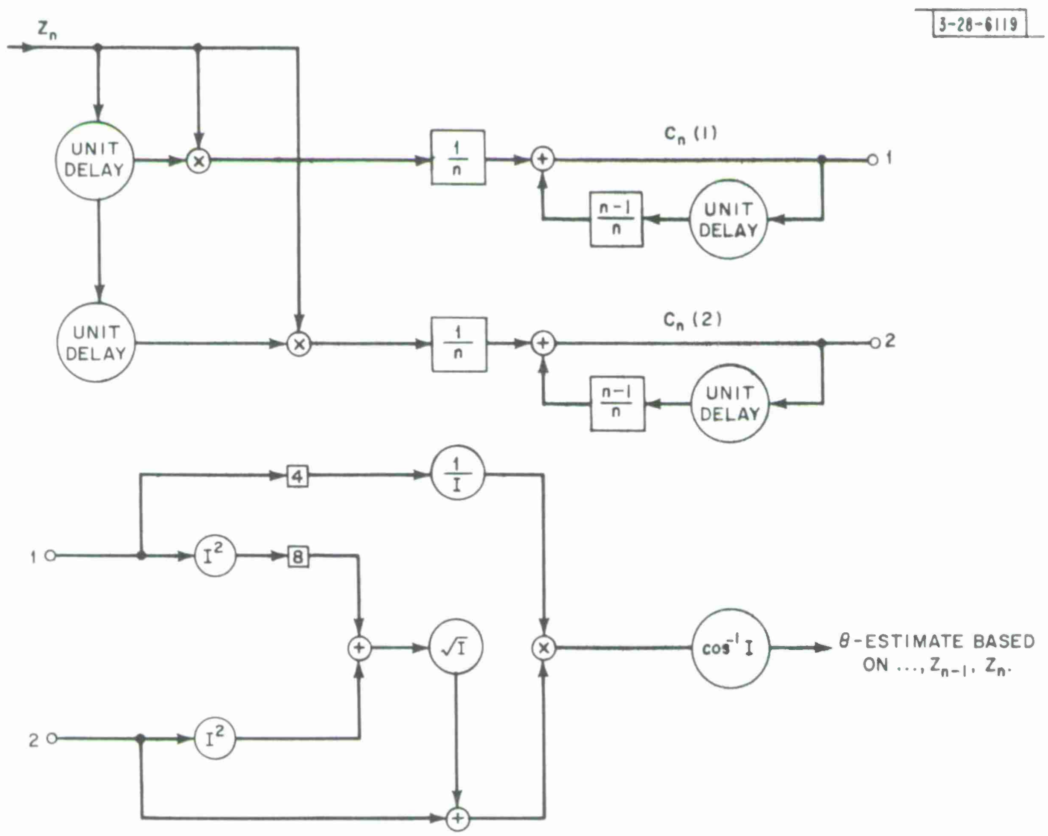

Fig. 1. Method II in the white noise case.

The methods are applied to two illustrative situations:

1. $\mathfrak{L} = \Delta^p$ and $u_t = \mathfrak{L} w_t$ with $\{w_t\}$ white, which results from removing a polynomial trend by differencing.
2. $\mathfrak{L} = \{1, -a_0, \ldots, -a_p\}$ and $u_t = w_t$ versus $\mathfrak{L} = \{1\}$ and $u_t = a_1 u_{t-1} + \cdots + a_p u_{t-p} + w_t$, i.e. a comparison of prewhitening and no prewhitening when a regression $\rho \cos(\theta t - \varphi)$ is disturbed by an autogressive scheme.

Numerical results are presented for these applications.

2. SOME LIMIT THEOREMS

In this section we consider time series of the form

$$z_t = \mathfrak{L}[\rho\cos(\theta t - \varphi)] + u_t$$

where $\mathfrak{L}[\cdot]$ denotes any realizable linear operator with summable impulse response coefficients, and $\{u_t\}$ is a centered stationary process whose moments up through order 4 obey certain restrictions. The parameters defining the regression sequence are arbitrary, with the exception that θ is presumed to be an interior point of $(0, \pi)$. Specifically, we are interested in the joint asymptotic statistical properties of sample covariances

$$\frac{1}{n}\sum_{t=1}^{n} z_{ct}\, z_{ct+h} \,,$$

corresponding to distinct choices of the lag variable h, when c is an arbitrarily fixed integer. We first deal with the deterministic components.

Theorem 1. Let $\mathfrak{L} = \{\alpha_0, \alpha_1, \alpha_2, \ldots\}$ be a linear operator defined by

$$\mathfrak{L}\, y_t = \alpha_0 y_t + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots \quad ,$$

for which $|\alpha_0| + |\alpha_1| + |\alpha_2| + \cdots < \infty$. Let

$$g_\theta(t) = \mathfrak{L}\,\rho\cos(\theta t - \varphi) \qquad (t = 0, \pm 1, \pm 2, \ldots)$$

where ρ is an arbitrary constant. Choose integers a, b and c (positive or negative). Then, for any $0 < \theta < \pi$ and φ, and some K which does not depend on a or b,

$$\left| \frac{1}{n}\sum_{t=1}^{n} g_\theta(ct+a)\, g_\theta(ct+b) - \tfrac{1}{2}\rho^2 |A(\theta)|^2 \cos(b-a)\theta \right| \le \frac{K}{n}$$

wherein

$$A(\theta) = \sum_{j=0}^{\infty} \alpha_j e^{ij\theta} \quad .$$

Proof. It is notationally convenient to work with the Tchebichev polynomials of the first and second kind

$$T_t(\xi) = \cos t\theta \qquad U_t(\xi) = \frac{\sin(t+1)\theta}{\sin\theta} \tag{2.1}$$

which are of degree $t \geq 0$ in

$$\xi = \cos\theta \qquad (-1 < \xi < 1). \tag{2.2}$$

Then

$$g_\theta(t) = \rho \cos\varphi \, \mathfrak{L}\, T_t(\xi) + \rho \sin\varphi \sqrt{1-\xi^2}\, \mathfrak{L}\, U_{t-1}(\xi) \tag{2.3}$$

where $T_{-t} = T_t$ and $U_{-t-1} = -U_{t-1}$. Without loss of generality we can obviously assume $\rho = 1$. Deleting the fixed arguments θ and ξ, we then have

$$\begin{aligned} g(t+a)\, g(t+b) = \; & \cos^2\varphi \, \mathfrak{L}\, T_{t+a} \mathfrak{L}\, T_{t+b} \\ & + \sin^2\varphi (1-\xi^2) \mathfrak{L}\, U_{t+a-1} \mathfrak{L}\, U_{t+b-1} \\ & + \cos\varphi \sin\varphi \sqrt{1-\xi^2} (\mathfrak{L}\, T_{t+a} \mathfrak{L}\, U_{t+b-1} + \mathfrak{L}\, T_{t+b} \mathfrak{L}\, U_{t+a-1}) \end{aligned} \tag{2.4}$$

for all negative and positive integers a, b and t. (It should be clear that by $\mathfrak{L}\, T_{t+a} \mathfrak{L}\, T_{t+b}$ we mean $\mathfrak{L}[T_{t+a}]$ times $\mathfrak{L}[T_{t+b}]$ and not $\mathfrak{L}[T_{t+a}\mathfrak{L}[T_{t+b}]]$.) In terms of the polynomials, standard trigonometric addition formulae become

$$\begin{aligned} 2\, T_t U_s &= T_{t+s} + T_{t-s} \\ 2(1-\xi^2) U_{t-1} U_{s-1} &= T_{t-s} - T_{t+s} \\ T_t U_{s-1} + T_s U_{t-1} &= U_{t+s-1} \quad . \end{aligned} \tag{2.5}$$

Using the first (resp. second) line of Eq. (2.5) in the first (resp. second) line of Eq. (2.4) we find, with the abbreviations

$$h = b-a \qquad \ell = b+a,$$

that

$$\mathfrak{L}T_{t+a}\mathfrak{L}T_{t+b} = \sum_{i,j\geq 0} \alpha_i\alpha_j T_{t+a-i}T_{t+b-j}$$

$$= \frac{1}{2}\sum_{i,j\geq 0} \alpha_i\alpha_j(T_{2t+\ell-i-j} + T_{h+i-j})$$

and

$$(1-\xi^2)\mathfrak{L}U_{t+a-1}\mathfrak{L}U_{t+b-1} = \frac{1}{2}\sum_{i,j\geq 0} \alpha_i\alpha_j(T_{h+i-j} - T_{2t+\ell-i-j}).$$

Similarly, in the third term of Eq. (2.4) we use the third line of Eq. (2.5) and get

$$\mathfrak{L}T_{t+a}\mathfrak{L}U_{t+b-1} + \mathfrak{L}T_{t+b}\mathfrak{L}U_{t+a-1}$$

$$= \sum_{i,j\geq 0} \alpha_i\alpha_j(T_{t+a-i}U_{t+b-j-1} + T_{t+b-i}U_{t+a-j-1})$$

$$= \sum_{i,j\geq 0} \alpha_i\alpha_j U_{2t+\ell-i-j-1}$$

because the interchange of i and j in the second product does not alter the value of the double sum. We substitute these last three equations into Eq. (2.4), replace the dummy t by ct, and then sum over $t = 1, 2, \ldots, n$. There results

$$\frac{1}{n}\sum_{t=1}^{n} g(ct+a)\,g(ct+b) = \frac{1}{2}\sum_{i,j\geq 0} \alpha_i \alpha_j T_{h+j-i}$$

$$+ \frac{1}{2}\cos 2\varphi \sum_{i,j\geq 0} \alpha_i \alpha_j \left[\frac{1}{n}\sum_{t=1}^{n} T_{2ct+\ell-i-j}\right] \tag{2.6}$$

$$+ \frac{1}{2}\sin 2\varphi \sum_{i,j\geq 0} \alpha_i \alpha_j \left[\frac{\sqrt{1-\xi^2}}{n}\sum_{t=1}^{n} U_{2ct+\ell-i-j-1}\right] \quad .$$

Now from the summation formulae (Knopp, p. 480, Ref. [6])

$$\begin{aligned} \sum_{t=1}^{n} \cos 2t\omega &= \frac{\sin(2n+1)\omega}{2\sin\omega} - \frac{1}{2} = \frac{\sin n\omega \cos(n+1)\omega}{\sin\omega} \\ \sum_{t=1}^{n} \sin 2t\omega &= \frac{\cos\omega - \cos(2n+1)\omega}{2\sin\omega} = \frac{\sin n\omega \sin(n+1)\omega}{\sin\omega} \quad , \end{aligned} \tag{2.7}$$

which hold for all real numbers ω, we see that

$$\left.\begin{array}{r} \left|\sum_{t=1}^{n} T_{2ct+r}(\xi)\right| \\ \\ \sqrt{1-\xi^2}\left|\sum_{t=1}^{n} U_{2ct+r-1}(\xi)\right| \end{array}\right\} \begin{aligned} &\leq \left|\sum_{t=1}^{n} \cos 2ct\theta\right| + \left|\sum_{t=1}^{n} \sin 2ct\theta\right| \\ &\leq 2/\left|\sin c\theta\right| \\ &= O(1) \quad \text{provided } \theta \neq 0, \pi \end{aligned} \tag{2.8}$$

as $n \to \infty$ independently of $r = 0, \pm 1, \ldots$. Since $\{\alpha_j\}$ is square summable we can take limits inside the double summations. Hence, both the second and third terms in

Eq. (2.6) go to 0 as 1/n, and do so uniformly in ℓ (but not in θ). The remaining term is

$$\frac{1}{2}\sum_{i\geq 0}\sum_{j\geq 0}\alpha_i\alpha_j\cos(h+j-i)\theta = \frac{1}{2}\cos h\theta\sum_{i\geq 0}\sum_{j\geq 0}\alpha_i\alpha_j\cos(i-j)\theta$$

$$= \frac{1}{2}\cos h\theta\left[\left(\sum_{j\geq 0}\alpha_j\cos j\theta\right)^2 + \left(\sum_{j\geq 0}\alpha_j\sin j\theta\right)^2\right] < \infty$$

which, since h=b-a, is the asserted result with ρ set to 1. Q.E.D.

Remark. Although we will have no need for the result, we note in passing what happens at the endpoints. From Eq. (2.6) we find the formula, after returning the multiplier ρ,

$$\frac{1}{n}\sum_{t=1}^{n} g_\theta(ct+a)\, g_\theta(ct+b) = \cos^2\varphi\,\rho^2|A(\theta)|^2\cos(b-a)\theta \qquad (\theta = 0 \text{ or } \pi)$$

which is valid for every n.

Remark. The conclusion of Theorem 1 is invariant under translations of the time index on which $\mathfrak{L}$ operates. Thus, we could just as well have started with $\mathfrak{L}y_t$ written forward in time, or even as a two-sided operator,

$$\mathfrak{L}y_t = \sum_{j=-\infty}^{\infty}\alpha_j y_{t-j}\ .$$

We have taken $\mathfrak{L}$ to be realizable since we will be interested only in statistical applications. An important special case is

$$\mathfrak{L} = \Delta^p$$

where Δ is the first backward difference operator. We have

$$\alpha_j = \begin{cases} (-1)^j \binom{p}{j} & \text{for } j = 0, 1, \ldots, p \\ 0 & \text{for } j \geq p+1 \end{cases} \quad . \tag{2.9}$$

The spectrum of this operator at θ is

$$\begin{aligned} |A(\theta)|^2 &= \left[\sum_{j=0}^{p} (-1)^j \binom{p}{j} e^{ij\theta} \right] \left[\sum_{j=0}^{p} (-1)^j \binom{p}{j} e^{-ij\theta} \right] \\ &= (1-e^{i\theta})^p (1-e^{-i\theta})^p \\ &= \left[\left(e^{-i\frac{\theta}{2}} - e^{i\frac{\theta}{2}} \right) \left(e^{i\frac{\theta}{2}} - e^{-i\frac{\theta}{2}} \right) \right]^p \\ &= \left(4 \sin^2 \frac{\theta}{2} \right)^p = (2-2\cos\theta)^p . \end{aligned} \tag{2.10}$$

This particular filter removes from a series $\{y_t\}$ any polynomial trend of degree not exceeding $p-1$.

Theorem 2. Let $\{u_t\}$ be a stationary time series with

$$\mathcal{E} u_t = 0 \qquad \mathcal{E} u_t u_{t+|k|} = \sigma(k)$$

and $\sigma(\cdot)$ summable. Assume also that $\{u_t\}$ has a summable fourth order moment sequence. Let

$$z_t = g_\theta(t) + u_t \qquad (t = 0, \pm 1, \pm 2, \ldots)$$

where $g_\theta(t)$ is defined in Theorem 1. Then, for any fixed integers h and c,

$$\frac{1}{n}\sum_{t=1}^{n} z_{ct} z_{ct+h} \rightarrow \tfrac{1}{2}\rho^2 |A(\theta)|^2 \cos h\theta + \sigma(h)$$

in mean square and with probability one as $n \rightarrow \infty$.

Proof. We introduce abbreviations

$$C_n(h) = \frac{1}{n}\sum_{t=1}^{n} z_{ct} z_{ct+h}$$

$$\gamma_\theta(h) = \tfrac{1}{2}\rho^2 |A(\theta)|^2 \cos h\theta \;, \tag{2.11}$$

which will be used throughout. We further let

$$X_n(h) = \frac{1}{n}\sum_{t=1}^{n} g_\theta(ct+h)\, u_{ct}$$

$$Y_n(h) = \frac{1}{n}\sum_{t=1}^{n} g_\theta(ct)\, u_{ct+h} \tag{2.12}$$

$$Z_n(h) = \frac{1}{n}\sum_{t=1}^{n} u_{ct} u_{ct+h} - \sigma(h).$$

By hypothesis these latter random variables are centered at expectations. It follows from the definitions and Theorem 1 that

$$C_n(h) - \gamma_\theta(h) - \sigma(h) = X_n(h) + Y_n(h) + Z_n(h) + O(1/n) \tag{2.13}$$

where the order term is a sure one as $n \rightarrow \infty$. The conclusion will be at hand if we can show that each of the averages in Eq. (2.12) tends to 0 both in mean square and with probability one. Such is easily accomplished with the help of the following

Law of Large Numbers for dependent random variables (see Parzen, p. 419, Ref. [7]):

Let $\overline{x}_n = \sum_{t=1}^{n} x_t/n$ where $x_1, x_2, \ldots$ is a sequence of centered random variables with uniformly bounded variances. Then a sufficient (and also necessary) condition that $\mathcal{E}\overline{x}_n^2 \to 0$ as $n \to \infty$ is $\mathcal{E}\overline{x}_n x_n \to 0$. Furthermore, if $\mathcal{E}\overline{x}_n x_n = O(1/n^{\epsilon})$ for some $\epsilon > 0$, then $\overline{x}_n \to 0$ with probability one.

The first two averages in Eq. (2.12) are trivial to handle. For the first we have

$$|\mathcal{E}X_n(h)\, g_\theta(cn+h) u_{cn}| = \frac{1}{n}\, \left|\sum_{t=1}^{n} g_\theta(ct+h)\, g_\theta(cn+h)\, \sigma(c(n-t))\right|$$

$$\leq \frac{\text{const.}}{n} \sum_{t=0}^{n-1} |\sigma(ct)| = O(1/n)$$

because $\sum_t |\sigma(ct)| < \infty$ obviously follows from $\sum_t |\sigma(t)| < \infty$. The same kind of bound obviously holds for $Y_n(h)$. For the remaining sequence we find

$$|\mathcal{E}Z_n(h)\,[u_{cn}u_{cn+h} - \sigma(h)]| \leq \frac{1}{n}\sum_{t=1}^{n} |\mathcal{E}u_{ct}u_{ct+h}u_{cn}u_{cn+h} - \sigma^2(h)|. \qquad (2.14)$$

The fourth order moment sequence

$$\mu(k_1, k_2, k_3) = \mathcal{E}u_t u_{t+k_1} u_{t+k_2} u_{t+k_3}$$

can be written in the form

$$\mu(k_1, k_2, k_3) = \sigma(k_1)\sigma(k_2-k_3) + \sigma(k_2)\sigma(k_1-k_3) + \sigma(k_3)\sigma(k_1-k_2) + \mu_{NG}(k_1, k_2, k_3) \qquad (2.15)$$

where μ_{NG} is the fourth cumulant sequence. It is called the non-Gaussian contribution to μ because it vanishes identically when $\{u_t\}$ is a Gaussian process. If in Eq. (2.15) we set $k_1 = h$, $k_2 = c(n-t)$ and $k_3 = c(n-t) + h$, then the $\sigma^2(h)$ term in Eq. (2.14) is cancelled by the leading term in the Gaussian contribution to μ. After reversing the direction of summation, we therefore have

$$|\mathcal{E} Z_n(h)[u_{cn}u_{cn+h} - \sigma(h)]|$$

$$\leq \frac{1}{n}\sum_{t=0}^{n-1} |\sigma^2(ct) + \sigma(ct+h)\sigma(ct-h) + \mu_{NG}(h, ct, ct+h)|$$

which by the summability assumption is $O(1/n)$. Q.E.D.

<u>Theorem 3.</u> In the notation of Theorem 2, introduce the deviations

$$D_n(h) = \sqrt{n}\left[\frac{1}{n}\sum_{t=1}^{n} z_{ct}z_{ct+h} - \tfrac{1}{2}\rho^2|A(\theta)|^2\cos h\theta - \sigma(h)\right]$$

where h is an arbitrary fixed integer. In addition to the restrictions of Theorem 2, suppose that

$$\mathcal{E} u_t u_{t+k_1} u_{t+k_2} = 0 \qquad (t, k_1, k_2 = 0, \pm 1, \pm 2, \ldots) .$$

Then, for all integers a, b and c,

$$\lim_n \mathcal{E} D_n(a) D_n(b) = \left[2\rho^2|A(\theta)|^2 \sum_{k=-\infty}^{\infty} \sigma(ck)\cos ck\theta\right]\cos a\theta\cos b\theta$$

$$+ \sum_{k=-\infty}^{\infty} [\sigma(ck)\sigma(ck+b-a) + \sigma(ck+b)\sigma(ck-a)]$$

$$+ \sum_{k=-\infty}^{\infty} \mu_{NG}(a, ck, ck+b)$$

where μ_{NG} is the fourth order cumulant sequence of $\{u_t\}$.

Proof. From Eq. (2.12), whether a is equal or unequal to b,

$$\mathcal{E} X_n(a)\, Z_n(b) = \frac{1}{n^2} \sum_{t=1}^{n} \sum_{s=1}^{n} g_\theta(ct+a)\, [\mathcal{E} u_{ct} u_{cs} u_{cs+b} - \mathcal{E} u_{ct} \sigma(b)] \tag{2.16}$$

$$= 0$$

for every n by our simplifying assumption concerning third moments. From Eq. (2.13) we therefore have the formula

$$\begin{aligned} \mathcal{E} D_n(a) D_n(b) = n\, \mathcal{E}[& X_n(a) X_n(b) + Y_n(a) Y_n(b) \\ & + X_n(a) Y_n(b) + X_n(b) Y_n(a) \\ & + Z_n(a) Z_n(b)] \ . \end{aligned} \tag{2.17}$$

We handle all the terms, except the last, in the following fashion. Let $m = m(n)$ tend to infinity over positive integers with n in such a way that

$$m^2/n \to 0,$$

and define

$$\sigma_n(ck) = \begin{cases} \sigma(ck) & \text{for } |k| \leq m \\ 0 & \text{for } |k| \geq m+1 \ . \end{cases} \tag{2.18}$$

Put

$$S_n = \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{n} g(ct+a)\, g(cs+b)\, \sigma_n(c(s-t)) \tag{2.19}$$

where θ is fixed and for the moment now shown. If we write $\sigma(\cdot)$ in place of $\sigma_n(\cdot)$, this becomes the first quantity of interest, $n\mathcal{E} X(a) X_n(b)$. When we make the index change from s to k=s-t, we obtain

$$S_n = \frac{1}{n} \sum_{t=1}^{n} g(ct+a)\, f_n(t)$$

with

$$f_n(t) = \sum_{k=-t+1}^{n-t} g(c(t+k)+b)\, \sigma_n(ck).$$

There are three ranges of t to be considered in accordance with Eq. (2.18). Since $n-2m \to \infty$, we have for all large enough n

$$f_n(t) = \left.\begin{matrix} \sum_{k=-t+1}^{m} \\ \\ \sum_{k=-m}^{m} \\ \\ \sum_{k=-m}^{n-t} \end{matrix}\right\} g(c(t+k)+b)\, \sigma(ck) \qquad \text{for} \left\{\begin{matrix} 1 \le t \le m \\ \\ m+1 \le t \le n-m \\ \\ n-m+1 \le t \le n\,. \end{matrix}\right. \tag{2.20}$$

Therefore,

$$S_n = \frac{1}{n} \sum_{t=m+1}^{n-m} g(ct+a) \sum_{k=-m}^{m} g(c(t+k)+b)\, \sigma(ck)$$
$$+ \frac{1}{n} \left(\sum_{t=1}^{m} + \sum_{t=n-m+1}^{n} \right) g(ct+a) f_n(t)\ .$$

Now in both the first and third sums in Eq. (2.20) there are at most 2m summands possessing a uniform bound. Hence

$$S_n = \sum_{k=-m}^{m} \sigma(ck) \left[\frac{1}{n} \sum_{t=m+1}^{n-m} g(ct+a)\, g(c(t+k)+b) \right] + O(m^2/n)\ .$$

We write the sum in square brackets as

$$\left(\sum_{t=1}^{n} - \sum_{t=1}^{m} - \sum_{t=n-m+1}^{n} \right) g(ct+a)\, g(ct+ck+b) \; .$$

Herein the second and third sums are bounded in absolute value by m times a constant independent of k. We therefore have

$$S_n = \sum_{k=-m}^{m} \sigma(ck) \left[\frac{1}{n} \sum_{t=1}^{n} g(ct+a)\, g(ct+ck+b) \right]$$

$$+ O(m/n) \sum_{k=-m}^{m} \sigma(ck) + O(m^2/n).$$

But according to Theorem 1 the average within square brackets, using the notation in Eq. (2.11), differs in absolute value from $\gamma_\theta(ck+b-a)$ by no more than K/n where the constant can be taken independent of k. From the summability of $\sigma(\cdot)$ and the choice of m, there follows

$$\lim_n S_n = \sum_{k=-\infty}^{\infty} \sigma(ck)\, \gamma_\theta(ck+b-a) < \infty \; . \tag{2.21}$$

Finally, since there are $n-|k|$ indices $t, s = 1, 2, \ldots, n$ for which the difference $s-t$ equals k, we have from Eq. (2.12), Eq. (2.18) and Eq. (2.19)

$$\left| n \,\mathcal{E} X_n(a) X_n(b) - S_n \right| \le \frac{K'}{n} \sum_{t=1}^{n} \sum_{s=1}^{n} \left| \sigma(c(s-t)) - \sigma_n(c(s-t)) \right|$$

$$= K' \sum_{|k| \le n-1} \left(1 - \frac{|k|}{n} \right) \left| \sigma(ck) - \sigma_n(ck) \right|$$

$$= K' \sum_{m+1 \le |k| \le n-1} \left(1 - \frac{|k|}{n} \right) \left| \sigma(ck) \right| \; .$$

This goes to 0 as $n > m \to \infty$ because the infinite summation is finite.

Precisely the same limit in Eq. (2.21) is found to hold for the second term in Eq. (2.17). Therefore, for any a and b

$$\lim_n n\,\mathcal{E} X_n(a) X_n(b) = \lim_n n\,\mathcal{E} Y_n(a) Y_n(b)$$

$$= \sum_{k=-\infty}^{\infty} \sigma(ck)\,\gamma_\theta(ck+b-a)$$

$$= \left[\tfrac{1}{2}\rho^2 |A(\theta)|^2 \sum_{k=-\infty}^{\infty} \sigma(ck)\cos ck\theta\right] \cos(b-a)\theta$$

where the last equality is a consequence of the definition in Eq. (2.11) of $\gamma_\theta(\cdot)$ and the fact that $\sigma(ck)\sin ck\theta$ is an odd function of integers k. In a similar fashion, we obtain

$$\lim_n n\,\mathcal{E} X_n(a) Y_n(b) = \lim_n n\,\mathcal{E} X_n(b) Y_n(a)$$

$$= \sum_{k=-\infty}^{\infty} \sigma(ck)\,\gamma_\theta(ck+b+a) \quad .$$

Therefore, taking limits in Eq. (2.17),

$$\lim_n \mathcal{E} D_n(a) D_n(b) = \left[2\rho^2 |A(\theta)|^2 \sum_{k=-\infty}^{\infty} \sigma(ck)\cos ck\theta\right] \cos a\theta \cos b\theta$$
$$+ \lim_n n\,\mathcal{E} Z_n(a) Z_n(b) \ . \qquad (2.22)$$

With regard to the remaining limit we have from Eqs. (2.12) and (2.15)

$$n\,\mathcal{E} Z_n(a) Z_n(b) = \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{n} \mathcal{E}[u_{ct} u_{ct+a} - \sigma(a)]\,[u_{cs} u_{cs+b} - \sigma(b)]$$

$$= \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{n} [\mathcal{E} u_{ct} u_{ct+a} u_{cs} u_{cs+b} - \sigma(a)\,\sigma(b)\,]$$

$$= \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{n} [\,\sigma(c(s-t))\,\sigma(c(s-t)+b-a) + \sigma(c(s-t)+b)\,\sigma(c(s-t)-a)$$
$$+ \mu_{NG}(a,\, c(s-t),\, c(s-t)+b)]$$

$$= \sum_{|k| \leq n-1} \left(1 - \frac{|k|}{n}\right) [\,\sigma(ck)\,\sigma(ck+b-a) + \sigma(ck+b)\,\sigma(ck-a) + \mu_{NG}(a, ck, ck+b)]\,.$$

Using the summability assumptions, and letting n go to infinity, the resulting limit combines with Eq. (2.22) to give the asserted formula. Q.E.D.

Remark. The restriction to processes $\{u_t\}$ with vanishing third order moment sequences was made solely for the purpose of simplifying the final formula. When the assumption fails, the result will have two additional terms (cf. Eq. (2.16)) involving the third order moment sequence of two indices.

Remark. The non-Gaussian contribution to the limiting value of $\mathcal{E} D_n(a) D_n(b)$ takes on a particularly simple form when $\{u_t\}$ is a one-sided (for physical realizability) linear process in independent random variables $\{w_t\}$, i.e.

$$u_t = \beta_0 w_t + \beta_1 w_{t-1} + \beta_2 w_{t-2} + \cdots \qquad (\sum_{j \geq 0} |\beta_j| < \infty). \qquad (2.23)$$

If we express the input process moments in terms of cumulants

$$\mathcal{E} w_t = 0$$

$$\mathcal{E} w_t^2 = \kappa_2 = \sigma^2$$

$$\mathcal{E} w_t^3 = \kappa_3$$

$$\mathcal{E} w_t^4 = \kappa_4 + 3\sigma^4 ,$$

then we have

$$\sigma(k) = \sigma^2 \sum_j \beta_j \beta_{j+k}$$

$$\mathcal{E} u_t u_{t+k_1} u_{t+k_2} = \kappa_3 \sum_j \beta_j \beta_{j+k_1} \beta_{j+k_2} \tag{2.24}$$

$$\mu_{NG}(k_1, k_2, k_3) = \kappa_4 \sum_j \beta_j \beta_{j+k_1} \beta_{j+k_2} \beta_{j+k_3}$$

where $\beta_j \equiv 0$ for all $j < 0$. From these we find the representation

$$\sum_{k=-\infty}^{\infty} \mu_{NG}(a, ck, ck+b) = \frac{\kappa_4}{\sigma^4} \sigma(a)\, \sigma(b) \tag{2.25}$$

independent of c. κ_4/σ^4 is the familiar "coefficient of excess" over a $N(0, \sigma^2)$ distribution of $\{w_t\}$, and it vanishes for such a distribution.

In summary, we specialize our preceding results to linear Gaussian process, since in such cases it follows almost immediately that the random variables $D_n(h)$, corresponding to distinct choices of the lag variable, tend to normality.

Theorem 4. Let $\{u_t\}$ be a one-sided linear process in independently identically distributed 0 mean normal random variables, and suppose its covariance sequence

$$\sigma(k) = \mathcal{E} u_t u_{t+|k|}$$

is absolutely summable. Let $\mathfrak{L} = \{\alpha_0, \alpha_1, \alpha_2, \ldots\}$ be any linear operator with absolutely summable impulse response coefficients, and let

$$A(\omega) = \sum_{j=0}^{\infty} \alpha_j e^{ij\omega}$$

be its complex valued transfer function. Introduce the time series

$$z_t = \mathfrak{L}[\rho \cos(\theta t - \varphi)] + u_t \qquad (t = 0, \pm 1, \pm 2, \ldots)$$

where $0 < \theta < \pi$ and ρ, φ are arbitrary. Fix an integer c, and define the random variables

$$C_n(h) = \frac{1}{n} \sum_{t=1}^{n} z_{ct} z_{ct+h},$$

and sure quantities

$$\gamma_\theta(h) = \tfrac{1}{2}\rho^2 |A(\theta)|^2 \cos h\theta$$

$$\psi_\theta(a, b) = [2\rho^2 |A(\theta)|^2 \sum_{k=-\infty}^{\infty} \sigma(ck) \cos ck\theta] \cos a\theta \cos b\theta + \sum_{k=-\infty}^{\infty} [\sigma(ck)\sigma(ck+b-a) + \sigma(ck+b)\sigma(ck-a)]$$

where h, a and b are integers. Finally, put

$$D_n(h) = \sqrt{n}\,[C_n(h) - \gamma_\theta(h) - \sigma(h)] .$$

Then

(i) $C_n(h) \to \gamma_\theta(h) + \sigma(h)$ as $n \to \infty$ both in mean square and with probability one.

(ii) If $\underline{h} = (h_1, h_2, \ldots, h_r)$ is any finite collection of distinct integers, then the distribution of the vector of random variables

$$D_n(h_1), D_n(h_2), \ldots, D_n(h_r)$$

tends to that of a centered r-variate normal distribution as $n \to \infty$. The entry of the covariance matrix of the limiting distribution corresponding to $h_i, h_j \in \underline{h}$ is equal to the limiting value of the corresponding covariance, and their common value is $\psi_\theta(h_i, h_j)$. $(i, j = 1, 2, \ldots, r)$.

Proof. From Eq. (2.12), $\sqrt{n}\, X_n(h)$ and $\sqrt{n}\, Y_n(h)$ are normal for every n. It is a known result in time series analysis (Parzen, Ref. [8]) that

$$\sqrt{n}\, Z_n(h_1),\ \sqrt{n}\, Z_n(h_2), \ldots, \sqrt{n}\, Z_n(h_r)$$

tend to joint normality as $n \to \infty$ (for linear processes $\{u_t\}$ but not necessarily normal). Moreover, the covariances of the limiting distribution are given by the limiting values of $n \mathcal{E} Z_n(h_i) Z_n(h_j)$. It remains, therefore, to multiply Eq. (2.13) through by $\sqrt{n}$ and evaluate second order moments. (The bias is $O(1/\sqrt{n})$ so mean square and variance are asymptotically the same.) Theorem 3 gives the result of these calculations, where μ_{NG} is 0 by hypothesis. Q.E.D.

Remark. The function

$$f_c(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \sigma(ck) \cos ck\omega\ , \tag{2.26}$$

whose value at $\omega = \theta$ appears in the limiting covariance $\psi_\theta(a, b)$, becomes the spectral density function of $\{u_t\}$ for c=1. (It is an interesting problem to try to express $f_c(\cdot)$ in terms of c and $f_1(\cdot)$.) From Eq. (2.23) it follows that

$$f(\omega) = f_1(\omega) = \left| \sum_{j \geq 0} \beta_j e^{ij\omega} \right|^2 \frac{\sigma^2}{2\pi}\ . \tag{2.27}$$

Thus, in case $u_t = \mathfrak{L} w_t$, i.e. $\beta_j = \alpha_j$, we will have

$$f(\omega) = \frac{\sigma^2}{2\pi} \left| A(\omega) \right|^2 \tag{2.28}$$

for all ω. The second summation in $\psi_\theta(a, b)$, for c=1, can be written in terms of the spectral density (see (4.32) below).

3. METHOD I

We are now going to apply the results of the previous section to the following statistical problem: given observations

$$z_t = \mathcal{L}[\rho\cos(\theta t - \varphi)] + u_t \qquad (t = t_0, t_0+1, \ldots) \tag{3.1}$$

estimate $\theta \in (0, \pi)$ without knowledge of φ or $\rho \neq 0$. In accordance with our basic ground rule, we want to consider methods involving finite-memory computations which can be performed rapidly (the latter, of course, being a relative requirement). The model in Eq. (3.1) can arise in different ways, but more often than not it is the output of a linear filter whose input is disturbed by additive noise. That is, $u_t = \mathcal{L}v_t$ for some process $\{v_t\}$. To be concrete, we introduce two such situations which subsequently are used to illustrate our results.

Problem 1. Estimate θ from data

$$\begin{aligned} y_t = {} & \text{(polynomial in } t \text{ of degree at most } p-1) \\ & + \rho\cos(\theta t - \varphi) + w_t \qquad (t = 1, 2, \ldots) \end{aligned} \tag{3.2}$$

where $\{w_t\}$ is a white zero mean process with variance σ^2, and the unknown polynomial trend is not of interest. The latter can be removed, as already remarked after the proof of Theorem 1, by applying the p^{th} order difference operator $\mathcal{L} = \Delta^p$ to Eq. (3.2). Thus,

$$z_t = \Delta^p y_t \qquad u_t = \Delta^p w_t$$

with $t_0 = p+1$ when we take Δ to be the first backward difference. The covariance between u_t and u_{t+k} for a general $\mathcal{L}$ with memory of order p applied to white noise is, for $k \geq 0$,

$$\sigma(k) = \sum_{i=0}^{p}\sum_{j=0}^{p} \alpha_i \alpha_j \, \mathcal{E} w_{t-i} w_{t+k-j} = \sigma^2 \sum_{j=0}^{p-k} \alpha_j \alpha_{j+k} \, .$$

Upon substituting Eq. (2.9), this reduces to

$$\sigma(k) = \begin{cases} \sigma^2(-1)^k \binom{2p}{p-k} & k=0, \pm 1, \dots, \pm p \\ 0 & |k| > p \end{cases} \tag{3.3}$$

(see Feller, p. 62, Ref. [3]).

Problem 2. Estimate θ from data

$$y_t = \rho \cos(\theta t - \varphi) + v_t \qquad (t = 1, 2, \dots) \tag{3.4}$$

where $\{v_t\}$ is an autogressive scheme of order p driven by white noise with variance σ^2, i.e.

$$v_t = a_1 v_{t-1} + \dots + a_p v_{t-p} + w_t \, .$$

If the coefficients are known, we can apply the operator $\mathcal{L}$ defined by

$$\alpha_0 = 1 \quad \alpha_j = -a_j \qquad (j = 1, 2, \dots, p).$$

We then have Eq. (3.1) with

$$\sigma(k) = \begin{cases} \sigma^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \tag{3.5}$$

because $u_t = \mathcal{L} v_t = w_t$. On the other hand, if the coefficients are not known we perform no filtering, and impose conditions on the a_j's so that $u_t = v_t$ obeys the moment restrictions imposed by the theorems of Sec. 2. In this case, we would have

$$\sigma(k)/\sigma(0) = b_1 \lambda_1^k + \dots + b_p \lambda_p^k \tag{3.6}$$

with $\lambda_1, \ldots, \lambda_p$ (the roots of the characteristic equation) assumed less than 1 in modulus. From the point of view of estimating θ, it is not clear whether Eq. (3.4) should in every case be "prewhitened," even when the a_j's are known. As a by-product of our calculations we will be able to investigate this question.

Both in this section and the next we consider estimator sequences $\{\xi_n\}$, where n will always be linearly related to the number of observations in Eq. (3.1), of the transformed parameter value

$$\xi = \cos\theta \qquad (-1 < \xi < 1).$$

In practice, one takes the arccosine of these because it is the angular frequency θ which is of interest. This operation entails no difficulty from the standpoint of large sample distributions. Indeed, suppose the distribution of the random variable $\sqrt{n}\,(\xi_n - \xi)$ is known to converge as $n \to \infty$ to that of a normal random variable with mean 0 and variance $V^2(\xi)$. We express this property in symbols by

$$\sqrt{n}\,(\xi_n - \xi) \sim N(0, V^2(\xi)). \tag{3.7}$$

Let $g(\cdot)$ be a given function whose derivative $g'(\cdot)$ is continuous and nonzero at the true parameter point ξ. Then (the "delta" method)

$$\sqrt{n}(g(\xi_n) - g(\xi)) \sim N(0, g'^2(\xi)V^2(\xi)).$$

Suppose, furthermore, that $\xi_n \to \xi$ with probability one. Since the true parameter point by hypothesis is an interior point of (-1, 1) it follows, with probability arbitrarily close to 1, that

$$\theta_n = \cos^{-1}\xi_n, \tag{3.8}$$

becomes and remains defined beyond some index n (which by Egorov's Theorem is independent of the points in the underlying sample space). Taking $g(x) = \cos^{-1}x$, we get

$$\sqrt{n}(\theta_n - \theta) \sim N(0, Q^2(\theta)) \qquad Q^2(\theta) = \frac{V^2(\cos\theta)}{\sin^2\theta} \ . \tag{3.9}$$

This function, for the estimates we will be considering, blows up as θ approaches the endpoints. This is not a desirable feature, nor is it an unexpected one. It is difficult to see how one can avoid the situation without resorting to more computationally expensive schemes, such as those based on the Fourier transform of the appropriately weighted covariance sequence. Section 5 contains a brief discussion of one of these other approaches to the problem.

* * * * * * *

Our first method assumes that the lag 0 and 1 covariances of $\{u_t\}$ are known. Under the assumptions, and in the notation, of Theorem 4 the ratio

$$\boxed{\xi_n = \frac{C_n(1) - \sigma(1)}{C_n(0) - \sigma(0)}} \tag{3.10}$$

is a (strongly) consistent estimate of $\xi = \gamma_\theta(1)/\gamma_\theta(0)$ no matter what value we fix for the integer c when computing $C_n(0)$ and $C_n(1)$. (Nothing is changed asymptotically of course, when we start the summations at $t = t_0$ rather than $t=1$. Furthermore, we need only consider $c \geq 1$.) According to Theorem 2, consistency holds under weaker noise conditions. However, we will assume the restrictions imposed on $\{u_t\}$ by Theorem 4 are satisfied, in the interest of uniformity. After we derive the variance function V^2 for Eq. (3.10), we will consider the problem of choosing c when the z's are generated as in the illustrative problems.

Let us discard some excess notational baggage, and for the purposes of algebraic manipulation write

$$\left.\begin{array}{lll} C_h & \text{in place of} & C_n(h) \\ D_h & \text{"} & D_n(h) \\ \gamma_h & \text{"} & \gamma_\theta(h) \\ \psi_{ab} & \text{"} & \psi_\theta(a, b) \end{array}\right\} \text{ defined in Theorem 4.} \tag{3.11}$$

The first two are random variables depending on the sample size or, what is the essential thing, the highest data index ,

$$N = cn + h , \tag{3.12}$$

on the z's. The second pair in Eq. (3.11) depends on the parameter θ under estimation. Keeping this in mind, Eq. (3.10) is seen to be equivalent to

$$\sqrt{n}(\xi_n - \xi) = \frac{\gamma_0 D_1 - \gamma_1 D_0}{\gamma_0[C_0 - \sigma(0)]}$$

$$= \frac{D_1 - \xi D_0}{C_0 - \sigma(0)}$$

because γ_1 is $\xi \gamma_0$. The numerator tends to normality and the denominator converges in probability to γ_0. By a well-known theorem in large sample distribution theory (Cramér, Sec. 20.6, Ref. [2]) Eq. (3.7) follows with

$$V^2(\xi) = \frac{1}{\gamma_0^2}[\psi_{11} - 2\xi\psi_{01} + \xi^2\psi_{00}] \tag{3.13}$$

under the assumption that γ_0 is strictly positive. Now we can write, introducing the spectral notation in Eq. (2.26) and indication of the dependence on our selection of c,

$$\psi_{ab} = [4\pi\rho^2 |A(\theta)|^2 f_c(\theta)] \cos a\theta \cos b\theta + \varphi_{ab}^{(c)}$$

$$\boxed{\varphi_{ab}^{(c)} = \sum_{k=-\infty}^{\infty} [\sigma(ck)\sigma(ck+b-a) + \sigma(ck+b)\sigma(ck-a)]} \quad . \tag{3.14}$$

Upon substituting into Eq. (3.13) we find the first terms in ψ_{ab} depend on ξ in such a way that they do not contribute. Therefore

$$\boxed{\begin{aligned} V^2(\xi) &= \frac{1}{\gamma_0^2} [\varphi_{00}^{(c)} \xi^2 - 2\varphi_{01}^{(c)} \xi + \varphi_{11}^{(c)}] \\ \gamma_0 &= \tfrac{1}{2}\rho^2 |A(\theta)|^2 > 0 \ . \end{aligned}} \tag{3.15}$$

We note that $\varphi_{aa}^{(c)}$ must always be positive, because it is the limiting value of $n\mathcal{E} Z_n^2(a)$. $\gamma_0 > 0$ places restriction on the filter coefficients $\alpha_0, \alpha_1, \ldots, \alpha_p$ or θ.

By Eq. (3.9), now, $\sqrt{n}(\theta_n - \theta)$ has a large sample normal distribution with mean 0 and variance

$$Q^2(\theta) = \frac{\lambda^2(\xi)}{\gamma_0^2} \tag{3.16}$$

when we set

$$\lambda^2(\xi) = \frac{\varphi_{00}^{(c)} \xi^2 - 2\varphi_{01}^{(c)} \xi + \varphi_{11}^{(c)}}{1-\xi^2} \qquad (|\xi| < 1) \ . \tag{3.17}$$

If the entire covariance sequence $\sigma(\cdot)$ of $\{u_t\}$ is given, we can set up large sample confidence intervals on θ, which are free of unknowns, by consistently estimating the parameter values γ_0 and λ. (Weak consistency suffices, although our estimates are strong.) $C_0 - \sigma(0)$ is just such an estimate of the former. The numerator in Eq. (3.17) is positive (for, as a matter of fact, all real numbers ξ). When $|\xi_n| < 1$, therefore, we can certainly compute the real positive square root $\lambda(\xi_n)$. Since $\lambda(\cdot)$ is a continuous function at every interior point of $(-1, 1)$, it follows from $\xi_n \to \xi$ that $\lambda(\xi_n) \to \lambda(\xi)$ with probability one. Therefore, if we take k_ϵ such that

$$\sqrt{\frac{2}{\pi}} \int_0^{k_\epsilon} e^{-\frac{1}{2}x^2} dx = 1-\epsilon ,$$

the interval

$$\theta_n \pm \frac{k_\epsilon}{\sqrt{n}} \frac{\lambda(\xi_n)}{C_n(0) - \sigma(0)} \tag{3.18}$$

will have confidence coefficient $1-\epsilon$.

* * * * * *

The computations involved in Method I, i.e., Eq. (3.10), are relatively trivial. Indeed, the basic statistics are $C_n(0)$ and $C_n(1)$ which, being averages, can be generated recursively. Only the previous z(obtained c time units ago) need be retained. The inversion of the cosine at each step to get the estimate of frequency is an inherent part of our approach, and presents no real computational difficulty. The price we pay for computational rapidity is reflected in the variance of the limiting distribution. A quantitative examination of this function is made below for the illustrative problems.

With regard to determining a choice for c, the expression to be investigated in any given situation is not V^2 alone, but rather the large sample variance of ξ_n written in terms of the total number of z-observations used to achieve it. By Eq. (3.12), this is

$$V^2/n \cong c\,V^2/N$$

in large samples. It is c times the quadratic in Eq. (3.15), therefore, i.e.

$$L_\xi(c) = c[\,\varphi_{00}^{(c)}\xi^2 - 2\varphi_{01}^{(c)}\xi + \varphi_{11}^{(c)}] \,, \tag{3.19}$$

which we would like to minimize in some sense with respect to $c \geq 1$. Up to the multiplier $1/\gamma_0^2$, this is the loss incurred when we use c and ξ is the true state of nature. The coefficients are given in Eq. (3.14) as a function of c corresponding to the particular covariances $\sigma(\cdot)$ of $\{u_t\}$. We now consider this aspect, among others, for the two illustrative problems.

Application to Problem 1. The coefficients of interest are

$$\varphi_{00}^{(c)} = 2 \sum_{k=-\infty}^{\infty} \sigma^2(ck)$$

$$\varphi_{01}^{(c)} = 2 \sum_{k=-\infty}^{\infty} \sigma(ck)\,\sigma(ck+1) \tag{3.20}$$

$$\varphi_{11}^{(c)} = \sum_{k=-\infty}^{\infty} [\,\sigma^2(ck) + \sigma(ck+1)\,\sigma(ck-1)\,]$$

with $\sigma(\cdot)$ given by Eq. (3.3). From the same combinatorial summation formula used to obtain Eq. (3.3), we find the general expression

$$\varphi_{ab}^{(1)} = \sigma^4(-1)^{b-a}\left[\binom{4p}{2p-b+a} + \binom{4p}{2p-b-a}\right] . \tag{3.21}$$

This is valid for all integers $b \geq a \geq 0$, with the usual convention that $\binom{n}{r} \equiv 0$ for $r < 0$ or $r > n$. Let us consider Eq. (3.20) with $c=p+2$. Since $\sigma(\cdot)$ vanishes for all arguments exceeding p in absolute value, only the k=0 terms remain and

$$\varphi_{00}^{(p+2)} = 2\sigma^4 \binom{2p}{p}^2$$

$$\varphi_{01}^{(p+2)} = -2\sigma^4 \binom{2p}{p}\binom{2p}{p-1}$$

$$\varphi_{11}^{(p+2)} = \sigma^4 \left[\binom{2p}{p}^2 + \binom{2p}{p-1}^2\right].$$

The same expressions obviously hold for any larger setting of c. The calculation of closed expressions for the three sums in Eq. (3.20) over $2 \le c \le p+1$ we leave to the reader, and content ourselves with a comparison of Eq. (3.19) at the extreme values. We find

$$L_\xi(1) < L_\xi(p+2)$$

for every $|\xi| < 1$ and $p \ge 1$. In fact, as $p \to \infty$,

$$\frac{L_\xi(p+2)}{L_\xi(1)} \sim \frac{p \cdot \binom{2p}{p}}{1 \cdot \binom{4p}{2p}} \sim \sqrt{p}$$

(see Feller, p. 63, Ref. [3]). We will consider this sufficient reason for setting c=1.

The variance of the limiting normal distribution of $\sqrt{n}(\theta_n - \theta)$ for Problem 1 using Method I with c=1 is

$$Q_p^2(\theta) = \frac{4\sigma^4}{\rho^4} \cdot \frac{1}{2^{4p}}\binom{4p}{2p} \frac{2\cos^2\theta + \frac{8p}{2p+1}\cos\theta + \frac{4p^2+2p+1}{2p^2+3p+1}}{\sin^{4p}(\frac{1}{2}\theta)\sin^2\theta} \qquad (3.22)$$

where we have added indication of the dependence on the differencing order. This is obtained by using Eqs. (2.10) and (3.21) in the general formula of Eq. (3.15), factoring, and dividing by $\sin^2\theta$. The constant multiplier in Eq. (3.22) is the inverse of the square of the " signal to noise" ratio $\frac{1}{2}\rho^2/\sigma^2$. The value of Q^2 becomes extremely large with increasing differencing order over the lower end of the parameter range, as is clear from Table 1. For p=0, which means there is no polynomial trend in Eq. (3.2), the variance function is proportional to

$$2\,\mathrm{ctn}^2\theta + \csc^2\theta \ .$$

This is symmetric about $\pi/2$, and takes on its minimum value of 1 there (and this was our reason for the placement of the numerical constant). As $\theta \to 0$ the function increases to infinity like $1/\theta^2$. In space applications, for example, the trend is usually taken to be parabolic, i.e. p=3, so the curves corresponding to large p are mainly of academic interest. On the other hand, at the upper end of the range the variance initially decreases with increasing p, and then, of course, increases. Thus, if certain prior knowledge exists concerning θ, it may pay to difference the data considerably more times than the degree of the unwanted polynomial. For example, if it is known that $\theta \geq .8\pi$ we should difference the data 5 times no matter what the trend, of degree less than 4. Figure 2 shows this behavior over the upper half of the interval. The reason for this is that differencing tends to pull large frequencies towards the center of the interval, where estimation is easier.

θ/π \ p	0	1	2	3	4	5	6	7	8	9
.1	2.942(1)	3.614(4)	4.992(7)	7.265(10)	1.089(14)	1.665(17)	2.578(20)	4.032(23)	6.355(26)	1.008(30)
.2	6.683	5.515(2)	5.043(4)	4.838 (6)	4.775 (8)	4.799(10)	4.886(12)	5.023(14)	5.202(16)	5.420(18)
.3	2.584	4.620(1)	9.915(2)	1.907 (4)	4.057 (5)	8.777 (6)	1.922 (8)	4.247 (9)	9.453(10)	2.116(12)
.4	1.317	7.578	5.474(1)	4.086 (2)	3.114 (3)	2.409 (4)	1.884 (5)	1.485 (6)	1.178 (7)	9.401 (7)
.5	1.000	1.750	6.125	2.217 (1)	8.155 (1)	3.034 (2)	1.139 (3)	4.305 (3)	1.637 (4)	6.250 (4)
.6	1.317	5.164(-1)	9.919(-1)	2.117	4.606	1.012 (1)	2.235 (1)	4.966 (1)	1.108 (2)	2.480 (2)
.7	2.584	2.683(-1)	2.209(-1)	2.911(-1)	4.246(-1)	6.372(-1)	9.664(-1)	1.473	2.254	3.457
.8	6.683	4.223(-1)	1.421(-1)	8.461(-2)	6.944(-2)	6.818(-2)	7.335(-2)	8.290(-2)	9.610(-2)	1.130(-1)
.9	2.942(1)	1.814	5.237(-1)	2.302(-1)	1.246(-1)	7.670(-2)	5.158(-2)	3.706(-2)	2.807(-2)	2.220(-2)

Table 1. $\frac{\rho^4}{4\sigma^4} Q_p^2(\theta)$ given by Eq. (3.22).

The number within parentheses is the exponent of the power of 10 by which the entry is to be multiplied.

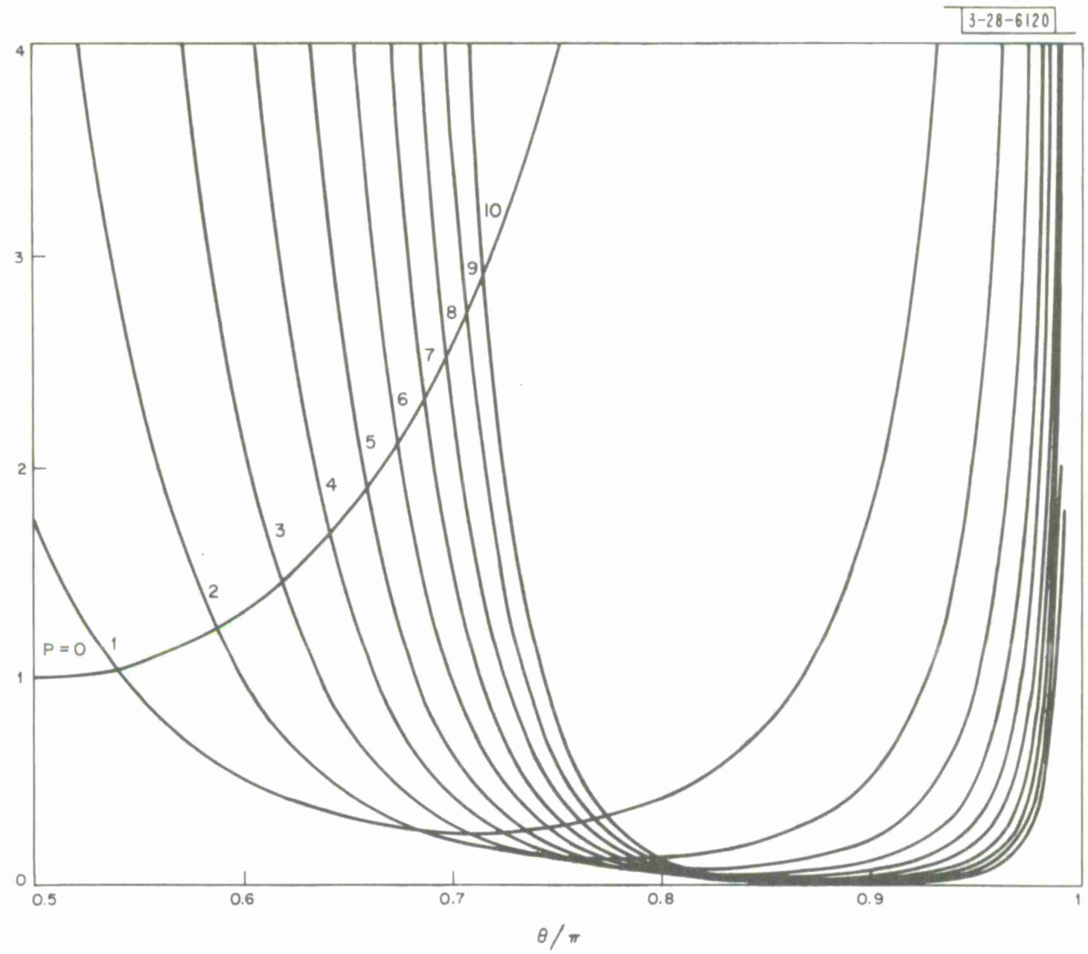

Fig. 2. $\frac{\rho^4}{4\sigma^4} Q_p^2(\theta)$ given by Eq. (3.22).

Application to Problem 2. Let us consider the simplest nontrivial situation in which the driving sequence in Eq. (3.4) obeys

$$v_t = a\,v_{t-1} + w_t \qquad (|a| < 1).$$

If a is known, we can take $\mathfrak{L} = \{1, -a\}$. Substituting Eq. (3.5) into Eq. (3.20) we get

$$\varphi_{00}^{(c)} = 2\sigma^4 \qquad \varphi_{01}^{(c)} = 0 \qquad \varphi_{11}^{(c)} = \sigma^4$$

independent of c. Since we want to keep the loss in Eq. (3.19) small, we should therefore take c=1 when computing the estimate in Eq. (3.10). The squared modulus of the transfer function of $\mathfrak{L}$ is

$$\begin{aligned} |A(\theta)|^2 &= \left[\sum_{j=0}^{p} \alpha_j \cos j\theta\right]^2 + \left[\sum_{j=0}^{p} \alpha_j \sin j\theta\right]^2 \\ &= (1-a\cos\theta)^2 + a^2\sin^2\theta \\ &= 1-2\xi a + a^2 \end{aligned}$$

in terms of $\xi = \cos\theta$. The function in Eq. (3.15) is therefore

$$V^2 = \frac{4\sigma^4}{\rho^4} \cdot \frac{2\xi^2 + 1}{(1-2\xi a + a^2)^2} \quad . \tag{3.23}$$

Now suppose we perform no filtering (as, for example, when a is not known). The process $u_t = v_t$ (assumed to have been arbitrarily initialized at $t = -\infty$) has the covariance sequence

$$\sigma(k) = \sigma(0)\,a^{|k|} \qquad \sigma(0) = \frac{\sigma^2}{1-a^2} \qquad (k=0, \pm1, \pm2, \dots).$$

The sums in Eq. (3.20) are found to be

$$\varphi_{00}^{(c)} = 2\sigma^2(0) \sum_{k=-\infty}^{\infty} a^{2c|k|} = 2\sigma^2(0) \frac{1+b}{1-b}$$

$$\varphi_{01}^{(c)} = 2\sigma^2(0) \sum_{k=-\infty}^{\infty} a^{c|k|} a^{|ck+1|} = 2\sigma^2(0) \frac{a^2+b}{a(1-b)}$$

$$\varphi_{11}^{(c)} = \sigma^2(0) \sum_{k=-\infty}^{\infty} \left[a^{2c|k|} + a^{|ck+1|} a^{|ck-1|}\right]$$

$$= \sigma^2(0) \left[2 \frac{1+b}{1-b} + a^2 - 1\right]$$

wherein

$$b = a^{2c}$$

belongs to (0, 1) no matter what our choice of $c \geq 1$. These combine to give for Eq. (3.15), since $|A(\theta)| = 1$,

$$\overline{V}^2 = \frac{4\sigma^4}{\rho^4} \cdot \frac{1}{(1-a^2)^2 (1-b)} \left[2(1+b)\xi^2 - 4\frac{a^2+b}{a}\xi + 1 + 3b + a^2(1-b)\right] \qquad (3.24)$$

where the bar indicates no prewhitening. The loss in using c, when ξ and a are the true parameter values, is

$$L_{\xi, a}(c) = c\,\overline{V}^2 .$$

We can adopt as a criterion for choosing c minimizing the worst that can happen to us. The result of maximizing over $|\xi| \leq 1$ is proportional to

$$c\,\frac{3 + 4\,|a| + a^2 + 4\,|a|^{2c-1} + 5a^{2c} - a^{2(c+1)}}{1 - a^{2c}} .$$

If a is known, we then choose the integer $c=c(a)$ for which this ratio is smallest.

One might believe that

$$E_{\xi,a}(c) = \frac{1 \cdot V^2}{c \cdot \bar{V}^2}$$

would not exceed 1 no matter what value we fix for c, because the numerator is the variance of the estimate which makes use of the assumed knowledge of the value of the nuisance parameter a. This is erroneous, since for any fixed integer $c \geq 1$ there exists a ξ and a (in fact, a continuum of values) for which $E > 1$. Indeed, one need only take $\xi=a$. Then the square bracketed quantity in Eq. (3.24) is just $(1-a^2)(1-b)$. This combines with Eq. (3.23) to yield

$$E_{a,a}(c) = \frac{2a^2 + 1}{c(1-a^2)} \quad .$$

Given the value of c, it remains to choose a^2 sufficiently close to 1 so that $E > 1$. The continuum of values is obtained by letting ξ range through a sufficiently small neighborhood of a, and appealing to continuity. It is important to remember that this comparison is based solely on the particular estimate in Eq. (3.10).

4. METHOD II

Our second procedure for estimating the parameter value θ in Eq. (3.1) requires a bit more computation than does the estimate based on Eq. (3.10). The new ξ-estimate is an unambiguous solution of a certain quadratic equation whose coefficients involve $C_n(2)$ as well as $C_n(0)$ and $C_n(1)$. The dependence on population quantities is via the correlations $\sigma(1)/\sigma(0)$ and $\sigma(2)/\sigma(0)$. Thus, when $\{u_t\}$ is a prescribed linear operation on white noise (as e.g. in Problem 1), there is no need to know the common variance. Furthermore, the coefficients have the property of being invariant under the replacement of $C_n(h)$ by $C_n(h) - \sigma(h)$. This feature greatly facilitates calculation of the limiting distribution of the solution.

The observation which suggests the procedure is that the Tchebichev polynomials in Eq. (2.1) both satisfy, for the appropriate pair of initial conditions, the same second order difference equation; viz.,

$$f_{t+1} - 2\xi f_t + f_{t-1} = 0 \qquad (t = 1, 2, \ldots).$$

Since $\mathcal{L}$ is a linear operator, the sequence of regression functions in Eq. (2.3) also obeys this recursion:

$$g_\theta(t+2) - 2\xi g_\theta(t+1) + g_\theta(t) = 0 . \tag{4.1}$$

It thus follows for Eq. (3.1), no matter what our choice of the integer $c \geq 1$, that

$$z_{ct+2} - 2\xi z_{ct+1} + z_{ct} = u_t^* \tag{4.2}$$

wherein

$$u_t^* = u_{ct+2} - 2\xi u_{ct+1} + u_{ct} . \tag{4.3}$$

The former is an identity in $\xi = \cos\theta$ relating three successive observables to hypothetical random variables. The covariances of $\{u_t^*\}$ in terms of those of $\{u_t\}$ are

$$\sigma_\xi(k) = 4\sigma(ck)\xi^2 - 4[\sigma(ck+1) + \sigma(ck-1)]\xi$$
$$+ 2\sigma(ck) + \sigma(ck+2) + \sigma(ck-2) \qquad (k = 0, \pm 1, \pm 2, \ldots). \tag{4.4}$$

Let $\underline{\Sigma}$, now, denote the $n \times n$ matrix with entry $\sigma_\xi(t-s)$ in row s and column t, and let σ_ξ^{st} be the st^{th} entry of $\underline{\Sigma}^{-1}$, which we presume exists for all $|\xi| < 1$. If $\underline{u}^*$ is the column vector with components $u_1^*, u_2^*, \ldots, u_n^*$, then the components of $\underline{u}^{**} = \underline{\Sigma}^{-\frac{1}{2}}\underline{u}^*$ are uncorrelated. If we apply the principle of Least Squares to the linear combinations of Eq. (4.2), an estimate of ξ is obtained by minimizing $u_1^{**2} + \cdots + u_n^{**2}$, i.e. finding a root of the equation

$$0 = \frac{\partial}{\partial \xi} \sum_{s=1}^{n} \sum_{s=1}^{n} \sigma_\xi^{st}(z_{cs+2} - 2\xi z_{cs+1} + z_{cs})(z_{ct+2} - 2\xi z_{ct+1} + z_{ct}), \tag{4.5}$$

where for notational convenience we are assuming $t_0 = 1$ in Eq. (3.1). Generally, this is highly nonlinear and, in addition, the inverse matrix changes in a nontrivial fashion with sample size.

Nonetheless, the above considerations together with our freedom to choose c as we please, suggest a computationally simple method for estimating ξ. First of all, the variance of $\{u_t^*\}$, which does not depend on c, is

$$\sigma_\xi(0) = 2\sigma(0)P^2(\xi) \qquad P^2(\xi) = 2\xi^2 - 4\rho_1\xi + 1 + \rho_2 \tag{4.6}$$

where

$$\rho_1 = \sigma(1)/\sigma(0) \qquad \rho_2 = \sigma(2)/\sigma(0)$$

are the first two correlations of $\{u_t\}$. We will assume

$$\xi \neq \rho_1 \pm \sqrt{\rho_1^2 - \tfrac{1}{2}(1+\rho_2)} \tag{4.7}$$

so that $P^2(\xi) > 0$. This may or may not place restriction on the unknown parameter in (-1, 1). For example, if $\{u_t\}$ is any stable autogressive scheme the discriminant is always negative, and Eq. (4.7) amounts to no restriction at all. Equivalently, we could assume that $\rho_1 \pm \sqrt{\rho_1^2 - \frac{1}{2}(1-\rho_2)}$ does not belong to (-1, 1). (Note that if ξ coincides with one of these points, then $u_t^* = 0$ with probability one and Eq. (4.2) can be solved exactly.) In any event, let us imagine, for the moment, that $\{u_t\}$ is a moving average of some given order $0 \le q < \infty$, i.e. all coefficients in Eq. (2.23) are 0 except $\beta_0, \beta_1, \ldots, \beta_q$. Then $\sigma(k) = 0$ for all $|k| \ge q+1$. For the selection

$$c = q+3$$

we see that $\sigma_\xi(k)$ in Eq. (4.4) is 0 for all $k \neq 0$ and all ξ. The "errors" in Eq. (4.2) are thereby made white. Since σ_ξ^{st} is 0 unless s=t, Eq. (4.5) reduces to

$$0 = \frac{\partial}{\partial \xi} \frac{1}{P^2(\xi)} \sum_{t=1}^{n} (z_{ct+2} - 2\xi z_{ct+1} + z_{ct})^2 . \tag{4.8}$$

We are going to use a solution of Eq. (4.8) to estimate ξ in any case, i.e. even when $\sigma(\cdot)$ does not vanish above some index. Furthermore, we can and will leave free the choice of c until we have calculated the variance function V^2.

Before proceeding, it should be pointed out that the technique known in numerical analysis as Prony's method (Hildebrand, p. 382, Ref. [5]), for approximating the angular frequency θ from data

$$z_t = \rho \cos(\theta t - \varphi) + u_t \qquad (t = 1, 2, \ldots, n)$$

when the errors are independent and σ^2 is "small", is equivalent to solving Eq. (4.8) with c=1 and the factor $1/P^2(\xi)$ deleted (and then taking the arccosine). This "least squares" estimate (cf. Eq. (4.2) with c=1) is

$$\xi^* = \frac{\sum_{t=1}^{n} z_{t+2} z_{t+1} + \sum_{t=1}^{n} z_{t+1} z_t}{2 \sum_{t=1}^{n} z_{t+1}^2} ,$$

and when n is large it will statistically behave like the ratio $2C_1/2C_0$. Thus, with probability one (see Eq. (4.11) below)

$$\xi^* \rightarrow \frac{\gamma_0 \xi + \sigma(1)}{\gamma_0 + \sigma(0)} = \frac{\frac{1}{2}\rho^2}{\frac{1}{2}\rho^2 + \sigma^2} \xi ,$$

so that ξ^* always underestimates ξ. For large values of the signal to noise ratio $R = \frac{1}{2}\rho^2/\sigma^2$, the right side will be

$$\left[1 - \frac{1}{R} + O\left(\frac{1}{R^2}\right)\right] \xi \quad .$$

On the other hand, the relative bias error can be quite large when R is small. Prony's method, of course, presumes the contrary. Nonetheless, if the error variance is known (as is usually the case in numerical approximations) the estimate ξ^*, simply modified by subtracting σ^2 out of the denominator, converges to the correct value (i.e. Method I). As we will see below in Eq. (4.25), we get a consistent ξ-estimate without knowing σ^2, or making any assumption on R, by using C_1 and C_2 rather than C_1 and C_0.

Let us return to Eq. (4.8), and simplify the notation by writing

$$a_t = z_{ct+2} \qquad b_t = z_{ct+1} \qquad c_t = z_{ct} .$$

After performing the differentiation, using Eq. (4.6) and multiplying through by $-P^4(\xi)/4 \neq 0$, we find that Eq. (4.8) is equivalent to

$$0 = -2(\Sigma b^2)\,\xi P^2(\xi) + (\Sigma(a+c)b)\,P^2(\xi)$$

$$+ (\xi - \rho_1)\,[\,4(\Sigma b^2)\xi^2 - 4\,(\Sigma(a+c)b)\,\xi + \Sigma(a+c)^2\,]$$

wherein Σb^2 is $\sum_{t=1}^{n} b_t^2$, etc. The leading term in $P^2(\xi)$ is $2\xi^2$, so happily the cubic term cancels. If we collect the coefficients and multiply through by $-1/2n$, then

$$0 = [\,\frac{1}{n}\,\Sigma(a+c)b - \frac{2\rho_1}{n}\,\Sigma b^2\,]\,x^2$$

$$- [\frac{1}{2n}\,\Sigma(a+c)^2 - \frac{1+\rho_2}{n}\,\Sigma b^2\,]\,x \qquad (4.9)$$

$$+ [\,\frac{\rho_1}{2n}\,\Sigma(a+c)^2 - \frac{1+\rho_2}{2n}\,\Sigma(a+c)b\,]$$

where we have written x in place of ξ (the true parameter value). Now we have, using the abbreviations in Eq. (3.11) ,

$$\frac{1}{n}\Sigma a^2\,,\quad \frac{1}{n}\Sigma b^2\,,\quad \frac{1}{n}\Sigma c^2 = C_0$$

$$\frac{1}{n}\Sigma ab\,,\quad \frac{1}{n}\Sigma bc = C_1$$

$$\frac{1}{n}\Sigma ac = C_2\;.$$

These are not algebraic equalities, but rather probabilistic ones valid in large samples (which suffices). If we substitute in Eq. (4.9), there results

$$0 = 2[\,C_1 - \rho_1 C_0]\,x^2 - [\,C_2 - \rho_2 C_0]\,x + [\,\rho_1(C_0 + C_2) - (1+\rho_2)C_1]\,. \qquad (4.10)$$

The solution of this equation corresponding to the positive square root is a (strongly) consistent estimate of ξ, as we now proceed to show.

Continuing to use the abbreviations in Eq. (3.11), we have from Theorem 4

$$
\begin{aligned}
C_0^* &= C_0 - \sigma(0) \to \gamma_0 \\
C_1^* &= C_1 - \sigma(1) \to \gamma_1 = \gamma_0 \xi \\
C_2^* &= C_2 - \sigma(2) \to \gamma_2 = \gamma_0(2\xi^2 - 1)
\end{aligned}
\tag{4.11}
$$

with probability one as $n \to \infty$. Upon replacing C_h by $\sigma(h)$ in Eq. (4.10) we see that all coefficients vanish by the definition of ρ_1 and ρ_2. Therefore the solutions of Eq. (4.10) must also be the solutions of

$$0 = 2[C_1^* - \rho_1 C_0^*]\, x^2 - [C_2^* - \rho_2 C_0^*]\, x + [\rho_1(C_0^* + C_2^*) - (1+\rho_2)C_1^*]\,. \tag{4.12}$$

Let ξ_n^+ (resp. ξ_n^-) denote the root of this equation corresponding to the positive (resp. negative) square root. Since the probability one limit of a solution as $n \to \infty$ is the corresponding solution of the limiting equation, it follows from Eq. (4.11) and Eq. (4.12) that

$$\xi_n^\pm \to \xi^\pm$$

where the numbers ξ^+ and ξ^- solve

$$0 = 2(\xi - \rho_1)x^2 - (2\xi^2 - 1 - \rho_2)x + [2\xi^2 \rho_1 - (1+\rho_2)\xi]\ . \tag{4.13}$$

The product of the roots is given by

$$\xi^+ \xi^- = \xi \cdot \frac{2\xi\rho_1 - (1+\rho_2)}{2(\xi - \rho_1)}\ .$$

Direct substitution shows the indicated factorization is the correct one, so either ξ^+ or ξ^- is ξ. Since this solution does not depend on ρ_1 or ρ_2, we can set them both to 0 and solve Eq. (4.13) to resolve the sign question. We get

$$\xi^+ = \frac{(2\xi^2 - 1) + \sqrt{(2\xi^2 - 1)^2 + 8\xi^2}}{4\xi} = \xi .$$

Returning to the original notation, our estimate of ξ is thus

$$\boxed{\begin{gathered}\xi_n = \text{positive square root solution of} \\ 2[C_n(1) - \rho_1 C_n(0)]\,x^2 - [C_n(2) - \rho_2 C_n(0)]\,x + [\rho_1(C_n(0)) + C_n(2) - (1+\rho_2)C_n(1)] = 0\end{gathered}}\, . \tag{4.14}$$

The equation can, of course, be divided through by $C_n(0) > 0$. The coefficients are then simpler expressions involving $C_n(h)/C_n(0) = R_n(h)$ $(h = 1, 2)$.

We next establish Eq. (3.7) for the estimate in Eq. (4.14), and evaluate the variance function. Let us now use the abbreviations

$$\begin{aligned} b_0 &= C_1^* - \rho_1 C_0^* & \qquad \beta_0 &= \gamma_1 - \rho_1\gamma_0 \\ b_1 &= C_2^* - \rho_2 C_0^* & \qquad \beta_1 &= \gamma_2 - \rho_2\gamma_0 \\ b_2 &= \rho_1(C_0^* + C_2^*) - (1+\rho_2)C_1^* & \qquad \beta_2 &= \rho_1(\gamma_0 + \gamma_2) - (1+\rho_2)\gamma_1 \end{aligned}$$

which, of course, have nothing to do with the β_j's in Eq. (2.23). In the notation of Theorem 4 and Eq. (3.11), $\sqrt{n}(C_h^* - \gamma_h)$ is D_h. Thus

$$\begin{aligned} B_0 &= \sqrt{n}(b_0 - \beta_0) = D_1 - \rho_1 D_0 \\ B_1 &= \sqrt{n}(b_1 - \beta_0) = D_2 - \rho_2 D_0 \\ B_2 &= \sqrt{n}(b_2 - \beta_2) = \rho_1(D_0 + D_2) - (1+\rho_2)D_1 \ . \end{aligned} \tag{4.15}$$

The statements that the estimate ξ_n and parameter ξ respectively satisfy Eqs. (4.12) and (4.13) are

$$2b_0\xi_n^2 - 2b_1\xi_n + b_2 = 0$$

$$2\beta_0\xi^2 - 2\beta_1\xi + \beta_2 = 0$$

after multiplying the latter by γ_0. If we subtract these two equations and multiply through by $\sqrt{n}$ we find the equality

$$[2b_0(\xi_n + \xi) - b_1]\ \sqrt{n}(\xi_n - \xi) = -2B_0\xi^2 + B_1\xi - B_2\ . \tag{4.16}$$

According to Eq. (4.15) and Theorem 4, the right side tends to normality as $n \to \infty$ with a variance

$$\lim_n \mathcal{E}(2B_0\xi^2 - B_1\xi + B_2)^2 = F^4(\xi)\ . \tag{4.17}$$

As shown in the next paragraph, this is a polynomial of degree 4 in the unknown parameter (hence the notation), even though the limiting second moments of the D_h's involve powers of ξ as high as 4. We have already shown that $\xi_n \to \xi$ with probability one. This, together with Eqs. (4.11) and (4.6), proves that the random variable in Eq. (4.16) multiplying the scaled deviation of interest converges to

$$4\beta_0\xi - \beta_1 = \gamma_0 P^2(\xi) \neq 0.$$

It follows that Eq. (3.7) is true, and the variance function

$$\boxed{V^2(\xi) = \frac{1}{\gamma_0^2}\ \frac{F^4(\xi)}{[2\xi^2 - 4\rho_1\xi + 1+\rho_2]^2}} \tag{4.18}$$

is finite provided Eq. (4.7) holds. (The correlations ρ_1 and ρ_2 are not to be confused with the amplitude ρ in $\gamma_0 = \frac{1}{2}\rho^2 |A(\theta)|^2$.) It remains to evaluate the numerator by substituting Eq. (4.15) into Eq. (4.17), and using the formula

$$\lim_n \mathcal{E} D_a D_b = \psi_{ab} \qquad (a, b = 0, 1, 2)$$

given in Eq. (3.14). The reduction entails several pages of rather frustrating algebraic manipulations, and we present only the highlights.

We first express Eq. (4.17) as a sum of quartics in ξ weighted by the 6 different ψ's. We find

$$F^4(\xi) = \psi_{00}[4\rho_1^2\xi^4 - \cdots + \rho_1^2] + \cdots + \psi_{22}[\xi^2 - 2\rho_1\xi + \rho_1^2] \ . \qquad (4.19)$$

The manner in which this is to be completed is clear from Table 2, reading ψ's in place of φ's. Now the first terms in $\psi_{00}, \psi_{01}, \ldots, \psi_{22}$ are proportional to $1, \xi, \ldots, (2\xi^2-1)^2$. If we multiply the corresponding rows of the table by these polynomials and collect the coefficients of $\xi^6, \xi^5, \ldots, \xi^0$ we find that they

	ξ^4	ξ^3	ξ^2	ξ^1	ξ^0
φ_{00}	$4\rho_1^2$	$-4\rho_1\rho_2$	$-4\rho_1^2 + \rho_2^2$	$2\rho_1\rho_2$	ρ_1^2
φ_{01}	$-8\rho_1$	$4\rho_2$	$4\rho_1(2+\rho_2)$	$-2\rho_2(1+\rho_2)$	$-2\rho_1(1+\rho_2)$
φ_{02}	0	$4\rho_1$	$-2(2\rho_1^2 + \rho_2)$	$2\rho_1(\rho_2 - 1)$	$2\rho_1^2$
φ_{11}	4	0	$-4(1+\rho_2)$	0	$(1+\rho_2)^2$
φ_{12}	0	-4	$4\rho_1$	$2(1+\rho_2)$	$-2\rho_1(1+\rho_2)$
φ_{22}	0	0	1	$-2\rho_1$	ρ_1^2

Table 2. Coefficients of the quartic numerator, Eq. (4.20), of the variance function.

all vanish. Thus, as was previously the case for Eq. (3.15), there is no contribution from $f_c(\theta)$, and the formula is indeed given by Eq. (4.19) with ψ's replaced by φ's (which we anticipated in forming the table). These latter are independent of the parameter, so working with the columns we obtain

$$
\boxed{\begin{aligned}
F^4(\xi) &= [4\rho_1^2\varphi_{00}^{(c)} - 8\rho_1\varphi_{01}^{(c)} + 4\varphi_{11}^{(c)}]\,\xi^4 \\
&\quad + [-4\rho_1\rho_2\varphi_{00}^{(c)} + 4\rho_2\varphi_{01}^{(c)} + 4\rho_1\varphi_{02}^{(c)} - 4\varphi_{12}^{(c)}]\,\xi^3 + \cdots \\
&\quad \cdots + [\rho_1^2\varphi_{00}^{(c)} - 2\rho_1(1+\rho_2)\,\varphi_{01}^{(c)} + \cdots + \rho_1^2\varphi_{22}^{(c)}]
\end{aligned}} \qquad (4.20)
$$

where in Table 2 $\varphi_{ab} = \varphi_{ab}^{(c)}$ is computed from Eq. (3.14).

One may obtain a neater expression by dividing through by $\sigma^2(0)$. The coefficients in Eq. (4.20) then become dimensionless functions of the correlations

$$
\rho_k = \sigma(k)/\sigma(0) \qquad k = 0, \pm 1, \pm 2, \ldots
$$

of $\{u_t\}$. When this entire sequence is known, we can set up large sample confidence intervals on θ in the same way as we did with Method I in Eq. (3.18). Indeed, it suffices to replace Eq. (3.17) by

$$
\lambda^2(\xi) = \frac{F^4(\xi)}{[2\xi^2 - 4\rho_1\xi + 1 + \rho_2]^2\,(1-\xi^2)} .
$$

* * * * * * *

It is not too difficult to compute Eq. (4.20) when $\{u_t\}$ is a moving average of known order q if (as was suggested initially for motivational reasons) we take c=q+3. Then the only contribution in Eq. (3.14) is from the k=0 term, so that

$$\varphi_{ab}^{(q+3)} = \sigma^2(0)\,(\rho_{b-a} + \rho_a \rho_b) \qquad a, b = 0, 1, 2 \ . \tag{4.21}$$

We get the coefficients in Eq. (4.20) by taking the "inner product" between the 6-vector created with this formula and the respective colums of Table 2. These turn out to be relatively simple functions of ρ_1 and ρ_2:

$$\begin{aligned} F^4(\xi)/\sigma^2(0) = {} & 4(1-\rho_1^2)\xi^4 - 4\rho_1(1-\rho_2)\xi^3 \\ & - (3+4\rho_2 + \rho_2^2 - 8\rho_1^2)\xi^2 \\ & + (1+\rho_2)\,(1+\rho_2 - 2\rho_1^2) \ . \end{aligned}$$

We note that there is no linear term.

Comparison with Method I in the white noise case. When $\{u_t\}$ in Eq. (3.1) is white the formula for F^4 becomes particularly simple. It is worthwhile comparing our two methods in this special case. We see that

$$\varphi_{00}^{(c)} = 2\sigma^4 \qquad \varphi_{11}^{(c)} = \sigma^4 \qquad \varphi_{22}^{(c)} = \sigma^4 \ ,$$

and that the other three φ's with different subscripts are 0. This is true independent of c, so we want to use c=1 (and not q+3=3). Setting $\rho_1 = \rho_2 = 0$ in Table 2, we obtain from Eq. (4.18) and Eq. (4.20)

$$V_{II}^2 = \frac{\sigma^4}{\gamma_0^2} \; \frac{4\xi^4 - 3\xi^2 + 1}{(2\xi^2 + 1)^2} \ . \tag{4.22}$$

For Method I with c=1 we have from Eq. (3.15)

$$V_I^2 = \frac{\sigma^4}{\gamma_0^2}(2\xi^2 + 1) \tag{4.23}$$

where $\gamma_0 = \frac{1}{2}\rho^2 |A(\theta)|^2$ is the same in both equations and varies with the operator $\mathcal{L}$. (We have given this before in Eq. (3.23) for a particular $\mathcal{L}$.) The relative efficiency of Method I to Method II is the ratio

$$e_\xi = \frac{V_{II}^2}{V_I^2} = \frac{4\xi^4 - 3\xi^2 + 1}{(2\xi^2 + 1)^3} \leq 1 , \tag{4.24}$$

which we have plotted in Fig. 3. This depends only on ξ^2, so as a function of θ it is symmetric about the midpoint $\pi/2$. The estimates in Eqs. (4.14) and (3.10) with these variance functions are

$$\xi_n^{II} = \frac{C_2 + \sqrt{C_2^2 + 8C_1^2}}{4C_1}$$

$$\xi_n^{I} = \frac{C_1}{C_0 - \sigma^2} \tag{4.25}$$

where C_h again abbreviates $C_n(h)$. The additional amount of computation required by Method II is evidently sufficient to overcome the knowledge of σ^2 utilized in Method I. In case there is no filtering, i.e. z_t is $\rho\cos(\theta t - \varphi)$ plus white noise with variance σ^2, numerical values of the variance function for θ_n^{II} can be obtained from Eq. (4.14) and Table 1. Indeed, we have

$$Q_{II}^2(\theta) = \frac{V_{II}^2(\cos\theta)}{\sin^2\theta} = \frac{4\sigma^4}{\rho^4} e_{\cos^{-1}\theta} \cdot \frac{\rho^4}{4\sigma^4} Q_0^2(\theta)$$

in the notation of Sec. 3.

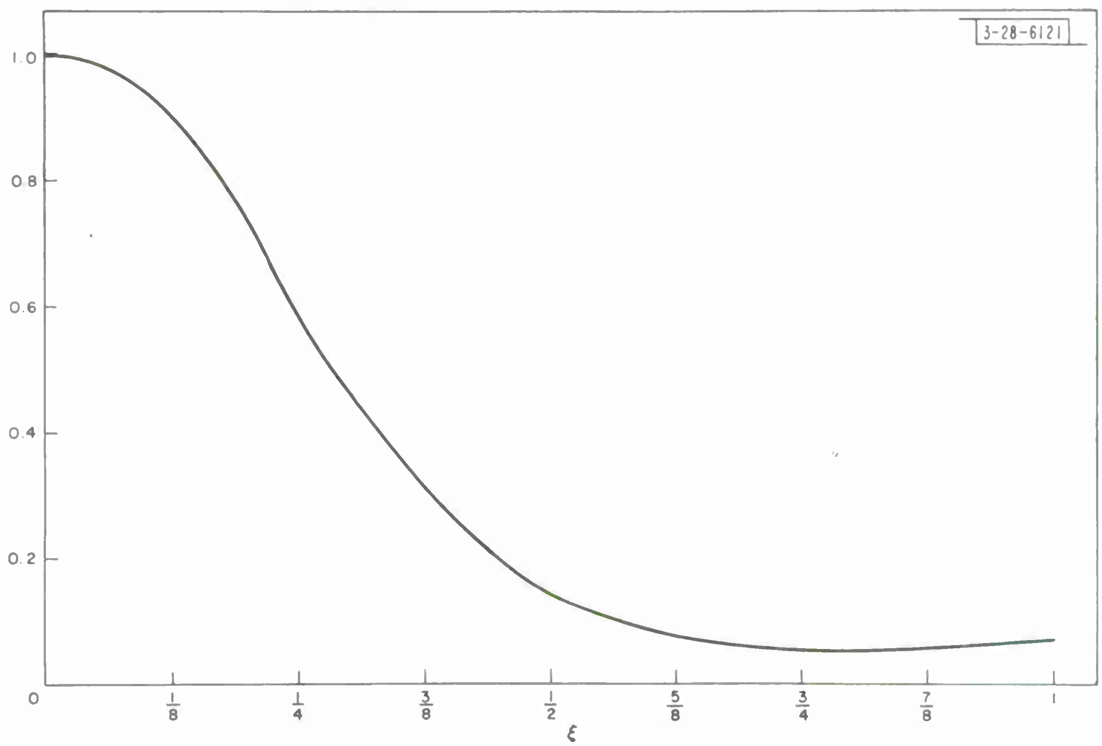

Fig. 3. Efficiency of Method I relative to Method II, in the white noise case, given by Eq. (4.24).

The problem of how to pick c for Method II in the illustrative problems is an order of magnitude more difficult than it is for Method I. We leave this question aside, since it can be resolved in any specific situation (in Problem 1 by using Eq. (3.21)).

* * * * * *

When the noise is a moving average of order q, we can write down a simple estimate of ξ which does not require knowledge of ρ_1 and ρ_2. The price paid is that the variance function has q+2 infinites at the roots of $T_{q+2}(\xi) = 0$. The restriction that ξ does not fall in the neighborhood of one of these points is of a different nature and obviously more severe than Eq. (4.7). Such an estimate is a limiting case of the one presented below, which illustrates the trade-off between knowledge of noise statistics and statistical accuracy. It is included mainly for academic reasons.

In generalization of Eq. (4.11) we have, as a limit with probability one as $n \to \infty$,

$$C_n^*(h) = C_n(h) - \sigma(h) \to \gamma_\theta(h) = J(\theta)\, T_h(\xi) \tag{4.26}$$

where we now denote $\gamma_\theta(0)$ by

$$J(\theta) = \tfrac{1}{2}\rho^2 |A(\theta)|^2 > 0. \tag{4.27}$$

Using the abbreviations in Eq. (3.11), the limit sequence obeys the recursion

$$\gamma_{h+1} - 2\xi\gamma_h + \gamma_{h-1} \equiv 0 \tag{4.28}$$

which defines the Tchebichev polynomials, and upon which we based Method II. Thus, for any fixed integer $r \geq 1$,

$$\boxed{\xi_n = \frac{C^*_{r+1} + C^*_{r-1}}{2C^*_r}} \tag{4.29}$$

is a strongly consistent estimate of ξ, and it can be computed once we know the values of $\sigma(r-1)$, $\sigma(r)$ and $\sigma(r+1)$. (For r=0, Eq. (4.29) can be interpreted as Eq. (3.10), and in this sense generalizes Method I.) Let us assume that $\gamma_r \neq 0$, i.e. that θ does not coincide with one of the r roots of $\cos r\theta = 0$:

$$\theta \neq \frac{2i+1}{2r}\pi \qquad (i=0, 1, \ldots, r-1). \tag{4.30}$$

Then

$$\sqrt{n}(\xi_n - \xi) = \frac{\sqrt{n}(C^*_{r+1} + C^*_{r-1})}{2C^*_r} - \frac{\sqrt{n}(\gamma_{r+1} + \gamma_{r-1})}{2\gamma_r}$$

$$\sim \frac{1}{2\gamma_r}(D_{r+1} + D_{r-1})$$

in the sense of equality of asymptotic probability distributions. According to Theorem 4, the variance of the limiting normal distribution is

$$V^2(\xi) = \frac{1}{4\gamma_r^2}(\psi_{r+1,r+1} + 2\psi_{r+1,r-1} + \psi_{r-1,r-1}).$$

From Eqs. (3.14), (4.27) and $\cos h\theta = T_h(\xi) = \gamma_h/J(\theta)$ we have

$$\psi_{ab} = 8\pi \frac{f_c(\theta)}{J(\theta)} \gamma_a\gamma_b + \varphi^{(c)}_{ab} \quad .$$

But according to Eq. (4.28)

$$\frac{1}{4\gamma_r^2}(\gamma^2_{r+1} + 2\gamma_{r+1}\gamma_{r-1} + \gamma^2_{r-1}) = \left(\frac{\gamma_{r+1} + \gamma_{r-1}}{2\gamma_r}\right)^2 = \xi^2 .$$

Therefore

$$V^2(\xi) = 8\pi\frac{f_c(\theta)}{J(\theta)}\xi^2 + \frac{1}{J^2(\theta)} \quad \frac{\varphi^{(c)}_{r+1,r+1} + 2\varphi^{(c)}_{r+1,r-1} + \varphi^{(c)}_{r-1,r-1}}{4\,T_r^2(\xi)} . \tag{4.31}$$

Unlike the situation in Methods I and II, the leading term in ψ_{ab} contributes to the variance function of the estimate.

For the choice c=1, we can write $\varphi_{ab}^{(1)}$ in terms of the spectral density $f(\omega) = f_1(\omega)$ given in Eq. (2.26). Using the inversion formula,

$$\sigma(k) = 2\int_0^{\pi} \cos k\omega \; f(\omega)d\omega ,$$

we obtain for Eq. (3.14), after translating the index in the second sum by b,

$$\begin{aligned}\varphi_{ab}^{(1)} &= \sum_{k=-\infty}^{\infty} \sigma(k)[\sigma(k+b-a) + \sigma(k+b+a)] \\ &= 8\pi \int_0^{\pi} \cos a\omega \cos b\omega \, f^2(\omega)d\omega .\end{aligned} \tag{4.32}$$

For the numerator of the second term in Eq. (4.31) we therefore have, using once again the basic recursion,

$$\begin{aligned}&\varphi_{r+1,r+1}^{(1)} + 2\varphi_{r+1,r-1}^{(1)} + \varphi_{r-1,r-1}^{(1)} \\ &\qquad = 8\pi \int_0^{\pi} [T_{r+1}(\cos\omega) + T_{r-1}(\cos\omega)]^2 f^2(\omega)d\omega \\ &\qquad = 8\pi \int_0^{\pi} [2\cos\omega \; T_r(\cos\omega)]^2 f^2(\omega)d\omega .\end{aligned}$$

Consequently, in terms of cosines,

$$V^2(\xi) = 8\pi \frac{f(\theta)\cos^2\theta}{J(\theta)} + \frac{8\pi}{J^2(\theta)} \frac{\int_0^{\pi} \cos^2 r\omega \cos^2\omega \, f^2(\omega)d\omega}{\cos^2 r\theta} . \tag{4.33}$$

The special case in which $u_t = \mathfrak{L}w_t$ is of interest. Letting

$$R = \frac{\rho^2}{2\sigma^2}$$

denote the signal to noise ratio, we have from Eq. (2.28)

$$J(\theta) = 2\pi R f(\theta).$$

Thus

$$V^2(\xi) = \frac{4\cos^2\theta}{R}\left[1 + \frac{1}{2\pi R} \cdot \frac{\int_0^\pi \cos^2 r\omega \cos^2\omega f^2(\omega)d\omega}{\cos^2 r\theta \cos^2\theta f^2(\theta)}\right] . \qquad (4.34)$$

The ratio involving the integral will become smaller as the spectrum of the operator $\mathfrak{L}$ becomes more peaked at $\omega=\theta$, as one would expect.

5. A DIFFERENT APPROACH

At present there is considerable interest in the subject of statistical spectral analysis. In particular, there is the problem of choosing and interpreting sample spectra which are computed from data modeled as a superposition of trigonometric terms additively disturbed by a 0 mean stationary noise process. Such processes are asymptotically stationary (cf. (5.2) below), and are said to possess a mixed spectrum. The time series with which we have been dealing,

$$z_t = \mathcal{L}[\rho \cos(\theta t - \varphi)] + u_t ,$$

is essentially a case in point. Using existing spectral estimation theory, we will present in this section another method for estimating the unknown parameter $0 < \theta < \pi$. Although the technique is computationally expensive, and not designed for real time usage, we include it for comparative reasons.

Let us redefine our sample covariances by

$$C_n(k) = \begin{cases} \dfrac{1}{n} \displaystyle\sum_{t=1}^{n-|k|} z_t z_{t+|k|} & k = 0, \pm 1, \ldots, \pm(n-1) \\ 0 & |k| \geq n . \end{cases} \tag{5.1}$$

We have without loss of generality assumed $t_0 = 1$, and we could (if desired) retain the integer $c \geq 1$. According to Theorem 2 we have, in the notation of Eq. (4.27),

$$\begin{aligned} \lim_n \mathcal{E} C_n(k) &= J(\theta) \cos k\theta + \sigma(k) \\ &\equiv C(k) \end{aligned} \tag{5.2}$$

for every fixed k. $A(\theta)$ is the value of the transfer function of the linear operator $\mathcal{L}$, and $\sigma(\cdot)$ is the (summable) covariance sequence of $\{u_t\}$ with spectral density function

$f(\omega) = f_1(\omega)$ given in Eq. (2.26). Thus,

$$C(k) = J(\theta)\cos k\theta + \int_{-\pi}^{\pi} \cos k\omega\, f(\omega)d\omega$$

$$= \int_{-\pi}^{\pi} \cos k\omega\, dF(\omega)$$

where $F(\cdot)$ is the nondecreasing spectral distribution function of $\{z_t\}$ consisting of a jump at $\omega=\theta$. of magnitude $J(\theta)$, plus the absolutely continuous spectral distribution function of $\{u_t\}$.

Following Parzen (Ref. [9]), we consider the following function of the first n observations $z_1, z_2, \ldots z_n$:

$$S_{n,m}(\omega) = \frac{1}{2\pi} \sum_{|k|<m} w(k/m)\, C_n(k) \cos k\omega \;. \tag{5.3}$$

This is periodic in ω with period 2π, and even about $\omega=0$ (as will be all functions of angular frequency, both random and nonrandom). The integer $m=m(n)$, called the truncation point, is to be chosen (for reasons which will become clear) so that

$$m \to \infty \qquad m^2/n \to 0$$

as $n \to \infty$. The function $w(\cdot)$ is called a covariance weight generator, and it is assumed to have the properties

$$w(0) = 1 \qquad w(-x) = w(x)$$

$$w(x) = 0 \qquad \text{for all } |x| \geq 1 \;.$$

Further conditions will be imposed on $w(\cdot)$ as we proceed. The Fourier transform

$$W_m(\omega) = \frac{1}{2\pi} \sum_{|k|<m} w(k/m) \cos k\omega, \tag{5.4}$$

is called the spectral window. It is said to be generated by the (aperiodic) function

$$W(z) = \frac{1}{2\pi} \int_{-1}^{1} w(x) \cos zx dx \qquad (-\infty < z < \infty), \qquad (5.5)$$

because

$$W_m(\omega) = mW(m\omega) \qquad (5.6)$$

holds in the limit of large m.

We first investigate the expectation of Eq. (5.3) for large sample sizes n. From Theorem 1 it follows that

$$\mathcal{E} C_n(k) = \frac{1}{n} \sum_{t=1}^{n-|k|} [g_\theta(t) g_\theta(t+|k|) + \sigma(k)]$$

$$= J(\theta) \cos k\theta + \left(1 - \frac{|k|}{n}\right) \sigma(k) + O(1/n),$$

where the order term is independent of the lag k. Thus, for fixed ω, we have

$$\mathcal{E} S_{n,m}(\omega) = \frac{J(\theta)}{2\pi} \sum_{|k|<m} w(k/m) \cos k\theta \cos k\omega + \frac{1}{2\pi} \sum_{|k|<m} w(k/m) \left(1 - \frac{|k|}{n}\right) \sigma(k) \cos k\omega$$

$$+ O(m/n)$$

because $w(\cdot)$ is bounded. Using Eqs. (2.5) and (5.4) in the first term we get

$$\mathcal{E} S_{n,m}(\omega) = J(\theta) \tfrac{1}{2} [W_m(\omega-\theta) + W_m(\omega+\theta)] + s_n(\omega) + O(m/n) \qquad (5.7)$$

where $s_n(\omega)$ stands for the second sum. We have

$$f(\omega) - s_n(\omega) = \frac{1}{2\pi} \sum_{|k|<m} [1-w(k/m)]\,\sigma(k)\cos k\omega$$

$$+ \frac{1}{2\pi n} \sum_{|k|<m} |k|\,\sigma(k)\cos k\omega + \frac{1}{2\pi} \sum_{|k|\geq m} \sigma(k)\cos k\omega \tag{5.8}$$

$$= 1° + 2° + 3° \ .$$

Suppose now that $\Sigma k^2 |\sigma(k)| < \infty$, so that

$$f''(\omega) = -\frac{1}{2\pi} \sum_{k=-\infty}^{\infty} k^2 \sigma(k) \cos k\omega$$

exists for all ω. Then $k\sigma(k)$ is certainly summable, and

$$|2°| = O(1/n)$$

$$|3°| \leq \frac{1}{2\pi m^2} \sum_{|k|\geq m} k^2 |\sigma(k)| = o(1/m^2) \ . \tag{5.9}$$

Suppose, further, that the weight generator is such that

$$\lim_{x \to 0} \frac{1 - w(x)}{x^2} = a \qquad (0 < |a| < \infty).$$

Most w's will have at least a local maximum at 0, so we might as well consider $a > 0$ (although this is not necessary). In any case, there exists an integer m' which depends on m in such a way that $m' \to \infty$, $m'/m \to 0$, and

$$|1 - w(x) - a x^2| < \frac{1}{m^2 \log m} \qquad \text{for all } |x| < \frac{m'}{m} \ .$$

We rewrite the leading term in Eq. (5.8) as

$$
\begin{aligned}
1^\circ &= \frac{a}{2\pi m^2} \sum_{|k|<m'} k^2 \sigma(k) \cos k\omega \\
&+ \frac{1}{2\pi} \sum_{|k|<m'} [1-w(k/m) - a(k^2/m^2)]\, \sigma(k) \cos k\omega \\
&+ \frac{1}{2\pi m^2} \sum_{m' \le |k| < m} \frac{1-w(k/m)}{k^2/m^2}\, k^2 \sigma(k) \cos k\omega \\
&= 4^\circ + 5^\circ + 6^\circ \; .
\end{aligned}
\tag{5.10}
$$

Since $|1-w(x)|/x^2$ is bounded for all x by(say) B,

$$
\begin{aligned}
|5^\circ| &< \frac{1}{2\pi m^2 \log m} \sum_{|k|<m'} |\sigma(k)| = o(1/m^2) \\
|6^\circ| &< \frac{B}{2\pi m^2} \sum_{m' \le |k| < m} k^2 |\sigma(k)| = o(1/m^2) \; .
\end{aligned}
\tag{5.11}
$$

Finally, with regard to 4°, we have

$$
\lim_{m' \to \infty} \frac{a}{2\pi} \sum_{|k|<m'} k^2 \sigma(k) \cos k\omega = -a f''(\omega) \; .
$$

Multiplying Eqs. (5.8) - (5.11) through by m^2, there results

$$
\lim_n m^2 [s_n(\omega) - f(\omega)] = a f''(\omega)
$$

from our supposition that $m^2/n \to 0$. We therefore have for Eq. (5.7)

$$\mathcal{E} S_{n,m}(\omega) = J(\theta)\,\tfrac{1}{2}[\,W_m(\omega-\theta) + W_m(\omega+\theta)\,]$$

$$+ f(\omega) + \frac{a f''(\omega)}{m^2} + O(m/n) \tag{5.12}$$

plus terms of smaller order than $1/m^2$.

Let us now restrict attention to weight generators $w(\cdot)$ which are twice differentiable throughout some neighborhood of the origin. Then, after inverting Eq. (5.5) and doing the differentiations, we obtain

$$a = -\tfrac{1}{2} w''(0) = \tfrac{1}{2} \int_{-\infty}^{\infty} z^2 W(z)dz \quad .$$

We must have $z^3 W(z) = O(1)$ as $|z| \to \infty$, for otherwise the integral would not be finite. There results from Eqs. (5.6) and (5.12)

$$\mathcal{E} S_{n,m}(\omega) = \tfrac{1}{2} W(0) J(\omega) m + f(\omega) + o(1),$$

$$J(\omega) \equiv \begin{cases} \tfrac{1}{2}\rho^2 |A(\theta)|^2 & \text{if } \omega = \theta \\ 0 & \text{otherwise} \end{cases} \quad . \tag{5.13}$$

The order term gives to 0 as the slower of $1/m^2$ and m/n (usually the former).

Consider next sequences m_1 and m_2 going to infinity with n in such a way that each is $o(\sqrt{n})$. Then it can be shown by working in the frequency domain (Eq. (5.10) in Ref. [9]) that

$$n\mathrm{Cov}\{S_{n,m_1}(\omega), S_{n,m_2}(\omega)\} = 2\pi W^2(0) J(\omega)\, f(\omega) m_1 m_2$$

$$+ 2\pi f^2(\omega)\, m_1 m_2 \int_{-\infty}^{\infty} W(m_1 z)\, W(m_2 z) dz + o(1). \tag{5.14}$$

In particular, for the variance at $0 < \omega < \pi$, we have

$$\operatorname{Var} S_{n,m}(\omega) \cong 2\pi W^2(0) J(\omega) f(\omega) \frac{m^2}{n} + f^2(\omega) \int_{-1}^{1} w^2(x) dx \frac{m}{n} \tag{5.15}$$

which goes to 0 as $n \to \infty$.

The method suggested in Ref. [9] for estimating, at a prescribed ω, the value of the (possibly zero) jump as well as the value of the spectral density function is based on choosing a small number of truncation points $m_1 < m_2 < m_3 < m_4$ (say). Writing y_i for $S_{n,m_i}(\omega)$, α for $\frac{1}{2} W(0) J(\omega)$ and β for $f(\omega)$, Eq. (5.13) gives in large samples a "regression model"

$$y_i = \alpha m_i + \beta + \epsilon_i$$

with correlated "errors". One can thus consider the computation of the Least Squares estimate of α and β; that is to say, locating the minimum of

$$\sum_{i=1}^{4} \sum_{j=1}^{4} \sigma_{\alpha,\beta}^{ij} (y_i - \alpha m_i - \beta)(y_j - \alpha m_j - \beta) \tag{5.16}$$

with respect to the unknown parameter pair. Here $\sigma_{\alpha,\beta}^{ij}$ denotes the ij^{th} element of the inverse of the matrix of quantities $\mathcal{E} \epsilon_i \epsilon_j$ derivable from Eq. (5.14).

But in the problem as originally posed, we assume there exists a line at some unknown θ. We wish to estimate this angular frequency; there is no interest in the continuous spectrum of the disturbing process $\{u_t\}$. For such situations, the following procedure was introduced as a method for estimating the absolute maximum of a spectral density function (Gardner, Ref. [4]). However, it can be applied to the present situation, and, in fact, to mixed spectra in general where there are a finite number of lines interior to $(0, \pi)$.

Let us replace m by m+1 in Eq. (5.3), and denote the coefficients by

$$\psi_0(k) = \begin{cases} w\left(\frac{k}{m+1}\right) C_n(k) & \text{for } |k| \le m \\ 0 & \text{for } |k| > m . \end{cases}$$

The weight generator $w(\cdot)$, in addition to the properties previously cited, is to have a transform in Eq. (5.5) which is positive at every point of the real line. Then $S_{n,m}(\omega)$ is nonnegative on $(0, \pi)$, because it can be written as a (scaled) convolution of $W(\cdot)$ with the positive-valued periodogram of the data $z_1, z_2, \ldots, z_n$. Fix (large) values of n and m. Then

$$\Psi_0(\omega) = S_{n,m}(\omega) = \frac{1}{2\pi} \sum_{|k| \le m} \psi_0(k) \cos k\omega \qquad (5.17)$$

has an absolute maximum at some point θ^*:

$$\Psi_0(\omega) < \Psi_0(\theta^*)$$

for all $\omega \ne \theta^*$ (the probability being 0 that there will be two or more such points). We now iteratively generate Fourier coefficients $\psi_j(\cdot)$ for $j = 1, 2, \ldots, J$ by means of

$$\psi_j(k) = \frac{\sum_{|h| \le 2^{j-1}m} \psi_{j-1}(h)\, \psi_{j-1}(h+k)}{\sum_{|h| \le 2^{j-1}m} \psi_{j-1}^2(h)} \qquad (|k| \le 2^j m). \qquad (5.18)$$

For each j, this is even in k and vanishes for all integer lags $|k| > 2^j m$. Furthermore, the $\psi_j(\cdot)$ are the Fourier coefficients of a nonnegative function

$$\Psi_j(\omega) = \frac{1}{2\pi} \sum_{|k| \le 2^j m} \psi_j(k) \cos k\omega \quad .$$

Using the convolution formula, it is not difficult to see that

$$\Psi_J(\omega) = \frac{\Psi_0^{2^J}(\omega)}{\int_{-\pi}^{\pi} \Psi_0^{2^J}(\lambda) d\lambda} \quad , \tag{5.19}$$

which is valid for any ω. Dividing both sides by $\Psi_J(\theta^*)$, we see that the ratio tends to 0 as $J \to \infty$, unless $\omega = \theta^*$, in which case the limit is 1. Therefore

$$\lim_{J \to \infty} \psi_J(k) = \cos k\theta^* \quad . \tag{5.20}$$

The iteration in Eq. (5.18) can provide an arithmetic scheme for approximating the mode of the estimate in Eq. (5.3), which does not entail its calculation over a grid of frequencies. One way to devise a method for "extracting" an approximation to θ^* from the ψ_J's, when J is finite, is to consider the expression

$$\psi_J(k) = a_J(k) \cos k\theta^* + b_J(k)$$

for $k = 0, 1, \ldots, K$. In view of Eq. (5.20), coefficient sequences (which of course depend on Eq. (5.17)) can be found such that

$$\lim_J b_J(k) = 1 - \lim_J a_J(k) = 0 \quad .$$

For a proper choice of $K = o(2^J)$ the approach can be made uniform in the lag variable k. Suppose, further, that we are given a real number $\ell \in (0, 1)$, which also depends on

$\Psi_0(\cdot)$ and J but not on θ^*. Now let ν be the number of "ℓ-threshold axis crossings" in the series

$$\psi_J(0) = 1, \quad \psi_J(1), \quad \psi_J(2), \ldots, \psi_J(K) \; .$$

That is to say, ν is the number of times the series goes from a value $> \ell$ (resp. $< \ell$) to a value $< \ell$ (resp. $> \ell$). The ratio

$$\theta^*_J = \frac{\pi\nu}{K} \tag{5.21}$$

in $[0, \pi]$ then converges to θ^* as $J \to \infty$, for an appropriate definition of ℓ.

We can estimate θ with Eq. (5.21) because the random variable $\theta^* = \theta^*_{n,m}$ is a consistent estimate of θ as $n \to \infty$. In view of

$$|\theta^*_J - \theta| \leq |\theta^*_J - \theta^*| + |\theta^* - \theta| \quad ,$$

the problem centers around properly relating the number of iterations J and the sample size n in order to (statistically) balance the rates at which the bounds go to 0. The design formulae given in Ref. [4] are not applicable because they are derived under the assumption that $\lim_n \mathcal{E} S_{n,m}(\omega)$ is sufficiently differentiable. If computation time is not a serious consideration, it appears worthwhile investigating the above method for the line case in Eq. (5.13).

ACKNOWLEDGMENT

The author wishes to thank Mrs. Margaret G. Gardner of MIT Lincoln Laboratory for doing the numerical calculations.

REFERENCES

[1] A. Albert, L.A. Gardner, Jr., "Stochastic Approximation and Nonlinear Regression." To be published in the Institute of Mathematical Statistics - University of Chicago Monograph Series. (Available as MIT Lincoln Laboratory preprint JA-2430, August 1964.)

[2] H. Cramér, Mathematical Methods of Statistics. Princeton University Press, New Jersey, (1946).

[3] W. Feller, An Introduction to Probability Theory and its Applications, 1, 2nd edition. Wiley, New York (1957).

[4] L. A. Gardner, Jr., "A Consistent Time-Domain Procedure for Estimating the Mode of an Unknown Multi-peak Power Spectrum." To appear in the Journal of the Society for Industrial and Applied Mathematics. (Available as MIT Lincoln Laboratory preprint, JA-2248, December 1963.)

[5] F. B. Hildebrand, Introduction to Numerical Analysis, McGraw-Hill, New York, (1956).

[6] K. Knopp, Theory and Applications of Infinite Series, Hafner, New York, (1947).

[7] E. Parzen, Modern Probability Theory and its Applications, Wiley, New York, (1960).

[8] E. Parzen, "An Approach to Time Series Analysis." Annals of Mathematical Statistics, 32, pp. 951-989, (1961).

[9] E. Parzen, "On Statistical Spectral Analysis." Stochastic Processes in Mathematical Physics and Engineering. Symposia of the American Mathematical Society, 16, pp. 221-246, (1964).

DISTRIBUTION LIST

Division 2

F.C.Frick
S.H.Dodd
J.T.Huntington

Group 21

H.L.Kasnitz
G.M.Shannon
C.W.Uskavitch

Group 22

V.A.Nedzel
W.Z.Lemnios
W.I.Wells
A.Freed
A.A.Grometstein
J.H.Hartogensis
C.E.Nielsen,Jr.
J.Rheinstein
A.F.Smith
P.A.Willmann

Group 28

J.A.Arnow
J.F.Nolan
A.W.Armenti
C.R.Arnold
T.C.Bartee
R.N.Davis
P.L.Falb
L.A.Gardner,Jr. (15)
P.L.Grant
H.K.Knudsen
O.A.Z.Leneman
J.B.Lewis
H.E.Meily
A.J.Morency
H.C.Peterson
F.C.Schweppe(5)

Group 28(Con't.)

J.M.Winett
P.E.Wood,Jr.

Division 3

W.R.Romaine

Group 35

M.Balser
L.R.Martin
J.Penhune
K.E.Ralston
A.R.Warwas

Division 4

J.Freedman
H.G.Weiss

Group 41

M.Axelbank
J.P.Hawkes
H.M.Jones
H.O.Schneider

Group 42

D.A.Cahlander

Group 45

W.W.Ward
J.P.Perry
A.Bertolini
S.F.Catalano
D.F.DeLong
H.W.Eklund
R.W.Straw

Group 62

F.E.Heart
K.Jordan
I.Stiglitz

Group 64

P.E.Green
H.W.Briscoe
R.Price

WSMR

S.J.Miller
G.R.Armstrong
R.L.Hardy
C.J.McKenney

Kwajalein

T.H.Einstein
E.Korngold
B.Nanni
R.P.Wishner

Hq.ESD
(Maj. S.Jeffery ESRC)
L.G.Hanscom Field
Bedford, Mass.

Hq.ESD (Lt.J.Egan)
L.G.Hanscom Field
Bedford, Mass.

LAG:smm
11 March 1965

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA – R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Lincoln Laboratory, M.I.T.	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP
3. REPORT TITLE Computationally Feasible Frequency Estimates in the Presence of Unknown Phase and Amplitude	
4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)* Technical Note	
5. AUTHOR(S) *(Last name, first name, initial)* Gardner, L.A., Jr.	

6. REPORT DATE 11 March 1965	7a. TOTAL NO. OF PAGES 78	7b. NO. OF REFS 9
8a. CONTRACT OR GRANT NO. AF 19(628)-500 b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Note 1965-4 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* ESD-TDR-65-58	
10. AVAILABILITY/LIMITATION NOTICES		
11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Air Force Systems Command, USAF	

13. ABSTRACT

The problem of estimating the angular frequency in a single trigonometric regression function, observed in the presence of correlated noise, can be approached by the method of Least Squares or procedures based on the theory of empirical spectral analysis. For real time applications, such methods generally are prohibitively time consuming.

We introduce two very simple and consistent frequency estimates which use at most the first three sample covariances, and derive and compare their large sample distributions. One interesting by-product of our calculations is a precise analysis of the asymptotic behavior of the (inconsistent) technique known in numerical analysis as Prony's method.

14. KEY WORDS

frequency measurement	stochastic approximation	random variables
spectral analysis	Gaussian processes	sequences
Prony's method	signal-to-noise ratio	Fourier transform
re-entry physics	time series	

UNCLASSIFIED
Security Classification