Please cite as: Mackisack, M. S., Osborne, M. R., and Smyth, G. K. (1994). A modified Prony algorithm for estimating sinusoidal frequencies. *J. Statist. Comput. Simul.* 49, 111-124.

A MODIFIED PRONY ALGORITHM FOR ESTIMATING SINUSOIDAL FREQUENCIES

M. S. Mackisack
Department of Mathematics
Queensland University of Technology

M. R. Osborne
Statistics Research Section, School of Mathematical Sciences
Australian National University

G. K. Smyth
Department of Mathematics
University of Queensland

Abstract. We investigate the modified Prony algorithm of Osborne and Smyth (1991, 1993) applied to the estimation of the frequency of oscillation of a sinusoidal signal with added white noise, and compare it with the algorithms usually associated with the names of Prony and Pisarenko. We show that the Pisarenko algorithm returns $O(N^{-\frac{1}{2}})$ consistent estimates, and that the modified Prony algorithm returns the least squares estimates which are $O(N^{-\frac{3}{2}})$ consistent. The transfer function parameter estimates that are returned by these algorithms converge to the true values at the same rate as the frequency estimates in each case. A concise expressions for the asymptotic covariance matrix of these estimates is given in the simplest case. Small sample properties, starting values and the importance of correctly specifying the number of sinusoids, are investigated in a simulation study.

Keywords. Prony's method; Pisarenko harmonic decomposition; frequency estimation; efficiency

1. Introduction

One of the recurring problems of time series analysis has proved to be the estimation of the frequencies present in a sinusoidal signal corrupted by the addition of white noise. This problem has spawned a variety of procedures arising from the various ways in which such a process may be represented. Suppose we observe a sequence $\{y(n)\}$, $n = 1, \ldots, N$, where

$$y(n) = \sum_{j=1}^{p} \rho_j \cos(\omega_j n + \phi_j) + \epsilon(n), \tag{1}$$

$\{\epsilon(n)\}$ is a sequence of uncorrelated random variables with zero mean and variance σ^2, and $\{\rho_j\}$ and $\{\omega_j\}$ are constants. Assuming p is known and $\{\phi_j\}$ non-random, the maximum likelihood estimates for ω_j may be obtained by non-linear least squares procedures from (1) or from the equivalent parametrisation $y(n) = \sum_{j=1}^{p} \alpha_j \cos \omega_j n + \beta_j \sin \omega_j n + \epsilon(n)$ (Hannan 1971, Walker 1971), or asymptotically from the location of maxima of the periodogram (Bloomfield 1976, p.18). These procedures include an optimisation step which requires initial estimates of the frequencies, and return estimated frequencies which are $O(N^{-\frac{3}{2}})$ consistent, while the estimates of the linear parameters (ρ, or α, β) are $O(N^{-\frac{1}{2}})$ consistent (Hannan 1973).

Other methods make use of the recurrence relation formulation of (1), derived in, for example, Feller (1971, p.93) or Chan, Lavoie and Plant (1981):

$$\sum_{j=0}^{2p} c(j) y(n-j) = \sum_{j=0}^{2p} c(j) \epsilon(n-j), \quad n = 2p+1, \ldots, N \tag{2}$$

where $c(0) = c(2p) = 1$, $c(j) = c(2p-j)$ for $j = 0, \ldots, 2p$. The relation (2) is formally equivalent to saying that a linear combination of p sinusoidal signals (or $2p$ pure imaginary complex conjugate signals) can be modelled as an ARMA(2p,2p) process. Since the coefficients of both parts of the transfer function are identical, so that the transfer function reduces to 1, this might be thought to be vacuous, but a variety of procedures have been evolved based on estimating the parameters $\mathbf{c} = \{c(0), \ldots, c(2p)\}^T$. The frequency estimates $\hat{\omega}_j$ are then found as phase components of the zeroes of the polynomial $\hat{c}(z) = \sum_{j=0}^{2p} \hat{c}(j) z^j$, the estimate of the transfer function of the recurrence relation (2).

In the following section we develop the Prony and Pisarenko algorithms, based on (2), as constrained optimisation procedures for the same objective function, and prove that the parameter estimates derived from the Pisarenko are $O(N^{-\frac{1}{2}})$ consistent. In section 3 we use the representations developed previously to derive an alternative algorithm for estimating $\mathbf{c}$ and show that the frequency estimates found in this way are the maximum likelihood estimates. In section 4 we prove that the $\mathbf{c}$ estimates are $O(N^{-\frac{3}{2}})$ consistent and derive a concise expression for the asymptotic covariance matrix in the simplest case. We display some results of simulation experiments in section 5, and defer proofs of theorems until section 6.

2. Algorithms of Pisarenko and Prony.

In order to present the Prony and Pisarenko procedures in a concise manner we introduce the following matrix notations for the set of equations (2). Let $\mathbf{y}$ and $\mathbf{e}$ be the vectors $\{y(1),\ldots,y(N)\}^T$ and $\{\epsilon(1),\ldots,\epsilon(N)\}^T$, and define $\mathbf{C}$ and $\mathbf{Y}$ to be the $N\times(N-2p)$ and $(N-2p)\times(2p+1)$ matrices

$$\mathbf{C}=\begin{pmatrix} c(2p) & 0 & \ldots & \ldots & 0\\ \vdots & \ddots & \ddots & \ldots & 0\\ c(0) & & c(2p) & 0 & \vdots\\ 0 & \ddots & & \ddots & \ddots\\ \vdots & 0 & c(0) & & c(2p)\\ \vdots & \ldots & \ddots & \ddots & \vdots\\ \mathbf{0} & \ldots & \ldots & 0 & c(0) \end{pmatrix},$$

$$\mathbf{Y}=\begin{pmatrix} y(2p+1) & \ldots & y(1)\\ \vdots & \ddots & \vdots\\ y(N) & \ldots & y(N-2p) \end{pmatrix},$$

and let $\mathbf{E}$ be a matrix of the same form as $\mathbf{Y}$ with $\epsilon(j)$ replacing $y(j)$ throughout. Then we can write (2) in either of the two equivalent forms

$$\mathbf{Yc}=\mathbf{Ec} \tag{3}$$

or

$$\mathbf{y}^T\mathbf{C}=\mathbf{e}^T\mathbf{C}\,. \tag{4}$$

Write $\mathbf{R}$ for the $(2p+1)\times(2p+1)$ matrix of expected squares and products

$$\mathbf{R}=E\{(\mathbf{Y}^T\mathbf{Y})/(N-2p)\}\,,$$

and observe that $E(\mathbf{Y}^T\mathbf{E})=(N-2p)\sigma^2\mathbf{I}_{2p+1}$. In the signal processing literature, $\mathbf{R}$ is called the autocovariance matrix of $\mathbf{y}$, and this relates to an alternative model for the data in which the phases ϕ_j are uniformly distributed on $[0,2\pi)$. If we premultiply both sides of (3) by $\mathbf{Y}^T$ and take expectations we find that $(N-2p)\mathbf{Rc}=(N-2p)\sigma^2\mathbf{I}_{2p+1}\mathbf{c}$, hence $(\mathbf{R}-\sigma^2\mathbf{I})\mathbf{c}=0$. That is, $(\mathbf{R}-\sigma^2\mathbf{I})$ has an eigenvalue at 0, whose eigenvector is the parameter vector $\mathbf{c}$. Estimating $\mathbf{R}$ by some appropriate estimators of the autocovariances, finding σ^2 as the smallest eigenvalue and $\mathbf{c}$ as its eigenvector is the algorithm of Pisarenko. The procedure concludes by finding the phase components of the zeroes of the polynomial $c(z)$.

The right hand sides of (2), (3) and (4) represent a noise sequence with a complicated correlation structure. The fact that $c(0)=1$ enables a third form of the estimating equations to be written,

$$\mathbf{y}_*=-\mathbf{Y}_*\mathbf{c}_*+\mathbf{f}, \tag{5}$$

where $\mathbf{Y}_*$ is the matrix consisting of the last $2p$ columns of $\mathbf{Y}$, $\mathbf{c}_*$ is the vector $\{c(1), \ldots, c(2p)\}$, and $\mathbf{y}_*$ is the first column of $\mathbf{Y}$, so that $\mathbf{Y} = [\mathbf{y}_* \vdots \mathbf{Y}_*]$, and $\mathbf{f} = \mathbf{E}\mathbf{c}$. Ignoring the correlation structure of $\mathbf{f}$ and solving (5) as if it were an ordinary least squares problem (minimising $\mathbf{f}^T\mathbf{f}$) gives

$$\hat{\mathbf{c}}_* = -(\mathbf{Y}_*^T\mathbf{Y}_*)^{-1}\mathbf{Y}_*^T\mathbf{y}$$

which is one member of a class of algorithms known by the name of Prony. There are a number of variants of this algorithm, which attempt to reduce the inefficiency that arises from ignoring the correlation structure of the $\mathbf{f}$ sequence, and to incorporate information about the symmetric structure of the $\mathbf{c}$ parameters (see Chan, Lavoie and Plant 1981, Marple 1987).

Expressed as an optimisation problem, Pisarenko's algorithm finds $\mathbf{c}$ by minimising $\mathbf{c}^T\widehat{\mathbf{R}}\mathbf{c}$ subject to $\mathbf{c}^T\mathbf{c} = 1$ (Golub and van Loan 1983 Theorem 8.1–2), where $\widehat{\mathbf{R}}$ is some estimate of $\mathbf{R}$. With the particular choice of $\widehat{\mathbf{R}} = \mathbf{Y}^T\mathbf{Y}/(N-2p)$, the alternative constraint $c(0) = c(2p) = 1$ leads to Prony's algorithm, as follows. Letting λ be a Lagrange multiplier, the normal equations are $\mathbf{Y}^T\mathbf{Y}\mathbf{c} = \lambda\mathbf{e}_1$ where $\mathbf{e}_1$ is the first coordinate vector. Putting $c(0) = 1$, the last $2p$ rows give

$$\mathbf{Y}_*^T\mathbf{y}_* + \mathbf{Y}_*^T\mathbf{Y}_*\mathbf{c}_* = 0$$

that is

$$\mathbf{c}_* = -(\mathbf{Y}_*^T\mathbf{Y}_*)^{-1}\mathbf{Y}_*^T\mathbf{y}$$

as above. Kahn, Mackisack, Osborne and Smyth (1991) show that $\mathbf{c}^T\mathbf{c} = 1$ is the only scaling of this optimisation problem which leads to consistent estimates for $\mathbf{c}$.

The error in the estimates of the elements of $\hat{\mathbf{c}}$ in the Pisarenko algorithm is of the same order of magnitude as the error in the estimates of the autocovariances, so that none of the parameter estimates can be consistent to order greater than $N^{-\frac{1}{2}}$. To see this, note that the estimators $\hat{\mathbf{c}}$ and $\hat{\sigma}^2$ satisfy $\psi(\widehat{\mathbf{R}}; \hat{\mathbf{c}}, \hat{\sigma}^2) = 0$ where $\psi(\widehat{\mathbf{R}}; \mathbf{c}, \sigma^2)^T = [c^T(\widehat{\mathbf{R}} - \sigma^2 I) \;\; (1 - \mathbf{c}^T\mathbf{c})/2]$. The asymptotic covariance matrix of $\hat{\mathbf{c}}$ and $\hat{\sigma}^2$ is $[E(\psi')]^{-1}\mathrm{var}(\psi)[E(\psi')]^{-1}$ where $\psi' = [\partial\psi/\partial\mathbf{c}, \; \partial\psi/\partial\sigma^2]$ and ψ and ψ' are evaluated at $\widehat{\mathbf{R}}$ and the true $\mathbf{c}$ and σ^2. Since $E(\psi')$ is a constant nonsingular matrix, the covariance matrix of $\hat{\mathbf{c}}$ and $\hat{\sigma}^2$ is proportional to $\mathrm{var}(\psi) = \mathrm{var}(\widehat{\mathbf{R}}\mathbf{c})$. This suggests that an improved algorithm in respect of convergence properties will need to in some way avoid explicitly estimating the autocovariances, and in particular will need to use a matrix that is of smaller order than $\widehat{\mathbf{R}}$ in the direction of $\mathbf{c}$.

3. The Modified Prony Algorithm

We derive an algorithm which includes specific treatment of the dependence in the error sequence and which avoids the inefficiency of the Pisarenko algorithm, starting from the representation (4) for the data, and following the arguments of Osborne and Smyth (1991, 1993). The maximum likelihood estimates for $\{\rho_j\}$, $\{\phi_j\}$ and $\{\omega_j\}$, or equivalently $\{\alpha_j\}$, $\{\beta_j\}$ and $\{\omega_j\}$, minimise the sum of squares $S = (\mathbf{y} - \boldsymbol{\mu})^T(\mathbf{y} - \boldsymbol{\mu})$, where $\boldsymbol{\mu} = (\mu(1), \ldots, \mu(N))^T$ is the vector of signals $\mu(n) =$

$\sum_{j=1}^{p} \alpha_j \cos \omega_j n + \beta_j \sin \omega_j n$. We can write $\boldsymbol{\mu} = \mathbf{A}(\boldsymbol{\omega})\boldsymbol{\gamma}$ where $\mathbf{A}(\boldsymbol{\omega})$ is the $N \times 2p$ matrix

$$\mathbf{A}(\boldsymbol{\omega}) = \begin{pmatrix} \cos\omega_1 1 & \sin\omega_1 1 & \dots & \cos\omega_p 1 & \sin\omega_p 1 \\ \vdots & \vdots & & \vdots & \vdots \\ \cos\omega_1 N & \sin\omega_1 N & \dots & \cos\omega_p N & \sin\omega_p N \end{pmatrix}$$

and $\boldsymbol{\gamma} = (\alpha_1, \beta_1, \dots, \alpha_p, \beta_p)^T$; then S is minimised for fixed $\boldsymbol{\omega}$ by $\widehat{\boldsymbol{\gamma}}(\boldsymbol{\omega}) = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}$. Substituting $\widehat{\boldsymbol{\gamma}}(\boldsymbol{\omega})$ into S gives the reduced sum of squares

$$S(\boldsymbol{\omega}) = S\big(\boldsymbol{\omega}, \widehat{\boldsymbol{\gamma}}(\boldsymbol{\omega})\big) = \mathbf{y}^T(\mathbf{I} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)\mathbf{y}\,.$$

Now $\mathbf{I} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is the orthogonal projection onto the null space of $\mathbf{A}$, and since $\mathbf{C}^T\mathbf{A} = 0$ and $\mathbf{C}$ and $\mathbf{A}$ have full column rank, this projection can be written also as $\mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T$. We have shown that

$$S(\boldsymbol{\omega}) = S(\mathbf{c}) = \mathbf{y}^T\mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{y}\,,$$

and minimising $S(\mathbf{c})$ gives the maximum likelihood estimator of $\mathbf{c}$ (in the case when $\{\epsilon(n)\}$ are independent and identically distributed as $\mathcal{N}(0, \sigma^2)$).

The frequency estimates derived from $\hat{c}(z)$ are the maximum likelihood estimates (Osborne and Smyth 1993 Theorem 1), and have the asymptotic distribution given in, for example, Hannan (1971, Theorem 4′); for $p = 1$, that is, for $y(n) = \rho\cos(\omega n + \phi) + \epsilon(n)$, $N^{-\frac{3}{2}}(\hat{\omega} - \omega) \to \mathcal{N}(0, 2\pi f(\omega)(24/\rho^2))$ where $f(\omega)$ is the noise process spectral density at the true frequency.

The derivative of the objective function $S(\mathbf{c})$ is found (using results in, for example, Dwyer and McPhail 1948 pp. 517–534) to be

$$\begin{aligned} \frac{ds}{d\mathbf{c}} &= 2\mathbf{B}(\mathbf{c})\mathbf{c} \\ &= 2\mathbf{y}^T\frac{\partial\mathbf{C}}{\partial\mathbf{c}}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{y} - 2\mathbf{y}^T\mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\frac{\partial\mathbf{C}^T}{\partial\mathbf{c}}\mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{y} \\ &= 2\left[\sum_{j=0}^{2p} c(j)\mathbf{B}_{ij}\right]_{i=0,\dots,2p}. \end{aligned}$$

Here, $\partial\mathbf{C}/\partial\mathbf{c}$ is a three-dimensional array, so $\partial\mathbf{C}/\partial c(i)$ and $\partial\mathbf{C}/\partial c(j)$ are matrices, and

$$\mathbf{B}_{ij} = \mathbf{y}^T\frac{\partial\mathbf{C}}{\partial c(i)}(\mathbf{C}^T\mathbf{C})^{-1}\frac{\partial\mathbf{C}^T}{\partial c(j)}\mathbf{y} - \mathbf{y}^T\mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\frac{\partial\mathbf{C}^T}{\partial c(i)}\frac{\partial\mathbf{C}}{\partial c(j)}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{y} \quad (7)$$

(Osborne and Smyth 1991) so as in that paper the necessary condition for a minimum of the least squares problem is $\mathbf{B}(\mathbf{c})\mathbf{c} = \lambda\mathbf{c}$, where λ is a Lagrange multiplier for the scaling constraint. This is a nonlinear eigenproblem, and the algorithm which solves it is called here the modified Prony algorithm. It starts by forming the matrix $\mathbf{B}$ using an initial estimate of $\mathbf{c}$, and proceeds by inverse iteration to

locate the $\mathbf{c}$ which is the eigenvector corresponding to the zero eigenvalue of $\mathbf{B}$ (see Osborne and Smyth 1991 for details).

4. Asymptotic variances of the Prony parameter estimators

In this section we seek an explicit expression for the covariance matrix of the transfer function parameters $\hat{\mathbf{c}}$. The $\mathbf{c}$ are of interest independently of $\boldsymbol{\omega}$ as the parameters in the transfer function and because the estimated signal values $\hat{\boldsymbol{\mu}}$ can be obtained directly from them without rooting the polynomial $\hat{c}(z)$ (Osborne and Smyth 1993). There is also reason to expect that an expression for the asymptotic covariance of $\hat{\mathbf{c}}$ may remain valid for smaller N and lower signal-to-noise ratio than the corresponding expression for $\hat{\boldsymbol{\omega}}$. This is because, for small N or low signal-to-noise ratios, the estimated polynomial may yield roots which are partly or entirely real. When the roots are real, the corresponding frequency estimates are meaningless, but the signal obtained directly from $\hat{\mathbf{c}}$ is still meaningful as estimating the solution of a homogeneous differential equation, as described in Osborne and Smyth (1991).

Estimation of the covariance matrix of $\hat{\mathbf{c}}$ is based on the matrix $\mathbf{B}$. For the case of real exponential signals considered by Osborne and Smyth (1991 Theorem 3), $E(\mathbf{B}(\mathbf{c}))$ cannot be expressed more explicitly than as a matrix whose terms are products like the first term of $\mathbf{B}_{ij}$ in (7); they prove (Osborne and Smyth 1993 Theorem 2) that $N^{-1}\mathbf{B}(\hat{\mathbf{c}})$ converges almost surely to a positive semi-definite limit, when the asymptotic sequence is taken in the sense of progressively finer dissections of the same finite interval. They also show that

$$E\{\frac{\partial^2 S}{\partial \mathbf{c}\partial \mathbf{c}^T} - \mathbf{B}(\mathbf{c})\} = 0.$$

The methods used can be adapted to the sinusoidal case to prove the following lemma.

Lemma. *Let* $\Omega = \lim_{N\to\infty} \frac{1}{N^3} E\{\mathbf{B}(\mathbf{c})\}$. *Then*

$$\lim_{N\to\infty} \frac{1}{N^3}\mathbf{B}(\mathbf{c}) = \Omega$$

$$\text{and } \lim_{N\to\infty} \frac{1}{N^3}\left(\frac{\partial^2 S}{\partial \mathbf{c}\partial \mathbf{c}^T} - \mathbf{B}(\mathbf{c})\right) = 0\,.$$

Theorem 1. *Let* $\hat{\mathbf{c}}$ *minimize* $S(\mathbf{c})$ *subject to* $\mathbf{c}^T\mathbf{c} = 1$. *Then* Ω^+ *is the asymptotic covariance matrix of* $\sigma^{-1}N^{-\frac{3}{2}}\hat{\mathbf{c}}$.

Theorem 2. *If* $p = 1$ *then* $\Omega = (12\sin^2\omega)^{-1}(\mathbf{R} - \sigma^2\mathbf{I})$.

Theorem 2 gives an expression for Ω in the case of a single sinusoid, $p = 1$ in (1), where it is possible to evaluate it explicitly in terms of the autocovariance matrix

$$\mathbf{R} = \frac{\rho^2}{2}\begin{pmatrix} 1+\sigma^2 & \cos\omega & \cos 2\omega \\ \cos\omega & 1+\sigma^2 & \cos\omega \\ \cos 2\omega & \cos\omega & 1+\sigma^2 \end{pmatrix}.$$

Proofs of both theorems are given in Section 6. We conjecture that the relationship between Ω and $\mathbf{R}$ is true for larger p, but the algebraic devices that enable Theorem 2 to be derived are not obviously generalised.

For $p = 1$ the matrix $(2/\rho^2)(\mathbf{R} - \sigma^2\mathbf{I})$ has eigenvectors $(1, -2\cos\omega, 1)^T$, $(1, 0, -1)^T$ and $(\cos\omega, 1, \cos\omega)^T$ with eigenvalues 0, $2\sin^2\omega$ and $2\cos^2\omega + 1$ respectively. This allows us to obtain

$$\Omega^+ = \frac{24a^2}{\rho^2(2b^2+1)^2} \begin{pmatrix} (8b^2+1)/(4a^2) & b & (8b^2-1)/(4a^2) \\ b & 1 & b \\ (8b^2-1)/(4a^2) & b & (8b^2-1)/(4a^2) \end{pmatrix}$$

where $a = \sin\omega$ and $b = \cos\omega$. Alternatively, if we prefer to rescale $\hat{\mathbf{c}}$ so that $c(3) = 1$, then we can find the asymptotic covariance matrix of $\sigma^{-1}N^{-\frac{3}{2}}\{\hat{c}(1)/\hat{c}(3), \hat{c}(2)/\hat{c}(3)\}^T$ as $\mathbf{D}^T\Omega^+\mathbf{D}$ where

$$\mathbf{D}^T = \begin{pmatrix} 1/c(3) & 0 & -c(1)/c(3)^2 \\ 0 & 1/c(3) & -c(2)/c(3)^2 \end{pmatrix}$$

and $\mathbf{D}$ is evaluated at the asymptotic values $\mathbf{c} = (1, -2\cos\omega, 1)^T/\sqrt{2 + 4\cos^2\omega}$. After some algebra this gives

$$\mathbf{D}^T\Omega^+\mathbf{D} = \frac{48(2\cos^2\omega + 1)}{\rho^2} \begin{pmatrix} 1 & -\cos\omega \\ -\cos\omega & 1 \end{pmatrix}.$$

For $p > 1$ the algebra is too complicated to write down concisely the explicit form of Ω, but we conjecture that it remains of the form $k(\mathbf{R} - \sigma^2\mathbf{I}_{2p+1})$ for larger p, with some constant k depending on ω and p.

5. Simulation Experiments.

There are several issues to be considered in connection with the performance of the algorithm on finite data sets: one would like to know how large N needs to be for the asymptotic rate of convergence to become evident; whether the algorithm is sensitive to choice of starting values and if so whether a method for selecting starting values can be recommended; how the algorithm behaves if the assumption that the number of signals p is known correctly is violated; and whether the conjecture about the form of $\lim_{N\to\infty} N^{-3}\mathbf{B}(\hat{\mathbf{c}})$ seems to be supported by computational results, for $p > 1$. An exhaustive investigation of these questions is beyond the scope of this paper, which has a primarily theoretical intention. We present a selection of simulation results which we believe indicate the important features of the algorithm's small-sample performance.

Simulated data sequences were constructed with one and two sinusoids,

$$(A) \qquad y(n) = 20\cos(0.34n + \phi) + \epsilon(n)$$

$$(B) \qquad y(n) = 20\cos(0.34n + \phi_1) + 20\cos(1.3n + \phi_2) + \epsilon(n)$$

The $\epsilon(n)$ were generated as an independent, standard Gaussian sequence using algorithm AS70 of Odeh and Evans (1974). The modified Prony algorithm was

implemented as described in Osborne and Smyth (1991, 1993), and all calculations were performed in double precision arithmetic.

Estimation of the frequency ($\omega = 0.34$) and the transfer function parameter ($c(1) = -2\cos(\omega) = -1.885509$) in (A) was very satisfactory even with N as small as 16, when the true frequency was used as the starting value for the algorithm and with the variance of the noise sequence $\sigma^2 = 1$. The mean and standard deviation from 100 replications of the simulation with $N = 16$ were $\mu_{\hat{\omega}} = 0.3405$ and $\sigma_{\hat{\omega}} = 0.0037$ respectively. When the variance of the noise sequence was increased to $\sigma^2 = 100$, the estimates deteriorated rapidly for $N = 16$ ($\mu_{\hat{\omega}} = 0.3233$, $\sigma_{\hat{\omega}} = 0.1056$) although they were still satisfactory for $N = 32$ ($\mu_{\hat{\omega}} = 0.349$, $\sigma_{\hat{\omega}} = 0.071$). The values predicted by the asymptotics for the three $\sigma_{\hat{\omega}}$ given above are 0.0054, 0.054 and 0.019 respectively. For $N \geq 64$ the agreement between predicted and actual standard deviations was within sampling error.

The importance of an adequate starting value for the usual Maximum Likelihood frequency estimation procedures has been demonstrated by Rice and Rosenblatt (1988), and the same need can be seen in the case of the modified Prony algorithm. For a single short data sequence and a single long one ($N = 32$ and $N = 512$), we examined empirically how far away from the true value a starting value could be chosen which would still lead to convergence. We ran the algorithm repeatedly on this sequence for starting values in an interval around the true value, separated by steps of 0.0006. For $N = 32$, convergence occurred reliably for starting values between 0.2664 and 0.4254; outside this range, convergence still occurred for isolated sets of values as far away from the true value as 0.2490 and 0.4548, and then beyond this region convergence apparently never occurred. As N increases, the starting value needs to be closer to the true value, and for $N = 512$, the reliable range of starting values was between 0.3354 and 0.3450, and hardly any values outside this region led to convergence. The absence of a clear-cut boundary to the domain of attraction of the maximum likelihood solution is to be expected for such a non-linear algorithm.

There are several algorithms in common use for frequency estimation which do not require starting values (unmodified Prony, Pisarenko) or which are much less sensitive to their choice (autoregressive spectral estimation), and it would clearly be an advantage if such an algorithm could provide an adequate starting value for the modified Prony algorithm. However, since we know that the approximate likelihood has maxima which are $O(N^{-1})$ apart (Rice and Rosenblatt 1988), we probably need a method better than that of Prony or Pisarenko which has estimation errors $O(N^{-\frac{1}{2}})$. We may expect that these will be unreliable, whereas the autoregressive method which has estimation error $O(N^{-\frac{5}{4}})$ (Mackisack and Poskitt 1989) is a more promising candidate. The autoregressive method has the added advantage that it does not require prior input of the number of sinusoidal frequencies in the signal for a sensible estimate to emerge, whereas the Pisarenko and Prony methods depend critically on correct model specification.

A study of 20 input sequences of type (A) with $N = 32$ confirmed that the (inconsistent) Prony method was not useful, particularly at high noise levels, giv-

ing estimates consistently far from the domain of attraction of the maximum of the likelihood surface. In the same circumstances the (consistent) Pisarenko algorithm returned an acceptable starting value about half the time, while the autoregressive procedure was always close to the true value and well within the domain of attraction established earlier. The autoregressive method was implemented using Burg's algorithm and choosing the order by the AIC criterion, as described in Mackisack and Poskitt (1989).

When $p > 1$, care must be taken not only to have starting values close to the true frequencies, but also to ensure that the value of p is correct. When these conditions are satisfied the modified Prony algorithm again performs well, as expected, but if p is mis-specified the results may be misleading, even when the true frequency values are given as starting values. Fitting a model with $p = 1$ to an input sequence of type (B) starting from one of the true frequency values, the algorithm performs as if there were only that one frequency in the signal. With $N = 32$ for example the frequency estimate agrees with that found using $p = 1$ to about two decimal places. Adding the extra frequency in the sequence biases the estimates slightly. If an input sequence of type (A) is used and $p = 2$ is fitted, as long as a starting value is given which is very close to the true (single) frequency, the algorithm will estimate that frequency, agreeing for $N = 32$ to about four decimal places with the estimate obtained for the same sequence using $p = 1$. Sometimes, the algorithm returns a second frequency quite close to the second starting value given (although on other occasions there is no relation between the two); if one actually believed that there ought to be a second signal component, this can give a misleading impression that the algorithm has in fact located such a frequency. There is clearly a need for a test to be carried out, such as for example that of Quinn (1986), to establish the correct value for p before proceeding with the frequency estimation.

We finally estimated the correlation matrix of the parameters $\mathbf{c}$ in model (B) to see whether the conjecture of the asymptotic form seems to be supported. $\mathbf{D}^T(\mathbf{R} - \sigma^2\mathbf{I})^+\mathbf{D}$ can be computed for $p = 2$ and the values from (B), following the same arguments as were used for the $p = 1$ case. Computing the correlation matrix derived from this removes the influence of the unknown constant, and gives

$$\begin{pmatrix} 1.0000 & -0.9531 & 0.9251 & -0.8705 \\ -0.9531 & 1.0000 & -0.9726 & 0.9251 \\ 0.9251 & -0.9726 & 1.0000 & -0.9531 \\ -0.8705 & 0.9251 & -0.9531 & 1.0000 \end{pmatrix}$$

compared with the observed correlation matrix of the estimated parameters, which is similar in pattern and approximate relative size of its elements:

$$\begin{pmatrix} 1.0000 & -0.9428 & 0.8348 & -0.7326 \\ -0.9428 & 1.0000 & -0.9467 & 0.8178 \\ 0.8348 & -0.9467 & 1.0000 & -0.9297 \\ -0.7326 & 0.8178 & -0.9297 & 1.0000 \end{pmatrix}.$$

Scaling $\mathbf{D}^T(\mathbf{R}-\sigma^2\mathbf{I})^+\mathbf{D}$ by dividing by the largest diagonal element gives

$$\begin{pmatrix} 0.3781 & -0.5861 & 0.5688 & -0.3292 \\ -0.5861 & 1.0000 & -0.9726 & 0.5688 \\ 0.5688 & -0.9726 & 1.0000 & -0.5861 \\ -0.3292 & 0.5688 & -0.5861 & 0.3781 \end{pmatrix}$$

which can be compared with the similarly scaled version of the covariance matrix of the sample:

$$\begin{pmatrix} 0.2607 & -0.4814 & 0.4090 & -0.1861 \\ -0.4814 & 1.0000 & -0.9085 & 0.4069 \\ 0.4090 & -0.9085 & 0.9207 & -0.4439 \\ -0.1861 & 0.4069 & -0.4439 & 0.2476 \end{pmatrix}$$

from which we see that the relative sizes of the variances of the parameter estimates also are within reasonable distance of each other (using an approximate χ^2 distribution for the simulation variance estimates, the coefficient of variation is about 14%). This all seems consistent with the idea that the covariance matrix is of the form conjectured, with some constant which we have not been able to identify.

6. Proofs

Proof of Theorem 1. This theorem is an extension of standard least squares results, which show that the second derivative matrix of the sum of squares, S, approximates the inverse covariance matrix of the estimates. Introducing a Lagrange multiplier, the least squares estimators minimize

$$F(\boldsymbol{\gamma}, \mathbf{c}, \lambda) = S(\boldsymbol{\gamma}, \mathbf{c}) + \lambda(\mathbf{c}^T\mathbf{c} - 1)\ ,$$

where $\boldsymbol{\gamma}$ is as in Section 3. Applying a Taylor series expansion to the derivative of F, using $\mathcal{D}$ for the derivative operator, leads to

$$0 = \mathcal{D}F(\hat{\boldsymbol{\gamma}}, \hat{\mathbf{c}}, \hat{\lambda}) \approx \mathcal{D}F(\boldsymbol{\gamma}, \mathbf{c}, \lambda) + \mathcal{D}^2F(\boldsymbol{\gamma}, \mathbf{c}, \lambda) \begin{pmatrix} \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \\ \hat{\mathbf{c}} - \mathbf{c} \\ \hat{\lambda} - \lambda \end{pmatrix}$$

and

$$\begin{pmatrix} \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \\ \hat{\mathbf{c}} - \mathbf{c} \\ \hat{\lambda} - \lambda \end{pmatrix} \approx -\mathcal{D}^2F(\boldsymbol{\gamma}, \mathbf{c}, \lambda)^{-1}\mathcal{D}F(\boldsymbol{\gamma}, \mathbf{c}, \lambda)\,. \tag{11}$$

Now

$$\mathcal{D}F = \begin{pmatrix} -(\mathcal{D}_{\boldsymbol{\gamma}}\mu)^T(\mathbf{y} - \boldsymbol{\mu}) \\ -(\mathcal{D}_{\mathbf{c}}\mu)^T(\mathbf{y} - \boldsymbol{\mu}) + \lambda\mathbf{c} \\ \mathbf{c}^T\mathbf{c} - 1 \end{pmatrix}$$

where $\mathcal{D}_{\boldsymbol{\gamma}}$ and $\mathcal{D}_{\mathbf{c}}$ are partial derivative operators, and since $\mathbf{c}^T(\mathcal{D}_{\mathbf{c}}\boldsymbol{\mu})^T = 0$ we see that $\lambda = 0$ at the solution. So

$$\mathcal{D}^2 F = \begin{pmatrix} \mathcal{D}^2_{\boldsymbol{\gamma}\boldsymbol{\gamma}^T} S & \mathcal{D}^2_{\boldsymbol{\gamma}\mathbf{c}^T} S & 0 \\ \mathcal{D}^2_{\mathbf{c}\boldsymbol{\gamma}^T} S & \mathcal{D}^2_{\mathbf{c}\mathbf{c}^T} S & \mathbf{c} \\ 0 & \mathbf{c}^T & 0 \end{pmatrix}.$$

The eigenvectors of $\mathcal{D}^2 F$ are $(0, \mathbf{c}^T, 1)$, $(0, \mathbf{c}^T, -1)$ and $(\mathbf{v}_j^T, 0)$, $j = 1, \ldots, 2p$, where $\mathbf{v}_j$ are the eigenvectors of $\mathcal{D}^2 S$ orthogonal to $\mathbf{c}$, and the first two eigenvalues are 1 and -1, so

$$(\mathcal{D}^2 F)^{-1} = \begin{pmatrix} (\mathcal{D}^2 S)^+ & \mathbf{g} \\ \mathbf{g}^T & 0 \end{pmatrix},$$

where $\mathbf{g}^T = (0, \mathbf{c})$ Therefore (11) becomes

$$\begin{pmatrix} \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \\ \hat{\mathbf{c}} - \mathbf{c} \\ 0 \end{pmatrix} = \begin{pmatrix} (\mathcal{D}^2 S)^+ \mathcal{D} S \\ 0 \end{pmatrix}$$

or equivalently

$$\begin{pmatrix} N^{-\frac{1}{2}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \\ N^{-\frac{3}{2}}(\hat{\mathbf{c}} - \mathbf{c}) \end{pmatrix} \approx \begin{pmatrix} N^{-1}\mathcal{D}^2_{\boldsymbol{\gamma}\boldsymbol{\gamma}^T} S & N^{-2}\mathcal{D}^2_{\boldsymbol{\gamma}\mathbf{c}^T} S \\ N^{-2}\mathcal{D}^2_{\mathbf{c}\boldsymbol{\gamma}^T} S & N^{-3}\mathcal{D}^2_{\mathbf{c}\mathbf{c}^T} S \end{pmatrix}^+ \begin{pmatrix} N^{-\frac{1}{2}}\mathcal{D}_{\boldsymbol{\gamma}} S \\ N^{-\frac{3}{2}}\mathcal{D}_{\mathbf{c}} S \end{pmatrix}.$$

Now $N^{-\frac{1}{2}}\mathcal{D}_{\boldsymbol{\gamma}} S$ and $N^{-\frac{3}{2}}\mathcal{D}_{\mathbf{c}} S$ have expectation zero and an asymptotic covariance matrix, $\mathbf{H}\sigma^2$ say, that has null space spanned by $(\mathbf{0}^T, \mathbf{c}^T)^T$. Also the matrix of standardized second derivatives of S can be shown to converge to $\mathbf{H}$ by the law of large numbers. Therefore asymptotically,

$$\operatorname{var}\begin{pmatrix} N^{-\frac{1}{2}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \\ N^{-\frac{3}{2}}(\hat{\mathbf{c}} - \mathbf{c}) \end{pmatrix} = \mathbf{H}^+ \sigma^2.$$

To complete the proof we note that

$$\mathcal{D}_{\mathbf{c}\mathbf{c}^T} S(\mathbf{c}, \hat{\boldsymbol{\gamma}}(\mathbf{c})) = \mathcal{D}^2_{\mathbf{c}\mathbf{c}^T} S - \mathcal{D}^2_{\mathbf{c}\boldsymbol{\gamma}^T} S (\mathcal{D}^2_{\boldsymbol{\gamma}\boldsymbol{\gamma}^T} S)^{-1} \mathcal{D}^2_{\boldsymbol{\gamma}\mathbf{c}^T} \big|_{\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}(\mathbf{c})}$$

(Richards, 1961), which is the Moore-Penrose inverse of the trailing $(p+1)\times(p+1)$ submatrix of $\mathcal{D}^2 S$. This shows that $(\lim_{N\to\infty} N^{-3}\mathcal{D}^2_{\mathbf{c}\mathbf{c}^T} S(\mathbf{c}, \hat{\boldsymbol{\gamma}}(\mathbf{c})))^+$ is the trailing $(p+1)\times(p+1)$ submatrix of H^+, but from the Lemma this is Ω^+ and the Theorem is proved.

Proof of Theorem 2. $E(\mathbf{y}\mathbf{y}^T)$ has as its (r, c) element $\sigma^2\delta_{rc} + \frac{\rho^2}{2}\cos(r-c)\omega + \frac{\rho^2}{2}\cos\{(r+c)\omega + 2\phi\} = \sigma^2\delta_{rc} + \mathbf{W}_{1,rc} + \mathbf{W}_{2,rc}$, where δ_{rc} is the Kronecker delta

function. We shall write $E(\mathbf{y}\mathbf{y}^T) - \sigma^2\mathbf{I} = \mathbf{W}_1 + \mathbf{W}_2$. To evaluate the expectation of (7), use (4) and trace identities to write

$$\begin{aligned} E(\mathbf{B}_{ij}) &= E\left[\mathrm{tr}\left\{ \frac{\partial \mathbf{C}}{\partial c(i)}(\mathbf{C}^T\mathbf{C})^{-1}\frac{\partial \mathbf{C}^T}{\partial c(j)}\mathbf{y}\mathbf{y}^T - \frac{\partial \mathbf{C}}{\partial c(j)}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{e}\mathbf{e}^T\mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\frac{\partial \mathbf{C}^T}{\partial c(i)}\right\}\right] \\ &= \mathrm{tr}\left\{ \frac{\partial \mathbf{C}}{\partial c(i)}(\mathbf{C}^T\mathbf{C})^{-1}\frac{\partial \mathbf{C}^T}{\partial c(j)}(\mathbf{W}_1 + \mathbf{W}_2)\right\} \\ &= \mathrm{tr}\left\{ \frac{\partial \mathbf{C}}{\partial c(i)}(\mathbf{C}^T\mathbf{C})^{-1}\frac{\partial \mathbf{C}^T}{\partial c(j)}\frac{\rho^2}{4}(\xi_1\xi_1^* + \eta_1\eta_1^* + \xi_2\xi_2^* + \eta_2\eta_2^*)\right\} \end{aligned} \tag{9}$$

where $\xi_1^* = (1, e^{-i\omega}, \ldots, e^{-i(N-3)\omega})$, $\eta = \overline{\xi}$ and $*$ denotes conjugate transpose, so that $\mathbf{W}_1 = \frac{\rho^2}{2}(\xi_1\xi_1^* + \eta_1\eta_1^*)$. The vectors ξ_2 and η_2 are defined similarly to represent the decomposition of $\mathbf{W}_2$.

We need only consider the first and third terms of (9),

$$A_k = \frac{\rho}{2}\xi_k^*\frac{\partial \mathbf{C}}{\partial \mathbf{c}(i)}(\mathbf{C}^T\mathbf{C})^{-1}\frac{\partial \mathbf{C}^T}{\partial \mathbf{c}(j)}\frac{\rho}{2}\xi_k$$

for $k = 1, 2$, since the analysis of the terms involving η_k is identical except for signs. The key to the proof is a triangular decomposition of $(\mathbf{C}^T\mathbf{C})$ as $\boldsymbol{\mathcal{L}}\boldsymbol{\mathcal{L}}^T$ where $\boldsymbol{\mathcal{L}}$ is lower triangular. For $p = 1$, $\mathbf{c} = (1, -2\cos\omega, 1)^T$ and hence we can show that

$$\boldsymbol{\mathcal{L}} = \frac{1}{\sin\omega}\begin{pmatrix} \sin\omega & & & \mathbf{0} \\ \sin 2\omega & \sin\omega & & \\ \vdots & \vdots & \ddots & \\ \sin(N-2)\omega & \cdots & \cdots & \sin\omega \end{pmatrix} + \begin{matrix}\text{a matrix of} \\ \text{negligible terms.}\end{matrix}$$

The terms which are ignored in $\boldsymbol{\mathcal{L}}$ arise from the fact that $\mathbf{C}$ is not square; the contribution of the $2p = 2$ extra rows is dealt with by partitioning $\mathbf{C}^T = [\mathbf{L}^{-T} \mid \mathbf{N}]$, so that $\mathbf{C}^T\mathbf{C} = \mathbf{L}^{-T}\mathbf{L}^{-1} + \mathbf{N}\mathbf{N}^T$ where $\mathbf{N}\mathbf{N}^T = [\mathbf{0} \mid \mathbf{I}_2]^T\mathbf{M}[\mathbf{0} \mid \mathbf{I}_2]$, with

$$\mathbf{M} = \begin{pmatrix} 1 & -2\cos\omega \\ -2\cos\omega & 1 + 4\cos^2\omega \end{pmatrix}.$$

Straightforward algebra shows that

$$\begin{aligned} (\mathbf{C}^T\mathbf{C})^{-1} &= \mathbf{L}\{\mathbf{I} - \mathbf{K}(\mathbf{M}^{-1} + \mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T\}\mathbf{L}^T \\ &= \mathbf{L}\mathbf{L}^T + \mathbf{S} \end{aligned} \tag{10}$$

where $\mathbf{K} = \mathbf{L}[\mathbf{0} \mid \mathbf{I}_2]^T$ and $(\mathbf{M}^{-1} + \mathbf{K}^T\mathbf{K})^{-1}$ is only a 2×2 matrix, whose row sums are $O(N^{-1})$.

The matrices $\partial\mathbf{C}/\partial\mathbf{c}(j)$ are $[\mathbf{I}_{N-2}, \mathbf{0}, \mathbf{0}]$, $[\mathbf{0}, \mathbf{I}_{N-2}, \mathbf{0}]$, and $[\mathbf{0}, \mathbf{0}, \mathbf{I}_{N-2}]$, for $j =$ 1, 2 and 3, so that $\partial\mathbf{C}/\partial\mathbf{c}(j)\,\mathbf{L}$ is of the same form with $\mathbf{L}$ replacing $\mathbf{I}_{N-2}$ in each case. A_1 in the expression below (9) can be seen to be the squared length of the

vector $(\rho/2)\xi_1^*\partial\mathbf{C}/\partial\mathbf{c}(j)\,\mathbf{L}$; writing $\sin(j\omega) = (e^{ij\omega} - e^{-ij\omega})/2i$ we can write these vectors as sums of products of complex exponentials, say v_1 v_2 and v_3, in which the k'th element of v_1 will be the same as the $(k+1)$st element of v_2 and the $(k+2)$nd element of v_3, namely

$$\begin{aligned} v_1(k) &= \frac{\rho}{4i\sin\omega}\sum_{j=k-1}^{n-3} e^{-ij\omega}\big(e^{i(j-k+2)\omega} - e^{-i(j-k+2)\omega}\big) \\ &= \frac{\rho}{4i\sin\omega}\sum_{j=k-1}^{N-3} e^{-(k-2)\omega} - e^{-(k-2)\omega}e^{-ij(2\omega)} \\ &= \frac{\rho(N-k-2)e^{-(k-2)\omega}}{4i\sin\omega} + o(N). \end{aligned}$$

Observe that the leading term is independent of j, but depends on the position k of the element in the vector. Forming the inner product, for example $v_1 v_2^*$ using the fact that $v_2(1) = 0$ we obtain

$$\begin{aligned} v_1 v_2^* &= \frac{\rho^2}{4}\sum_{k=2}^{N-2}\frac{(N-k-2)}{2i\sin^2\omega}e^{-i(k-2)\omega} \times \frac{(N-k-3)}{-2i}e^{i(k-3)\omega} \\ &= \frac{\rho^2}{4}\,\frac{e^{-i\omega}}{4\sin^2\omega}\sum_{k=2}^{N-2}(N-k-2)(N-k-3) \\ &= \frac{\rho^2 N^3 e^{-i\omega}}{48\sin^2\omega} + O(N^2). \end{aligned}$$

Following the same algebra through with the second term of (10) shows that $\xi_1^*\mathbf{S}\xi_1 = o(N^3)$ and hence can be ignored as required.

Repeating the above argument using the third term of (9), we see that A_2 ia the squared length of the vector $(\rho/2)\xi_2^*\partial\mathbf{C}/\partial\mathbf{c}(j)\,\mathbf{L}$; each element of this vector can be shown to be $o(N)$ by algebra parallel to that used in evaluating $v_1(k)$, so that the contributions to (9) involving $\mathbf{W}_2$ can be shown to be $o(N^3)$ and hence to be asymptotically negligible in this context.

Combining the inner product involving ξ_1 with that involving η_1 leads to

$$E(\mathbf{B}_{12}) = \frac{N^3}{12\sin^2\omega}\left(\frac{\rho^2}{2}\cos\omega\right) + O(N^2).$$

The evaluation of the remaining terms of $\mathbf{B}$ follows exactly the same lines and is omitted.

Acknowledgement

This paper was prepared while the first author was visiting the Statistics Research Section of the Australian National University.

REFERENCES

Bloomfield, P. (1976) *The Spectral Analysis of Time Series: An Introduction.* Wiley, New York.

Chan, Y.T., Lavoie, J.M.M. and Plant, J.B. (1981) A parameter estimation approach to estimation of frequencies of sinusoids. *IEEE Trans. Acoust., Speech and Sig. Proc.* **ASSP-29**, 214–219.

Dwyer, P.S and MacPhail, M.S. (1948) Symbolic matrix derivatives. *AMS* **19**, 517–534.

Feller, W. (1971) *An Introduction to Probability Theory and its Applications.* Vol.2. Wiley, New York.

Golub, G.H. and van Loan, C.F. (1983) *Matrix Computations* 1st Ed. Johns Hopkins, Baltimore.

Hannan, E.J. (1971) Non-linear time series regression. *J. Appl. Prob.* **8**, 767–780.

Hannan, E.J. (1973) The estimation of frequency. *J. Appl. Prob.* **10**, 510–519.

Kahn, M., Mackisack, M.S., Osborne, M.R. and Smyth, G.K. (1991) On the consistency of Prony's method and related algorithms. *J. Comp. Graph. Statist.* In press.

Mackisack, M.S. and Poskitt, D.S. (1989) Autoregressive frequency estimation. *Biometrika* **76**, 565–576

Mackisack, M.S. and Poskitt, D.S. (1990) Some Properties of autoregressive estimates for processes with mixed spectra. *J. Time Series Anal.* **11**, 325–337.

Marple, S.L. (1987) *Digital Spectral Analysis with Applications* Prentice Hall, Englewood Cliffs N.J.

Odeh, R.E and Evans, J.O. (1974) Algorithm AS70: The percentage points of the Normal distribution. *Appl. Statist.* **23**, 96–97

Osborne, M.R and Smyth, G.K. (1991) A modified Prony algorithm for fitting functions defined by difference equations. *SIAM J. Sci. Statist. Comp.* **12**, 362–382

Osborne, M.R and Smyth, G.K. (1993) A modified Prony algorithm for fitting sums of exponential functions. Submitted.

Priestley, M.B. (1981) *Spectral Analysis and Time Series* vol.1 Academic Press, London.

Quinn, B.J. (1986) Testing for the presence of sinusoidal components. Contribution to *Essays in Time Series and Allied Processes: Papers in honour of E.J. Hannan* J. Gani and M.B. Priestley (editors), Applied Probability Trust, Sheffield p.201–210

Rice, J.A. and Rosenblatt, M. (1988) On frequency estimation. *Biometrika* **75**, 477–484

Richards, F.S.G. (1961) A method of maximum likelihood estimation. *J. Roy. Statist. Soc. Ser. B* **23**, 469–475

Sakai, H (1984) Statistical analysis of Pisarenko's method for sinusoidal frequency estimation.*IEEE Trans. Acoust., Speech and Sig. Proc.* **ASSP-32**, 95–101

Walker, A.M. (1971) On the estimation of a harmonic component in a time series with stationary independent residuals. *Biometrika,* **58**, 21–36.