UNITED STATES
DEPARTMENT OF THE INTERIOR
GEOLOGICAL SURVEY

MAXIMUM LIKELIHOOD ESTIMATION OF TRANSFER FUNCTION PARAMETERS WHEN INPUT AS WELL AS OUTPUT OBSERVATIONS ARE SUBJECT TO ERROR

by David D. Goodman

Open-File Report 81-548

Reston, Virginia
December 1980

UNITED STATES DEPARTMENT OF THE INTERIOR

JAMES G. WATT, Secretary

GEOLOGICAL SURVEY

Doyle G. Frederick, Acting Director

For additional information write to:

U.S. Geological Survey
Water Resources Division
National Center, Mail Stop 430
12201 Sunrise Valley Drive
Reston, Virginia 22092

CONTENTS

Page

Abstract iv

Introduction 1

Formulation of the Likelihood Function and Preliminary Minimization Steps 3

A Computationally Efficient Expression of the Likelihood Function 8

The Computational Cost of Evaluating the Likelihood Function 10

The Convergence of the Diagonally Positive, Upper-triangular Square Root Decomposition of Covariance Matrices of Finite Contiguous Segments of Stationary Timer Series 14

References 17

Appendices

A. Proof of the Equivalence of Li and Li´ for the Case of a Pure Moving Average Process 18

B. Decomposition of Positive Definite Banded Matrices into Symmetric LU Form 20

C. Roots of Symmetric Polynomials of Even Degree 22

D. The Symmetric Decomposition of Specially Formed Symmetric Polynomials of Even Degree 25

E. The Square Root Decomposition of Covariance Matrices of N Consecutive Random Variables Generated from a Moving Average Process 29

TABLES

Page

1.--Number of multiplications per operation required to evaluate the likelihood function . . . 9

2.--Computational cost per operation to evaluate the likelihood function 11

ABSTRACT

An efficient procedure is described for computing the maximum likelihood estimates of parameters associated with a transfer-function model defining a linear relationship between two discrete time series. Two equivalent expressions are developed for the likelihood of the transfer function parameters, given observations of the input and output series that are each normally distributed about their true values. Several algebraic theorems are developed which provide shortcuts for the numerical evaluation of one of these expressions.

Maximum Likelihood Estimation of Transfer Function Parameters When Input as Well as Output Observations are Subject to Error

by

David D. Goodman

ABSTRACT

An efficient procedure is described for computing the maximum likelihood estimates of parameters associated with a transfer-function model defining a linear relationship between two discrete time series. Two equivalent expressions are developed for the likelihood of the transfer function parameters, given observations of the input and output series that are each normally distributed about their true values. Several algebraic theorems are developed which provide shortcuts for the numerical evaluation of one of these expressions.

INTRODUCTION

Consider the problem of estimating the behavior of a discrete time series, α_i, as a linear function of a related time series β_i. When the two signals are random and jointly weakly stationary (see Bloomfield, 1976 for a precise definition of this property), the linear least squares estimate can be computed from the first and second moments of the joint probability distribution. Specifically,

$$\alpha_i = \vec{\Sigma}^T_{\alpha\beta}\Sigma^{-1}_{\beta\beta} \ (\vec{\beta} - \vec{\mu}_\beta) + \mu_\alpha \qquad (1)$$

where $\vec{\beta}$ is a vector of the β_j variables used to form the estimate, $\vec{\mu}_\beta$ is a vector whose elements are all identical and are equal in value to the mean of the β_i series and in number to the number of variables in $\vec{\beta}$, μ_α is the mean of the α_i series, $\vec{\Sigma}_{\alpha\beta}$ is the covariance vector between α_i and $\vec{\beta}$, and $\Sigma_{\beta\beta}$ is the covariance matrix of the β_j (Whittle, 1963); the superscript T and -1, respectively, indicate the transform and inverse of a matrix. Estimates of the above moments can be used at some cost in prediction accuracy whenever the precise values are not known.

When randomness and(or) stationarity is in doubt, a logical strategy is to fit a functional relationship to concurrent observations of the two series that can be used, subsequently, for estimation purposes when observations of only one series are available. The most general linear relationship between the α_i and β_i is given by:

$$\alpha_1 + \sum_{j=1}^{n} a_j \alpha_{i-j} = \sum_{j=0}^{m} b_j \beta_{i-j} + \mu \tag{2}$$

where the a_j and b_j and μ are parameters that are adjusted to achieve the fit. In most time series applications that employ eq. (2), observations of α_i are assumed to be normally distributed about α_i:

$$y_i = \alpha_i + \varepsilon_i \tag{3}$$

where the ε_i derive from a zero mean, unit variance, time dependent Gaussian random process. Observations of β_i are usually assumed to be exact:

$$x_i = \beta_i. \tag{4}$$

Theory dictates that both the maximum likelihood and the least squares estimates of the a_j, b_j, α_i, β_i, and μ are those parameter values that minimize the sums of squares of every ε_i that is associated with an observed α_i. If the ε_i are not normal these values will still be the least squares estimates. If the ε_i are not independent in time, computation of the maximum likelihood and(or) least squares estimates is more complicated.

The popularity of the observation model defined above probably stems from the ease with which approximate maximum likelihood estimates (MLEs) can be computed from observations. (See Box and Jenkins, 1970, for an exhaustive treatment of this procedure.) However, the inherent assumptions in eqs. (2) and (4) are often not physically justified by measurement errors or other physical phenomena. When these assumptions are not justified, the observation model and the minimum $\Sigma\varepsilon_i^2$ objective really do no more than define criteria for fitting eq. (2) to observations. It is the contention of the author that in applications where there is no basis for discriminating between the nature of the observations of the two series, it is more reasonable to assume that both x_i and y_i are normally distributed about their means;

$$x_i = \beta_i + \delta_i \tag{5}$$

$$y_i = \alpha_i + \varepsilon_i \tag{6}$$

where ε and δ are normally distributed with mean, 0, and standard deviation, σ. In addition the ε and δ are uncorrelated, so that

$$E(\delta_i \delta_j) = 0 \quad (i \neq j),$$

$$E(\varepsilon_i, \varepsilon_j) = 0 \quad (i \neq j),$$

$$E(\varepsilon_i, \delta_j) = 0 \quad (\text{all } i,j).$$

The major purpose of this paper is to develop an efficient procedure for computing the maximum likelihood estimates of the a_j, b_j and μ for the model defined by eqs. (2), (5), and (6).

FORMULATION OF THE LIKELIHOOD FUNCTION AND PRELIMINARY MINIMIZATION STEPS

Assume one has N consecutive observations of α_i beginning at time k and M observations of β_i beginning at time p. The likelihood of the parameters given these data is the joint probability density of the observations x_i and y_i:

$$Li\ (a_j,\ b_j,\ \alpha_i,\ \beta_i,\ \mu) =$$

$$(2\Pi\sigma^2)^{-\frac{(N+M)}{2}} \exp - \frac{1}{2\sigma^2} \left(\sum_{i=p}^{M+p-1} (x_i - \beta_i)^2 + \sum_{i=k}^{N+k-1} (y_i - \alpha_i)^2 \right) \quad (7)$$

The likelihood function $Li(a_j,b_j,\alpha_i,\beta_i,\mu)$, that will be subsequently used is the portion of the negative log of eq. (7) that varies with the parameters:

$$Li(a_j,b_j,\alpha_i,\beta_i,\mu) = \sum_{i=p}^{M+p-1} (x_i - \beta_i)^2 + \sum_{i=k}^{N+k-1} (y_i - \alpha_i)^2. \quad (8)$$

Clearly, minimizing eq. (8) maximizes eq. (7).

The first step in minimizing Li is to express the function in terms of an independent set of parameters. Using eq. (2), some of the α_i can be expressed in terms of the β_i, b_j, a_j, μ and other α_i. Before reexpression, part of the likelihood function will be put in matrix notation. This separation will facilitate the further analysis.

Let j_1 and j_2 delimit the segment of α_i that can be expressed strictly in terms of other parameters. It should be clear that j_1 is the larger of $k + n$ and $p + m$ and that j_2 is the smaller of $N + k - 1$ and $M + p - 1$. The α_i in this range can be expressed through eq. (2) in terms of each other, μ, the a_j and b_j, the β_j between $j_1 - m$ and j_2 inclusive and the α_i between $j_1 - n$ and $j_1 - 1$.

Now let $P = j_2 - j_1 + 1$, $\vec{\beta}$ be a vector comprised of the $P + m$ β_i between $j_1 - m$ and j_2 inclusive, $\vec{x}$ be a vector of corresponding x_i, $\vec{\alpha}$ be a vector composed of the $P + n$ α_i between $j_1 - n$ and j_2 inclusive, and $\vec{y}$ be a vector of corresponding y_i. Using these constructs the likelihood function becomes:

$$\begin{aligned} Li(a_j,b_j,\alpha_i,\beta_i,\mu) &= (\vec{x} - \vec{\beta})^T(\vec{x} - \vec{\beta}) + (\vec{y} - \vec{\alpha})^T(\vec{y} - \vec{\alpha}) \\ &+ \sum_{i=p}^{j_1-m-1} (x_i - \beta_i)^2 + \sum_{i=j_2+1}^{M} (x_i - \beta_i)^2 \\ &+ \sum_{i=k}^{j_1-n-1} (y_i - \alpha_i)^2 + \sum_{i=j_2+1}^{N} (y_i - \alpha_i)^2 . \qquad (9) \end{aligned}$$

Note that at least two and possibly more of the four summations at the end of eq. (9) will not exist, depending on the overlap between the two observation segments.

The parameters involved in the vector portion of the equation are interdependent while the remaining parameters are not. Specifically, a total of P linear relationships can be established between the elements of $\vec{\alpha}$ and $\vec{\beta}$ through eq. (2). In the following developments, the first P elements of $\vec{\alpha}$ will be eliminated by this equation.

Define a vector $\vec{\theta}$ as the augmentation of $\vec{\beta}$ with the last n elements of $\vec{\alpha}$ and let IP be a $(P + m) \times (P + m + n)$ matrix such that if ip_{ij} defines the elements of IP,

$$ip_{ij} = 0 \qquad i \neq j$$
$$= 1 \qquad i = j. \tag{10}$$

Consequently

$$\vec{\beta} = IP\vec{\theta}\ . \tag{11}$$

Next define the matrices A and B in the following manner. If a_{ij} defines the jth element in the ith row of the (P + n) x (P + n) matrix A, then:

$$\begin{aligned} a_{ij} &= 0 && j < i \\ &= 1 && j = i \\ &= a_{j-i} && i + n \geq j > i,\ i \leq p \\ &= 0 && j > i,\ i > P\ . \end{aligned} \tag{12}$$

If b_{ij} defines the jth element in the ith row of the (P + n) x (P + n ÷ m) matrix B, then

$$\begin{aligned} b_{ij} &= 0 && j < i \\ &= b_{j-i} && i + m \geq j \geq i,\ i \leq p \\ &= 1 && j = i + m + 1,\ i > p \\ &= 0 && j > i,\ i < P\ . \end{aligned} \tag{13}$$

According to eq. (2) and the above definitions,

$$A\vec{\alpha} = B\vec{\theta} + \vec{u} \tag{14}$$

where $\vec{u}$ is a vector of length P + n, the first P elements of which are u and the least n elements of which are zero. As A is an upper-triangular square matrix with non-zero elements on its diagonal, its determinant is non-zero and its inverse exists. Consequently,

$$\vec{\alpha} = A^{-1}(B\vec{\theta} + \vec{u}) \tag{15}$$

and the likelihood function becomes:

$$Li(a_j, b_j, \beta_i, \alpha_i, \mu) = [\vec{x} - IP\vec{\theta}]^T[\vec{x} - IP\vec{\theta}] + [\vec{y} - A^{-1}(B\vec{\theta}+\vec{u})]^T[\vec{y} - A^{-1}(B\vec{\theta}+\vec{u})]$$
$$+ \sum_{i=p}^{j_1-m-1} (x_i - \beta_i)^2 + \sum_{i=j_2+1}^{M} (x_i - \beta_i)^2$$
$$+ \sum_{i=k}^{j_1-n-1} (y_i - \alpha_i)^2 + \sum_{i=j_2+1}^{N} (y_i - \alpha_i)^2 \ . \tag{16}$$

Note that eq. (16) is quadratic in $\vec{\theta}$, β_i and α_i. This means that the first derivative of Li with respect to $\vec{\theta}$, β_i, or α_i will be linear in these parameters. It makes sense, then, to eliminate these parameters from eq. (16) as a first step in minimizing the likelihood function.

Setting the derivative of eq. (16) with respect to the α_i and β_i not included in $\vec{\theta}$ zero generates

$$x_i = \beta_i$$

and

$$y_i = \alpha_i \ . \tag{17}$$

This result demonstrates that "outlying" observations do not contribute to the likelihood function at its maximum.

Setting the derivative of eq. (16) with respect to $\vec{\theta}$ to zero gives:

$$-IP^T x + IP^T IP\vec{\theta} - B^T(A^{-1})^T\vec{y} + B^T(A^TA)^{-1}B\vec{\theta} + B^T(A^{-1})^T\vec{u} = 0$$

or

$$[B^T(A^TA)^{-1}B + IP^TIP]\vec{\theta} = IP^T\vec{x} + B^T(A^{-1})^T(\vec{y} - A^{-1}\vec{u}). \tag{18}$$

Now let

$$\Gamma = [B^T(A^TA)^{-1}B + IP^TIP]. \tag{19}$$

Then

$$\vec{\theta} = \Gamma^{-1}[IP^T\vec{x} + B^T(A^{-1})^T(\vec{y} - A^{-1}\vec{u})]. \tag{20}$$

Replacing eq. (20) in eq. (16) generates:

$$\begin{aligned} Li(a_j,b_j,\mu) = {} & [\vec{x} - IP\Gamma^{-1}[IP^T\vec{x} + B^T(A^{-1})^T(\vec{y} - A^{-1}\vec{u})]]^T \\ & [\vec{x} - IP\Gamma^{-1}[IP^T\vec{x} + B^T(A^{-1})^T(\vec{y} - A^{-1}\vec{u})]] \\ & + [\vec{y} - A^{-1}[B\Gamma^{-1}[IP^T\vec{x} + B^T(A^{-1})^T(\vec{y} - A^{-1}\vec{u})] + \vec{u}]]^T \\ & [\vec{y} - A^{-1}[B\Gamma^{-1}[IP^T\vec{x} + B^T(A^{-1})^T(\vec{y} - A^{-1}\vec{u})] + \vec{u}]] \end{aligned} \tag{21}$$

where the summation terms at the end of eq. (16) have been dropped off due to eq. (17). After considerable algebra and cancellation, Li can be reexpressed as:

$$\begin{aligned} Li(a_j,b_j,u) = {} & \vec{x}^T[I - IP\Gamma^{-1}IP^T]\vec{x} + \vec{y}^T[I - A^{-1}B\Gamma^{-1}B^T(A^{-1})^T]\vec{y}^T \\ & - 2\vec{x}^T[IP\Gamma^{-1}B^T(A^{-1})^T]\vec{y} + 2\vec{x}^T[IP\Gamma^{-1}B^T(A^TA)^{-1}]\vec{u} \\ & + 2\vec{y}^T[A^{-1}B\Gamma^{-1}B^T(A^TA)^{-1} - I]\vec{u} + \vec{u}[(A^{-1})^T[I - A^{-1}B\Gamma^{-1}B^TA^{-1}]]\vec{u}. \end{aligned} \tag{22}$$

Note that the above expression is highly non-linear in the parameters a_j and b_j. As such, it is probably best minimized through iterative procedures which require repetitive evaluation of eq. (22). In the following paragraphs, an equivalent expression is developed which is computationally easier to evaluate.

A COMPUTATIONALLY EFFICIENT EXPRESSION OF THE LIKELIHOOD FUNCTION

Consider the following vector of random variables:

$$\vec{\psi} = \hat{A}\vec{y} - \hat{B}\vec{x} - \vec{u} = \hat{A}(\vec{\alpha} + \vec{\varepsilon}) - \hat{B}(\vec{\beta} + \vec{\delta}) - \vec{u} \tag{23}$$

where $\hat{A}$ is a P x (P + m) matrix formed from the first P rows of the matrix A, $\hat{B}$ is a P x (P + m) matrix formed from the first P rows and p + m columns of the matrix B, $\vec{\varepsilon}$ is a vector containing the ε_i that are associated with the y_i in $\vec{y}$, and $\vec{\delta}$ is a vector containing the δ_i that are associated with the x_i in $\vec{x}$. From eq. (2) it should be clear that

$$\vec{\psi} = \hat{A}\vec{\varepsilon} - \hat{B}\vec{\delta} \ . \tag{24}$$

Now let

$$\hat{\Gamma} = E[\vec{\psi}^T\vec{\psi}] \ . \tag{25}$$

Due to the independence of the ε_i and δ_i

$$\hat{\Gamma} = [\hat{A}\hat{A}^T + \hat{B}\hat{B}^T] \ . \tag{26}$$

Finally, construct a vector of P uncorrelated, unit variance random variables $\vec{\theta}$ by decomposing $\hat{\Gamma}$ according to

$$G^TG = \hat{\Gamma} \tag{27}$$

and defining $\vec{\theta}$ by

$$\vec{\theta} = (G^{-1})^T\vec{\psi} \ . \tag{28}$$

The existence of a matrix G that satisfies eq. (27) is guaranteed because Γ is positive definite, as can be seen from eq. (26). The desired alternative expression of the likelihood function is the portion of the log of the density function of $\vec{\theta}$ that varies with the parameters a_j, b_j, and μ:

$$\begin{aligned} Li'(a_j,b_j,\mu) &= \vec{\theta}^T\vec{\theta} = \vec{\psi}^T\hat{\Gamma}^{-1}\vec{\psi} \\ &= [\hat{A}\vec{y} - \hat{B}\vec{x} - \vec{u}]^T\hat{\Gamma}^{-1}[\hat{A}\vec{y} - \hat{B}\vec{x} - \vec{u}] \\ &= \vec{x}^T[\hat{B}^T\hat{\Gamma}^{-1}\hat{B}]\vec{x} + \vec{y}^T[\hat{A}^T\hat{\Gamma}^{-1}\hat{A}]\vec{y} \\ &\quad -2\vec{x}^T[\hat{B}^T\hat{\Gamma}^{-1}\hat{A}]\vec{y} + 2\vec{x}^T[\hat{B}^T\hat{\Gamma}^{-1}]\vec{u} \\ &\quad - 2\vec{y}^T[\hat{A}^T\hat{\Gamma}^{-1}]\vec{u} + \vec{u}^T\hat{\Gamma}^{-1}\vec{u} \ . \end{aligned} \tag{29}$$

Although the general equivalence of Li and Li′ has not yet been proven analytically, the two expressions produced identical values in each of twenty separate numerical experiments. An analytical proof of the equivalence is given in Appendix A (p. 17-18) for the special case of a pure moving average model, i.e., $a_i = 0$, for all $i \geq 1$.

THE COMPUTATIONAL COST OF EVALUATING THE LIKELIHOOD FUNCTION

Inspection of eq. (21) reveals that among all the operations required to compute Li, those involving the construction and inversion of Γ require the largest number of multiplications to complete. A procedure for completing each of these operations is described in the following three paragraphs. Although these procedures are chosen to offer an efficient approach to the computation of their associated operations, no claim to optimality is made. The number of multiplications required to complete each procedure is listed in Table 1.

Table 1

	Operation	Number of Multiplications
1)	$[A^T A]^{-1} B$	$3/2 P^2 n$
2)	$B^T [A^T A]^{-1} B$	$P^2 m$
3)	$\Gamma = [B^T [A^T A]^{-1} B + IP^T IP]$	0
4)	$\Gamma^{-1} [IP^T \vec{x} + B^T (A^{-1})^T (\vec{y} - \vec{u})]$	$1/6 P^3$
	Total Multiplications Required:	$1/6 P^3 + P^2 \ (3/2\ n + m)$

The product in the first step is computed by solving the following two systems of linear equations:

$$A \vec{c}_j = \vec{b}_j \tag{30}$$

$$A^T \vec{d}_j = \vec{c}_j \tag{31}$$

where $\vec{b}_j$ is the jth column of B and $\vec{c}_j$ and $\vec{d}_j$ are the vectors to be computed. It should be clear that $\vec{d}_j$ will correspond to the jth column of the desired product. To understand the figure given in table 1 recall that A is by definition an upper triangular, P + n degree, square matrix of bandwidth n. Also, the jth column of B contains only zero entries below the jth element. It follows that the last P + n - j entries in $\vec{c}_j$ will be zero and that eq. (30) can be solved using back substitution in approximately nj multiplications. Solution of eq. (31) through back substitution without shortcuts requires about (P + n)n multiplications. As B contains P + n + m columns the total number of multiple operations required is approximately

$$\sum_{j=1}^{P+n+m} (P + n)n + jn \approx 3/2P^2n$$

whenever P>>n,m.

The product formed above will be a (P + n) x (P + n + m) matrix that is in general not banded or triangular. As B has exactly m entries in nearly every row, the second step will require approximately

$$(P + n + m)(P + n)m \approx P^2m$$

multiplications.

The product in the fourth step is computed by solving the following system of linear equations:

$$\Gamma\vec{c} = \vec{d} \qquad (32)$$

where $\vec{d} = IP^T\vec{x} + B^T(A^{-1})^T(\vec{y} - \vec{u})$ and is assumed to have been computed previously. It should be clear that $\vec{c}$ is the desired product. That Γ is positive definite can be seen by viewing the matrix as the sum of two positive definite matrices $[A^{-1}B]^T[A^{-1}B]$ and IP^TIP. This property insures that eq. (32)

can be solved analytically through symmetric Gaussian elimination (explained in Appendix B), or iteratively through the Gauss-Seidel method (Forsythe and Moler, 1967). As Γ is a (P + n + m) x (P + n + m) matrix, the former process requires on the order of $1/6\ P^3$ multiplications when $P >> n,m$. The latter process requires exactly $(P + n + m)^2$ k operations where k is the number of iterations required to achieve satisfactory convergence.

Table II summarizes the computational cost of evaluating the likelihood function using eq. (29). The particular sequence of matrix operations is chosen to minimize the total number of multiplications.

TABLE II

	Operation	Multiplications Required
1)	$\hat{A}\vec{y}$	Pn
2)	$\hat{B}\vec{x}$	Pm
3)	$[\hat{A}\vec{y} - \hat{B}\vec{x} - \vec{u}]$	0
4)	$\hat{A}\hat{A}^T$	$\frac{1}{2}n^2$
5)	$\hat{B}\hat{B}^T$	$\frac{1}{2}m^2$
6)	$\hat{\Gamma} = \hat{A}\hat{A}^T + \hat{B}\hat{B}^T$	0
7)	$\hat{\Gamma}^{-1}[\hat{A}\vec{y} - \hat{B}\vec{x} - \vec{u}]$	$P(\frac{1}{2}p^2 + 2p)$
8)	$[\hat{A}\vec{y} - \hat{B}\vec{x} - \vec{u}]^T\ \hat{\Gamma}^{-1}[\hat{A}\vec{y} - \hat{B}\vec{x} - \vec{u}]$	P
	Approximate Total Number of Multiplications	$P(n+m+\frac{1}{2}p^2+2p+1)$

The figures given for the multiplications required by the first two steps are based on the fact that $\hat{A}$ is a P x (P + n) matrix with exactly n non-zero elements in each row while $\hat{B}$ is P x (P + m) with exactly m non-zero elements per row. To understand the figures for steps four and five let c_{ij} define the jth element in the ith row of $\hat{A}\hat{A}^T$. From the definition of $\hat{A}$ it can be seen that:

$$a_{i,i+k} = \sum_{j=0}^{n-|k|} a_j a_{j+|k|} \tag{33}$$

where $a_0 = 1$. It follows that there are only n + 1 different non-zero values among all the entries of $\hat{A}\hat{A}^T$. The numbers of multiplications necessary to compute them is given by

$$\sum_{k=0}^{n-1} n - k \approx \frac{1}{2} n^2 .$$

The figure given for $\hat{B}\hat{B}^T$ is similarly explained.

The figure given for step seven is based on the assumption that the product is computed through the following procedure. First $\hat{\Gamma}$ is decomposed according to

$$\hat{\Gamma} = G^T G \tag{34}$$

where G is upper triangular with positive diagonal elements. The process is completed by solving the following two systems of equations:

$$G\vec{c} = [\hat{A}\vec{y} - \hat{B}\vec{x} - \vec{u}] \tag{35}$$

$$G^T\vec{d} = \vec{c} \tag{36}$$

It should be clear that $\vec{d}$ is the desired product.

It can be seen from eq. (26) that $\hat{\Gamma}$ is a positive definite square matrix of degree P and bandwidth p = max(m,n). The positive definiteness guarantees the existence and uniqueness of G (Forsythe and Moler, 1967). The bandedness implies that G will also have a bandwidth of p. An algorithm is described in Appendix B that can perform the decomposition indicated by eq. (34) on matrices with these properties using approximately $\frac{1}{2}Pp^2$ multiplications. The systems eqs. (35) and (36) can be solved through back substitution using approximately Pp multiplications, due to the triangularity and bandedness of G.

Table II shows that the $\frac{1}{2}Pp^2$ multiplications required by the decomposition of $\hat{\Gamma}$ makes that step the most critical link in the evaluation of the likelihood function through eq. (29). The last section of this paper demonstrates how this decomposition procedure can often be cut short due to the convergence of the rows of G.

THE CONVERGENCE OF THE DIAGONALLY POSITIVE, UPPER-TRIANGULAR SQUARE ROOT DECOMPOSITION OF COVARIANCE MATRICES OF FINITE CONTIGUOUS SEGMENTS OF STATIONARY TIME SERIES.

Consider the vector of random variables $\vec{\psi}$ defined by eq. (23). Let ψ_i denote the ith element in $\vec{\psi}$. It should be clear that the covariance structure of the ψ_i is defined by:

$$E[\psi_i\psi_{i+k}] \equiv \gamma_k = [\sum_{j=0}^{n-|k|} a_j a_{j+k} + \sum_{j=0}^{m-|k|} b_j b_{j+k}]\sigma^2 \qquad 0\leq|k|\leq\max(m,n)$$

$$= 0 \qquad |k|>\max(n,m) \qquad (37)$$

It can be seen from eq. (26) that the γ_k correspond to the elements of the covariance matrix $\hat{\Gamma}$. Specifically, if $\hat{\Gamma}_{n,n+|k|}$ defines the $n+|k|$th entry in the nth row of $\hat{\Gamma}$, then $\hat{\Gamma}_{n,n+|k|} = \gamma_k$.

Now define

$$q(z) = \sum_{i=-\infty}^{\infty} q_i z^i \qquad (38)$$

as the z transform of any series q_i. It can be seen from eq. (37) that

$$\gamma(z) = a(z)z(z^{-1}) + b(z)b(z^{-1}) \qquad (39)$$

Next express ψ_i as a moving average of an independent series of zero mean unit variance Gaussian random variables ϕ_i:

$$\psi_i = \sum_{j=-\infty}^{\infty} c_j\phi_{i-j} \qquad (40)$$

where the c_j are such that c(z) is a polynomial in z of degree p = max(n,m) with roots that all lie outside the unit circle. Computing the covariance structure of the ψ_i using eq. (40) gives:

$$\gamma_k = \sum_{j=0}^{p-|k|} c_j c_{j+|k|} \qquad 0 \leq |k| \leq p$$

$$= 0 \qquad |k| > p \tag{41}$$

or

$$\gamma(z) = c(z)c(z^{-1}) \tag{42}$$

The existence and uniqueness of the c_j is proven in Appendices C and D.

Finally consider the following decomposition of $\hat{\Gamma}$:

$$\hat{\Gamma} = G^T G \tag{43}$$

where G is upper triangular with positive diagonal elements. As mentioned in the previous section, the existence and uniqueness of G follows from the fact that $\hat{\Gamma}$ is positive definite. Also, in Appendix B it is shown that G has a bandwidth of p. Now let $g_{i,j}$ correspond to the jth element in the ith row of G. The critical result,

$$\lim_{n\to\infty} [g_{n,n+j}] = c_j \tag{44}$$

is proven in Appendix E. In words, eq. (44) indicates that the entries in G on and to the right of the diagonal in the nth row converge to the c_j series as n gets large. Fortunately, the decomposition algorithm described in Appendix C computes G one row at a time, starting at the top. Also, the number of multiplications required to compute the next row remains constant at $\frac{1}{2}p^2$ after the first p rows have been computed. More importantly, examination of the decomposition algorithm reveals that corresponding entries of all matrices, G, with the same "seed polynomial", $\gamma(z)$, are identical and independent of the magnitude of P! This means that the number of rows required to achieve satisfactory convergence is also independent of P. It follows that for problems with large enough concurrent observation segments, the matrix decomposition requires $\frac{1}{2}QP^2$ multiplications where Q<<P, and the limiting operations in computing $\Gamma^{-1}\vec{c}$ are the solutions of eqs. (35) and (36).

REFERENCES

Bloomfield, Peter, 1976, Fourier analysis of time series: An introduction: John Wiley and Sons, New York, N.Y.

Forsythe, George, and Moler, Cleve B., 1967, Computer solution of linear algebraic systems: Prentice-Hall, Englewood Cliffs, N.J.

Box, G. E. P., and Jenkins, Gwilym M., 1970, Time series analysis: Forecasting and control: Holden-Day, San Francisco, Ca.

Whittle, P., 1963, Prediction and regulation by linear least-squares methods: D. Van Nostrand, Inc., Princeton, N. J.

APPENDIX A

PROOF OF THE EQUIVALENCE OF Li AND Li´ FOR THE CASE OF A PURE MOVING AVERAGE PROCESS

From previous definitions it can be seen that when eq. (2) defines a pure moving average model, i.e., $a_0 = 1$ and $a_n = 0$ for all $n>0$,

1) A is a PxP identity matrix;
2) B is a Px(P+m) matrix whose elements are defined by

$$b_{ij} = 0, \quad j<i$$

$$b_{ij} = b_{j-i}, \quad i \geq j;$$

3) IP is (P+m) x (P+m) identity matrix;
4) $\hat{A}$ is identical to A; and
5) $\hat{B}$ is identical to B.

When used in eq. (22), these simplifications generate:

$$\begin{aligned} Li(b_j,\mu) = {} & \vec{x}^T[I-\Gamma^{-1}]\vec{x} + \vec{y}^T[I-B\Gamma^{-1}B^T]\vec{y} \\ & -2\vec{x}^T[\Gamma^{-1}B^T]\vec{y} + 2\vec{x}^T[\Gamma^{-1}B^T]\vec{u} \\ & + 2y^T[(B\Gamma^{-1}B^T-I]\vec{u} + \vec{u}^T[I-B\Gamma^{-1}B^T]\vec{u} \end{aligned} \qquad A(1)$$

where $\Gamma = [I+B^TB]$.

The alternate expression, eq. (29), becomes

$$\begin{aligned} Li'(b_j,\mu) & = [\vec{y}-B\vec{x}-\vec{u}]^T\hat{\Gamma}^{-1}[\vec{y}-B\vec{x}-\vec{u}] \\ & = \vec{x}^T[B^T\hat{\Gamma}^{-1}B]\vec{x} + \vec{y}^T[\hat{\Gamma}^{-1}]\vec{y} \\ & \quad -2\vec{x}^T[\Gamma^{-1}B^T]\vec{y} + 2x^T[\Gamma^{-1}b^T]\vec{u} \\ & \quad + 2\vec{y}^T[B\Gamma^{-1}B^T-I]\vec{u} + u^T[I-B\Gamma^{-1}B^T]\vec{u} \end{aligned}$$

where $\hat{\Gamma} = [I+B^TB]$.

The equivalence of eqs. A(1) and A(2) can be established by demonstrating the following:

a) $[I - \Gamma^{-1}] = [B^T\hat{\Gamma}^{-1}B]$

b) $[I - B\Gamma^{-1}B^T] = \hat{\Gamma}^{-1}$ and

c) $\Gamma^{-1}B^T = B^T\hat{\Gamma}^{-1}$.

To show (c), start with the following identity:

$$B^T[I + BB^T] = [I + B^TB]B^T$$

and substitute definitions of Γ and $\hat{\Gamma}$;

$$B^T\hat{\Gamma} = \Gamma B^T$$

$$\rightarrow \Gamma^{-1}B^T = B^T\hat{\Gamma}^{-1} \ .$$

To show (a), post multiply (c) by B and reexpress the right side of the equation:

$$\begin{aligned} B^T\hat{\Gamma}^{-1}B &= \Gamma^{-1}B^TB \\ &= \Gamma^{-1}[\Gamma-I] \\ &= [I-\Gamma^{-1}] \ . \end{aligned}$$

Finally, demonstrate (b) by premultiplying (c) by B, postmultiplying by $\hat{\Gamma}$, and using some algebra:

$$BB^T = B\Gamma^{-1}B^T\hat{\Gamma}$$

$$\rightarrow [\hat{\Gamma}-I] = B\Gamma^{-1}B^T\hat{\Gamma}$$

$$\rightarrow \ I = [I-B\Gamma^{-1}B^T]\hat{\Gamma}$$

$$\rightarrow \Gamma^{-1} = [I-B\Gamma^{-1}B^T] \ .$$

APPENDIX B

DECOMPOSITION OF POSITIVE DEFINITE BANDED MATRICES INTO SYMMETRIC LU FORM:

Let Γ be an Nth-degree positive definite matrix with a bandwidth of p. As shown in Forsythe and Moler (1967), Γ can be decomposed according to

$$\Gamma = G^T G \qquad B(1)$$

where G is unique, upper-triangular, and has positive diagonal elements. This appendix consists of a theorem concerning the bandedness of G, and an algorithm for computing its entries.

Theorem: G has a bandwidth of at most p.

Proof: Let g_{ij} be the jth element in the ith row of G and define a set ϕ consisting of all $g_{i,j}$ such that $j>i+p-1$. Also define subsets ϕ_m, $m=1,\ldots N-p$ consisting of all the elements of ϕ that fall in the mth row of G. Finally, let $\gamma_{i,j}$ denote the jth element in the ith row of Γ.

a) If $g_{m,n}\,\varepsilon\phi$, then $\gamma_{m,n}=0$. This follows from the bandedness of Γ and the definition of ϕ.

b) If $g_{m.n}\,\varepsilon\phi$, $g_{m-k,n}\,\varepsilon\phi$ for all k s.t. $1\le k<m$. This follows from the defintion of ϕ. In other words, all of the elements in the column above an element of ϕ are also in ϕ.

c) If $\gamma_{m,n}=0$, and $g_{m-k,n}=0$ for all k s.t. $1\le k<m$, then $g_{m,n}=0$. This result follows directly from the general equation relating the elements of Γ and G:

$$\gamma_{m,n} = \sum_{j=1}^{N} g_{j,m} \times g_{j,n}$$
$$= \sum_{j=1}^{m} g_{j.m} \times g_{j,n}$$

$$= \sum_{j=0}^{m-1} g_{m-j,m} \times g_{m-j,n} \qquad B(2)$$

$$\rightarrow \quad g_{m,n} = \frac{1}{g_{m,m}} [\gamma_{m,n} - \sum_{j=1}^{m-1} g_{m-j,m} \times g_{m-j,n} \qquad B(3)$$

Note that $g_{m.m}$ cannot be zero as G by definition has strictly positive diagonal entries.

Now consider the set ϕ_m. From (a), corresponding elements in Γ will be zero. According to (b), all elements in the columns above the elements of ϕ_m will be elements of ϕ. It follows from (c) that if the elements of ϕ_j, j=1,...,m-1 are all zero, the elements of ϕ_m will be zero as well. The elements of ϕ_1 are zero because they meet the requirements of (c). By induction, then the elements of every subset ϕ_m are all zero, and the bandwidth of G is no greater than p.

The algorithm proposed here uses eq. B(3) to solve for each $g_{m,n}$. The limits of this equation are changed slightly to take advantage of the bandedness of G:

$$g_{m,n} = \frac{1}{g_{m,m}} [\gamma_{m,n} - \sum_{\substack{j=1 \\ \text{and} \\ j<m}}^{m-n+p-1} g_{m-j,m} \times g_{m-j,n}] \qquad B(4)$$

Note that each $g_{m,n}$ is expressed in terms of a diagonal entry plus entries of previous rows. It follows that $g_{m,n}$ will be the only undetermined variable in B(4) if computations start with $g_{1,1}$ and proceed from left to right across each row.

Except for the first p rows where the k<m requirement sometimes restricts the summation limits, the computation of each element $g_{m,n}$ requires m-n+p-1 multiplications to complete. The total for each row m is

$$\sum_{j=m}^{m+p} m-j+p-1 \approx \frac{1}{2}p^2 .$$

APPENDIX C

ROOTS OF SYMMETRIC POLYNOMIALS OF EVEN DEGREE

Let p(x) be a polynomial of degree 2n,

$$p(x) \equiv a_0 + a_1 x + a_2 x^2 \ldots + a_{2n} x^{2n} \qquad C(1)$$

such that

$$a_{2n} = a_0,\ a_{2n-1} = a_1 \ldots a_{n-1} = a_{n+1}\ . \qquad C(2)$$

Theorem: The 2n roots of p(x) can be divided into n pairs, with each root in every pair being the reciprocal of its associate.

Proof: If p(r) = 0, then

$$a_0 + a_1 r + a_2 r \ldots + a_{2n} r^{2n} = 0 \qquad C(3)$$

Dividing by r^{2n} gives

$$a_0 r^{-2n} + a_1 r^{-(2n-1)} \ldots + a_{2n} = 0 \qquad C(4)$$

or

$$a_{2n} r^{-2n} + a_{2n-1} r^{-(2n-1)} \ldots + a_0 = 0. \qquad C(5)$$

Consequently $p(r^{-1}) = 0$.

It follows that r and r^{-1} form a reciprocal root pair except for when r=1 or -1. It will now be shown that if -1 or 1 are roots of p(x), they are roots of even multiplicity.

First assume that 1 is a root of p(x). It follows that

$$\sum_{i=0}^{2n} a_i = 0 = \sum_{i=0}^{n-1} 2a_i + a_n \qquad C(7)$$

Now consider the derivative of p(x) evaluated at x=1:

$$p'(x)\Big|_{x=1} = \sum_{i=0}^{2n} i a_i = \sum_{i=0}^{n-1} (i + (2n-i))\ a_i + n a_i$$

$$= n\ [\sum_{i=0}^{n-1} 2a_i + a_n] = 0 \qquad \text{C(8)}$$

Thus 1 is a root of $p'(x)$ if it is a root of p(x). It follows that 1 has a multiplicity of at least 2 if it is a root at all.

Next assume that -1 is a root. It follows that

$$\sum_{i=0}^{2n} (-1)^i\ a_i = 0 = \sum_{i=0}^{n-1} (-1)^i\ 2a_i + (-1)^n\ a_n. \qquad \text{C(9)}$$

Now consider the derivative of p(x) evaluated at x = -1.

$$p'(x)\Big|_{x=-1} = \sum_{i=0}^{2n} i a_i\ (-1)^{i-1} = \sum_{i=0}^{n-1} (-1)^{i-1}\ (i + 2n-i))\ a_i + (-1)^{n-1} n a_n$$

$$= -n[\sum_{i=0}^{n-1} (-1)^i\ 2a_i + (-1)^n\ a_n]$$

$$= 0 \qquad \text{C(10)}$$

Thus -1 has a multiplicity of at least 2 if it is a root of p(x).

Now define $q(x) = (x-r)\ (x - \frac{1}{r})$

$$= x^2 - (r + \frac{1}{r})x + 1 \qquad \text{C(11)}$$

and let

$$p'(x) = p(x)/q(x) \qquad \text{C(12)}$$

where r is a root of p(x). It should be clear that p(x) is divisible by q and that p' will be a polynomial of degree 2n-2. If $p'(x)$ is symmetric, then by induction the roots of p consist of n reciprocal pairs.

To show that p' is symmetric, let $a'_1 \ldots a'_{2(n-1)}$ denote the coefficients of p' and consider the equation

$$q(x)\, p'(x) = p(x). \qquad C(13)$$

Equating coefficients of x gives

1) $a_{2n} = a'_{2n-1}$

$a_0 = a'_0 \qquad \rightarrow \quad a'_{2(n-1)} = a'_0$

2) $a_{2n-1} = -(r + \frac{1}{r})\, a'_{2(n-1)} + a'_{2(n-1)-1}$

$a_1 = -(r + \frac{1}{r})\, a'_0 + a'_1 \qquad \rightarrow \quad a'_{2(n-1)-1} = a'_1$

3) $a_{2n-2} = a'_{2(n-1)} - (r + \frac{1}{r})\, a'_{2(n-1)-1} + a'_{2(n-1)-2}$

$a_2 = a'_0 - (r + \frac{1}{r})\, a'_1 + a'_2 \qquad \rightarrow \quad a'_{2(n-1)-2} = a'_2$

.

.

.

n-2) $a_{n+2} = a'_{n+2} - (r + \frac{1}{r})\, a'_{n+1} + a'_n$

$a_{n-2} = a'_{n-4} - (r + \frac{1}{r})\, a'_{n-3} + a'_{n-2} \qquad a'_n = a'_{n-2}\ . \qquad C(14)$

APPENDIX D

THE SYMMETRIC DECOMPOSITION OF SPECIALLY FORMED SYMMETRIC POLYNOMIALS OF EVEN DEGREE

Let $g^i(x)$ define a series of real polynomials of degree n_i. Let

$$f(x) = \sum_{i=1}^{p} g^i(x)\; g^i(x^{-1}) \qquad D(1)$$

Theorm: f(x) can be expressed in the form

$$f(x) = c(x)\; c(x^{-1}) \qquad D(2)$$

where c(x) is a unique real polynomial of degree $m = \max\,(n_1 \ldots n_p)$ whose roots fall on or outside the unit circle.

Proof: Let $h(x) = x^m f(x)$ D(3)

From D(1) it follows that h(x) is a real, symmetric polynomial of degree 2m. From the previous theorem, h(x) has m reciprocal root pairs. Thus,

$$f(x) = x^{-m}\, h_m \prod_{i=1}^{m} (x-d_i)(x-d_i^{-1}) = x^{-m}\, f_m \prod_{i=1}^{m} (x-d_i)(x-d_i^{-1}) \qquad D(4)$$

where the d_i are the roots of h(x), h_m is the constant multiplying the x^{2m} term in h(x), and f_m is the constant multiplying the x^m term in f(x).

Lemma:

All roots of f(x) that fall on the unit circle are of even multiplicity.

Proof:

Let r be a point on the unit circle. Let $\bar{r}$ be the complex conjugate of r. It follows that $\bar{r} = r^{-1}$ and that

$$g^i(r)\; g^i(r^{-1}) = g^i(r)\; g^i(\bar{r})\;. \qquad D(5)$$

Now express g^i as a product of its roots:

$$g^i(x) = K \prod_{j=1}^{a} (x-R_j) \prod_{k=1}^{b} (x-I_k)(x-\overline{I}_k) \qquad D(6)$$

where a is the number of real roots of g^i,

b is the number of complex-conjugate root pairs of g^i,

$a + 2b = n_j$,

R_j is the jth real root of $g^i(x)$,

I_k is one member of the kth complex root pair of $g^i(x)$, and

K is a constant.

Thus,

$$g^i(r)\ g^i(\overline{r}) = K^2 \prod_{j=1}^{a} (r-R_j)(\overline{r}-R_j) \prod_{k=1}^{b} (r-I_k)(\overline{r}-I_k)(\overline{r}-\overline{I}_k)(r-\overline{I}_k)$$

$$= K^2 \prod_{j=1}^{a} |r-R_j|^2 \prod_{k=1}^{b} |r-I_k|^2\ |\overline{r}-I_k|^2$$

$$\geq 0. \qquad D(7)$$

Now suppose that f(x) has a root, r, which lies on the unit circle, i.e.,

$$f(r) = \sum_{i=1}^{p} g^i(r)\ g^i(r)^{-1} \qquad D(8)$$

$$= \sum_{i=1}^{p} g^i(r)\ g^i(\overline{r})$$

$$= 0.$$

As all the terms in the above summation are non-negative by D(7), it follows that $g^i(r)\ g^i(\overline{r}) = 0$ for all i. Thus, r is a root of either $g^i(x)$ or $g^i(x^{-1})$. Suppose r is not complex, i.e., r=1 or -1. It follows that r is a root of multiplicity k of both $g^i(x)$ and $g^i(x^{-1})$, and is a root of $g^i(x)\ g^i(x^{-1})$ of multiplicity 2k. If r is complex and is a root or $g^i(x)$ of multiplicity k, then $\bar{r}$ or r^{-1} is a root of $g(x)$ of multiplicity k because $g^i(x)$ is real. It follows that r will be a root of $g^i(x^{-1})$ of multiplicity k and of $g^i(x)\ g^i(x^{-1})$ of multiplicity 2k. A similar argument produces the identical conclusion when one begins with r being complex and a root of $g(x^{-1})$ of multiplicity k. It follows that if r is a root of f(x), it is also a root of even multiplicity of each of the products $g^i(x)\ g^i(x^{-1})$. By the distributive property then, r is a root of f(x) of even multiplicity.

Now separate the 2m roots of f(x) into two groups. In one group include those roots that fall outside the unit circle, plus one half of each set of roots that fall on the unit circle. By the reciprocal property of even symmetric polynomials, both groups will have m members, each with a reciprocal counterpart in the second group. It should be clear that the division is unique. Now, let the d_i in D(4) correspond to the m d_i in the "outer" group, and carry out the following algebraic manipulations:

$$f(x) = x^{-m} f_m \prod_{i=1}^{m} (x-d_i)(x-d_i^{-1})$$

$$= f_m \prod_{i=1}^{m} (x-d_i)(1-(xd_i)^{-1})$$

$$= f_m \prod_{i=1}^{m} d_i^{-1} (-1)^m (x-d_i)(x^{-1}-d_i) \qquad D(9)$$

Let

$$c(x) = \sqrt{(-1)^m f_m \prod_{i=1}^{m} d_i^{-1}} \prod_{j=1}^{m} (x-d_j) \quad . \qquad D(10)$$

Thus,

$$f(x) = c(x)\ c(x^{-1}). \qquad D(11)$$

To see that the coefficients of $c(x)$ are real, reason as follows. All complex roots of $f(x)$ outside the unit circle have complex conjugates that are also outside the unit circle. Also, these complex roots will pair off because $f(x)$ is real. All complex roots that lie on the unit circle will pair off with conjugates that also lie on the unit circle. Because exactly half of these roots and half of their conjugates are included among the m d_i group, all the complex roots in the group will pair off with their conjugates. It follows that the coefficients of $\prod_{j=1}^{m} (x-d_i)$ are real.

It remains to be shown that $(-1)^m f_m \prod_{i=1}^{m} d_i^{-1}$ is positive. Let f_0 denote the constant coefficient in $f(x)$ and let e_j denote the coefficients of $\prod_{j=1}^{m} (x-d_i)$. From D(1):

$$f_0 = \sum_{i=1}^{p} \sum_{j=1}^{n_i} g_j^{i\,2} > 0. \qquad \text{D(12)}$$

Thus,

$$f_0 = (-1)^m f_m \prod_{i=1}^{m} d_i^{-1} \prod_{j=0}^{m} e_j^{2} > 0$$

$$= (-1)^m f_m \prod_{i=1}^{m} d_i^{-1} > 0. \qquad \text{D(13)}$$

APPENDIX E

THE SQUARE ROOT DECOMPOSITION OF COVARIANCE MATRICES OF N CONSECUTIVE RANDOM VARIABLES GENERATED FROM A MOVING AVERAGE PROCESS

Let h(x) be a real polynomial of order m with no roots on the unit circle:

$$h(x) = \sum_{i=o}^{m} h_i \; x^i \qquad E(1)$$

and let

$$\gamma(x) = \sum_{i=-m}^{m} \gamma_i \; x^i$$

$$= h(x) \; h(x^{-1}). \qquad E(2)$$

Let Γ be a matrix of order N>m where the jth element in the ith row of Γ is γ_{i-j}. Defined as such, Γ represents the covariance matrix of N consective samples of a random series y_i generated from

$$y_i = \sum_{j=0}^{m} h_j \phi_{i-j} \qquad E(3)$$

where ϕ_i represents a unit variance white noise process.

Now define an Nx(N+m) matrix H such that if h_{ij} defines the elements of H,

$$h_{i,j} = 0 \qquad i>j$$

$$= h_{j-i} \qquad j \geq i. \qquad E(4)$$

From the above definitions

$$\Gamma = H^T H \qquad E(5)$$

Equation E(5) demonstrates that Γ is positive definite. According to Forsythe and Moler (1967), this property insures that Γ can be uniquely decomposed into the product of an upper triangular matrix with positive diagonal elements, G, and its transpose:

$$\Gamma = G^T G \qquad E(6)$$

Let g_{ij} denote the jth element in the ith row of G.

Now from the previous theorem, there exists a unique polynomial g(x) of order m, with all roots on or outside the unit circle, such that

$$g(x)\ g(x^{-1}) = \gamma(x). \qquad E(7)$$

Note that as h(x) has no roots on the unit circle, neither do $\gamma(x)$ or g(x). Let g_j denote the jth coefficient of g(x).

Theorem:

$$\lim_{n\to\infty} [g_{n,n+j}] = g_j \qquad E(8)$$

In other words, the entries on and above the diagonal in the nth column (or on and to the right of the diagonal in the nth row) of G approach the coefficients of g(x) as n grows large.

Proof: As G is upper triangular with non-zero diagonal elements, G has an inverse, G^{-1}.

From E(6)

$$\Gamma G^{-1} = G^T \qquad E(9)$$

Now consider the following series of linear equations:

$$\sum_{j=1}^{N} \gamma_{i-j}\ \phi_j = U_i \qquad i=1,\ldots\ldots,N \qquad E(10)$$

where $\phi_j = 0$ for $j\leq 0$, $j>N$, $U_1,\ldots\ldots U_n$ correspond to the elements in the nth column of G^T, and all the U_j outside these bounds remain unrestricted. From E(9) it should be clear that the ϕ_j will correspond to the entries in the nth column of G^{-1}. As G^{-1} is upper triangular, $\phi_j=0$ for $j>n$. Also, $U_1,\ldots\ldots U_{n-1}$ will be zero due to the triangularity of G.

Taking the z transform of E(10) gives:

$$\gamma(z)\ \phi(z) = U(z) \qquad E(11)$$

or

$$\phi(z) = U(z)/\gamma(z). \qquad E(12)$$

The zero structure the γ_i and ϕ_i and the triangularity of G^T implies that $U_i=0$ for $0<i<n$, $i<-m+1$, or $i>n+m$. Furthermore, as a polynomial in Z, $\phi(z)$ can have no poles. It follows that the numerator in E(12) must have the same roots as the denominator. This provides 2m equations for the 2m+1 unknown U_j. Specifically, the U_j must satisfy

$$\sum_{j=0}^{-m+1} U_j a^j + \sum_{j=n}^{n+m} U_j a^j = 0 \qquad E(13)$$

where a is a particular root of $\gamma(x)$. It should be clear that

$$\lim_{n\infty} \left[\sum_{j=0}^{-m+1} U_j a^j + \sum_{j=n}^{n+m} U_j a^j \right] = \sum_{j=-0}^{-m+1} U_j a^j = 0 \qquad E(14)$$

for $|a|<1$. If any root with a norm less than 1 has a multiplicity $k>1$ the following equations must also be satisfied

$$\sum_{j=-0}^{-m+1} [jx(j-1)x\ldots(j-p+2)] \; U_j a^{j-p+1} = 0 \qquad p=2,\ldots.k \qquad E(15)$$

The m roots of $\gamma(x)$ within the unit circle will provide m linearly independent, homogenous equations for $U_0 \ldots U_{-m+1}$ through E(14) and E(15). Obviously,

$$\lim_{n\to\infty} (U_0 \ldots U_{-m+1}) = (0,\ldots,0) \qquad E(16)$$

Thus

$$\sum_{j=n}^{n+m} U_j a^j = 0 \qquad E(17)$$

for all a such that $|a|>1$, i.e., the remaining U_j form a polynomial with the same roots as g(x).

Consequently,

$$\lim_{n\to\infty} [U(z)] = kz^n g(z) \qquad E(18)$$

and

$$\lim_{n\to\infty} [\phi(x)] = kZ^n g(z)/\gamma(z)$$
$$= kz^n g(z)/g(z)g(z^{-1})$$
$$= kz^n g(z^{-1})^{-1} \qquad E(19)$$

Equating coefficients of z gives:

$$\lim_{n\to\infty} [U_{n+j}] = \lim_{n\to\infty} [g_{n,n+j}] = kg_j \qquad E(20)$$

and

$$\lim_{n\to\infty} [\phi_{n-j}] = \lim_{n\to\infty} [g^{-1}_{n-j,n}] = kg_j^{-1} \qquad E(21)$$

where g_j^{-1} is the jth coefficient of $g(x)^{-1}$ and g^{-1}_{ij} is the jth entry in the ith row of G^{-1}.

It remains to be shown that k=1. By definition, $U_n \ldots U_{n+m}$ will correspond to the elements on and to the right of the diagonal in the nth row of G. Also, $\phi_1 \ldots \phi_n$ will correspond to the first n elements in the nth column of G^{-1}. Thus $\phi_n U_n=1$ because $GG^{-1}=I$. From E(20), $U_n=kg_0$ and from E(21) $\phi_n = k^{-1}g_U$.Consequently,

$$k^2 = 1/g_0 g_0^{-1}.$$

Finally,

$$g(x)\ g(x)^{-1} = 1. \qquad E(22)$$

Equating the constant coefficients on either side of E(22) gives the desired result:

$$g_0 g_0^{-1} = 1.$$