ESTIMATING AND TESTING THE PARAMETERS OF A GENERALIZATION OF THE FIRST ORDER NONSTATIONARY AUTOREGRESSIVE PROCESS

by

Darryl Jon Downing

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF
THE UNIVERSITY OF FLORIDA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA
1974

TO BARBARA, DARREN, AND KELLY
FOR THEIR LOVE, UNDERSTANDING,
AND CONSTANT SUPPORT.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude and heartfelt thanks to Dr. John Saw. He suggested this topic to me, was always available for assistance, and without his help I would have never completed this work. Appreciation is expressed also to the other members of my supervisory committee, Professors D. T. Hughes, M. C. K. Yang, and Z. R. Pop-Stojanovic.

A special thank you is given to Dr. William Mendenhall, who made it possible for me to come to the University of Florida. His concern for my welfare and my family's will always be remembered and appreciated.

To Libby Coker who typed this manuscript from a rough draft, I owe more than a simple thank you can express. Her dedication and perseverance will always be remembered and appreciated.

The years of study here were consummated by the whole of the Statistics Department. To all of the faculty, secretaries, and students I extend my thanks for making me feel wanted and welcome.

TABLE OF CONTENTS

(Table of Contents Continued)

Abstract of Dissertation Presented to the
Graduate Council of the University of Florida in Partial
Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

ESTIMATING AND TESTING THE PARAMETERS OF A
GENERALIZATION OF THE FIRST ORDER NONSTATIONARY
AUTOREGRESSIVE PROCESS

by

Darryl Jon Downing

August, 1974

Chairman: Dr. J. G. Saw
Major Department: Statistics

A stochastic process is represented as having two components. The first component is called drift and measures location. The second component, called noise, measures the variability of the stochastic process. This paper is concerned with estimating the noise process when the noise process is assumed to follow what we shall call a generalized first order nonstationary autoregressive process. The generalized first order autoregressive process is defined similar to the first order autoregressive process, except that the parameter relating two observations is different for each time point. In order to estimate these parameters it is necessary that the stochastic process be replicated a sufficient number of times.

A method of estimating the parameters is proposed and the broad attendant distribution theory is delineated, both in a general setting and for specific situations. The prop-

erties of these estimators are given and some tests of hypothesis concerning the parameters are investigated. In order to comment further on the value of the proposed estimators, we use as a benchmark the maximum likelihood estimators. Their properties are given and a critical comparison is made between them and the proposed estimators.

In any practical situation it will be necessary to decide whether or not the first order generalized autoregressive process is sufficiently accurate to describe the data. Therefore, a test of the adequacy of the model is given.

Finally, numerical results are obtained using a computer simulation. The proposed estimators and the maximum likelihood estimators are compared. Also a practical application is given.

Chapter I

STATEMENT OF THE PROBLEM

1.1 Introduction

The statistical model in this dissertation is a stochastic process $\{Y(t):t\varepsilon T\}$. Usually T will denote a time interval and we shall suppose that replications of the process can be monitored during T at times $t_0<t_1<t_2<\ldots<t_m$, a typical replicate yielding the random sample $y_j=Y(t_j):0\leq j\leq m$. It is not necessary that the time increments t_1-t_0, t_2-t_1, $\ldots$, t_m-t_{m-1} be of equal length.

If we write $\mu(t)=EY(t)$ and $X(t)=Y(t)-\mu(t)$, then we may think of $\mu(t)$ as the "drift" of the sample paths of Y(t) and X(t) as "noise". Clearly $EX(t)=0$. Various schemes have, classically, been used to describe the noise process. In particular one may assume that, with $y_j=\mu(t_j)+x_j:0\leq j\leq m$,

$$x_j = a_1x_{j-1}+a_2x_{j-2}+\ldots+a_px_{j-p}+\varepsilon_j \; : \; p\leq j\leq m, \tag{1.1.1}$$

where ε_p, ε_{p+1}, $\ldots$, ε_m are independent identically distributed random variables.

In order that the data lend themselves to analysis under this classical model, several assumptions must be made. The most restrictive of these is that of stationarity. Expressed informally, stationarity assumes that the process has been running a sufficiently long time so that it has settled down.

Putting this into a probabilistic context, stationarity implies that the probability distribution of x_{t_1}, x_{t_2}, ..., x_{t_k} is the same as the probability distribution of x_{t_1+t}, x_{t_2+t}, ..., x_{t_k+t} for every finite set of values $(t_1, t_2, \ldots, t_k)$ and for every finite t.

The classical analysis of the model of equation (1.1.1), known as the p-th order autoregressive model is likely to be inappropriate in many cases due to the requirement of stationarity. For example, consider observing the effect of a diet on weight loss. Initially the weight loss will be greatest and will tend toward zero as time goes on and the subject tends to some constant weight. Obviously since the larger values appear first the probability distribution of the initial observations is not the same as that occurring later. A second example is the effect of drug infusion. A patient is given a dose of some drug, either orally or intravenously, and blood samples are drawn at various times t_0, t_1, ..., t_m thereafter. The amount of drug in the blood is then measured for each sample at each time. Again the initial readings will be larger than the later ones since the drug will be absorbed into the system or discharged as time goes on.

It may be argued in both examples cited, that successive differences (or perhaps successive second differences) have, approximately, a stationary distribution. Rather than concede to ad hoc procedures we prefer to replace the stationary autoregressive process by a nonstationary process. We gain this generality in the model for $\{Y(t) : t\varepsilon T\}$ at the expense of requiring (for our analysis of, and estimation of the parameters of the process) several replications of this process. Fortunately, in many instances of interest, replicates will be available.

The simplest alternative to the p-th order autoregressive scheme is what we shall call the first order generalized autoregressive scheme. Formally we shall assume that the errors $x_1, x_2, \ldots, x_m$ satisfy

$$x_j = \alpha_j x_{j-1} + \varepsilon_j \; : \; 1 \le j \le m, \tag{1.1.2}$$

where again $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m$ are independent identically distributed random variables. It will be assumed that the joint distribution of the errors is multivariate normal. The assumption that the process can be replicated is needed in order to estimate the unknown parameters $\alpha_1, \alpha_2, \ldots, \alpha_m$.

In the analysis of this model we shall be concerned with three major problems: (1) providing estimators for the unknown parameters, (2) finding the distribution of the estimators and comparing them to other estimators (typically likelihood estimators) and, (3) providing methodology for the testing the goodness of fit of the model.

1.2 Summary of Results

A method of estimating the parameters is proposed in Chapter 2 and the broad attendant distribution theory is delineated, both in a general setting and for specific situations. The properties of these estimators are provided in Chapter 3. In Chapter 4 a critical comparison is made between the maximum likelihood estimators and the estimators which we propose as an alternative. In any practical situation it will be necessary to decide whether or not the first order generalized autoregressive scheme is sufficiently accurate to describe the data. A decision procedure bearing on this aspect is investigated in Chapter 5.

An application of the theory is given in Chapter 6, with a comparison between the estimators derived in this dissertation and the usual maximum likelihood estimators.

1.3 Notation

In almost all areas of statistics notation is very important - consistant notation aids in solving and understanding the material presented. Also it is very convenient to abbreviate the distributional properties of random variables. For these reasons, certain notational conventions have been adopted. Although many of these are standard, they will be listed here for reference.

1. An underscored lower case letter invariably represents a column vector. Its row dimension will be given the first time a vector appears. Thus $\underline{x}$:(m) denotes a column vector consisting of m elements. The vector which has all its elements zero will be denoted by $\underline{0}$.

2. Matrices will be denoted by capital letters and the first time a matrix appears, its row and column dimensions will be given. Thus M:(rxc) denotes a

matrix with r rows and c columns. Denote the zero matrix by (0). The symbol "I" will be reserved for the identity matrix.

3. The elements of a matrix will be denoted by the corresponding small letter with subscripts to denote their row and column position. Thus m_{ij} denotes the element in the i^{th} row and j^{th} column of the matrix M. The symbol $(M)_{ij}$ is equivalent to m_{ij} and is sometimes substituted for convenience.

4. It is sometimes convenient to form row and column vectors from a given matrix. The symbol $(M)_{i\cdot}$ will represent the row vector formed from the i^{th} row of M. Similarly $(M)_{\cdot j}$ will denote the column vector formed from the j^{th} column of M.

5. The matrix formed from the first j rows and columns of M will be denoted by $M_{(j)}$. $M_{(j)}$ is commonly called the j^{th} principal submatrix.

6. The Kronecher product of M:(mxm) and N:(nxn) will be denoted by $M \otimes N$ and is a matrix P:(mnxmn) with

$$p_{(k-1)s+i,(\ell-1)s+j} = m_{k\ell} n_{ij} .$$

7. Diagonal matrices will be denoted by $D=\mathrm{diag}(d_1,d_2,\ldots,d_m)$, where diag is short for diagonal and $d_1,d_2,\ldots,d_m$ are the elements on the diagonal.

8. In keeping with conventional notation we shall write etr(.) to denote the constant "e" (Euler's constant) raised to the tr(.) power.

9. In distribution theory transformations are often made use of. To denote the Jacobian of the transformation the following notation $J\{X \to Y\}$ will represent the Jacobian of the transformation from the X-space into the Y-space.

10. If y is a random variable having a normal density, with mean μ, and variance σ^2, we will write

$$y \sim N(\mu, \sigma^2) \; .$$

11. If y is a random variable defined on (o, ∞) with density

$$f(y) = \{\Gamma(\tfrac{1}{2}\nu)\; 2^{\frac{1}{2}(\nu-2)}\}^{-1} y^{\nu-1} \exp\{-\tfrac{1}{2} y^2\} \quad ,$$

we will denote this by

$$y \sim \chi_\nu(0) .$$

This is to be read as "y has the central Chi density on ν degrees of freedom."

12. If y is a random variable defined on (o, ∞) with density

$$f(y) = \{\Gamma(\tfrac{1}{2}\nu)(2\sigma^2)^{\frac{1}{2}\nu}\}^{-1} \; y^{\frac{1}{2}\nu-1} \; e^{-y/2\sigma^2} \quad ,$$

we shall abbreviate this by

$$y \sim \sigma^2\chi^2(0) .$$

13. If $\underline{y}$ is an m-dimensional column vector whose elements have a joint normal density with mean vector, $\underline{\mu}$:(m) and dispersion matrix V:(mxm), this will be denoted by

$$\underline{y} \sim N_m(\underline{\mu}, V) \; .$$

14. If $\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_n$ are mutually independent m-variate column vectors with

$$\underline{y}_i \sim N_m(\underline{\mu}, V) \quad i=1,\ldots,n \; .$$

Then, with $Y = (\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_n)$, we will write

$$Y \sim N_{m\times n}(M, V\otimes I) \; ,$$

where $M = (\underline{\mu}, \underline{\mu}, \ldots, \underline{\mu})$.

15. With Y an mxn matrix such that

$$Y \sim N_{m\times n}(M, V\otimes I) \; .$$

the mxm matrix

$$W=YY'$$

has a noncentral Wishart distribution with dispersion matrix V, degrees of freedom n, and noncentrality matrix MM'. We will write

$$W \sim \mathcal{W}_m(V,n,MM')$$

16. If W is an mxm symmetric matrix whose $m(m+1)/2$ mathematically independent elements have the density

$$f(W) = k_m(V,\nu) \; |W|^{\frac{1}{2}(\nu-m-1)} \; \mathrm{etr}\{-\tfrac{1}{2}V^{-1}W\}$$

over the group of positive definite matrices, where

$$K_m^{-1}(V,\nu) = 2^{\frac{1}{2}m\nu}\pi^{\frac{1}{4}m(m-1)} \; |V|^{\frac{1}{2}\nu} \prod_{j=1}^{m} \Gamma(\tfrac{1}{2}(\nu-j+1)),$$

we will write

$$W \sim \mathcal{W}_m(V,\nu,(0)) .$$

17. For referencing within the text [$\cdot$] will denote bibliographical references, while ($\cdot$) will denote references to equations. Thus [4] refers to the fourth entry in the bibliography, while (1.2.3) refers to equation 3 in section 2 of Chapter I.

Chapter II

THE DOOLITTLE DECOMPOSITION AND ASSOCIATED DISTRIBUTION THEORY

2.1 Introduction

In the event that the "noise" process, $X(t)=Y(t)-\mu(t)$, is nonstationary, the general models and methods of estimation, like those given by Box and Jenkins [5] and Anderson [2], are no longer valid. We ignore the cases where the nonstationarity is caused by trend, since this can be removed and the resulting series is stationary and usual methods apply. An appropriate model for a "noise" process with this irregular behavior is

$$x_j = \alpha_j x_{j-1} + \varepsilon_j \ : \ 1 \le j \le m \ .$$

Our purpose in this chapter is to find a method of estimating the parameters of this model and to establish the attendent distribution theory.

Throughout, $\{y_j : 0 \le j \le m\}$, will denote the observed values of $Y(t)$ at times $t_0 < t_1 < \ldots < t_m$. We shall assume that y_j is normally distributed.

2.2 $\tilde{V}$: A Class of Dispersion Matrices

If $\{X(t) = Y(t) - \mu(t) : t \varepsilon T\}$ is a nonstationary time series, whose realizations, x_j, satisfy the relationship

$$x_j = \alpha_j x_{j-1} + \varepsilon_j \ : \ 1 \le j \le m, \qquad (2.2.1)$$

then $\{X(t)\}$ is said to follow the first order generalized autoregressive sequence.

Suppose we arrive at the (m+1)-variate column vector $\underline{x} = (x_0, x_1, \ldots, x_m)'$, obtained from random sampling from the above process. We assume that $\underline{x}$ is an observation from the (m+1)-variate normal distribution, that is,

$$\underline{x} \sim N_{m+1}(\underline{\mu}, V) \quad , \tag{2.2.2}$$

where $\underline{\mu} = \mathcal{E}\underline{x} = \underline{o}$, since $\mathcal{E}x(t) = o$, and

$$V = \text{Var } \underline{x} = \mathcal{E}\underline{xx}' \ .$$

In order to determine V we assume that x_o, ε_1, ..., ε_m are uncorrelated random variables with

$$\text{Var}(x_o) = \sigma^2\beta^2 \tag{2.2.3}$$

and

$$\text{Var}(\varepsilon_j) = \sigma^2 \ : 1 \le j \le m \ .$$

Since the process follows equation (2.2.1) we can write $\{x_0, x_1, \ldots, x_m\}$ in terms of $\{x_0, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m\}$. Applying equation (2.2.1) recursively we find the relationship between x_j and $x_0, \varepsilon_1, \ldots, \varepsilon_m$. Letting $\underline{x}:(m+1) = (x_0, x_1, \ldots, x_m)'$, $\underline{\varepsilon}:(m) = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m)'$ and $A:(m+1 \times m+1)$ having elements

$$\begin{aligned} a_{ii} &= 1 && : o \le i \le m \\ a_{ij} &= \alpha_j \alpha_{j-1} \cdots \alpha_i && : o \le j < i \le m \\ a_{kl} &= o && : \text{elsewhere} \ , \end{aligned} \tag{2.2.4}$$

then it is easily verified that

$$\underline{x} = A \begin{bmatrix} x_o \\ -- \\ \underline{\varepsilon} \end{bmatrix} . \tag{2.2.5}$$

From expression (2.2.5) we see that

$$V = AUA' \ , \tag{2.2.6}$$

where $U:(m+1 \times m+1) = \text{diag}(\sigma^2\beta^2, \sigma^2, \sigma^2, \ldots, \sigma^2)$. (2.2.7)

Writing out the elements of V explicitly we have

$$v_{oo} = \sigma^2\beta^2$$

$$v_{jj} = \sigma^2 + \sigma_j^2 v_{j-1,j-1} \qquad : 1\le j\le m \tag{2.2.8}$$

$$v_{jk} = v_{kj} = \alpha_{j+1}\alpha_{j+2} \cdots \alpha_k v_{jj} \quad : 1\le j<k\le m \ .$$

In the density of $\underline{x}$ we need $\Lambda=V^{-1}$, because of the form of V, Λ has a particularly simple form. By taking the inverse of the product in (2.2.6) we find

$$\Lambda = (A^{-1})'U^{-1}A^{-1} \ . \tag{2.2.9}$$

The inverse of U is trivial and since A is lower triangular its inverse is easily shown to have elements:

$$a^{ii} = 1 \qquad : o\le j\le m$$

$$a^{j+1,j} = -\alpha_{j+1} \quad : o\le j\le m-1 \tag{2.2.10}$$

$$a^{jk} = o \qquad : \text{elsewhere} \ .$$

Hence the elements of Λ are given by

$$\sigma^2\lambda_{oo} = \beta^{-2} + \alpha_1^2 \ ; \ \sigma^2\lambda_{mm} = 1$$

$$\sigma^2\lambda_{jj} = 1 + \alpha_{j+1}^2 \qquad : 1\le j\le m-1$$

$$\sigma^2\lambda_{j+1,j} = \sigma^2\lambda_{j',j+1} = -\alpha_{j+1} \quad : o\le j\le m-1 \tag{2.2.11}$$

$$\sigma^2\lambda_{jk} = o \qquad : \text{elsewhere} \ .$$

A square matrix M, with $m_{ij} = 0 : |i-j|>1$, is called a "Jacobi matrix" in the literature. This matrix can be factored into $\sigma^2\Lambda = R'R$, where R is lower triangular with

$$r_{oo} = \beta^{-1}$$

$$r_{jj} = 1 \qquad : 1\le j\le m$$

$$r_{j,j-1} = -\alpha_j \quad : 1\le j\le m \tag{2.2.12}$$

$$r_{jk} = 0 \qquad : \text{elsewhere} \ .$$

This result can be obtained from (2.2.9) very easily.

One further property of V is: let $V_{(j)}$ be the $(j+1 \times j+1)$ principal submatrix formed from V, then

$$|V_{(j)}| = \sigma^{2(j+1)}\beta^2 \quad : \ 0 \le j \le m \ . \qquad (2.2.13)$$

This result can be obtained by partitioning the matrices in (2.2.6) and taking the determinant of the corresponding product of the partitioned matrices. Since A is lower triangular with unit diagonal elements any square partition has determinant equal to unity. The square partition of U is diagonal with determinant equal to (2.2.13).

We note that the special form of V and its properties are due to the model. We shall let $\tilde{V}$ denote the class of matrices with this special form. Specifically, $\tilde{V}$ is the class of all positive definite matrices V such that V^{-1} is a Jacobi matrix. To see that $V \varepsilon \tilde{V}$ implies V^{-1} is a Jacobi matrix we note that V may always be represented by AUA' where A is lower triangular given by (2.2.4). Since A^{-1} is given by (2.2.10) then the resulting product $(A^{-1})'U^{-1}A^{-1}$ is always a Jacobi matrix.

2.3 The Doolittle Decomposition and Its Jacobian

Suppose we observe a number of (m+1)-variate column vectors $\underline{y}_j$: $1 \le j \le n$ $(n > m+1)$, obtained by random sampling from an (m+1)-variate normal population with mean $\underline{\mu}$ and dispersion matrix V. It is well known that the maximum likelihood estimates of $\underline{\mu}$ and V are given by

$$\underline{\hat{\mu}} = \underline{\bar{y}} = \frac{1}{n}\sum_{j=1}^{n} \underline{y}_j \qquad (2.3.1)$$

and

$$\hat{V} = \frac{1}{n} W \tag{2.3.2}$$

where W is the (m+1 x m+1) matrix

$$W = \sum_{j=1}^{n} (\underline{y}_j - \bar{\underline{y}})(\underline{y}_j - \bar{\underline{y}})' \quad . \tag{2.3.3}$$

W has the central Wishart distribution with $\nu=n-1$ degrees of freedom and dispersion matrix V. It is well known that $\mathcal{E}W = \nu V$, so that $\nu^{-1}W$ is an unbiassed estimate of V. In later sections we shall assume that $V \epsilon \tilde{V}$ so that V may be written as $V=AUA'$, where A and U are defined in section 2. We note that the sub-diagonal of A is $(\alpha_1, \alpha_2, \ldots, \alpha_m)$ with other elements being products of the α's. Since A contains all the information on the α's and U contains information on σ^2 and β^2, if we could estimate these matrices we would have estimates of the unknown parameters. Since W estimates V, perhaps a transformation on the element of W will give us estimates of A and U. It is with this intuitive notion in mind that we proceed. We assume that a matrix W is available with ν degrees of freedom, and for the moment, that V is an arbitrary positive definite matrix.

For convenience we label the rows and columns of W zero through m (rather than 1 through m+1) and let $W_{(j)}$ be the (j+1 x j+1) principal submatrix of W. Define

$$\begin{aligned} d_o &= |W_{(o)}| \\ d_j &= |W_{(j)}| \cdot |W_{(j-1)}|^{-1} \ : \ 1 \le j \le m \quad . \end{aligned} \tag{2.3.4}$$

With $D:(m+1 \times m+1) = \mathrm{diag}(d_0, d_1, \ldots, d_m)$ define $G:(m+1 \times m+1)$, a lower triangular matrix with unit diagonal elements (uniquely) by

$$W = GDG' . \tag{2.3.5}$$

Since the (m+1) random diagonal elements of D and the $\frac{1}{2}m(m+1)$ random elements $\{g_{ij}: o \le j < i \le m\}$ of G combine to give $\frac{1}{2}(m+2)(m+1)$ random variables we see that the transformation from W into G and D is nonsingular. The actual decomposition of W into G and D can be obtained using the forward Doolittle procedure outlined in Rao [14] and Saw [16].

We wish to determine the joint density of G and D. This can be obtained by using the density of W, denoted by f(W), and obtaining the Jacobian of the transformation from W into G and D defined by (2.3.5). Denoting this Jacobian by $J\{W \to G,D\}$, and the joint density on G and D by h(G,D) then

$$h(G,D) = f(GDG')J\{W \to G,D\} . \tag{2.3.6}$$

Direct evaluation of the Jacobian is cumbersome and the method used here is due to Hsu as reported by Deemer and Olkin [7]. Since the derivation of the Jacobian is rather long the rest of this section is devoted to it.

We seek the Jacobian of the transformation from W to G and D defined by

$$W = GDG', \tag{2.3.7}$$

with all matrices (m+1 x m+1) and G lower triangular with unit diagonals. Let $[\delta G]$ and $[\delta D]$, both (m+1 x m+1), denote small changes in G and D, respectively. Suppose that the changes $[\delta G]$ in G and $[\delta D]$ in D bring about a change $[\delta W]$ in W so that (2.3.5) is preserved. That is

$$W+[\delta W] = (G+[\delta G])(D+[\delta D])(G+[\delta G])' . \tag{2.3.8}$$

Expanding equation (2.3.8) and dropping terms of second order in the $[\delta *][*\varepsilon(G,D)]$, we find that

$$W+[\delta W] = GDG' + [\delta G]DG' + G[\delta D]G' + GD[\delta G]' \ . \qquad (2.3.9)$$

Since $W = GDG'$, we see that

$$[\delta W] = [\delta G]DG' + G[\delta D]G' + GD[\delta G]' \ . \qquad (2.3.10)$$

Hsu has shown that

$$J\{W \to G,D\} = J\{[\delta W] \to [\delta G],[\delta D]\}, \qquad (2.3.11)$$

where $J\{[\delta W] \to [\delta G],[\delta D]\}$ is the Jacobian of the transformation defined by (2.3.10), in which G and D are considered to be fixed (m+1 x m+1) matrices. In essence we have gone from a non-linear transformation in G and D into a linear transformation in the differential elements $[\delta G]$ and $[\delta D]$.

Pre and post multiplying (2.3.10) by G^{-1} and $(G')^{-1}$, respectively gives

$$G^{-1}[\delta W](G')^{-1} = G^{-1}[\delta G]D + [\delta D] + D[\delta G]'(G')^{-1} \ . \qquad (2.3.12)$$

Let $A = G^{-1}[\delta W](G')^{-1}$,

$B = G^{-1}[\delta G]$, (2.3.13)

and $C = [\delta D]$.

We note that A is symmetric, B is lower triangular with $(B)_{ii} = 0{:}0 \le i \le m$ and C is diagonal. We may rewrite (2.3.12) as

$$A = BD + C + DB' \ . \qquad (2.3.14)$$

From equations (2.3.10), (2.3.12), (2.3.13), and (2.3.14) we have

$$\begin{aligned} J\{W \to G,D\} &= J\{[\delta W] \to [\delta G], [\delta D]\} \\ &= J\{[\delta W] \to A\} \cdot J\{A \to B,C\} \cdot J\{B,C \to [\delta G],[\delta D]\}. \end{aligned} \qquad (2.3.15)$$

We shall evaluate the last three Jacobians separately.

The Jacobian, $J\{[\delta W] \to A\}$, is the Jacobian of the first transformation defined by (2.3.13). This can be evaluated by usual methods and we find

$$J\{[\delta W]\to A\} = |G|^{(m+2)} = 1 \quad , \tag{2.3.16}$$

since G is lower triangular with unit diagonals.

The Jacobian, $J\{B,C\to[\delta G],[\delta D]\}$, is the Jacobian of the transformation defined by the last two equations in (2.3.13). Hence it may be factored into the product of two Jacobians, namely,

$$J\{B,C\to[\delta G],\ [\delta D]\} = J\{B\to[\delta G]\}\cdot J\{C\to[\delta D]\}. \tag{2.3.17}$$

By the usual methods for determining Jacobians we find

$$J\{B\to[\delta G]\} = |G^{-1}|^{(m+1)} = 1 \ , \tag{2.3.18}$$

and $$J\{C\to[\delta D]\} = |I|^{(m+1)} = 1 \tag{2.3.19}$$

so that equation (2.3.17) is unity.

Finally we need to determine $J\{A\to B,C\}$. Writing out the equations given by (2.3.14) and using the fact that B is lower triangular with zero diagonal elements we find

$$a_{ji} = a_{ij} = b_{ij}d_j \ : \ 0\le j<i\le m \tag{2.3.20}$$

and $$a_{jj} = c_{jj} \qquad : \ 0\le j\le m \quad .$$

Hence we find that

$$\frac{\partial a_{jj}}{\partial c_{jj}} = 1 \qquad : \ 0\le j\le m$$

and

$$\frac{\partial a_{ij}}{\partial b_{ij}} = d_j \qquad : \ 0\le j<i\le m \ , \tag{2.3.21}$$

so that the Jacobian is

$$J\{A\to B,C\}= \left\|\frac{\partial(a_{oo}a_{10}a_{20}\cdots a_{m0}a_{11}a_{21}\cdots a_{m1}\cdots a_{mm})}{\partial(c_{oo}b_{10}b_{20}\cdots b_{mo}c_{11}b_{21}\cdots b_{m1}\cdots c_{mm})}\right\|$$

$$= \prod_{j=o}^{m} d_j^{m-j} \quad . \tag{2.3.22}$$

Following equation (2.3.15) we obtain

$$J\{W \to G,D\} = \prod_{j=o}^{m} d_j^{m-j} \quad . \tag{2.3.23}$$

2.4 The Joint Distribution of G and D for Arbitrary V and in Particular When $V \varepsilon \tilde{V}$.

In section 3 we derived the Jacobian of the non-linear transformation $W = GDG'$. We now suppose that, for arbitrary positive definite V,

$$W \sim \mathcal{W}_{m+1}(V,\nu,(0)) . \tag{2.4.1}$$

The joint density on G and D is then obtained from (2.4.1), (2.3.6), and (2.3.23).

$$h(G,D) = K_{m+1}(V,\nu)\,\mathrm{etr}\{-\tfrac{1}{2}V^{-1}GDG'\} \prod_{j=0}^{m} d_j^{\frac{1}{2}(\nu+m)-j-1} \tag{2.4.2}$$

on

$$d_j > o \;:\; o \leq j \leq m$$

and $-\infty < g_{ij} < \infty \;:\; o \leq j < i \leq m$.

The term $K_{m+1}(V,\nu)$ is defined by

$$K_{m+1}^{-1}(V,\nu) = |V|^{\frac{1}{2}\nu} (2)^{\frac{1}{2}\nu(m+1)} \pi^{\frac{1}{4}m(m+1)} \prod_{j=0}^{m} \Gamma(\tfrac{1}{2}(\nu-j)) . \tag{2.4.3}$$

With $\Lambda = V^{-1}$ we may write,

$$\begin{aligned} \mathrm{tr}\{-\tfrac{1}{2}V^{-1}GDG'\} &= \mathrm{tr}\{-\tfrac{1}{2}\Lambda GDG'\} \\ &= \mathrm{tr}\{-\tfrac{1}{2}DG'\Lambda G\} \\ &= -\tfrac{1}{2}\sum_{j=0}^{m} d_j (G'\Lambda G)_{jj} \;. \end{aligned} \tag{2.4.4}$$

Hence we see that the density in (2.4.2) partitions into the subsets $\{d_0, g_{10}, g_{20}, \ldots, g_{m0}\}$; $\{d_1, g_{21}, g_{31}, \ldots, g_{m1}\}$; ...; $\{d_{m-1}, g_{m,\,m-1}\}$; $\{d_m\}$ which are mutually independent, but variables within a subset are dependent.

In the case that $V\varepsilon\tilde{V}$ then we may write $\sigma^2\Lambda = R'R$ from equation (2.2.12) and we find

$$(G'\Lambda G)_{jj} = \frac{1}{\sigma^2}(G'R'RG)_{jj} = \frac{1}{\sigma^2}(RG)'_{\cdot j}(RG)_{\cdot j}\,, \qquad (2.4.5)$$

where

$$(RG)'_{\cdot o} = (\beta^{-1}; g_{10}-\alpha_1; g_{20}-\alpha_2 g_{10}; \ldots; g_{mo}-\alpha_m g_{m,m-1})$$

and for $i \le j \le m-1$

$$(RG)'_{\cdot j} = (0;0;\ldots;1;g_{j+1,j}-\alpha_{j+1}; g_{j+2,j}-\alpha_{j+2}g_{j+1,j};\ldots; g_{mj}-\alpha_m g_{m-1,j})\,. \qquad (2.4.6)$$

The "1" appears as the j^{th} element in $(RG)_{\cdot j}$.

Now we may write

$$(G'\Lambda G)_{oo} = \frac{1}{\sigma^2}\{\beta^{-2} + \sum_{k=1}^{m}(g_{ko} - \alpha_k g_{k-1,o})^2\}$$

and for $1 \le j \le m-1$ (2.4.7)

$$(G'\Lambda G)_{jj} = \frac{1}{\sigma^2}\{1 + \sum_{k=j+1}^{m}(g_{k,j}-\alpha_k g_{k-1,j})^2\}\,.$$

The density $h(G,D)$ factors into

$$h(G,D) = \{\prod_{j=0}^{m-1} h_j(d_j, g_{j+1,j}, g_{j+2,j},\ldots,g_{mj})\}h_m(d_m)\,, \qquad (2.4.8)$$

where

$h_o(d_0, g_{10}, g_{20},\ldots,g_{mo}) =$

$$\frac{d_o^{\frac{1}{2}(\nu+m)-1}\,\mathrm{etr}\{-\tfrac{1}{2}d_o[\beta^{-2}+\sum_{k=1}^{m}(g_{ko}-\alpha_k g_{k-1,0})^2]\}}{\Gamma(\tfrac{1}{2}\nu)\,2^{\frac{1}{2}(\nu+m)}\,\pi^{\frac{1}{2}m}\,|V|^{\frac{1}{2}}\,|V_{(0)}|^{\frac{1}{2}(\nu-1)}} \qquad (2.4.9)$$

and for $1 \le j \le m-1$

$$h_j(d_j, g_{j+1,j}, \ldots, g_{mj}) =$$

$$\frac{d_j^{\frac{1}{2}(\nu+m)-j-1} \operatorname{etr}\{-\tfrac{1}{2} d_j [1 + \sum_{k=j+1}^{m} (g_{kj} - \alpha_k g_{k-1,j})^2]\}}{\Gamma(\tfrac{1}{2}(\nu-j)) 2^{\frac{1}{2}(\nu+m)-j} \pi^{\frac{1}{2}(m-j)} |V|^{\frac{1}{2}} |V_{(j)}|^{\frac{1}{2}(\nu-j-1)} |V_{(j-1)}|^{-\frac{1}{2}(\nu-j)}},$$

(2.4.10)

and $$h_m(d_m) = \{\Gamma(\tfrac{1}{2}(\nu-m))(2\sigma^2)^{-\frac{\nu-m}{2}}\}^{-1} d_m^{\frac{1}{2}(\nu-m)-1} \exp[-\tfrac{1}{2}\sigma^{-2} d_m].$$

(2.4.11)

2.5 The Distribution of G, of D, and of G Conditional on D When V Is Arbitrary and When $V \varepsilon \tilde{V}$.

The necessity of knowing the distribution of $d_0; d_1; \ldots; d_m$ and the subsets $\{g_{10}; g_{20}; \ldots; g_{mo}\}$; $\{g_{21}; g_{31}; \ldots; g_{m1}\}; \ldots; \{g_{m-1,m}\}$ arises from the fact that functions of these statistics will be estimators for the parameters σ^2, β^2, and $\{\alpha_1, \alpha_2, \ldots, \alpha_m\}$. A knowledge of the distribution of the estimators gives us the information we need to talk about the "goodness" of the estimators. It is to this end that we derive the distribution of G, D, and G conditional on D.

The distribution on the elements of D follow directly from a theorem given in class lecture notes and in Saw [15].

Theorem 2.5.1

If $W \sim \mathcal{W}_{m+1}(V, \nu, (0))$, ν integer with $\nu > m$; and $W_{(r)}$ is defined by

$$W_{(r)} = \begin{bmatrix} w_{00} & w_{01} & \cdots & w_{0r} \\ w_{10} & w_{11} & \cdots & w_{1r} \\ \vdots & \vdots & & \vdots \\ w_{r0} & w_{r1} & & w_{rr} \end{bmatrix} \quad : 0 \le r \le m .$$

Then $\{|W_{(0)}|;\ \{|W_{(r)}|/|W_{(r-1)}|\}_{r=1}^{r=m}\}$ are independent chi-square variates such that

$$|W_{(0)}| \sim |V_{(0)}| \chi_{\nu}^{2}\ (o) \tag{2.5.2}$$

and for $1 \leq r \leq m$

$$|W_{(r)}|/|W_{(r-1)}| \sim |V_{(r)}|/|V_{(r-1)}|\ \chi_{\nu-r}^{2}\ (o). \tag{2.5.3}$$

Since from (2.3.4) we have

$$d_o = |W_{(0)}|$$

$$d_j = |W_{(j)}|/|W_{(j-1)}| :\ 1 \leq j \leq m\ ,$$

then by direct application of Theorem 2.5.1 we have

$$d_o \sim |V_{(0)}|\ \chi_{\nu}^{2}\ (o)$$

and for $1 \leq j \leq m$

$$d_j \sim |V_{(j)}|/|V_{(j-1)}|\ \chi_{\nu-j}^{2}\ (o)\ . \tag{2.5.4}$$

Now if we allow $V \epsilon \tilde{V}$, then since $|V_{(j)}| = \sigma^{2(j+1)}\beta^2$ we find

$$d_o \sim \sigma^2\beta^2\chi_{\nu}^{2}\ (o)$$

and for $1 \leq j \leq m$ (2.5.5)

$$d_j \sim \sigma^2\chi_{\nu-j}^{2}\ (o)\ .$$

To find the density on the subsets $\{g_{10};g_{20};\ldots;g_{m0}\}$; $\{g_{21};g_{31};\ldots;g_{m1}\};\ldots;\ \{g_{m,m-1}\}$ we refer back to equations (2.4.9) and (2.4.10). Using those equations we may write for $o \leq j \leq m-1$,

$$h_j(d_j, g_{j+1,j};\ldots;g_{m,j}) = C\ d_j^{\frac{1}{2}(\nu+m)-j-1} etr\{-\tfrac{1}{2}d_j(G'\Lambda G)_{jj}\}, \tag{2.5.6}$$

for some constant C. Performing the integration over d_j and replacing $(G'\Lambda G)_{jj}$ by $\sum_{k=j}^{m}\sum_{\ell=j}^{m}\lambda_{k\ell}g_{kj}g_{\ell j}$ we have,

$$h_j(g_{j+1,j};\ldots;g_{mj}) = C_m(\nu:V,j)\{\sum_{k=j}^{m}\sum_{\ell=j}^{m}\lambda_{k\ell}g_{kj}g_{\ell j}\}^{-\frac{1}{2}(\nu+m)+j}, \tag{2.5.7}$$

where, with $V_{(-1)}$ taken as unity

$$C_m(\nu:V,j) = \frac{\Gamma(\frac{1}{2}(\nu+m)-j)\,|V_{(j-1)}|^{\frac{1}{2}(\nu-j)}}{\Gamma(\frac{1}{2}(\nu-j))\,\Pi^{\frac{1}{2}(m-j)}\,|V|^{\frac{1}{2}}|V_{(j)}|^{\frac{1}{2}(\nu-j-1)}} . \tag{2.5.8}$$

Remembering that $g_{jj}=1$, a transformation shows the variables in (2.5.7) have a multivariate t-distribution. The form and properties of the multivariate t-distribution were found by Cornish [6] in 1954 and also in the same year by Dunnett and Sobel [8].

Now if we allow $V\varepsilon\tilde{V}$, we find

$$h_0(g_{10};g_{20};\ldots;g_{m0})=\beta^m C_m(\nu:I,0)\{1+\beta^2\sum_{k=1}^{m}(g_{k0}-\alpha_k g_{k-1,0})^2\}^{-\frac{1}{2}(\nu+m)} \tag{2.5.9}$$

and for $1\leq j\leq m-1$

$$h_j(g_{j+1,j};\ldots;g_{m,j})=C_m(\nu:I,j)\{1+\sum_{k=j+1}^{m}(g_{k,j}-\alpha_k g_{k-1,j})^2\}^{-\frac{1}{2}(\nu+m)+j} \tag{2.5.10}$$

Evidently $g_{10},g_{21},\ldots,g_{m,m-1}$ are mutually independent and

$$\nu^{\frac{1}{2}}\beta(g_{10}-\alpha_1) \sim t_\nu(0) \tag{2.5.11}$$

and

$$(\nu-k+1)^{\frac{1}{2}}(g_{k,k-1}-\alpha_k) \sim t_{\nu-k+1} \ :\ 2\leq k\leq m . \tag{2.5.12}$$

To find the distribution of G conditional on D let $|V_{(-1)}| = 1$, then the marginal distribution on D is

$$\begin{aligned} h(D) &= h(d_0)\cdot h(d_1)\ldots h(d_m) \\ &= \prod_{j=0}^{m} \frac{d_j^{\frac{1}{2}(\nu-j)-1}\exp\{-\frac{1}{2}\,|V_{(j)}|^{-1}|V_{(j-1)}|d_j\}}{(2|V_{(j)}|\,|V_{(j-1)}|^{-1})^{\frac{1}{2}(\nu-j)}\,\Gamma(\frac{1}{2}(\nu-j))} . \end{aligned} \tag{2.5.13}$$

We have from equations (2.4.2) and (2.4.4) that the joint density on (G,D) is

$$h(G,D)=k_{m+1}(V,\nu)\prod_{j=0}^{m}\{d_j^{\frac{1}{2}(\nu+m)-j-1}\exp\{-\tfrac{1}{2}d_j(G'\Lambda G)_{jj}\}\} .$$

The conditional distribution of G given D is

$$h(G|D) = \sum_{j=0}^{m} \left\{ \frac{d_j^{\frac{1}{2}(m-j)} \exp\{-\frac{1}{2}d_j[(G'\Lambda G)_{jj} - |V_{(j)}|^{-1}|V_{(j-1)}|]\}}{(|V||V_{(j)}|^{-1})^{\frac{1}{2}}(2\pi)^{\frac{1}{2}(m-j)}} \right\}. \quad (2.5.14)$$

Although this does not have a very pleasing form if we let $V\varepsilon\tilde{V}$, we find the conditional density simplifies greatly. If we write $G'G = (RG)'(RG)$ and use the fact that $|V_{(j)}| = \sigma^{2(j+1)}\beta^2$ we find

$$h(G|D) = \prod_{j=o}^{m-1} \{(2\pi\sigma^2 d_j^{-1})^{-\frac{1}{2}(m-j)} \exp[-\tfrac{1}{2}\sigma^{-2} d_j \sum_{k=j+1}^{m} (g_{kj} - \alpha_k g_{k-1,j})^2]\}. \quad (2.5.15)$$

Let $\underline{g}_j$ be the $(m-j)$-variate column vector given by

$$\underline{g}_j = (g_{j+1,j}; g_{j+2,j}; \ldots; g_{mj})' : 0 \le j \le m-1, \quad (2.5.16)$$

$\underline{\mu}_j$ the $(m-j)$ dimensional column vector defined by

$$\underline{\mu}_j = (\alpha_{j+1}; \alpha_{j+1}\alpha_{j+2}; \ldots; \alpha_{j+1}\alpha_{j+2}\cdots\alpha_m)' : o \le j \le m-1, \quad (2.5.17)$$

and V_j the $(m-j \times m-j)$ matrix whose elements are given by

$$(V_j)_{11} = \frac{\sigma^2}{d_j} \quad : 0 \le j \le m-1 ,$$

$$(V_j)_{kk} = \frac{\sigma^2}{d_j} + \alpha_{j+k}^2 (V_j)_{k-1,k-1} \quad : 2 \le k \le m-j;\ o \le j \le m-1, \quad (2.5.18)$$

and $(V_j)_{kl} = (V_j)_{lk} = \alpha_{j+k}\alpha_{j+k+1}\cdots\alpha_{j+l}(V_j)_{kk} : 1 \le k < l \le m-j;$
$o \le j \le m-1.$

Then equation (2.5.15) may be written as

$$h(G|D) = \prod_{j=o}^{m-1} h_j(\underline{g}_j | d_j) , \quad (2.5.19)$$

where for $o \le j \le m-1$

$$h_j(\underline{g}_j|d_j) = (2\pi)^{-\frac{1}{2}(m-j)} |V_j|^{-\frac{1}{2}} \exp[-\tfrac{1}{2}(\underline{g}_j - \underline{\mu}_j)' V_j^{-1} (\underline{g}_j - \underline{\mu}_j)].$$

That is (2.5.20)

$$\underline{g}_j | d_j \sim N_{m-j}(\underline{\mu}_j, V_j) \quad : o \le j \le m-1 . \quad (2.5.21)$$

2.6 Verification of the Distribution of G.

In finding the density of the subset $(g_{k+1,k};\ldots;g_{mk})$ the constant $C_m(\nu:V,k)$ was given, but was not verified. In this section we show how the value of the constant can be obtained.

We need to evaluate the integral

$$I_g = \int_{\{-\infty<u_i<+\infty:1\leq i\leq p\}} \Big(\sum_{i=0}^{p} \sum_{j=0}^{p} \alpha_{ij}u_iu_j\Big)^{-\beta} du_1du_2\ldots du_p \tag{2.6.1}$$

where $u_o=1$ and $\alpha_{ij}=\alpha_{ji}$.

We may write

$$\sum_{i=0}^{p} \sum_{j=0}^{p} \alpha_{ij}u_iu_j = (1\ \underline{u}')\begin{pmatrix}\alpha_{oo} & \underline{\alpha}' \\ \underline{\alpha} & A\end{pmatrix}\begin{pmatrix}1 \\ \underline{u}\end{pmatrix} \tag{2.6.2}$$

where $\underline{\alpha} = (\alpha_{01},\ \alpha_{02},\ \ldots,\alpha_{op})'$ (2.6.3)

and

$$A = \begin{bmatrix}\alpha_{11} & \alpha_{12} & \cdots & \alpha_{1p} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2p} \\ \vdots & \vdots & & \vdots \\ \alpha_{p1} & \alpha_{p2} & \cdots & \alpha_{pp}\end{bmatrix} . \tag{2.6.4}$$

Now equation (2.6.2) may be rewritten as

$$\sum_{i=0}^{p} \sum_{j=0}^{p} \alpha_{ij}u_iu_j = \alpha_{00}+\underline{u}'A\underline{u}+2\underline{u}'\underline{\alpha} \tag{2.6.5}$$

$$= \alpha_{00}+(\underline{u}+A^{-1}\underline{\alpha})'A(\underline{u}+A^{-1}\underline{\alpha})-\underline{\alpha}'A\underline{\alpha}. \tag{2.6.6}$$

Write $\underline{u} + A^{-1}\underline{\alpha} = (\alpha_{00}-\underline{\alpha}'A^{-1}\underline{\alpha})^{\frac{1}{2}}K\underline{w}$. (2.6.7)

where $KAK'=I$, so that $|K| = |A|^{-\frac{1}{2}}$. The Jacobian from $\underline{u}$

into $\underline{w}$ can be found by standard methods and is

$$J\{\underline{u}\to\underline{w}\} = (\alpha_{00}-\underline{\alpha}'A^{-1}\underline{\alpha})^{\frac{1}{2}p}|K| \tag{2.6.8}$$

$$= (\alpha_{00}-\underline{\alpha}'A^{-1}\underline{\alpha})^{\frac{1}{2}p}|A|^{-\frac{1}{2}}. \tag{2.6.9}$$

Hence we have

$$I_g = (\alpha_{00}-\alpha'A^{-1}\underline{\alpha})^{\frac{1}{2}p-\beta}|A|^{-\frac{1}{2}}\int_{\{-\infty<w_i<\infty:1\le i\le p\}}(1+\underline{w}'\underline{w})^{-\beta}\,dw_1\ldots dw_p\ . \tag{2.6.10}$$

We compare this with the following version of the multivariate t-distribution

$$f(\underline{t})=C_p|\Sigma|^{-\frac{1}{2}}[1+(\underline{t}-\underline{\theta})'\Sigma^{-1}(\underline{t}-\underline{\theta})]^{-\frac{1}{2}(n+p)}:-\infty<t_j<\infty,\ n>o\ ,\ 1\le j\le p \tag{2.6.11}$$

$$\text{where } C_p = \Gamma(\tfrac{1}{2}(n+p))/\{\pi^{\frac{1}{2}p}\Gamma(\tfrac{1}{2}n)\}\ . \tag{2.6.12}$$

Let $\Sigma=I$ and $\underline{\theta}=\underline{o}$, then setting $\beta=\frac{1}{2}(n+p)$ we see that $n=2\beta-p$, since $n>o$ we must have $\beta>\frac{1}{2}p$ and we find

$$\int_{\{-\infty<w_i<\infty:1\le i\le p\}}(1+w'w)^{-\beta}dw_1\ldots dw_p = \frac{\Gamma(\beta-\frac{1}{2}p)\pi^{\frac{1}{2}p}}{\Gamma(\beta)} \tag{2.6.13}$$

Hence,

$$I_g = (\alpha_{oo}-\underline{\alpha}'A^{-1}\alpha)^{\frac{1}{2}p-\beta}\ A^{-\frac{1}{2}}\ \frac{\Gamma(\beta-\ p)\pi^{\frac{1}{2}p}}{\Gamma(\beta)} \tag{2.6.14}$$

$$= \frac{\begin{vmatrix}\alpha_{oo} & \underline{\alpha}' \\ \underline{\alpha} & A\end{vmatrix}^{\frac{1}{2}p-\beta}\Gamma(\beta-\frac{1}{2}p)\pi^{\frac{1}{2}p}}{|A|^{\frac{1}{2}(p+1)-\beta}\Gamma(\beta)}\ . \tag{2.6.15}$$

Referring back to (2.5.8) we see that $p=m-j$; $\beta=\frac{1}{2}(\nu+m)-j$, so that

$$C_m(\nu:V,j) = \frac{\Gamma(\frac{1}{2}(\nu+m)-j)}{\Gamma(\frac{1}{2}(\nu-j))\pi^{\frac{1}{2}(m-j)}} \frac{|\Lambda_1|^{\frac{1}{2}(\nu-j)}}{|\Lambda_2|^{\frac{1}{2}(\nu-j-1)}} \tag{2.6.16}$$

where

$$\Lambda_1 = \begin{bmatrix} \lambda_{j,j} & \lambda_{j,j+1} & \cdots & \lambda_{j,m} \\ \lambda_{j+1,j} & \lambda_{j+1,j+1} & \cdots & \lambda_{j+1,m} \\ \vdots & \vdots & & \vdots \\ \lambda_{m,j} & \lambda_{m,j+1} & \cdots & \lambda_{m,m} \end{bmatrix} \tag{2.6.17}$$

and

$$\Lambda_2 = \begin{bmatrix} \lambda_{j+1,j+1} & \lambda_{j+1,j+2} & \cdots & \lambda_{j+1,m} \\ \lambda_{j+2,j+1} & \lambda_{j+2,j+2} & \cdots & \lambda_{j+2,m} \\ \vdots & \vdots & & \vdots \\ \lambda_{m,j+1} & \lambda_{m,j+2} & \cdots & \lambda_{m,m} \end{bmatrix} \tag{2.6.18}$$

Now with $\Lambda=V^{-1}$, then by an application of the clockwise rule

$$\begin{aligned} |V|^{-1} = |\Lambda| = \begin{vmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{vmatrix} &= |\Lambda_{22}|\,|\Lambda_{11}-\Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}| \\ &= |\Lambda_{22}|\,|V_{11}|^{-1}\ , \end{aligned} \tag{2.6.19}$$

hence we find

$$|\Lambda_{22}| = |V|^{-1}|V_{11}|\ . \tag{2.6.20}$$

So that

$$|\Lambda_1| = |V|^{-1} \cdot |V_{(j-1)}| \tag{2.6.21}$$

$$|\Lambda_2| = |V|^{-1} \cdot |V_{(j)}|\ , \tag{2.6.22}$$

and we confirm that

$$C_m(\nu:V,j) = \frac{\Gamma(\frac{1}{2}(\nu+m)-j)}{\Gamma(\frac{1}{2}(\nu-j))\pi^{\frac{1}{2}(m-j)}} \frac{|V_{(j-1)}|^{\frac{1}{2}(\nu-j)}}{|V|^{\frac{1}{2}}|V_{(j)}|^{\frac{1}{2}(\nu-j-1)}}\ .$$

Chapter III

THE ESTIMATORS σ_*^2, β_*^2, and $\{\alpha_{*j}:1\leq j\leq m\}$

3.1 Introduction

In Chapter II the unknown parameters σ^2, β^2, and $\{\alpha_j:1\leq j\leq m\}$ were introduced to define the generalized first order autoregressive process and thence the underlying distribution of the observations. In this chapter we propose estimators (alternatives to the maximum likelihood estimators) for these parameters and discuss their properties. (To distinguish between the estimators proposed in Chapter II and the maximum likelihood estimator, we shall reserve the hat (^) notation for the latter and star (*) notation for the former.)

Finally, tests of hypothesis concerning the parameters are discussed. Special attention is given to the case of testing $\alpha_k = \alpha_{k+1} = \cdots = \alpha_m = \eta_o$, for $k = 2,3,\ldots,m$. A method, due to Fisher, of combining independent tests is likely to be appropriate.

3.2 The Distribution and Properties of σ_*^2, β_*^2, and $\{\alpha_{*j}: 1\leq j\leq m.\}$

Suppose that we observe a number of (m+1)-variate column vectors $\underline{y}_j$: $1\leq j\leq n$ $(n>m+1)$, obtained by random sampling from an (m+1)-variate normal population with mean $\underline{\mu}$ and dispersion matrix V. We assume that these observations come from the process $Y(t) = \mu(t) + X(t)$, during times $t_0 < t_1 < \ldots < t_m$, where $X(t)$ follows the first order generalized autoregressive process. Since $\mathcal{E}\, Y(t) = \mu(t)$ we have, letting $\mu(t_i) = \mu_i$: $0 \leq i \leq m$, that

$$\hat{\mu}_i = \bar{y}_i = \frac{1}{n} \sum_{j=1}^{n} y_{ij} \; : \; 0 \leq i \leq m, \tag{3.2.1}$$

or alternatively with the $(m+1)$ dimensional column vector $\mu = (\mu_0, \mu_1, \ldots, \mu_m)'$, that

$$\hat{\underline{\mu}} = \bar{\underline{y}} = \frac{1}{n} \sum_{j=1}^{n} \underline{y}_j \,. \tag{3.2.2}$$

Estimates of the noise process $X(t)$ for $t_0 < t_1 < \ldots < t_m$ are

$$\underline{x}_j = \underline{y}_j - \bar{\underline{y}} \; : \; 1 \leq j \leq n \,. \tag{3.2.3}$$

From these we may arrive at

$$W = \sum_{j=1}^{n} \underline{x}_j \underline{x}_j' \tag{3.2.4}$$

$$= \sum_{j=1}^{n} (\underline{y}_j - \bar{\underline{y}})(\underline{y}_j - \bar{\underline{y}})' \,, \tag{3.2.5}$$

where W has the central Wishart distribution on $\nu = n-1$ degrees of freedom and dispersion matrix V (where $V \varepsilon \tilde{V}$). Hence the theory of Chapter II is applicable and we have from equation (2.5.5), (2.5.11), and (2.5.12) that

$$d_0 \sim \sigma^2 \beta^2 \chi_\nu^2 (0) \,, \tag{3.2.6}$$

$$d_j \sim \sigma^2 \chi_{\nu-j}^2 (0) \; : \; 1 \leq j \leq m \,, \tag{3.2.7}$$

$$\nu^{\frac{1}{2}} \beta (g_{10} - \alpha_1) \sim t_\nu (0) \,, \tag{3.2.8}$$

and $$(\nu-k+1)^{\frac{1}{2}} (g_{k,k-1} - \alpha_k) \sim t_{\nu-k+1}(0) : 2 \leq k \leq m \,. \tag{3.2.9}$$

Hence the estimators for $\{\alpha_j : 1 \leq j \leq m\}$ are:

$$\alpha_{*j} = g_{j,j-1} \; : \; 1 \leq j \leq m, \tag{3.2.10}$$

with densities given in equations (3.2.8) and (3.2.9).

The estimators for σ^2 and β^2 are:

$$\sigma_*^2 = \frac{1}{mc_1} \sum_{j=1}^{m} d_j \, , \tag{3.2.11}$$

where $c_1 = \nu - \frac{1}{2}(m+1)$ (3.2.12)

and

$$\beta_*^2 = d_0 (c_2 \sum_{j=1}^{m} d_j)^{-1} \tag{3.2.13}$$

where $c_2 = \nu(mc_1 - 2)^{-1}$. (3.2.14)

We note that σ_*^2 has a Gamma density and β_*^2 has a Beta Type 2 density.

In order to evaluate the "goodness" of these estimators the following properties are investigated: (1) the first two moments, (2) the consistency of the estimators, and (3) the efficiency of the estimators.

The first moment of the estimators are given by

$$\mathcal{E}(\alpha_{*j}) = \alpha_j \; : \; 1 \le j \le m \, , \tag{3.2.15}$$

$$\mathcal{E}(\sigma_*^2) = \sigma^2 \, , \tag{3.2.16}$$

and

$$\mathcal{E}(\beta_*^2) = \beta^2 \, , \tag{3.2.17}$$

so that the estimators are all unbiassed.

Letting Σ_* denote the (m+2 x m+2) variance-covariance matrix of the (m+2) dimensional vector of unbiassed estimators $(\alpha_{*1}, \alpha_{*2}, \ldots, \alpha_{*m}, \sigma_*^2, \beta_*^2)$, we find

$$(\Sigma_*)_{11} = \beta^{-2}(\nu-2)^{-1}, \tag{3.2.18}$$

$$(\Sigma_*)_{jj} = (\nu-j+1)^{-1} \; : \; 2 \le j \le m \, , \tag{3.2.19}$$

$$(\Sigma_*)_{m+1,m+1} = 2(mc_1)^{-1}\sigma^4 \, , \tag{3.2.20}$$

$$(\Sigma_*)_{m+2,m+2} = 2(\nu+mc_1-2)[\nu(mc_1-4)]^{-1}\beta^4 \, , \tag{3.2.21}$$

$$(\Sigma_*)_{m+1,m+2} = (\Sigma_*)_{m+2,m+1} = -2(mc_1)^{-1}\sigma^2\beta^2, \tag{3.2.22}$$

and

$$(\Sigma_*)_{ij} = 0 \; : \; \text{elsewhere} \; . \tag{3.2.23}$$

In Fisz [10] and Feller [9] it is shown that a statistic t, based on ν observations, with mean θ and variance τ_ν, will be consistent for θ if

$$\lim_{\nu\to\infty} \tau_\nu = 0 . \tag{3.2.24}$$

In the equations (3.2.18) thru (3.2.21) if we let $\nu\to\infty$, we find the variance of the estimator goes to zero. Hence the estimators $(\alpha_{*1}, \alpha_{*2}, \ldots, \alpha_{*m}, \sigma_*^2, \beta_*^2)$ are consistent.

The likelihood function, from which the estimators are derived, is the density of W, given by

$$L = f(W) = K_{m+1}(V,\nu)\,|W|^{\frac{1}{2}(\nu-m-2)}\operatorname{etr}\{-\tfrac{1}{2}V^{-1}W\} . \tag{3.2.25}$$

Using this we may obtain the "Information Matrix", F, whose elements are defined by

$$(F)_{j\ell} = \mathcal{E}\left[-\frac{\partial^2 \log L}{\partial\theta_j\ \partial\theta_\ell}\right] \tag{3.2.26}$$

where θ_j and θ_ℓ are any two of the parameters $\alpha_1,\ldots,\alpha_m$, σ^2, and β^2. The j^{th} diagonal element of F^{-1} gives the minimum varinace bound of any estimator of θ_j. Letting $(\theta_1,\theta_2, \ldots,\theta_m,\theta_{m+1},\theta_{m+2}) \equiv (\alpha_1,\alpha_2, \ldots,\alpha_m,\sigma^2,\beta^2)$ we find

$$(F)_{jj} = \mathcal{E}(\frac{1}{\sigma^2}w_{j-1,j-1})$$

$$= \frac{\nu}{\sigma^2}v_{j-1,j-1} \qquad : 1\le j\le m , \tag{3.2.27}$$

with $v_{j-1,j-1}=\sigma^2(1+\alpha_{j-1}^2+\alpha_{j-1}^2\alpha_{j-2}^2+\cdots+\alpha_{j-1}^2\alpha_{j-2}^2\cdots\alpha_2^2$

$$+\alpha_{j-1}^2\alpha_{j-2}^2\cdots\alpha_2^2\alpha_1^2\beta^2) , \tag{3.2.28}$$

$$(F)_{m+1,m+1} = \mathcal{E}\,[\frac{-\nu(m+1)}{2\sigma^4} + \frac{1}{\sigma^4}\operatorname{tr}V^{-1}W]$$

$$= \frac{\nu(m+1)}{2\sigma^4} , \tag{3.2.29}$$

$$(F)_{m+2,m+2} = \mathcal{E}\left[\frac{-\nu}{2\beta^4} + \frac{w_{oo}}{\sigma^2\beta^6}\right]$$

$$= \frac{\nu}{2\beta^4} \ , \qquad (3.2.30)$$

$$(F)_{m+1,m+2} = (F)_{m+2,m+1} = \mathcal{E}\left[\frac{w_{oo}}{2\sigma^4\beta^4}\right]$$

$$= \frac{\nu}{2\sigma^2\beta^2} \ , \qquad (3.2.31)$$

and

$$(F)_{j\ell} = 0\text{:elsewhere.} \qquad (3.2.32)$$

To see that $(F)_{j\ell} = 0$ elsewhere we note that

$$\frac{\partial^2 \log L}{\partial\theta_j\ \partial\theta_\ell} = 0 \quad \text{for } 1\le j\le m;\ 1\le \ell\le m;\ j \ne \ell , \qquad (3.2.33)$$

$$\frac{\partial^2 \log L}{\partial\theta_j\ \partial\theta_{m+1}} = -\frac{1}{\sigma^4}\left(\alpha_j w_{j-1,j-1} - w_{j-1,j}\right) \ : \ 1\le j\le m , \qquad (3.2.34)$$

and

$$\frac{\partial^2 \log L}{\partial\theta_j\ \partial\theta_{m+2}} = 0 \ : \ 1\le j\le m . \qquad (3.2.35)$$

Now $$\mathcal{E}\left(\frac{1}{\sigma^4}\left[\alpha_j w_{j-1,j-1} - w_{j-1,j}\right]\right) = \frac{\nu}{\sigma^4}\left(\alpha_j v_{j-1,j-1} - v_{j-1,j}\right)$$

$$= 0 , \qquad (3.2.36)$$

since $v_{j-1,j} = \alpha_j v_{j-1,j-1}$. Hence $(F)_{j\ell}$ is zero as was to be shown.

With $\Sigma = F^{-1}$ we find that the minimum variance bounds are

$$(\Sigma)_{jj} = (\nu\ v_{j-1,j-1})^{-1}\sigma^2 \ : \ 1\le j\le m , \qquad (3.2.37)$$

$$(\Sigma)_{m+1,m+1} = 2(\nu m)^{-1}\sigma^4 , \qquad (3.2.38)$$

$$(\Sigma)_{m+2,m+2} = 2(m+1)(\nu m)^{-1}\beta^4 , \qquad (3.2.39)$$

$$(\Sigma)_{m+1,m+2} = (\Sigma)_{m+2,m+1} = -2(\nu m)^{-1}\sigma^2\beta^2 , \qquad (3.2.40)$$

and $\quad (\Sigma)_{jk} = 0 \quad$: elsewhere . (3.2.41)

Since the efficiency of an estimator is the ratio of the minimum variance bound to the variance of the estimator, we have from equations (3.2.18) thru (3.2.23) and equations (3.2.37) thru (3.2.41) that the efficiency of the starred estimator, denoted by $\mathrm{Eff}(\theta_*)$, is

$$\mathrm{Eff}(\alpha_{*1}) = (\nu-2)\nu^{-1} , \tag{3.2.42}$$

$$\mathrm{Eff}(\alpha_{*j}) = (\nu-j-1)(\nu v_{j-1,j-1})^{-1}\sigma^2 : 2\le j \le m , \tag{3.2.43}$$

$$\mathrm{Eff}(\sigma_*^2) = c_1\nu^{-1} , \tag{3.2.44}$$

and

$$\mathrm{Eff}(\beta_*^2) = (m+1)\nu(mc_1-4)[\nu m(\nu+mc_1-2)]^{-1} . \tag{3.2.45}$$

The estimators α_{*1}, σ_*^2, and β_*^2 are asymptotically efficient while the asymptotic efficiency of α_{*j}: $2\le j\le m$ is given by

$$\lim_{\nu\to\infty} \mathrm{Eff}(\alpha_{*j}) = v_{j-1,j-1}^{-1}\sigma^2 : 2\le j\le m . \tag{3.2.46}$$

Replacing $v_{j-1,j-1}$ by the right hand side of equation (3.2.28) gives

$$\lim_{\nu\to\infty} \mathrm{Eff}(\alpha_{*j}) = (1+\alpha_{j-1}^2\alpha_{j-2}^2+\ldots+\alpha_{j-1}^2\alpha_{j-2}^2\ldots\alpha_{j-2}^2+\alpha_{j-1}^2\alpha_{j-2}^2\alpha_2^2+ \alpha_{j-1}^2\alpha_{j-2}^2\ldots\alpha_2^2\alpha_1^2\beta^2)^{-1} : 2\le j\le m . \tag{3.2.47}$$

Hence the asymptotic efficiency of α_{*j} depends on the true values of the previous α's and β^2. If the true value of α_{j-1} is zero then α_{*j} is asymptotically efficient regardless of the values of any of the previous α's. That is, the asymptotic efficiency of α_j is dependent most upon the true value of α_{j-1} the previous α, next upon α_{j-2}, and so forth with the least dependence on α_1 and β^2. If we suppose that all the parameters are less than unity in absolute value and write,

$$a = \max(|\alpha_1|, |\alpha_2|, \ldots, |\alpha_{j-1}|, |\beta|), \tag{3.2.48}$$

then we may arrive at the inequality

$$(1+a^2+a^4+\ldots+a^{2(j-2)}+a^{2j}) \geq (1+\alpha_{j-1}^2+\alpha_{j-1}^2\alpha_{j-2}^2+\ldots+ \alpha_{j-1}^2\alpha_{j-2}^2\ldots\alpha_2^2+\alpha_{j-1}^2\alpha_{j-2}^2\ldots\alpha_2^2\alpha_1^2\beta^2), \tag{3.2.49}$$

where equality holds only when $|\alpha_1| = |\alpha_2| = \ldots = |\alpha_{j-1}| = |\beta| = |a|$. The inequality reverses upon taking reciprocals and we find the asymptotic efficiency of α_{*j} is at least as great as

$$\min \mathrm{Eff}(\alpha_{*j}) = (1-a^{2(j-1)}+a^{2j}-a^{2(j+1)})^{-1}(1-a^2) \quad : \quad 2\leq j\leq m \;. \tag{3.2.50}$$

Although $\{\alpha_{*i}: 2\leq i\leq m\}$ is inefficient this loss in efficiency is more than made up for by the fact that the distribution of $\{(\alpha_{*i}-\alpha_i): 2\leq i\leq m\}$ contains no unknown parameters.

3.3 Tests of Hypothesis

Since the distribution of the estimators are known, tests of hypothesis may be carried out with ease. We tabulate here a few hypothesis of interest.

It is desired to test the hypothesis

$$H_o: \alpha_k = \alpha_{k+1} = \ldots = \alpha_m = \eta_o (k=2,3,\ldots,m)$$

against (3.3.1)

H_a: at least one of the equalities does not hold.

With

$$t_j = (\nu-j+1)^{\frac{1}{2}}(g_{j,j-1}-\eta_0) \; : \; 2\leq j\leq m \;, \tag{3.3.2}$$

define

$$p_j = P\{|t_{(\nu-j+1)}| \geq t_j\} \; : \; 2 \leq j \leq m. \qquad (3.3.3)$$

An appropriate test statistic for testing (3.3.1) is

$$L = -2 \sum_{j=k}^{m} \log_e(p_j). \qquad (3.3.4)$$

The quantity, L, has the chi-square distribution on 2(m-k+1) degrees of freedom. The hypothesis is rejected at significance level α if

$$L > \ell \qquad (3.3.5)$$

where ℓ is chosen so that

$$P\{\chi^2_{2(m-k+1)} \geq \ell\} = \alpha \; . \qquad (3.3.6)$$

This procedure is called Fisher's method of combining independent tests. It has been shown by Littell and Folks [12] to be asymptotically optimal over other tests as judged by Bahadur relative efficiency. The Bahadur relative efficiency compares the rates at which the competing test statistics observed significance levels converge to zero, in some sense, when the null hypothesis is false. The interested reader is referred to Bahadur [3] and Littell and Folks [12] .

The above hypothesis has some interesting interpretations for choicesof η_o. If $\eta_o = 0$, we are testing whether the process is white noise from some point k on. In the case where η_o is a constant, not equal to zero, we are hypothesizing that the time series is stationary.

Hypothesis concerning individual parameters can be carried out in the usual manner since the distribution of the estimator is known.

An hypothesis of importance, concerning a single parameter would be

$$H_o: \beta^2 = \beta_0^2$$

against (3.3.7)

$$H_a: \beta^2 \neq \beta_0^2.$$

An appropriate test statistic is

$$F_0 = mc_1\beta_*^2/(mc_1-2)\beta_0^2 \tag{3.3.8}$$

which has an F distribution on ν and mc_1 degrees of freedom, where c_1 is defined in equation (3.2.12) as $\nu-\frac{1}{2}(m+1)$. The null hypothesis is rejected at the α level of significance if

$$F_0 > F^{\nu}_{mc_1,\alpha} \tag{3.3.9}$$

where $F^{\nu}_{mc_1,\alpha}$ is chosen so that

$$P\ \{F^{\nu}_{mc_1} > F^{\nu}_{mc_1,\alpha}\} = \alpha\ . \tag{3.3.10}$$

Choosing $\beta_0^2 = 1$, the hypothesis implies homoscedasticity between the initial observation and the errors of the "noise" process.

Chapter IV

THE MAXIMUM LIKELIHOOD ESTIMATORS

4.1 Introduction

In order to comment further on the value of the estimators given in Chapter III some standard of comparison must be employed. To this end we study the maximum likelihood estimators. In this chapter we obtain the maximum likelihood estimators and examine their sampling properties. A comparison is then made between the maximum likelihood estimators and the starred estimators of Chapter III.

4.2 The Maximum Likelihood Estimators and Their Distribution

As in section 2 of Chapter III, we suppose that we observe a number of (m+1)-variate column vectors $\underline{y}_j$: $1 \le j \le n$ $(n>m+1)$, obtained by random sampling from an (m+1)-variate normal population with mean $\underline{\mu}$ and dispersion matrix V. As in Chapter III, we estimate $\underline{\mu}$ by $\bar{\underline{y}}$ and form the (m+1 x m+1) matrix W by

$$W = \sum_{j=1}^{n} (\underline{y}_j-\bar{\underline{y}})(\underline{y}_j-\bar{\underline{y}})' \ . \qquad (4.2.1)$$

W has the central Wishart distribution with $\nu=n-1$ degrees of freedom and dispersion matrix V. W may also be represented by

$$W = \sum_{j=1}^{\nu} \underline{z}_j\underline{z}_j' \qquad (4.2.2)$$

where $\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_\nu$ ($\nu=n-1$) are mutally independent and

$$\underline{z}_j \sim N_{m+1}(\underline{0}, V) \;:\; 1 \le j \le \nu . \tag{4.2.3}$$

To see this let the (m+1 x n) matrix Y be defined by

$$Y = (\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_n) \;, \tag{4.2.4}$$

and let B be any orthogonal (n x n) matrix with last column

$$\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right)' . \tag{4.2.5}$$

Define the (m+1 x n) matrix $Z = (\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_n)$ by

$$Z = YB . \tag{4.2.6}$$

We note that

$$\underline{z}_n = \sqrt{n}\,\bar{\underline{y}} . \tag{4.2.7}$$

Now W may be written as

$$W = YY' - n\bar{\underline{y}}\bar{\underline{y}}' , \tag{4.2.8}$$

and since B is orthogonal (BB'=I) we may write

$$\begin{aligned} YY' &= (YB)(YB)' \\ &= ZZ' , \end{aligned} \tag{4.2.9}$$

and upon substituting $YY' = ZZ'$ and $\underline{z}_n = \sqrt{n}\,\bar{\underline{y}}$ into (4.2.8) we find

$$\begin{aligned} W &= Z Z' - \underline{z}_n \underline{z}_n' \\ &= \sum_{j=1}^{n} \underline{z}_j \underline{z}_j' - \underline{z}_n \underline{z}_n' \\ &= \sum_{j=1}^{n-1} \underline{z}_j \underline{z}_j \end{aligned} . \tag{4.2.10}$$

Hence we have the representation

$$w_{ij} = \sum_{\ell=1}^{n} (y_{i\ell} - \bar{y}_i)(y_{j\ell} - \bar{y}_j)$$

$$= \sum_{k=1}^{\nu} Z_{ik} Z_{jk} \quad : \; o \leq i \leq m; \; o \leq j \leq m \; . \tag{4.2.11}$$

Hence forth we shall use $W = ZZ'$ keeping in mind the representation given in (4.2.11).

Assuming that $Y(t) = \mu(t) + X(t)$, where $X(t)$ follows the first order generalized autoregressive process then $V \varepsilon \tilde{V}$ and has the properties given in section 2 of Chapter II. Since it is through W that we obtain estimates, the elements of W serve as observations and the likelihood of this set of observations is $L = f(W)$ given in (3.2.25). Taking the logarithm of (3.2.25) and utilizing the form and properties of V we obtain

$$\log L = C - \tfrac{1}{2}\nu(m+1) \log \sigma^2 - \frac{\nu}{2} \log \beta^2$$

$$+ \tfrac{1}{2}(\nu-m-2) \log |W| - \tfrac{1}{2} \mathrm{tr} V^{-1} W \; , \tag{4.2.12}$$

where C is a constant. Recalling that $\Lambda = V^{-1}$ has the special form given by (2.2.11) we may write

$$\mathrm{tr} V^{-1} W = \mathrm{tr} \Lambda W$$

$$= \frac{1}{\sigma^2} \{ (\beta^2)^{-1} w_{00} + \sum_{j=1}^{m} w_{jj} + \sum_{j=1}^{m} (\alpha_j^2 w_{j-1,j-1} - 2\alpha_j w_{j-1,j}) \} \; . \tag{4.2.13}$$

Substituting (4.2.13) into (4.2.12) we find

$$\log L = C - \tfrac{1}{2}\nu(m+1) \log \sigma^2 - \frac{\nu}{2} \log \beta^2 + \tfrac{1}{2}(\nu-m-2) \log |W|$$
$$- \frac{1}{2\sigma^2} \{ (\beta^2)^{-1} w_{00} + \sum_{j=1}^{m} w_{jj} + \sum_{j=1}^{m} (\alpha_j^2 w_{j-1,j-1} - 2\alpha_j w_{j-1,j}) \} \tag{4.2.14}$$

Differentiation of (4.2.14) with respect to α_j yields

$$\frac{\partial \log L}{\partial \alpha_j} = -\frac{1}{\sigma^2} \{ \alpha_j w_{j-1,j-1} - w_{j-1,j} \} \quad : \; 1 \leq j \leq m \; , \tag{4.2.15}$$

differentiation with respect to σ^2 yields

$$\frac{\partial \log L}{\partial \sigma^2} = -\frac{\nu(m+1)}{2\sigma^2} + \frac{1}{2\sigma^4}\{(\beta^2)^{-1}w_{00} + \sum_{j=1}^{m} w_{jj}$$

$$+ \sum_{j=1}^{m} (\alpha_j^2 w_{j-1,j-1} - 2\alpha_j w_{j-1,j})\} , \tag{4.2.16}$$

and finally differentiation with respect to β^2 gives

$$\frac{\partial \log L}{\partial \beta^2} = \frac{-\nu}{2\beta^2} + \frac{w_{00}}{2\sigma^2\beta^4} . \tag{4.2.17}$$

Setting $\{\frac{\partial \log L}{\partial \alpha_j} : 1 \le j \le m\}$, $\frac{\partial \log L}{\partial \sigma^2}$, and $\frac{\partial \log L}{\partial \beta^2}$ equal to zero and solving we obtain the maximum likelihood estimators:

$$\hat{\alpha}_j = w_{j-1,j} w_{j-1,j-1}^{-1} \quad : 1 \le j \le m , \tag{4.2.18}$$

$$\hat{\sigma}^2 = [\nu(m+1)]^{-1}\{\beta^{-2}w_{00} + \sum_{j=1}^{m} (w_{jj} - w_{j-1,j}^2 w_{j-1,j-1}^{-1})\}, \tag{4.2.19}$$

and

$$\hat{\beta}^2 = (\nu\hat{\sigma}^2)^{-1} w_{00} . \tag{4.2.20}$$

Eliminating $\hat{\beta}^2$ from equation (4.2.19) yields

$$\hat{\sigma}^2 = (\nu m)^{-1} \sum_{j=1}^{m} (w_{jj} - w_{j-1,j}^2 w_{j-1,j-1}^{-1}) . \tag{4.2.21}$$

We now proceed to determine the distribution of the maximum likelihood estimators. In order to do this we shall use conditional arguments frequently. We shall write

$$y|z \sim f(\cdot) \tag{4.2.22}$$

in order to imply that "y conditional on z has the density ... ." Using the representation of w_{ij} given in (4.2.11) one has

$$\hat{\alpha}_j = \frac{\sum_{k=1}^{\nu} z_{j-1,k} z_{jk}}{\sum_{k=1}^{\nu} z_{j-1,k}^2} : 1 \le j \le m . \tag{4.2.23}$$

Letting

$$\phi_{jk} = \frac{z_{jk}}{\sum_{k=1}^{\nu} z_{jk}^2} : o \le j \le m ; 1 \le k \le \nu , \tag{4.2.24}$$

we may write

$$\hat{\alpha}_j = \sum_{k=1}^{\nu} \phi_{j-1,k} z_{jk} : 1 \le j \le m . \tag{4.2.25}$$

Recalling that

$$\underline{z}_k \sim N_{m+1}(\underline{0}, V) : 1 \le k \le \nu \tag{4.2.26}$$

when $V \varepsilon \tilde{V}$, we must be able to represent z_{jk} by

$$z_{jk} = \alpha_j z_{j-1,k} + \varepsilon_{jk} : 1 \le j \le m; 1 \le k \le \nu , \tag{4.2.27}$$

where ε_{jk} are independent identically distributed normal random variables with mean zero and variance σ^2.

Hence

$$z_{jk} | z_{j-1,k} \sim N(\alpha_j z_{j-1,k}, \sigma^2) : 1 \le j \le m; 1 \le k \le \nu . \tag{4.2.28}$$

Using (4.2.28) in (4.2.25),

$$\hat{\alpha}_j | \{z_{j-1,k} : 1 \le k \le \nu\} \sim N(\alpha_j \sum_{k=1}^{\nu} \phi_{j-1,k} z_{j-1,k} ; \sigma^2 \sum_{k=1}^{\nu} \phi_{j-1,k}^2) . \tag{4.2.29}$$

Since

$$\sum_{k=1}^{\nu} \phi_{j-1,k} z_{j-1,k} = \frac{\sum_{k=1}^{\nu} z_{j-1,k}^2}{\sum_{k=1}^{\nu} z_{j-1,k}^2} = 1 \tag{4.2.30}$$

and

$$\sum_{k=1}^{\nu} \phi_{j-1,k}^2 = \frac{1}{\sum_{k=1}^{\nu} z_{j-1,k}^2} , \tag{4.2.31}$$

we find upon their substitution into (4.2.29) that

$$\hat{\alpha}_j|\{z_{j-1,k}: 1\le k\le\nu\}\sim N(\alpha_j,\sigma^2(\sum_{k=1}^{\nu} z_{j-1,k}^2)^{-1}). \qquad (4.2.32)$$

To complete the derivation we note the following.

Let

$$u \sim N(\mu,\sigma^2) \qquad (4.2.33)$$

and

$$v \sim \sigma^2\chi_\nu^2(0) , \qquad (4.2.34)$$

then Student's t-distribution is defined as

$$t = (u-\mu)(\frac{1}{\nu} v)^{-\frac{1}{2}} . \qquad (4.2.35)$$

We note that the distribution of t conditional on v is

$$t|v\sim N(0,\nu\sigma^2 v^{-1}) . \qquad (4.3.36)$$

Hence, by analogy, the distribution of $\hat{\alpha}_j$ is also t and we have

$$\nu^{\frac{1}{2}}\left(\frac{v_{j-1,j-1}}{\sigma^2}\right)^{\frac{1}{2}}(\hat{\alpha}_j-\alpha_j)\sim t_\nu(0): 1\le j\le m , \qquad (4.2.37)$$

where $\frac{v_{oo}}{\sigma^2} = \beta^2$,

and

$$\frac{v_{j,j}}{\sigma^2} = (1+\alpha_{j-1}^2+\alpha_{j-1}^2\alpha_{j-2}^2+\ldots+\alpha_{j-1}^2\alpha_{j-2}^2\ldots\alpha_2^2 +$$
$$\alpha_{j-1}^2\alpha_{j-2}^2\ldots\alpha_2^2\alpha_1^2\beta^2) : 1\le j\le m . \qquad (4.2.38)$$

To find the distribution of $\hat{\sigma}^2$ we define the (νx1) column vectors $\underline{\theta}_j = (\theta_{j1},\theta_{j2}, \ldots, \theta_{j\nu})'$ by

$$\theta_{jk} = (\sum_{k=1}^{\nu} z_{jk}^2)^{-\frac{1}{2}} z_{jk} : 1\le k\le\nu;\ o\le j\le m , \qquad (4.2.39)$$

and note that

$$\underline{\theta}_j'\ \underline{\theta}_j = 1 : 0\le j\le m . \qquad (4.2.40)$$

In terms of the (m+1 x ν) matrix Z we have

$$\underline{\theta}_j = \{(Z)_{j.}(Z)_{j.}'\}^{-\frac{1}{2}}(Z)_{j.}' \qquad (4.2.41)$$

For convenience we take

$$\eta_j = w_{jj} - w^2_{j-1,j} w_{j-1,j-1} \quad : \; 1 \le j \le m \; , \qquad (4.2.42)$$

and write

$$\hat{\sigma}^2 = \frac{1}{m\nu} \sum_{j=1}^{m} \eta_j \; . \qquad (4.2.43)$$

Using $\underline{\theta}_j$ we may write

$$\eta_j = (Z)_{j.} (I_\nu - \underline{\theta}_{j-1} \underline{\theta}'_{j-1}) (Z)'_{j.} \quad : \; 1 \le j \le m \; , \qquad (4.2.44)$$

where I_ν is the $(\nu x \nu)$ identity matrix. Consider the distribution of η_m conditional on $\{(Z)_{j.} : 0 \le j \le m-1\}$. Now

$$(Z)_{m.} | \{(Z)_{j.} : \; 0 \le j \le m-1\} \sim N_{\nu x 1} (\alpha_m (Z)_{m-1,.} ; \sigma^2 I_\nu) . \qquad (4.2.45)$$

Conditional on $\{(Z)_{j.} : \; o \le j \le m-1\}$ the matrix $(I_\nu - \underline{\theta}_{m-1} \underline{\theta}'_{m-1})$ is symmetric and idempotent with rank $(\nu-1)$, and the quadratic form η_m follows the non-central chi-square distribution, that is

$$\eta_m | \{(Z)_{j.} : \; 0 \le j \le m-1\} \sim \sigma^2 \chi^2_{\nu-1} (\gamma_m) \qquad (4.2.46)$$

where

$$\gamma_m = \frac{1}{2\sigma^2} \alpha^2_m \; (Z)_{m-1,.} (I_\nu - \underline{\theta}_{m-1} \underline{\theta}'_{m-1}) (Z)'_{m-1,.} \quad . \qquad (4.2.47)$$

Upon replacing $\underline{\theta}_{m-1}$ by its definition in (4.2.47) we find

$$\gamma_m = 0 \; . \qquad (4.2.48)$$

Since none of $\{(Z)_{j.} \; : \; 0 \le j \le m-1\}$ enter into the distribution of η_m we have the unconditional distribution is the same and hence η_m is independent of $\{(Z)_{j.} \; : \; o \le j \le m-1\}$, that is

$$\eta_m \sim \sigma^2 \chi^2_{(\nu-1)} (0) \quad . \qquad (4.2.49)$$

The distribution of $\eta_{(m-1)}$ conditional on $\{(Z)_{j.} : \; o \le j \le m-2\}$ can be obtained in exactly the same manner and since it depends only on $\{(Z)_{j\cdot} : 0 \le j \le m-2\}$ it is independent of η_m. By the same arguments as above $\eta_{(m-1)}$ can be shown to be independent of $\{(Z)_{j\cdot} : o \le j \le m-2\}$ and hence

$$\eta_{(m-1)} \sim \sigma^2 \chi^2_{(\nu-1)} \ (0) \ . \tag{4.2.50}$$

In exactly the same manner we can show η_{k+1} is independent of η_k and hence so is $\eta_{(k+2)}, \eta_{(k+3)}, \ldots, \eta_m$ since they are independent of $\eta_{(k+1)}$. Again the distribution of η_k is free of $\{(Z)_{j\cdot} : 0 \le j \le k-1\}$ so that the unconditional distribution of η_k is identical to η_m . In this way we argue that $\{\eta_j : 1 \le j \le m\}$ are mutually independent and identically distributed with

$$\eta_j \sim \sigma^2 \chi^2_{(\nu-1)} \ (0) \ : 1 \le j \le m \ . \tag{4.2.51}$$

Since $\hat{\sigma}^2 = \frac{1}{m\nu} \sum_{j=1}^{m} \eta_j$ we have

$$\hat{\sigma}^2 \sim \frac{\sigma^2}{m\nu} \chi^2_{m(\nu-1)} (0) \ . \tag{4.2.52}$$

In the above arguement we note that the variables $\{\eta_m, \eta_{(m-1)}, \ldots, \eta_k\}$ are independent of $\{(Z)_{j\cdot} : o \le j \le k-1\}$. Hence we have $\{\eta_m, \eta_{(m-1)}, \ldots, \eta_1\}$ are independent of $(Z)_{0\cdot}$, and hence of

$$w_{00} = (Z)_{0\cdot} (Z)'_{0\cdot} \ . \tag{4.2.53}$$

Since

$$Z_{0k} \sim N(0, \sigma^2 \beta^2) \ : 1 \le k \le \nu \ , \tag{4.2.54}$$

then

$$w_{00} \sim \sigma^2 \beta^2 \chi^2_{\nu} \ (0) \ . \tag{4.2.55}$$

Since $\hat{\beta}^2 = (\nu \hat{\sigma}^2)^{-1} w_{00}$, then we have

$$\frac{(\nu-1)}{\nu\beta^2}\;\hat{\beta}^2 \sim F^{\nu}_{m(\nu-1)}(0) \; . \qquad (4.2.56)$$

4.3 Properties of the Maximum Likelihood Estimators

Since the distribution of the maximum likelihood estimators is known their properties are easily obtained. We find using equations (4.2.37), (4.2.52), and (4.2.56) that $\mathcal{E}(\hat{\alpha}_j) = \alpha_j \; : \; 1 \le j \le m$, (4.3.1)

$$\mathcal{E}(\hat{\sigma}^2) = (\nu-1)\nu^{-1}\sigma^2 , \qquad (4.3.2)$$

and $\mathcal{E}(\hat{\beta}^2) = m\nu[m(\nu-1)-2]^{-1}\beta^2$. (4.3.3)

Hence the $\hat{\alpha}_j$ are unbiassed estimators while $\hat{\sigma}^2$ and $\hat{\beta}^2$ are biassed. Since unbiassedness is a desirable property we shall use the unbiassed estimators of σ^2 and β^2 in calculating the rest of the properties.

Letting $\hat{\Sigma}$ denote the (m+2 x m+2) variance-covariance matrix of the (m+2) dimensional vector of unbiassed estimators $(\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_m, \nu(\nu-1)^{-1}\hat{\sigma}^2, [m(\nu-1)-2](m\nu)^{-1}\hat{\beta}^2)$, we find

$$(\hat{\Sigma})_{11} = (\nu-2)^{-1}\beta^{-2} , \qquad (4.3.4)$$

$$(\hat{\Sigma})_{jj} = (\nu-2)^{-1}v^{-1}_{j-1,j-1}\sigma^2 : \; 2 \le j \le m , \qquad (4.3.5)$$

where $v_{j-1,j-1}$ is defined in (4.2.38) ,

$$(\hat{\Sigma})_{m+1,m+1} = 2[m(\nu-1)]^{-1}\sigma^4 , \qquad (4.3.6)$$

$$(\hat{\Sigma})_{m+2,m+2} = 2[m(\nu-1)+\nu-2][m\nu(\nu-1)-4\nu]^{-1}\beta^4 \qquad (4.3.7)$$

$$(\hat{\Sigma})_{m+1,m+2} = (\hat{\Sigma})_{m+2,m+1} = -2[m(\nu-1)]^{-1}\sigma^2\beta^2 \qquad (4.3.8)$$

and

$$(\hat{\Sigma})_{jk} = 0 \; : \; \text{elsewhere} \; . \qquad (4.3.9)$$

To see that the set of estimators $\{\hat{\alpha}_j : 1 \le j \le m\}$ are independent of the set $\{\hat{\sigma}^2, \hat{\beta}^2\}$ we note that the distribution of $\hat{\sigma}^2$ and $\hat{\beta}^2$ are free of the elements of $\hat{\alpha}_j$ and hence the two sets of estimators are independent. To see that the covariances between the $\hat{\alpha}$'s are zero we note that

$$\hat{\alpha}_j = \sum_{k=1}^{\nu} \phi_{j-1,k} z_{j,k}$$

$$= \sum_{k=1}^{\nu} \phi_{j-1,k}(\alpha_j z_{j-1,k} + \varepsilon_{jk})$$

$$= \alpha_j + \sum_{k=1}^{\nu} \phi_{j-1,k}\varepsilon_{j,k} \quad : 1 \le j \le m , \tag{4.3.10}$$

where $\phi_{j,k}$ is defined by equations (4.2.24) as

$$\phi_{j,k} = \frac{z_{j,k}}{\sum_{k=1}^{\nu} z_{j,k}^2} .$$

Hence the covariance between $\hat{\alpha}_j$ and $\hat{\alpha}_\ell$ $(j \ne \ell)$ is given by

$$(\hat{\Sigma})_{j\ell} = \mathcal{E}(\hat{\alpha}_j \hat{\alpha}_\ell) - \mathcal{E}(\hat{\alpha}_j)\,\mathcal{E}(\hat{\alpha}_\ell)$$

$$= \mathcal{E}[(\alpha_j + \sum_{k=1}^{\nu} \phi_{j-1,k}\varepsilon_{jk})(\alpha_\ell + \sum_{k=1}^{\nu} \phi_{\ell-1,k}\varepsilon_{\ell k})] - \alpha_j\alpha_\ell$$

$$= \mathcal{E}(\alpha_j\alpha_\ell + \alpha_j \sum_{k=1}^{\nu} \phi_{\ell-1,k}\varepsilon_{\ell k} + \alpha_\ell \sum_{k=1}^{\nu} \phi_{j-1,k}\varepsilon_{jk}$$

$$+ \sum_{i=1}^{\nu} \sum_{k=1}^{\nu} \phi_{j-1,i}\varepsilon_{ji}\phi_{\ell-1,k}\varepsilon_{\ell k}) - \alpha_j\alpha_\ell \quad . \tag{4.3.11}$$

Since the $\{\varepsilon_{st} : 1 \le s \le m,\ 1 \le t \le \nu\}$ are independent identically distributed normal random variables with zero mean, we have taking expectations first with respect to the ε_{st}'s that the last three terms in brackets in (4.3.11) vanish and we are left with

$$(\hat{\Sigma})_{j\ell} = \alpha_j\alpha_\ell - \alpha_j\alpha_\ell$$
$$= 0 \qquad : 1 \le j \ne \ell \le m \; . \qquad (4.3.12)$$

Although this shows the estimators are uncorrelated it is not true that they are independent. To see this we examine the case for m=2. Write

$$W = GDG'$$

$$= \begin{bmatrix} d_0 & d_0g_{10} & d_0g_{20} \\ d_0g_{10} & d_1+d_0g_{10}^2 & d_1g_{21}+d_0g_{10}g_{20} \\ d_0g_{20} & d_1g_{21}+d_0g_{10}g_{20} & d_2+d_1g_{21}^2+d_0g_{20}^2 \end{bmatrix} . \qquad (4.3.13)$$

Hence,

$$\hat{\alpha}_1 = \frac{w_{01}}{w_{00}}$$

$$= g_{10} \; , \qquad (4.3.14)$$

and

$$\hat{\alpha}_2 = \frac{w_{12}}{w_{11}}$$

$$= \frac{(d_0g_{10}g_{20}+d_1g_{21})}{d_0g_{10}^2+d_1} \; . \qquad (4.3.15)$$

We shall show that

$$\hat{\alpha}_2 \,|\, (g_{10}, d_0, d_1) \sim N(\alpha_2, \frac{\sigma^2}{d_1+d_0g_{10}^2}) \; , \qquad (4.3.16)$$

by equation (4.3.14) this is equivalent to

$$\hat{\alpha}_2 | (\hat{\alpha}_1, d_0, d_1) \sim N(\alpha_2, \frac{\sigma^2}{d_1 + d_0 \hat{\alpha}_1^2}) \ . \tag{4.3.17}$$

Since the conditional distribution of $\hat{\alpha}_2$ depends on $\hat{\alpha}_1$ we have that they are dependent. To show (4.3.16) we need the distribution of (g_{10}, g_{20}, g_{21}) conditional on (d_0, d_1, d_2). Referring back to (2.5.21) we have

$$\underline{g}_0 = (g_{10}, g_{20})' | (d_0, d_1) \sim N_2(\underline{\mu}_0, V_0) \tag{4.3.18}$$

where $\underline{\mu}_0 = (\alpha_1, \alpha_1\alpha_2)'$ (4.3.19)

and

$$V_0 = \begin{bmatrix} \frac{\sigma^2}{d_0} & \alpha \frac{\sigma^2}{d_0} \\ \alpha_2 \frac{\sigma^2}{d_0} & \frac{\sigma^2}{d_0} + \alpha_2^2 \frac{\sigma^2}{d_0} \end{bmatrix} , \tag{4.3.20}$$

and

$$g_{21} | (d_0, d_1) \sim N(\alpha_2, \frac{\sigma^2}{d_0}) \tag{4.3.21}$$

independent of (g_{10}, g_{20}). Now the distribution of g_{20} conditional on (g_{10}, d_0, d_1) is easily shown to be

$$g_{20} | (g_{10}, d_0, d_1) \sim N(\alpha_2 g_{10}, \frac{\sigma^2}{d_0}) \ . \tag{4.3.22}$$

Since g_{21} is independent of g_{10}, conditioning on g_{10} does not affect the distribution of g_{21}, that is,

$$g_{21} | (g_{10}, d_0, d_1) \sim N(\alpha_2, \frac{\sigma^2}{d_1}) \ .$$

We note, by equation (4.3.15), that $\hat{\alpha}_2$ conditional on (g_{10}, d_0, d_1) is simply a linear combination of g_{20} and g_{21}. Since they are normally distributed (conditionally) so will $\hat{\alpha}_2$

(conditionally). All we need do is calculate the mean and variance to find the conditional distribution of $\hat{\alpha}_2$.

$$\mathcal{E}\hat{\alpha}_2|(g_{10},d_0,d_1) = \mathcal{E}\{\frac{(d_0 g_{10} g_{20}+d_1 g_{21})}{d_0 g_{10}^2+d_1}\} \,|(g_{10},d_0,d_1)$$

$$= \frac{\alpha_2(d_0 g_{10}^2+d_1)}{d_0 g_{10}^2+d_1}$$

$$= \alpha_2 \ . \tag{4.3.23}$$

$$\mathrm{Var}(\hat{\alpha}_2)|(g_{10},d_0,d_1) = \mathrm{Var}\{\frac{(d_0 g_{10} g_{20}+d_1 g_{21})}{d_0 g_{10}^2+d_1}\}|(g_{10},d_0,d_1)$$

recalling that g_{20} and g_{21} are independent we have,

$$= \frac{d_0^2 g_{10}^2 \ \mathrm{Var}\{(g_{20})|(g_{10},d_0,d_1)\}+d_1^2\mathrm{Var}\{(g_{21})|(g_{10},d_0,d_1)\}}{(d_0 g_{10}^2+d_1)^2}$$

$$= \frac{d_0^2 g_{10}^2 \frac{\sigma^2}{d_0} + d_1^2 \frac{\sigma^2}{d_1}}{(d_0 g_{10}^2+d_1)^2}$$

$$= \frac{\sigma^2}{(d_0 g_{10}^2+d_1)} \ . \tag{4.3.24}$$

Hence

$$\hat{\alpha}_2|(g_{10},d_0,d_1) \sim N(\alpha_2, \frac{\sigma^2}{(d_0 g_{10}^2+d_1)})$$

as was to be shown.

Furthermore it can be shown that $\hat{\alpha}_1$ and $\hat{\alpha}_2$ do not have a bivariate t-distribution. To see this we find $f(\hat{\alpha}_2|\hat{\alpha}_1 = 0)$. We have that

$$\hat{\alpha}_2 \mid (\hat{\alpha}_1, d_o, d_1) \sim N(\alpha_2, \frac{\sigma^2}{\hat{d}_0 \hat{\alpha}_1^2 + d_1})$$

and hence

$$\hat{\alpha}_2 \mid (\hat{\alpha}_1 = 0, d_0, d_1) \sim N(\alpha_2, \frac{\sigma^2}{d_1}) \quad . \tag{4.3.25}$$

Since the distribution does not depend on d_0 and since d_0 and d_1 are independent we have

$$\hat{\alpha}_2 \mid (\hat{\alpha}_1 = 0, d_1) \sim N(\alpha_2, \frac{\sigma^2}{d_1}) \quad . \tag{4.3.26}$$

and from (2.5.5)

$$d_1 \sim \sigma^2 \chi^2_{\nu-1}(0) \quad , \tag{4.3.27}$$

so that

$$f(\hat{\alpha}_2, d_1 \mid \hat{\alpha}_1 = 0) = \frac{1}{\sqrt{2\pi}\,(\frac{\sigma^2}{d_1})^{\frac{1}{2}}} e^{-\frac{d_1}{2\sigma^2}(\hat{\alpha}_2 - \alpha_2)^2} \frac{d_1^{\frac{1}{2}(\nu-1)-1} e^{-\frac{d_1}{2\sigma^2}}}{(2\sigma^2)^{\frac{1}{2}(\nu-1)} \Gamma(\frac{1}{2}(\nu-1))}$$

$$: \begin{array}{l} -\infty < \hat{\alpha}_2 < \infty \\ d_1 > 0 \\ (\nu-1) > 0 \end{array} \quad . \tag{4.3.28}$$

Integrating over d_1, we have

$$f(\hat{\alpha}_2 \mid \hat{\alpha}_1 = 0) = \int_0^\infty \frac{d_1^{\frac{1}{2}\nu-1} e^{-\frac{d_1}{2\sigma^2}[1+(\hat{\alpha}_2-\alpha_2)^2]}}{\sqrt{2\pi}\,(2\sigma^2)^{\frac{1}{2}\nu}\Gamma(\frac{1}{2}(\nu-1))} d(d_1)$$

$$= \frac{\Gamma(\frac{1}{2}\nu)\ [1+(\hat{\alpha}_2-\alpha_2)^2]^{-\frac{1}{2}\nu}}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}(\nu-1))} \quad : \begin{array}{l} -\infty < \hat{\alpha}_2 < \infty \\ \nu > 0 \end{array} \quad . \tag{4.3.29}$$

Hence we find $\hat{\alpha}_2$ conditional on $\hat{\alpha}_1 = 0$ has a t-distribution with $(\nu-1)$ degrees of freedom, but we know $\hat{\alpha}_2$ has the t-distribution with ν degrees of freedom, this will show a contradiction and $\hat{\alpha}_1$ and $\hat{\alpha}_2$ cannot have a bivariate t-distribution.

Now suppose t_1 and t_2 have a bivariate t-distribution with ν degrees of freedom. Their joint density is given by

$$f(t_1,t_2) = \frac{\Gamma(\frac{1}{2}(\nu+2))\,|\Sigma|^{-\frac{1}{2}}}{\pi\ \Gamma\ (\frac{1}{2}\nu)\,[1+\begin{pmatrix}t_1-\mu_1\\t_2-\mu_2\end{pmatrix}'\Sigma^{-1}\begin{pmatrix}t_1-\mu_1\\t_2-\mu_2\end{pmatrix}]^{\frac{1}{2}(\nu+2)}} \quad : -\infty<t_i<\infty\ ,\ i=1,2\ ;\ \nu>0\ , \tag{4.3.30}$$

where μ_1 and μ_2 are the expected values of t_1 and t_2 respectively. Also Σ is the (2x2) variance-covariance matrix of (t_1,t_2). Relating this to $\hat{\alpha}_1$ and $\hat{\alpha}_2$ we would replace (μ_1,μ_2) by (α_1,α_2) and Σ would be a diagonal matrix since we have shown the covariance of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ to be zero. Without any loss of generality we may take $\mu_1=\mu_2=0$ and $\Sigma=I$. The marginal density of t_1 is

$$f(t_1) = \frac{\Gamma(\frac{1}{2}(\nu+1)}{\pi^{\frac{1}{2}}\Gamma(\frac{1}{2}\nu)\,[1+t_1^2]^{\frac{1}{2}(\nu+1)}} \quad : -\infty<t_1<\infty;\ \nu>0\ . \tag{4.3.31}$$

Hence the conditional density of t_2 given t_1 is

$$f(t_2|t_1) = \frac{f(t_1,t_2)}{f(t_1)}$$

$$= \frac{\Gamma(\frac{1}{2}(\nu+2))\quad [1+t_1^2]^{\frac{1}{2}(\nu+1)}}{\pi^{\frac{1}{2}}\Gamma(\frac{1}{2}(\nu+1))\ [1+t_1^2+t_2^2]^{\frac{1}{2}(\nu+2)}} \quad : -\infty<t_2<\infty\ ;\ \nu\geq 0\ . \tag{4.3.32}$$

In particular, suppose $t_1=0$, then

$$f(t_2|t_1=0) = \frac{\Gamma(\frac{1}{2}(\nu+2))}{\pi^{\frac{1}{2}}\,\Gamma(\frac{1}{2}(\nu+1))[1+t_2^2]^{\frac{1}{2}(\nu+2)}} \quad : \quad -\infty<t_2<\infty; \quad \nu\geq 0 \; . \tag{4.3.33}$$

That is, t_2 conditional on $t_1 = 0$ has the Student t-distribution with $(\nu+1)$ degrees of freedom. Hence we see that if $(\hat{\alpha}_1,\hat{\alpha}_2)$ are bivariate t-distributed, then since $\hat{\alpha}_2$ conditional on $\hat{\alpha}_1 = 0$ has the Student t-distribution with $(\nu-1)$ degrees of freedom it must be that $\hat{\alpha}_2$ has the Student t-distribution with $(\nu-2)$ degrees of freedom, but this is a contradiction since $\hat{\alpha}_2$ has the Student t-distribution with ν degrees of freedom. Hence $(\hat{\alpha}_1,\hat{\alpha}_2)$ do not have a bivariate t-distribution.

Hence we see that the maximum likelihood estimators are not independently distributed and their joint distribution is not multivariate t. This of course is a drawback in using the maximum likelihood estimators and accentuates the benefits of using the starred estimator, which are independent and have the t-distribution.

It is easily seen that the unbiased estimators are consistent since the variance tends to zero as the sample size increases without bound. To find the efficiency of the maximum likelihood estimators we compare their variance to the minimum variance bounds given in equations (3.2.37) through (3.2.41). We find that

$$\mathrm{Eff}(\hat{\alpha}_j)=(\nu-2)\nu^{-1} \quad : \quad 1\leq j\leq m \; , \tag{4.3.34}$$

$$\mathrm{Eff}(\nu(\nu-1)^{-1}\hat{\sigma}^2) = (\nu-1)\nu^{-1} \; , \tag{4.3.35}$$

and $\mathrm{Eff}([m(\nu-1)-2](m\nu)^{-1}\hat{\beta}^2)=[(m+1)(m(\nu-1)-4)][m(m(\nu-1)+\nu-2]^{-1}$.

(4.3.36)

It is obvious that the estimators are asymptotically efficient. Unlike the starred estimators the efficiency of the maximum likelihood estimators does not depend on the unknown parameters, but the distribution of the maximum likelihood estimators $\{\hat{\alpha}_j: 1 \leq j \leq m\}$ depend on the unknown parameters so that tests of hypothesis of α_j depends on knowing the values of $\alpha_1, \alpha_2, \ldots, \alpha_{j-1}$. This clearly shows the trade-off between the starred estimators and the maximum likelihood estimators. While the starred estimators are inefficient, tests of hypotheses are performed with no difficulty, and vice versa the maximum likelihood estimators are efficient (asymptotically), but test of hypotheses are complicated since their distribution depends on several unknown parameters. Moreover the dependence between the $\hat{\alpha}$'s also causes complications in making tests of hypotheses concerning two or more of the parameters since the joint distribution may be very complex.

Chapter V

A TEST OF THE ADEQUACY OF THE MODEL

5.1 Introduction

Throughout we have assumed that the process is adequately described by the first order autoregressive model. In this chapter we propose a method of testing the validity of this assumption. Due to the assumption of the first order autoregressive process a class of dispersion matrices arose which we identified by $\tilde{V}$. Since this class of dispersion matrices is a consequence of the model, a test to validate the model is equivalent to a test of H_0: $V \varepsilon \tilde{V}$ against H_1: V is an arbitrary positive definite matrix.

In order to arrive at a test statistic for testing this hypothesis we recall that if $V \varepsilon \tilde{V}$ then $V=AUA'$ where A and U were defined in equations (2.2.4) and (2.2.7). In particular U was given as the $(m+1)x(m+1)$ diagonal matrix

$$U=\mathrm{diag}(\sigma^2\beta^2,\sigma^2,\sigma^2,\ldots,\sigma^2) \ . \qquad (5.1.1)$$

We also showed that

$$d_j \sim \sigma^2 \chi^2_{\nu-j}(0) \ : \ 1 \le j \le m \ , \qquad (5.1.2)$$

and with ν large compared to m each of the d_j's should be nearly the same. Ignoring the first row and column of U we have that the remaining diagonal elements of U are σ^2 and $\{d_j: 1 \le j \le m\}$ are independent estimators of this quantity. If

H_0 is true then all of the d_j's should be equal. Another way of putting this is that the arithmetic mean of $d_1, d_2, \ldots, d_m$ is equal to the geometric mean, that is,

$$\lambda_1 = \frac{\prod_{i=1}^{m} d_i}{\left(\frac{1}{m}\sum_{j=1}^{m} d_j\right)^m}$$

$$= \prod_{i=1}^{m} \left(\frac{d_i}{\frac{1}{m}\sum_{j=1}^{m} d_j} \right) = 1, \ (V\varepsilon\tilde{V}) \ . \qquad (5.1.3)$$

If H_0 is false then the d_j's will not be equal and the arithmetic mean will be larger than the geometric mean and λ_1 will be less than one. Hence we see that we reject H_0 for small values of λ_1. The asymptotic distribution of λ_1 will be investigated in section 2. This test has an interesting geometrical interpretation. If we consider the $\{d_j: 1 \le j \le m\}$ to be the squared lengths of the principal axes of an m dimensional ellipsoid, the above hypothesis specifies that these are all equal, that is, that the ellipsoid is a sphere. Hence this test is the sphericity test on the $\{d_j: 1 \le j \le m\}$.

Besides the sphericity test we consider another test, independent of λ_1, based on $\{g_{ij}: 0 \le j < i \le m\}$. The statistics $\{g_{ij}: 0 \le j < i \le m\}$ contain a great deal of information about the process since they are used as estimators of $\{\alpha_j: 1 \le j \le m\}$ and hence of V. In section 3 we shall investigate the distribution of a function of these statistics under the hypothesis that $V \varepsilon \tilde{V}$. In section 4 we discuss combining the two tests given in sections 2 and 3 and the asymptotic equivalence of the combined tests as compared to the likelihood ratio.

5.2 An Approximation to the Distribution of $-\rho_0 \log \lambda_1$.

Referring to equation (5.1.3) we see that λ_1 may be written as the product of $u_1, u_2, \ldots, u_m$ where

$$u_i = \frac{d_i}{\frac{1}{m}\sum_{j=;}^{m} d_j} \quad : 1 \le i \le m \, , \tag{5.2.1}$$

that is,

$$\lambda_1 = \prod_{i=1}^{m} u_i \; . \tag{5.2.2}$$

Rather than consider the distribution of λ_1 we shall consider the distribution of

$$\eta = -\rho \log \lambda_1 \quad : 0 \le \eta < \infty \quad , \tag{5.2.3}$$

where ρ is some constant. The moment generating function of η is

$$\begin{aligned} \phi_\eta(\theta) &= \mathcal{E} e^{\theta\eta} \\ &= \mathcal{E}(\lambda_1)^{-\theta\rho} \\ &= \mathcal{E}\left(\prod_{i=1}^{m} u_i\right)^{-\theta\rho} \; . \end{aligned} \tag{5.2.4}$$

In order to find this expectation we need to find the joint distribution of $(u_1, u_2, \ldots, u_m)$. To find the joint distribution we shall transform from $(d_1, d_2, \ldots, d_m)$ into $(u_1, u_2, \ldots, u_m, S)$

where u_i is defined by (5.2.1) and $S = \frac{1}{m} \sum_{j=1}^{m} d_j$. Hence we seek the Jacobian of the transformation from $(d_1, d_2, \ldots, d_m)$ to $(u_1, u_2, \ldots, u_m, S)$ defined by

$$d_i = u_i S \quad : \; 1 \le i \le m \; , \tag{5.2.5}$$

with

$$\sum_{i=1}^{m} u_i = m. \tag{5.2.6}$$

Let $[\delta u_i]$ and $[\delta S]$ denote small changes in u_i and S, respectively. Suppose that the changes $[\delta u_i]$ in u_i and $[\delta S]$ in S bring about a change $[\delta d_i]$ in d_i so that (5.2.5) and (5.2.6) is preserved. That is

$$d_i + [\delta d_i] = (u_i + [\delta u_i])(S + [\delta S]) \quad : \; 1 \le i \le m \; , \tag{5.2.7}$$

and

$$\sum_{i=1}^{m} (u_i + [\delta u_i]) = m \quad . \tag{5.2.8}$$

Expanding the above equations and dropping terms of second order in the $[\delta *][*\varepsilon(d_i, S)]$, we find that

$$d_i + [\delta d_i] = u_i S + [\delta u_i] S + u_i [\delta S] \quad : \; 1 \le i \le m \; , \tag{5.2.9}$$

and

$$\sum_{i=1}^{m} u_i + \sum_{i=1}^{m} [\delta u_i] = m \; . \tag{5.2.10}$$

Since $d_i = u_i S$ and $\sum_{i=1}^{m} u_i = m$ we see that

$$[\delta d_i] = [\delta u_i] S + u_i [\delta S] \quad : \; 1 \le i \le m \; , \tag{5.2.11}$$

and

$$\sum_{i=1}^{m} [\delta u_i] = 0 \; . \tag{5.2.12}$$

To write the above in vector notation we define the (mx1) column vectors

$$[\delta \underline{d}] = ([\delta d_1],[\delta d_2],\ldots,[\delta d_m])' , \tag{5.2.13}$$

$$\underline{d} = (d_1,d_2,\ldots,d_m)' \tag{5.2.14}$$

$$[\delta \underline{u}] = ([\delta u_1],[\delta u_2],\ldots,[\delta u_m])' , \tag{5.2.15}$$

$$\underline{u} = (u_1,u_2,\ldots,u_m)' \tag{5.2.16}$$

and

$$\underline{1}_m = (1,1,\ldots,1)' \ . \tag{5.2.17}$$

Equations (5.2.11) and (5.2.12) may now be written

$$[\delta \underline{d}] = [\delta \underline{u}]S + \underline{u}[\delta S] \tag{5.2.18}$$

and

$$\underline{1}_m'[\delta \underline{u}] = 0 \ . \tag{5.2.19}$$

Equations (5.2.11) can be thought of as a singular transformation from $\{[\delta d_1], [\delta d_2], \ldots, [\delta d_m]\}$ to $\{[\delta u_1], [\delta u_2], \ldots, [\delta u_m], [\delta S]\}$ made one-to-one through use of equation (5.2.12). Saw [17] has shown that

$$J\{\underline{d}\to\underline{u},S\} = J\{[\delta \underline{d}]\to[\delta \underline{u}],[\delta S]\}, \tag{5.2.20}$$

where $J\{[\delta \underline{d}]\to[\delta \underline{u}], [\delta S]\}$ is the Jacobian of the transformation defined by (5.2.18), in which $\underline{u}$ and S are considered fixed.

Choose P to be an orthogonal mxm matrix with the first row equal to $\underline{1}_m'$ and pre-multiplying equation (5.2.18) by it gives

$$\begin{aligned}\underline{v} = P[\delta \underline{d}] &= P[\delta \underline{u}]S + P\,\underline{u}[\delta S] \\ &= \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} S + \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} [\delta S] \ ,\end{aligned} \tag{5.2.21}$$

where $y_1 \equiv 0$ since $1'[\delta \underline{u}] = 0$.

From equations (5.2.20) and (5.2.21) we have

$$J\{\underline{d}\rightarrow\underline{u},\ S\} = J\{[\delta\underline{d}] \rightarrow [\delta u],\ [\delta S]\}$$

$$= J\{[\delta\underline{d}] \rightarrow \underline{v}\}\cdot J\{\underline{v}\rightarrow y_2,\ldots,y_m,[\delta S]\}. \qquad (5.2.22)$$

The Jacobian, $J\{[\delta\underline{d}]\rightarrow\underline{v}\}$, is unity since P is an orthogonal matrix. The Jacobian, $J\{\underline{v}\rightarrow y_2,\ldots,y_m,[\delta S]\}$, is the modulus of the determinant of the (mxm) matrix K with elements

$$(K)_{11} = \frac{\partial v_1}{\partial[\delta S]} = 1\ ,$$

$$(K)_{1j} = \frac{\partial v_j}{\partial[\delta S]} = 0\ :\ 2\le j\le m\ ,$$

$$(K)_{jj} = \frac{\partial v_j}{\partial y_j} = S\quad :\ 2\le j\le m\ ,$$

$$(K)_{kj} = \frac{\partial v_j}{\partial y_k} = 0\quad :\ \text{elsewhere}\ . \qquad (5.2.23)$$

Hence K is a diagonal matrix and

$$J\{\underline{v}\rightarrow y_2,\ldots,y_m,[\delta S]\} = ||K|| = S^{m-1} \qquad (5.2.24)$$

and finally

$$J\{\underline{d}\rightarrow\underline{u},\ S\} = S^{m-1}\ . \qquad (5.2.25)$$

Since the d_j's are independent with

$$d_j \sim \sigma^2\chi^2_{\nu-j}\ (0)\ :\ 1\le j\le m\ , \qquad (5.2.26)$$

then

$$f(d_1,d_2,\ldots,d_m) = \prod_{j=1}^{m} \frac{d_j^{\frac{1}{2}\nu_j-1}\, e^{-\frac{1}{2\sigma^2}d_j}}{(2\sigma^2)^{\frac{1}{2}\nu_j}\,\Gamma(\frac{1}{2}\nu_j)}\quad \begin{array}{l}: d_j>0\\ ;\ 1\le j\le m\end{array} \qquad (5.2.27)$$

where $\nu_j = \nu - j :\ 1\le j\le m$. Hence we find the joint distribution $\underline{u} = (u_1,u_2,\ldots,u_m)'$ and S is

$$f(u_1,u_2,\ldots,u_m,S) = \{\Gamma(\tfrac{1}{2}\sum_{j=1}^{m}\nu_j)\prod_{j=1}^{m}\left(\frac{u_j^{\frac{1}{2}\nu_j-1}}{\Gamma(\frac{1}{2}\nu_j)m^{\frac{1}{2}\nu_j}}\right)\}$$

$$\frac{(S)^{\frac{1}{2}\sum_{j=1}^{m}\nu_j-1}\; e^{-\frac{m}{2\sigma^2}S}}{\left(\frac{2\sigma^2}{m}\right)^{\frac{1}{2}\sum_{j=1}^{m}\nu_j}\Gamma(\frac{1}{2}\sum_{j=1}^{m}\nu_j)}$$

$$: \; 0\le u_j\le m,\; \sum_{j=1}^{m}u_j=m;\; S>0\;. \qquad (5.2.28)$$

Hence we find $\underline{u}$ and S are independently distributed with

$$S \sim \frac{\sigma^2}{m}\chi^2_{(\sum_{j=1}^{m}\nu_j)}(0)\;, \qquad (5.2.29)$$

and $\underline{u}$ is distributed as $m\underline{Z}$ where $\underline{Z}$ has the Dirichlet distribution with (mxl) dimensional parameter vector $(\frac{1}{2}\nu_1,\frac{1}{2}\nu_2,\ldots,\frac{1}{2}\nu_m)' = (\frac{1}{2}(\nu-1),\frac{1}{2}(\nu-2),\ldots,\frac{1}{2}(\nu-m))'$.

If $(y_1, y_2, \ldots, y_n)$ has the Dirichlet distribution with parameters $(\alpha_1,\alpha_2,\ldots,\alpha_n)$, then the moments about the origin are given by

$$\mathcal{E}(y_1^{r_1}, y_2^{r_2}\cdots y_n^{r_n}) = \frac{\{\prod_{j=1}^{n}\Gamma(\alpha_j+r_j)\}\Gamma(\sum_{j=1}^{n}\alpha_j)}{\Gamma(\sum_{j=1}^{n}(\alpha_j+r_j))\{\prod_{j=1}^{n}\Gamma(\alpha_j)\}} \qquad (5.2.30)$$

Hence we find the moment generating function of η is given by

$$\phi_\eta(\theta) = \mathcal{E}e^{\theta\eta}$$

$$= \mathcal{E}(\prod_{i=1}^{m}u_i)^{-\theta\rho}$$

and letting $u_i = mZ_i$ gives

$$\phi_\eta(\theta) = \mathcal{E}\left(\prod_{i=1}^{m} mZ_i\right)^{-\theta\rho}$$

$$= m^{-m\theta\rho} \cdot \mathcal{E}\left(\prod_{i=1}^{m} Z_i^{-\theta\rho}\right)$$

and since $Z_i : 1 \le i \le m$ are Dirichlet we have from (5.2.30) that

$$\phi_\eta(\theta) = m^{-m\theta\rho} \frac{\left\{\prod_{j=1}^{m} \Gamma(\tfrac{1}{2}\nu_j - \theta\rho)\right\} \cdot \Gamma(\tfrac{1}{2}\sum_{j=1}^{m} \nu_j)}{\Gamma\left(\sum_{j=1}^{m} (\tfrac{1}{2}\nu_j - \theta\rho)\right) \left\{\prod_{j=1}^{m} \Gamma(\tfrac{1}{2}\nu_j)\right\}} . \qquad (5.2.31)$$

Since the moments of η are functions of gamma functions we can apply Box's [4] method to obtain an approximation to the distribution of η. A good discussion of Box's method is also contained in Anderson [1].

Using equation (5.2.31) the cumulant generating function for η is

$$\Psi_\eta(\theta) = \log\phi_\eta(\theta)$$
$$= k - m\theta\rho\log m + \sum_{j=1}^{m} \log\Gamma(\tfrac{1}{2}(\nu-j) - \theta\rho) - \log\Gamma(\tfrac{1}{2}(m\nu - \tfrac{1}{2}m(m+1)$$
$$-m\theta\rho)), \qquad (5.2.32)$$

where k has a value independent of θ. Rewriting this as

$$\Psi_\eta(\theta) = k - m\rho\theta\log m + \sum_{j=1}^{m} \log\Gamma(\alpha_j + \tfrac{1}{2}\rho(1-2\theta)) - \log\Gamma(\beta + \tfrac{1}{2}m\rho(1-2\theta)) \qquad (5.2.33)$$

where

$$\alpha_j = \tfrac{1}{2}(\nu - \rho - j) \ : \ 1 \le j \le m,$$

and $\qquad (5.2.34)$

$$\beta = \tfrac{1}{2}[m\nu - m\rho - \tfrac{1}{2}m(m+1)] .$$

We use the expansion formula

$$\log\Gamma(x+h) = \tfrac{1}{2}\log 2\pi + (x+h-\tfrac{1}{2})\log x - x - \sum_{r=1}^{\infty} \frac{(-1)^r B_{r+1}(h)}{r(r+1)x^r} \tag{5.2.35}$$

where $B_s(h)$ is the s-th Bernoulli polynomial defined by

$$\frac{\tau e^{h\tau}}{(e^{\tau}-1)} = \sum_{s=0}^{\infty} \frac{\tau^s}{s!} B_s(h) , \tag{5.2.36}$$

for example

$$B_1(h) = h-\frac{1}{2} \; ; \; B_2(h) = h^2-h+\frac{1}{6} \; ;$$

$$B_3(h) = h^3 - \frac{3}{2} h^2 + \frac{1}{2} h .$$

Using the expansion formula (5.2.35) on (5.2.33) we find that

$$\Psi_\eta(\theta) = k+\frac{(m-1)}{2}(\log 2\pi-\log \tfrac{\rho}{2}) - (\beta+\tfrac{m\rho}{2}-\tfrac{1}{2})\log m - \frac{1}{2}(m-1)\log(1-2\theta) + \sum_{r=1}^{\infty} \pi_r^* \left\{\frac{1}{(1-2\theta)^r}\right\} \tag{5.2.37}$$

with

$$\pi_r^* = \frac{(-1)^r\left\{\dfrac{B_{r+1}(\beta)}{m^r} - \sum_{j=1}^{m} B_{r+1}(\alpha_j)\right\}}{r(r+1)\;(\tfrac{1}{2}\rho)^r} . \tag{5.2.38}$$

By virtue of the fact that $\Psi_\eta(\theta=0)=0$ we must have

$$k+\frac{(m-1)}{2} (\log 2\pi- \log \tfrac{\rho}{2}) - (\beta+\tfrac{m\rho}{2}-\tfrac{1}{2}) \log m = \sum_{r=1}^{\infty} \pi_r^* \tag{5.2.39}$$

so that we may write

$$\Psi_\eta(\theta) = -\tfrac{1}{2}(m-1)\log(1-2\theta) + \sum_{r=1}^{\infty} \pi_r^* \left\{\frac{1}{(1-2\theta)^r} - 1\right\}. \tag{5.2.40}$$

If r has the chi-square distribution on e degrees of freedom then its cumulant generating function is

$$\Psi_r(\theta) = -\frac{e}{2}\log(1-2\theta), \tag{5.2.41}$$

we see that equation (5.2.37) has the same form with e=(m-1) degrees of freedom and an additional sum which may be called the remainder. This remainder may be reduced by choosing $\rho=\rho_0$ so that $\pi_1^* = 0$ and the approximation is improved. For $\pi_1^* = 0$ we must have

$$B_2(\beta) = m\sum_{j=1}^{m} B_2(\alpha_j) \tag{5.2.42}$$

or

$$\beta^2-\beta+\frac{1}{6}=m\sum_{j=1}^{m}(\alpha_j^2-\alpha_j+\frac{1}{6}). \tag{5.2.43}$$

Recall that

$$\beta=\frac{1}{2}m(\nu-\rho)-\frac{1}{4}m(m+1)$$

and

$$\alpha_j=\tfrac{1}{2}(\nu-\rho)-\tfrac{1}{2}j \ : \ 1\leq j\leq m,$$

letting $\delta = \frac{1}{2}(\nu-\rho)$ then

$$\beta= m\delta-\tfrac{1}{4}m(m+1) \tag{5.2.44}$$

and

$$\alpha_j=\delta-\tfrac{1}{2}j \ : \ 1\leq j\leq m. \tag{5.2.45}$$

Substituting (5.2.44) and (5.2.45) into equation (5.3.43) gives

$$[m\delta-\frac{1}{4}m(m+1)]^2-m\delta+\frac{1}{4}m(m+1)+\frac{1}{6} =$$

$$m\sum_{j=1}^{m}[(\delta-\frac{1}{2}j)^2-\delta+\frac{1}{2}j+\frac{1}{6}]. \tag{5.2.46}$$

Expanding the left hand side of equation (5.2.46) one has

$$m^2\delta^2-\frac{1}{2}m^2(m+1)\delta+\frac{1}{16}m^2(m+1)^2-m\delta+\frac{1}{4}m(m+1)+\frac{1}{6} \tag{5.2.47}$$

and expanding and summing the right hand side of equation (5.2.46) one has

$$m^2\delta^2 -\frac{1}{2}m^2(m+1)\delta+ \frac{1}{24}m^2(m+1)(2m+1)-m^2\delta +\frac{1}{4}m^2(m+1)+\frac{m}{6} . \tag{5.2.48}$$

Collecting like terms we find

$$\delta = \frac{(m+1)(m^2+12m+8)}{48m} \tag{5.2.49}$$

hence

$$\rho=\rho_0=\nu- \frac{(m+1)(m^2+12m+8)}{24m} . \tag{5.2.50}$$

We find then that

$$\phi_\eta(\theta)\Big|_{\rho=\rho_0} = (1-2\theta)^{-\frac{1}{2}(m-1)} \exp\{\sum_{r=2}^{\infty} \pi_r[(1-2\theta)^{-r}-1]\} \tag{5.2.51}$$

where

$$\pi_2=\pi_2^*\Big|_{\rho=\rho_0} = \frac{(m+1)[3m^6-36m^5-583m^4-336m^3+160m^2-192]}{6912\rho_0^2m^2}. \tag{5.2.52}$$

Thus the cumulative distribution function of $\eta=-\rho_0\log\lambda_1$ is found from

$$\Pr\{-\rho_0\log\lambda_1\le\lambda\}= \Pr\{\chi^2_{(m-1)} \le\lambda\}+ \pi_2(\Pr\{\chi^2_{(m+3)}\le\lambda\}- \Pr\{\chi^2_{(m-1)}\le\lambda\})+ R'(\rho_0^{-3}) \tag{5.2.53}$$

with $R'(\rho_0^{-3})$ a remainder involving terms in $\rho_0^{-3},\rho_0^{-4},\ldots$.

Asymptotically we have that the distribution of

$$-\rho_0 \log\lambda_1 = -\rho_0\sum_{i=1}^{m}\log\left(\frac{d_i}{\frac{1}{m}\sum_{j=1}^{m}d_j}\right) \tag{5.2.54}$$

tends to that of a chi-square variate with (m-1) degrees of freedom.

5.3 The Distribution of T a Function of $\{g_{ij}:0\le j<i\le m\}$.

In section 5 of Chapter II we derived the distribution of $\{g_{ij}:0\le j<i\le m\}$ conditional on $\{d_j:0\le j\le m\}$ when V was arbitrary and when $V\varepsilon\tilde{V}$. In particular, when $V\varepsilon\tilde{V}$ and m=4 we find from equation (2.5.15)

$$
\begin{aligned}
f(g_{10},g_{20},g_{30},g_{40},g_{21},g_{31},g_{41},g_{32},g_{42},g_{43}|d_0,d_1,d_2,d_3) = {} \\
\frac{1}{(2\pi\sigma^2 d_0^{-1})^2} \exp\{-\frac{d_0}{2\sigma^2}[(g_{10}-\alpha_1)^2+(g_{20}-\alpha_2 g_{10})^2 \\
+ (g_{30}-\alpha_3 g_{20})^2 + (g_{40}-\alpha_4 g_{30})^2]\} \\
\frac{1}{(2\pi\sigma^2 d_1^{-1})^{\frac{3}{2}}} \exp\{-\frac{d_1}{2\sigma^2}[(g_{21}-\alpha_2)^2+(g_{31}-\alpha_3 g_{21})^2+(g_{41}-\alpha_4 g_{31})^2]\} \\
\frac{1}{(2\pi\sigma^2 d_2^{-1})} \exp\{-\frac{d_2}{2\sigma^2}[(g_{32}-\alpha_3)^2+(g_{42}-\alpha_4 g_{32})^2]\} \\
\frac{1}{(2\pi\sigma^2 d_3^{-1})^{\frac{1}{2}}} \exp\{-\frac{d_3}{2\sigma^2}(g_{43}-\alpha_4)^2\} \quad : \begin{array}{l} -\infty<g_{ij}<\infty, \\ 0\le j<i\le 4 . \end{array}
\end{aligned}
\tag{5.3.1}
$$

If in (5.3.1) we replace $(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$ by their estimators $(g_{10},g_{21},g_{32},g_{43})$ we obtain the statistics $\{(g_{20}-g_{21}g_{10})$, $(g_{30}-g_{32}g_{20})$, $(g_{40}-g_{43}g_{30})$, $(g_{31}-g_{32}g_{21})$, $(g_{41}-g_{43}g_{31})$, $(g_{42}-g_{43}g_{32})\}$. These statistics indicate a departure from the model. We shall investigate the distribution of a function of these statistics. No compact expressions have been found for general m so we present here the case for m=4. The case for general m follows directly from the case m=4. In what follows we shall use conditional distributions and for convenience shall let $\mathcal{H}$ denote a set of fixed variables.

Consider first the statistics $\{(g_{40}-g_{43}g_{30}), (g_{41}-g_{43}g_{31}), (g_{42}-g_{43}g_{32})\}$. We shall determine their joint distribution conditional on $\mathcal{H} = \{d_0, d_1, d_2, d_3, d_4, g_{30}, g_{31}, g_{32}\}$. We have from (5.3.1) that g_{40}, g_{41}, g_{42}, and g_{43} are normally distributed conditional on $\mathcal{H}$ so we need to find the moments of $\{(g_{40}-g_{43}g_{30}), (g_{41}-g_{43}g_{31}), (g_{42}-g_{43}g_{32})\}$ conditional on $\mathcal{H}$ to determine their joint distribution.

We have that

$$\mathcal{E}(g_{40}-g_{43}g_{30})\,|\,\mathcal{H} = \mathcal{E}\,[\mathcal{E}(g_{40}-g_{43}g_{30})]\,|\,\mathcal{H}\ , \tag{5.3.2}$$

where the inner expectation is with respect to g_{40} and the outer expectation is with respect to g_{43}. From (5.3.1) we obtain $\mathcal{E}(g_{40}|\mathcal{H})$ by inspection to be $\alpha_4 g_{30}$, hence

$$\mathcal{E}(g_{40}-g_{43}g_{30})\,|\,\mathcal{H} = \mathcal{E}\,[(\alpha_4-g_{43})g_{30}]\,|\,\mathcal{H} \tag{5.3.3}$$

$$= 0 \tag{5.3.4}$$

since by inspection of (5.3.1) $\mathcal{E}(g_{43})\,|\,\mathcal{H} = \alpha_4$.

In the same way

$$\begin{aligned}\mathcal{E}(g_{41}-g_{43}g_{31})\,|\,\mathcal{H} &= \mathcal{E}\,[\mathcal{E}(g_{41}-g_{43}g_{31})]\,|\,\mathcal{H}\\ &= \mathcal{E}\,[(\alpha_4-g_{43})g_{31}]\,|\,\mathcal{H}\\ &= 0\ ,\end{aligned} \tag{5.3.5}$$

and

$$\begin{aligned}\mathcal{E}(g_{42}-g_{43}g_{32})\,|\,\mathcal{H} &= \mathcal{E}\,[\mathcal{E}(g_{42}-g_{43}g_{32})]\,|\,\mathcal{H}\\ &= \mathcal{E}\,[(\alpha_4-g_{43})g_{32}]\,|\,\mathcal{H}\\ &= 0\ .\end{aligned} \tag{5.3.6}$$

The second moments are handled identically and we find

$$\mathcal{E}(g_{40}-g_{43}g_{30})^2|\mathcal{H} = \mathcal{E}[\mathcal{E}(g_{40}^2-2g_{40}g_{43}g_{30}+g_{43}^2g_{30}^2)]|\mathcal{H}$$

$$= \mathcal{E}[\frac{\sigma^2}{d_0} +\alpha_4^2g_{30}^2-2\alpha_4g_{43}g_{30}^2+g_{43}^2g_{30}^2]|\mathcal{H}$$

$$= [\frac{\sigma^2}{d_0} + \alpha_4^2g_{30}^2-2\alpha_4^2g_{30}^2+g_{30}^2(\frac{\sigma^2}{d_3} + \alpha_4^2)]$$

$$= \frac{\sigma^2}{d_0} + g_{30}^2\frac{\sigma^2}{d_3}, \tag{5.3.7}$$

$$\mathcal{E}(g_{41}-g_{43}g_{31})^2|\mathcal{H} = \mathcal{E}[\mathcal{E}(g_{41}^2-2g_{41}g_{43}g_{31}+g_{43}^2g_{31}^2)]|\mathcal{H}$$

$$= \mathcal{E}[\frac{\sigma^2}{d_1} + \alpha_4^2g_{31}^2-2\alpha_4g_{43}g_{31}^2+g_{43}^2g_{31}^2]|\mathcal{H}$$

$$= \frac{\sigma^2}{d_1} + \alpha_4^2g_{31}^2-2\alpha_4^2g_{31}^2+g_{31}^2(\frac{\sigma^2}{d_3} + \alpha_4^2)$$

$$= \frac{\sigma^2}{d_1} + g_{31}^2\frac{\sigma^2}{d_3}, \tag{5.3.8}$$

and

$$\mathcal{E}(g_{42}-g_{43}g_{32})^2|\mathcal{H} = \mathcal{E}[\mathcal{E}(g_{42}^2-2g_{42}g_{43}g_{32}+g_{43}^2g_{32}^2)]|\mathcal{H}$$

$$= \mathcal{E}[\frac{\sigma^2}{d_2} +\alpha_4^2g_{32}^2-2\alpha_4g_{43}g_{32}^2+g_{43}^2g_{32}^2]|\mathcal{H}$$

$$= \frac{\sigma^2}{d_2} + \alpha_4^2g_{32}^2-2\alpha_4^2g_{32}^2+g_{32}^2(\frac{\sigma^2}{d_3} + \alpha_4^2)$$

$$= \frac{\sigma^2}{d_2} + g_{32}^2\frac{\sigma^2}{d_3} . \tag{5.3.9}$$

The cross product terms are handled similarly and we have that

$$\mathcal{E}(g_{40}-g_{43}g_{30})(g_{41}-g_{43}g_{31})|\mathcal{H} = \mathcal{E}\{\mathcal{E}[\mathcal{E}(g_{40}g_{41}-g_{40}g_{43}g_{31}-g_{41}g_{43}g_{30}$$
$$+g_{43}^2g_{30}g_{31})]\}|\mathcal{H}, \tag{5.3.10}$$

where the inner, middle, and outer expectations are with respect to g_{40}, g_{41}, and g_{43}, respectively. Continuing we find

$$\mathcal{E}(g_{40}-g_{43}g_{30})(g_{41}-g_{43}g_{31})|\mathcal{H} = \mathcal{E}\{\mathcal{E}[\alpha_4 g_{30}g_{41}-\alpha_4 g_{30}g_{43}g_{31}-g_{41}g_{43}g_{30} + g_{43}^2 g_{30}g_{31}]\}|\mathcal{H}$$

$$= \mathcal{E}\{\alpha_4^2 g_{30}g_{31}-\alpha_4 g_{30}g_{43}g_{31}-\alpha_4 g_{43}g_{30}g_{31} + g_{43}^2 g_{30}g_{31}\}|\mathcal{H}$$

$$= \alpha_4^2 g_{30}g_{31}-2\alpha_4^2 g_{30}g_{31}+(\frac{\sigma^2}{d_3}+\alpha_4^2)g_{30}g_{31}$$

$$= \frac{\sigma^2}{d_3} g_{30}g_{31}, \tag{5.3.11}$$

$$\mathcal{E}(g_{40}-g_{43}g_{30})(g_{42}-g_{43}g_{32})|\mathcal{H} = \mathcal{E}\{\mathcal{E}[\mathcal{E}(g_{40}g_{42}-g_{40}g_{43}g_{32}-g_{42}g_{43}g_{30}+ + g_{43}^2 g_{30}g_{32})]\}|\mathcal{H}$$

$$= \mathcal{E}\{\mathcal{E}[\alpha_4 g_{30}g_{42}-\alpha_4 g_{30}g_{43}g_{32}-g_{42}g_{43}g_{30} + g_{43}^2 g_{30}g_{32}]\}|\mathcal{H}$$

$$=\mathcal{E}\{\alpha_4^2 g_{30}g_{32}-\alpha_4^2 g_{30}g_{32}-\alpha_4 g_{43}g_{30}g_{32} + g_{43}^2 g_{30}g_{32}\}|\mathcal{H}$$

$$= \alpha_4^2 g_{30}g_{32}-\alpha_4^2 g_{30}g_{32}-\alpha_4^2 g_{30}g_{32} + (\frac{\sigma^2}{d_3} + \alpha_4^2)g_{30}g_{32}$$

$$= \frac{\sigma^2}{d_3} g_{30}g_{32}, \tag{5.3.12}$$

and

$$\mathcal{E}(g_{41}-g_{43}g_{31})(g_{42}-g_{43}g_{32}) \mid \mathcal{H} = \mathcal{E}\{\mathcal{E}[\mathcal{E}(g_{41}g_{42}-g_{41}g_{43}g_{32}-g_{42}g_{43}g_{31}$$
$$+g_{43}^2g_{31}g_{32})]\}|\mathcal{H}$$

$$= \mathcal{E}\{\mathcal{E}[\alpha_4 g_{31}g_{42}-\alpha_4 g_{31}g_{43}g_{32}-g_{42}g_{43}g_{31}$$
$$+g_{43}^2g_{31}g_{32}]\}|\mathcal{H}$$

$$= \mathcal{E}\{\alpha_4^2 g_{31}g_{32}-\alpha_4 g_{31}g_{43}g_{32}-\alpha_4 g_{32}g_{43}g_{31}$$
$$+g_{43}^2g_{31}g_{32}\}|\ \mathcal{H}$$

$$= \alpha_4^2 g_{31}g_{32}-\alpha_4^2 g_{31}g_{32}-\alpha_4^2 g_{31}g_{32}+$$
$$+(\frac{\sigma^2}{d_3}+\alpha_4^2)\, g_{31}g_{32}$$

$$= \frac{\sigma^2}{d_3} g_{31}g_{32}\ . \tag{5.3.13}$$

Hence we find

$$\begin{pmatrix} g_{40}-g_{43}g_{30} \\ g_{41}-g_{43}g_{31} \\ g_{42}-g_{43}g_{32} \end{pmatrix} |\mathcal{H} \ \sim N_3(\underline{\mu},\ \sigma^2 V_3) \tag{5.3.14}$$

where

$$\underline{\mu} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\ , \tag{5.3.15}$$

and

$$V_3 = \begin{bmatrix} (\frac{1}{d_0} + \frac{g_{30}^2}{d_3}) & g_{30}g_{31}d_3^{-1} & g_{30}g_{32}d_3^{-1} \\ g_{30}g_{31}d_3^{-1} & (\frac{1}{d_1} + \frac{g_{31}^2}{d_3}) & g_{31}g_{32}d_3^{-1} \\ g_{30}g_{32}d_3^{-1} & g_{31}g_{32}d_3^{-1} & (\frac{1}{d_2} + \frac{g_{32}^2}{d_3}) \end{bmatrix} . \tag{5.3.16}$$

It is well known that if $\underline{x}:(n\text{x}1) \sim N_n(\underline{0},\Sigma)$ then $\underline{x}'\Sigma^{-1}\underline{x} \sim \chi_n^2(0)$. Therefore it follows that

$$r_4 = \begin{pmatrix} g_{40}-g_{43}g_{30} \\ g_{41}-g_{43}g_{31} \\ g_{42}-g_{43}g_{32} \end{pmatrix}' V_3^{-1} \begin{pmatrix} g_{40}-g_{43}g_{30} \\ g_{41}-g_{43}g_{31} \\ g_{42}-g_{43}g_{32} \end{pmatrix} \Big| \mathcal{H} \sim \sigma^2\chi_3^2(0) , \tag{5.3.17}$$

where the subscript on r equals the first subscript on g. Since the distribution above is functionally independent of the variables in $\mathcal{H}$ we have the unconditional distribution also, that is,

$$r_4 \sim \sigma^2\chi_3^2(0) . \tag{5.3.18}$$

Putting $\mathcal{H} = \{d_0,d_1,d_2,d_3,d_4,g_{20},g_{21}\}$, we now find the joint distribution of $\{(g_{30}-g_{32}g_{20}),\ (g_{31}-g_{32}g_{21})\}$. Proceeding as before

$$\begin{aligned} \mathcal{E}(g_{30}-g_{32}g_{20}) \mid \mathcal{H} &= \mathcal{E}[\mathcal{E}(g_{30}-g_{32}g_{20})] \mid \mathcal{H} \\ &= \mathcal{E}[(\alpha_3-g_{32})g_{20}] \mid \mathcal{H} \\ &= 0 , \end{aligned} \tag{5.3.19}$$

and

$$\begin{aligned} \mathcal{E}(g_{31}-g_{32}g_{21}) \mid \mathcal{H} &= \mathcal{E}[\mathcal{E}(g_{31}-g_{32}g_{21})] \mid \mathcal{H} \\ &= \mathcal{E}[(\alpha_3-g_{32})g_{21}] \mid \mathcal{H} \\ &= 0 . \end{aligned} \tag{5.3.20}$$

The second moments are

$$\begin{aligned}
\mathcal{E}(g_{30}-g_{32}g_{20})^2|\mathcal{H} &= \mathcal{E}[\mathcal{E}(g_{30}^2-2g_{30}g_{32}g_{20}+g_{32}^2g_{20}^2)]|\mathcal{H} \\
&= \mathcal{E}[\frac{\sigma^2}{d_0}+\alpha_3^2g_{20}^2-2\alpha_3g_{32}g_{20}^2+g_{32}^2g_{20}^2]|\mathcal{H} \\
&= \frac{\sigma^2}{d_0}+\alpha_3^2g_{20}^2-2\alpha_3^2g_{20}^2+(\frac{\sigma^2}{d_2}+\alpha_3^2)g_{20}^2 \\
&= \frac{\sigma^2}{d_0}+\frac{\sigma^2}{d_2}g_{20}^2 \ ,
\end{aligned} \tag{5.3.21}$$

and

$$\begin{aligned}
\mathcal{E}(g_{31}-g_{32}g_{21})^2|\mathcal{H} &= \mathcal{E}[\mathcal{E}(g_{31}^2-2g_{31}g_{32}g_{21}+g_{32}^2g_{21}^2)]|\mathcal{H} \\
&= \mathcal{E}[\frac{\sigma^2}{d_1}+\alpha_3^2g_{21}^2-2\alpha_3g_{32}g_{21}^2+g_{32}^2g_{21}^2]|\mathcal{H} \\
&= \frac{\sigma^2}{d_1}+\alpha_3^2g_{21}^2-2\alpha_3^2g_{21}^2+(\frac{\sigma^2}{d_2}+\alpha_3^2)g_{21}^2 \\
&= \frac{\sigma^2}{d_1}+\frac{\sigma^2}{d_2}g_{21}^2 \ .
\end{aligned} \tag{5.3.22}$$

The cross product term is

$$\begin{aligned}
\mathcal{E}(g_{30}-g_{32}g_{20})(g_{31}-g_{32}g_{21})|\mathcal{H} &= \mathcal{E}\{\mathcal{E}[\mathcal{E}(g_{30}g_{31}-g_{30}g_{32}g_{21}-g_{31}g_{31}g_{20} \\
&\quad +g_{32}^2g_{20}g_{21})]\mathcal{H} \\
&= \mathcal{E}\{\mathcal{E}[\alpha_3g_{20}g_{31}-\alpha_3g_{20}g_{32}g_{21}-g_{31}g_{32}g_{20} \\
&\quad +g_{32}^2g_{20}g_{21}]\}|\mathcal{H} \\
&= \mathcal{E}\{\alpha_3^2g_{20}g_{21}-\alpha_3g_{20}g_{32}g_{21}-\alpha_3g_{21}g_{32}g_{20} \\
&\quad +g_{32}^2g_{20}g_{21}]\}|\mathcal{H} \\
&= \alpha_3^2g_{20}g_{21}-\alpha_3^2g_{20}g_{21}+\alpha_3^2g_{20}g_{21} \\
&\quad +(\frac{\sigma^2}{d_2}+\alpha_3^2)g_{20}g_{21} \\
&= \frac{\sigma^2}{d_2}g_{20}g_{21} \ .
\end{aligned} \tag{5.3.23}$$

Hence we find that

$$\begin{pmatrix} g_{30}-g_{32}g_{20} \\ g_{31}-g_{32}g_{21} \end{pmatrix} \Big| \mathcal{H} \sim N_2(\underline{\mu},\sigma^2 V_2) \tag{5.3.24}$$

where

$$\underline{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{5.3.25}$$

and

$$V_2 = \begin{bmatrix} (\frac{1}{d_0} + \frac{g_{20}^2}{d_2}) & g_{20}g_{21}d_2^{-1} \\ g_{20}g_{21}d_2^{-1} & (\frac{1}{d_1} + \frac{g_{21}^2}{d_2}) \end{bmatrix} . \tag{5.3.26}$$

Now form

$$r_3 = \begin{pmatrix} g_{30}-g_{32}g_{20} \\ g_{31}-g_{32}g_{21} \end{pmatrix}' V_2^{-1} \begin{pmatrix} g_{30}-g_{32}g_{20} \\ g_{31}-g_{32}g_{21} \end{pmatrix} \tag{5.3.27}$$

where

$$r_3 | \mathcal{H} \sim \sigma^2 \chi_2^2 (0) . \tag{5.3.28}$$

By the same arguments as before we have that the distribution of r_3 is functionally independent of the elements of $\mathcal{H}$ and hence

$$r_3 \sim \sigma^2 \chi_2^2 (0) \tag{5.3.29}$$

unconditionally. Moreover r_4 is independent of r_3 since the distribution of r_4 is functionally independent of the elements of r_3.

Finally we consider $(g_{20}-g_{21}g_{10})$ and put

$\mathcal{H} = \{d_0,d_1,d_2,d_3,d_4,g_{10}\}$.

$$\begin{aligned}\mathcal{E}(g_{20}-g_{21}g_{10})|\mathcal{H} &= \mathcal{E}[\mathcal{E}(g_{20}-g_{21}g_{10})]|\mathcal{H}\\ &= \mathcal{E}[(\alpha_2-g_{21})g_{10}]|\mathcal{H}\\ &= 0 \quad ,\end{aligned} \tag{5.3.30}$$

and

$$\begin{aligned}\mathcal{E}(g_{20}-g_{21}g_{10})^2|\mathcal{H} &= \mathcal{E}[\mathcal{E}(g_{20}^2-2g_{20}g_{21}g_{10}+g_{21}^2g_{10}^2)]|\mathcal{H}\\ &= \mathcal{E}[\frac{\sigma^2}{d_0}+\alpha_2^2g_{10}^2-2\alpha_2g_{21}g_{10}^2+g_{21}^2g_{10}^2]|\mathcal{H}\\ &= \frac{\sigma^2}{d_0}+\alpha_2^2g_{10}^2-2\alpha_2^2g_{10}^2+(\frac{\sigma^2}{d_1}+\alpha_2^2)g_{10}^2\\ &= \frac{\sigma^2}{d_0}+\frac{\sigma^2}{d_1}g_{10}^2 \quad .\end{aligned} \tag{5.3.31}$$

Hence

$$(g_{20}-g_{21}g_{10})|\mathcal{H} \sim N(\mu,\sigma^2V_1) \tag{5.3.32}$$

where

$$\mu = 0 \tag{5.3.33}$$

and

$$V_1 = \frac{1}{d_0}+\frac{g_{10}^2}{d_1} \quad . \tag{5.3.34}$$

With

$$r_2 = \frac{(g_{20}-g_{21}g_{10})^2}{(\frac{1}{d_0}+\frac{g_{10}^2}{d_1})} \tag{5.3.35}$$

then

$$r_2|\mathcal{H} \sim \sigma^2\chi_1^2(0) \quad . \tag{5.3.36}$$

Since the distribution of r_2 is functionally independent of the elements of $\mathcal{H}$ we have, unconditionally that, r_2 is chi-square with one degree of freedom. Also r_2 is independent of r_3 and r_4 since their distribution is functionally independent of the elements of r_2. Since the three statistics are independent we may add them to get

$$R = (r_2+r_3+r_4) \sim \sigma^2\chi_6^2(0) \ . \qquad (5.3.37)$$

We note that the distribution of R depends on the nuisance parameter σ^2, to eliminate this parameter we consider

$$T = \frac{(R/6)}{\sigma_*^2} \sim F_{4\nu-10}^{6}(0) \ . \qquad (5.3.38)$$

Since R is independent of $\{d_0,d_1,d_2,d_3,d_4\}$ then its independent of σ_*^2 and T is the ratio of two independent chi-square variates divided by their degrees of freeom, that is, T has the F distribution.

Before extending this result to general m we note that the dispersion matrix V_3 of equation (5.3.16) may be written

$$V_3 = D^{-1}(3) + \underline{\gamma}_3\underline{\gamma}_3' \qquad (5.3.39)$$

where

$$D(3) = \text{diag } (d_0,d_1,d_2) \qquad (5.3.40)$$

and

$$\underline{\gamma}_3 = \frac{1}{\sqrt{d_3}} (g_{30},g_{31},g_{32})' \ . \qquad (5.8.41)$$

Since V_3 may be written in this form its inverse can be obtained from the Binomial Inverse Theorem, found in Press [13], which states

$$(D^{-1}(3)+\underline{\gamma}_3\underline{\gamma}_3')^{-1} = D(3)-\frac{D(3)\underline{\gamma}_3\underline{\gamma}_3'D(3)}{1+\underline{\gamma}_3'D(3)\underline{\gamma}_3} \ . \qquad (5.3.42)$$

This is very useful in the actual computing of the statistic T. It follows that V_2 is the same form and can be written

$$V_2 = D^{-1}(2) + \underline{\gamma}_2\underline{\gamma}_2' \qquad (5.3.43)$$

where

$$D(2)=\text{diag}(d_0,d_1) \tag{5.3.44}$$

and

$$\underline{\gamma}_2=\frac{1}{\sqrt{d_2}}(g_{20},g_{21})' . \tag{5.3.45}$$

In general we can form

$$r_j=\underline{g}_j'V_{j-1}^{-1}\underline{g}_j : 2\le j\le m , \tag{5.3.46}$$

where

$$\underline{g}_j=[(g_{j0}-g_{j-1,0}g_{j,j-1}),(g_{j1}-g_{j-1,1}g_{j,j-1}),\ldots,$$
$$(g_{j,j-2}-g_{j-1,j-2}g_{j,j-1})]' : 2\le j\le m , \tag{5.3.47}$$

and

$$V_{j-1}=D^{-1}(j-1)+\underline{\gamma}_{j-1}\underline{\gamma}_{j-1}' : 2\le j\le m , \tag{5.3.48}$$

with

$$D(j-1)=\text{diag}(d_0,d_1,\ldots,d_{j-2}) : 2\le j\le m , \tag{5.3.49}$$

and

$$\underline{\gamma}_{j-1}=\frac{1}{\sqrt{d_{j-1}}}(g_{j-1,0};g_{j-1,1};\ldots;g_{j-1,j-2})' : 2\le j\le m . \tag{5.3.50}$$

Following the pattern given for m=4, when $V\varepsilon\tilde{V}$

$$r_j \sim \sigma^2\chi^2_{j-1}(0) : 2\le j\le m \tag{5.3.51}$$

and they are mutually independent so that

$$R = \sum_{j=2}^{m} r_j \sim \sigma^2\chi^2_{\frac{1}{2}m(m-1)}(0), \quad (V\varepsilon\tilde{V}) \tag{5.3.52}$$

and finally

$$T=\frac{(R/\frac{1}{2}m(m-1))}{\sigma_*^2} \sim F^{\frac{1}{2}m(m-1)}_{m\nu-\frac{1}{2}m(m+1)}(0), \quad (V\varepsilon\tilde{V}) . \tag{5.3.53}$$

No attempt has been made to find the distribution of T when $V\notin\tilde{V}$, but a computor simulation indicates, as we would expect, that T is stochastically larger in $V\notin\tilde{V}$ than in $V\varepsilon\tilde{V}$.

5.4 Asymptotic Performance of $(-\rho_0\log\lambda_1)$ and T.

In section 3 we derived the asymptotic distribution of $-\rho_0\log\lambda_1$ and showed it to have a chi-square distribution with (m-1) degrees of freedom. In that section we also showed that the distribution of $-\rho_0\log\lambda_1$ is independent of $S=\frac{1}{m}\sum_{j=1}^{m} d_j$ or equivalently of $\sum_{j=1}^{m} d_j$. In section 4 we showed that R has a chi-square distribution with $\frac{1}{2}m(m-1)$ degrees of freedom, independent of $\{d_0,d_1,\ldots,d_m\}$ and hence independent of $-\rho_0\log\lambda_1$ and $\sum_{j=1}^{m} d_j$. Since R is independent of $\sum_{j=1}^{m} d_j$ it is independent of σ_*^2 and hence we formed T equal to the ratio of R and σ_*^2 divided by the appropriate constants to form an F distribution with $\frac{1}{2}m(m-1)$ degrees of freedom in the numerator and $[m\nu-\frac{1}{2}m(m+1)]$ degrees of freedom in the denominator. Since both R and σ_*^2 are independent of $-\rho_0\log\lambda_1$ then so is T. Now the distribution of $\frac{1}{2}m(m-1)T$ tends to that of a chi-square variate with $\frac{1}{2}m(m-1)$ degrees of freedom as $\nu\to\infty$. Since T and $-\rho_0\log\lambda_1$ are independent we have

$$\lim_{\nu\to\infty} \{-\rho_0\log\lambda_1 + \tfrac{1}{2}m(m-1)T\} \sim \chi^2_{\frac{1}{2}(m-1)(m+2)}(0) \quad . \qquad (5.4.1)$$

It has been shown by Wilks [18] that under certain regularity conditions $-2\log\lambda$ will be asymptotically distributed as a chi-square with ℓ degrees of freedom under the null hypothesis, where λ denotes the likelihood ratio. The degrees of freedom, ℓ, may be computed from $(\ell_1-\ell_0)$ where ℓ_1 equals the number of parameters estimated under the alternative hypothesis (H_1) and ℓ_0 equals the number of parameters estimated under the null hypothesis (H_0). For the problem here we find that under H_1 V is arbitrary and we must estimate all $\frac{1}{2}(m+1)(m+2)=\ell_1$ different parameters. Under H_0 there are only $(m+2)=\ell_0$ unknown parameters to estimate and hence

$$\ell = \ell_1 - \ell_0$$

$$= \tfrac{1}{2}(m+2)(m-1) \quad . \tag{5.4.2}$$

That is, the asymptotic distribution of $-2\log\lambda$ and $(-\rho_0\log\lambda_1+\frac{1}{2}m(m-1)T)$ both agree, under the null hypothesis. Hence both methods are asymptotically equivalent under the null hypothesis.

We note that since $-\rho_0\log\lambda_1$ and T are independent that Fisher's method of combining independent tests may be used in place of $(-\rho_0\log\lambda_1+\frac{1}{2}m(m-1)T)$. Fisher's method would be especially appropriate if the sample size is small.

Chapter VI

COMPUTER SIMULATIONS AND AN APPLICATION

6.1 Introduction

A computer simulation of the generalized autoregressive process was performed thirty times. Each simulation had fifty vector observations with each vector observation having six measures including the initial measure. Specific values were given $(\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5)$, σ^2, and β^2 and they were (0.80, 0.60, 0.50, 0.30, 0.20), 1.00, and 4.00, respectively.

The simulations were made using a computer program written for the IBM 360 computer. The output from the program includes

(1) the data used in the analysis

(2) the mean for each time period

(3) the cross product matrix

(4) the G matrix

(5) the diagonal elements of D

(6) the starred estimates of $\{\alpha_i: 1\le i\le m\}$, σ^2, and β^2.

(7) the maximum likelihood estimates of $\{\alpha_i: 1\le i\le m\}$, σ^2, and β^2.

(8) the values of $-\rho_0\log\lambda_1$ and T used in testing the adequacy of the model.

The main purpose of the simulations was to see if the starred estimators would perform well. In keeping with this we present only the starred and maximum likelihood estimates for $\{\alpha_i: 1 \le i \le m\}$, σ^2, and β^2.

An application of the theory was made using data from a drug study at the University of Florida. This study was directed by Dr. Arlan L. Rosenbloom. Each patient was infused with glucose and observations were taken on the patient's level of calcium prior to infusion and at 90 minute intervals thereafter for four additional observations.

6.2 Computer Simulation Results

Each of the estimates was tested against its true value at the .05 level of significance. On the average then we would expect to reject two out of the thirty estimates by chance alone. Those that were significantly different from the actual value are listed with an asterisk. Counting the number of tests that were accepted as a measure of the estimator's goodness we find α_{*1} gave 28 acceptable estimates out of 30. Since α_{*1} is identical to the maximum likelihood estimator, $\hat{\alpha}_1$, there is no comparison. α_{*2} gave acceptable estimates in all 30 runs while $\hat{\alpha}_2$ gave 28. Estimating $\alpha_3=.50$, the starred estimators did slightly better with α_{*3} giving 28 acceptable estimates and $\hat{\alpha}_3$ giving 27. α_{*4} gave acceptable estimates in all runs while $\hat{\alpha}_4$ gave 29. The last estimators, α_{*5} and $\hat{\alpha}_5$, both gave 28 acceptable estimates. We note that whenever the starred estimate was rejected so was the maximum likelihood estimate, but not conversely.

Tests were also performed on the estimates of σ^2 and β^2. In order to test both σ_*^2 and $\hat{\sigma}^2$ an approximation to the distribution of chi-square given by Wilson and Hilferty [19] was used. Their result is that $(\chi^2/\nu)^{1/3}$ is approximately normally distributed with mean, $1-2/(9\nu)$, and variance, $2/(9\nu)$. This result and a discussion are also given in Kendall and Stuart [11]. The results of the tests showed that the starred estimator gave 25 acceptable estimates while the maximum likelihood estimator gave 24. Again both estimates were rejected on the same runs, with one exception, when the maximum likelihood estimate was too high. All of the rejections for the starred estimates were caused by under estimating the true value.

The starred estimates and maximum likelihood estimates performed equally well in estimating β^2. Both gave acceptable estimates 26 out of the 30 runs. Of the four incorrect estimates both were high on three and low on one. They both gave poor estimates on the same runs.

Overall the starred estimators performed as well or better than the maximum likelihood estimators. As can be seen by the means and standard deviations at the bottom of Tables 1 through 3, both estimates are very close to the true value. The mean of the maximum likelihood estimates is closer to the true value for α_2, α_4, and α_5, but not for α_3, σ^2, or β^2. Also we note that the sample standard deviations are smaller for the maximum likelihood estimates except for β^2. None of the differences seem to be appreciable in any case.

Table 1

ESTIMATES OF α_1, α_2, AND α_3 FOR

COMPUTER SIMULATED PROCESS

Run Number	$\alpha_1=.80$ $\alpha_{*1}=\hat{\alpha}_1$	$\alpha_2=.60$ α_{*2}	$\alpha_2=.60$ $\hat{\alpha}_2$	$\alpha_3=.50$ α_{*3}	$\alpha_3=.50$ $\hat{\alpha}_3$
1	0.766	0.484	0.503	0.415	0.421
2	0.770	0.691	0.625	0.274	0.414
3	0.767	0.639	0.647	0.623	0.564
4	0.747	0.427	0.475	0.620	0.518
5	0.643*	0.576	0.611	0.420	0.356
6	0.842	0.749	0.722	0.423	0.553
7	0.723	0.490	0.650	0.469	0.452
8	0.839	0.488	0.661	0.485	0.521
9	0.906	0.373	0.492	0.837*	0.719*
10	0.902	0.565	0.514	0.556	0.540
11	0.892	0.428	0.554	0.851*	0.691*
12	0.674	0.730	0.663	0.422	0.493
13	0.799	0.503	0.528	0.389	0.333
14	0.747	0.754	0.655	0.775	0.541
15	0.826	0.725	0.643	0.579	0.643
16	0.748	0.696	0.592	0.277	0.381
17	0.795	0.591	0.624	0.411	0.415
18	0.815	0.634	0.626	0.389	0.508
19	0.861	0.546	0.659	0.650	0.573
20	0.848	0.506	0.557	0.559	0.398
21	0.810	0.458	0.532	0.258	0.335
22	0.722	0.702	0.636	0.375	0.368
23	0.746	0.861	0.667	0.610	0.709
24	0.747	0.819	0.571	0.522	0.603
25	0.826	0.456	0.575	0.663	0.632
26	0.927	0.695	0.784*	0.255	0.429
27	0.741	0.459	0.514	0.529	0.600
28	0.652*	0.601	0.566	0.607	0.642
29	0.719	0.596	0.683	0.508	0.510
30	0.706	0.352	0.435*	0.568	0.521
Mean	0.784	0.586	0.599	0.511	0.513
Standard Deviation	0.074	0.134	0.079	0.158	0.112

*indicates estimate is significantly different from the true value, at the .05 level of significance.

Table 2

ESTIMATES OF α_4 AND α_5 FOR

COMPUTER SIMULATED PROCESSES

Run	$\alpha_4=.30$	$\alpha_4=.30$	$\alpha_5=.20$	$\alpha_5=.20$
Number	α_{*4}	$\hat{\alpha}_4$	α_{*5}	$\hat{\alpha}_5$
1	0.360	0.292	-0.032	0.060
2	0.353	0.358	-0.077	0.090
3	0.241	0.282	-0.042	-0.013
4	0.257	0.256	-0.156*	-0.071*
5	0.366	0.449	0.334	0.204
6	0.195	0.202	0.105	0.109
7	0.010	0.073	-0.050	-0.048
8	0.223	0.267	0.080	0.076
9	0.133	0.234	0.347	0.338
10	0.312	0.351	0.268	0.300
11	0.352	0.264	0.209	0.212
12	0.063	0.070*	0.053	0.093
13	0.461	0.383	0.119	0.175
14	0.269	0.225	0.096	0.102
15	0.431	0.295	0.259	0.224
16	0.364	0.271	0.225	0.187
17	0.444	0.352	-0.095	0.027
18	0.261	0.147	0.209	0.196
19	0.099	0.194	0.040	0.108
20	0.388	0.425	0.214	0.199
21	0.070	0.168	0.022	-0.026
22	0.425	0.405	0.301	0.307
23	0.262	0.349	-0.106*	-0.121*
24	0.180	0.284	0.236	0.250
25	0.195	0.170	0.390	0.363
26	0.264	0.207	0.330	0.327
27	0.018	0.102	0.135	0.110
28	0.392	0.420	0.381	0.294
29	0.276	0.260	0.084	0.099
30	0.168	0.261	-0.033	0.029
Mean	0.261	0.267	0.128	0.140
Standard Deviation	0.129	0.101	0.160	0.129

*indicates estimate is significantly different from the true value, at the .05 level of significance.

Table 3

ESTIMATES OF σ^2 AND β^2 FOR

COMPUTER SIMULATED PROCESSES

Run	σ^2=1.00	σ^2=1.00	β^2=4.00	β^2=4.00
Number	σ_*^2	$\hat{\sigma}^2$	β_*^2	$\hat{\beta}^2$
1	0.842	0.826	4.982	5.127
2	0.947	0.928	3.490	3.594
3	0.805*	0.788*	6.847*	7.055*
4	0.791*	0.788*	3.915	3.965
5	0.920	0.899	4.256	4.394
6	1.037	1.015	5.268	5.432
7	0.993	0.975	4.070	4.179
8	0.940	0.925	3.120	3.278
9	1.180	1.163*	2.396*	2.451*
10	0.818*	0.811*	5.888*	5.991*
11	0.811*	0.795*	3.164	3.255
12	0.969	0.935	4.235	4.426
13	1.029	1.003	3.194	3.305
14	0.925	0.926	4.096	4.129
15	1.016	0.992	4.700	4.853
16	1.028	0.992	3.227	3.426
17	0.963	0.929	4.122	4.309
18	0.816*	0.797*	5.684*	5.876*
19	1.011	0.981	3.191	3.317
20	0.989	0.993	3.904	3.922
21	1.082	1.058	3.458	3.569
22	1.010	0.988	4.296	4.429
23	0.840	0.831	3.004	3.061
24	1.081	1.053	3.637	3.768
25	1.141	1.103	4.413	4.608
26	0.994	0.971	3.494	3.608
27	0.996	0.957	4.023	4.223
28	0.894	0.861	4.517	4.733
29	0.986	0.942	4.209	4.444
30	0.865	0.867	4.217	4.243
Mean	0.957	0.936	4.102	4.232
Standard Deviation	0.102	0.096	0.944	0.971

*indicates estimate is significantly different from the true value, at the .05 level of significance.

6.3 Application

As discussed in section 1, patients were infused with glucose and measurements were taken on their calcium level, prior to infusion and four times later at 90-minute periods. The data are given in Table 4. Inspecting the means at each period given at the bottom of Table 4 we see that, on the average, initially the calcium reading was highest and infusion of glucose caused it to drop continually until the last time period where there is a mild increase in the level of calcium. In Table 5, both the starred estimates and the maximum likelihood estimates are given. Both estimators gave similar results for all of the parameters with α_4 having the largest value, probably reflecting the increase in the level of calcium from time period 3 to time period 4. Table 6 shows the standard deviations and 95% confidence intervals for $\alpha_{*1}, \alpha_{*2}, \alpha_{*3}$, and α_{*4}. The confidence intervals for α_{*1}, α_{*2}, and α_{*3} contain zero implying the parameters do not differ significantly from zero. This could have been guessed by noting the relatively small change in the mean level of calcium from one period to the next. Since the mean level rose in the last period the parameter α_{*4} is large and, as noted by the 95% confidence interval, is significantly different from zero.

In testing the adequacy of the model we found $-\rho_0 \log\lambda_1 =$ 4.91 and T = 1.77. Since the distribution of $-\rho_0 \log\lambda_1$ is approximately chi-square with 3 degrees of freedom we compare the calculated value against the tabulated value at the .05

level of significance. We find $\chi^2_{3,.05} = 7.81$, since the calculated value is less than this we accept the hypothesis of sphericity. The distribution of T is F with 6 degrees of freedom in the numerator and 114 degrees of freedom in the denominator. The upper 5% point of this distribution is $F^6_{114,.05} = 2.18$. Since the tabulated value is greater than the calculated value we accept the adequacy of the model.

To see how well the model fits the data we randomly selected patient number 18 and calculated his measurements for the four time periods using his pervious readings. Letting y_{*jk} denote the predicted value at time j of patient k we have that

$$y_{*jk} = \bar{y}_j + x_{jk} \quad : 1 \le j \le 4 \ ; \ 1 \le k \le 32 \ , \tag{6.3.1}$$

where

$$x_{jk} = \alpha_{*j} x_{j-1,k} : 1 \le j \le 4 \ ; \ 1 \le k \le 32 \ , \tag{6.3.2}$$

and

$$x_{jk} = y_{jk} - \bar{y}_j \quad : 1 \le j \le 4 \ ; \ 1 \le k \le 32 \ . \tag{6.3.3}$$

Hence we may write

$$y_{*jk} = \bar{y}_j - \alpha_{*j} \bar{y}_{j-1} + \alpha_{*j} y_{j-1,k} : 1 \le j \le 4 \ ; \ 1 \le k \le 32 \ . \tag{6.3.4}$$

Given that patient 18 had an initial reading of 9.9, the prediction for his 90-minute reading is

$$\begin{aligned} y_{*1,18} &= 9.15 - .137(9.64) + .137\ (9.9) \\ &= 9.20 \ . \end{aligned}$$

Similarly for the rest of the readings we find that

$$y_{*2,18} = 9.02 \ ,$$

$$y_{*3,18} = 9.06 \ ,$$

and $y_{*4,18} = 9.29$.

Table 4

LEVEL OF CALCIUM IN GRAMS PER LITER IN

PATIENTS INFUSED WITH GLUCOSE

Patient Number	Initial Period	90 Minutes	180 Minutes	270 Minutes	360 Minutes
1	9.1	9.2	8.9	8.5	8.3
2	10.0	9.2	9.5	9.2	8.4
3	10.1	9.8	10.1	8.0	9.7
4	10.0	10.1	9.1	9.4	9.5
5	9.7	8.9	9.1	9.1	9.2
6	9.5	8.7	9.1	8.3	8.6
7	9.5	9.4	9.3	9.6	9.3
8	10.1	9.2	9.3	9.1	8.7
9	9.6	8.9	9.3	9.4	8.9
10	9.1	9.3	9.0	9.0	9.0
11	9.6	8.8	8.9	8.8	9.4
12	9.3	9.4	9.3	9.4	9.7
13	10.2	9.5	9.8	9.9	9.8
14	9.2	8.8	9.4	8.9	8.2
15	9.6	9.4	8.9	8.9	9.0
16	10.1	9.0	9.1	9.2	9.1
17	9.4	8.5	8.5	8.6	8.7
18	9.9	8.9	9.5	9.5	9.8
19	10.4	8.9	9.4	8.3	8.1
20	9.0	8.8	8.5	8.5	8.4
21	9.7	9.6	9.4	8.4	8.8
22	10.2	8.1	9.0	8.9	9.4
23	9.2	10.3	9.0	8.7	8.7
24	9.7	8.9	9.1	9.1	9.2
25	9.0	8.4	8.1	8.7	8.7
26	9.4	9.2	9.2	9.2	9.0
27	9.4	8.9	8.8	8.7	9.0
28	10.1	9.8	9.1	8.8	9.0
29	9.8	9.3	9.5	9.3	9.6
30	9.6	9.1	8.4	8.6	8.5
31	9.5	8.8	8.5	8.7	9.2
32	9.4	9.6	9.4	9.4	9.5
Mean	9.64	9.15	9.11	8.94	9.01

Table 5

ESTIMATES OF THE PARAMETERS

FOR THE GLUCOSE STUDY

Type of Estimate	Parameters α_1	α_2	α_3	α_4	σ^2	β^2
Starred Estimate	0.137	0.334	0.305	0.506	0.173	0.843
Maximum Likelihood	0.137	0.383	0.312	0.590	0.174	0.854

Table 6

STANDARD DEVIATIONS AND 95% CONFIDENCE

INTERVALS FOR α_{*1}, α_{*2}, α_{*3}, AND α_{*4}

	Parameters			
	α_{*1}	α_{*2}	α_{*3}	α_{*4}
Standard Deviation	0.203	0.182	0.186	0.189
95% Confidence Interval	(-0.278, 0.652)	(-0.038, 0.706)	(-0.075, 0.685)	(0.119, 0.893)

Comparing these to the actual measurements of 8.9, 9.5, and 9.8 we see that the model gives reasonable predictions.

BIBLIOGRAPHY

[1] Anderson, T. W. (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York .

[2] Anderson, T. W. (1971). The Statistical Analysis of Time Series. Wiley, New York.

[3] Bahadur, R. R. (1960). Stochastic Comparison of Tests. Ann. Math. Statist., 31, 276-295.

[4] Box, G. E. P. (1949). A General Distribution Theory for a Class of Likelihood Criteria. Biometrika, 36, 317-346.

[5] Box, G. E. P. and Jenkins, G. M. (1970). Time Series Analysis (Forecasting and Control). Holden-Day, San Francisco.

[6] Cornish, E. A. (1954). The Multivariate t-Distribution Associated with a Set of Normal Sample Deviates. Australian Journal of Physics, 7, 531-542.

[7] Deemer, W. L. and Olkin, I. (1951). The Jacobians of Certain Matrix Tranformations Useful In Multivariate Analysis. Biometrika, 38, 345-367.

[8] Dunnett, C. W. and Sobel, M. (1954). A Bivariate Generalization of Students t-Distribution with Tables for Certain Special Cases. Biometrika, 41, 153-169.

[9] Feller, W. (1966). An Introduction to Probability Theory and Its Application, Vol. 2. Wiley, New York.

[10] Fisz, M. (1967). Probability Theory and Mathematical Statistics. Wiley, New York.

[11] Kendall, M. G. and Stuart, A. (1967). The Advanced Theory of Statistics, Vol. 2. Hafner, New York.

[12] Littell, R. C. and Folks, J. L. (1971). Asymptotic Optimality of Fisher's Method of Combining Independent Tests. Journal of the American Statistical Association, 66, 802-806.

[13] Press, S. J. (1972). Applied Multivariate Analysis. Holt, New York.

[14] Rao, C. R. (1952). Advanced Statistical Methods in Biometric Research. Wiley, New York.

[15] Saw, J. G. (1964). Likelihood Ratio Tests of Hypothesis on Multivariate Populations, Volume I: Distribution Theory. Virginia Polytechnical Institute, Blacksbury, Virginia.

[16] Saw, J. G. (1964). Likelihood Ratio Tests of Hypothesis on Multivariate Populations, Volume II: Tests of Hypothesis. Virginia Polytechnical Institute, Blacksburg, Virginia.

[17] Saw, J. G. (1973). Jacobians of Singular Transformations with Applications to Statistical Distribution Theory. Communications In Statistics, 1, 81-91.

[18] Wilks, S. S. (1938). The Large Sample Distribution of the Likelihood Ratio for Testing Composite Hypothesis Ann. Math. Statist., 9, 60.

[19] Wilson, E. B. and Hilferty, M. M. (1931). The Distribution of Chi-square. Proc. Nat. Acad. Sci., U.S.A., 17, 684.

BIOGRAPHICAL SKETCH

Darryl Jon Downing was born January 4, 1947 in Beaver Dam, Wisconsin, and was the youngest of the five children who William and Roberta Downing had. He spent most of his youth in Janesville, Wisconsin where he graduated from high school in 1965.

Shortly after high school he married Barbara Ann Fisher. It was through Barbara's coaxing that Darryl applied to Whitewater State University where he obtained a Bachelor of Science degree with a major in mathematics, in January of 1970. While attending Whitewater State Universtiy Darryl met Dr. David Stoneman who introduced him to the field of statistics. Dr. Stoneman was also instrumental in helping Darryl go to graduate school.

After graduating from Whitewater State University Darryl attended graduate school at Michigan Technological University, majoring in mathematics. He attended Michigan for six months and left for the University of Florida in the Fall of 1970. In June, 1972 Darryl received the Master of Statistics degree. From 1972 until the present he has been working towards the degree of Doctor of Philosophy with a major in Statistics.

Darryl and Barbara have two children: Darren Jon, age 8 and Kelly Ann, age 6. Both children were born in Janesville, Wisconsin while Darryl was attending Whitewater State University.

Darryl has been hired as an Assistant Professor of Statistics at Marquette University's Mathematics and Statistics Department in Milwaukee, Wisconsin and will start teaching there in August, 1974.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

J. G. Saw, Chairman
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the Degree of Doctor of Philosophy.

D. T. Hughes
Assistant Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the Degree of Doctor of Philosophy.

M. C. K. Yang
Assistant Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the Degree of Doctor of Philosophy.

Z. R. Pop-Stojanovic
Professor of Mathematics

This dissertation was submitted to the Department of Statistics in the College of Arts and Sciences and the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August, 1974

Dean, Graduate School