7·10·75

ANALYSIS OF PARAMETRIC STRUCTURES FOR VARIANCE MATRICES

BY

ANTHONY J. SWAIN

B.Sc. (Hons.) (Qld.), Dip. Comp.Sc.

Thesis submitted for the Degree of

Doctor of Philosophy

in the University of Adelaide,

Department of Statistics,

January, 1975.

CONTENTS

SUMMARY

Parametric models for variance matrices are of wide practical and theoretical interest. In this thesis we consider aspects of the statistical analysis for such variance structures.

In chapter 1 we provide some background to the problems considered, and a more complete description of this work. The rest of the thesis falls naturally into two parts.

In the first part (chapters 2, 3 and 4) we are concerned with general theory for parameter estimation and tests for the structure, with inference based on the sample variance matrix. In chapter 2 we study a general class, F of analysis procedures. This class includes as special cases maximum Wishart likelihood, some generalised least squares procedures, and other methods which arise naturally for variance structure analysis. Large sample properties, common to all members of F, are established for general variance structures. The case where the parameters need to be identified by imposed constraints is included. Estimates obtained by some of the procedures for parametric models for the eigen values of a variance matrix are derived and compared.

Further illustrations of and comparisons between procedures of F are given in chapters 3 and 4. The first problem considered in chapter 3 is the basic one of a completely specified variance matrix. Assuming the sum of squares matrix has a Wishart distribution, we consider theoretically and numerically the null

distribution of the test statistics provided by members of F. Improvement of the basic χ^2 approximation using Box type factors, and by computation of the next term of the asymptotic distribution is examined. It is found that the approximations are more accurate for maximum likelihood than for the other procedures considered. Extensions of the factor approach to the tests for more general structures are then suggested and examined for some simple structures. In chapter 4, we compare theoretically and by numerical example the estimates obtained for the unrestricted factor model.

In the second part (chapter 5) we consider the application of a particular variance structure, the direct product model, which, although it does not seem to have been examined in the literature, does appear useful for many situations. A rapidly convergent iterative procedure for determining the maximum likelihood parameter estimates is proposed and some of its properties examined. Expressions are obtained for the elements of the asymptotic variance matrix of the estimates. Three sets of data are used to demonstrate the usefullness of the structure.

We include, as appendices, computer programs which implement procedures described in chapters 2, 4 and 5.

SIGNED STATEMENT

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University. To the best of my knowledge and belief, the thesis contains no material previously published or written by any other person, except where due reference is made in the text of the thesis.

(A.J. Swain)

ACKNOWLEDGEMENTS

I am indebted to my supervisor, Dr. W.N. Venables, for suggesting the topics of research, and for his valuable advice and encouragement throughout the whole period of preparation of this work. The contents of chapters 2 and 4 have also benefited considerably from the constructive comments of the reviewers of a paper submitted by the author to Psychometrika.

I am pleased to acknowledge the financial support of a Commonwealth Postgraduate Research Award and of the Queensland Department of Primary Industries. My sincere thanks go to Mrs. D. Darwent and Mrs. K. Halsey for their proficient typing of the manuscript.

Chapter 1: Introduction

1.1 Background

By a parametric structure for a variance matrix we mean a specification of the variables and covariances of the elements of a random vector, $\underset{\sim}{y}$, as functions of a vector, $\underset{\sim}{\gamma}$, of parameters $\gamma_1 \ldots \gamma_q$. Denoting the variance matrix of $\underset{\sim}{y}$ by Σ, we write such variance structures as $\Sigma = \Sigma(\underset{\sim}{\gamma})$.

In many situations the parametric form arises naturally from the nature of the variates, and $\underset{\sim}{\gamma}$ provides a relatively simple and often readily interpretable description of the variance matrix of $\underset{\sim}{y}$. Structures may also be defined simply in order to reduce the degree of parametrisation and so to increase the precision of subsequent data analysis [Dempster (1972)]. Numerous examples occur, ranging from simple forms such as independence of the variates, or even complete specification of Σ, to much more complex forms such as the widely-used factor analysis model. Other examples are mentioned later.

Basic statistical analysis for variance structures is concerned with the estimation of the parameters and lack of fit of the structure. In this case, inference from observed data is naturally based on the sample variance matrix, S. Other related questions (which we will not be concerned with) may also arise, e.g. tests for equality of the parameters in different populations, and simultaneous analysis of both variance and mean structures.

The usual inferential approach has been to assume that $\underset{\sim}{y}$ follows the multinormal distribution, or almost equivalently, that the sum of squares matrix, nS, follows the Wishart distribution, and to apply maximum likelihood procedures. Early contributors to the statistical theory for particular structures included Lawley (1940, 1941) for factor analysis, Mauchly (1940) sphericity, Wilks (1946) complete symmetry, Votaw (1948) compound symmetry; and there are many more.

While investigation of the theory for particular structures using special features of the model continues (e.g. Dempster (1972) zeros in specified positions of Σ^{-1}, Olkin (1973) circular symmetry in blocks), general models which incorporate many of the specific forms which arise in practice have been recently formulated in order to unify both theory and computational procedures. Thus linear models have been considered by Bock and Bargmann (1966), Anderson (1969, 1970, 1973) and Mukherjee (1970, 1973), a general type of nonlinear model has been proposed and extensively applied by Jöreskog (1970a, 1972) and a completely general form has been considered by Browne (1974).

The commonly applied maximum likelihood procedures have well-known optimal large-sample properties, and the small-sample theory is well developed in many cases where explicit formulae for the maximum likelihood estimates are available [see e.g. Anderson 1958, chapters 9 and 10]. In other cases however, iterative methods are required to determine the estimates, making computation awkward and improvement over the basic asymptotic distribution theory difficult. Thus interest has turned recently to generalised least squares methods of analysis [Jöreskog and Goldberger (1972), Anderson (1973), Browne (1974)]. These procedures have the same large sample properties as maximum likelihood, and at least for linear models explicit formulae for the estimates are obtained, even though this is not the case in general for the maximum likelihood estimates.

In many respects there is a close similarity between the analysis of variance structures and the problems of inference on structural parameters for the multinomial distribution. For this case a general class of procedures, known as minimum chi-square methods, which are equivalent in large samples to maximum likelihood, were first proposed by Neyman (1949). They have since been considered by many authors, e.g. Taylor (1953), Ferguson (1958), Rao (1955, 1963).

This thesis is intended as a contribution to the theory for analysis of variance structures. The research falls naturally into two parts.

In the first part (chapters 2, 3 and 4), a general class of estimation procedures is studied. This class includes as special cases maximum likelihood, some generalised least squares procedures and other procedures proposed either explicitly or implicitly in the literature for the analysis of variance structures. It is shown that for general structures the large-sample properties of the estimates corresponding to any member of the class are the same for all members of the class. The performances of some of the procedures are then compared theoretically and numerically for some commonly applied structures.

In the second part (chapter 5), a particular variance structure, the direct product structure, is considered. This model, which has not previously been studied in the literature, proves useful in some applications.

1.2 The problems considered

In chapter 2 we define the general form of variance structure to be considered and present a general class of estimation procedures, which includes maximum likelihood when the sample variance matrix has the Wishart distribution, some generalised least squares methods and other procedures which also arise naturally for variance structure estimation. The common large-sample properties of the estimates are established for the case when the parameters are identified, using methods similar to those adopted by Rao (1973) for maximum likelihood estimation of structural parameters for the multinomial distribution, and are extended to the nonidentified case. Numerical determination of the estimates is also considered. To illustrate similarities and differences in the procedures, a theoretical comparison is made between the estimates obtained for particular types of model which constrain the eigen values of Σ but not its eigen vectors.

Associated with each of the procedures is a goodness of fit statistic which is asymptotically distributed as χ^2. The worth of this approximation for small samples is in general unknown, and as an accurate knowledge of the distribution of the test statistic could be a good reason for preferring a particular procedure, methods for evaluating and improving the approximation are considered in chapter 3. Since, at least, a poor approximation in this case is likely to indicate poor performance more generally, particular attention is given to the test for specified Σ when the sample variance matrix has the

Wishart distribution, with the χ^2 approximation to the distribution of the test statistic being improved by both multiplying factors [Box (1949)] and by computation of the order n^{-1} term through an extension of the method of Nagao (1973). An indication of performance is given by a numerical comparison. Possible extensions of the factor approach to structures for which the expectation of the test statistic is unavailable are also suggested.

The effect of the use of different procedures on the parameter estimates for the widely-used unrestricted factor analysis model is considered both theoretically and by numerical example in chapter 4.

In chapter 5 the direct product structure for a variance matrix is introduced as a model which may be appropriate where the elements of $\underset{\sim}{y}$ are naturally arranged in matrix form. A rapidly convergent iterative procedure for determination of the maximum likelihood estimates of the parameters is proposed and some of its properties examined. The asymptotic distribution of the estimates is obtained, and methods proposed in chapter 3 for the improvement of the χ^2 approximation to the distribution of the goodness of fit statistic are examined numerically. Three sets of actual data are used to illustrate the usefulness of the structure.

Computer programs which implement procedures described in chapters 2, 4 and 5 are included as appendices.

Chapter 2: A general approach to the analysis of variance structures

2.1 Introduction

We are concerned with the situation where N independent observations are made on a $p\times1$ random vector, $\underset{\sim}{y}$, which is distributed with mean $\underset{\sim}{\mu}$ and variance matrix Σ, and it is assumed that Σ has some prescribed parametric form. By far the most common approach to the statistical problems of parameter estimation and testing for the structure is to assume that $\underset{\sim}{y}$ has the multinormal distribution, i.e. $\underset{\sim}{y}\sim N(\underset{\sim}{\mu},\Sigma)$, and to apply maximum likelihood (ML) methods. An essentially equivalent approach when no assumptions are made on $\underset{\sim}{\mu}$, is to assume that, with S the sample variance matrix and $n=N-1$, nS has the Wishart distribution with n degrees of freedom and variance matrix Σ, i.e. $nS\sim W_p(n,\Sigma)$, and to base the ML procedures on S.

Under certain restrictions on the structure the method has a number of well-known large sample properties. In particular the parameter estimates are consistent and asymptotically normally distributed with large sample variance matrix being given by the inverse of the information matrix. Further, a convenient test of the assumed form against a suitable alternative is provided by $-2\log\lambda$, with λ the likelihood ratio, which under the null hypothesis is asymptotically distributed as χ^2.

When $nS\sim W_p(n,\Sigma)$, maximisation of the likelihood of Σ given S is easily shown to be equivalent to the minimisation of

$$F(\Sigma;S) = \log|\Sigma| - \log|S| + \mathrm{tr}(S\Sigma^{-1}) - p \qquad (1.1)$$

which at the minimum is n^{-1} times $-2 \log \lambda$, for the alternative Σ unstructured. The function (1.1) has the property $F(\Sigma;S) \geq 0$, with equality only if $\Sigma = S$, and may be regarded as a measure of distance of Σ from S, with the ML estimates being determined so that the distance is minimised.

As an alternative to ML for the particular case of the unrestricted factor model, Jöreskog and Goldberger (1972) considered the use of a generalised least squares method in which estimates are obtained by the minimisation of

$$F(\Sigma;S) = \tfrac{1}{2}\,\mathrm{tr}[S^{-1}(\Sigma-S)]^2. \qquad (1.2)$$

They showed that if S has

$$E[s_{ij}] = \sigma_{ij};\ \mathrm{Cov}(s_{ij}s_{k\ell}) = n^{-1}(\sigma_{ik}\sigma_{j\ell}+\sigma_{i\ell}\sigma_{jk}) \qquad (1.3)$$

and its elements asymptotically normally distributed, which is true in particular when $nS \sim W_p(n,\Sigma)$, then the estimates have the same large sample properties as those obtained by ML. This approach has been extended by Browne (1974) who has shown that the same is true for general variance structures for estimates determined by the minimisation of

$$F(\Sigma;S) = \tfrac{1}{2}\,\mathrm{tr}[V(\Sigma-S)]^2 \qquad (1.4)$$

provided V is a consistent estimator of Σ^{-1}. Browne also remarks that (1.3) is true provided that the fourth order cumulants of the distribution of $\underset{\sim}{y}$ are zero.

After giving in section 2.2 some basic notation and results, we define in section 2.3 the general model to be considered, and present a class of "distance" functions which includes (1.1) and (1.2), and some other $F(\Sigma;S)$ which arise naturally for estimation, but not all those of the form (1.4). In section 2.4 we consider the large sample properties of estimates obtained by minimisation of any $F(\Sigma;S)$, with these properties proving to be the same for all $F(\Sigma;S)$. A method of determining the estimates is then considered in section 2.5. In section 2.6 we extend the theory of section 2.4 to the case where the parameters are not identified. Finally, as an example of the effects which the choice of F can have on the estimates, we consider in section 2.7 estimation for models which specify structure to the eigen values of Σ but impose no constraint on its eigen vectors.

2.2 Basic notation

If A is an $m \times n$ matrix we denote by $\underset{\sim}{A}$, or vec(A), the $mn \times 1$ vector formed by writing the columns of A one under the other. Thus

$$\underset{\sim}{A}' = (a_{11}a_{21}\ldots a_{m1}a_{12}a_{22}\ldots a_{mn}). \tag{2.1}$$

The elements of this vector will usually be denoted by their double subscripts. Similarly double subscript notation will sometimes be used for rows or columns of matrices, in which case, as above, the first subscript is always nested within the second.

Similarly if S is a $p \times p$ symmetric matrix, we denote by $\underset{\sim}{s}$, or ve(S), the $\frac{1}{2}p(p+1) \times 1$ vector formed in the same way as $\underset{\sim}{S}$ but using only the lower triangle of S. In this case we may write

$$\underset{\sim}{S} = D_p \underset{\sim}{s} \tag{2.2}$$

where D_p is the $p^2 \times \frac{1}{2}p(p+1)$ matrix with

$$(D_p)_{ij\ k\ell} = \begin{cases} 1 & \text{if } i=k \text{ and } j=\ell \text{ or } i=\ell \text{ and } j=k \\ 0 & \text{otherwise,} \end{cases} \tag{2.3}$$

where $i,j,k,\ell = 1\ldots p$, $k \geq \ell$. Further, if D_p^- is any generalised inverse of D_p, then $\underset{\sim}{s} = D_p^- \underset{\sim}{S}$ [see e.g. Rao (1973, p.24)]. Of particular interest is the Moore-Penrose g-inverse D_p^+, which, since D_p has rank $\frac{1}{2}p(p+1)$, is given by

$$D_p^+ = (D_p' D_p)^{-1} D_p'. \tag{2.4}$$

In subscript notation, for $i \geq j$

$$(D_p^+)_{ij\ k\ell} = \tfrac{1}{2}(\delta_{ik}\delta_{j\ell} + \delta_{i\ell}\delta_{jk}) \tag{2.5}$$

If A and B are m×n and p×q respectively, the direct or Kronecker product $A \otimes B$ of A and B is the mp×nq matrix with

$$(A \otimes B)_{ij\ k\ell} = a_{j\ell} b_{ik}, \tag{2.6}$$

[see e.g. Anderson, 1958, p.347]. We also consider a new product, $A \boxtimes B$, of A and B, defined by

$$(A \boxtimes B)_{ij\ k\ell} = a_{i\ell} b_{jk} \tag{2.7}$$

The relationship between $A \otimes B$ and $A \boxtimes B$ may be expressed as

$$(A \boxtimes B) = L_{mp}(A \otimes B) = (B \otimes A)L_{nq} \tag{2.8}$$

where $L_{mp} = I_m \boxtimes I_p$. This matrix and some of the properties associated with it have recently been considered by MacRae (1974). We note that L_{mp} is a permutation matrix, and so orthogonal, with $L'_{mp} = L_{pm}$. Further, since from (2.2) the effect of premultiplying a matrix by D_p is, for $i \geq j$, to create a matrix with its ij and ji rows equal to the ij row of the original matrix, from (2.5)

$$D_p D_p^+ = \tfrac{1}{2}(I_{p^2} + L_{pp}). \tag{2.9}$$

It is then easily shown that, from (2.4), (2.8) and (2.9), for A and B each pxp

$$D'_p(A \otimes B)D_p = D'_p(B \otimes A)D_p = D'_p(A \boxtimes B)D_p = D'_p(B \boxtimes A)D_p, \tag{2.10}$$

and if A is nonsingular,

$$[D'_p(A^{-1} \otimes A^{-1})D_p]^{-1} = \tfrac{1}{2}D_p^+(A \otimes A + A \boxtimes A)(D_p^+)' \tag{2.11}$$

Many further properties of the products $\otimes$ and $\boxtimes$ and of the matrices L_{mp} are set out in appendix A.

2.3 The general variance structure model and the class F of estimation functions.

We consider* a general model of variance structure,

$$\Sigma = \Sigma(\underset{\sim}{\gamma}),\ \underset{\sim}{\gamma}\in\Gamma,$$

where Σ is a $p\times p$ variance matrix and the parameter space Γ is contained in R^q, the space of all real $q\times 1$ vectors, and write

$$A = \{\Sigma | \Sigma = \Sigma(\underset{\sim}{\gamma}), \underset{\sim}{\gamma}\in\Gamma\} = \Sigma(\Gamma).$$

We will assume that for the population being sampled $\underset{\sim}{\gamma} = \underset{\sim}{\gamma}_o$, with $\underset{\sim}{\gamma}_o$ an interior point of Γ, i.e. with $\|\underset{\sim}{\gamma}\|$ defined by

$$\|\underset{\sim}{\gamma}\| = \max_{i=1,q} |\gamma_i|$$

for $\underset{\sim}{\gamma}\in R^q$, there exists $\delta>0$ defining a neighbourhood (nbd.), N, of $\underset{\sim}{\gamma}_o$ in R^q by

$$N = \{\underset{\sim}{\gamma}|\ \|\underset{\sim}{\gamma} - \underset{\sim}{\gamma}_o\| < \delta\}$$

such that $N\subset\Gamma$. We will also assume the following regularity conditions on the model and $\underset{\sim}{\gamma}_o$.

(a) $\Sigma_o = \Sigma(\underset{\sim}{\gamma}_o)$ is positive definite.

(b) all $\sigma_{ij}(\underset{\sim}{\gamma})$ and all partial derivatives of the first two orders with respect to the elements of $\underset{\sim}{\gamma}$ are continuous in a nbd. of $\underset{\sim}{\gamma}_o$.

(c) the $\frac{1}{2}p(p+1)\times q$ matrix

$$\Delta(\underset{\sim}{\gamma}) = \frac{\partial\underset{\sim}{\sigma}(\underset{\sim}{\gamma})}{\partial\underset{\sim}{\gamma}'} \tag{3.1}$$

---

* Much of the material in this section and section 2.4 has been presented in a less rigorous way in a paper by the author [Swain (1975)].

has rank q at $\underset{\sim}{\gamma}_o$.

(d) The parameters are identified in Γ at all points in some nbd. of $\underset{\sim}{\gamma}_o$, i.e. if $\underset{\sim}{\gamma}_1$ is some point of this nbd. and $\Sigma(\underset{\sim}{\gamma}_2) = \Sigma(\underset{\sim}{\gamma}_1)$ for some $\underset{\sim}{\gamma}_2 \in \Gamma$, then $\underset{\sim}{\gamma}_2 = \underset{\sim}{\gamma}_1$.

We consider the situation where a sample variance matrix S based on $n>p$ degrees of freedom is observed, and an estimate of $\underset{\sim}{\gamma}_o$ is formed by minimisation of some "distance" function.

$$F(\Sigma:S) = \sum_{i=1}^{p} f(\theta_i) \tag{3.2}$$

where $\theta_1,\ldots,\theta_p$ are the eigen values of $S^{-\frac{1}{2}}\Sigma S^{-\frac{1}{2}}$ and $f(\theta)$ is some function with a continuous second order derivative on $\theta>0$, satisfying the conditions

$$\text{(i)} \quad f(1) = f'(1) = 0,\ f''(1) = 1 \tag{3.3}$$

$$\text{(ii)} \quad f'(\theta) \begin{cases} < 0 & 0 < \theta < 1 \\ > 0 & 1 < \theta \end{cases} \tag{3.4}$$

We denote by $\mathcal{F}$ the class of all such $F(\Sigma;S)$, and shall sometimes refer to the elements of $\mathcal{F}$ as estimation functions.

The conditions (i) and (ii) ensure that $f(\theta)$ has a minimum at $\theta=1$, and that for arbitary positive definite Σ, $F(\Sigma;\Sigma_o) \geq 0$, with equality if and only if $\Sigma=\Sigma_o$. Thus, by assumption (d) on the model, the estimates, $\hat{\underset{\sim}{\gamma}}$, of the parameters $\underset{\sim}{\gamma}_o$ formed by minimisation of $F(\Sigma(\underset{\sim}{\gamma});S)$ will necessarily be Fisher consistent, i.e. if $S=\Sigma_o$ then $\hat{\underset{\sim}{\gamma}}=\underset{\sim}{\gamma}_o$.

We note that for any ~~arbitary~~ non-singular p×p matrix B, $|B\Sigma B'-\theta BSB'| = |B||\Sigma-\theta S||B'|$, $= 0$ if and only if

$|\Sigma-\theta S| = 0$, and so

$$F(\Sigma;S) = F(B\Sigma B'; BSB').$$

We also note that if $f(\theta)$ satisfies the conditions (3.3) and (3.4) then so too does $f(\frac{1}{\theta})$. Thus, since for non-singular Σ the eigen values of $\Sigma^{-\frac{1}{2}}S\Sigma^{-\frac{1}{2}}$ are θ_i^{-1}, $i = 1\ldots p$, if $F(\Sigma;S)\in\mathcal{F}$ then so too is $F^*(\Sigma;S) = F(S;\Sigma)$, which is in general different from $F(\Sigma;S)$.

In the next section we will consider the asymptotic properties of the estimates of $\underset{\sim}{\gamma}_o$. Before doing so we present some of the elements of $\mathcal{F}$ which are of special interest. These are shown in table 1, and arise from the following approaches to estimation.

TABLE 1

Some special estimation functions $F(\Sigma;S)$

CODE	DEFINITION	$f'''(1)$
TGLS	$F_1 = \frac{1}{2}\sum_{i=1}^{p}\left(\frac{1}{\theta_i}-1\right)^2 = \frac{1}{2}\,tr[\Sigma^{-1}(\Sigma-S)]^2$	-6
ML	$F_2 = \sum_{i=1}^{p}\left(\frac{1}{\theta_i}+\log\theta_i-1\right) = \log\lvert\Sigma\rvert-\log\lvert S\rvert + tr(S\Sigma^{-1})-p$	-4
GD	$F_3 = \frac{1}{2}\sum_{i=1}^{p}(\log\theta_i)^2$	-3
DIV	$F_4 = \frac{1}{2}\sum_{i=1}^{p}\left(\frac{1}{\theta_i}+\theta_i-2\right) = \frac{1}{2}\left(tr(S\Sigma^{-1})+tr(S^{-1}\Sigma)-2p\right)$	-3
GLS	$F_5 = \frac{1}{2}\sum_{i=1}^{p}(\theta_i-1)^2 = \frac{1}{2}\,tr[S^{-1}(\Sigma-S)]^2$	0
GLSE	$F_6 = \frac{1}{2}\sum_{i=1}^{p}(\theta_i-1)^2e^{(\theta_i-1)}$	3

TGLS: True generalised least squares. Estimation of $\underset{\sim}{\gamma}$ by the application of the generalised least squares principle to the observed S is equivalent to determining $\underset{\sim}{\gamma}$ to minimise

$$(\underset{\sim}{s}-\underset{\sim}{\sigma}(\underset{\sim}{\gamma}))'[V(\underset{\sim}{s})]^{-1}(\underset{\sim}{s}-\underset{\sim}{\sigma}(\underset{\sim}{\gamma})) \tag{3.5}$$

where $V(\underset{\sim}{s})$ is the variance matrix of $\underset{\sim}{s}$. From (1.3) and (2.11)

$$[V(\underset{\sim}{s})]^{-1} = \frac{n}{2} D_p'(\Sigma_o^{-1} \otimes \Sigma_o^{-1})D_p \tag{3.6}$$

and so (3.5) becomes

$$\begin{aligned} &\frac{n}{2}(\underset{\sim}{s}-\underset{\sim}{\sigma}(\underset{\sim}{\gamma}))'D_p'(\Sigma_o^{-1} \otimes \Sigma_o^{-1})D_p(\underset{\sim}{s}-\underset{\sim}{\sigma}(\underset{\sim}{\gamma})) \\ &= \frac{n}{2}(\underset{\sim}{S}-\underset{\sim}{\Sigma}(\underset{\sim}{\gamma}))'(\Sigma_o^{-1} \otimes \Sigma_o^{-1})(\underset{\sim}{S}-\underset{\sim}{\Sigma}(\underset{\sim}{\gamma})) \\ &= \frac{n}{2}\operatorname{tr}[\Sigma_o^{-1}(S-\Sigma)]^2, \end{aligned} \tag{3.7}$$

from (xii) of Appendix A. When Σ_o is replaced by its estimate we obtain $F_1(\Sigma;S)$.

ML: Maximum likelihood. When $nS \sim W_p(n,\Sigma_o)$, then maximisation of the likelihood of Σ is equivalent to the minimisation of $F_2(\Sigma;S)$.

GD: Geodesic distance. James (1973) has considered the metric differential form

$$(ds)^2 = \tfrac{1}{2}\operatorname{tr}(\Sigma^{-1}d\Sigma\ \Sigma^{-1}d\Sigma)$$

on the variance information manifold and shown that $F_3(\Sigma;S)$ is the square of the corresponding geodesic distance between S and Σ. He has also suggested its possible use in variance structure estimation. We note that $F_3(\Sigma;S) = F_3(S;\Sigma)$, and that $(ds)^2 = F_3(\Sigma+d\Sigma;\Sigma)$.

DIV: Divergence. In the context of information theory, the divergence between two density functions $f_1(\underset{\sim}{x};\underset{\sim}{\theta}_1)$ and $f_2(\underset{\sim}{x};\underset{\sim}{\theta}_2)$, of the same form and with parameters $\underset{\sim}{\theta}_1$ and $\underset{\sim}{\theta}_2$ respectively, is defined by

$$J(1;2) = I(1;2) + I(2;1) \tag{3.8}$$

where

$$I(i;j) = \int \log \frac{f_i}{f_j} f_i \, d\underset{\sim}{x} \tag{3.9}$$

is the mean information in $\underset{\sim}{x}$ for discrimination in favour of $\underset{\sim}{\theta}_1$ against $\underset{\sim}{\theta}_2$ (Kullback (1959)). $I(2;1)$ is essentially the Kullback-Leibler measure of distance from $\underset{\sim}{\theta}_1$ to $\underset{\sim}{\theta}_2$. When the theory is applied to the Wishart distribution

$$I(2;1) = \frac{n}{2} F_2(\Sigma_2;\Sigma_1); \; I(1;2) = \frac{n}{2} F_2(\Sigma_1;\Sigma_2) \tag{3.10}$$

$$J(1,2) = nF_4(\Sigma_1;\Sigma_2) = nF_4(\Sigma_2;\Sigma_1). \tag{3.11}$$

Although proposed as a measure of divergence between the populations, here $J(1,2)$ can be interpreted as a measure of the difference between Σ_1 and Σ_2, and $F_4(\Sigma;S)$ as a convenient measure of the departure of Σ from the observed S.

GLS: Generalised least squares. This has been proposed for estimation in factor analysis by Jöreskog and Goldberger (1972) who derived it from (3.7) by replacing Σ_o^{-1} by S^{-1}. We note that $F_5(\Sigma;S) = F_1(S;\Sigma)$.

GLSE: Generalised least squares extended. This was introduced by the author [Swain (1975)] merely in order to extend the range of functions considered when applying the general theory to the factor analysis model.

2.4 Asymptotic properties of minimum distance estimates of the parameters.

In this section we investigate the large sample properties of the estimates of γ_0 obtained by minimisation of some $F(\Sigma;S)\in F$, when the model $\Sigma=\Sigma(\gamma)$ satisfies the conditions (a) to (d) of section 2.3. The approach adopted is similar to that used by Rao (1973, p.355) to obtain corresponding results for the maximum likelihood estimates of multinomial distribution parameters.

In establishing the results we shall use the following theory on probability order relationships.

Definition If $\{r_n\}$ is a sequence of positive real numbers and $\{X_n\}$ a sequence of $p\times 1$ random vectors, we say

$$X_n = o_p(r_n) \text{ if } \|X_n\|/r_n \xrightarrow{p} 0 \quad \text{as} \quad n\to\infty$$

[i.e. if $\|X_n\|/r_n \to 0$ in probability] and

$X_n = O_p(r_n)$ if, given $\epsilon>0$ there exists $M<\infty$ such that $\Pr(\|X_n\|/r_n<M) > 1-\epsilon$ for all sufficiently large n [i.e. if $\|X_n\|/r_n$ is bounded in probability].

Lemma 1 If $f(X)$ is a function with continuous derivatives of order s in a neighbourhood of $X=a$, if $\{r_n\}$ is a sequence of positive numbers such that $r_n \to 0$ as $n\to\infty$, and if $\{X_n\}$ is a sequence of random variables with $X_n = a + O_p(r_n)$, then

$$f(X_n) = f(a) + \sum_{k=1}^{s} f^{(k)}(a)\,\frac{(X_n-a)^k}{k!} + o_p(r_n{}^s).$$

Proof: See Mann and Wald (1943) who present it as an application to one form of Taylors theorem of a general result which states, in effect, that the algebra of the usual order notation o and O extends to o_p and O_p [Chernoff (1956)]. For convenience lemma 1 has been stated in univariate form. We will, however, use the corresponding multivariate result, which is presented by Mann and Wald.

<u>Corollary</u> Under the conditions of lemma 1,

$$f(X_n) = f(a) + \sum_{k=1}^{s-1} f^{(k)}(a) \frac{(X_n - a)^k}{k!} + O_p(r_n{}^s).$$

In particular, when s=1, i.e. f(X) has continuous first order derivatives in a neighbourhood of X = a,

$f(X_n) = f(a) + O_p(r_n)$.

Mann and Wald have also given the following results which are well known [see e.g. Rao (1973, p.122)].

<u>Lemma 2</u> If $f(\underset{\sim}{X})$ is a continuous function in a neighbourhood of $\underset{\sim}{X}=\underset{\sim}{a}$ and $\{\underset{\sim}{X}_n\}$ is a sequence of random vectors such that $\underset{\sim}{X}_n \overset{p}{\to} \underset{\sim}{a}$, then $f(\underset{\sim}{X}_n) \overset{p}{\to} f(\underset{\sim}{a})$.

<u>Lemma 3</u> If $\{\underset{\sim}{X}_n\}$ and $\{\underset{\sim}{Y}_n\}$ are two sequences of random vectors such that $\underset{\sim}{X}_n - \underset{\sim}{Y}_n \overset{p}{\to} 0$ and $\underset{\sim}{X}_n \overset{L}{\to} \underset{\sim}{X}$ (i.e. $\underset{\sim}{X}_n \to \underset{\sim}{X}$ in distribution) as $n\to\infty$, then $\underset{\sim}{Y}_n \overset{L}{\to} \underset{\sim}{X}$.

We will now assume that a sample variance matrix S is observed with n>p degrees of freedom and asymptotic distribution given by

$$\sqrt{n}\ (\underset{\sim}{S}-\underset{\sim}{\Sigma}_o) \ \dot{\sim}\ N(\underset{\sim}{o},V) \tag{4.1}$$

with V defined by (1.3) i.e. $V = \Sigma_o \otimes \Sigma_o + \Sigma_o \boxtimes \Sigma_o$, and that the model $\Sigma(\underset{\sim}{\gamma})$ and the true value $\underset{\sim}{\gamma}_o$ satisfy the conditions (a) to (d) of section 2.3. Let $F(\Sigma;S)$ be any element of F. Then we have the following.

Definition We say that $\underset{\sim}{\gamma}^* \in \Gamma$ is an approximate minimum distance estimator of $\underset{\sim}{\gamma}_o$ if

$$F(\Sigma(\underset{\sim}{\gamma}^*);S) \leq c \inf_{\underset{\sim}{\gamma} \in \Gamma} F(\Sigma(\underset{\sim}{\gamma});S) \tag{4.2}$$

where c is a constant with $1<c<\infty$. If the inf is achieved at some point $\hat{\underset{\sim}{\gamma}} \in \Gamma$, then we say $\hat{\underset{\sim}{\gamma}}$ is a minimum distance estimator (m.d.e.). We note that while an approximate m.d.e. must always exist for any $1<c<\infty$, an m.d.e. need not.

Theorem 1 If $\underset{\sim}{\gamma}^*$ is any approximate m.d.e. of $\underset{\sim}{\gamma}_o$, then $\underset{\sim}{\gamma}^*$ is consistent for $\underset{\sim}{\gamma}_o$ with $\underset{\sim}{\gamma}^* - \underset{\sim}{\gamma}_o = O_p(n^{-\frac{1}{2}})$.

Proof We write $\Sigma^* = \Sigma(\underset{\sim}{\gamma}^*)$ and let $\theta_i^{(o)}$ and θ_i^*, $i = 1 \dots p$, be the ordered eigen values of $S^{-\frac{1}{2}}\Sigma_o S^{-\frac{1}{2}}$ and $S^{-\frac{1}{2}}\Sigma^* S^{-\frac{1}{2}}$ respectively. Now since $S-\Sigma_o = O_p(n^{-\frac{1}{2}})$, we have

$$\Sigma_o^{-\frac{1}{2}} S \Sigma_o^{-\frac{1}{2}} - I = O_p(n^{-\frac{1}{2}}) \Rightarrow tr(\Sigma_o^{-\frac{1}{2}} S \Sigma_o^{-\frac{1}{2}} - I)^2 = O_p(n^{-1})$$

$$\Rightarrow \sum_{i=1}^{p} (\theta_i^{(o)}-1)^2 = O_p(n^{-1}) \Rightarrow (\theta_i^{(o)}-1)^2 = O_p(n^{-1})$$

$$\Rightarrow (\theta_i^{(o)}-1) = O_p(n^{-\frac{1}{2}}),$$

$i = 1 \dots p$. Further, since $f(\theta_i)$ has continuous second order derivatives, by the corollary to lemma 1, $f(\theta_i^{(o)}) = f(1) + f'(1)(\theta_i^{(o)}-1) + O_p(n^{-1}) = O_p(n^{-1})$ by condition (3.3). Thus, since $F(\Sigma_o;S) = \sum_i f(\theta_i^{(o)})$, $F(\Sigma_o;S) = O_p(n^{-1})$.

By the definition of $\underset{\sim}{\gamma}^*$,

$$F(\Sigma^*;S) \leq c \inf_{\underset{\sim}{\gamma} \in \Gamma} F(\Sigma(\underset{\sim}{\gamma});S) \leq c\, F(\Sigma_o;S) \tag{4.3}$$

so $F(\Sigma^*;S) = O_p(n^{-1})$, and since $F(\Sigma^*;S) \geq f(\theta_i^*)$, $f(\theta_i^*) = O_p(n^{-1})$. Accordingly $\theta_i^* \xrightarrow{p} 1$, and since

$$f(\theta_i^*) = \frac{(\theta_i^*-1)^2}{2} + \text{higher order terms} \tag{4.4}$$

and the order of the lowest order terms on both left and right hand sides must agree, $\theta_i^* -1 = O_p(n^{-\frac{1}{2}})$. It follows by effectively reversing the argument above that

$$(\Sigma^*-S) = O_p(n^{-\frac{1}{2}}). \tag{4.5}$$

Thus

$$(\Sigma^*-\Sigma_o) = (\Sigma^*-S) + (S-\Sigma_o) = O_p(n^{-\frac{1}{2}}).$$

In particular then $\Sigma^* \xrightarrow{p} \Sigma_o$, and so, since conditions (c) and (d) on the model together with the continuity of the derivatives of $\Sigma(\underset{\sim}{\gamma})$ imply strong identifiability in a nbd. of $\underset{\sim}{\gamma}_o$ (see section 2.6, result (iii)) $\underset{\sim}{\gamma}^* \xrightarrow{p} \underset{\sim}{\gamma}_o$ i.e. $\underset{\sim}{\gamma}^*$ is consistent for $\underset{\sim}{\gamma}_o$.

Now since $\Sigma(\underset{\sim}{\gamma})$ has continuous derivatives in a nbd. of $\underset{\sim}{\gamma}_o$, we have

$$\underset{\sim}{\sigma}(\underset{\sim}{\gamma}^*) - \underset{\sim}{\sigma}(\underset{\sim}{\gamma}_o) = \Delta(\underset{\sim}{\gamma}_o)(\underset{\sim}{\gamma}^*-\underset{\sim}{\gamma}_o) + \text{higher order terms} \tag{4.6}$$

and again since the order of the lowest order terms on both left and right hand sides must agree, and since $\Delta(\underset{\sim}{\gamma}_o)$ has rank q, it follows that $\underset{\sim}{\gamma}^*-\underset{\sim}{\gamma}_o = O_p(n^{-\frac{1}{2}})$.

<u>Theorem 2</u> The probability that an m.d.e. exists $\to 1$ as $n \to \infty$.

<u>Proof</u> By the assumptions, there exists $\delta>0$ such that the nbd. Γ_1 of $\underset{\sim}{\gamma}_o$ defined by $\Gamma_1 = \{\underset{\sim}{\gamma} \mid \|\underset{\sim}{\gamma}-\underset{\sim}{\gamma}_o\| \leq \delta\}$ is in the interior of Γ and conditions (b) and (d) hold on Γ_1. Since this is a closed and bounded region in R^q

and $F(\Sigma(\underset{\sim}{\gamma});S)$ is a continuous function of $\underset{\sim}{\gamma}$ on it, $\inf_{\underset{\sim}{\gamma}\in\Gamma_1} F(\Sigma(\underset{\sim}{\gamma});S)$ must be attained for some $\underset{\sim}{\gamma}\in\Gamma_1$. To prove the theorem it is thus sufficient to show that, as $n\to\infty$,

$$\Pr\left\{\inf_{\underset{\sim}{\gamma}\in\Gamma} F(\Sigma(\underset{\sim}{\gamma});S) = \inf_{\underset{\sim}{\gamma}\in\Gamma_1} F(\Sigma(\underset{\sim}{\gamma});S)\right\} \to 1 \qquad (4.7)$$

We prove this result by assuming that it does not hold, and then showing that this leads to a contradiction of theorem 1. Thus, suppose (4.7) is false. Then there exists $\varepsilon>0$ and at least a subsequence of $\{n\}$, say $\{n_\nu\}$ with the property that

$$\Pr\left\{\inf_{\underset{\sim}{\gamma}\in\Gamma} F(\Sigma(\underset{\sim}{\gamma});S) < \inf_{\underset{\sim}{\gamma}\in\Gamma_1} F(\Sigma(\underset{\sim}{\gamma});S)\right\} > \varepsilon > 0 \qquad (4.8)$$

for all n_ν. Now if we define an approximate m.d.e. in such a way that whenever $\inf_{\underset{\sim}{\gamma}\in\Gamma} F(\Sigma(\underset{\sim}{\gamma});S) < \inf_{\underset{\sim}{\gamma}\in\Gamma_1} F(\Sigma(\underset{\sim}{\gamma});S)$ $\underset{\sim}{\gamma}^*\in\Gamma_1^c$ (the complement of Γ_1 in Γ), then for each n_ν, $\Pr(\underset{\sim}{\gamma}^*\in\Gamma_1{}^c) > \varepsilon$, i.e. $\Pr(\underset{\sim}{\gamma}^*\in\Gamma_1) < 1-\varepsilon$, and so does not approach 1 as $n_\nu\to\infty$, i.e. $\underset{\sim}{\gamma}^*$ is not consistent for $\underset{\sim}{\gamma}_0$, contradicting theorem 1.

<u>Corollary</u> The probability that an m.d.e. exists and is given as a solution of

$$\frac{\partial F}{\partial \underset{\sim}{\gamma}}(\Sigma(\underset{\sim}{\gamma});S) = \underset{\sim}{0} \qquad (4.9)$$

$\to 1$ as $n\to\infty$.

<u>Proof</u> Choosing Γ_1 as in theorem 2, whenever $\inf_{\underset{\sim}{\gamma}\in\Gamma_1} F(\Sigma(\underset{\sim}{\gamma});S) = \inf_{\underset{\sim}{\gamma}\in\Gamma} F(\Sigma(\underset{\sim}{\gamma});S)$ the point at which the inf is achieved is an interior point of Γ at which the derivatives of $F(\Sigma(\underset{\sim}{\gamma});S)$ are continuous, and so must vanish.

Theorem 3 Let an estimator $\hat{\underset{\sim}{\gamma}}$ be defined as an m.d.e. whenever an m.d.e. exists, and otherwise as an approximate m.d.e. Then

$$\sqrt{n}(\hat{\underset{\sim}{\gamma}}-\underset{\sim}{\gamma}_o) \to N(\underset{\sim}{0}, 2(\Delta_o' E\Delta_o)^{-1}) \tag{4.10}$$

where $\Delta_o = \Delta(\underset{\sim}{\gamma}_o)$ and

$$E = D_p'(\Sigma_o^{-1} \otimes \Sigma_o^{-1})D_p \tag{4.11}$$

with D_p defined by (2.3).

Proof: With Γ_1 as defined in theorem 2, as $n\to\infty$ the probability that $\hat{\underset{\sim}{\gamma}}\in\Gamma_1\to 1$, and so, in order to determine the asymptotic distribution of $\hat{\underset{\sim}{\gamma}}$ we can assume that $\hat{\underset{\sim}{\gamma}}\in\Gamma_1$ and satisfies

$$\frac{\partial F}{\partial \underset{\sim}{\gamma}}(\Sigma(\hat{\underset{\sim}{\gamma}});S) = \underset{\sim}{0}. \tag{4.12}$$

In the course of the proof of theorem 1, we showed that, taking $\underset{\sim}{\gamma}^* = \hat{\underset{\sim}{\gamma}}$, $\hat{\theta}_i - 1 = O_p(n^{-\frac{1}{2}})$ where $\hat{\theta}_i$ are the ordered eigen values of $S^{-\frac{1}{2}}\hat{\Sigma}S^{-\frac{1}{2}}$. Then applying lemma 1 to $f'(\theta)$ we have

$$f'(\hat{\theta}_i) = (\hat{\theta}_i - 1) + o_p(n^{-\frac{1}{2}}). \tag{4.13}$$

Now, from (5.11) and (4.13) we find that

$$\left.\frac{\partial F}{\partial \underset{\sim}{\sigma}}\right|_{\hat{\Sigma}} = D_p' \operatorname{vec}(U'(\hat{\Theta}-I)U) + o_p(n^{-\frac{1}{2}}) \tag{4.14}$$

where, if $S^{-\frac{1}{2}}\hat{\Sigma}S^{-\frac{1}{2}} = \hat{H}\hat{\Theta}\hat{H}'$ with $\hat{\Theta} = \operatorname{diag}(\hat{\theta}_i)$ and $\hat{H}'\hat{H} = I$, $U = S^{-\frac{1}{2}}\hat{H}$. Thus

$$\begin{aligned}\left.\frac{\partial F}{\partial \underset{\sim}{\sigma}}\right|_{\hat{\Sigma}} &= D_p' \operatorname{vec}(S^{-1}(\hat{\Sigma}-S)S^{-1}) + o_p(n^{-\frac{1}{2}}) \\ &= D_p'(S^{-1} \otimes S^{-1})(\hat{\underset{\sim}{\Sigma}}-\underset{\sim}{S}) + o_p(n^{-\frac{1}{2}})\end{aligned}$$

cont.

$$= D_p'(\Sigma_o^{-1} \otimes \Sigma_o^{-1})(\hat{\underset{\sim}{\Sigma}} - \underset{\sim}{S}) + o_p(n^{-\frac{1}{2}})$$

$$= E(\hat{\underset{\sim}{\sigma}} - \underset{\sim}{s}) + o_p(n^{-\frac{1}{2}}). \qquad (4.15)$$

Thus from (4.12) and the continuity of the derivatives of $\Sigma(\underset{\sim}{\gamma})$,

$$\underset{\sim}{0} = \frac{\partial F}{\partial \underset{\sim}{\gamma}}\bigg|_{\hat{\underset{\sim}{\gamma}}} = \left[\frac{\partial \underset{\sim}{\sigma}'}{\partial \underset{\sim}{\gamma}} \frac{\partial F}{\partial \underset{\sim}{\sigma}}\right]_{\hat{\underset{\sim}{\gamma}}}$$

$$= \Delta'(\hat{\underset{\sim}{\gamma}}) E(\hat{\underset{\sim}{\sigma}} - \underset{\sim}{s}) + o_p(n^{-\frac{1}{2}})$$

$$= \Delta_o' E(\hat{\underset{\sim}{\sigma}} - \underset{\sim}{s}) + o_p(n^{-\frac{1}{2}})$$

$$= \Delta_o' E(\hat{\underset{\sim}{\sigma}} - \underset{\sim}{\sigma}_o) + \Delta_o' E(\underset{\sim}{\sigma}_o - \underset{\sim}{s}) + o_p(n^{-\frac{1}{2}}). \qquad (4.16)$$

Further, from (4.6)

$$(\hat{\underset{\sim}{\sigma}} - \underset{\sim}{\sigma}_o) = \Delta_o(\hat{\underset{\sim}{\gamma}} - \underset{\sim}{\gamma}_o) + o_p(n^{-\frac{1}{2}}), \qquad (4.17)$$

so substituting in (4.16) we have, since Δ_o has rank q,

$$(\hat{\underset{\sim}{\gamma}} - \underset{\sim}{\gamma}_o) = (\Delta_o' E \Delta_o)^{-1} \Delta_o' E(\underset{\sim}{s} - \underset{\sim}{\sigma}_o) + o_p(n^{-\frac{1}{2}}). \qquad (4.18)$$

Thus by lemma 3 and (4.1), $\sqrt{n}(\hat{\underset{\sim}{\gamma}} - \underset{\sim}{\gamma}_o)$ is asymptotically normally distributed with mean $\underset{\sim}{o}$, and variance matrix given by

$$V\left(\sqrt{n}(\hat{\underset{\sim}{\gamma}} - \underset{\sim}{\gamma}_o)\right) \doteq (\Delta_o' E \Delta_o)^{-1} \Delta_o' E D_p^+ (\Sigma_o \otimes \Sigma_o + \Sigma_o \boxtimes \Sigma_o) (D_p^+)' E \Delta_o (\Delta_o' E \Delta_o)^{-1} \qquad (4.19)$$

and, since, by (2.11), $D_p^+(\Sigma_o \otimes \Sigma_o + \Sigma_o \boxtimes \Sigma_o)(D_p^+)' = 2E^{-1}$,

$$V\left(\sqrt{n}(\hat{\underset{\sim}{\gamma}} - \underset{\sim}{\gamma}_o)\right) \doteq 2(\Delta_o' E \Delta_o)^{-1} \qquad (4.20)$$

<u>Corollary 1</u> Under the conditions of theorem 3

$$\sqrt{n}(\hat{\underset{\sim}{\sigma}} - \underset{\sim}{\sigma}_o) \doteq N\left(\underset{\sim}{0}, 2\Delta_o(\Delta_o' E \Delta_o)^{-1} \Delta_o'\right) \qquad (4.21)$$

<u>Proof</u> This follows immediately from (4.10) and (4.17).

Corollary 2   Under the conditions of theorem 3,

$$nF(\hat{\Sigma};S) \to \chi^2(d) \tag{4.22}$$

where $d = \frac{1}{2}p(p+1) - q$.

Proof   From the proof of theorem 2 we have $(\hat{\theta}_i - 1) = O_p(n^{-\frac{1}{2}})$. Thus applying lemma 1 to $f(\theta)$ we have

$$f(\hat{\theta}_i) = \tfrac{1}{2}(\hat{\theta}_i - 1)^2 + o_p(n^{-1}). \tag{4.23}$$

Hence

$$\begin{aligned} F(\hat{\Sigma};S) &= \tfrac{1}{2}\sum_i(\hat{\theta}_i - 1)^2 + o_p(n^{-1}) \\ &= \tfrac{1}{2}\,\mathrm{tr}[S^{-1}(\hat{\Sigma}-S)]^2 + o_p(n^{-1}) \\ &= \tfrac{1}{2}\,\mathrm{tr}[\Sigma_o^{-1}(\hat{\Sigma}-S)]^2 + o_p(n^{-1}) \\ &= \tfrac{1}{2}(\hat{\underset{\sim}{\sigma}}-\underset{\sim}{s})'E(\hat{\underset{\sim}{\sigma}}-\underset{\sim}{s}) + o_p(n^{-1}), \end{aligned} \tag{4.24}$$

from (3.7). Thus by lemma 3, to determine the asymptotic distribution of $nF(\hat{\Sigma};S)$ it is sufficient to determine the asymptotic distribution of $\frac{n}{2}(\hat{\underset{\sim}{\sigma}}-\underset{\sim}{s})'E(\hat{\underset{\sim}{\sigma}}-\underset{\sim}{s})$.

Now from (4.17) and (4.18) we have

$$\begin{aligned} (\hat{\underset{\sim}{\sigma}}-\underset{\sim}{s}) &= (\hat{\underset{\sim}{\sigma}}-\underset{\sim}{\sigma}_o) + (\underset{\sim}{\sigma}_o-\underset{\sim}{s}) \\ &= \left(\Delta_o(\Delta_o'E\Delta_o)^{-1}\Delta_o'E-I\right)(\underset{\sim}{s}-\underset{\sim}{\sigma}_o) + o_p(n^{-\frac{1}{2}}). \end{aligned} \tag{4.25}$$

That is

$$\begin{aligned} \left(\tfrac{n}{2}\right)^{\frac{1}{2}}E^{\frac{1}{2}}(\hat{\underset{\sim}{\sigma}}-\underset{\sim}{s}) = {} & \left(I-E^{\frac{1}{2}}\Delta_o(\Delta_o'E\Delta_o)^{-1}\Delta_o'E^{\frac{1}{2}}\right) \\ & \left(\tfrac{n}{2}\right)^{\frac{1}{2}}E^{\frac{1}{2}}(\underset{\sim}{\sigma}-\underset{\sim}{s}) + o_p(n^{-\frac{1}{2}}), \end{aligned} \tag{4.26}$$

which is of the form $\underset{\sim}{z} = (I-P)\underset{\sim}{y} + o_p(1)$, with $\underset{\sim}{y} = \left(\frac{n}{2}\right)^{\frac{1}{2}}E^{\frac{1}{2}}(\underset{\sim}{\sigma}_o - s) \to N(\underset{\sim}{o}, I)$ and $P = E^{\frac{1}{2}}\Delta_o(\Delta_o'E\Delta_o)^{-1}\Delta_o'E^{\frac{1}{2}}$, which is idempotent and of rank q. Thus, from the usual theory for the distribution of quadratic forms of normal variables [see e.g. Rao (1973, p.187)], and by lemma 3

$$\underset{\sim}{z}'\underset{\sim}{z} = \frac{n}{2}(\hat{\underset{\sim}{\sigma}}-\underset{\sim}{s})'E(\hat{\underset{\sim}{\sigma}}-\underset{\sim}{s}) \doteq \chi^2(d).$$

Remark Although condition (b) on the model specifies second order derivatives of $\sigma_{ij}(\underset{\sim}{\gamma})$, the results of this section have been established using only the continuity of the first order partial derivatives. This condition is much weaker than those generally adopted for maximum likelihood theory, but is in agreement with the multinomial situation [Rao (1973, p.363)].

2.5 Numerical determination of the estimates

In practice, the m.d.e. will usually be obtained by solution of the equations

$$\frac{\partial F}{\partial \underset{\sim}{\gamma}}(\hat{\underset{\sim}{\gamma}}) = \underset{\sim}{0}. \tag{5.1}$$

As these are frequently non-linear and do not admit an explicit solution, recourse must often be made to iterative numerical methods, such as the Fletcher-Powell method [Fletcher and Powell (1963)] which minimises $F(\Sigma(\underset{\sim}{\gamma});S)$ directly using first order derivative information, or the classical Newton-Raphson method which seeks a solution of (5.1) and uses second order derivatives. In this section we obtain formulae for the first and second order derivatives which can be used for the general model of section 2.3.

Although for particular cases the derivatives w.r.t. $\underset{\sim}{\gamma}$ may be more simply obtained directly from $F(\Sigma(\underset{\sim}{\gamma});S)$, in general cases they can be determined by

$$\frac{\partial F}{\partial \underset{\sim}{\gamma}} = \frac{\partial \underset{\sim}{\sigma}'}{\partial \underset{\sim}{\gamma}} \frac{\partial F}{\partial \underset{\sim}{\sigma}} \tag{5.2}$$

$$\frac{\partial^2 F}{\partial \underset{\sim}{\gamma} \partial \underset{\sim}{\gamma}'} = \frac{\partial \underset{\sim}{\sigma}'}{\partial \underset{\sim}{\gamma}} \frac{\partial^2 F}{\partial \underset{\sim}{\sigma} \partial \underset{\sim}{\sigma}'} \frac{\partial \underset{\sim}{\sigma}}{\partial \underset{\sim}{\gamma}'} + \left(\sum_{i \geq j} \frac{\partial F}{\partial \sigma_{ij}} \frac{\partial^2 \sigma_{ij}}{\partial \gamma_r \partial \gamma_s} \right) \tag{5.3}$$

Accordingly we seek the derivatives of F w.r.t. the elements of Σ. In determining these we shall use the following well known result [see e.g. Jöreskog and Goldberger (1972)].

Lemma 4 If A is any p×p symmetric matrix with distinct eigen values $\theta_1 < \ldots < \theta_p$ and corresponding eigen vectors

$\underset{\sim}{h}_1 \ldots \underset{\sim}{h}_p$, then

$$d\theta_i = \underset{\sim}{h}_i' dA \underset{\sim}{h}_i \tag{5.4}$$

$$d\underset{\sim}{h}_i = \sum_{\substack{j=1 \\ j\neq i}}^{p} \frac{1}{\theta_i-\theta_j} \underset{\sim}{h}_j \underset{\sim}{h}_i' dA \underset{\sim}{h}_j. \tag{5.5}$$

Now let $S^{-\frac{1}{2}}\Sigma S^{-\frac{1}{2}} = H\Theta H'$, with $\theta_1 < \ldots < \theta_p$ the eigen values of $S^{-\frac{1}{2}}\Sigma S^{-\frac{1}{2}}$ and $H'H = I$. Then, applying lemma 4 to $F(\Sigma;S) = \sum_{i=1}^{p} f(\theta_i)$, and writing $U = S^{-\frac{1}{2}}H$, $\alpha_i = f'(\theta_i)$, and $A = \text{diag}\ (\alpha_i)$, we have

$$dF = \sum_{i=1}^{p} f'(\theta_i) \underset{\sim}{h}_i' S^{-\frac{1}{2}} d\Sigma S^{-\frac{1}{2}} \underset{\sim}{h}_i$$

$$= \sum_{i=1}^{p} \alpha_i \underset{\sim}{u}_i' d\Sigma \underset{\sim}{u}_i$$

$$= \text{tr}(UAU'd\Sigma). \tag{5.6}$$

$$ddF = \sum_{i=1}^{p} [df'(\theta_i) \underset{\sim}{u}_i' d\Sigma_1 \underset{\sim}{u}_i + 2\alpha_i \underset{\sim}{u}_i' d\Sigma_1 S^{-\frac{1}{2}} d\underset{\sim}{h}_i]$$

$$= \sum_{i=1}^{p} [f''(\theta_i) \underset{\sim}{u}_i' d\Sigma_2 \underset{\sim}{u}_i \underset{\sim}{u}_i' d\Sigma_1 \underset{\sim}{u}_i$$

$$+ 2\alpha_i \underset{\sim}{u}_i' d\Sigma_1 \sum_{\substack{j=1 \\ j\neq i}}^{p} \frac{1}{\theta_i-\theta_j} \underset{\sim}{u}_j \underset{\sim}{u}_i' d\Sigma_2 \underset{\sim}{u}_j]$$

$$= \sum_{i=1}^{p} \sum_{j=1}^{p} \beta_{ij} \underset{\sim}{u}_i' d\Sigma_1 \underset{\sim}{u}_j \underset{\sim}{u}_i' d\Sigma_2 \underset{\sim}{u}_j \tag{5.7}$$

where

$$\beta_{ij} = \begin{cases} f''(\theta_i) & i=j \\ \dfrac{f'(\theta_i)-f'(\theta_j)}{\theta_i-\theta_j} & i\neq j \end{cases} \tag{5.8}$$

Thus, since by (xi) of Appendix A

$$\underset{\sim}{u}_i' d\Sigma \underset{\sim}{u}_j = (\underset{\sim}{u}_j' \otimes \underset{\sim}{u}_i') d\underset{\sim}{\Sigma} = (\underset{\sim}{u}_j' \otimes \underset{\sim}{u}_i') D_p d\underset{\sim}{\sigma}, \tag{5.9}$$

we have from (5.6) and (5.7),

$$\frac{\partial F}{\partial \underset{\sim}{\sigma}} = D_p'\Big(\sum_{i=1}^{p} \alpha_i \underset{\sim}{u}_i \otimes \underset{\sim}{u}_i\Big) \tag{5.10}$$

$$= D_p'[\mathrm{vec}(UAU')] \tag{5.11}$$

$$\frac{\partial^2 F}{\partial \underset{\sim}{\sigma}\,\partial \underset{\sim}{\sigma}'} = D_p'\Big(\sum_{i=1}^{p} \sum_{j=1}^{p} \beta_{ij}(\underset{\sim}{u}_j\underset{\sim}{u}_j') \otimes (\underset{\sim}{u}_i\underset{\sim}{u}_i')D_p\Big) \tag{5.12}$$

In subscript form these are

$$\frac{\partial F}{\partial \sigma_{k\ell}} = \begin{cases} \sum_{i=1}^{p} \alpha_i u_{ki}^2 & k=\ell \\ 2\sum_{i=1}^{p} \alpha_i u_{ki} u_{\ell i} & k\neq \ell \end{cases} \tag{5.13}$$

$$\frac{\partial^2 F}{\partial \sigma_{k\ell}\,\partial \sigma_{mn}} = \begin{cases} \sum_{i=1}^{p} \sum_{j=1}^{p} \beta_{ij} u_{ki} u_{mi} u_{\ell j} u_{nj} & k=\ell,\ m=n \\ 2\sum_{i=1}^{p} \sum_{j=1}^{p} \beta_{ij} u_{ki} u_{mi} u_{\ell j} u_{nj} & k>\ell,\ m=n \\ 2\sum_{i=1}^{p} \sum_{j=1}^{p} \beta_{ij}(u_{ki} u_{mi} u_{\ell j} u_{nj} + u_{ki} u_{ni} u_{\ell j} u_{mj}) & k>n,\ m>n \end{cases} \tag{5.14}$$

We note that although the derivations of (5.6) and (5.7) are dependent on the assumption that the θ's are distinct, the formulae themselves remain defined when there are equalities in the θ's, provided that when $\theta_i = \theta_j$ we take

$$\beta_{ij} = f''(\theta_i) = \lim_{\theta_j \to \theta_i} \frac{f'(\theta_i) - f'(\theta_j)}{\theta_i - \theta_j} .$$

Further it is easily verified that the values of (5.6) and (5.7) are then independent of the particular set of eigen vectors chosen to correspond to the equal θ's, so that (5.6), (5.7) and thence (5.11) and (5.12) are true generally.

For some of the estimation functions given in table 1, the formulae (5.11) and (5.12) reduce to fairly simple forms whose calculation does not require the computation of eigen values and vectors (in these cases the formulae may also be found directly), while for others (notably GD) this is not the case. For completeness we present in table 2 formulae for each of the estimation functions of table 1, except GLSE, for which there is no simplification.

We note that it was shown in the proof of theorem 1 that, for i = 1 ... p, $\hat{\theta}_i \xrightarrow{p} 1$ as $n \to \infty$. Thus $\hat{\alpha}_i \xrightarrow{p} 0$, and, from (5.8) $\hat{\beta}_{ij} \xrightarrow{p} 1$. Further, since $\sum_i \underset{\sim}{u}_i \underset{\sim}{u}_i' = S^{-1} \xrightarrow{p} \Sigma_o^{-1}$, from (5.3) we have

$$\left.\frac{\partial^2 F}{\partial \underset{\sim}{\gamma} \partial \underset{\sim}{\gamma}'}\right|_{\hat{\underset{\sim}{\gamma}}} \xrightarrow{p} \Delta_o' D_p' (\Sigma_o^{-1} \; \Sigma_o^{-1}) D_p \; \Delta_o$$

$$= \Delta_o' E \Delta_o . \tag{5.15}$$

Thus from (4.10), $2\left(\left.\frac{\partial^2 F}{\partial \underset{\sim}{\gamma} \partial \underset{\sim}{\gamma}'}\right|_{\hat{\underset{\sim}{\gamma}}}\right)^{-1}$ is a consistent estimate of the asymptotic variance matrix of $\sqrt{n}(\hat{\underset{\sim}{\gamma}} - \underset{\sim}{\gamma}_o)$.

A computer routine for the determination of estimates for general structures, utilizing the formulae (5.2) and (5.3) and the Fletcher-Powell method of minimisation, has been written, and is provided in Appendix B. Since the Fletcher-Powell method achieves its convergence properties by building up on each iteration an approximation to the inverse matrix of second derivatives of F at $\hat{\underset{\sim}{\gamma}}$, the program also provides an estimate of the asymptotic variance matrix of the estimates. To implement the

TABLE 2

Derivative formulae for some special estimation functions

FN	α_i	β_{ij}	$\dfrac{\partial F}{\partial \underset{\sim}{\sigma}}$	$\dfrac{\partial^2 F}{\partial \underset{\sim}{\sigma} \partial \underset{\sim}{\sigma}'}$
TGLS	$\frac{1}{\theta_i^2} - \frac{1}{\theta_i^3}$	$\dfrac{\theta_i^2+\theta_i\theta_j+\theta_j^2-\theta_i^2\theta_j-\theta_i\theta_j^2}{\theta_i^3\theta_j^3}$	$D_p' \text{ vec}[\Sigma^{-1}S\Sigma^{-1}-\Sigma^{-1}S\Sigma^{-1}S\Sigma^{-1}]$	$D_p'(2\Sigma^{-1} \otimes \Sigma^{-1}S\Sigma^{-1}S\Sigma^{-1} + \Sigma^{-1}S\Sigma^{-1} \otimes \Sigma^{-1}S\Sigma^{-1}-2\Sigma^{-1}\otimes \Sigma^{-1}S\Sigma^{-1})D_p$
ML	$\frac{1}{\theta_i} - \frac{1}{\theta_i^2}$	$\dfrac{\theta_i+\theta_j-\theta_i\theta_j}{\theta_i^2\theta_j^2}$	$D_p' \text{ vec}[\Sigma^{-1}-\Sigma^{-1}S\Sigma^{-1}]$	$D_p'(2\Sigma^{-1} \otimes \Sigma^{-1}S\Sigma^{-1}-\Sigma^{-1}\otimes \Sigma^{-1})D_p$
GD	$\dfrac{\log \theta_i}{\theta_i}$	$\dfrac{1-\log \theta_i}{\theta_i^2}$ $i=j$ $\left(\dfrac{\log \theta_i}{\theta_i}-\dfrac{\log \theta_j}{\theta_j}\right)\dfrac{1}{\theta_i-\theta_j}$ $i\neq j$		
DIV	$\frac{1}{2}\left(1-\frac{1}{\theta_i^2}\right)$	$\frac{1}{2}\,\dfrac{\theta_i+\theta_j}{\theta_i^2\theta_j^2}$	$\frac{1}{2}\, D_p' \text{ vec}[S^{-1}-\Sigma^{-1}S\Sigma^{-1}]$	$D_p'(\Sigma^{-1}\otimes \Sigma^{-1}S\Sigma^{-1})D_p$
GLS	θ_i-1	1	$D_p' \text{ vec}[S^{-1}\Sigma S^{-1}-S^{-1}]$	$D_p'(S^{-1}\otimes S^{-1})D_p$

routine for any particular structure, it is necessary for the user to provide subroutines for the calculation of $\underset{\sim}{\sigma}(\underset{\sim}{\gamma})$ and $\frac{\partial\underset{\sim}{\sigma}}{\partial\underset{\sim}{\gamma}'}$. On option any of the estimation functions of table 1 may be used, although use of either ML or GLS will be considerably faster since in these cases the formulae of table 2 are used, avoiding eigenvalue calculation. Further details for implementation are contained in appendix B. Unfortunately experience with the factor model and the direct product structure suggests that convergence to the final estimate can be quite slow, so that for commonly used models routines which take advantage of the special nature of the structure would generally be preferable.

Example Bock and Bargmann (1966) have presented data due to Bilodeau (1957) and the corresponding ML estimates for the parameters of the quasi-simplex model

$$\Sigma = A\Gamma A' + \psi I \tag{5.16}$$

where $\Gamma = \mathrm{diag}(\gamma_i)$, and $a_{ij} = 1$ for $i \leq j$, 0 otherwise. The same example has also been considered by Jöreskog (1970b). Browne (1974) has given both the ML and GLS estimates. For the structure (5.16), $\underset{\sim}{\gamma} = (\gamma_1 \ldots \gamma_p \psi)'$ and the matrix $\Delta(\underset{\sim}{\gamma})$ is easily calculated with

$$\begin{aligned} \frac{\partial\sigma_{ij}}{\partial\gamma_k} &= \begin{cases} 1 & p \geq i \geq j \geq k \\ 0 & \text{otherwise} \end{cases} \\ \frac{\partial\sigma_{ij}}{\partial\psi} &= \begin{cases} 1 & i=j \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{5.17}$$

The estimates obtained by VARS for each of the estimation functions of Table 1 are shown in Table 3.

The ML estimates shown agree with those obtained by Bock and Bargmann but differ slightly from those presented by Browne. The same starting values were used for each estimation function, and iteration ceased when all partial derivatives were $<1\times10^{-6}$ in magnitude. The starting values were obtained by transforming S and taking the simple least squares estimates, as suggested by McDonald (1974).

We note that because the structure is linear, for GLS $F(\Sigma;S)$ is quadratic in the parameters and the initial approximation used by VARS to the second derivatives, viz. $\Lambda' D_p'(S^{-1}\otimes S^{-1})D_p\Lambda$, is exact. It follows that the procedure necessarily requires only one iteration for GLS estimation.

TABLE 3

Estimates for Bilodeau's data for the functions of Table 1

	Iterns.	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	ψ	$nF(\Sigma:S)^*$
START		475.0	51.0	13.0	78.0	24.0	5.0	46.0	
TGLS	4	497.5	55.0	16.2	85.0	22.1	2.7	45.8	8.96
ML	4	482.6	54.6	16.0	81.4	21.6	1.6	45.3	9.39
GD	3	475.0	54.3	15.8	79.6	21.4	1.0	45.1	9.48
DIV	3	474.9	54.3	15.8	79.6	21.4	1.0	45.1	9.51
GLS	1	452.3	53.1	15.2	74.3	20.6	-0.8	44.5	9.27
GLSE	4	433.7	51.7	14.4	69.8	20.2	-2.5	44.1	8.50

* n = 151.

2.6 The non-identified case

In many situations the natural expression of the variance structure model will result in a parametrisation for which the parameters $\underset{\sim}{\gamma}$ are not identified by $\Sigma(\underset{\sim}{\gamma})$, and in this section we consider the effects on the results of section 2.4 in this case. Key results for the extension of the theory are given in lemma 5 and its corollaries, which can also prove useful in the determination of the estimates [see e.g. sections 4.2, 5.3]. While these results do not seem to have previously been given explicitly in the estimation literature, they are implicit in the papers by Silvey (1959) and Aitchison and Silvey (1960), which are concerned with identification conditions within the framework of general parametrically constrained ML estimation, and in the ML procedures commonly adopted for non-identified models, e.g. factor analysis.

We commence by reviewing general identification theory [Koopmans and Reiersøl (1950), Rothenberg (1971)] as it pertains to variance structure models. Of particular note is the result (iii), which has already been used in section 2.4, but which again does not appear to have been previously stated in the literature.

As before, we shall assume a parametric model $\Sigma = \Sigma(\underset{\sim}{\gamma})$, with $\underset{\sim}{\gamma}$ a $q\times 1$ vector in some parameter space $\Gamma\subset R^q$, and write $A = \Sigma(\Gamma)$. We will be concerned with the identification of some $\underset{\sim}{\gamma}_o\in\Gamma$, given $\Sigma_o = \Sigma(\underset{\sim}{\gamma}_o)$. We assume that $\underset{\sim}{\gamma}_o$ is an interior point of Γ, and that assumptions (a) and (b) of section 2.3 apply. We let

$\Delta(\underset{\sim}{\gamma})$ be defined by (3.1), with $\Delta_o = \Delta(\underset{\sim}{\gamma}_o)$.

We can distinguish the following forms of identification.

(a) $\underset{\sim}{\gamma}_o$ is said to be identified (globally identified, weakly identified) in Γ if $\underset{\sim}{\gamma}_1 \in \Gamma$ and $\Sigma(\underset{\sim}{\gamma}_1) = \Sigma(\underset{\sim}{\gamma}_o) \Rightarrow \underset{\sim}{\gamma}_1 = \underset{\sim}{\gamma}_o$.

(b) $\underset{\sim}{\gamma}_o$ is said to be locally identified in Γ if there exists a nbd. of $\underset{\sim}{\gamma}_o$ in Γ on which $\underset{\sim}{\gamma}_o$ is identified.

(c) $\underset{\sim}{\gamma}_o$ is said to be strongly identified in Γ if for any sequence $\{\Sigma(\underset{\sim}{\gamma}_n)\}$ in A, $\Sigma(\underset{\sim}{\gamma}_n) \to \Sigma(\underset{\sim}{\gamma}_o) \Rightarrow \underset{\sim}{\gamma}_n \to \underset{\sim}{\gamma}_o$.

Clearly at $\underset{\sim}{\gamma}_o$, strong identification $\Rightarrow$ identification $\Rightarrow$ local identification.

As Rothenberg (1971) has pointed out, the identification problem with models of this type is simply a question of the uniqueness of solutions of systems of equations, so that the classical results of analysis may be applied to give the following.

(i) If Δ_o has rank q, then $\underset{\sim}{\gamma}_o$ is locally identified.

(ii) If $\underset{\sim}{\gamma}_o$ is a regular point of $\Delta(\underset{\sim}{\gamma})$ (i.e. if rank $(\Delta(\underset{\sim}{\gamma}))$ is constant in a nbd. of $\underset{\sim}{\gamma}_o$), then a necessary condition that $\underset{\sim}{\gamma}_o$ be locally identified is that rank $(\Delta_o) = q$.

We note that (i) is really a special case of the inverse function theorem [see e.g. Apostol (1957,p.144)] which in the present case states that there exists open nbd.'s Γ_1 of $\underset{\sim}{\gamma}_o$ in Γ and A_1 of Σ_o in A such that

(a) $A_1 = \Sigma(\Gamma_1)$

(b) $\Sigma(\underset{\sim}{\gamma})$ is 1-1 on Γ_1

(c) there exists a function $\underset{\sim}{g}(\Sigma)$ defined on A_1 such that $\underset{\sim}{g}(A_1) = \Gamma_1$ and $\underset{\sim}{g}(\Sigma(\underset{\sim}{\gamma})) = \underset{\sim}{\gamma}$ for every $\underset{\sim}{\gamma}$ in Γ_1.

(d) $\underset{\sim}{g}$ is continuous with continuous first order partial derivatives with respect to the elements of Σ on A_1.

We note that the expression of $\underset{\sim}{g}(\Sigma)$ is not in general uniquely determined in the sense that it may, for example, be written as a function of any q of the elements of Σ for which the corresponding rows of Δ_0 are linearly independent.

Now, since if $\{\Sigma(\underset{\sim}{\gamma}_n)\}$ is a sequence in A such that $\Sigma(\underset{\sim}{\gamma}_n) \to \Sigma_0$ there exists N such that for all n>N $\Sigma(\underset{\sim}{\gamma}_n) \in A_1$, so for all n>N there exists a sequence, $\{\underset{\sim}{\gamma}_n^*\}$ say, in Γ_1 such that $\Sigma(\underset{\sim}{\gamma}_n^*) = \Sigma(\underset{\sim}{\gamma}_n)$, and, by the continuity of $\underset{\sim}{g}$, $\underset{\sim}{\gamma}_n^* \to \underset{\sim}{\gamma}_0$. Thus if there exists a nbd. Γ_2 of $\underset{\sim}{\gamma}_0$ in Γ such that all $\underset{\sim}{\gamma} \in \Gamma_2$ are identified in Γ, since $\underset{\sim}{\gamma}_n^* \to \underset{\sim}{\gamma}_0$ implies that there exists N_1(>N say) such that for all $n>N_1$ $\underset{\sim}{\gamma}_n^* \in \Gamma_2$, then for all $n>N_1$ $\underset{\sim}{\gamma}_n = \underset{\sim}{\gamma}_n^*$ and so $\underset{\sim}{\gamma}_n \to \underset{\sim}{\gamma}_0$.

Conversely if rank $(\Delta_0) = q$ and there exists a sequence $\{\underset{\sim}{\gamma}_n\}$ in Γ such that $\underset{\sim}{\gamma}_n \to \underset{\sim}{\gamma}_0$ and the $\underset{\sim}{\gamma}_n$ are not identified in Γ, since any $\underset{\sim}{\gamma}_n \in \Gamma_1$ is identified in Γ_1, there exists a sequence $\{\underset{\sim}{\gamma}_n^*\}$ such that $\Sigma(\underset{\sim}{\gamma}_n) = \Sigma(\underset{\sim}{\gamma}_n^*)$ and $\underset{\sim}{\gamma}_n^* \notin \Gamma_1$. Then $\Sigma(\underset{\sim}{\gamma}_n^*) \to \Sigma(\underset{\sim}{\gamma}_0)$ but $\underset{\sim}{\gamma}_n^* \not\to \underset{\sim}{\gamma}_0$, i.e. $\underset{\sim}{\gamma}_0$ is not strongly identified.

Summarising then, we have

(iii) If Δ_0 has rank q, a necessary and sufficient condition that $\underset{\sim}{\gamma}_0$ be strongly identified in Γ is that there exist a nbd. Γ_2 of $\underset{\sim}{\gamma}_0$ in Γ such that all $\underset{\sim}{\gamma}\in\Gamma_2$ are identified in Γ.

When $\underset{\sim}{\gamma}_0$ is not a regular point of $\Delta(\underset{\sim}{\gamma})$, the condition rank $(\Delta_0) = q$ is not necessary for local identification, nor for strong identification. For estimation it is clear from the proof of theorem 1 that an approximate m.d.e. will be consistent for $\underset{\sim}{\gamma}_0$ provided $\underset{\sim}{\gamma}_0$ is strongly identified, so that the condition rank $(\Delta_0) = q$ is not necessary for consistency. However, when rank $(\Delta_0) < q$, from (5.5), $\left.\frac{\partial^2 F}{\partial\underset{\sim}{\gamma}\partial\underset{\sim}{\gamma}'}\right|_{\hat{\underset{\sim}{\gamma}}}$ will be "almost singular" for large n, so that special procedures are likely to be needed for numerical determination of the estimates. Further (4.6) now no longer implies that $(\hat{\underset{\sim}{\gamma}}-\underset{\sim}{\gamma}_0) = O_p(n^{-\frac{1}{2}})$, and the results of theorem 3 on the distribution of $\sqrt{n}(\hat{\underset{\sim}{\gamma}}-\underset{\sim}{\gamma}_0)$ will not apply.

We pass now to the case where $\underset{\sim}{\gamma}_0$ is not identified. If $\underset{\sim}{\gamma}_0$ is locally identified then global identification in Γ can often be achieved by simple restriction of the parameter space Γ. Accordingly we concentrate on the situation where $\underset{\sim}{\gamma}_0$ is not locally identified. In this case there is an infinite number of elements $\underset{\sim}{\gamma}$ of Γ with $\Sigma(\underset{\sim}{\gamma}) = \Sigma_0$, since otherwise $\underset{\sim}{\gamma}_0$ would be identified on a sufficiently small nbd. of $\underset{\sim}{\gamma}_0$. Where these $\underset{\sim}{\gamma}$ define a continuous manifold in R^q, it may be possible to express the model in terms of a smaller number of parameters and so reduce the situation to the identified case. Alternatively, the parameters may be identified for inference purposes by restriction of Γ through say t constraint equations $\underset{\sim}{v}(\underset{\sim}{\gamma}) = \underset{\sim}{0}$, and it is this case we shall now consider.

We say the equations $\underset{\sim}{v}(\underset{\sim}{\gamma}) = \underset{\sim}{0}$ define identification conditions (i.c.) on the model $\Sigma(\underset{\sim}{\gamma})$ if for each $\Sigma \in A$ the equations

$$\Sigma(\underset{\sim}{\gamma}) = \Sigma; \qquad \underset{\sim}{v}(\underset{\sim}{\gamma}) = \underset{\sim}{0} \tag{6.1}$$

have a solution $\underset{\sim}{\gamma} \in \Gamma$. That is, if we write

$$\Gamma_v = \{\underset{\sim}{\gamma} \mid \underset{\sim}{\gamma} \in \Gamma,\ \underset{\sim}{v}(\underset{\sim}{\gamma}) = \underset{\sim}{0}\},\ A_v = \Sigma(\Gamma_v) \tag{6.2}$$

then the equations $\underset{\sim}{v}(\underset{\sim}{\gamma}) = \underset{\sim}{0}$ define i.c. if $A_v = A$. Further we say $\underset{\sim}{\gamma}_o \in \Gamma_v$ is (globally, locally, strongly) identified in Γ by the i.c. $\underset{\sim}{v}(\underset{\sim}{\gamma}) = \underset{\sim}{0}$ if $\underset{\sim}{\gamma}_o$ is (globally, locally, strongly) identified in Γ_v.

We will assume that the functions $\underset{\sim}{v}(\underset{\sim}{\gamma})$ admit continuous first order partial derivatives in a nbd. of $\underset{\sim}{\gamma}_o$ and write

$$V(\underset{\sim}{\gamma}) = \frac{\partial \underset{\sim}{v}}{\partial \underset{\sim}{\gamma}'}(\underset{\sim}{\gamma}), \tag{6.3}$$

$$W(\underset{\sim}{\gamma}) = \begin{bmatrix} \Delta(\underset{\sim}{\gamma}) \\ V(\underset{\sim}{\gamma}) \end{bmatrix}, \tag{6.4}$$

$V_o = V(\underset{\sim}{\gamma}_o)$ and $W_o = W(\underset{\sim}{\gamma}_o)$. Then paralleling the results (i), (ii) and (iii) above we have

(iv) If W_o has rank q then $\underset{\sim}{\gamma}_o$ is locally identified by the i.c. $\underset{\sim}{v}(\underset{\sim}{\gamma}) = \underset{\sim}{0}$.

(v) If $\underset{\sim}{\gamma}_o$ is a regular point of $W(\underset{\sim}{\gamma})$, then a necessary condition that $\underset{\sim}{\gamma}_o$ be locally identified by the i.c. $\underset{\sim}{v}(\underset{\sim}{\gamma}) = \underset{\sim}{0}$ is that rank $(W_o) = q$.

(vi) If W_o has rank q a necessary and sufficient condition that $\underset{\sim}{\gamma}_o$ be strongly identified in Γ by the i.c. $\underset{\sim}{v}(\underset{\sim}{\gamma}) = \underset{\sim}{0}$ is that there exist a nbd. of $\underset{\sim}{\gamma}_o$ in Γ_v such that all $\underset{\sim}{\gamma}$ in this nbd. are identified in Γ by the identification conditions.

The results (iv) and (v) are given by Rothenberg (1971), while (vi) is a direct extension of (iii). We also have the following lemma which is of importance in the determination of m.d.e.'s.

<u>Lemma 5</u> If the equations $\underset{\sim}{v}(\underset{\sim}{\gamma}) = \underset{\sim}{0}$ define i.c. on the model $\Sigma(\underset{\sim}{\gamma})$, and Γ_v is defined by (6.2) then

$$\inf_{\underset{\sim}{\gamma}\in\Gamma_v} F\{\Sigma(\underset{\sim}{\gamma});S) = \inf_{\underset{\sim}{\gamma}\in\Gamma} F(\Sigma(\underset{\sim}{\gamma});S) \qquad (6.5)$$

<u>Proof</u> Since $\Gamma_v \subset \Gamma$

$$\inf_{\underset{\sim}{\gamma}\in\Gamma_v} F(\Sigma(\underset{\sim}{\gamma}):S) \geqslant \inf_{\underset{\sim}{\gamma}\in\Gamma} F(\Sigma(\underset{\sim}{\gamma}):S)$$

Suppose the inequality is strict. Then there exists $\underset{\sim}{\gamma}_1\in\Gamma$ such that

$$F(\Sigma(\underset{\sim}{\gamma}_1);S) < \inf_{\underset{\sim}{\gamma}\in\Gamma_v} F(\Sigma(\underset{\sim}{\gamma});S).$$

However, by our definition of i.c. there exists $\underset{\sim}{\gamma}_2\in\Gamma_v$ such that $\Sigma(\underset{\sim}{\gamma}_2) = \Sigma(\underset{\sim}{\gamma}_1)$ and so

$$F(\Sigma(\underset{\sim}{\gamma}_2);S) < \inf_{\underset{\sim}{\gamma}\in\Gamma_v} F(\Sigma(\underset{\sim}{\gamma});S),$$

which is a contradiction. Thus (6.5) holds.

<u>Corollary 1</u> If $\hat{\gamma}\in\Gamma_v$ is such that

$$F(\Sigma(\hat{\gamma});S) = \inf_{\gamma\in\Gamma_v} F(\Sigma(\gamma);S),$$

and $\hat{\gamma}$ is an interior point of Γ, then

$$\left.\frac{\partial F}{\partial \gamma}\right|_{\hat{\gamma}} = 0.$$

<u>Corollary 2</u> If under the conditions of corollary 1, $\hat{\gamma}$ is determined by the Lagrange multiplier technique, i.e. as a solution of

$$\frac{\partial F}{\partial \gamma}(\gamma) + V'(\gamma)\lambda = 0$$

$$v(\gamma) = 0 \qquad (6.6)$$

with λ a vector of Lagrange multipliers, and if $V(\hat{\gamma})$ has rank t, then $\lambda=0$.

<u>Proof</u> By corollary 1, $V'(\hat{\gamma})\lambda=0$ which if the columns of $V'(\hat{\gamma})$ are linearly independent implies $\lambda=0$.

In order to extend the results of section 2.4 to the case where the parameters are identified by i.c., we modify the assumptions (c) and (d) of section 2.3 on the model to give

(c) γ_o is a regular point of $\Delta(\gamma)$, with rank $(\Delta_o) = r<q$.

(d) there exists s = q-r identification conditions $v(\gamma) = 0$ on Γ, with $v(\gamma_o) = 0$, and $v_i(\gamma)$ possessing continuous first order partial derivatives in a nbd. of γ_o, i = 1 ... s. Further, with $W(\gamma)$ defined by (6.4), rank $(W_o) = q$.

(e) there exists a nbd. of $\underset{\sim}{\gamma}_o$ in Γ_v such that all $\underset{\sim}{\gamma}$ in this nbd. are identified in Γ by the i.c. $\underset{\sim}{v}(\underset{\sim}{\gamma}) = \underset{\sim}{0}$.

We shall also need the following well known result [see e.g. Rao, 1973, p.34]

Lemma 6 Let A and B be matrices with sizes $m\times q$ and $t\times q$ respectively, and $R(A')$ denote the range of A'. Then if rank $(A) = r$, rank $(B) = (q-r)$ and $R(A') \cap R(B') = \underset{\sim}{0}$

(a) $A'A + B'B$ is $q\times q$ and nonsingular

(b) $(A'A + B'B)^{-1}$ is a generalised inverse for $A'A$.

We note too that $R(A') \cap R(B') = \underset{\sim}{0}$ if any only if rank (A') + rank (B') = rank$(A' \vdots B')$, as is easily shown by consideration of basis vectors.

We assume that a sample variance matrix is observed with an asymptotic distribution defined by (4.1). Corresponding to the definitions of section 2.4 we now define an approximate m.d.e. $\underset{\sim}{\gamma}^*$ as an element of Γ_v satisfying

$$F\left(\Sigma(\underset{\sim}{\gamma}^*);S\right) \leqslant c \inf_{\underset{\sim}{\gamma}\in\Gamma} F\left(\Sigma(\underset{\sim}{\gamma});S\right) \tag{6.7}$$

with c a constant and $1<c<\infty$, and note that by lemma 5, an approximate m.d.e. always exists. Similarly an m.d.e. is now defined as an element of Γ_v such that

$$F\left(\Sigma(\hat{\underset{\sim}{\gamma}});S\right) = \inf_{\underset{\sim}{\gamma}\in\Gamma} F\left(\Sigma(\underset{\sim}{\gamma});S\right) \tag{6.8}$$

Writing $\Sigma^* = \Sigma(\underset{\sim}{\gamma}^*)$ for any approximate m.d.e. $\underset{\sim}{\gamma}^*$, we see by (4.5) (which is derived independently of the parametrisation) that $\Sigma^* - \Sigma_o = O_p(n^{-\frac{1}{2}})$. In particular then $\Sigma^* \xrightarrow{p} \Sigma_o$ and since, by (vi) above, conditions (d) and (e) ensure that $\underset{\sim}{\gamma}_o$ is strongly identified in Γ_v, so $\underset{\sim}{\gamma}^* \xrightarrow{p} \underset{\sim}{\gamma}_o$. Further, since

$$\underset{\sim}{\sigma}^* - \underset{\sim}{\sigma}_o = \Delta_o(\underset{\sim}{\gamma}^* - \underset{\sim}{\gamma}_o) + \text{higher order terms}$$

$$\underset{\sim}{0} = \underset{\sim}{v}(\underset{\sim}{\gamma}^*) - \underset{\sim}{v}(\underset{\sim}{\gamma}_o) = V_o(\underset{\sim}{\gamma}^* - \underset{\sim}{\gamma}_o) + \text{higher order terms}$$

we have

$$\Delta_o'(\underset{\sim}{\sigma}^* - \underset{\sim}{\sigma}_o) = (\Delta_o'\Delta_o + V_o'V_o)(\underset{\sim}{\gamma}^* - \underset{\sim}{\gamma}_o) + \text{higher order terms}$$

which, since $\underset{\sim}{\sigma}^* - \underset{\sim}{\sigma}_o = O_p(n^{-\frac{1}{2}})$, by lemma 6, implies $\underset{\sim}{\gamma}^* - \underset{\sim}{\gamma}_o = O_p(n^{-\frac{1}{2}})$. Summarising then, we have corresponding to theorem 1,

<u>Theorem 4</u> If under the i.c. $\underset{\sim}{v}(\underset{\sim}{\gamma}) = \underset{\sim}{0}$, $\underset{\sim}{\gamma}^*$ is any approximate m.d.e. of $\underset{\sim}{\gamma}_o$, then $\underset{\sim}{\gamma}^*$ is consistent for $\underset{\sim}{\gamma}_o$, with $\underset{\sim}{\gamma}^* - \underset{\sim}{\gamma}_o = O_p(n^{-\frac{1}{2}})$.

Similarly, corresponding to theorem 2, using lemma 5, corollary 1, it is easily shown that

<u>Theorem 5</u> The probability that an m.d.e. exists and is given as a solution of

$$\frac{\partial F}{\partial \underset{\sim}{\gamma}}\left(\Sigma(\underset{\sim}{\gamma}); S\right) = \underset{\sim}{0} \tag{6.9}$$

$\to 1$ as $n \to \infty$.

It then follows from (4.16) and (4.17) that with E defined by (4.11)

$$\Delta_o' E \Delta_o(\hat{\underset{\sim}{\gamma}} - \underset{\sim}{\gamma}_o) = \Delta_o' E(\underset{\sim}{s} - \underset{\sim}{\sigma}_o) + o_p(n^{-\frac{1}{2}}) \tag{6.10}$$

and so, since

$$\underset{\sim}{0} = \underset{\sim}{v}(\hat{\underset{\sim}{\gamma}}) - \underset{\sim}{v}(\underset{\sim}{\gamma}_o) = V_o(\hat{\underset{\sim}{\gamma}}-\underset{\sim}{\gamma}_o) + o_p(n^{-\frac{1}{2}}) \tag{6.11}$$

and since $R(\Lambda_o{}'E^{\frac{1}{2}}) = R(\Lambda_o{}')$, by lemma 6

$$(\hat{\underset{\sim}{\gamma}}-\underset{\sim}{\gamma}_o) = (\Lambda_o{}'E\Lambda_o+V_o{}'V_o)^{-1}\Lambda_o{}'E(\underset{\sim}{s}-\underset{\sim}{\sigma}_o) + o_p(n^{-\frac{1}{2}}) \tag{6.12}$$

Thus corresponding to theorem 3 we have

Theorem 6 Under the i.c. $\underset{\sim}{v}(\underset{\sim}{\gamma}) = \underset{\sim}{0}$, let an estimator $\hat{\underset{\sim}{\gamma}}$ be defined as an m.d.e. whenever an m.d.e. exists, and otherwise as an approximate m.d.e. Then

$$\sqrt{n}(\hat{\underset{\sim}{\gamma}}-\underset{\sim}{\gamma}_o) \div N(\underset{\sim}{0}, 2G^{-1}(\Lambda_o{}'E\Lambda_o)G^{-1}) \tag{6.13}$$

where

$$G = \Lambda_o{}'E\Lambda_o + V_o{}'V_o. \tag{6.14}$$

Corollary 1 Under the conditions of theorem 6,

$$\sqrt{n}(\hat{\underset{\sim}{\sigma}}-\underset{\sim}{\sigma}_o) \div N(\underset{\sim}{0}, 2\Lambda_o[G^{-1}(\Lambda_o{}'E\Lambda_o)G^{-1}]\Lambda_o{}') \tag{6.15}$$

Further, using similar reasoning to that of theorem 3, corollary 2 we have

Corollary 2 Under the conditions of theorem 6

$$nF(\hat{\Sigma};S) \div \chi^2(d) \tag{6.16}$$

where $d = \frac{1}{2}p(p+1) - r$.

We note that, with reference to the condition (d) on the model, it has been assumed that i.c. $\underset{\sim}{v}(\underset{\sim}{\gamma}) = \underset{\sim}{0}$ exist and are such that $\underset{\sim}{v}(\underset{\sim}{\gamma}_o) = \underset{\sim}{0}$. In practice $\underset{\sim}{v}(\underset{\sim}{\gamma})$ is usually chosen for convenience in the estimation procedure, and $\underset{\sim}{\gamma}_o$ becomes that member of the family of $\underset{\sim}{\gamma}$ with $\Sigma(\underset{\sim}{\gamma})=\Sigma(\underset{\sim}{\gamma}_o)$ which satisfies the condition. Although this $\underset{\sim}{\gamma}_o$ may no longer be the true (i.e. population) value of $\underset{\sim}{\gamma}$ it is indistinguishable from it when inference is via Σ. We also note that in assuming $(q-r)$ conditions with rank $(V_o) = (q-r)$ we are merely ensuring that the conditions are essentially independent. No real changes to the theory are required if there are more than $(q-r)$ i.c. provided that rank $(V_o) = (q-r)$ and rank $(W_o) = q$. Remarks similar to those following result (iii) above can also be made about models for which $\underset{\sim}{\gamma}_o$ is not a regular point of $\Delta(\underset{\sim}{\gamma}_o)$ or of $W(\underset{\sim}{\gamma}_o)$, with such cases usually requiring individual consideration.

2.7 Application to models of root structure

To illustrate some of the effects of the choice of $F \in \mathcal{F}$ we consider the estimation of the parameters of models which specify structure for the eigen values of Σ, but for which no constraint is imposed on the corresponding eigen vectors, i.e. if we assume the eigen values of Σ are $\lambda_1 \geq \ldots \geq \lambda_p > 0$, $\Lambda = \mathrm{diag}(\lambda_i)$, and

$$\Sigma = J\Lambda J', \quad J'J = I, \tag{7.1}$$

we are concerned with models which specify a structure for Λ, say $\Lambda \in \Gamma_\lambda$, but for which J may be any element of $O(p)$, the group of orthogonal $p \times p$ matrices. Models which specify sets of equal roots are common examples of this type.

The determination of the m.d.e. of Λ and J following the observation of a sample variance matrix S, with eigen values $\ell_1 > \ldots > \ell_p > 0$, $L = \mathrm{diag}(\ell_i)$ and

$$S = KLK', \quad K'K = I, \tag{7.2}$$

may be conveniently divided into two parts, viz., the minimisation of $F(J\Lambda J';S)$ with respect to J for given Λ, then the minimisation of this conditional minimum with respect to Λ. For maximum likelihood estimation it has been shown by Anderson (1963) that with $F_2(\Sigma;S)$ defined in table 1, $F_2(J\Lambda J';S)$ is minimised when $J=K$ independently of Λ, and so

$$\inf_{J \in O(p)} F_2(J\Lambda J';S) = F_2(\Lambda;L), \tag{7.3}$$

which may then be minimised with respect to Λ. We seek a corresponding result for general $F \in \mathcal{F}$. In doing so we make use of the following well known lemmas [see e.g. Bellman (1960)], which we restate here in a convenient notational form.

Lemma 7 Let $\Theta = \mathrm{diag}(\theta_i)$ with $\theta_1 \geq \ldots \geq \theta_p$ have k distinct diagonal elements with multiplicities $m_1 \ldots m_k$. Let H and J be elements of $O(p)$ and $\Phi = \mathrm{diag}(\phi_i)$. Then

$$J\Phi J' = H\Theta H' \tag{7.4}$$

if and only if J and Φ may be written in the forms

$$J = HMP', \quad \Phi = P\Theta P' \tag{7.5}$$

for some permutation matrix P and some matrix

$$M = \begin{bmatrix} M_1 & 0 & \ldots & 0 \\ 0 & M_2 & \ldots & 0 \\ . & . & \ldots & . \\ 0 & 0 & \ldots & M_k \end{bmatrix} = \mathrm{diag}\ (M_i)$$

with $M_i \in O(m_i)$.

Lemma 8 Let A and B be $p \times p$ symmetric matrices with $\theta_1 > \ldots > \theta_p$ the eigen values of A, $\phi_1 \geq \ldots \geq \phi_p$ the eigen values of B, with k distinct values of multiplicities $m_1 \ldots m_k$, and $A = H\Theta H'$, $B = J\Phi J'$ with $H, J \in O(p)$. Then

$$AB = BA \tag{7.6}$$

if and only if J can be written in the form

$$J = HPM' \tag{7.7}$$

for some permutation matrix P and some $M = \mathrm{diag}(M_i)$, with $M_i \in O(m_i)$.

Now corresponding to the maximum likelihood situation we have

<u>Theorem 7</u> Let S be a fixed $p \times p$ positive definite symmetric matrix and L and K be defined by (7.2). Let $\Lambda = \mathrm{diag}(\lambda_i)$ be a fixed diagonal matrix with $\lambda_1 \geq \ldots \geq \lambda_p > 0$ and k distinct diagonal elements with multiplicities $n_1 \ldots n_k$. Then for any $F \in \mathcal{F}$ such that $\theta f'(\theta)$ is monotone increasing on $\theta > 0$, J is a stationary point of $F(J\Lambda J'; S)$ on $O(p)$ if and only if J can be written in the form

$$J = KPN', \tag{7.8}$$

for some permutation matrix P, and $N = \mathrm{diag}\ (N_i)$, with $N_i \in O(n_i)$.

<u>Proof</u> For any J we write $\Sigma = J\Lambda J'$, and let the eigen values of $S^{-\frac{1}{2}}\Sigma S^{-\frac{1}{2}}$ be $\theta_1 \geq \ldots \geq \theta_p > 0$ with k distinct values with multiplicities $m_1 \ldots m_k$, $\Theta = \mathrm{diag}\ (\theta_i)$ and $S^{-\frac{1}{2}}\Sigma S^{-\frac{1}{2}} = H\Theta H'$, $H'H = I$. Then, with $\alpha_i = f'(\theta_i)$, $A = \mathrm{diag}(\alpha_i)$, by (5.6)

$$\begin{aligned} dF &= \mathrm{tr}\ HAH'S^{-\frac{1}{2}}d\Sigma S^{-\frac{1}{2}} \\ &= 2\,\mathrm{tr}\ HAH'S^{-\frac{1}{2}}dJ\Lambda J'S^{-\frac{1}{2}} \\ &= 2\,\mathrm{tr}\ J\Lambda J'S^{-\frac{1}{2}}HAH'S^{-\frac{1}{2}}dJJ', \end{aligned} \tag{7.9}$$

which, since $JJ' = I$ implies dJJ' is skew symmetric, is equal to zero for all dJJ' if and only if $J\Lambda J'S^{-\frac{1}{2}}HAH'S^{-\frac{1}{2}}$ is symmetric. That is, if and only if

$$SS^{-\frac{1}{2}}\Sigma S^{-\frac{1}{2}}HAH' = HAH'S^{-\frac{1}{2}}\Sigma S^{-\frac{1}{2}}S$$

$$\Leftrightarrow (KLK')(H\Theta AH') = (H\Theta AH')(KLK'). \tag{7.10}$$

Thus J is a stationary point of $F(J\Lambda J';S)$ on $O(p)$ iff the corresponding H and Θ satisfy (7.10), i.e. by lemma 8 and the assumed monotonicity of $\theta f'(\theta)$ if and only if H has the form

$$H = KP_1M' \tag{7.11}$$

for some permutation matrix P_1 and $M = \mathrm{diag}(M_i)$, with $M_i \in O(m_i)$. In such cases

$$\begin{aligned}\Sigma = J\Lambda J' &= S^{\frac{1}{2}}\, H\Theta H' S^{\frac{1}{2}} \\ &= KL^{\frac{1}{2}}K'KP_1M'\Theta MP_1'K'KL^{\frac{1}{2}}K' \\ &= KL^{\frac{1}{2}}P_1\Theta P_1'L^{\frac{1}{2}}K' \end{aligned} \tag{7.12}$$

which by lemma 7 is true if and only if for some P and N as in the statement of the theorem

$$K = JNP', \quad L^{\frac{1}{2}}P_1\Theta P_1'L \;\; = P\Lambda P' \tag{7.13}$$

i.e.

$$J = KPN', \quad \Theta = P_1'L^{-\frac{1}{2}}P\Lambda P'L^{-\frac{1}{2}}P_1 \tag{7.14}$$

Remark   For all $F \in \mathcal{F}$ the conditions (3.3) and (3.4) ensure that $\theta f'(\theta)$ is monotone increasing in a nbd. of $\theta=1$, but not necessarily on $\theta>0$. Where this condition fails it is easily verified that $J = KPN'$ gives H and Θ satisfying (7.10) and so is a stationary point of $F(J\Lambda J';S)$ on $O(p)$. However, the necessity argument fails for particular S and Λ. The situation is considered further later.

Corollary 1   Writing $J_p = KPN'$, with P and N as above, the minimum over all permutation matrices P of $F(J_p\Lambda J_p';S)$ occurs when $P = I$ for all Λ and S

if and only if

$$f\left(\frac{\lambda_j}{\ell_i}\right) + f\left(\frac{\lambda_i}{\ell_j}\right) - f\left(\frac{\lambda_i}{\ell_i}\right) - f\left(\frac{\lambda_j}{\ell_j}\right) \geq 0 \tag{7.15}$$

for all $\lambda_i \geq \lambda_j > 0$, $\ell_i > \ell_j > 0$.

Proof $F(J_p \Lambda J_p' ; S)$

$$= F(KPN'\Lambda NP'K';KLK')$$

$$= F(P\Lambda P';L). \tag{7.16}$$

Thus, if $\Lambda_p = P\Lambda P' = \mathrm{diag}(\lambda_i^*)$,

$$F(J_p \Lambda J_p' ; S) = F(\Lambda_p;L) = \sum_{i=1}^{p} \left(\frac{\lambda_i^*}{\ell_i}\right) \tag{7.17}$$

The condition (7.15) is clearly necessary. For suppose there existed $\lambda_i \geq \lambda_j > 0$; $\ell_i > \ell_j > 0$ such that (7.15) failed. Then taking Λ and S so that $\lambda_i, \lambda_j, \ell_i, \ell_j$ are the i and j eigen values of Λ and S respectively, and $\Lambda_p = (\lambda_1 \ldots \lambda_{i-1}\ \lambda_j\ \lambda_{i+1} \ldots \lambda_{j-1}\ \lambda_i\ \lambda_{j+1} \ldots \lambda_p)$, then

$$F(\Lambda_p;L) - F(\Lambda;L) = f\left(\frac{\lambda_j}{\ell_i}\right) + f\left(\frac{\lambda_i}{\ell_j}\right) - f\left(\frac{\lambda_i}{\ell_i}\right) - f\left(\frac{\lambda_j}{\ell_j}\right) \tag{7.18}$$

<0, so that the minimum is not achieved at P=I.

To demonstrate the sufficiency, suppose that (7.15) holds for all $\lambda_i \geq \lambda_j > 0$, $\ell_i > \ell_j > 0$. Suppose that $\Lambda_p \neq \Lambda$, and has $\lambda_r^* = \lambda_r$, $r<i$, $\lambda_i^* = \lambda_j \neq \lambda_i$. Then $\lambda_k^* = \lambda_i$ for some k>i, and if $\Lambda_{p_1} = \mathrm{diag}(\lambda_1 \ldots \lambda_i\ \lambda_{i+1}^* \ldots \lambda_{k-1}^*\ \lambda_j\ \lambda_{k+1}^* \ldots \lambda_p^*)$

$$F(\Lambda_p;L) - F(\Lambda_{p_1};L) = f\left(\frac{\lambda_j}{\ell_i}\right) + f\left(\frac{\lambda_i}{\ell_k}\right) - f\left(\frac{\lambda_i}{\ell_i}\right) - f\left(\frac{\lambda_j}{\ell_k}\right) \tag{7.19}$$

≥ 0 by (7.15). Applying the same argument to Λ_{p_1} and so on, it follows that $F(\Lambda_p;L) - F(\Lambda;L) \geq 0$.

The theorem and corollary together show that if $F(\Sigma;S)\in F$ is such that $\theta f'(\theta)$ is monotone increasing and $f(\theta)$ satisfies (7.15) for all $\lambda_i \geq \lambda_j > 0$, $\ell_i > \ell_j > 0$, then for all Λ and S of the assumed forms

$$\inf_{J\in O(p)} F(J\Lambda J' \cdot S) = F(\Lambda, L) \qquad (7.20)$$

with the minimum being achieved at $J = KN'$, with N as in theorem 7.

It is easily verified that of the estimation functions presented in table 1, ML, GD and DIV each satisfies both of the conditions and therefore (7.20). For TGLS $\theta f'(\theta)$ is monotone increasing on $0<\theta\leq 2$, while (7.15) holds if and only if

$$(\ell_i+\ell_j)(\lambda_i+\lambda_j) \geq 2\lambda_i\lambda_j. \qquad (7.21)$$

Similarly for GLS, $\theta f'(\theta)$ is monotone increasing on $\tfrac{1}{2}\leq\theta$, while (7.15) holds if and only if

$$(\ell_i+\ell_j)(\lambda_i+\lambda_j) \geq 2\ell_i\ell_j, \qquad (7.22)$$

and a similar result is also true for GLSE.

We note, however, that for any $F\in F$, if $\hat{\Sigma} = \hat{J}\hat{\Lambda}\hat{J}'$ is an m.d.e. of Σ, as $n\to\infty$

$$\Pr\left(\inf_{J\in O(p)} F(J\hat{\Lambda}J':S) = F(\hat{\Lambda};L)\right) \to 1. \qquad (7.23)$$

For, as shown in section 2.4, $F(\hat{\Sigma};S) \xrightarrow{p} 0$, $\hat{\Sigma}-S \xrightarrow{p} 0$, and $\hat{\Theta}-I \xrightarrow{p} 0$. If $F(J\hat{\Lambda}J';S)$ were to admit a stationary point on $O(p)$ not of the form KPN', then, by the argument of theorem 7, it must have at least one of the corresponding θ outside the nbd. of $\theta=1$ on which $\theta f'(\theta)$ is monotone increasing. Thus since $\hat{\theta}-I \xrightarrow{p} 0$ the probability that $\hat{J}$ has the form KPN' approaches 1 as $n\to\infty$.

To establish that for any $F\in\mathcal{F}$

$$\Pr\left(\min_{J=KPN'} F(J\hat{\Lambda}J';S) = F(\hat{\Lambda};L)\right)\to 1 \qquad (7.24)$$

as $n\to\infty$, is rather lengthy because of the difficulties raised by certain special cases of equality in the roots of Σ_o, the population variance matrix, and we demonstrate here only that

$$\Pr\left[f\left(\frac{\hat{\lambda}_j}{\ell_i}\right) + f\left(\frac{\hat{\lambda}_i}{\ell_j}\right) - f\left(\frac{\hat{\lambda}_i}{\ell_i}\right) - f\left(\frac{\hat{\lambda}_j}{\ell_j}\right) \geqslant 0\right] \to 1 \qquad (7.25)$$

as $n\to\infty$, for $i < j$, which by the arguement of corollary 1 is a necessary condition for (7.24). Similar reasoning can be used to establish a corresponding sufficient condition.

Since $\hat{\Sigma} - \Sigma_o \xrightarrow{p} 0$ and $S - \Sigma_o \xrightarrow{p} 0$, if the eigen values of Σ_o are $\lambda_1^{(o)} \geqslant \ldots \geqslant \lambda_p^{(o)} > 0$, and $\Lambda_o = \mathrm{diag}(\lambda_i^{(o)})$, then $\hat{\Lambda} - \Lambda_o \xrightarrow{p} 0$, and $L - \Lambda_o \xrightarrow{p} 0$. If the number of independent parameters in Λ, J is to be constant in a nbd. of Λ_o in Γ_λ, equalities in the elements of Λ_o must be matched by equalities in the elements of Λ for all Λ in a nbd of Λ_o in Γ_λ. Thus for all $i > j$ such that $\lambda_i^{(o)}=\lambda_j^{(o)}$, as $n\to\infty$ $\Pr(\hat{\lambda}_i = \hat{\lambda}_j)\to 1$, and hence

$$\Pr\left[f\left(\frac{\hat{\lambda}_j}{\ell_i}\right) + f\left(\frac{\hat{\lambda}_i}{\ell_j}\right) - f\left(\frac{\hat{\lambda}_i}{\ell_i}\right) - f\left(\frac{\hat{\lambda}_j}{\ell_j}\right) = 0\right] \to 1. \tag{7.26}$$

If $\lambda_i^{(o)} > \lambda_j^{(o)}$, then as $n\to\infty$

$$f\left(\frac{\hat{\lambda}_j}{\ell_i}\right) + f\left(\frac{\hat{\lambda}_i}{\ell_j}\right) - f\left(\frac{\hat{\lambda}_i}{\ell_i}\right) - f\left(\frac{\hat{\lambda}_j}{\ell_j}\right) \xrightarrow{p} f\left(\frac{\lambda_j^{(o)}}{\lambda_i^{(o)}}\right) + f\left(\frac{\lambda_i^{(o)}}{\lambda_j^{(o)}}\right) \tag{7.27}$$

> 0, and so (7.25) holds.

We also note that since, given $\hat{J} = KN'$,

$$F(\hat{\Lambda};L) = \inf_{\Lambda\varepsilon\Gamma_\lambda} F(\Lambda;L) \tag{7.28}$$

(assuming that the inf is achieved), by (7.23) for any $F\in\mathcal{F}$ the asymptotic distribution of the estimate of Λ obtained by miniimisation of $F(\Lambda;L)$ on Γ_λ is the same as that of the true m.d.e. $\hat{\Lambda}$.

As an illustration of the differences in estimates of Λ which can occur through varying choice of F, we present in Table 4 the estimates obtained by minimisation of $F(\Lambda;L)$ for the equal roots model

$$\Lambda = \mathrm{diag}(\lambda_i I_{m_i}),\ \lambda_1 >\ldots.> \lambda_k, \tag{7.29}$$

when the eigen values of S are $\ell_{11} >\ldots> \ell_{1m_1} > \ell_{21} >\ldots> \ell_{km_k}$, for each of the estimation functions of Table 1. We note that the harmonic mean estimate

$$\hat{\lambda}_i = \left(\sum_{j=1}^{m_i} \ell_{ij}^{-1}\right)^{-1}$$

which does not appear in Table 3 is the estimate obtained by the minimisation of $F_2^*(\Lambda;L) = F_2(L;\Lambda)$.

TABLE 4

Some estimates for the equal roots model

FUNCTION	$\hat{\lambda}_i$
TGLS	$\left(\sum_{j=1}^{m_i} \ell_{ij}^2\right) / \left(\sum_{j=1}^{m_i} \ell_{ij}\right)$
ML	$\left(\sum_{j=1}^{m_i} \ell_{ij}\right) / m_i$
GD	$\left(\prod_{j=1}^{m_i} \ell_{ij}\right)^{1/m_i}$
DIV	$\left(\left(\sum_{j=1}^{m_i} \ell_{ij}\right) / \left(\sum_{j=1}^{m_i} \ell_{ij}^{-1}\right)\right)^{\frac{1}{2}}$
GLS	$\left(\sum_{j=1}^{m_i} \ell_{ij}^{-1}\right) / \left(\sum_{j=1}^{m_i} \ell_{ij}^{-2}\right)$
GLSE	No explicit solution

Chapter 3: The test statistic

3.1 Introduction

We have shown in section 2.4 that for any $F \in \mathcal{F}$ and model $\Sigma = \Sigma(\underset{\sim}{\gamma})$ satisfying the conditions (a) to (d) of section 2.3,

$$T = nF(\hat{\Sigma};S) \doteq \chi^2(d) \tag{1.1}$$

where $\hat{\Sigma} = \Sigma(\hat{\gamma})$, with $\underset{\sim}{\hat{\gamma}}$ an m.d.e. of $\underset{\sim}{\gamma}$ and

$$d = \tfrac{1}{2}p(p+1) - q. \tag{1.2}$$

The statistic T thus provides a readily applied test of the appropriateness of the model. We have however at this stage no idea of how large n need be before the approximation to the distribution of T provided by (1.1) is reasonable. As an accurate knowledge of the distribution of the test statistic would be a good reason for prefering one element of $\mathcal{F}$ to another, in this chapter we examine ways in which the approximation in (1.1) may be assessed and improved.

Box (1949) derived for any statistic W with a prescribed form for its moments, and for constant ρ, an asymptotic expansion for the cumulant generating function (c.g.f.) of $\rho M = \rho(-2 \log W)$ [see also Anderson, 1958, p. 203]. When nS follows the Wishart distribution, for many standard hypotheses of variance structure the moments of λ have the required form [Anderson, 1958, chapters 8 and 9] and the expansion for the c.g.f. of $\rho M = \rho(-2 \log \lambda)$ has the form

$$\Phi(t) = -\frac{d}{2}\log(1-2it) + \sum_{r=1}^{m} \omega_r[(1-2it)^{-r} - 1] + O(n^{-(m+1)}) \tag{1.3}$$

where $\omega_r = O(n^{-r})$ and is independent of t. It follows that the distribution function (c.d.f.) of ρM also has the form

$$Pr(\rho M < x) = P_d + \omega_1(P_{d+2}-P_d) + O(n^{-2}), \qquad (1.4)$$

where

$$P_d = Pr(\chi^2(d)<x). \qquad (1.5)$$

If then ρ is chosen so that $\omega_1=0$,

$$Pr(\rho M<x) = P_d + O(n^{-2}). \qquad (1.6)$$

We say that the distribution of ρM agrees with $\chi^2(d)$ to $O(n^{-1})$, since the $O(n^{-1})$ terms on both sides of (1.6) agree.

This result was extended by Lawley (1956) who showed that under quite general conditions on the likelihood function and the form of the hypothesis there exists a constant ρ such that the distribution of $\rho(-2 \log \lambda)$ agrees with $\chi^2(d)$ to $O(n^{-1})$, where as before λ is the likelihood ratio. Lawley's assumptions would generally be satisfied for models of variance structure, where, with nS distributed Wishart, $-2 \log \lambda = nF_2(\hat{\Sigma};S) = T_2$ say, with $F_2(\Sigma;S)$ as defined in section 2.3.

In order to determine a suitable ρ it is then sufficient to determine $E[T_2]$ in the form

$$E[T_2] = d(1 + \frac{k_1}{n} + \frac{k_2}{n^2} + O(n^{-3})) \qquad (1.7)$$

whence any factor of the form

$$\rho = 1 - \frac{k_1}{n} + O(n^{-2}) \qquad (1.8)$$

will achieve the desired result. In practice ρ is usually chosen as either $(1-k_1/n)$ or $(1+k_1/n)^{-1}$, the choice being made to give the smaller $O(n^{-2})$ term, i.e. the former will be favoured if $2k_2 > k_1^2$.

The use of a factor in this way is very appealing in its simplicity and effectiveness for the ML test, and it might be hoped that similar properties apply for general $F \in \mathcal{F}$.

In order to explore the situation, in section 2, for a slightly restricted class of elements of $\mathcal{F}$, we consider in detail the test for specified Σ, i.e. we examine the distribution of $T = nF(\Sigma_o;S)$ for $nS \sim W_p(n,\Sigma_o)$ with Σ_o known, on the reasonable assumption that this will give some guide to more general cases and certainly that if for some $F \in \mathcal{F}$ the distribution approximation available in this case is poor then this is also likely to be true more generally.

While the results of section 3.2 were primarily developed as a guide to more general models, there being good approximation to the distribution of the ML test statistic already available [Korin (1968), Davis (1971)] they may still be of interest for testing $\Sigma = \Sigma_o$ for other reasons. Thus Nagao (1973) considered the use of $nF_1(\Sigma_o;S)$, because of its relationship with the asymptotic variance of the ML test statistic under the alternative hypothesis.

Two approaches are adopted to the distribution approximation. Firstly the expectation of the test statistic is considered with the aim of determining, in the manner of (1.7) and (1.8), a factor ρ for which $E[\rho T] = \frac{1}{2}p(p+1) + O(n^{-2})$. Secondly a technique used by Nagao (1973) is applied to obtain an expansion for the c.d.f. of T to $O(n^{-1})$, corresponding to (1.4). A consequence of the form of the expansion is that in general a result corresponding to (1.6) is not possible. A numerical experiment is then described which compares the two methods with each other and the simple approximation (1.1).

In section 3.3 we show that for models of a type frequently encountered in practice, the form of $nF(\hat{\Sigma};S)$ simplifies for many of the commonly used elements of F. The desirability of some improvement to the approximation (1.1), then leads us to suggest in section 3.4 two factors which might be used for general models when detailed knowledge of $E[nF(\hat{\Sigma};S)]$ is not available, as is frequently the case when estimates are determined iteratively. The results of section 3.3 suggest for the important ML case, two further approximate factors which are also presented in section 3.4. A comparison is then made between the proposed factors and the appropriate factors for two models for which these are available.

3.2 The test for specified Σ

3.2.1. Large sample expectations of the test statistic

Under the assumption $nS \sim W_p(n,\Sigma_o)$, we derive $E[nF(\Sigma_o;S)]$ to $O(n^{-2})$, and thence an appropriate correction factor, ρ, such that

$$E[n\rho F(\Sigma_o;S)] = \tfrac{1}{2}p(p+1) + O(n^{-2}) \tag{2.1}$$

with ρ being chosen either so that $O(n^{-2})$, the remainder, is zero, or in the manner described by (1.7) and (1.8). We restrict consideration to $F \in \mathcal{F}$ for which $f(\theta)$ admits a Taylor expansion of the form

$$f(\theta) = \sum_{j=2}^{6} a_j(\theta-1)^j + O(\theta-1)^7, \tag{2.2}$$

where $a_2 = \tfrac{1}{2}$. We denote this subset of $\mathcal{F}$ by $\mathcal{F}_1$, and note that all the estimation functions considered in section 2.3 are elements of $\mathcal{F}_1$. Writing $\phi_i = \theta_i^{-1}$, so that $\phi_1 \geq \dots \geq \phi_p$ are the eigen values of $\Sigma^{-\frac{1}{2}}S\Sigma^{-\frac{1}{2}}$, each $F \in \mathcal{F}_1$ may also be written in the form

$$F(\Sigma;S) = \sum_{j=2}^{6} b_j \sum_{i=1}^{p}(\phi_i - 1)^j + O(\phi_i - 1)^7 \tag{2.3}$$

and for this section we find this form more convenient than the expansion based on (2.2). We note that the a's and b's are simply related by

$$a_j = (-1)^j \sum_{r=2}^{j} \binom{j-1}{r-1} b_r; \tag{2.4}$$

$$b_j = (-1)^j \sum_{r=2}^{j} \binom{j-1}{r-1} a_r. \tag{2.5}$$

As shown in Theorem 1 of Chapter 2, since $S - \Sigma_o = O_p(n^{-\frac{1}{2}})$, when $\phi_1 \geq \dots \geq \phi_p$ are the eigen values of $\Sigma_o^{-\frac{1}{2}}S\Sigma_o^{-\frac{1}{2}}$,

$$(\phi_i - 1) = O_p(n^{-\frac{1}{2}}), \quad i=1\dots p, \tag{2.6}$$

so that to determine $E[nF(\Sigma_o;S)]$ to $O(n^{-1})$ we need only take the first three terms in (2.3) while the first five term will be sufficient to determine $E[nF(\Sigma_o;S)]$ to $O(n^{-2})$.

We seek the expectations of the individual terms of (2.3). In obtaining these, we shall use the following lemma due to Das Gupta (1968) [see also Kshirsagar, 1972, p.70].

Lemma 1 If $A \sim W_p(n,I)$, then

$$E[A^k] = c_k I_p \tag{2.7}$$

where c_k is a constant depending on k,n, and p.
The proof is given by Kshirsagar who also used direct methods to determine the values for k = -2, -1, 1 and 2.

Corollary. Under the conditions of lemma 1

$$E[tr(A^k)] = E[s_k] = c_k p \tag{2.8}$$

where $s_k = tr(A^k)$ is the k th power sum of the latent roots of A.

The general form for $E[s_k]$, $k > o$, and the expectations of the other basic power sum functions of a matrix ZA, with $A \sim W_p(n,\Sigma)$, for general fixed Σ and Z has been given by Fujikoshi (1973, formulae (3.9), (5.15)-(5.22)), who also gives tables for calculation for $k \leq 6$.
Applying his method for $Z = \Sigma = I$ and combining the results with those given by Kshirsagar, we obtain the following values for c_k:

$$c_{-2} = (n-1)/\{(n-p)(n-p-1)(n-p-3)\}, n>p+3$$

$$c_{-1} = 1/(n-p-1), n>p+1$$

$$c_o = 1$$

$c_1 = n$

$c_2 = n^2+n(p+1)$

$c_3 = n^3+n^2(3p+3)+n(p^2+3p+4)$

$c_4 = n^4+n^3(6p+6)+n^2(6p^2+17p+21)+n(p^3+6p^2+21p+20)$

$c_5 = n^5+n^4(10p+10)+n^3(20p^2+55p+65)+n^2(10p^3+55p^2+175p+160)$
$\quad + n(p^4+10p^3+65p^2+160p+148)$

$c_6 = n^6+n^5(15p+15)+n^4(50p^2+135p+155)+n^3(50p^3+262p^2+787p+701)$
$\quad + n^2(15p^4+135p^3+787p^2+1827p+1620)+n(p^5+15p^4+155p^3+701p^2$
$\quad +1620p+1348)$. (2.9)

We note that for $k>o$ these may be obtained more directly by inverting the relationships given by James (1964) to write $s_k = tr(A^k)$ as a linear combination of the zonal polynomials, $Z_\kappa(A)$, with κ a partition $(k_1 \ldots k_p)$ of k and the coefficient of $Z_\kappa(A)$ depending only on κ. Then, from James' formulae (21) and (24) [due to Constantine (1963)], for $A \sim W_p(n,I)$

$$E[Z_\kappa(A)] = 2^{2k} \left(\tfrac{n}{2}\right)_\kappa \left(\tfrac{p}{2}\right)_\kappa \tag{2.10}$$

where

$$(a)_\kappa = \prod_{i=1}^{p} (a-\tfrac{1}{2}(i-1))_{k_i}; \quad (b)_{k_i} = \prod_{j=1}^{k_i} (b+j-1). \tag{2.11}$$

Since $E[Z_\kappa(A)]$ is a symmetric function of n and p, we see that for $k>o$, pc_k will be a symmetric function of n and p - a result not apparant from the above expressions for c_k. Thus e.g.

$pc_1 = np$

$pc_2 = np(n+p+1)$

$pc_3 = np(n^2+3np+p^2+3n+3p+4)$ (2.12)

Since $n\Sigma_o^{-\frac{1}{2}}S\Sigma_o^{-\frac{1}{2}} \sim W_p(n,I)$ it follows that

$$E[\mathrm{tr}(\Sigma_o^{-\frac{1}{2}}S\Sigma_o^{-\frac{1}{2}})^k] = E[\sum_{i=1}^{p} \phi_i^k] = c_k p n^{-k}, \tag{2.13}$$

so that, for $k>0$,

$$\begin{aligned} e_k &= E[\mathrm{tr}(\Sigma_o^{-\frac{1}{2}}S\Sigma_o^{-\frac{1}{2}}-I)^k] \\ &= E[\sum_{i=1}^{p} (\phi_i-1)^k] \\ &= p\sum_{i=0}^{k} (-1)^{k-i}\binom{k}{i}\frac{c_i}{n^i} \end{aligned} \tag{2.14}$$

We thus obtain the following values for e_k:

$$\begin{aligned} e_1 &= 0 \\ e_2 &= p(p+1)/n \\ e_3 &= p(p^2+3p+4)/n^2 \\ e_4 &= p(2p^2+5p+5)/n^2+p(p^3+6p^2+21p+20)/n^3 \\ e_5 &= p(5p^3+25p^2+70p+60)/n^3+p(p^4+10p^3+65p^2+160p+148)/n^4 \\ e_6 &= p(5p^3+22p^2+52p+41)/n^3+p(9p^4+65p^3+397p^2+867p+732)/n^4 \\ &\quad + p(p^5+15p^4+155p^3+701p^2+1620p+1348)/n^5. \end{aligned} \tag{2.15}$$

Using the expressions for c_k we can calculate exact expressions for the expectations of some of the particular estimation functions which we have considered. These are presented in Table 1. In all cases the expectations may be calculated to $O(n^{-2})$ by

$$E[nF(\Sigma_o,S)] = n\sum_{j=2}^{6} b_k e_k + o(n^{-2}). \tag{2.16}$$

Writing

$$E[nF(\Sigma_o;S)] = \frac{p(p+1)}{2}\left(1+\frac{k_1}{n} + \frac{k_2}{n^2} + o(n^{-2})\right) \tag{2.17}$$

Table 2 presents the values of the b_k and the resulting

k_1 and k_2 for each of the functions considered. Table 3 then presents the corresponding factors ρ, determined to satisfy (2.1), together with values of k_2' where

$$E[n\rho F(\Sigma_o;S)] = \frac{p(p+1)}{2}\left(1+\frac{k_2'}{n^2} + o(n^{-2})\right) \tag{2.18}$$

TABLE 1

Exact expectations for some estimation functions.

FUNCTION	$E[nF(\Sigma_o;S)]$
TGLS	$\frac{p(p+1)}{2}$
DIV	$\frac{p(p+1)}{2} \cdot \frac{n}{n-(p+1)}$
GLS	$\frac{p(p+1)}{2} \cdot \frac{n^3+n^2(p^2+2p+3)/(p+1)-np(p+3)}{(n-p)(n-p-1)(n-p-3)}$

TABLE 2

Expectations of some estimation functions to $O(n^{-2})$

FN.	b_3	b_4	b_5	b_6	k_1	k_2
TGLS	0	0	0	0	0	0
ML	$-\frac{1}{3}$	$\frac{1}{4}$	$-\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{6}(2p+1-\frac{2}{p+1})$	$\frac{1}{6}(p^2+p-2)$
GD	$-\frac{1}{2}$	$\frac{11}{24}$	$-\frac{5}{12}$	$\frac{137}{360}$	$\frac{1}{12}(10p+9-\frac{2}{p+1})$	$\frac{1}{180}(100p^2+154p-65-\frac{18}{p+1})$
DIV	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$p+1$	p^2+2p+1
GLS	-1	$\frac{3}{2}$	-2	$\frac{5}{2}$	$4p+5+\frac{2}{p+1}$	$8p^2+20p+23+\frac{2}{p+1}$
GLSE	$-\frac{3}{2}$	$\frac{13}{4}$	$-\frac{73}{12}$	$\frac{501}{48}$	$\frac{1}{2}(20p+27+\frac{14}{p+1})$	$\frac{1}{24}(1201p^2+3457p+5431+\frac{710}{p+1})$

TABLE 3

Correction factors and corresponding $O(n^{-2})$ terms of expectation

FUNCTION	ρ	k_2'
TGLS	1	0
ML	$1 - \frac{2p^2+3p-1}{6n(p+1)}$	$\frac{1}{36}\left(2p^2 + 2p - 5 - \frac{4p+8}{(p+1)^2}\right)$
GD	$1 - \frac{10p^2+19p+7}{12n(p+1)}$	$-\frac{1}{720}\left(100p^2+284p+465+\frac{146p+112}{(p+1)^2}\right)$
DIV	$1 - \frac{p+1}{n}$	0
GLS	$\frac{(n-p)(n-p-1)(n-p-3)}{n^3+n^2(p^2+2p+3)/(p+1)-np(p+3)}$	0
GLSE	$\left(1 + \frac{20p^2+47p+41}{2n(p+1)}\right)^{-1}$	$\frac{1}{24}\left(1201p^2+3457p+5431 + \frac{710}{p+1}\right)$

We note that Korin (1968) obtained the ML results presented here using an approach closely paralleling that of Box (1949). The ML factor had been derived earlier by Bartlett (1954).

Considering the size of the terms involved in Tables 2 and 3, it seems that for GLS without a factor the χ^2 approximation to the distribution is going to be poor for anything but very large n, while for GLSE this is likely to be true even when the factor is used. Accordingly, we drop GLSE from further consideration. We note that, using the expectations as a guide, application of the χ^2 approximation without a factor is likely to lead, for all the estimation functions considered except TGLS, to too frequent rejection of the null hypothesis. We will see that this suggested behaviour is borne out in the numerical experiments described later.

3.2.2 Asymptotic distribution of the test statistic.

The asymptotic distribution of $nF_1(\Sigma_o;S)$ has been determined to $O(n^{-1})$ by Nagao (1973) by writing the test statistic in the form

$$T_1 = nF_1(\Sigma_o;S)$$

$$= \text{tr } Y^2 + \left(\frac{2}{n}\right)^{\frac{1}{2}} \text{tr } Y^3 + \frac{7}{6n} \text{tr } Y^4 + O_p(n^{-\frac{3}{2}}) \qquad (2.19)$$

where $Y = (n/2)^{\frac{1}{2}} \log (\Sigma_o^{-\frac{1}{2}} S \Sigma_o^{-\frac{1}{2}})$, with log S being defined as the real symmetric matrix B such that

$$S = e^B = I + B + \frac{B^2}{2!} + \frac{B^3}{3!} + \ldots \qquad (2.20)$$

From the asymptotic distribution of Y, he then shows, via the characteristic function of T_1, that

$$\Pr(T_1<x) = P_d + \frac{1}{n}\{f_3 P_{d+6} + f_2 P_{d+4}$$

$$+ f_1 P_{d+2} + f_o P_d\} + O(n^{-2}) \qquad (2.21)$$

here $d = \frac{1}{2}p(p+1)$, P_d is defined by (1.6), and the f_i are polynomials in p.

This approach may be extended to general estimation functions having an expansion of the form (2.3).

Since

$$\frac{n}{2}\sum_{i=1}^{p}(\phi_i-1)^2 = \text{tr } Y^2 + \left(\frac{2}{n}\right)^{\frac{1}{2}} \text{tr } Y^3 + \frac{7}{6n} \text{tr } Y^4 + O_p(n^{-\frac{3}{2}}) \qquad (2.22)$$

$$\frac{n}{2}\sum_{i=1}^{p}(\phi_i-1)^3 = \left(\frac{2}{n}\right)^{\frac{1}{2}} \text{tr } Y^3 + \frac{3}{n} \text{tr } Y^4 + O_p(n^{-\frac{3}{2}}) \qquad (2.23)$$

$$\frac{n}{2}\sum_{i=1}^{p}(\phi_i-1)^4 = \frac{2}{n} \text{tr } Y^4 + O_p(n^{-\frac{3}{2}}) \qquad (2.24)$$

we have, for $F(\Sigma;S)$ with an expansion (2.3),

$$T = nF(\Sigma_o;S)$$

$$= \operatorname{tr} Y^2 + (1+2b_3)\left(\frac{2}{n}\right)^{\frac{1}{2}}\operatorname{tr} Y^3 + \left(\frac{7}{6} + 6b_3 + 4b_4\right)\frac{1}{n}\operatorname{tr} Y^4$$

$$+ O_p(n^{-\frac{3}{2}}), \qquad (2.25)$$

whence, following through Nagao's working, $Pr(T<x)$ has the form (2.21) with

$$f_3 = \frac{p}{12}(4p^2+9p+7)(1+3b_3)^2$$

$$f_2 = -\frac{p}{8}[2(4p^2+9p+7)(1+2b_3)(1+3b_3)$$

$$- (2p^2+5p+5)(1+6b_3+4b_3)]$$

$$f_1 = \frac{p}{48}[9(4p^2+9p+7)(1+2b_3)^2$$

$$- (2p^2+5p+5)(7+36b_3+24b_4)$$

$$+ 2(p^2+p-2)]$$

$$f_0 = -\frac{p}{24}(2p^2+3p-1) \qquad (2.26)$$

The values of these coefficients for the estimation functions previously considered are given in Table 4.

Further, if $X \sim \chi^2(d)$ and ρ is some constant of the form

$$\rho = 1 - \frac{k_1}{n} + O(n^{-2}) \qquad (2.27)$$

then it is easily shown via the characteristic function that

$$Pr(\rho X < x) = \left(1 - \frac{k_1 d}{2n}\right)P_d + \frac{k_1 d}{2n} P_{d+2} + O(n^{-2}) \qquad (2.28)$$

Thus for ρ of the form (2.27)

$$Pr(\rho T < x) = P_d + \frac{1}{n}\{f_3 P_{d+6} + f_2 P_{d+4}$$

$$+ f_1' P_{d+2} + f_o' P_d\} + O(n^{-2}) \tag{2.29}$$

where

$$f_1' = f_1 - \tfrac{1}{2}dk_1$$

$$f_o' = f_o + \tfrac{1}{2}dk_1 \tag{2.30}$$

The values for the coefficients f_o' and f_1' for the estimation functions previously considered are given in Table 5, where ρ has been taken with the value given in Table 3. Probabilities calculated using (2.29) would of course agree to $O(n^{-1})$ with $Pr(T < x/\rho)$ from (2.21).

TABLE 4

The coefficients f_i for some estimation functions

FUNCTION	f_3	f_2	f_1
TGLS	$\frac{p}{12}(4p^2+9p+7)$	$-\frac{p}{8}(6p^2+13p+9)$	$\frac{p}{2}(p+1)^2$
ML	0	0	$\frac{p}{24}(2p^2+3p-1)$
GD	$\frac{p}{48}(4p^2+9p+7)$	$-\frac{p}{48}(2p^2+5p+5)$	$\frac{p}{24}(p^2+p-2)$
DIV	$\frac{p}{48}(4p^2+9p+7)$	0	$-\frac{p}{16}(p+3)$
GLS	$\frac{p}{3}(4p^2+9p+7)$	$-\frac{p}{8}(14p^2+31p+23)$	$\frac{p}{2}(p+1)^2$

TABLE 5

The coefficients f_1' f_o' for some estimation functions

FUNCTION	f_1'	f_o'
TGLS	$\frac{p}{2}(p^2+2p+1)$	$-\frac{p}{24}(2p^2+3p-1)$
ML	0	0
GD	$-\frac{p}{48}(8p^2+17p+11)$	$\frac{p}{48}(6p^2+13p+9)$
DIV	$-\frac{p}{16}(4p^2+9p+7)$	$\frac{p}{24}(4p^2+9p+7)$
GLS	$-\frac{p}{4}(2p^2+5p+5)$	$\frac{p}{24}(22p^2+51p+43)$

We note that, from (2.29), with a factor ρ of the form (2.27), $\Pr(\rho T < x)$ can agree with P_d to $O(n^{-1})$ only if $f_3 = f_2 = 0$, which by (2.26) is true only if $b_3 = -\frac{1}{3}$, and $b_4 = \frac{1}{4}$, and thus only if $nF(\Sigma_o;S)$ agrees with $nF_2(\Sigma_o;S)$ to $O_p(n^{-1})$.

3.2.3 Numerical comparisons

We have in the previous subsections given four methods by which the distribution of $T = nF(\Sigma_o;S)$ may be approximated when $nS \sim W_p(n,\Sigma_o)$. These, corresponding to (1.1), (2.21) and (2.29), may be summarised as

$$\text{(a)} \quad \Pr(T < x) \doteqdot P_d \tag{2.31}$$

$$\text{(b)} \quad \Pr(T < x) \doteqdot P_d + \frac{1}{n}(f_3P_{d+6}+f_2P_{d+4}+f_1P_{d+2}+f_oP_d) \tag{2.32}$$

$$\text{(c)} \quad \Pr(\rho T < x) \doteqdot P_d \tag{2.33}$$

$$\text{(d)} \quad \Pr(\rho T < x) \doteqdot P_d + \frac{1}{n}(f_3P_{d+6}+f_2P_{d+4}+f_1'P_{d+2}+f_o'P_d) \tag{2.34}$$

where ρ is an appropriate factor determined in the manner of (2.17) and (2.18). We will now describe a numerical experiment which was conducted in order to compare the effectiveness of these approximations.

The basic method was to generate a large number of sample matrices from a $W_p(n,\Sigma_o)$ population and calculate the corresponding values of T, thus obtaining a numerical approximation to the distribution of T with which to compare the theoretical approximations given above. The population matrix Σ_o was taken equal to I, without loss of generality. As each of (2.31) to (2.34) could be expected to perform well for large n, a moderate value of n (n=40) was chosen, corresponding to p=6.

Four hundred independent sample matrices from $W_p(n,I)$ were drawn, their eigen values determined, and thence the values of T calculated for each of the estimation functions previously considered. The samples from $W_p(n,I)$ were generated via the triangular decomposition, the necessary χ and N(0,1) random deviates being determined by transformation of uniform random deviates. Full details of the procedure are given e.g. by Browne (1968). The uniform random deviates required were obtained using the CDC library function RANF. The values of T were then ordered to give the empirical, or observed, c.d.f.. Corresponding values of the approximating c.d.f.'s given by (2.31) and (2.32) were also calculated, the necessary values of P_d being determined using a subroutine kindly made available by Dr. W.N. Venables.

The empirical c.d.f.'s, together with the theoretical approximations, were plotted using an incremental plotter, and the results are shown in figures 1, 2A - 5A. Similarly, with the values of ρ given in table 2, the empirical c.d.f.'s of ρT (for the same set of 400 samples) and their approximating c.d.f.'s were found, and plotted as shown in figures 2B - 5B. Note that for TGLS, ρ=1 so (2.33) and (2.34) are equivalent to (2.31) and (2.32), while for ML (2.33) and (2.34) are the same.

For T_1 to T_4 the figures suggest that, for the particular n and p used, the approximation to the c.d.f. of T given by (2.31) is not particularly good, but is improved by the inclusion of the $O(n^{-1})$ term as in (2.32). For T_2 to T_4 the use of the appropriate factor with (2.33) also gives considerable improvement over (2.31), while for T_3 and T_4 this is further improved by the inclusion of the $O(n^{-1})$ terms as in (2.34). For each of T_1 to T_4 there is a close resemblance between the observed c.d.f. and the best of the approximating curves. On the other hand, for T_5 (GLS), even though (2.33) represents a considerable improvement over (2.31), the approximation is still poor, while n is apparently too small for the inclusion of the $O(n^{-1})$ term of the asymptotic series to give any improvement.

In conclusion then, without further information on its distribution T_5 is likely to prove an unsatisfactory test statistic for the hypothesis $\Sigma=\Sigma_0$ for anything but very large n, although any of T_1 to T_4 with the appropriate approximating distribution should provide a reasonably accurate test for moderate n.

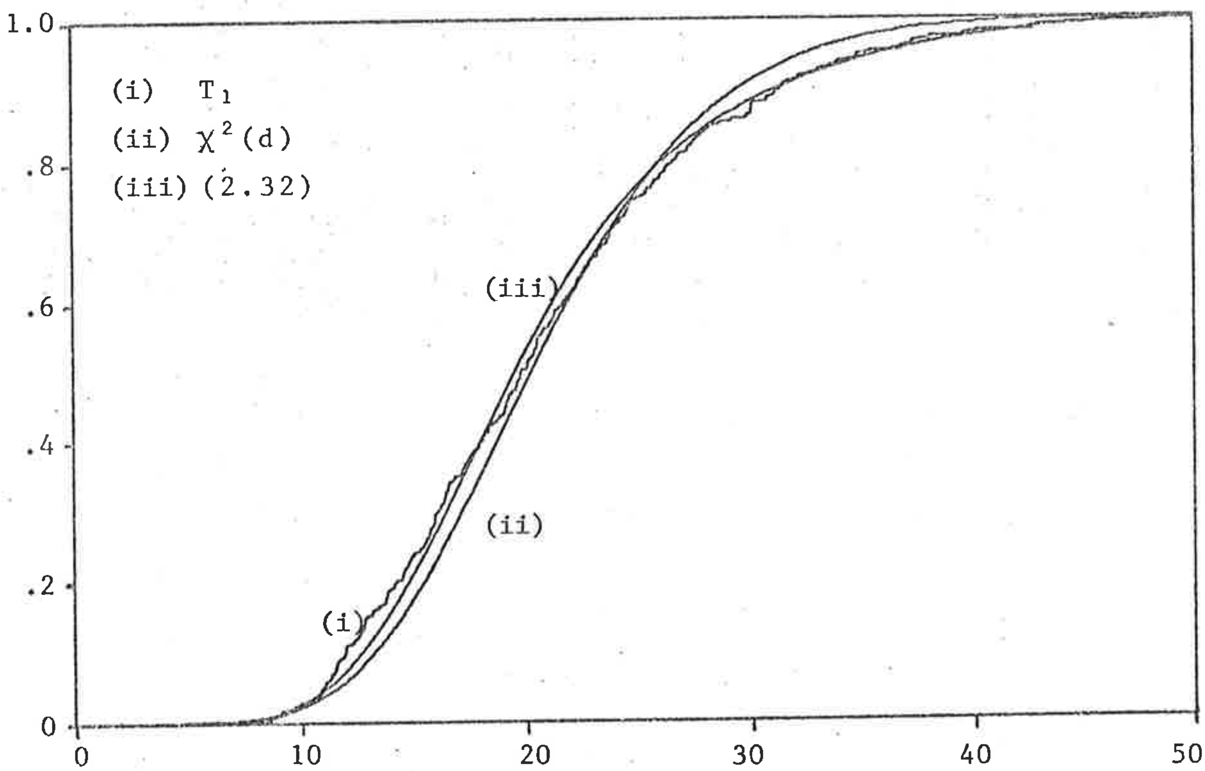

FIGURE 1 - Observed and approximating c.d.f. - T_1.

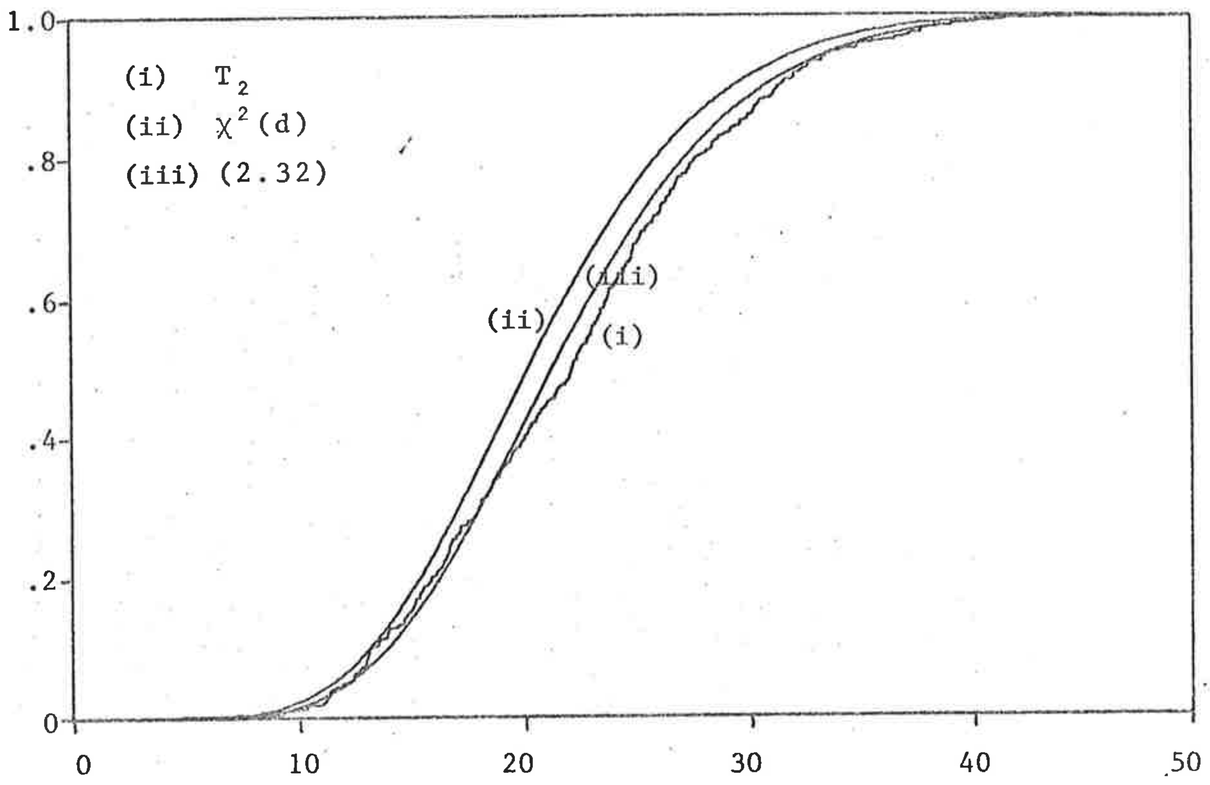

FIGURE 2A - Observed and approximating c.d.f. - T_2

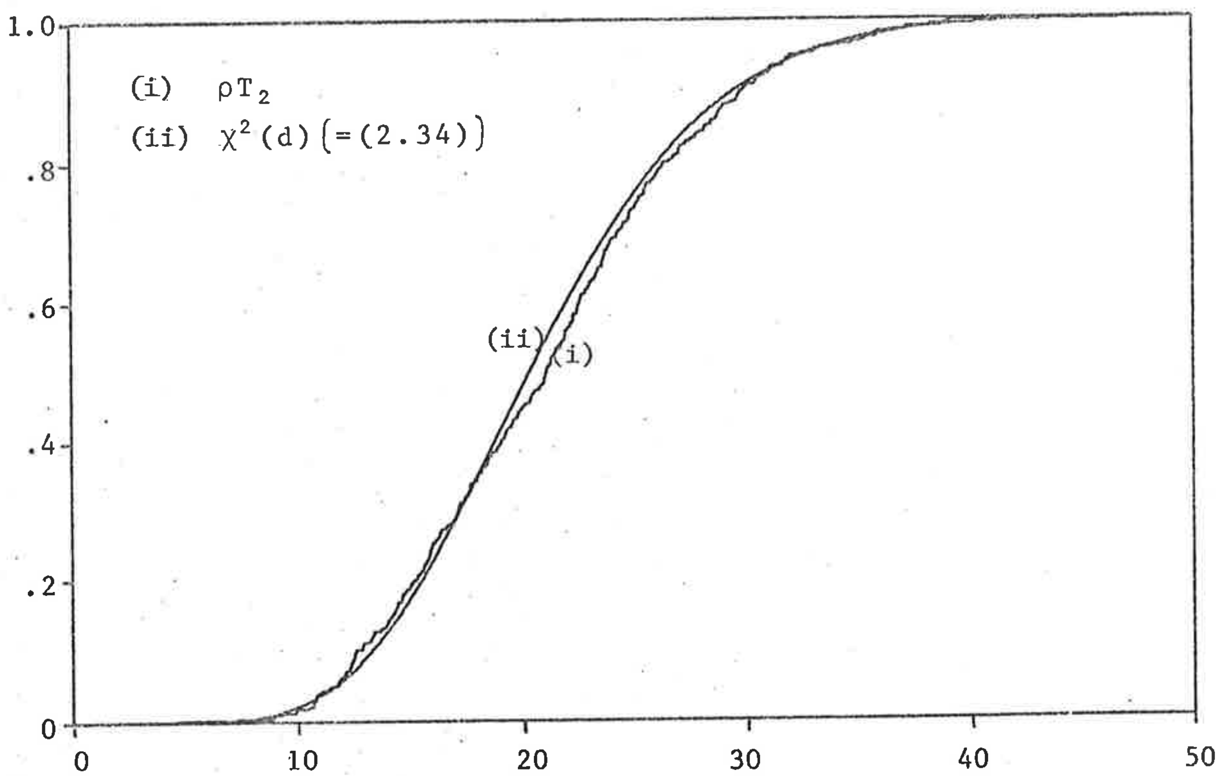

FIGURE 2B - Observed and approximating c.d.f. - ρT_2.

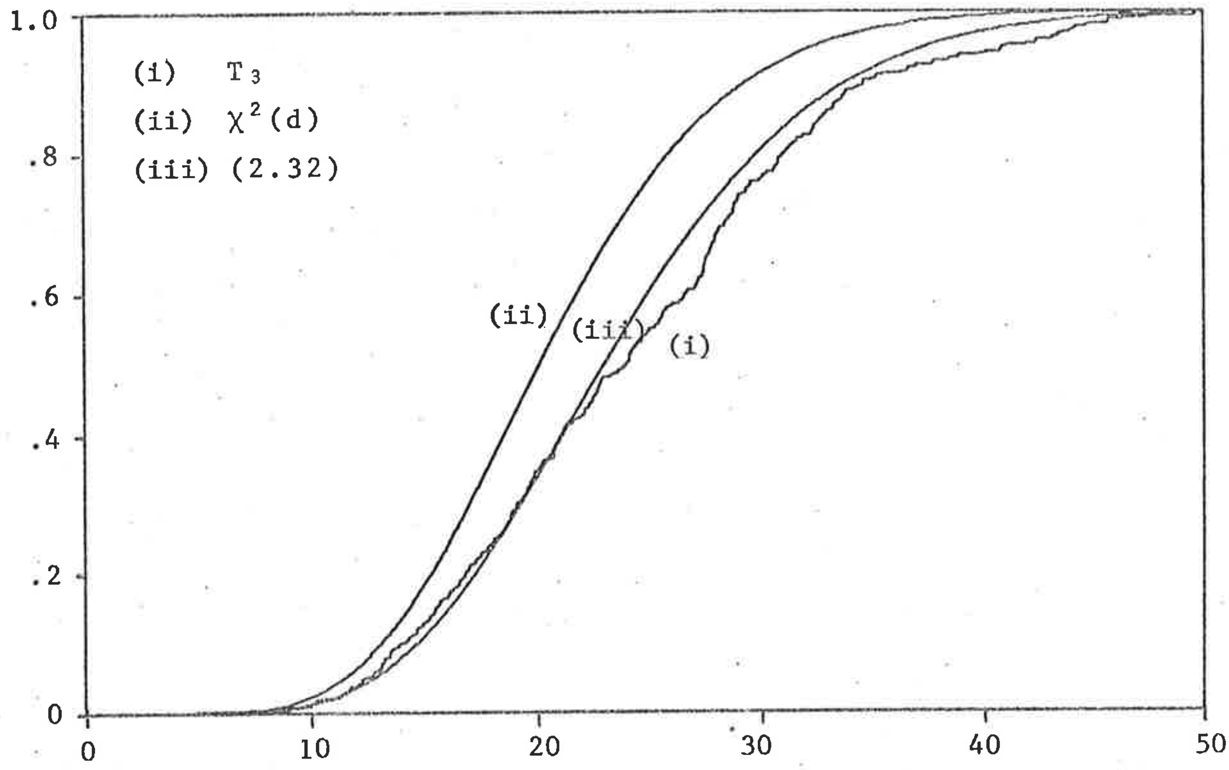

FIGURE 3A - Observed and approximating c.d.f. - T_3.

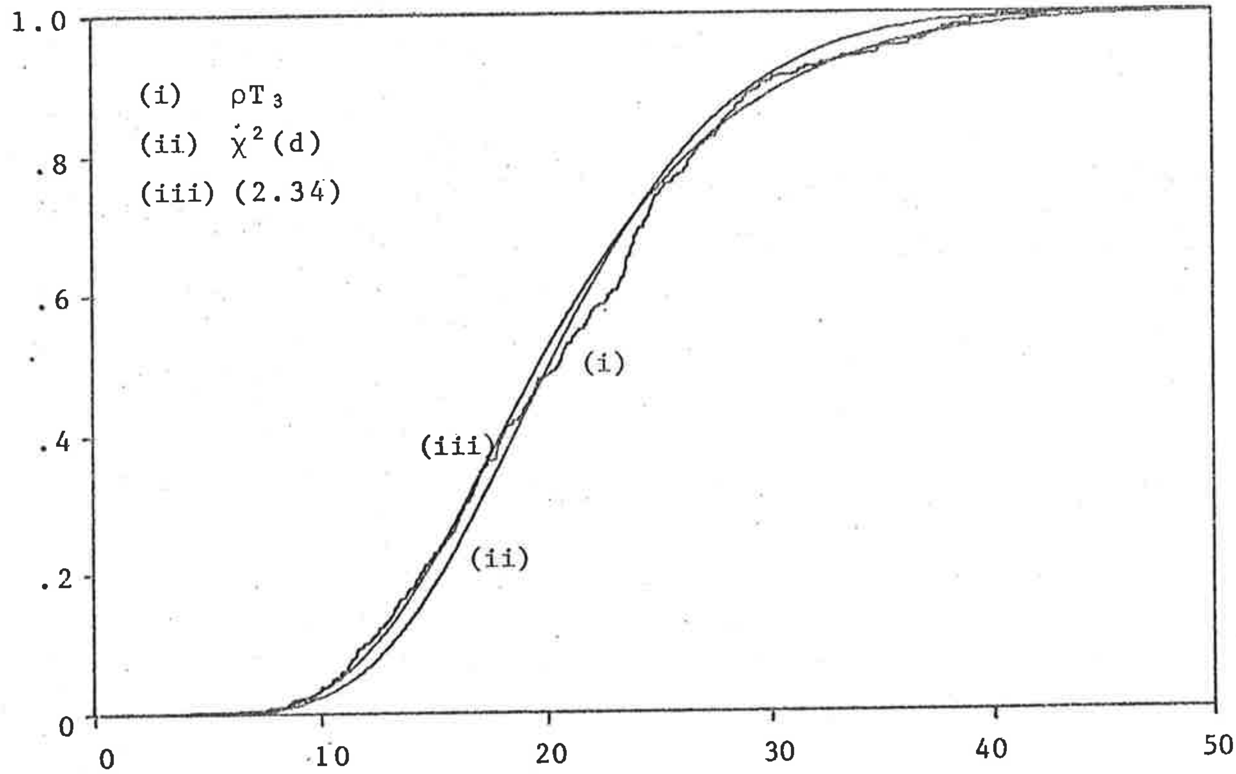

FIGURE 3B - Observed and approximating c.d.f. - ρT_3

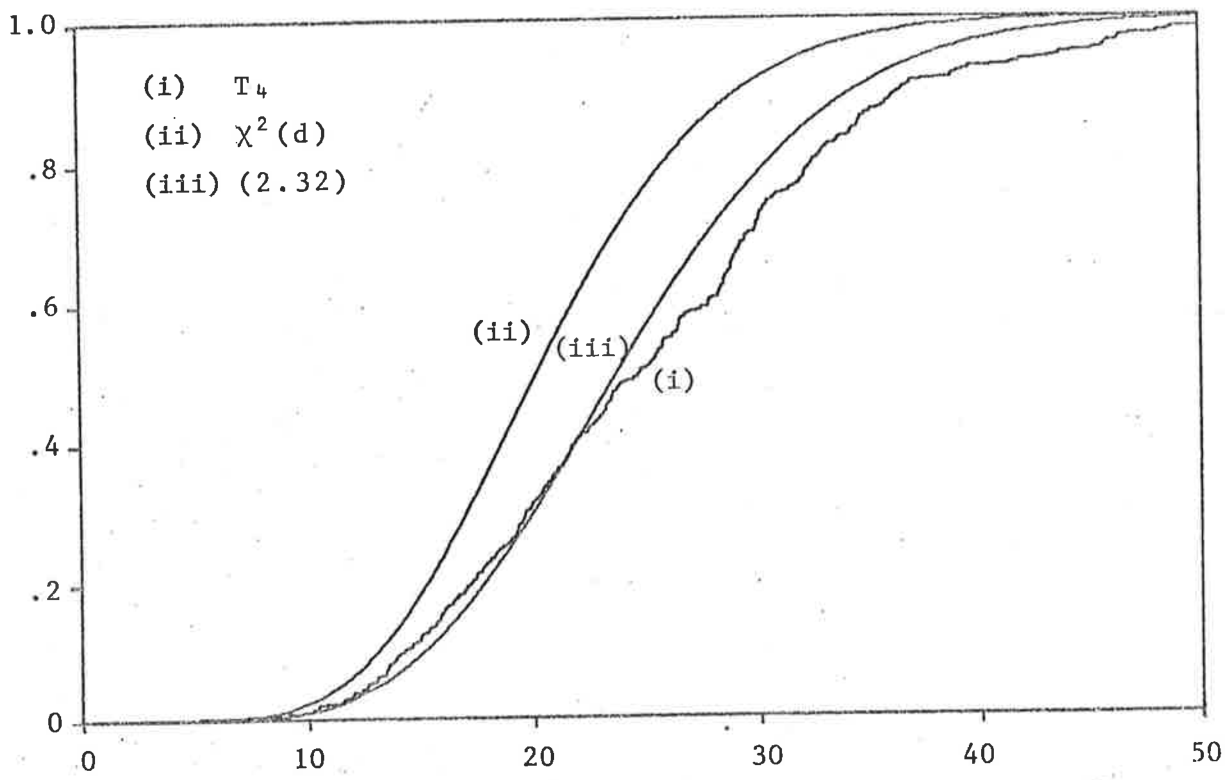

FIGURE 4A - Observed and approximating c.d.f. - T_4.

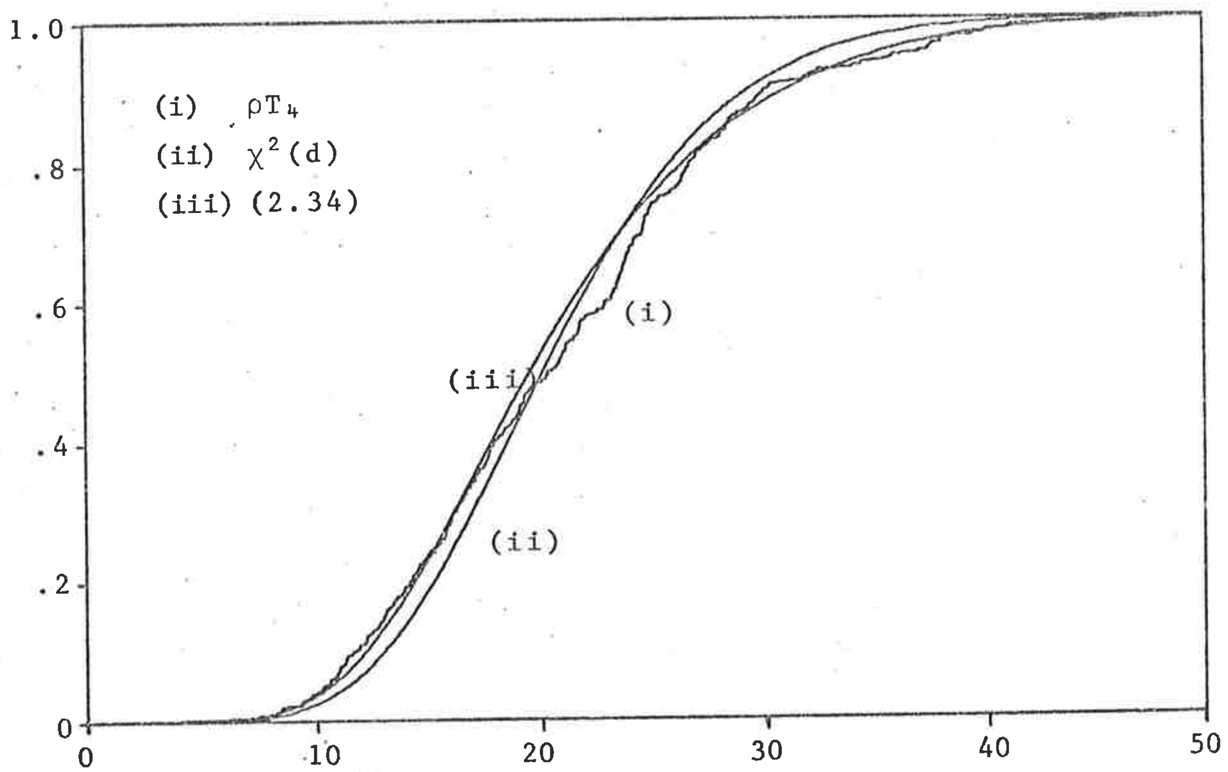

FIGURE 4B - Observed and approximating c.d.f. - ρT_4

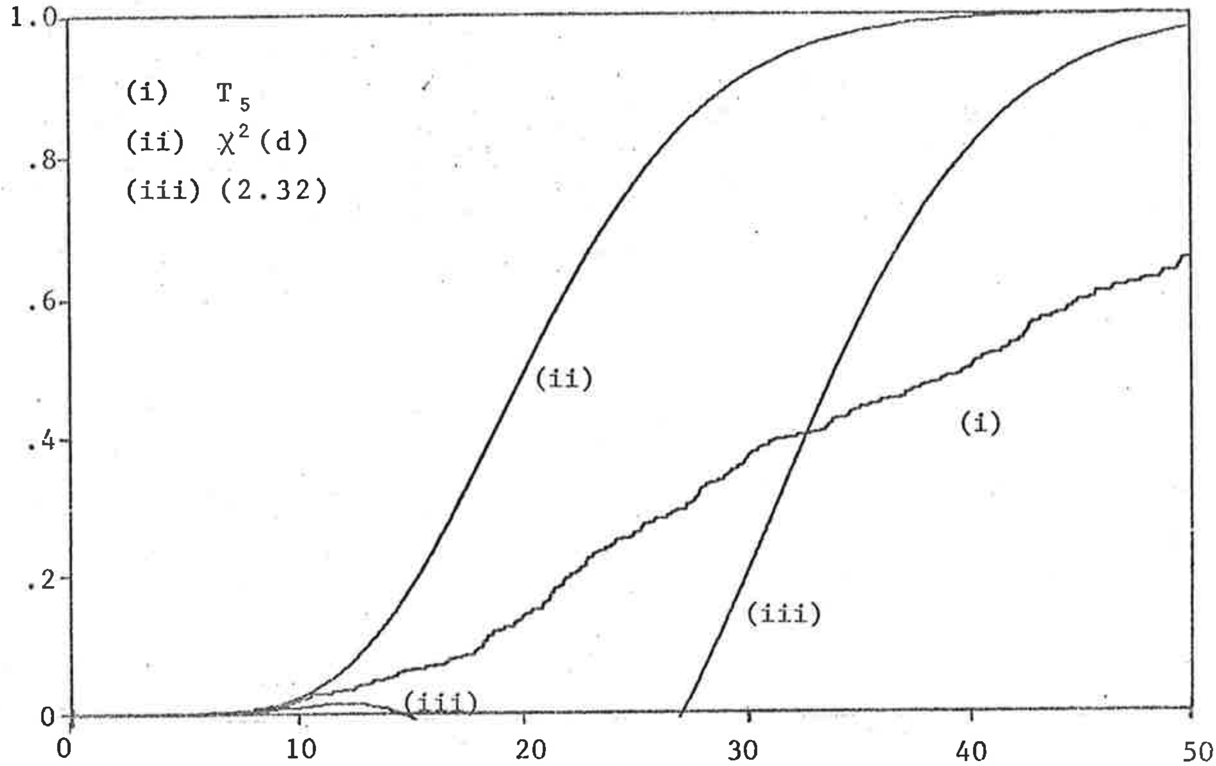

FIGURE 5A - Observed and approximating c.d.f. - T_5.

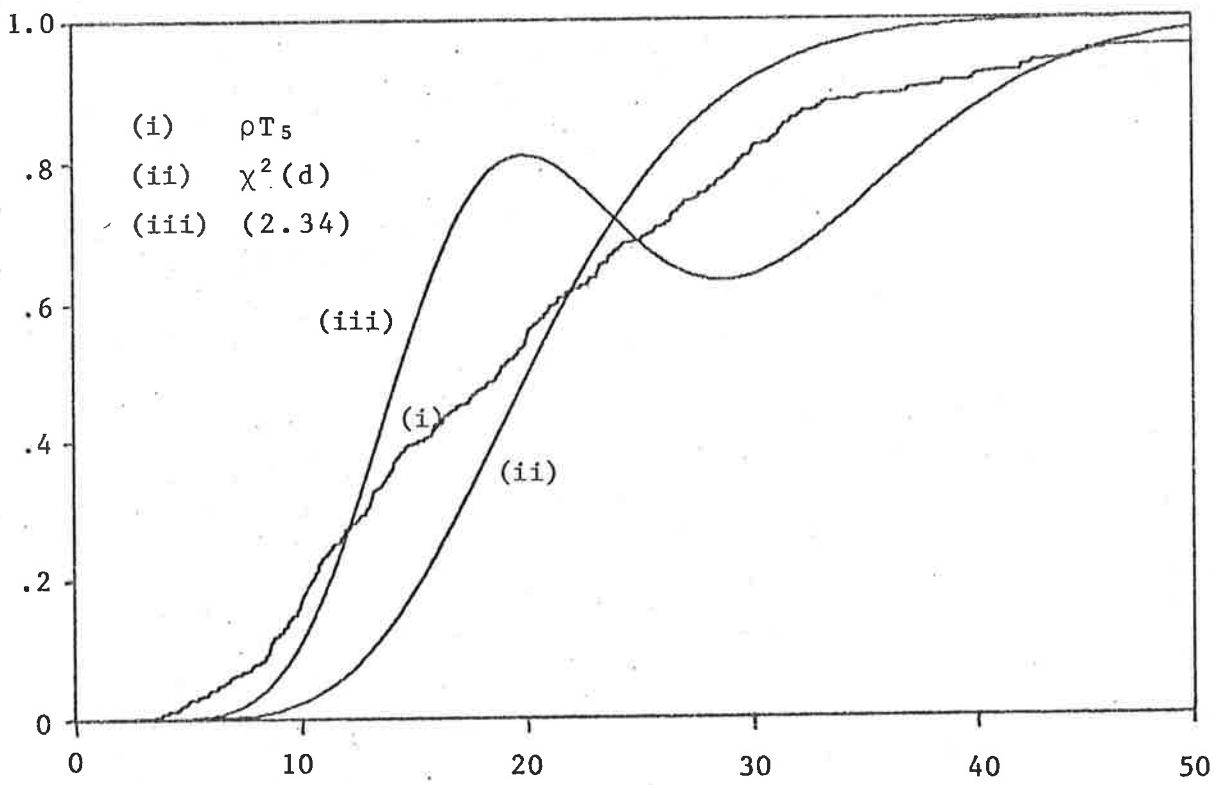

FIGURE 5B - Observed and approximating c.d.f. - ρT_5

3.3 The test statistic for some special models

As in chapter 2, for the model $\Sigma = \Sigma(\underset{\sim}{\gamma})$, $\underset{\sim}{\gamma}\in\Gamma$ we write $A = \Sigma(\Gamma)$. In this section we shall show that where the model has the property

$$\Sigma(\underset{\sim}{\gamma})\in A \Rightarrow c\Sigma(\underset{\sim}{\gamma})\in A, \ c>0 \tag{3.1}$$

the form of the test statistic $nF(\hat{\Sigma};S)$ simplifies for many $F\in\mathcal{F}$, with possible advantages in both computation and distribution approximation. The result we obtain for the particular case of maximum likelihood has been previously shown by Bock and Bargmann (1966). We note that many of the general structures mentioned in chapter 1 have the property (3.1), e.g. linear structures when the range of the coefficients is unrestricted, the unrestricted factor model, the direct product structure, and Jöreskog's general model of variance structures, when all parameters are free and also in some restricted cases. Similarly, models which essentially apply to the correlation matrix and leave the variances unconstrained will have the property (3.1).

Denoting, as usual, the eigen values of $S^{-\frac{1}{2}}\Sigma S^{-\frac{1}{2}}$ by $\theta_1 \leqslant \ldots \leqslant \theta_p$, and those of $S^{-\frac{1}{2}}\hat{\Sigma}S^{-\frac{1}{2}}$ by $\hat{\theta}_1 \leqslant \ldots \leqslant \hat{\theta}_p$, we have the following.

<u>Theorem 1</u> For any $F\in\mathcal{F}$ and model $\Sigma(\underset{\sim}{\gamma})$ with the property (3.1),

$$\sum_{i=1}^{p} f'(\hat{\theta}_i)\hat{\theta}_i = 0. \tag{3.2}$$

<u>Proof</u>. For any fixed Σ and $c>0$, the eigen values of $S^{-\frac{1}{2}}(c\Sigma)S^{-\frac{1}{2}}$ are $c\theta_1,\ldots,c\theta_p$ so that

$$F(c\Sigma;S) = \sum_{i=1}^{p} f(c\theta_i). \tag{3.3}$$

Differentiating with respect to c, at a minimum in the $c\Sigma$ direction

$$\sum_{i=1}^{p} f'(c\theta_i)\theta_i = 0. \tag{3.4}$$

Taking $\Sigma = \hat{\Sigma}$, the minimum occurs at c=1, and so (3.2) holds.

Corollary For the model $\Sigma = a\Sigma_0$ where Σ_0 is specified, the estimate of a will be given as a solution of

$$\sum_{i=1}^{p} f'(a\theta_i^{(o)})\theta_i^{(o)} = 0, \tag{3.5}$$

where $\theta_i^{(o)}$ are the eigen values of $S^{-\frac{1}{2}}\Sigma_0 S^{-\frac{1}{2}}$.

It is the expression (3.2) which frequently leads to simplification in the form of $F(\hat{\Sigma};S)$, as can be seen in Table 6 when the results of applying (3.2) to the particular estimation functions considered in section 3.2 are presented. The table also gives the value of a from (3.5), where for simplicity the (o)'s have been dropped. We note the similarity between this column and the estimates for hypothesis of equal roots (section 2.7). This is to be expected, since the hypothesis $\Sigma = a\Sigma_0$ is equivalent to the hypothesis that all the roots of $\Sigma_0^{-\frac{1}{2}}\ \Sigma\Sigma_0^{-\frac{1}{2}}$ are equal.

TABLE 6

Reduced form of the test statistic for some estimation functions.

FN	$F(\Sigma;S)$	(3.2)	$F(\hat{\Sigma};S)$	a, from (3.5)
TGLS	$\frac{1}{2}\sum_{i=1}^{p}\left(\frac{1}{\theta_i}-1\right)^2$ $=\frac{1}{2}tr[\Sigma^{-1}(\Sigma-S)]^2$	$\sum_{i=1}^{p}\frac{1}{\hat{\theta}_i^2}=\sum_{i=1}^{p}\frac{1}{\hat{\theta}_i}$ $tr(S\hat{\Sigma}^{-1})^2=tr(S\hat{\Sigma}^{-1})$	$\frac{1}{2}\sum_{i=1}^{p}\left(-\frac{1}{\hat{\theta}_i}+1\right)$ $=\frac{1}{2}(p-trS\hat{\Sigma}^{-1})$	$\left(\sum_{i=1}^{p}\frac{1}{\theta_i^2}\right)/\left(\sum_{i=1}^{p}\frac{1}{\theta_i}\right)$ $=tr(S\Sigma^{-1})^2/tr(S\Sigma^{-1})$
ML	$\sum_{i=1}^{p}\left(\frac{1}{\theta_i}+\log\theta_i-1\right)$ $=trS\Sigma^{-1}+\log\lvert\Sigma\rvert-\log\lvert S\rvert-p$	$\sum_{i=1}^{p}\frac{1}{\hat{\theta}_i}=p$ $tr(S\hat{\Sigma}^{-1})=p$	$\sum_{i=1}^{p}\log\hat{\theta}_i$ $=\log\lvert\hat{\Sigma}\rvert-\log\lvert S\rvert$	$\frac{1}{p}\sum_{i=1}^{p}\frac{1}{\theta_i}$ $=\frac{1}{p}tr(S\Sigma^{-1})$
GD	$\frac{1}{2}\sum_{i=1}^{p}(\log\theta_i)^2$	$\sum_{i=1}^{p}\log\hat{\theta}_i=0$ $\log\lvert\hat{\Sigma}\rvert=\log\lvert S\rvert$	$\frac{1}{2}\sum_{i=1}^{p}(\log\hat{\theta}_i)^2$	$\left(\prod_{i=1}^{p}\frac{1}{\theta_i}\right)^{1/p}$ $=\lvert S\Sigma^{-1}\rvert^{1/p}$
DIV	$\frac{1}{2}\sum_{i=1}^{p}\left(\theta_i+\frac{1}{\theta_i}-2\right)$ $=\frac{1}{2}(trS\Sigma^{-1}+trS^{-1}\Sigma-2p)$	$\sum_{i=1}^{p}\frac{1}{\hat{\theta}_i}=\sum_{i=1}^{p}\hat{\theta}_i$ $trS\hat{\Sigma}^{-1}=trS^{-1}\hat{\Sigma}$	$\sum_{i=1}^{p}(\hat{\theta}_i-1)=\sum_{i=1}^{p}\left(\frac{1}{\hat{\theta}_i}-1\right)$ $=tr(S^{-1}\hat{\Sigma})-p=tr(S\hat{\Sigma}^{-1})-p$	$\left(\left(\sum_{i=1}^{p}\frac{1}{\theta_i}\right)/\left(\sum_{i=1}^{p}\theta_i\right)\right)^{\frac{1}{2}}$ $=[tr(S\Sigma^{-1})/tr(S^{-1}\Sigma)]^{\frac{1}{2}}$
GLS	$\frac{1}{2}\sum_{i=1}^{p}(\theta_i-1)^2$ $=\frac{1}{2}tr[S^{-1}(\Sigma-S)]^2$	$\sum_{i=1}^{p}\hat{\theta}_i^2=\sum_{i=1}^{p}\hat{\theta}_i$ $tr(S^{-1}\hat{\Sigma})^2=tr(S^{-1}\hat{\Sigma})$	$\frac{1}{2}\sum_{i=1}^{p}(-\hat{\theta}_i+1)$ $=\frac{1}{2}(p-trS^{-1}\hat{\Sigma})$	$\left(\sum_{i=1}^{p}\theta_i\right)/\left(\sum_{i=1}^{p}\theta_i^2\right)$ $=tr(S^{-1}\Sigma)/tr(S^{-1}\Sigma)^2$

3.4 Approximation to the distribution of the test statistic for general models

For some models of the general form $\Sigma = \Sigma(\underset{\sim}{\gamma})$, $\underset{\sim}{\gamma}\in\Gamma$, it can be extremely difficult to obtain $E[nF(\hat{\Sigma};S)]$, and thence an appropriate factor ρ for the improvement of the distribution approximation (1.1). This can be especially true where no explicit formula for the estimates is available, and they are determined iteratively. However, since we still have

$$\rho nF(\hat{\Sigma};S) \doteq \chi^2(d) \tag{4.1}$$

for any ρ of the form

$$\rho = 1 - \frac{k_1}{n} + O(n^{-2}), \tag{4.2}$$

and experience suggests that for suitable ρ and moderate n (4.1) can give a much improved approximation over (1.1), we propose in this section forms for ρ which may prove useful for general structures. For any particular model the worth of the forms suggested would, of course, have to be carefully evaluated before routine application, e.g. in the manner of section 5.7 for the direct product model.

The following two factors for general structures are suggested by the known formulae for the test for specified Σ.

(i) The same factor, ρ_1, as is appropriate to the test for specified Σ might be used for general structures. This is equivalent to testing the hypothesis $\Sigma = \hat{\Sigma}$, but using degrees of freedom d instead of $\frac{1}{2}p(p+1)$.

(ii) The factor for the test $\Sigma = \Sigma_o$ is determined basically from the form

$$E[nF(\Sigma_o;S)] = \tfrac{1}{2}p(p+1)\left(1 + \frac{k_1(p)}{n} + O(n^{-2})\right) \qquad (4.3)$$

where $k_1(p)$ is some function of p. Since we know that

$$E[nF(\hat{\Sigma};S)] = d(1 + o(1)) \qquad (4.4)$$

we might assume that the o(1) term has the same form as that in (4.3) but with $\tfrac{1}{2}p(p+1)$ replaced by d, i.e. if

$$z = \tfrac{1}{2}(-1 + (1+8d)^{\frac{1}{2}}) \qquad (4.5)$$

we might assume that

$$E[nF(\hat{\Sigma};S)] = \tfrac{1}{2}z(z+1)\left(1 + \frac{k_1(z)}{n} + O(n^{-2})\right). \qquad (4.6)$$

This would then lead to a factor, ρ_2, which is obtained from ρ_1 by replacing p by z.

We note that both ρ_1 and ρ_2 reduce to the appropriate factor for the test for specified Σ, and for all the particular estimation functions considered above it can be shown that $1 \geqslant \rho_2 \geqslant \rho_1$.

For the important special case of maximum likelihood with models having the property (3.1), two further forms of ρ are suggested by the simplified expression for the test statistic, which from table 6, may be written as

$$nF(\hat{\Sigma};S) = n \log |\hat{\Sigma}\Sigma_o^{-1}| - n \log |S\Sigma_o^{-1}|, \qquad (4.7)$$

where Σ_o is the population variance matrix. Now, since from (2.13) $E[tr(S\Sigma_o^{-1})] = p$, from table 2

$$E[n \log |S\Sigma_o^{-1}|] = -\left(\frac{p(p+1)}{2} + \frac{p(2p^2+3p-1)}{12n} + O(n^{-2})\right) \qquad (4.8)$$

and so, by (4.4)

$$E[n \log |\hat{\Sigma}\Sigma_o^{-1}|] = -q + o(1). \tag{4.9}$$

When the o(1) term in (4.9) is unknown, we may reasonably approximate it either by zero, or, in the manner of (ii) above, by replacing p in the $O(n^{-1})$ term of (4.8) by

$$y = \tfrac{1}{2}(-1 + (1+8q)^{\frac{1}{2}}). \tag{4.10}$$

Based on these two approximations, we derive the two further approximate factors ρ_3 and ρ_4.

Explicitly then we now have the following four factors proposed for maximum likelihood and models with the property (3.1).

$$\rho_1 = 1 - \frac{2p^2+3p-1}{6n(p+1)} = 1 - \frac{p(2p^2+3p-1)}{12n(\frac{1}{2}p(p+1))} \tag{4.11}$$

$$\rho_2 = 1 - \frac{2z^2+3z-1}{6n(z+1)} = 1 - \frac{z(2z^2+3z-1)}{12n\ d} \tag{4.12}$$

$$\rho_3 = 1 - \frac{p(2p^2+3p-1)}{12n\ d} \tag{4.13}$$

$$\rho_4 = 1 - \frac{1}{d}\left(\frac{p(2p^2+3p-1)}{12n} - \frac{y(2y^2+3y-1)}{12n}\right) \tag{4.14}$$

with y and z defined by (4.10) and (4.5).

We note that for all $p \geq 1$, $q \geq o$, and $n > p$ it can be shown $1 > \rho_2 \geq \rho_1 \geq \rho_4 \geq \rho_3$, with equalities occurring only if $d = \frac{1}{2}p(p+1)$, i.e. $q=o$.

In order to get some idea of the performance of these factors, we now compare them with the known theoretical factors for two common situations.

For the model $\Sigma = \text{diag}\ (\sigma_{ii})$, the ML estimates are $\hat{\sigma}_{ii} = s_{ii}$. Setting p=1 in (4.8) we find

$$E[n \log |s_{ii}\, \sigma_o^{ii}|] = -\left(1+1/(3n) + 0(n^{-2})\right) \qquad (4.15)$$

and so

$$E[nF(\hat{\Sigma};S)] = \frac{p(p-1)}{2}\left(1+\frac{2p+5}{6n} + 0(n^{-2})\right) \qquad (4.16)$$

whence the appropriate factor is given by

$$\rho = 1 - \frac{(2p+5)}{6n} \qquad (4.17)$$

[c.f. Anderson, 1958, p.239]. By comparison,

$$\rho_1 = \rho + \frac{2p+3}{3n(p+1)} \qquad (4.18)$$

$$\rho_3 = \rho - \frac{2}{3n(p-1)} \qquad (4.19)$$

$$\rho_4 = \rho + \frac{2(2y+3)}{3n(y+1)(y+2)} \qquad (4.20)$$

Although ρ_2 does not reduce simply, we certainly have $1>\rho_2>\rho_1>\rho_4>\rho>\rho_3$, with ρ_3 and ρ_4 being quite close to ρ.

For the sphericity model $\Sigma = \sigma^2 I$, the ML estimate is $\hat{\sigma}^2 = \frac{1}{p}\,tr(S)$, and so $np\hat{\sigma}^2/\sigma^2 \sim \chi^2(np)$ under the model. Thus by (4.15) and (4.8)

$$E[n \log |\hat{\Sigma}\, \Sigma_o^{-1}|] = -\left(1 + \frac{1}{3np} + 0(n^{-2})\right) \qquad (4.21)$$

$$E[nF(\hat{\Sigma};S)] = \left(\frac{p(p+1)}{2} - 1\right)\left(1 + \frac{2p^2+p+2}{6np} + 0(n^{-2})\right) \qquad (4.22)$$

and the appropriate factor is given by

$$\rho = 1 - \frac{1}{6n}\left(2p+1 + \frac{2}{p}\right) \qquad (4.23)$$

[c.f. Anderson, 1958, p.263]. By comparison

$$\rho_1 = \rho + \frac{6p+2}{6np(p+1)} \qquad (4.24)$$

$$\rho_3 = \rho + \frac{2p^2+p+2}{6np(p+2)(p+1)} \qquad (4.25)$$

$$\rho_4 = \rho + \frac{2p^2+5p+2}{6np(p+2)(p+1)} \qquad (4.26)$$

Again there is no simple expression for ρ_2. However, we do have $1>\rho_2>\rho_1>\rho_4>\rho_3>\rho$.

Thus in both situations use of any of the approximate factors should lead to a closer approximation to the distribution of the test statistic by χ^2, with ρ_3 and ρ_4 giving better approximations than ρ_1 and ρ_2. This is also in agreement with the results of the numerical experiment described in section 5.7 for the direct product model. In practice ρ_4 is probably to be preferred to ρ_3, since it is always positive, has a more appealing derivation, and in the cases considered is conservative in the sense of rejecting the null hypothesis slightly more than the appropriate factor, and so erring in the same direction as no factor.

Chapter 4: General estimation functions and factor analysis.

4.1 Introduction

Probably the most widely applied of all variance structures is the factor analysis model which takes the form

$$\Sigma = \Lambda\Lambda' + \Psi \qquad (1.1)$$

where Σ, as before, is the $p \times p$ variance matrix of a random vector $\underset{\sim}{y}$, Λ is the $p \times k$ real matrix of loadings of the elements of $\underset{\sim}{y}$ on k independent common factors, and Ψ is the $p \times p$ diagonal matrix of residual variances. For a full description of the assumptions which underlie the form (1.1), we refer to Lawley and Maxwell (1971).

In this chapter we consider minimum distance estimation of the parameters Λ and Ψ for the unrestricted model, i.e. where no constraints (other than those needed for identification) are imposed on Λ and Ψ, following the observation of a sample variance matrix S.*

The particular case of maximum likelihood has received considerable attention, starting with Lawley (1940, 1941) who derived the likelihood equations and proposed an iterative procedure for their solution. The method however proved unsatisfactory in many cases and it is only comparatively recently that the computational difficulties have been overcome [Jöreskog (1967), Jennrich and Robinson (1969), Clarke (1970)], revealing the source of the problem, namely a surprising number of improper solutions, i.e. solutions with one or more of the elements of Ψ estimated as zero.

---

* Much of the material to be presented here is also contained in a paper by the author [Swain (1975)].

As little is known of the small sample properties of the ML estimates, while, as shown in chapter 2, the large sample properties, which are summarised in appendix 4A, are common to the m.d.e. for any $F \in \mathcal{F}$, investigation of other m.d.e. seems warranted. The particular case of GLS has been considered by Jöreskog and Goldberger (1972).

The parameters of (1.1) are clearly not identified, since Λ can be post-multiplied by any $k \times k$ orthogonal matrix without altering Σ. We will assume throughout this chapter that the population parameters Λ_0 and Ψ_0 are such that this is the only nonidentification present, and that it may be removed by the $\frac{1}{2}k(k-1)$ constraints defined by

$$\Lambda_0' \Sigma_0^{-1} \Lambda_0 = \text{diag.} \tag{1.2}$$

where $\Sigma_0 = \Lambda_0 \Lambda_0' + \Psi_0$, and a sign convention on the columns of Λ_0. Then the number of independent parameters in (1.1) is

$$q = pk - \tfrac{1}{2}k(k-1) + p \tag{1.3}$$

and the degrees of freedom associated with the model is

$$d = \tfrac{1}{2}[(p-k)^2 - (p+k)]. \tag{1.4}$$

We note that sufficient conditions, corresponding to the rank condition (d) of section 2.6, for the local identification of Ψ_0 and Λ_0 have been given by Anderson and Rubin (1956). These are included in the appendix to this chapter.

While the general approach of section 2.5 could be used for the determination of the estimates, the magnitude of the problem is considerably reduced by following the two stage minimisation procedure adopted by Jöreskog (1967) for ML, in which a conditional estimate, $\tilde{\Lambda}$, of Λ given Ψ is first determined as a function of Ψ, and the conditional minimum, $\tilde{F}(\Psi;S)$, then minimised with respect to Ψ. We show in section 4.2 that $\tilde{\Lambda}$ is in fact independent of the particular $F \in \mathcal{F}$ being used, while in section 4.3 we present the derivatives of $\tilde{F}$ w.r.t. the elements of Ψ, as required for the numerical determination of $\hat{\Psi}$. We note here that for the restricted factor model, in which further constraints are placed on the elements of Λ, there seems little alternative to using the general approach of section 2.5, which for ML estimation is essentially the same as that used by Jöreskog (1969).

In section 4.4 we examine theoretically the effect of varying choice of F on the estimates, while the estimates obtained for different F for two sets of data are presented in section 4.5. Finally in section 4.6 some of the practical implications are discussed.

4.2 The estimation of Λ given Ψ

Given Ψ and the sample variance matrix S, the conditional m.d.e., $\tilde{\Lambda}$, of Λ will be determined for some $F \in \mathcal{F}$ by the minimisation of $F(\Lambda\Lambda'+\Psi;S)$, with the minimum being taken over all $p\times k$ real matrices.

To identify Λ for estimation we shall impose the $\frac{1}{2}k(k-1)$ identification conditions

$$\Lambda' S^{-1}\Lambda = \text{diag.}, \tag{2.1}$$

remarking that for any $p\times k$ matrix Λ, with $d_1 \leqslant \ldots \leqslant d_k$ the nonzero eigen values of $S^{-\frac{1}{2}}\Lambda\Lambda' S^{-\frac{1}{2}}$ and $\underset{\sim}{v}_1 \ldots \underset{\sim}{v}_k$, $V'V = I$, a set of corresponding eigen vectors, $\Lambda_1 = S^{\frac{1}{2}}\, VD^{\frac{1}{2}}$ has $\Lambda_1\Lambda_1' = \Lambda\Lambda'$ and satisfies (2.1). We note that as (2.1) is dependent on the sample matrix the theory of section 2.6 cannot be immediately applied. However the extension needed is simple, and is indicated at the end of this section.

We can determine $\tilde{\Lambda}$ from the following lemma.

Lemma 1. Let S be a $p\times p$ positive definite symmetric matrix and T be some $p\times p$ positive semidefinite matrix with rank $(T) \geqslant p-k$. Let $\gamma_1 \leqslant \ldots \leqslant \gamma_p$ be the eigen values of $S^{-\frac{1}{2}}TS^{-\frac{1}{2}}$ with $\underset{\sim}{v}_1 \ldots \underset{\sim}{v}_p$, $V'V = I$ a set of corresponding eigen vectors, and suppose that $\gamma_{k_0} \leqslant 1$ while $\gamma_{k_0+1} > 1$. Let $k_1 = \min(k, k_0)$, $\Gamma_1 = \text{diag}(\gamma_i)$, $i = 1\ldots k_1$, and $V_1 = (\underset{\sim}{v}_1 \ldots \underset{\sim}{v}_{k_1})$. Then for any $F \in \mathcal{F}$

$$\inf F(T+\Lambda\Lambda';\, S) = \sum_{i=k_1+1}^{p} f(\gamma_i), \tag{2.2}$$

where the inf is taken over all real $p\times k$ matrices Λ and is achieved at

$$\tilde{\Lambda} = S^{\frac{1}{2}}(V_1(I-\Gamma_1)^{\frac{1}{2}}\ 0). \tag{2.3}$$

Proof We assume, for simplicity, that $\gamma_1 < \ldots < \gamma_p$, noting that only minor modification is required to extend it to the case where there are equalities in the γ's. We wish to minimise $F(T+\Lambda\Lambda';S)$ subject to the identification condition (2.1). Using lemma 5 of chapter 2, we therefore seek the solutions of

$$\frac{\partial F}{\partial \Lambda} = 0 \quad ; \quad \Lambda' S^{-1}\Lambda = \text{diag}. \tag{2.4}$$

For arbitary Λ we let $\theta_1 \ldots \theta_p$ be the eigen values of $S^{-\frac{1}{2}}(T+\Lambda\Lambda')S^{-\frac{1}{2}}$ with $\underset{\sim}{h}_1 \ldots \underset{\sim}{h}_p$, $H'H = I$, a set of corresponding eigen vectors, $A = \text{diag}\,(f'(\theta_i))$, and $U = S^{-\frac{1}{2}}H$. Then, from (5.6) of chapter 2,

$$\begin{aligned} dF &= \text{tr}[UAU'd(T+\Lambda\Lambda')] \\ &= 2\ \text{tr}\ [UAU'\Lambda\ d\Lambda'] \end{aligned} \tag{2.5}$$

Thus we seek the solutions of

$$UAU'\Lambda = 0; \quad \Lambda' S^{-1}\Lambda = \text{diag}. \tag{2.6}$$

For any solution, Λ, of (2.6) the ith column, $\underset{\sim}{\lambda}_i$, must satisfy

$$AH'S^{-\frac{1}{2}}\underset{\sim}{\lambda}_i = \underset{\sim}{0} \tag{2.7}$$

which for $\underset{\sim}{\lambda}_i \neq \underset{\sim}{0}$ implies at least one of the diagonal elements of A is equal to zero, i.e. at least one of $\theta_1 \ldots \theta_p$ is equal to 1. Suppose Λ is such that $\theta_1 = \ldots = \theta_r = 1$ where $r \geqslant 1$. Then partitioning H as $(H_1\ H_2)$ where H_1 is $p\times q$ and H_2 is $p\times(p-q)$ we have from (2.7)

$$H_2' S^{-\frac{1}{2}}\underset{\sim}{\lambda}_i = \underset{\sim}{0}. \tag{2.8}$$

Then, since

$$H\Theta H' = S^{-\frac{1}{2}}(T+\Lambda\Lambda')S^{-\frac{1}{2}} = S^{-\frac{1}{2}}TS^{-\frac{1}{2}} + \sum_{j=1}^{k} S^{-\frac{1}{2}}\underset{\sim}{\lambda}_j\underset{\sim}{\lambda}_j' S^{-\frac{1}{2}},$$

$$S^{-\frac{1}{2}}TS^{-\frac{1}{2}} = I + H_2(\Theta_2 - I)H_2' - \sum_{j=1}^{k} S^{-\frac{1}{2}}\underset{\sim}{\lambda}_j\underset{\sim}{\lambda}_j' S^{-\frac{1}{2}}. \tag{2.9}$$

Post-multiplying by $S^{-\frac{1}{2}}\underset{\sim}{\lambda}_i$, we have from (2.2) and (2.6)

$$(S^{-\frac{1}{2}}TS^{-\frac{1}{2}})\ (S^{-\frac{1}{2}}\underset{\sim}{\lambda}_i) = (S^{-\frac{1}{2}}\underset{\sim}{\lambda}_i)\ (1-\underset{\sim}{\lambda}_i'\ S\underset{\sim}{\lambda}_i). \tag{2.10}$$

Thus for some j and constant C_j, $S^{-\frac{1}{2}}\underset{\sim}{\lambda}_i + C_j\underset{\sim}{v}_j$, and since then $\gamma_j = 1 - \underset{\sim}{\lambda}_i' S^{-1}\underset{\sim}{\lambda}_i = 1 - C_j^2$, so $C_j = (1-\gamma_j)^{\frac{1}{2}}$. Thus any Λ which is a solution of (2.6) can be written in the form

$$\Lambda^* = S^{\frac{1}{2}} V^* D \tag{2.11}$$

where $V^* = (\underset{\sim}{v}_{i_1} \ldots \underset{\sim}{v}_{i_k})$ for some $i_1 \ldots i_k$, and D is a $k\times k$ diagonal matrix with diagonal element d_j equal to either $(1-\gamma_{i_j})^{\frac{1}{2}}$ or zero. At Λ^*

$$S^{-\frac{1}{2}} (T+\Lambda^*\Lambda^{*\prime})S^{-\frac{1}{2}} = V\Gamma V' + V^*D^2V^{*\prime}, \tag{2.12}$$

and so the θ's are a permutation of $\gamma_i + n_i(1-\gamma_i)$, where n_i' is the number of times $(1-\gamma_i)^{\frac{1}{2}}$ appears in the diagonal of D. Thus the value of F at $T + \Lambda^*\Lambda^{*\prime}$ is

$$F^* = \sum_{i=1}^{p} f\left(\gamma_i+n_i(1-\gamma_i)\right) \tag{2.13}$$

From the conditions on f it is clear that the minimum value of F^* for real Λ^* will occur when $n_1 = \ldots = n_{k_1} = 1$, i.e. when $\Lambda^* = \hat{\Lambda}$ given by (2.3), with the corresponding value of F being given by (2.2).

For the unrestricted factor model we can obtain the conditional estimate $\hat{\Lambda}$ given Ψ from lemma 1 by setting $T = \Psi$. In this case however we can always assume, as we will do from here on, that $k_1 = k$. For, suppose $k_1 > k$. Then $S^{-\frac{1}{2}} (\Psi/\gamma_k)S^{-\frac{1}{2}}$ has eigen values γ_i/γ_k at least k of which are $\leqslant 1$, and the conditional minimum satisfies

$$\tilde{F}(\Psi/\gamma_k;S) = \sum_{i=k+1}^{p} f(\gamma_i/\gamma_k) < \tilde{F}(\Psi;S), \tag{2.14}$$

i.e., since our ultimate aim is minimisation with respect to both Λ and Ψ, where $k_1 > k$ we can replace Ψ by Ψ/γ_k.

Summarising then, if $\gamma_1 \leqslant \ldots \leqslant \gamma_p$ are the eigen values of $S^{-\frac{1}{2}}\Psi S^{-\frac{1}{2}}$ with $\underset{\sim}{v}_1 \ldots \underset{\sim}{v}_p$, $V'V = I$, a set of corresponding eigen vectors and $\gamma_k \leqslant 1$, $\Gamma_1 = \text{diag }(\gamma_i)$, $i = 1 \ldots k$, $V_1 = (\underset{\sim}{v}_1 \ldots \underset{\sim}{v}_k)$, for any $F \in \mathcal{F}$

$$\inf F(\Lambda\Lambda'+\Psi;S) = \sum_{i=k+1}^{p} f(\gamma_i) = \tilde{F}(\Psi;S) \tag{2.15}$$

where the inf is taken over all real $p \times k$ matrices Λ and is achieved at

$$\tilde{\Lambda} = S^{\frac{1}{2}} V_1 (I-\Gamma_1)^{\frac{1}{2}} \tag{2.16}$$

In order to remove the indeterminancy due to rotation we have imposed the condition that $\Lambda' S^{-1}\Lambda$ be a diagonal matrix. Because of its relevance to the asymptotic distribution theory, we show that, under the assumption that $\psi_i \neq 0$, $i = 1 \ldots p$, our solution $\tilde{\Lambda}$ agrees with that obtained for ML estimation under the alternative condition $\Lambda'\Psi^{-1}\Lambda$ be diagonal. For, if $\varphi_1 \geqslant \ldots \geqslant \varphi_p$ are the eigen values of $\Psi^{-\frac{1}{2}}S\Psi^{-\frac{1}{2}}$ with corresponding eigen vectors $\underset{\sim}{w}_1 \ldots \underset{\sim}{w}_p$, $W'W = I$, then it is easily shown that

$$\Gamma = \Phi^{-1}; \quad V = S^{\frac{1}{2}}\Psi^{-\frac{1}{2}}W\Phi^{-\frac{1}{2}} = S^{-\frac{1}{2}}\Psi^{\frac{1}{2}}W\Phi^{\frac{1}{2}}. \tag{2.17}$$

Thus on partitioning W and Φ in the same manner as before for V and Γ and substituting in (2.16) we have

$$\tilde{\Lambda} = \Psi^{\frac{1}{2}}W_1(\Phi_1-I)^{\frac{1}{2}}, \tag{2.18}$$

which is the desired result [Lawley and Maxwell (1971, p.28)]. This agreement had been noted numerically for ML estimation by Jennrich and Robinson (1969).

We also note that if for given Ψ with $\psi_i > 0$, $i = 1 \ldots p$, the methods of lemma 1 are applied to the minimisation over real Λ of $F(S-\Lambda\Lambda';\Psi) = F(\Psi^{-\frac{1}{2}}(S-\Lambda\Lambda')\Psi^{-\frac{1}{2}};I)$, then the minimum also occurs at $\tilde{\Lambda}$, given by (2.16), with value

$$\tilde{F} = \sum_{i=k+1}^{p} f(\varphi_i) = \sum_{i=k+1}^{p} f(\gamma_i^{-1}). \tag{2.19}$$

Thus the estimates of Λ and Ψ obtained by minimisation of $F(\Lambda\Lambda'+\Psi;S)$ are the same as the estimates obtained by minimisation of $F(\Psi;S-\Lambda\Lambda')$, i.e. choosing Λ and Ψ to make Σ close to S in one sense is equivalent to choosing them to make $\Psi^{-\frac{1}{2}}(S-\Lambda\Lambda')\Psi^{-\frac{1}{2}}$ close to I in another, closely related, sense. This result was brought to the attention of the author by M.W. Browne (personal communication).

It was noted earlier that because of the dependence of the identification condition (2.1) on the sample matrix S, the theory of section 2.6 cannot be applied immediately. We now indicate the necessary extensions.

We use the notation and formula numbers of section 2.6. We suppose now that the i.c. are of the form $\underset{\sim}{v}(\underset{\sim}{\gamma};S) = \underset{\sim}{0}$, and that, for given S, $\underset{\sim}{v}(\underset{\sim}{\gamma};S)$ satisfies the modified conditions (c), (d) and (e). We suppose further that $\underset{\sim}{v}(\underset{\sim}{\gamma};S)$ has continuous derivatives with respect to the elements of S. Then it is clear that, as for theorem 4, any approximate m.d.e. satisfying the i.c. $\underset{\sim}{v}(\underset{\sim}{\gamma}^*;S) = \underset{\sim}{0}$ will be consistent for $\underset{\sim}{\gamma}_0$, where $\Sigma_0 = \Sigma(\underset{\sim}{\gamma}_0)$ and $\underset{\sim}{v}(\underset{\sim}{\gamma}_0;\Sigma_0) = \underset{\sim}{0}$. Further we have

$$\begin{aligned} \underset{\sim}{0} &= \underset{\sim}{v}(\underset{\sim}{\gamma}^*;S) - \underset{\sim}{v}(\underset{\sim}{\gamma}_0;\Sigma_0) \\ &= V_0\ (\underset{\sim}{\gamma}^*-\underset{\sim}{\gamma}_0) + X_0(\underset{\sim}{s} - \underset{\sim}{\sigma}_0) + \text{higher order terms}, \end{aligned} \tag{2.20}$$

where

$$X_0 = \left.\frac{\partial \underset{\sim}{v}}{\partial \underset{\sim}{s}'}\right|_{\underset{\sim}{\gamma}_0}, \tag{2.21}$$

whence, as for theorem 4, $\underset{\sim}{\gamma}^*-\underset{\sim}{\gamma}_0 = 0_p(n^{-\frac{1}{2}})$. The result corresponding to theorem 5 also follows. Corresponding to (6.12) we have

$$(\hat{\underset{\sim}{\gamma}}-\underset{\sim}{\gamma}_0) = (\Delta_0'E\Delta_0+V_0'V_0)^{-1}\ (\Delta_0'E-V_0'X_0)\ (\underset{\sim}{s}-\underset{\sim}{\sigma}_0) + o_p(n^{-\frac{1}{2}}), \tag{2.22}$$

whence the result corresponding to (6.13) can be obtained.

4.3 The estimation of Ψ

The estimate, $\hat{\Psi}$, of Ψ is now determined by minimisation of $\tilde{F}(\Psi;S)$. The system of equations obtained by setting the derivatives of $\tilde{F}$ with respect to the elements of Ψ to zero do not admit an explicit solution. However $\hat{\Psi}$ can be efficiently determined from an initial approximation by use of a direct minimisation algorithm such as that of Fletcher and Powell, as described for ML estimation by Jöreskog (1967), which uses only the first derivatives, or by solution of the system of equations by the Newton-Raphson method, as described for ML estimation by Jennrich and Robinson (1969) and Clarke (1970), which uses expressions for both first and second derivatives.

We now present the formulae for these derivatives. The notation has been chosen so that for the particular case of ML estimation the expressions reduce to those given by Jennrich and Robinson. From (2.15) and lemma 4 of chapter 2 we have

$$d\tilde{F} = \sum_{i=k+1}^{p} f'(\gamma_i)\, d\gamma_i$$

$$= \sum_{i=k+1}^{p} f'(\gamma_i)\, \underset{\sim}{v}_i'\, d(S^{-\frac{1}{2}}\Psi S^{-\frac{1}{2}})\underset{\sim}{v}_i$$

$$= \sum_{i=k+1}^{p} \alpha_i\, \underset{\sim}{u}_i'\, d\Psi \underset{\sim}{u}_i, \tag{3.1}$$

where $\alpha_i = f'(\gamma_i)$ and $U = S^{-\frac{1}{2}}V$. Thus

$$\frac{\partial\tilde{F}}{\partial\psi_r} = \sum_{i=k+1}^{p} \alpha_i\, u_{ri}^2. \tag{3.2}$$

Further

$$dd\tilde{F} = \sum_{i=k+1}^{p} [d\alpha_i \underset{\sim}{u}_i' d\Psi_1 \underset{\sim}{u}_i + 2\alpha_i \underset{\sim}{u}_i' d\Psi_1 d\underset{\sim}{u}_i]$$

$$= \sum_{i=k+1}^{p} [f''(\gamma_i) d\gamma_i \underset{\sim}{u}_i' d\Psi_1 \underset{\sim}{u}_i$$

$$+ 2f'(\gamma_i)\, \underset{\sim}{u}_i'\, d\Psi_1 S^{-\frac{1}{2}}\, d\underset{\sim}{v}_i]$$

$$= \sum_{i=k+1}^{p} [f''(\gamma_i)\ \underset{\sim}{u}_i'd\Psi_1\underset{\sim}{u}_i\underset{\sim}{u}_i'\ d\Psi_2\underset{\sim}{u}_i$$

$$+ 2f'(\gamma_i)\ \underset{\sim}{u}_i'd\Psi_1\ S^{-\frac{1}{2}} \sum_{\substack{j=1\\ j\neq i}}^{p} \frac{\underset{\sim}{v}_j}{\gamma_i-\gamma_j}\ \underset{\sim}{u}_j'\ d\Psi_2\underset{\sim}{u}_i]$$

$$= \sum_{i=k+1}^{p} \sum_{j=1}^{p} \beta_{ij}\ \underset{\sim}{u}_i'd\Psi_1\ \underset{\sim}{u}_j\underset{\sim}{u}_j'd\Psi_2\underset{\sim}{u}_i, \tag{3.3}$$

where

$$\beta_{ij} = \begin{cases} f''(\gamma_i) & i = j \\ 2f'(\gamma_i)/(\gamma_i-\gamma_j) & i \neq j \end{cases} \tag{3.4}$$

Thus, from (3.3)

$$\frac{\partial^2\tilde{F}}{\partial\psi_r\partial\psi_s} = \sum_{i=k+1}^{p} \sum_{j=1}^{p} \beta_{ij}\ u_{ri}u_{si}u_{rj}u_{sj}. \tag{3.5}$$

Straightforward application of the minimisation algorithms using these expressions can frequently lead to solutions for which one or more of the elements of Ψ is negative. In order to avoid special procedures which will restrict the minimisation to the region $\psi_r \geqslant 0$, $r = 1 \ldots p$, different specifications of the parameter for minimisation have been proposed. Thus Jennrich and Robinson (1969) have written $\psi_r = \delta_r^2$ and Jöreskog and Goldberger (1972) have considered $\psi_r = \exp(\delta_r)$, in both cases the minimisation being with respect to δ_r. The appropriate formulae are easily obtained from the above. For suppose we write $\psi_r = \psi_r(\delta_r)$. Then

$$\frac{\partial\tilde{F}}{\partial\delta_r} = \frac{\partial\tilde{F}}{\partial\psi_r}\frac{d\psi_r}{d\delta_r}$$

$$\frac{\partial^2 F}{\partial\delta_r\ \partial\delta_s} = \frac{\partial^2 F}{\partial\psi_r\partial\psi_s}\frac{d\psi_r}{d\delta_r}\frac{d\psi_s}{d\delta_s} + \delta_{rs}\frac{\partial\tilde{F}}{\partial\psi_r}\frac{d^2\psi_r}{d\delta_r^2}, \tag{3.7}$$

where δ_{rs} is the Kronecker delta. Similarly more complicated reparametrisations could be accommodated.

A computer program for the determination of the m.d.e. $\hat{\Lambda}$ and $\hat{\Psi}$ implementing the methods of this section and section 4.2 is presented in appendix C. Following Jennrich and Robinson (1969) the minimisation is with respect to δ_r rather than $\psi_r = \delta_r^2$, this having the advantage that no special computing procedures at all are required to account for improper solutions. Unless an initial approximation to $\hat{\Psi}$ is specified the method of Jöreskog (1967) is used to generate one. This is firstly improved by (two) steepest descent iterations, and then the final estimates obtained by the Newton-Raphson method. This has proved to be a generally satisfactory procedure, with the steepest descent iterations improving the approximation sufficiently to remove the problems of non-positive definiteness of the matrix of second derivatives and of convergence to saddle points, which frequently occur with direct application of the Newton-Raphson method when minimisation is with respect to δ_r.

4.4 The effect of the choice of F

We now examine theoretically the effect of varying choice of $F \in \mathcal{F}$ upon the estimates obtained. Since the conditional estimate of Λ given Ψ is the same for all $F \in \mathcal{F}$, the main interest is in how $\hat{\Psi}$ will vary with F. We note that, from (2.15) and the form of $f(\theta)$, the minimisation of $\tilde{F}(\Psi;S)$ is equivalent to determining $\hat{\Psi}$ so that the eigen values $\gamma_{k+1} \dots \gamma_p$ of $S^{-\frac{1}{2}}\Psi S^{-\frac{1}{2}}$ are as close as possible to one, in the sense determined by $f(\theta)$, i.e. the variation in estimates for differing F will be mainly determined by the differences in the behaviour of $f(\theta)$ in the nbd. of $\theta = 1$. Since our conditions require that $f'(1)$ and $f''(1)$ are the same for all $F \in \mathcal{F}$, most of this difference in behaviour will be accounted for by $f'''(1)$, where this exists. To examine the effects on the estimates we therefore consider the family of functions

$$f(\theta;\kappa) = \frac{(\theta-1)^2}{2!} + \kappa\,\frac{(\theta-1)^3}{3!}. \tag{4.1}$$

Suppose that we have the estimate $\hat{\Psi}(\kappa)$ which is the value of Ψ which minimises

$$\tilde{F}(\Psi;\kappa) = \sum_{i=k+1}^{p} f(\gamma_i;\kappa) \tag{4.2}$$

in the nbd. of $\gamma_i = 1$, $i = k+1 \dots p$. Since there is no explicit expression for $\hat{\Psi}(\kappa)$ we cannot say exactly how $\hat{\Psi}(\kappa)$ will vary with κ. However the following arguement suggests something of the behaviour.

Consider a change $\delta\kappa$ in κ. Then

$$\tilde{F}(\Psi;\kappa+\delta\kappa) = \tilde{F}(\Psi;\kappa) + \delta\kappa \sum_{i=k+1}^{p} \frac{(\gamma_i-1)^3}{3!} \tag{4.3}$$

Let

$$g_r(\Psi;\kappa) = \frac{\partial\tilde{F}}{\partial\psi_r}(\Psi;\kappa),\ r = 1 \dots p. \tag{4.4}$$

Then, from (3.2),

$$g_r\ (\Psi;\kappa+\delta\kappa) = g_r\ (\Psi;\kappa) + \delta\kappa \sum_{i=k+1}^{p} \frac{(\gamma_i-1)^2}{2}\ u_{ri}^{\ 2} \qquad (4.5)$$

and since $g_r\ (\hat{\Psi}(\kappa);\kappa) = 0$,

$$g_r\ (\hat{\Psi}(\kappa);\ \kappa+\delta\kappa) = \delta\kappa \left(\sum_{i=k+1}^{p} \frac{(\gamma_i-1)^2}{2!}\ u_{ri}^{\ 2}\right)_{\hat{\Psi}(\kappa)}. \qquad (4.6)$$

An approximation to $\hat{\Psi}(\kappa+\delta\kappa)$ may be obtained by a steepest descent adjustment to $\hat{\Psi}(\kappa)$, i.e.

$$\hat{\psi}_r\ (\kappa+\delta\kappa) \doteqdot \hat{\psi}_r(\kappa) - k\ g_r\ (\hat{\psi}(\kappa);\ \kappa+\delta\kappa) \qquad (4.7)$$

where $k>o$. For $\delta\kappa>o$, from (4.6), this will always lead to a reduction in all the elements $\hat{\psi}_r(\kappa)$. This suggests that most, if not all, of the elements $\hat{\psi}_r\ (\kappa+\delta\kappa)$ will be less than the corresponding elements $\hat{\psi}_r(\kappa)$. However, because $\hat{\psi}_r(\kappa)$ is a smooth function of κ, if for some r and $\delta\kappa>o$ $\hat{\psi}_r(\kappa+\delta\kappa) > \hat{\psi}_r(\kappa)$, then a small increase in $\delta\kappa$ could be expected to produce a still larger value of $\hat{\psi}_r\ (\kappa+\delta\kappa)$.

Values of the third derivative $\kappa = f'''(1)$ for the particular estimation functions considered in detail are given in the following Table

TABLE 0

Label	$f(\theta)$	$\kappa = f'''(1)$
TGLS	$\frac{1}{2}(\theta-1)^2/\theta^2$	-6
DIV	$1/\theta + \theta - 2$	-6
ML	$1/\theta + \log\theta - 1$	-4
GD	$\frac{1}{2}(\log\theta)^2$	-3
GLS	$\frac{1}{2}(\theta-1)^2$	0
GLSE	$\frac{1}{2}(\theta-1)^2 e^{\theta-1}$	3

4.5 Numerical Examples

To illustrate numerically the differences which occur we now present the results of using each of the estimation functions of table 1 of chapter 2 for the factor analysis of two sets of data.

Lawley and Maxwell (1971, p.43) present the first set of data (Data 1), using it to illustrate ML estimation. The same data has also been used by Jöreskog and Goldberger (1972) as an example for GLS estimation. There are nine variates and a three factor model is fitted. The matrix is based on a sample of 211 observations, i.e. n = 210.

The second matrix considered (Data 2) was generated artificially as a sample matrix from a Wishart distribution with a known population matrix using the method described in section 3.2. The population matrix Σ_0 factorises exactly as $\Sigma_0 = \Lambda_0\Lambda_0' + \Psi_0$, with eight variates and two factors, and the matrices Ψ_0, Λ_0 and Σ_0 are presented in Table 2, where Λ_0 has been rotated to satisfy the condition that $\Lambda_0' \Sigma_0^{-1} \Lambda_0$ be a diagonal matrix.

The purpose in choosing an artificial example, as opposed to real data, was so that the differences in the estimates corresponding to different elements of F could be compared with the expected variation in the estimates due to sampling. The population matrix was chosen because of its particularly simple form, and the regular nature of the asymptotic variance matrix of the estimated ψ's. This was calculated, using (A1), for the case when the sample matrix S is reduced to the corresponding correlation matrix, and is included in Table 2. The correlation matrix actually used as Data 2 is presented in Table 3, and corresponds to a sample matrix with $n = 60$. The probability of the chisquare value for the likelihood ratio test of the null hypothesis $\Sigma = \Sigma_0$, based on S, [Anderson (1958, p.264)] was .662.

The solutions obtained for Data 1 and Data 2 are presented in Tables 1 and 4 respectively. Computation time and rate of convergence was comparable for all functions used. The χ^2 values shown in the tables were calculated as $n\ F(\hat{\Sigma};S)$. For Data 1 there is little real difference in the estimates obtained for either Ψ or Λ. However, as predicted in section 4.4, in general the elements of $\hat{\Psi}$ decrease with increasing κ, while ψ_8 increases with increasing κ. For Data 2 the differences are more marked, particularly for $\hat{\Psi}$. Again the trends predicted in section 4.4 are present. The ranges in the estimates for the functions considered are of much the same size as the asymptotic standard deviations for $n = 60$, and as the departure from the true values given in Table 2.

TABLE 1

Estimates for Data 1

	VARIATE								
	1	2	3	4	5	6	7	8	9
Residual Variances									
TGLS	.453	.432	.626	.214	.386	.182	.407	.453	.233
ML	.450	.427	.617	.212	.381	.177	.400	.462	.231
GD	.449	.424	.612	.211	.378	.174	.396	.465	.230
DIV	.449	.424	.612	.211	.378	.174	.396	.465	.230
GLS	.445	.416	.600	.208	.370	.168	.387	.473	.227
GLSE	.442	.409	.590	.205	.363	.163	.379	.479	.225
Loadings - factor 1									
TGLS	.666	.689	.493	.836	.703	.816	.663	.460	.767
ML	.664	.689	.493	.837	.705	.819	.662	.458	.766
GD	.664	.689	.492	.838	.706	.820	.661	.457	.765
DIV	.664	.689	.492	.838	.706	.820	.661	.457	.765
GLS	.662	.688	.491	.839	.708	.823	.660	.454	.763
GLSE	.661	.688	.490	.841	.710	.825	.659	.452	.761
Loadings - factor 2									
TGLS	.318	.242	.297	-.296	-.319	-.377	.391	.299	.424
ML	.321	.247	.302	-.292	-.315	-.377	.396	.296	.427
GD	.322	.249	.304	-.291	-.313	-.377	.398	.294	.429
DIV	.322	.249	.304	-.291	-.313	-.377	.398	.294	.429
GLS	.325	.255	.310	-.286	-.309	-.376	.404	.290	.434
GLSE	.328	.260	.314	-.282	-.305	-.375	.409	.288	.437
Loadings - factor 3									
TGLS	-.068	.195	.221	.032	.147	-.104	.081	-.496	.018
ML	-.074	.193	.222	.035	.153	-.105	.078	-.491	.012
GD	-.076	.193	.223	.037	.155	-.105	.076	-.489	.009
DIV	-.076	.193	.223	.037	.155	-.105	.076	-.489	.009
GLS	-.082	.191	.225	.041	.162	-.106	.073	-.484	.001
GLSE	-.087	.190	.227	.045	.168	-.108	.070	-.480	-.005

χ^2 (12)

TGLS	ML	GD	DIV	GLS	GLSE
7.339	7.354	7.305	7.315	6.975	6.462

TABLE 2

Population Parameters for Data 2.

	VARIATE 1	2	3	4	5	6	7	8
Residual Variances								
	.2	.4	.6	.8	.8	.6	.4	.2
Loadings								
	.800	.693	.566	.400	.400	.566	.693	.800
	.400	.346	.283	.200	-.200	-.283	-.346	-.400
Variance Matrix								
	1.000							
	.693	1.000						
	.566	.490	1.000					
	.400	.346	.283	1.000				
	.240	.208	.170	.120	1.000			
	.339	.294	.240	.170	.283	1.000		
	.416	.360	.294	.208	.346	.490	1.000	
	.480	.416	.339	.240	.400	.566	.693	1.000
$\frac{n}{2}$ times Asymptotic Variance Matrix of $\hat{\Psi}^*$								
	.421							
	-.156	.400						
	-.049	.088	.396					
	-.012	.051	.044	.313				
	.009	.014	.014	.009	.313			
	.014	.021	.021	.014	.044	.396		
	.014	.021	.021	.014	.051	.088	.400	
	.009	.014	.014	.009	-.012	-.049	-.156	.421

TABLE 3

Data 2 Correlation Matrix

VARIATE 1	2	3	4	5	6	7	8
1.000							
.624	1.000						
.626	.573	1.000					
.271	.285	.120	1.000				
.400	.263	.301	.157	1.000			
.340	.185	.296	.239	.524	1.000		
.319	.340	.249	.270	.582	.563	1.000	
.496	.396	.380	.253	.560	.553	.651	1.000

TABLE 4

Estimates for Data 2

	VARIATE							
	1	2	3	4	5	6	7	8
Residual Variances								
TGLS	.274	.506	.472	.930	.507	.533	.350	.342
ML	.282	.442	.446	.884	.491	.507	.324	.346
GD	.283	.413	.433	.860	.484	.493	.308	.349
DIV	.283	.413	.433	.859	.484	.492	.309	.348
GLS	.279	.341	.401	.792	.468	.453	.264	.355
GLSE	.270	.303	.377	.738	.460	.424	.236	.359
Popln.	.2	.4	.6	.8	.8	.6	.4	.2
Loadings - factor 1								
TGLS	.750	.626	.619	.339	.654	.617	.696	.778
ML	.744	.634	.621	.341	.656	.619	.704	.776
GD	.741	.638	.622	.343	.656	.620	.708	.775
DIV	.741	.638	.622	.343	.656	.620	.708	.775
GLS	.737	.652	.624	.347	.655	.621	.721	.770
GLSE	.735	.660	.625	.350	.652	.621	.728	.767
Popln.	.800	.693	.566	.400	.400	.566	.693	.800
Loadings - factor 2								
TGLS	.411	.374	.400	-.006	-.279	-.326	-.424	-.233
ML	.407	.396	.410	-.004	-.281	-.331	-.425	-.227
GD	.405	.405	.414	-.003	-.281	-.333	-.428	-.224
DIV	.405	.405	.414	-.003	-.282	-.334	-.428	-.224
GLS	.402	.429	.420	-.002	-.282	-.341	-.439	-.218
GLSE	.402	.440	.423	-.002	-.282	-.346	-.449	-.216
Popln.	.400	.346	.283	.200	-.200	-.283	-.346	-.400

χ^2 (13)

TGLS	ML	GD	DIV	GLS	GLSE
8.373	9.268	9.398	9.475	8.581	7.000

4.6 Discussion

The examples given illustrate that substantial differences in the estimates can occur when different estimation functions are used. None of the functions considered however has any real computational advantage over the others, and it also appears that the choice of estimation function will have little effect on the occurrence of improper solutions; in the author's experience any sample variance matrix with an improper ML solution has also yielded corresponding improper solutions for the other functions.

As the large-sample properties of the m.d.e. are the same for all $F \in \mathcal{F}$, the best choice would thus be that which gave the best small sample properties. Unfortunately no theoretical results are available for these small sample properties, even for the ML solution. Browne (1968) has reported some results from numerical experiments based on Wishart distributed samples with known population variance matrices. These suggest that the ML estimates of the residual variances are biased downward, with the degree of bias being larger for higher values of the true residual variances. If this is indeed the case, the results of section 4.4 suggest that bias correction could be achieved by choosing a function with a lower value of $f'''(1)$, e.g. TGLS, although there is no indication of how much change would be needed before overcorrection occurred.

However, when nS can be assumed to follow the Wishart distribution, ML estimation remains a logical approach. Further, the ML estimates have the pleasing, if unimportant property that the diagonal elements of S are reproduced in $\hat{\Sigma}$ [Lawley and Maxwell (1971, p.30)]. On the other hand each of the other estimation functions considered (except for GLSE) arises in a natural way, and so may merit consideration, especially where little is known about the distribution of S. In conclusion then, in the absence of further information on small-sample properties, there seems no clear reason for favouring any particular estimation function.

Appendix 4A Large sample variances and covariances of the estimates.

We summarise here the formulae for the large sample variances and covariances of the estimates $\hat{\Lambda}$ and $\hat{\Psi}$ using notation which is consistent with that of sections 4.2 and 4.3. The formulae given are essentially translations, using (2.17), of those of Lawley and Maxwell (1971, chapter 5) with the modification suggested by Jennrich and Thayer (1973) incorporated in (A6) below.

For convenience we now denote the population parameters by Λ and Ψ, with $\Sigma = \Lambda\Lambda' + \Psi$, and assume Λ rotated to satisfy $\Lambda'\Sigma^{-1}\Lambda$ diagonal. We let $\gamma_1 \leqslant \ldots \leqslant \gamma_p$ be the eigenvalues of $\Sigma^{-\frac{1}{2}}\Psi\Sigma^{-\frac{1}{2}}$ with corresponding eigen-vectors $\underset{\sim}{v}_1 \ldots \underset{\sim}{v}_p$, $V'V = I$ and V partitioned as $(V_1\ V_2)$ with V_1 $p\times k$. We assume that the matrix E defined below is nonsingular and that $\gamma_1 \ldots \gamma_k$ are distinct. As shown by Anderson and Rubin (1956) these together imply the local identification of Ψ, and Λ under the identification condition $\Lambda'\Sigma^{-1}\Lambda$ diagonal. We will also assume that $\psi_i > o$ $i = 1 \ldots p$.

Let S be distributed as in section 2.1. Then writing

$\underset{\sim}{\Psi} = (\psi_1 \ldots \psi_p)'$, $\hat{\underset{\sim}{\Psi}} = (\hat{\psi}_1 \ldots \hat{\psi}_p)'$

$$\sqrt{n}\left(\begin{pmatrix}\hat{\underset{\sim}{\Psi}}\\ \hat{\underset{\sim}{\Lambda}}\end{pmatrix} - \begin{pmatrix}\underset{\sim}{\Psi}\\ \underset{\sim}{\Lambda}\end{pmatrix}\right) \dot{\sim} N\left(\begin{pmatrix}0\\ \underset{\sim}{0}\end{pmatrix}, \begin{pmatrix}2E^{-1} & 2E^{-1}B\\ 2B'E^{-1} & A+2B'E^{-1}B\end{pmatrix}\right) \qquad \text{(A1)}$$

where $E = (e_{rs}) = (\xi_{rs}^2)$ with

$$\Xi = \Sigma^{-\frac{1}{2}}V_2V_2'\Sigma^{-\frac{1}{2}} = \Sigma^{-1} - \Sigma^{-1}\Lambda(\Lambda'\Sigma^{-1}\Lambda)^{-1}\Lambda'\Sigma^{-1}, \qquad \text{(A2)}$$

and the matrices A and B are given by

$$a_{ir\ jr} = \frac{1}{1-\gamma_r}\left(\sigma_{ij} - \sum_{m=1}^{k} \mu_{rm}\lambda_{im}\lambda_{jm}\right) \qquad \text{(A3)}$$

$$\mu_{rm} = \begin{cases} \frac{1}{2}\frac{1}{1-\gamma_r} & m=r \\ \frac{1}{1-\gamma_m} - \frac{\gamma_m^2}{1-\gamma_m}\left(\frac{1-\gamma_r}{\gamma_m-\gamma_r}\right)^2 & m\neq r \end{cases} \tag{A4}$$

$$a_{ir\ js} = -\frac{\gamma_r\gamma_s}{(\gamma_r-\gamma_s)^2}\ \lambda_{is}\lambda_{jr} \qquad r\neq s \tag{A5}$$

$$b_{j\ ir} = -\frac{\lambda_{jr}\gamma_r}{1-\gamma_r}\ \psi_j^{-2}\ (\delta_{ij}\ \psi_j - \sum_{m=1}^{k} \nu_{rm}\ \lambda_{im}\ \lambda_{jm}) \tag{A6}$$

$$\nu_{rm} = \begin{cases} \frac{1}{2}\frac{\gamma_r}{1-\gamma_r} & m=r \\ -\frac{\gamma_m}{\gamma_m-\gamma_r} & m\neq r \end{cases} \tag{A7}$$

Note that A is the large sample conditional variance matrix of Λ given Ψ.

When estimates are standardised, i.e. the corresponding correlation matrix is used in place of S in estimation yielding estimates

$$\hat{\psi}_r^* = \hat{\psi}_r/s_{rr}; \quad \hat{\lambda}_{rs}^* = \hat{\lambda}_{rs}/(s_{rr})^{\frac{1}{2}}, \tag{A8}$$

the formulae given by Lawley and Maxwell (1971, formulae (5.48) - (5.51), (5.45)) apply, with A and B as defined above. In particular with $\psi_r^* = \psi_r/\sigma_{rr}$ asymptotically

$$n\ \mathrm{cov}\ (\hat{\psi}_r^*, \hat{\psi}_s^*) = 2\left(\frac{e^{rs}}{(\sigma_{rr}\sigma_{ss})^{\frac{1}{2}}} + \psi_r^*\psi_s^*\ \frac{\sigma_{rs}^2}{\sigma_{rr}\sigma_{ss}} - 2\delta_{rs}\psi_r^{*3}\right). \tag{A9}$$

Chapter 5: The direct product model of variance structure

5.1 Introduction

In many situations the p observed variates may be naturally arranged in matrix form, e.g. the variates may represent the results of p_2 psychological tests each administered to the same individual on p_1 occasions or by p_1 different methods. In each case the data for a particular individual could be presented as a $p_2 \times p_1$ matrix Y.

For data of this type, Σ, the variance matrix of $\underset{\sim}{Y}$, could be expected to exhibit special structure. Campbell and Fiske (1959), concerned with the determination of test validity, suggested a number of features which would be exhibited by R, the corresponding correlation matrix, in the multitrait-multimethod situation under valid test structure. More formal models for Σ for the examples above have been suggested by a number of authors, e.g. Tucker (1963, 1966) three mode factor analysis, Evans (1967), McDonald (1969, 1970), Boruch (1970) and Jöreskog (1971), Corballis (1973), and Hakstian (1973). While these differ from each other in form and assumptions made, in all cases the models are of the factor type, i.e. imply an additive structure to Σ, with, generally, some factors specific to the row variates of Y and others specific to the column variates.

In this chapter we consider the use of a direct product model for the variance structure for $\underset{\sim}{Y}$, i.e. we consider

$$\Sigma = \Sigma_1 \otimes \Sigma_2 \tag{1.1}$$

where Σ_1 and Σ_2 are $p_1 \times p_1$ and $p_2 \times p_2$ respectively. As will be seen in section 5.2 where the implications of the model are examined, when the model applies Σ_1 can be interpreted as the variance matrix between the column variates, while Σ_2 represents that for the row variates, thus providing a convenient separation of the sources of variance.

Sections 5.3, 5.4, and 5.5 are concerned with the estimation of the parameters Σ_1 and Σ_2. While the general procedure of section 2.5 could be applied, this proves to be very time consuming, and a procedure which takes advantage of the form of model is preferable. In section 5.3 the likelihood equations are derived, and a simple iterative scheme, which is non-Newton-Raphson but proves rapidly convergent, is proposed for their solution, while in 5.5 the asymptotic distribution of these estimates is obtained. The form of the likelihood equations suggests that good estimates would be obtained from just one iteration of the suggested procedure and in 5.4 it is shown that estimates obtained in this way possess the same asymptotic distribution as the ML estimates. Some properties of the iterative procedure also emerge.

Use of the structure is illustrated in section 5.6 by application to three sets of actual data. Two of these are from biological situations, while the third is of the multitrait-multimethod type mentioned earlier. Approximation to the distribution of the test statistic, as suggested in chapter 3, is considered in section 5.7. Finally in section 5.8, application of the model is discussed.

5.2 Features of the model

Suppose that Y is a $p_2 \times p_1$ random matrix with $V(\underset{\sim}{Y})$ given by (1.1). The implications of this form of Σ is perhaps most easily seen by considering the corresponding correlation matrix R. Since $\mathrm{diag}(\sigma_{ii}) = \mathrm{diag}(\sigma_{ii}^{(1)}) \otimes \mathrm{diag}(\sigma_{ii}^{(2)})$, $R = R_1 \otimes R_2$ where R_1 and R_2 are the correlation matrices corresponding to Σ_1 and Σ_2. From (2.6) of chapter 2, therefore

$$r_{ij\ k\ell} = r_{j\ell}^{(1)} r_{ik}^{(2)} \tag{2.1}$$

In particular $r_{ij\ i\ell} = r_{j\ell}^{(1)}$, i.e. the correlation between y_{ij} and $y_{i\ell}$ is the same for all i. Similarly, the correlation between y_{ij} and y_{kj} is $r_{ik}^{(2)}$, independent of the column j, while from (2.1) the correlation between elements of different rows and columns of Y is simply the product of the row and column correlations. *Thus R is the direct product of the correlation matrix of any row of Y and the correlation matrix of any column of Y.* Similarly, the variance matrix for any row of Y is proportional to Σ_1 and the variance matrix for any column of Y is proportional to Σ_2.

Σ_1, for example, may thus be interpreted as a variance matrix between the columns of Y. We note that a method of constructing such matrices in quite general situations has been proposed by Escoufier (1973). If $\mathrm{Cov}(\underset{\sim}{X}_i, \underset{\sim}{X}_j) = \Sigma_{ij}$, he defines the "covariance" between $\underset{\sim}{X}_i$ and $\underset{\sim}{X}_j$ by $\mathrm{Covv}(\underset{\sim}{X}_i, \underset{\sim}{X}_j) = \mathrm{tr}(\Sigma_{ij}\Sigma_{ji})$. When applied to the columns of Y, where $V(\underset{\sim}{Y})$ is given by (1.1), a matrix with elements $(\sigma_{ij}^{(1)})^2\ (\mathrm{tr}\Sigma_2)^2$ is obtained. Since $\mathrm{covv}(x_i, x_j) = (\mathrm{cov}(x_i, x_j))^2$, we see that his technique leads essentially to Σ_1.

The structure also has the following properties:

(i) Σ is positive definite if and only if both Σ_1 and Σ_2 are positive definite, i.e. if and only if there is no fixed linear relation between either the rows or the columns of Y. We shall assume throughout that this is the case.

(ii) The structure is invariant under linear transformation of Y, in the sense that if Z = AYB, then

$$V(\underset{\sim}{Z}) = (B'\Sigma_1 B) \otimes (A\Sigma_2 A'). \tag{2.2}$$

(iii) Σ_1 and Σ_2 are not identified, since for any $a>0$,

$$\Sigma_1 \otimes \Sigma_2 = (a\Sigma_1) \otimes (\tfrac{1}{a}\Sigma_2). \tag{2.3}$$

To effect identification for estimation purposes we will assume that $\sigma_{11}^{(1)} = 1$. Under this condition $\Sigma = \Sigma_1 \otimes \Sigma_2$ is easily solved for Σ_1 and Σ_2 since then

$$\sigma_{ij}^{(2)} = \sigma_{i1\ j1};\ \sigma_{ij}^{(1)} = \sigma_{1i\ 1j}/\sigma_{11}^{(2)}, \tag{2.4}$$

showing that (2.3) expresses the only form of non-identification in the model. We note that the condition $\sigma_{11}^{(1)} = 1$ will often also be an appropriate condition for interpretation; e.g. in the multitrait-multimethod situation above, while there may be a natural scale for the traits, that for the methods would usually be arbitrary, so that the variance of method 1 could be taken as 1, with the variances of the other methods then being relative to the first.

(iv) The correlation matrices R_1 and R_2 are identified, since these are independent of the particular a in (2.3). Similarly, other functions of Σ_1 and Σ_2 which are invariant under transformations of the type $\Sigma_1 \to a\Sigma_1$, $\Sigma_2 \to \frac{1}{a}\Sigma_2$, $a>0$, are identified, e.g. the principal components of Σ_1 and Σ_2 are identified, even though their variances will not be.

(v) The number of independent parameters is

$$q = \tfrac{1}{2}[p_1(p_1+1) + p_2(p_2+1)] - 1 \qquad (2.5)$$

so that the number of degrees of freedom in estimation is

$$d = \tfrac{1}{2}[p(p+1) - p_1(p_1+1) - p_2(p_2+1)] + 1 \qquad (2.6)$$

We note that the parameter reduction in the model can be quite considerable, e.g. if $p_1 = 3$ and $p_2 = 5$, then $q = 20$ while $d = 100$.

(vi) We also note that of the four features of multitrait-multimethod correlation matrices suggested by Campbell and Fiske (1959) as indicators of test validity, with Y as in Section 5.1, $R = R_1 \otimes R_2$ will automatically exhibit the second [for any fixed $j \neq \ell$, $r_{ij\ i\ell} > r_{ij\ k\ell}$ and $r_{kj\ i\ell}, k \neq i$] and the fourth [similar patterns in the $p_2 \times p_2$ matrices $(r_{ij\ k\ell})$, $i,k=1 \ldots p_2$, for all j and ℓ] while the first [high values of $r_{ij\ i\ell}$] will be true if the $r^{(1)}_{j\ell}$ are high, and the third [for fixed i and j, $k \neq i$ and $l \neq j$, $r_{ij\ i\ell} > r_{ij\ kj}$] will be true if $r^{(1)}_{j\ell} > r^{(2)}_{k\ell}$. Thus if the structure holds it will provide a convenient means of assessing validity. Further, the model possesses the multiplicative structure observed in multitrait-multimethod matrices by Campbell and O'Connell (1967).

5.3 Maximum likelihood estimation of the parameters

With the identification condition $\sigma_{11}^{(1)}=1$ the general method of estimation of section 2.5 can be applied, with

$$\underset{\sim}{\gamma}' = (\sigma_{21}^{(1)} \ldots \sigma_{p_1 1}^{(1)} \; \sigma_{22}^{(1)} \ldots \sigma_{p_1 p_1}^{(1)} \; \underset{\sim}{\sigma}_2') \qquad (3.1)$$

where $\underset{\sim}{\sigma}_2 = \mathrm{ve}(\Sigma_2)$. While it is difficult to write down a complete expression for $\Delta(\underset{\sim}{\gamma})$, defined by (3.1) of chapter 2, [see Neudecker (1969)] it is easily computed since $\sigma_{ij\;k\ell} = \sigma_{j\ell}^{(1)}\sigma_{ik}^{(2)}$ and so

$$\frac{\partial\sigma_{ij\;k\ell}}{\partial\sigma_{mn}^{(1)}} = \begin{cases} \sigma_{ik}^{(2)} & m=j,\ n=\ell \text{ or } m=\ell,\ n=j \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial\sigma_{ij\;k\ell}}{\partial\sigma_{rs}^{(2)}} = \begin{cases} \sigma_{j\ell}^{(1)} & r=i,\ s=k \text{ or } r=k,\ s=i \\ 0 & \text{otherwise} \end{cases} \qquad (3.2)$$

with $i,k=1 \ldots p_2$; $j,\ell=1 \ldots p_1$; $(j-1)p_2+i \geq (\ell-1)p_2+k$; $m,n=2 \ldots p_1$; $m \geq n$; $r,s=1 \ldots p_2$; $r \geq s$. However, because of the typically large number of parameters involved and the sparse nature of Δ, the general procedure is unnecessarily time consuming, and an approach which takes advantage of the form of the model is needed.

Although any other element of F could be used as an estimation function to give the same asymptotic sampling properties, we will consider only maximum likelihood estimation, under the assumption $nS \sim W_p(n,\Sigma=\Sigma_1\otimes\Sigma_2)$, i.e. Σ_1 and Σ_2 are to be estimated by minimising

$$F(\Sigma_1\otimes\Sigma_2;S) = \log|\Sigma| - \log|S| + \mathrm{tr}\, S\Sigma^{-1} - p$$

$$= p_2\log|\Sigma_1| + p_1\log|\Sigma_2| - \log|S| + \mathrm{tr}\, S(\Sigma_1^{-1}\otimes\Sigma_2^{-1}) - p, \qquad (3.3)$$

from (vii) and (x) of Appendix A. We will assume throughout that $n>p$ so that S can be assumed positive definite.

We define the $p_2^2 \times p_1^2$ matrix G_S by

$$(G_S)_{ij\,k\ell} = s_{ik\,j\ell}, \tag{3.4}$$

$i,j=1 \ldots p_2$; $k,\ell=1 \ldots p_1$, i.e. if S is partitioned as $(S_{k\ell})$ with each $S_{k\ell}$ $p_2 \times p_2$, then

$$G_S = (\underset{\sim}{S}_{11}\ \underset{\sim}{S}_{21} \ldots \underset{\sim}{S}_{p_1}\ \underset{\sim}{S}_{12} \ldots \underset{\sim}{S}_{p_1 p_1}). \tag{3.5}$$

We will also denote by A and B the $p_1 \times p_1$ and $p_2 \times p_2$ matrices defined by

$$a_{ij} = \sum_{k=1}^{p_1} \sum_{\ell=1}^{p_1} s_{ik\,j\ell}\, \sigma^{k\ell}_{(1)}$$

$$b_{ij} = \sum_{k=1}^{p_2} \sum_{\ell=1}^{p_2} s_{ki\,\ell j}\, \sigma^{k\ell}_{(2)}, \tag{3.6}$$

where $\sigma^{k\ell}_{(i)}$ is the $k\ell$ element of Σ_i^{-1}, $i = 1,2$. We note that because of the symmetry of S, Σ_1^{-1} and Σ_2^{-1}, A and B are also symmetric. Further

$$\underset{\sim}{A} = G_S \underset{\sim}{\Sigma}_1^{-1}; \quad \underset{\sim}{B} = G_S \underset{\sim}{\Sigma}_2^{-1} \tag{3.7}$$

where $\underset{\sim}{\Sigma}_i^{-1} = \mathrm{vec}(\Sigma_i^{-1})$, $i = 1,2$.

Now, since for a symmetric matrix Σ, $d(\log|\Sigma|) = \mathrm{tr}\,\Sigma^{-1}d\Sigma$ and $d\Sigma^{-1} = -\Sigma^{-1}d\Sigma\Sigma^{-1}$, from (3.3)

$$dF = p_2 \mathrm{tr}\Sigma_1^{-1}d\Sigma_1 - \mathrm{tr}\, S(\Sigma_1^{-1}d\Sigma_1\Sigma_1^{-1} \otimes \Sigma_2^{-1})$$

$$+ p_1 \mathrm{tr}\, \Sigma_2^{-1}d\Sigma_2 - \mathrm{tr}\, S(\Sigma_1^{-1} \otimes \Sigma_2^{-1}d\Sigma_2\Sigma_2^{-1}) \tag{3.8}$$

Further

$$\mathrm{tr}\, S(\Sigma_1^{-1}d\Sigma_1\Sigma_1^{-1} \otimes \Sigma_2^{-1})$$

$$= \sum_{i,j,k,\ell} s_{ij\,k\ell}\, (\Sigma_1^{-1}d\Sigma_1\Sigma_1^{-1})_{\ell j}\, \sigma^{ki}_{(2)}$$

cont.

$$= \sum_{j,\ell} b_{j\ell} \, (\Sigma_1^{-1} d\Sigma_1 \Sigma_1^{-1})_{\ell j}$$

$$= \text{tr } B\Sigma_1^{-1} \, d\Sigma_1 \Sigma_1^{-1} \tag{3.9}$$

and similarly

$$\text{tr } S(\Sigma_1^{-1} \otimes \Sigma_2^{-1} d\Sigma_2 \Sigma_2^{-1}) = \text{tr } A\Sigma_2^{-1} d\Sigma_2 \Sigma_2^{-1}$$

so from (3.8)

$$dF = \text{tr}(p_2 \Sigma_1^{-1} - \Sigma_1^{-1} B \Sigma_1^{-1}) d\Sigma_1$$

$$+ \text{ tr } (p_1 \Sigma_2^{-1} - \Sigma_2^{-1} A \Sigma_2^{-1}) d\Sigma_2 \tag{3.10}$$

Thus, since $\Sigma_1^{-1}, \Sigma_2^{-1}$, A and B are symmetric

$$\frac{\partial F}{\partial \sigma_{ij}^{(1)}} = \begin{cases} (p_2 \Sigma_1^{-1} - \Sigma_1^{-1} B \Sigma_1^{-1})_{ii} & i=j \\ 2(p_2 \Sigma_1^{-1} - \Sigma_1^{-1} B \Sigma_1^{-1})_{ij} & i>j \end{cases} \tag{3.11}$$

$$\frac{\partial F}{\partial \sigma_{k\ell}^{(2)}} = \begin{cases} (p_1 \Sigma_2^{-1} - \Sigma_2^{-1} A \Sigma_2^{-1})_{kk} & k=\ell \\ 2(p_1 \Sigma_2^{-1} - \Sigma_2^{-1} A \Sigma_2^{-1})_{k\ell} & k>\ell \end{cases} \tag{3.12}$$

In order to determine estimates which satisfy the identification condition $\sigma_{ii}^{(1)}=1$, we seek a solution to the equations formed by setting (3.11) (for $i>2$) and (3.12) to zero and $\sigma_{ii}^{(1)}$ to 1. From lemma 5 of chapter 2, we may alternatively seek a solution to the equations formed by setting all the derivatives in (3.11) and (3.12) to zero, i.e. of

$$\Sigma_1^{-1} = \Sigma_1^{-1} B \Sigma_1^{-1} / p_2 \tag{3.13}$$

$$\Sigma_2^{-1} = \Sigma_2^{-1} A \Sigma_2^{-1} / p_1, \tag{3.14}$$

and of the equation $\sigma_{11}^{(1)} = 1$. By retaining the symmetry with respect to Σ_1 and Σ_2 the second approach proves much simpler. We note that if $(\tilde{\Sigma}_1,\tilde{\Sigma}_2)$ is any solution of (3.13) and (3.14), then $(\tilde{\Sigma}_1/\tilde{\sigma}_{11}^{(1)},\tilde{\sigma}_{11}^{(1)}\tilde{\Sigma}_2)$ satisfies both these equations and $\sigma_{11}^{(1)} = 1$.

Using (3.7) we can write (3.13) and (3.14) as

$$\underset{\sim}{\Sigma}_1 = G_s' \, \underset{\sim}{\Sigma}_2^{-1}/p_2 \tag{3.15}$$

$$\underset{\sim}{\Sigma}_2 = G_s \, \underset{\sim}{\Sigma}_1^{-1}/p_1 \tag{3.16}$$

This suggests the following simple iterative scheme for the determination of $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ starting with some initial approximation $\Sigma_2^{(o)}$ to $\hat{\Sigma}_2$:

(i) Insert $\Sigma_2^{(i-1)}$ and use (3.15) to determine an approximation $\tilde{\Sigma}_1^{(i)}$ and $\hat{\Sigma}_1$.

(ii) If $i \geqslant 2$, compare $\tilde{\Sigma}_1^{(i)}$ and $\Sigma_1^{(i-1)}$. If the difference is sufficiently small end at step (iv).

(iii) Form $\Sigma_1^{(i)} = \tilde{\Sigma}_1^{(i)}/(\tilde{\Sigma}_1^{(i)})_{11}$.

(iv) Insert $\Sigma_1^{(i)}$ and use (3.16) to determine $\Sigma_2^{(i)}$.

(v) Return to step (i).

Some of the theoretical properties of this process are presented in the next section. While it may not be necessary for the normalisation (step (iii)) to be included in each iteration, this does make convergence test easier since then the quality of the initial approximation $\Sigma_2^{(o)}$ has little effect on the comparison in step (ii).

In practice a convenient initial approximation to $\hat{\Sigma}_2$ is given by S_{11}, the leading $p_2 \times p_2$ submatrix of S. Further the computation involved can be reduced by using the symmetry of Σ_1 and Σ_2 in (3.15) and (3.16) to obtain

$$\underset{\sim}{\sigma}_1 = G_1 \underset{\sim}{\sigma}_2^{-1}/p_2 \tag{3.17}$$

$$\underset{\sim}{\sigma}_2 = G_2 \underset{\sim}{\sigma}_1^{-1}/p_1 \tag{3.18}$$

where $\underset{\sim}{\sigma}_i^{-1} = \mathrm{ve}(\Sigma_i^{-1})$; $i = 1,2$, G_1 is the $\frac{1}{2}p_1(p_1+1) \times \frac{1}{2}p_2(p_2+1)$ matrix formed from G_S' by deleting rows ij for j>i, and adding together columns $k\ell$ and ℓk for $k \neq \ell$, and G_2 is similarly formed from G_S.

The procedure has been programmed (see Appendix D) and has proved highly successful in all cases tried. Although the convergence appears to be linear, differences of 1×10^{-9} between all the elements of $\Sigma_1^{(i-1)}$ and $\Sigma_1^{(i)}$ are typically achieved in less than ten iterations (see e.g. section 5.6).

5.4 Estimation from a single iteration of the likelihood equations.

In this section we shall examine some of the properties of estimates obtained from a single iteration of the likelihood equations, i.e. we will assume that some estimate $\Sigma_2^{(o)}$ of Σ_2 is available and consider the properties of the estimates determined by steps (i), (iii) and (iv) of the procedure proposed in the previous section. For convenience we write $(\Sigma_2^{(o)})^{-1} = V$.

Theorem 1 If V is any $p_2 \times p_2$ positive definite symmetric matrix, G_S is defined by (3.4) and the $p_1 \times p_1$ matrix A is defined by $\underset{\sim}{A} = G_S' \underset{\sim}{V}$, then A is positive definite.

Proof: For any $p_1 \times 1$ vector $\underset{\sim}{x}$

$$\underset{\sim}{x}'A\underset{\sim}{x} = \sum_{i,j} x_i \, a_{ij} \, x_j$$

$$= \sum_{i,j} x_i \sum_{k,\ell} (G_S')_{ij\;k\ell} \, v_{k\ell} \, x_j.$$

Since V is symmetric and positive definite it may be written as $H\Lambda H'$, where $\Lambda = \text{diag}\,(\lambda_r)$ with $\lambda_r > 0$, and H is orthogonal. Using this and (3.4)

$$\underset{\sim}{x}'A\underset{\sim}{x} = \sum_r \lambda_r \sum_{i,j,k,\ell} (x_j h_{\ell r}) s_{j\ell\;ik} (x_i h_{kr})$$

$$= \sum_r \lambda_r \, \underset{\sim}{y}_r' S \underset{\sim}{y}_r \geqslant 0$$

with equality if and only if $\underset{\sim}{y}_r = \underset{\sim}{0}$, $r = 1 \ldots p_2$, where $\underset{\sim}{y}_r = \underset{\sim}{x} \otimes \underset{\sim}{h}_r$ and $\underset{\sim}{h}_r$ is the r^{th} column of H. But $\underset{\sim}{y}_r' \otimes \underset{\sim}{y}_r = (\underset{\sim}{x}' \otimes \underset{\sim}{h}_r')(\underset{\sim}{x} \otimes \underset{\sim}{h}_r) = \underset{\sim}{x}'\underset{\sim}{x} = 0$ if and only if $\underset{\sim}{x} = \underset{\sim}{0}$. Thus A is positive definite.

Similarly it may be shown that if $\underset{\sim}{B} = G_S\underset{\sim}{U}$, then the $p_2 \times p_2$ matrix B is positive definite for any $p_1 \times p_1$ positive definite symmetric matrix U.

Theorem 2. If V is any $p_2 \times p_2$ positive definite symmetric matrix, G_S is defined by (3.4) and $S = S_1 \otimes S_2$, with S_1 and S_2 $p_1 \times p_1$ and $p_2 \times p_2$ respectively, and with $s_{11}^{(1)}=1$, then a single iteration of the procedure defined above will determine S_1 and S_2, i.e.

$$\underset{\sim}{S}_1 = (\frac{1}{p_2} G_S' \underset{\sim}{V}) \;/\; (\frac{1}{p_2} G_S' \underset{\sim}{V})_1 \tag{4.1}$$

$$\underset{\sim}{S}_2 = \frac{1}{p_1} G_S \underset{\sim}{S}_1^{-1} \tag{4.2}$$

Proof: From theorem 1, if $\underset{\sim}{A} = G_S'\underset{\sim}{V}$, A is positive definite, and so in particular $(\underset{\sim}{A})_1 \neq 0$. Now from (3.4) and (2.6) of chapter 2, $(G_S)_{ij\;k\ell} = s_{ik\;j\ell} = s_{k\ell}^{(1)}\; s_{ij}^{(2)}$, i.e. $G_S = \underset{\sim}{S}_2\underset{\sim}{S}_1'$. Thus $G_S'\underset{\sim}{V} = \underset{\sim}{S}_1(\underset{\sim}{S}_2'\underset{\sim}{V})$, and since $s_{11}^{(1)} = 1$, $(G_S'\underset{\sim}{V})_1 = (\underset{\sim}{S}_2'\underset{\sim}{V})$, and so $(G_S'\underset{\sim}{V}) \;/\; (G_S'\underset{\sim}{V})_1 = \underset{\sim}{S}_1$. Further $(G_S\underset{\sim}{S}_1^{-1})/p_1 = \underset{\sim}{S}_2(\underset{\sim}{S}_1'\underset{\sim}{S}_1^{-1})/p_1 = \underset{\sim}{S}_2$.

In particular then, we have on setting $S = \Sigma = \Sigma_1 \otimes \Sigma_2$, with $\sigma_{11}^{(1)} = 1$,

$$\underset{\sim}{\Sigma}_1 = G_\Sigma' \; \underset{\sim}{\Sigma}_2^{-1}/p_1; \; \underset{\sim}{\Sigma}_2 = G_\Sigma \underset{\sim}{\Sigma}_1^{-1}/p_2 \tag{4.3}$$

The above theorem suggests that good estimates of Σ_1 and Σ_2 should be obtained from a single iteration of the equations even where S is the sample variance matrix and so will not possess the structure exactly. The possibility of getting asymptotically efficient estimates from the first iteration of an iterative process was first realised by Fisher (1925) who remarked that for ordinary cases, the first iteration of the method-of-scoring iterative process for the solution of likelihood equations yields asymptotically efficient estimates. Formal conditions and proof of the result for Newton-Raphson type iterative processes is given by Zacks (1971, p.250). Recently Anderson (1973) has proposed an iterative process for maximum likelihood estimation of the parameters of linear variance structures and has shown that the estimates obtained by a single iteration are asymptotically efficient, provided that the initial estimate is consistent. We will show in theorem 3 a similar result for the iterative process defined above for the direct product model.

While these one-iteration estimates, being dependent as they are on the initial estimates, might not usually be considered in practice, they may prove useful where either or both of p_1 and p_2 is large, or where large scale computing facilities are unavailable. In any case the results are of interest because of their information on the nature of the iterative process.

<u>Lemma 1</u>. Let G_S be defined by (3.4) with $nS \sim W_p(n,\Sigma=\Sigma_1 \otimes \Sigma_2)$, and V_1 and V_2 be $p_2 \times p_2$ positive definite symmetric matrices estimating Σ_2^{-1} such that $V_1 - V_2 = o_p(n^{-k})$, $k \geq 0$, and $V_2 - \Sigma_2^{-1} = O_p(n^{-\frac{1}{2}})$. Let $\underset{\sim}{w}$, $\underset{\sim}{x}$, $\underset{\sim}{y}$ and $\underset{\sim}{z}$ be defined by

$$\underset{\sim}{w} = G_S' \underset{\sim}{V}_1; \quad \underset{\sim}{x} = G_S' \underset{\sim}{V}_2;$$

$$\underset{\sim}{y} = G_\Sigma' \underset{\sim}{V}_1; \quad \underset{\sim}{z} = G_\Sigma' \underset{\sim}{V}_2. \qquad (4.4)$$

Then the following relationships hold:

(i) each of $\underset{\sim}{V}_1, \underset{\sim}{V}_2, \underset{\sim}{w}, \underset{\sim}{x}, \underset{\sim}{y}, \underset{\sim}{z}$ is $O_p(1)$

(ii) $\underset{\sim}{y}-\underset{\sim}{z} = o_p(n^{-k})$

(iii) $\underset{\sim}{x}-\underset{\sim}{z} = O_p(n^{-\frac{1}{2}})$

(iv) $\underset{\sim}{w}-\underset{\sim}{y} = O_p(n^{-\frac{1}{2}})$

(v) $\underset{\sim}{w}-\underset{\sim}{x} = o_p(n^{-k})$

(vi) $\underset{\sim}{w}-\underset{\sim}{x}-\underset{\sim}{y}+\underset{\sim}{z} = o_p(n^{(k+\frac{1}{2})})$

(vii) $\underset{\sim}{w}\underset{\sim}{x}' - \underset{\sim}{y}\underset{\sim}{z}' = O_p(n^{-\frac{1}{2}})$

(viii) $z_1 - p_2 = O_p(n^{-\frac{1}{2}})$

(ix) $x_1 - p_2 = O_p(n^{-\frac{1}{2}})$

<u>Proof</u>: (i)-(v) are immediate from definitions.

(vi) $\underset{\sim}{w}-\underset{\sim}{x}-\underset{\sim}{y}+\underset{\sim}{z} = (G_S'-G_\Sigma')(\underset{\sim}{V}_1-\underset{\sim}{V}_2)$

$$= O_p(n^{-\frac{1}{2}})o_p(n^{-k}) = o_p(n^{-(k+\frac{1}{2})}).$$

(vii) $\underset{\sim}{w} = \underset{\sim}{y}+(G_S'-G_\Sigma')\underset{\sim}{V}_1$; $\underset{\sim}{x} = \underset{\sim}{z}+(G_S'-G_\Sigma')\underset{\sim}{V}_2$.

Thus $\underset{\sim}{w}\underset{\sim}{x}'-\underset{\sim}{y}\underset{\sim}{z}' = (G_S'-G_\Sigma')\underset{\sim}{V}_1\underset{\sim}{z}' + \underset{\sim}{y}\underset{\sim}{V}_2'(G_S-G_\Sigma)$

$$+ (G_S'-G_\Sigma')\underset{\sim}{V}_1\underset{\sim}{V}_2'(G_S-G_\Sigma)$$

$$= O_p(n^{-\frac{1}{2}}).$$

(viii) $\underset{\sim}{z}-G_\Sigma'\underset{\sim}{\Sigma}_2^{-1} = G_\Sigma'(\underset{\sim}{v}_2-\underset{\sim}{\Sigma}_2^{-1}) = O_p(n^{-\frac{1}{2}})$

But from (4.3) $G_\Sigma'\underset{\sim}{\Sigma}_2^{-1} = p_2\underset{\sim}{\Sigma}_1$, while $\sigma_{11}^{(1)} = 1$, so $z_1-p_2 = O_p(n^{-\frac{1}{2}})$.

(ix) $x_1-p_2 = (x_1-z_1)+(z_1-p_2) = O_p(n^{-\frac{1}{2}})$ from (iii) and (viii).

<u>Theorem 3</u>. With G_S, V_1 and V_2 as in Lemma 1

$$\frac{G_S' \underset{\sim}{V_1}}{(G_S'\underset{\sim}{V_1})_1} - \frac{G_S'\underset{\sim}{V_2}}{(G_S'\underset{\sim}{V_2})_1} = o_p(n^{-\frac{1}{2}(k+1)}) \tag{4.5}$$

<u>Proof</u>: Using the notation of lemma 1, (4.5) is

$$\underset{\sim}{w}/w_1 - \underset{\sim}{x}/x_1 = o_p(n^{-\frac{1}{2}(k+1)})$$

Now, since from theorem 2 $\underset{\sim}{y}/y_1 = \underset{\sim}{z}/z_1 = \underset{\sim}{\Sigma}_1$, using results from lemma 1,

$$\begin{aligned}
\underset{\sim}{w}/w_1 - \underset{\sim}{x}/x_1 &= \underset{\sim}{w}/w_1 - \underset{\sim}{x}/x_1 - \underset{\sim}{y}/y_1 + \underset{\sim}{z}/z_1 \\
&= \underset{\sim}{w}\left(\frac{1}{w_1} - \frac{1}{p_2}\right) - \underset{\sim}{x}\left(\frac{1}{x_1} - \frac{1}{p_2}\right) - \underset{\sim}{y}\left(\frac{1}{y_1} - \frac{1}{p_2}\right) + \underset{\sim}{z}\left(\frac{1}{z_1} - \frac{1}{p_2}\right) \\
&\qquad + (\underset{\sim}{w}-\underset{\sim}{x}-\underset{\sim}{y}+\underset{\sim}{z})/p_2 \\
&= \underset{\sim}{w}\left(\frac{1}{w_1} - \frac{1}{x_1}\right) - \underset{\sim}{y}\left(\frac{1}{y_1} - \frac{1}{z_1}\right) + (\underset{\sim}{w}-\underset{\sim}{x})\left(\frac{1}{x_1} - \frac{1}{p_2}\right) \\
&\qquad - (\underset{\sim}{y}-\underset{\sim}{z})\left(\frac{1}{z_1} - \frac{1}{p_2}\right) + o_p(n^{-(k+\frac{1}{2})}) \\
&= \underset{\sim}{y}\left(\frac{1}{w_1} - \frac{1}{x_1} - \frac{1}{y_1} - \frac{1}{z_1}\right) + (\underset{\sim}{w}-\underset{\sim}{y})\left(\frac{1}{w_1} - \frac{1}{x_1}\right) + o_p(n^{-(k+\frac{1}{2})}) \\
&= \underset{\sim}{y}(1/w_1 - 1/x_1 - 1/y_1 + 1/z_1) + o_p(n^{-(k+\frac{1}{2})}).
\end{aligned}$$

Further

$$\begin{aligned}
&(x_1y_1z_1 - w_1y_1z_1 - w_1x_1z_1 + w_1x_1y_1) \\
&= \{(x_1-w_1)y_1z_1 - w_1x_1(z_1-y_1)\} \\
&= \{-(w_1-x_1-y_1+z_1)y_1z_1 + (w_1x_1 - y_1z_1)(y_1-z_1)\} \\
&= o_p(n^{-(k+\frac{1}{2})})
\end{aligned}$$

whence the result.

Corollary 1. Under the conditions of theorem 3, writing $\underset{\sim}{\Sigma}_{1,1} = \underset{\sim}{w}/w_1$; $\underset{\sim}{\Sigma}_{1,2} = \underset{\sim}{x}/x_1$

$$\frac{1}{p_2} G_S \underset{\sim}{\Sigma}_{1,1}^{-1} - \frac{1}{p_2} G_S \underset{\sim}{\Sigma}_{1,2}^{-1} = o_p(n^{-(k+\frac{1}{2})}) \qquad (4.6)$$

Proof: L.H.S. $= \frac{1}{p_2} G_S(\underset{\sim}{\Sigma}_{1,1}^{-1} - \underset{\sim}{\Sigma}_{1,2}^{-1}) = o_p(n^{-(k+\frac{1}{2})})$ from theorem.

Corollary 2. If $\Sigma_2^{(o)}$ is a $p_2 \times p_2$ positive definite symmetric matrix, consistent for Σ_2, $\Sigma_1^{(i)}$ and $\Sigma_2^{(i)}$ are defined as in section 4.3, and $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ are the maximum likelihood estimates of Σ_1 and Σ_2, then, writing

$$\underset{\sim}{\Sigma}_3^{(i)} = \begin{pmatrix} \underset{\sim}{\Sigma}_1^{(i)} \\ \underset{\sim}{\Sigma}_2^{(i)} \end{pmatrix}$$

$$\hat{\underset{\sim}{\Sigma}}_3 = \begin{pmatrix} \hat{\underset{\sim}{\Sigma}}_1 \\ \hat{\underset{\sim}{\Sigma}}_2 \end{pmatrix}, \text{ and } \Sigma_3 = \begin{pmatrix} \underset{\sim}{\Sigma}_1 \\ \underset{\sim}{\Sigma}_2 \end{pmatrix}$$

$$\underset{\sim}{\Sigma}_3^{(i)} - \hat{\underset{\sim}{\Sigma}}_3 = o_p(n^{-\frac{1}{2}}) \qquad (4.7)$$

In particular then

$$\sqrt{n}(\underset{\sim}{\Sigma}_3^{(1)} - \hat{\underset{\sim}{\Sigma}}_3) \overset{p}{\to} \underset{\sim}{0} \qquad (4.8)$$

Hence $\sqrt{n}(\underset{\sim}{\Sigma}_3^{(1)} - \underset{\sim}{\Sigma}_3)$ and $\sqrt{n}(\hat{\underset{\sim}{\Sigma}}_3 - \underset{\sim}{\Sigma}_3)$ will have the same asymptotic distribution.

Proof: Taking $V_1 = (\Sigma_2^{(o)})^{-1}$, $V_2 = \hat{\Sigma}_2^{-1}$, the result for i=1 follows from the theorem and corollary 1. Formula (4.7) then follows by induction on i.

Note. In the same manner as above it may be shown that if $\underset{\sim}{V}_1 - \underset{\sim}{V}_2 = O_p(n^{-(k+\frac{1}{2})})$, $k>o$, then

$$\underset{\sim}{\Sigma}_{1,1} - \underset{\sim}{\Sigma}_{1,2} = O_p(n^{-(k+1)}) \tag{4.9}$$

so that if $\Sigma_2^{(o)} - \Sigma_2 = O_p(n^{-\frac{1}{2}})$, then

$$\underset{\sim}{\Sigma}_3^{(i)} - \hat{\underset{\sim}{\Sigma}}_3 = O_p(n^{-\frac{i+1}{2}}) \tag{4.10}$$

This will be the case where $\Sigma_2^{(o)}$ is chosen as S_{11}.

5.5 Distributions of the estimates

In this section we consider the distributions of the maximum likelihood estimates $\hat{\Sigma}_1,\hat{\Sigma}_2$ of Σ_1,Σ_2. We will distinguish two situations, considering firstly the case where either Σ_1 or Σ_2 is known, and secondly the case where both are unknown. In the first case the exact distribution when $nS \sim W_p(n,\Sigma_1\otimes\Sigma_2)$ is derived, while in the second case only the asymptotic distribution is obtained.

Suppose firstly that Σ_1 is known. Then we may assume that $\Sigma_1=I$, since the value of Σ_2 which minimises $F(\Sigma_1 \otimes \Sigma_2;S)$ must also minimise
$F\left((\Sigma_1^{-\frac{1}{2}} \otimes I)(\Sigma_1 \otimes \Sigma_2)(\Sigma_1^{-\frac{1}{2}} \otimes I);(\Sigma_1^{-\frac{1}{2}} \otimes I)S(\Sigma_1^{-\frac{1}{2}} \otimes I)\right)$
$= F(I \otimes \Sigma_2;S_1)$ where $nS_1 = (\Sigma_1^{-\frac{1}{2}} \otimes I)S(\Sigma_1^{-\frac{1}{2}} \otimes I)$
$\sim W_p(n,I \otimes \Sigma_2)$. Now if S_1 is partitioned as $\left((S_1)_{kl}\right)$ where each $(S_1)_{kl}$ is $p_2 \times p_2$, from (3.6) and (3.10), $\hat{\Sigma}_2$ will be given by

$$\hat{\Sigma}_2 = \sum_{i=1}^{p_1}(S_1)_{ii}/p_1. \tag{5.1}$$

Since the $n(S_1)_{ii}$ are independently $\sim W_{p_2}(n,\Sigma_2)$ it follows that $np_1\hat{\Sigma}_2 \sim W_{p_2}(np_1,\Sigma_2)$. Summarising we have:

Theorem 4. If $nS \sim W_p(n,\Sigma_1 \otimes \Sigma_2)$ with Σ_1 known then $np_1\hat{\Sigma}_2 \sim W_{np_1}(np_1,\Sigma_2)$.

It is also easily seen that if it is desired to estimate the parameters of some model $\Sigma_2 = \Sigma_2(\underset{\sim}{\gamma})$ of variance structure for Σ_2, the same results would be obtained by using either the distribution of S as the likelihood function or by using the distribution of $\hat{\Sigma}_2$.

When Σ_2 is known and Σ_1 estimated, by similar reasoning to the above we find $np_2\hat{\Sigma}_1 \sim W_{p_1}(np, \Sigma_1)$. Note that the original, arbitarily chosen, identification condition $\sigma_{11}^{(1)} = 1$ no longer applies since the model is identified by the knowledge of Σ_2. As above, inferences on Σ_1 could be made using the marginal distribution of $\hat{\Sigma}_1$, and in particular if it was known that $\sigma_{11}^{(1)} = 1$, then the remaining parameters of Σ_1 could be estimated using the methods of Chapter 2.

We will now consider the case where both Σ_1 and Σ_2 are unknown, and determine the asymptotic distributions of $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$, where these are defined as solutions of (3.15) and (3.16) with $\hat{\sigma}_{11}^{(1)} = 1$, under the assumption that $\Sigma = \Sigma_1 \otimes \Sigma_2$ with $\sigma_{11}^{(1)} = 1$ and that $\sqrt{n}\,(\underset{\sim}{S} - \underset{\sim}{\Sigma}) \doteq N(\underset{\sim}{0}, V_\Sigma)$ where

$$V_\Sigma = (\Sigma \otimes \Sigma + \Sigma \boxtimes \Sigma). \qquad (5.2)$$

<u>Lemma 2</u>. With G_s defined by (3.4), and the asymptotic distribution of $\underset{\sim}{S}$ as above, $\sqrt{n}(\underset{\sim}{G}_s - \underset{\sim}{G}_\Sigma) \doteq N(\underset{\sim}{0}, U_\Sigma)$, where

$$U_\Sigma = \left[(\Sigma_1 \otimes \Sigma_1) \otimes (\Sigma_2 \otimes \Sigma_2) + (\Sigma_1 \boxtimes \Sigma_1) \otimes (\Sigma_2 \boxtimes \Sigma_2)\right]. \qquad (5.3)$$

<u>Proof:</u> Since the elements of $\underset{\sim}{G}_s$ are just a permutation of those of $\underset{\sim}{S}$, the asymptotic distribution of $\sqrt{n}(\underset{\sim}{G}_s - \underset{\sim}{G}_\Sigma)$ is clearly normal with mean $\underset{\sim}{0}$. Further

$$n \operatorname{cov}(g_{ij\ k\ell}, g_{mn\ rs}) = n \operatorname{cov}(s_{ik\ j\ell}, s_{mr\ ns})$$

$$= \sigma_{ik\ mr}\,\sigma_{j\ell\ ns} + \sigma_{ik\ ns}\,\sigma_{j\ell\ mr}$$

$$= \sigma_{kr}^{(1)}\,\sigma_{im}^{(2)}\,\sigma_{\ell s}^{(1)}\,\sigma_{jn}^{(2)} + \sigma_{ks}^{(1)}\,\sigma_{in}^{(2)}\,\sigma_{\ell r}^{(1)}\,\sigma_{jm}^{(2)}$$

$$= (U_\Sigma)_{ij\ k\ell,\ mn\ rs}.$$

Theorem 5. If $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ are defined as the solution of (3.15) and (3.16) with $\hat{\sigma}_{11}^{(1)} = 1$, and if the asymptotic distribution of S is as above, then

$$\sqrt{n} \left(\begin{bmatrix} \hat{\underset{\sim}{\Sigma}}_1 \\ \hat{\underset{\sim}{\Sigma}}_2 \end{bmatrix} - \begin{bmatrix} \underset{\sim}{\Sigma}_1 \\ \underset{\sim}{\Sigma}_2 \end{bmatrix} \right) \stackrel{\cdot}{\sim} N \left[\underset{\sim}{0}, \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \right] \tag{5.4}$$

where

$$\begin{aligned} V_{11} &= (\Sigma_1 \otimes \Sigma_1 + \Sigma_1 \boxtimes \Sigma_1 + 2\underset{\sim}{\Sigma}_1\underset{\sim}{\Sigma}_1' - 2\underset{\sim}{\Sigma}_1\underset{\sim}{c}' - 2\underset{\sim}{c}\underset{\sim}{\Sigma}_1')/p_2 \\ V_{21} &= 2\underset{\sim}{\Sigma}_2(\underset{\sim}{c}-\underset{\sim}{\Sigma}_1)'/p_2 \\ V_{22} &= (\Sigma_2 \otimes \Sigma_2 + \Sigma_2 \boxtimes \Sigma_2)/p_1 + 2(p_1-1)\underset{\sim}{\Sigma}_2\underset{\sim}{\Sigma}_2'/(p_1 p_2) \end{aligned} \tag{5.5}$$

and $\underset{\sim}{c}$ is the $p_1^2 \times 1$ vector with $(\underset{\sim}{c})_{ij} = \sigma_{i1}^{(1)}\sigma_{j1}^{(1)}$, i.e. $\underset{\sim}{c}$ is the first column of both $\Sigma_1 \otimes \Sigma_1$ and $\Sigma_1 \boxtimes \Sigma_1$.

Proof: From (3.15), (4.3) and theorem 2 we have

$$\begin{aligned} p_2(\hat{\underset{\sim}{\Sigma}}_1 - \underset{\sim}{\Sigma}_1) &= G_S' \hat{\underset{\sim}{\Sigma}}_2^{-1} - G_\Sigma' \underset{\sim}{\Sigma}_2^{-1} \\ &= (G_S' - G_\Sigma')\hat{\underset{\sim}{\Sigma}}_2^{-1} + G_\Sigma'(\hat{\underset{\sim}{\Sigma}}_2^{-1} - \underset{\sim}{\Sigma}_2^{-1}) \\ &= (G_S' - G_\Sigma')\hat{\underset{\sim}{\Sigma}}_2^{-1} + \alpha \underset{\sim}{\Sigma}_1 \end{aligned} \tag{5.6}$$

where α is some scalar. Looking only at the first element and denoting by $\underset{\sim}{g}_S$ and $\underset{\sim}{g}_\Sigma$ the first columns of G_S and G_Σ respectively we have, since $\hat{\sigma}_{11}^{(1)} = \sigma_{11}^{(1)} = 1$,

$$\alpha = -(\underset{\sim}{g}_S' - \underset{\sim}{g}_\Sigma')\hat{\underset{\sim}{\Sigma}}_2^{-1} = -(\hat{\underset{\sim}{\Sigma}}_2^{-1})'(\underset{\sim}{g}_S - \underset{\sim}{g}_\Sigma) \tag{5.7}$$

and so from (5.6)

$$\begin{aligned} p_2(\hat{\underset{\sim}{\Sigma}}_1 - \underset{\sim}{\Sigma}_1) &= (G_S' - G_\Sigma')\hat{\underset{\sim}{\Sigma}}_2^{-1} - \underset{\sim}{\Sigma}_1(\hat{\underset{\sim}{\Sigma}}_2^{-1})'(\underset{\sim}{g}_S - \underset{\sim}{g}_\Sigma) \\ &= \mathrm{vec}[((G_S' - G_\Sigma')\hat{\underset{\sim}{\Sigma}}_2^{-1})'] - \underset{\sim}{\Sigma}_1(\hat{\underset{\sim}{\Sigma}}_2^{-1})'(\underset{\sim}{g}_S - \underset{\sim}{g}_\Sigma) \end{aligned}$$

cont.

$$= \text{vec}[(\hat{\underset{\sim}{\Sigma}}_2^{-1})'(G_S - G_\Sigma)] - \underset{\sim}{\Sigma}_1(\hat{\underset{\sim}{\Sigma}}_2^{-1})'(\underset{\sim}{g}_S - \underset{\sim}{g}_\Sigma)$$

$$= (I_{p_1^2} \otimes (\underset{\sim}{\Sigma}_2^{-1})')(\underset{\sim}{G}_S - \underset{\sim}{G}_\Sigma) - \underset{\sim}{\Sigma}_1(\underset{\sim}{\Sigma}_2^{-1})'(\underset{\sim}{g}_S - \underset{\sim}{g}_\Sigma) + O_p(n^{-1}). \tag{5.8}$$

Now, from (3.16) and (4.1) we have

$$p_1(\hat{\underset{\sim}{\Sigma}}_2 - \underset{\sim}{\Sigma}_2) = G_S \hat{\underset{\sim}{\Sigma}}_1^{-1} - G_\Sigma \underset{\sim}{\Sigma}_1^{-1}$$

$$= (G_S - G_\Sigma)\hat{\underset{\sim}{\Sigma}}_1^{-1} + G_\Sigma(\hat{\underset{\sim}{\Sigma}}_1^{-1} - \underset{\sim}{\Sigma}_1^{-1}) \tag{5.9}$$

But, as in theorem 2, $G_\Sigma = \underset{\sim}{\Sigma}_2 \underset{\sim}{\Sigma}_1'$ and, using the fact that $\underset{\sim}{\Sigma}_1' \underset{\sim}{\Sigma}_1^{-1} = (\hat{\underset{\sim}{\Sigma}}_1^{-1})'\hat{\underset{\sim}{\Sigma}}_1 = p_1$,

$$G_\Sigma(\hat{\underset{\sim}{\Sigma}}_1^{-1} - \underset{\sim}{\Sigma}_1^{-1}) = -(\hat{\underset{\sim}{\Sigma}}_1^{-1})'(\hat{\underset{\sim}{\Sigma}}_1 - \underset{\sim}{\Sigma}_1) \tag{5.10}$$

and so, on substituting in (5.9) and using (5.8)

$$p_1(\hat{\underset{\sim}{\Sigma}}_2 - \underset{\sim}{\Sigma}_2) = (I - \frac{1}{p_2}\underset{\sim}{\Sigma}_2(\underset{\sim}{\Sigma}_2^{-1})')(G_S - G_\Sigma)\underset{\sim}{\Sigma}_1^{-1}$$

$$+ \frac{1}{p_2}\underset{\sim}{\Sigma}_2(\underset{\sim}{\Sigma}_2^{-1})'(\underset{\sim}{g}_s - \underset{\sim}{g}_\Sigma). \tag{5.11}$$

Taking the vector form of the first term and combining with (5.8) we have

$$\begin{bmatrix} \hat{\underset{\sim}{\Sigma}}_1 - \underset{\sim}{\Sigma}_1 \\ \hat{\underset{\sim}{\Sigma}}_2 - \underset{\sim}{\Sigma}_2 \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} \underset{\sim}{G}_S - \underset{\sim}{G}_\Sigma \\ \underset{\sim}{g}_S - \underset{\sim}{g}_\Sigma \end{bmatrix} + O_p(n^{-1}) \tag{5.12}$$

where

$$M_{11} = I_{p_1^2} \otimes [(\underset{\sim}{\Sigma}_2^{-1})']/p_2$$

$$M_{12} = -\underset{\sim}{\Sigma}_1(\underset{\sim}{\Sigma}_2^{-1})'/p_2$$

$$M_{21} = [(\underset{\sim}{\Sigma}_1^{-1})' \otimes (I - \underset{\sim}{\Sigma}_2(\underset{\sim}{\Sigma}_2^{-1})'/p_2)]/p_1$$

$$M_{22} = \underset{\sim}{\Sigma}_2(\underset{\sim}{\Sigma}_2^{-1})'/p_2 \tag{5.13}$$

Thus, since from lemma 2 we find

$$n \operatorname{Var}\begin{bmatrix} \underset{\sim}{G}_s - \underset{\sim}{G}_\Sigma \\ \underset{\sim}{g}_s - \underset{\sim}{g}_\Sigma \end{bmatrix} = \begin{bmatrix} U_\Sigma & \underset{\sim}{c} \otimes V_{\Sigma_2} \\ \underset{\sim}{c}' \otimes V_{\Sigma_2} & V_{\Sigma_2} \end{bmatrix} \tag{5.14}$$

where V_{Σ_2} is defined by (5.2), we eventually obtain (5.4) and (5.5).

For computational purposes the formulae (5.5) may be written

$$n \operatorname{cov}(\hat{\sigma}_{ij}^{(1)}, \hat{\sigma}_{k\ell}^{(1)}) \doteqdot (\sigma_{j\ell}^{(1)}\sigma_{ik}^{(1)} + \sigma_{i\ell}^{(1)}\sigma_{jk}^{(1)} + 2\sigma_{ij}^{(1)}\sigma_{k\ell}^{(1)}$$

$$-2\sigma_{ij}^{(1)}\sigma_{k1}^{(1)}\sigma_{\ell 1}^{(1)} - 2\sigma_{i1}^{(1)}\sigma_{j1}^{(1)}\sigma_{k\ell}^{(1)})/p_2$$

$$n \operatorname{cov}(\hat{\sigma}_{ij}^{(1)}, \hat{\sigma}_{k\ell}^{(2)}) \doteqdot 2(\sigma_{i1}^{(1)}\sigma_{j1}^{(1)} - \sigma_{ij}^{(1)})\sigma_{k\ell}^{(2)} / p_2$$

$$n \operatorname{cov}(\hat{\sigma}_{ij}^{(2)}, \hat{\sigma}_{k\ell}^{(2)}) \doteqdot (\sigma_{j\ell}^{(2)}\sigma_{ik}^{(2)} + \sigma_{i\ell}^{(2)}\sigma_{jk}^{(2)})/p_1$$

$$+ 2(p_1-1)\sigma_{ij}^{(2)}\sigma_{k\ell}^{(2)}/(p_1p_2). \tag{5.15}$$

We note that (5.15) yields $\operatorname{cov}(\hat{\sigma}_{11}^{(1)}, \hat{\sigma}_{k\ell}^{(1)}) = \operatorname{cov}(\hat{\sigma}_{11}^{(1)}, \hat{\sigma}_{k\ell}^{(2)}) = 0$, as expected since $\hat{\sigma}_{11}^{(1)} = \sigma_{11}^{(1)} = 1$.

Since the asymptotic marginal distribution for $\hat{\Sigma}_2$ is dependent only on Σ_2, it might be taken as a natural basis for the estimation of the parameters, $\underset{\sim}{\gamma}$, of a variance structure model $\Sigma_2 = \Sigma_2(\underset{\sim}{\gamma})$. However, although the first term is of the form (5.2), the presence of the second term means that the general formulae of chapter 2 for estimation and tests of significance cannot be applied. On the other hand, since if $\underset{\sim}{u} = \sqrt{n}(\hat{\underset{\sim}{\sigma}}_2 - \underset{\sim}{\sigma}_2)$,

$$T = \underset{\sim}{u}'(\text{var}(\underset{\sim}{u}))^{-1}\underset{\sim}{u} \mathrel{\dot{\sim}} \chi^2(\tfrac{1}{2}p_2(p_2+1)), \tag{5.16}$$

and since for general vector $\underset{\sim}{x}$ and nonsingular matrix A

$$(A+\underset{\sim}{x}\underset{\sim}{x}')^{-1} = A^{-1} - A^{-1}\underset{\sim}{x}(1+\underset{\sim}{x}'A^{-1}\underset{\sim}{x})^{-1}\underset{\sim}{x}'A^{-1} \tag{5.17}$$

so, from (3.6) of chapter 2,

$$T = \frac{np_1}{2} \{tr[(\Sigma_2-\hat{\Sigma}_2)\Sigma_2^{-1}]^2 - \frac{(p_1-1)}{p_1p_2} [tr(\Sigma_2-\hat{\Sigma}_2)\Sigma_2^{-1}]^2\}$$

$$= \frac{np_1}{2} \{\sum_{i=1}^{p_2} (1-\phi_i)^2 - \frac{(p_1-1)}{p_1p_2} [\sum_{i=1}^{p_2}(1-\phi_i)]^2\} \tag{5.18}$$

where $\phi_1 \ldots \phi_{p_2}$ are the eigen values of $\Sigma_2^{-\frac{1}{2}}\hat{\Sigma}_2\Sigma_2^{-\frac{1}{2}}$. Minimisation of T would also asymptotically yield generalised least squares estimates of the parameters $\underset{\sim}{\gamma}$ [c.f. section 2.3].

We also note that because of the method used for identifying Σ_1 and Σ_2, i.e. $\sigma_{11}^{(1)} = 1$, estimation based on the asymptotic marginal distribution of $\hat{\Sigma}_2$ would only be appropriate either when the condition $\sigma_{11}^{(1)} = 1$ is realistic, or, if this is not the case, when the model for Σ_2 is essentially invariant under transformations of the type $\Sigma_2 \to a\Sigma_2$, e.g. correlation structure models with no constraint on variances.

Similar remarks apply to the fitting of models of variance structure for Σ_1. Further, where models for both Σ_1 and Σ_2 are proposed, because $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ are not asymptotically independent, tests based on the marginal distributions would not be independent. On the other hand, purely descriptive techniques which are either free from the identification problem or usually based on correlation matrices, e.g. principle components, could readily be applied to $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$.

5.6 Numerical examples

In this section we illustrate the use of the model as a descriptive technique by applying it to three sets of actual data. The procedure described above was used for estimation, the iterative process being terminated when the differences between all the elements of successive estimates of Σ_1 were less than 1×10^{-9}. For each example computation time on the CDC6400 was less than 3 seconds. The likelihood ratio statistic was multiplied by the factor ρ_4, defined by (4.14) of chapter 3 and further considered in section 5.7, to give a test statistic, T, more closely distributed to χ^2.

Our first example, which provides quite a meaningful illustration, uses some unpublished data kindly made available by Miss M. Rose. It comes from a uniformity experiment in which three wool measurements were made at seven different sites on twenty sheep for four successive years. The measurements were fibre diameter (FD), coefficient of variation of fibre diameter (CV), and crimp (CR), while the sites were upper shoulder (1), upper midside (2), upper hip (3), shoulder (4), midside (5), hip (6), and belly (7). A multivariate analysis of variance with animals and years as factors yielded the residual variance matrix shown in Table 1. After allowing for the estimation of six missing vectors, this will have 51 d.f.

The estimates $\hat{R}_1$ and $\hat{R}_2$ together with the corresponding variance estimates are presented in table 2. Because site 5 is the one commonly used in experimentation, and as the only meaning to be attached to site variance is a relative

TABLE 1

Data 1 - Variance matrix for wool measurements at different sites *

	FD							CV							CR						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
FD 1	11813																				
2	5731	9113																			
3	6511	5577	14635																		
4	7208	5253	8585	15346																	
5	7679	8269	10779	10516	17705																
6	4147	4172	6971	9971	8918	17989															
7	3972	5020	10239	7525	8823	4882	18181														
CV 1	861	-144	-7670	1047	1525	1490	-6325	53296													
2	-2592	3447	-4317	900	-520	-1558	-3862	15379	48035												
3	-4472	-3137	-2278	-4275	2692	4199	-9577	16636	12224	71307											
4	2354	-2409	-1305	1826	6137	388	-3249	19576	10071	27064	55072										
5	3902	1084	2849	7776	5497	-1344	818	18508	11614	16141	27649	65506									
6	2533	538	-933	706	4888	3277	-1909	11322	-5191	12722	2602	5954	52264								
7	8327	6862	8400	6475	11164	3495	5185	3745	-6267	-5542	3925	16453	8959	87166							
CR 1	-2518	-2124	-2828	-1650	-4371	-262	-1912	-848	602	-1576	-9356	-5372	595	-8869	7434						
2	-2237	-3652	-3637	-3030	-4616	-2903	-3354	-310	476	-5346	-3495	-3876	-4036	-3702	3287	8042					
3	-3862	-4833	-7103	-6619	-8621	-5543	-6217	-2032	3179	-968	-2250	-2816	-3661	-6840	3703	5115	11199				
4	-617	-1956	-2969	-831	-2532	125	-1605	-448	683	-1318	1443	437	1866	-7117	2083	2152	2230	8422			
5	377	-1922	-4079	227	-2850	582	-4014	5718	2194	-954	-412	-502	2881	1007	2424	2973	3752	1789	9924		
6	-1072	-1876	-4943	-1969	-5187	-1850	-2641	37	3098	-2862	-5920	-5864	-1994	-5167	4176	3559	6088	2618	5333	11627	
7	-1082	-4272	-2514	1152	-3943	3651	-1125	-3598	-3667	981	1821	-9504	975	-6037	2282	1684	2140	1789	4233	3275	22218

* x 10^4

one, the variance estimates have been identified by the condition $\sigma_{55}^{(2)} = 1$. The fit of the model is acceptable, and the estimates appear to provide a useful description, with uniform trends in the site variances, and the estimate of the site correlation matrix also exhibiting a natural pattern, with each site correlating most highly with those nearest to it. We also note that the FD-CR correlation agrees well with the -.13 given by Newton Turner and Young (1969, p.135).

TABLE 2

Estimates for Data 1

$\hat{R}_1$

	FD	CV	CR
FD	1.0000		
CV	.0948	1.0000	
CR	- .1082	.0266	1.0000
Variances			
	1.3311	7.5069	1.1915

$\hat{R}_2$

	1	2	3	4	5	6	7
1	1.0000						
2	.4326	1.0000					
3	.4020	.4104	1.0000				
4	.4066	.3047	.4162	1.0000			
5	.3851	.4083	.4175	.4531	1.0000		
6	.3098	.2123	.3792	.3591	.3882	1.0000	
7	.1702	.1282	.2297	.2284	.3185	.2142	1.0000
Variances							
	.7359	.6471	.9749	.8684	1.0000	1.0080	1.4735

11 iterations $T = 221.8016$ $Pr(\chi^2(198)>T) = .1181$

The second example (Data 2) comes from Tudor (1971), who conducted an experiment in which 58 calves were allocated at birth to one of sixteen nutrition treatment groups, and slaughtered on reaching approximately 400kg liveweight. Various measurements on body composition were taken. In particular we will consider the weights (in grams) recorded for five muscles dissected from both left and right hind legs. In this case it is expected that the relationship between the muscles should be the same on both sides of the animal and that the correlation between sides for each muscle could be the same.

After deleting 2 animals for which there was missing data, the within groups matrix, W, was calculated with covariance adjustment for age and fasted body weight prior to slaughter. Thus W is on 38 d.f.. As for our purpose here, there is little interest in the muscle variances, and as both the structure and test statistic are preserved under transformation of the type $S \rightarrow (D_1 \otimes D_2)S(D_1 \otimes D_2)$ where D_1 and D_2 are $p_1 \times p_1$ and $p_2 \times p_2$ diagonal matrices, for convenience of presentation we rescale W, choosing $D_1 = I_2$ and $D_2 = \{\mathrm{diag}(W_{11})\}^{-\frac{1}{2}}$, where W_{11} is the leading 5x5 submatrix of W. This rescaled variance matrix is presented in table 3, together with $\mathrm{diag}(W_{11})$.

The corresponding estimates are presented in table 4. The reason for the poor fit appeared to be a quite large difference between the left-right correlation for muscle E and that for the other muscles, the actual values being, in order A, B, C, D, E, .929, .919, .946, .922, and .597. This in turn may be attributable to a

TABLE 3

Data 2 - Rescaled variance matrix for hind leg muscles

	LEFT					RIGHT				
LEFT	A*	B	C	D	E	A	B	C	D	E
A	1.0000									
B	.5989	1.0000								
C	.6112	.4797	1.0000							
D	.5811	.5130	.6006	1.0000						
E	-.0513	.2656	.0245	.0399	1.0000					
RIGHT										
A	1.1091	.6636	.7905	.7181	-.1580	1.4262				
B	.6866	.9765	.5406	.6193	.2219	.8473	1.1288			
C	.6390	.4458	.9710	.6319	-.0556	.8782	.5727	1.0524		
D	.6620	.6221	.6301	.9494	-.0052	.8334	.7649	.7049	1.0611	
E	.4421	.5408	.2675	.4152	.7471	.3851	.6008	.1944	.3676	1.5642
diag (W_{11})										
	36508	30559	19249	8778	4232					

*A = M. biceps femoris B = M. semimembranosus
C = M. semitendinosus D = M. gastrocnemius
E = M. adductor femoris

higher measurement error associated with this muscle which is smaller and harder to dissect out than the others. When the analysis is redone without this muscle, the estimates shown in table 5 are obtained, giving quite an acceptable fit to the data. As an illustration of the formulae of section 5.5, and as these estimates are used as a basis for the simulation experiment described in section 5.7, we present in table 6, the estimated asymptotic variance matrix for $\hat{\Sigma}_1$, $\hat{\Sigma}_2$ calculated from (5.5) with Σ_1, Σ_2 replaced by $\hat{\Sigma}_1$, $\hat{\Sigma}_2$.

The third example (data 3) is taken from Campbell and Fiske (1959) who have presented data from an assessment study of clinical psychologists, considered earlier

TABLE 4

Estimates for Data 2 - all muscles

$\hat{\Sigma}_1$			$\hat{\Sigma}_2$					
	L	R		A	B	C	D	E
L	1.0000		A	.9766				
R	.8762	1.0305	B	.5357	.8524			
			C	.4863	.3857	.7257		
			D	.4243	.3916	.4149	.8140	
			E	.1585	.3923	.0647	.0941	2.4460

13 iterations $T = 66.2084$ $Pr(\chi^2(38)>T) = .0031$

TABLE 5

Estimate for Data 2 - muscle E excluded

$\hat{\Sigma}_1$			$\hat{\Sigma}_2$				
	L	R		A	B	C	D
L	1.0000		A	1.1202			
R	.9032	1.0082	B	.6021	.9692		
			C	.5284	.4297	.7968	
			D	.4490	.4179	.4433	.9205

11 iterations $T = 31.0427$ $Pr(\chi^2(24)>T) = .1526$

by Fiske (1949) and Kelly and Fiske (1951). Five personality traits of 124 subjects were assessed by three different methods viz. staff, teammate and self ratings. We will present only the results of fitting the direct product structure for the teammate and self ratings, this particular data being regarded by Campbell and Fiske as exhibiting good test validity.

TABLE 6

Estimated asymptotic variance matrix for $\hat{\Sigma}_1 \hat{\Sigma}_2$ of table 4*

	$\hat{\Sigma}_1$		$\hat{\Sigma}_2$									
$\hat{\Sigma}_1$	21	22	11	21	31	41	22	32	42	33	43	44
21	127											
22	229	511										
$\hat{\Sigma}_2$												
11	0	-284	4128									
21	0	-152	2219	2144								
31	0	-134	1947	1261	1726							
41	0	-114	1655	1150	1122	1755						
22	0	-245	1668	1920	1018	948	3090					
32	0	-109	1154	1184	1079	769	1370	1381				
42	0	-106	1019	1069	750	1100	1332	920	1519			
33	0	-202	1322	913	1385	852	994	1126	720	2088		
43	0	-112	951	720	933	1033	755	814	886	1162	1353	
44	0	-233	1209	858	844	1360	1047	748	1265	1000	1342	2787

* $\times 10^5$

In common with most published data of this type, only the correlation matrix is available, so that although the estimation procedure may still be applied, the associated significance test is not strictly appropriate. It should however, in this case, give some guide, provided that the trait variances of the teammate ratings do not differ greatly from those for the self ratings.

The estimated correlation matrices are given in table 7. We note that $\hat{R}_2$ agrees fairly closely with the individual monomethod matrices, and that the apparently good fit supports the Campbell and Fiske visual assessment of good validity.

The model was also fitted to the full correlation

TABLE 7

Estimated correlation matrices for Data 3

$\hat{R}_1$

	Te*	Se
Te	1.0000	
Se	.2823	1.0000

$\hat{R}_2$

	A	C	S	U	B
A	1.0000				
C	.2641	1.0000			
S	-.0832	-.1137	1.0000		
U	.1132	.2526	.1520	1.0000	
B	.2265	.1887	.2003	.3192	1.0000

8 iterations $T = 43.9295$ $Pr(\chi^2(38)>T) = .2347$

*Te = Teammate Se = Self
A = Assertive C = Cheerful S = Serious
U = Unshakable poise B = Broad issues

matrix with all three methods of rating included. Convergence occured in 9 iterations, with the value of the test statistic T now 157.9147 which is highly significant, with $Pr(\chi^2(100)>T) = .0002$. On examining the original correlation matrix it seems that the poor fit may be at least partly attributed to the rather large differences in the staff-teammate correlations for the different characteristics, and some noticeable departures from symmetry in the staff-teammate submatrix.

5.7 Approximation to the distribution of the test statistic

Although under the null hypothesis that the population variance matrix has the direct product structure $T_o = nF(\hat{\Sigma}_1 \otimes \hat{\Sigma}_2;S) \mathrel{\dot{\sim}} \chi^2(d)$, with F and d defined by (3.3) and (2.6), as shown in chapter 3 the approximation to the χ^2 distribution may be improved by multiplying T_o by some suitable factor ρ. In section 3.4 methods of constructing ρ are suggested for cases where $E[T_o]$ is unknown, and in this section we will examine, by a numerical experiment, the performance of these methods when applied to the direct product structure. In doing this we will also obtain some indication of how large n need be for the asymptotic formulae of section 5.5 for the mean and variance of $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ to be applicable.

The method was as follows. Four hundred samples were generated from $W_p(n,\Sigma=\Sigma_1 \otimes \Sigma_2)$ using the method described in section 3.2. In order that Σ_1 and Σ_2 be representative of those which occur in practice they were taken to be equal to the ML estimates for data 2, as given in table 5. Further, as improved approximation to the distribution of the test statistic is likely to be most important for moderate n, and since there is particular interest in the distribution for data 2, n=38 was taken. For each sample $\hat{\Sigma}_1$, $\hat{\Sigma}_2$ and T_o were calculated. The empirical c.d.f.'s of T_o and $\rho_i T_o$, i = 1...4, were then determined, where the ρ_i are given by (4.11) to (4.14) of chapter 3, with q and d given by (2.5) and (2.6), and plotted together with the approximating χ^2 c.d.f. using an incremental plotter.

The sample mean and variance matrix for the 400 sets of estimates are given in tables 8 and 9. These are quite close to the mean and variance of the asymptotic normal distribution as given in tables 4 and 5, suggesting that even with $n=38$ the asymptotic formulae may be used.

The c.d.f.'s plotted are shown in figures 1-5. For comparison the empirical c.d.f. of $\rho n F(\Sigma;S)$ with $\rho=1-(2p^2+3p-1)/(6n(p+1))$, which is the corresponding test statistic for the likelihood ratio test of hypothesis that the population matrix equals Σ, is shown in figure 6. The graphs suggest that the use of any of the factors for T_0 gives a statistic more closely distributed to $\chi^2(d)$ than T_0, with the better approximation being given by ρ_3 and ρ_4. Although this experiment has only been conducted for one particular n, Σ_1 and Σ_2, the results are in agreement with the theoretical comparisons made in section 3.4, and it appears that the general use of either ρ_3 or ρ_4 would be appropriate. Since ρ_4 is the more conservative, in the sense of section 3.4, and the form of figure 5 looks closer to that of figure 6 than does that of figure 4, ρ_4 has been preferred for the examples of section 5.6.

TABLE 8

Mean of estimates from 400 samples

$\hat{\Sigma}_1$			$\hat{\Sigma}_2$			
1.0000			1.1264			
.9044	1.0115		.6047	.9737		
			.5268	.4324	.7992	
			.4541	.4169	.4491	.9200

TABLE 9

Variance matrix of estimates from 400 samples*

		$\hat{\Sigma}_1$		$\hat{\Sigma}_2$									
		21	22	11	21	31	41	22	32	42	33	43	44
$\hat{\Sigma}_1$	21	141											
	22	262	593										
$\hat{\Sigma}_2$	11	-58	-376	4327									
	21	-41	-208	2356	2254								
	31	-30	-184	2081	1337	1837							
	41	-53	-184	1615	1216	1178	1859						
	22	-27	-305	1622	1837	907	852	2947					
	32	-30	-177	1207	1197	1081	789	1261	1331				
	42	-45	-189	993	1138	725	1102	1299	923	1479			
	33	-59	-356	1570	1004	1491	948	918	1130	715	2309		
	43	-29	-181	1040	760	966	1112	683	799	870	1211	1363	
	44	-49	-311	1131	931	761	1370	1148	749	1213	958	1249	2627

* $\times 10^5$

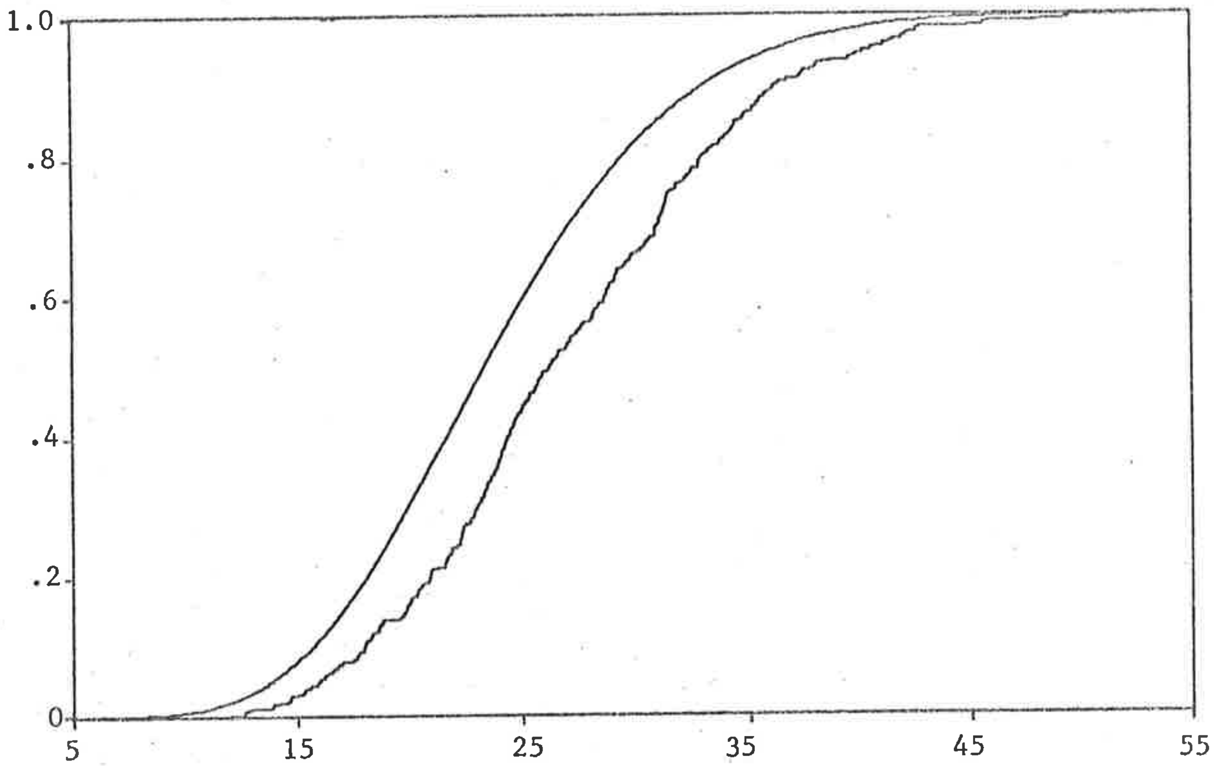

FIGURE 1 - Empirical c.d.f. for T_o

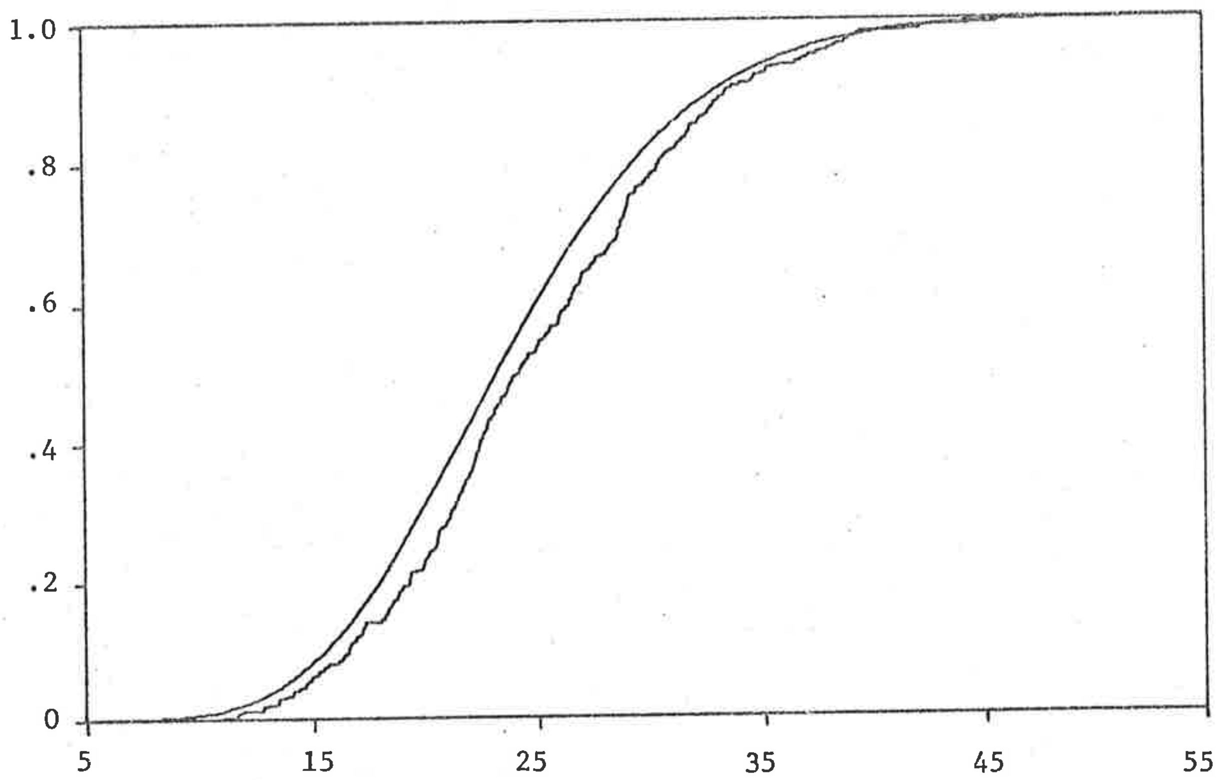

FIGURE 2 - Empirical c.d.f. for $\rho_1 T_o$

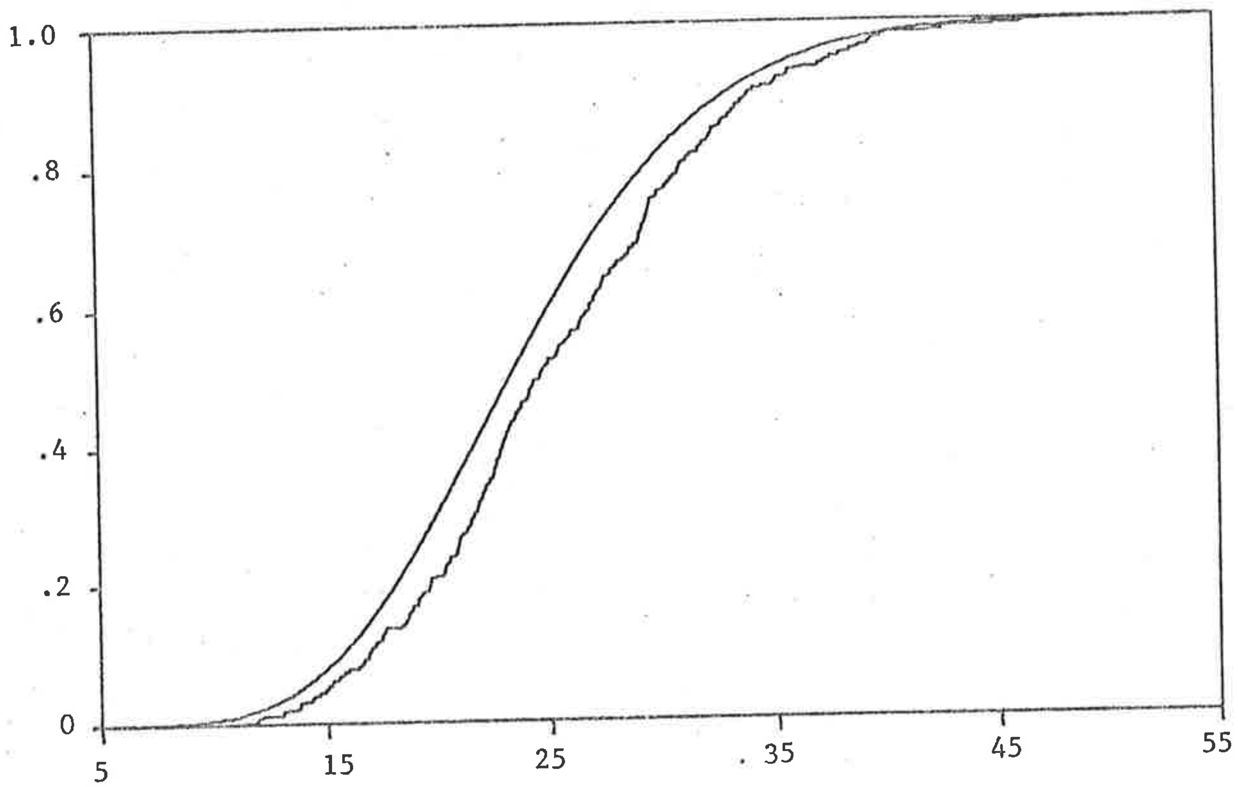

FIGURE 3 - Empirical c.d.f. for $\rho_2 T_o$

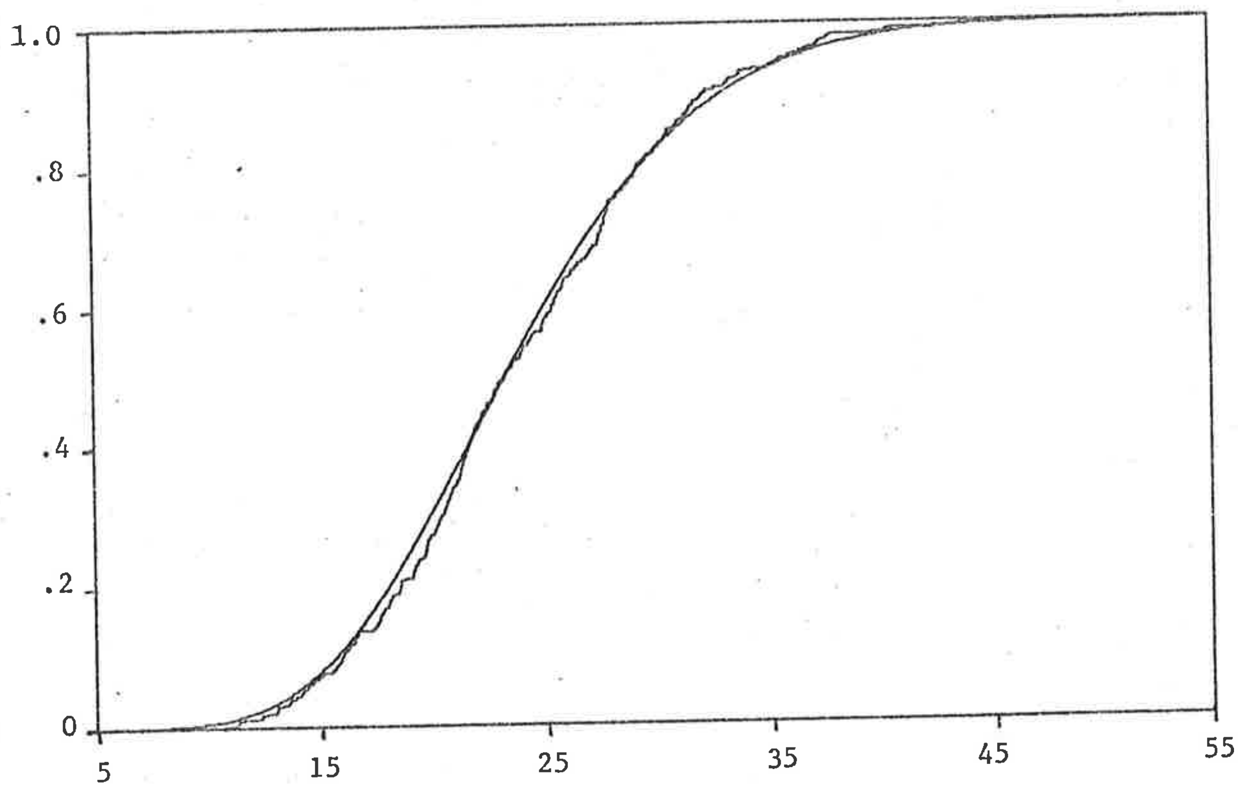

FIGURE 4 - Empirical c.d.f. for $\rho_3 T_o$

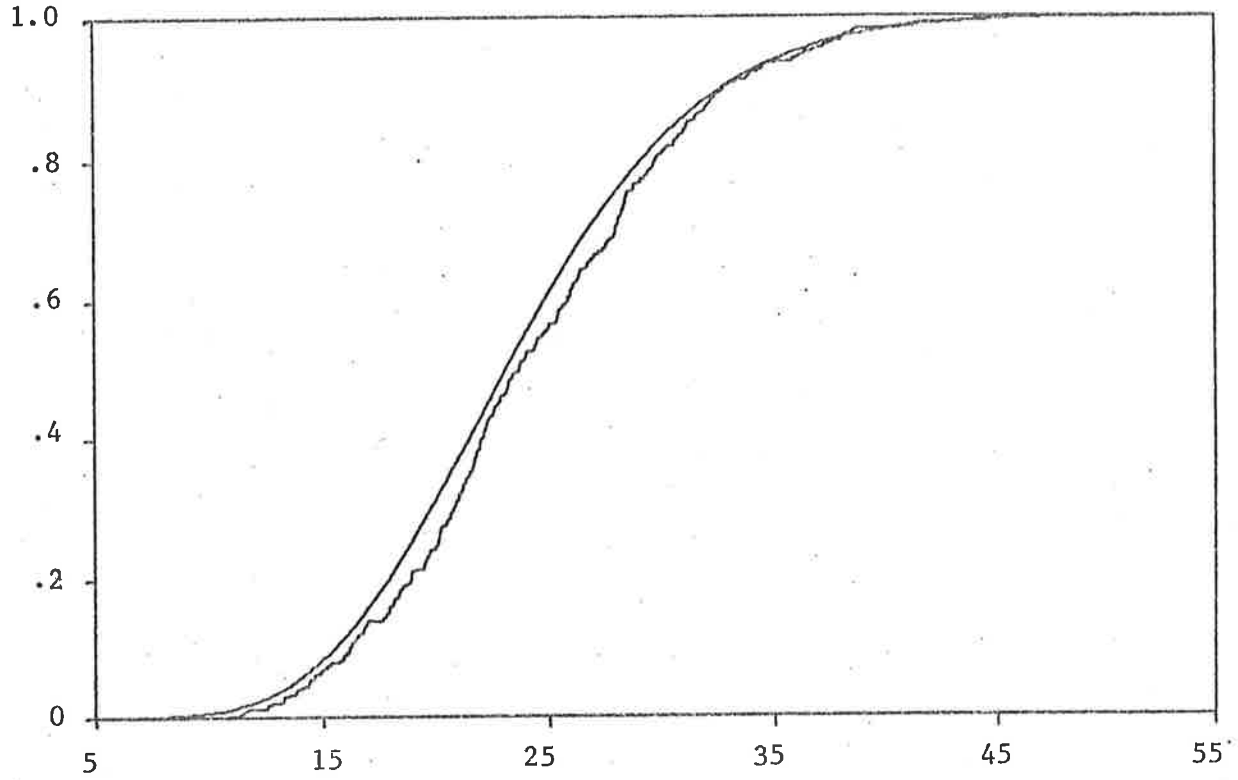

FIGURE 5 - Empirical c.d.f. for $\rho_4 T_o$

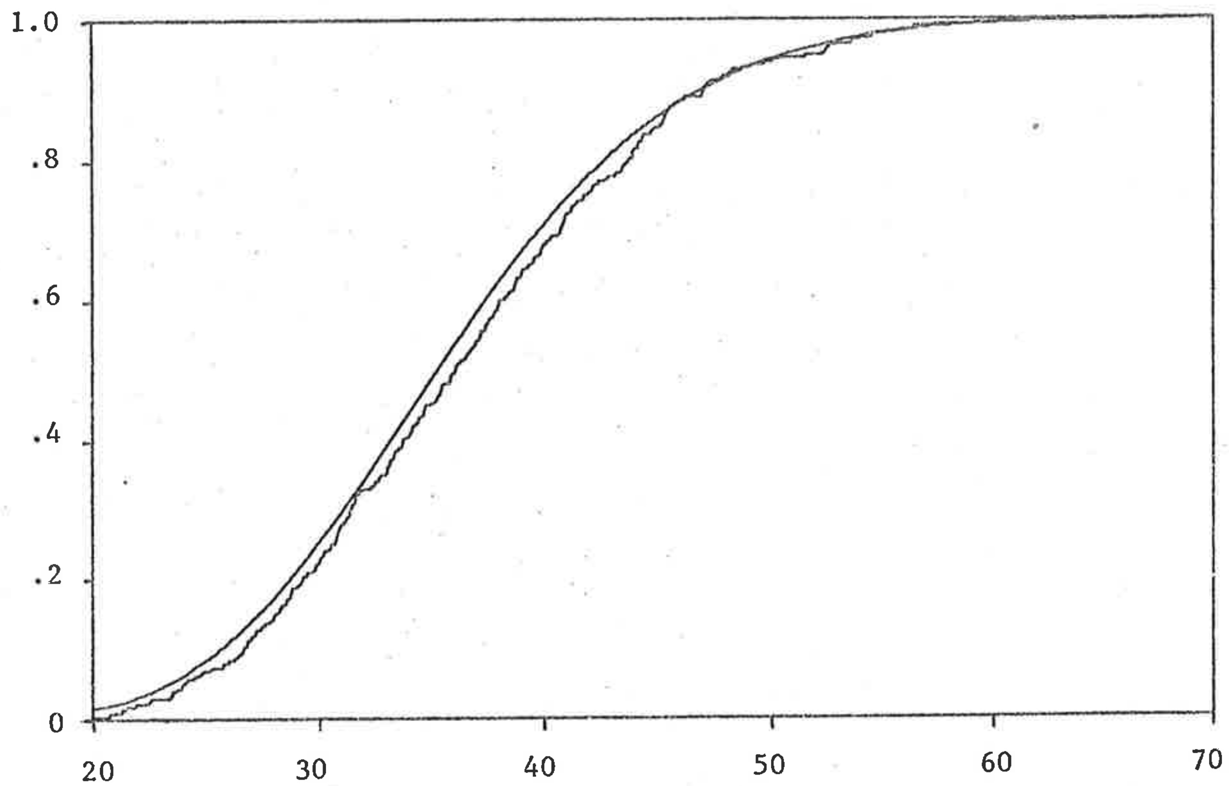

FIGURE 6 - Empirical c.d.f. for ρT

5.8 Discussion

In this chapter we have proposed the use of a direct product model of variance structure for certain types of situations, and considered some of its properties. A simple method, which has proved highly successful in applications, of determining the ML estimates of the parameters has been presented, together with some properties of the procedure and the asymptotic distribution of the estimates. We have also examined by numerical experiment ways in which the distribution of the test statistic for goodness of fit may be approximated.

In section 5.6 we have demonstrated the usefulness of the model by three numerical examples, with the direct product structure giving a convenient, meaningful description of the overall relationships between the variates through the separation of the sources of variance. In each case an acceptable fit to the data was obtained. However, for some other similar sets of data the fit was sufficiently bad to lead to the rejection of the model. This was particularly true for much of the psychology data tried, e.g. for data 3 of section 5.6 with all these assessment methods included. We note that for this particular set of data a good fit has been obtained for a restricted factor type model by Jöreskog (1971), and it may be that a direct product model is in fact inappropriate to much of this type of data. On the other hand, the reduction in the number of parameters from the unstructured alternative to the direct product structure is quite severe, and it may be that some other models which still incorporate the direct product are relevant. Further investigation seems warranted.

Of course, as an *exploratory* technique for greatly reducing the size of the interpretation problem in a useful manner, the structure may still, with caution, prove convenient in some situations where the fit would otherwise be regarded as unacceptable.

Appendix A: Properties of the matrix products $\otimes$ and $\boxtimes$

The following summarises some of the properties of the matrix products $\otimes$ and $\boxtimes$ defined in section 2.2. The properties of the direct product $\otimes$ are well known [see e.g. Anderson, 1958, Bellman, 1960, Neudecker, 1969]. Proofs of the results concerning the product $\boxtimes$ follow immediately from either the definition, or the relationship between $\otimes$ and $\boxtimes$ (see (xvi) below). The matrices A, B, C, and D are taken to be m×n, p×q, r×s, and t×u respectively, square and conforming where necessary, while the mp×mp matrix L_{mp} is defined in section 2.2.

(i) $(A \otimes B) \otimes C = A \otimes (B \otimes C)$

(ii) $(A+B) \otimes C = A \otimes C + B \otimes C; \quad A \otimes (B+C) = A \otimes B + A \otimes C$

(iii) $A \otimes B \neq B \otimes A$ in general. However

$\underset{\sim}{a} \otimes \underset{\sim}{b}' = \underset{\sim}{b}' \otimes \underset{\sim}{a} = \underset{\sim}{a}\underset{\sim}{b}'$

(iv) rank $(A \otimes B)$ = (rank A) (rank B)

(v) $(A \otimes B)' = A' \otimes B'$

(vi) $(A \otimes B)(C \otimes D) = AC \otimes BD = (\underset{\sim}{C}' \otimes I_{pm})(I_s \otimes \underset{\sim}{A}\underset{\sim}{D}' \otimes I_p)(I_{st} \otimes \underset{\sim}{B})$

(vii) $A^{-} \otimes B^{-}$ is a generalised inverse for $A \otimes B$

(viii) If $A\underset{\sim}{x} = \lambda \underset{\sim}{x}$, $B\underset{\sim}{y} = \mu \underset{\sim}{y}$, then

$(A \otimes B)(\underset{\sim}{x} \otimes \underset{\sim}{y}) = \lambda\mu(\underset{\sim}{x} \otimes \underset{\sim}{y})$

(ix) $\mathrm{tr}(A \otimes B) = (\mathrm{tr}A)(\mathrm{tr}B)$

(x) $|A \otimes B| = |A|^p \, |B|^m$

(xi) $\mathrm{vec}\,(ACB) = (B' \otimes A)\underset{\sim}{C}$

(xii) $\mathrm{tr}(C'A) = \underset{\sim}{C}'\underset{\sim}{A}$

$\mathrm{tr}(C'ADB) = \underset{\sim}{C}'(B' \otimes A)\underset{\sim}{D} = \underset{\sim}{D}'(B \otimes A')\underset{\sim}{C}$

(xiii) $L_{mp}^{-1} = L_{mp}' = L_{pm}$

(xiv) $|L_{mp}| = (-1)^{\frac{m(m-1)}{2} \cdot \frac{p(p-1)}{2}}$

(xv) $\mathrm{vec}(A') = L_{nm}\underset{\sim}{A}$

(xvi) $A \boxtimes B = L_{mp}(A \otimes B) = (B \otimes A)\, L_{nq} = L_{mp}(B \boxtimes A)\, L_{nq}$

$A \otimes B = L_{pm}(A \boxtimes B) = (B \boxtimes A)\, L_{nq} = L_{pm}(B \otimes A)\, L_{nq}$

(xvii) $\underset{\sim}{a}' \boxtimes B = \underset{\sim}{a}' \otimes B; \quad A \boxtimes \underset{\sim}{b} = \underset{\sim}{b} \otimes A;$

$\underset{\sim}{a} \boxtimes B = B \otimes \underset{\sim}{a}; \quad A \boxtimes \underset{\sim}{b}' = A \otimes \underset{\sim}{b}'$

(xviii) $(A \boxtimes B) \boxtimes C = A \boxtimes (B \boxtimes C)$

(xix) $(A+B) \boxtimes C = A \boxtimes C + B \boxtimes C$: $\quad A \boxtimes (B+C) = A \boxtimes B + A \boxtimes C$

(xx) $A \boxtimes B \neq B \boxtimes A$ in general, for any size A and B except 1×1.

(xxi) $(A \boxtimes B)' = B' \boxtimes A'$

(xxii) $(A \boxtimes B)(C \boxtimes D) = BC \otimes AD$

(xxiii) $(A \boxtimes B)(C \otimes D) = AC \boxtimes BD$

$(A \otimes B)(C \boxtimes D) = BC \boxtimes AD$

(xxiv) $B^- \boxtimes A^-$ is a generalised inverse for $A \boxtimes B$

(xxv) $|A \boxtimes B| = (-1)^{\frac{m(m-1)}{2} \frac{p(p-1)}{2}} \quad |A|^p \; |B|^m$

Proof of property (xiv):

Since L_{mp} is a permutation matrix, $|L_{mp}| = \pm 1$. We assume w.l.o.g. $m \geq p$. Then the sign is given by

$$\delta \begin{pmatrix} 1 & 2 & \cdots & m & m+1 & m+2 & \cdots & mp \\ 1 & p+1 & & (m-1)p+1 & 2 & p+2 & & mp \end{pmatrix}$$

$$= (-1)^{I(m,p)}$$

where $I(m,p)$ is the number of interchanges needed to get agreement between the first and second row by successively moving to the left the elements of the second row. Then

$$I(m,p) = (m-1) + 2(m-1) + \ldots + (p-1)(m-1) + I(m-1,p)$$

$$= (m-1)\frac{p(p-1)}{2} + I(m-1,p)$$

$$= [(m-1) + (m-2)]\frac{p(p-1)}{2} + I(m-2,p)$$

$$= \ldots$$

$$= \frac{m(m-1)}{2}\frac{p(p-1)}{2} .$$

Appendix B: Subroutines for fitting a general model of variance structure.

Foreword We include here and in Appendices C and D listings of a number of FORTRAN IV subroutines which implement procedures described in sections 2.5, 4.3 and 5.3. Some numerical routines have also been included, and except for an assumed eigenvalue/eigenvector routine (with specifications given in the VARS listing) the collection is complete in this regard.

In each of the appendices a sample calling program has been included, with sample input corresponding to examples given in the text. Some small discrepancies may appear between results obtained using the sample data and those in the text however, because of rounding to the input data.

The subroutines have been tested on the University of Adelaide CDC6400, under operating system SCOPE 3.4.1, using the FTN compiler. The coding has been written to ANSI specifications, with, for convenience and clarity, the following exceptions:

(i) the PRINT command has been used

(ii) multiple replacement statements have been used

(iii) the method of presetting data in COMMON blocks using DATA statements is apparently not in accordance with specifications.

```
      SUBROUTINE VARS
C
C PURPOSE.  TO DETERMINE BY MINIMISATION OF A GENERAL ESTIMATION
C FUNCTION THE ESTIMATES CORRESPONDING TO A  P X P  SAMPLE VARIANCE
C MATRIX S OF THE PARAMETERS GAMMA OF A GENERAL VARIANCE STRUCTURE
C MODEL SIGMA=SIGMA(GAMMA).
C
C METHOD.  AN INITIAL APPROXIMATION TO THE MINIMISING GAMMA IS
C IMPROVED USING THE METHOD OF FLETCHER AND POWELL.
C
C PROGRAMMER.  A.J.SWAIN,  JAN 1975.
C
C USAGE.  THE MAIN ARGUEMENTS ARE COMMUNICATED BETWEEN THE CALLING
C ROUTINE AND VARS THROUGH THE COMMON BLOCK
C /ARGS/NDFS    DEGREES OF FREEDOM OF S (0 IF UNKNOWN)
C        P      (INTEGER) SIZE OF S (AS LISTED .LE. 10)
C        Q      (INTEGER) NUMBER OF PARAMETERS (AS LISTED .LE.20)
C        ILSMF  DETERMINES THE ESTIMATION FUNCTION TO BE USED
C               1  TGLS - TRUE GENERALISED LEAST SQUARES
C               2  ML - MAXIMUM WISHART LIKELIHOOD
C               3  GD - GEODESIC DISTANCE
C               4  DIV - DIVERGENCE
C               5  GLS - GENERALISED LEAST SQUARES
C               6  GLSE - GENERALISED LEAST SQUARES EXTENDED
C               LOADED WITH VALUE 2 SO ML USED ON DEFAULT
C        S,NS   SAMPLE VARIANCE MATRIX (NS IS THE ROW DIMENSION OF S)
C               (PRESERVED BY VARS)
C        GA     VECTOR OF PARAMETERS - ON CALL INITIAL APPROX. TO
C               MINIMISING GAMMA - ON RETURN GIVES FINAL ESTIMATES
C        SI,NSI ON RETURN ESTIMATE OF SIGMA
C        X2     APPROX. CHISQUARE TEST STATISTIC FOR GOODNESS OF FIT
C        NDFM   ON RETURN NOMINAL DEGREES OF FREEDOM OF MODEL (AND X2)
C        NRP    ON RETURN SUSPECTED NUMBER OF REDUNDANT PARAMETERS
C               (SEE NOTE4)
C        IE     ERROR INDICATOR
C               0 SUCESSFUL CONVERGENCE TO ESTIMATES
C               1 FAILURE TO CONVERGE IN SPECIFIED MAX. NO. OF ITERNS.
C               2-5 FAILURE OF FLETCHER-POWELL METHOD TO LOCATE MINIMUM
C                 CODES CORRESPOND TO THE FLEPO CODES (SEE LISTING)
C               6 S NOT POSITIVE DEFINITE
C
C THE USER IS ALSO REQUIRED TO PROVIDE A SUBROUTINE SIGA(IOD) TO
C EVALUATE SIGMA(GAMMA) AND THE MATRIX OF FIRST DERIVATIVES
C AT THE POINT GA GIVEN IN /ARGS/.
C IOD SPECIFIES THE ORDER OF DERIVATIVE REQUIRED
C IOD=0  SIGMA(GAMMA) TO BE RETURNED IN SI OF /ARGS/
C IOD=1  DSIGMA/DGAMMA  (P*(P+1)/2 X Q)  TO BE RETURNED IN ARRAY D2 OF
C THE COMMON BLOCK /WORK2/ - SEE DECLARATION
C ERROR CONDITIONS, E.G. PARAMETERS OUT OF RANGE, SHOULD BE INDICATED
C BY SIGA BY SETTING F=FMAX, WHERE F AND FMAX ARE THE LOCATIONS IN THE
C COMMON BLOCK /WFUNCT/ - SEE BELOW
C
C VARS IS DESIGNED SO THAT THE FLETCHER-POWELL ITERATIONS MAY BE
C PRECEEDED BY STEEPEST DESCENT ITERATIONS.  THE CONTROLLING PARAMETERS
C ARE CONTAINED IN THE COMMON BLOCK
C /ITERL/MSD      MAXIMUM NUMBER OF STEEPEST DESCENT ITERATIONS
C        MFP      MAXIMUM NUMBER OF FLETCHER-POWELL ITERATIONS
C        EPSSD    CONVERGENCE PARAMETER FOR STEEPEST DESCENT
C        EPSFP    CONVERGENCE PARAMETER FOR FLETCHER-POWELL
C
C OTHER ARGUEMENTS AVAILABLE ON RETURN ARE CONTAINED IN THE COMMON BLOCK
C /WFUNCT/FMAX  USED WITHIN MINIMISATION ROUTINES TO INDICATE ERRORS
C         NP     =P
C         X      =GA
C         F      VALUE OF F(SIGMA,S)AT X
C         G      VALUE OF DERIVATIVES OF F AT X
C         H,NH   APPROXIMATION TO INVERSE OF MATRIX OF SECOND DERIVATIVES
C                OF F AT X.  NOTE THAT (2./NDFS)*H ESTIMATES THE
C                ASYMPTOTIC VARIANCE MATRIX OF GAMMA.
C
C THE COMMON BLOCK /WMINR/ IS USED BY VARS FOR COMMUNICATION WITH THE
C MINIMISATION ROUTINE.  PRINTING BY THIS ROUTINE MAY BE STOPPED BY
```

```
C SETTING IPNT TO ZERO
C
C NOTE1.  THE METHOD USED TO EVALUATE THE DERIVATIVES IS DEPENDENT ON
C THE ESTIMATION FUNCTION USED.  FOR GENERAL ESTIMATION FUNCTIONS THE
C GENERAL FORMULAE OF SECTION 2.5 ARE USED, AND REQUIRE EIGENVALUE
C CALCULATION.  FOR ML AND GLS THE FORMULAE OF TABLE 2 ARE USED, AND
C REQUIRE NO EIGENVALUES.  THUS ML AND GLS ESTIMATION COULD BE EXPECTED
C TO BE NOTICABLY FASTER THAN FOR GENERAL FUNCTIONS.
C
C NOTE2.  FOR EIGENVALUE AND EIGENVECTOR DETERMINATION VARS ASSUMES
C A SUBROUTINE EISSOL WITH SPECIFICATIONS GIVEN BELOW
C SUBROUTINE EISSOL(N,A,NA,D,Z,NZ,IERR) - A SUBROUTINE TO DETERMINE THE
C EIGENVALUES AND EIGENVECTORS OF A REAL SYMMETRIC N X N MATRIX A.
C ARGUEMENTS.  N      SIZE OF A
C              A,NA   MATRIX A
C              D      ARRAY FOR EIGENVALUES - RETURNED IN INCR. ORDER.
C                     IF CALLED WITH N=-N, RETURNED IN DECR. ORDER
C              Z,NZ   ARRAY FOR CORRESPONDING EIGENVECTORS
C                     MAY BE THE SAME AS A, OTHERWISE A IS PRESERVED
C              IERR   ERROR INDICATOR. 0=NO ERROR.
C
C NOTE3.  IN ADDITION TO THE USER-SUPPLIED SUBROUTINES SIGA AND EISSOL
C DESCRIBED ABOVE THE FOLLOWING SUBROUTINES ARE USED BY VARS AND EACH
C OTHER - VAFN, DFDGA, RDPA, IDENT, SMXM, FLEPO, DOT, GREAT, CHOL
C EACH OF THESE IS LISTED BELOW, WITH PURPOSE INDICATED.
C WORKING STORAGE IS CONTAINED IN THE COMMON BLOCKS /WORK1/, /WORK2/ AND
C /WORK3/ WITH, AS LISTED, 212, 1156 AND 212 LOCATIONS RESPECTIVELY.
C THESE AREAS COULD BE USED ARBITARILY OUTSIDE VARS (THOUGH NOT BY SIGA)
C
C NOTE4.  THE IDENTIFICATION OF THE MODEL IS CHECKED BY DETERMINING THE
C RANK, R SAY, OF DSIGMA/DGAMMA AT THE INITIAL GAMMA.  THE ARGUEMENT NRP
C =Q-R.  THIS WOULD COMMONLY GIVE THE NO. OF REDUNDANT PARAMETERS, BUT,
C BEING DEPENDENT ON THE INITIAL GAMMA, MAY NOT ALWAYS
C WHILE THE SUBROUTINES SHOULD WORK SATISFACTORALLY FOR THE
C NON-IDENTIFIED CASE, RESULTS SHOULD BE EXAMINED CLOSELY WHENEVER NRP
C IS NONZERO (CERTAINLY NO ADJUSTMENT IS MADE TO NDFM)
C
C NOTE5.  TO EXTEND THE RANGE OF P, IT IS NECESSARY TO CHANGE THE
C DIMENSIONS IN /ARGS/ IN ALL SUBROUTINES WHERE IT APPEARS AND THE DATA
C STATEMENT OF VARS.  SIMILARLY THE DIMENSIONS IN /WORK1/, /WORK2/ AND
C /WORK3/ AND THE ASSIGNMENT STATEMENTS IN VAFN
C
C NOTE6.  TO EXTEND THE NUMBER OF PARAMETERS IT IS NECESSARY TO
C EXTEND DIMENSIONS IN /WORK2/ (ASSIGNMENTS IN VAFN), /WFUNCT/ (DATA
C STATEMENT IN FLEPO), /ARGS/ AND THE WORKING ARRAYS IN FLEPO (SEE FLEPO
C LISTING)
C
C NOTE7.  TO INCLUDE OTHER ESTIMATION FUNCTIONS IT IS ONLY NECESSARY TO
C INCLUDE THE EVALUATING FUNCTIONS IN SMF AND SMFD, AND TO ALTER THE
C COMPUTED GO TO STATEMENTS IN THESE SUBROUTINES AND VAFN.
C
      COMMON /ARGS/NDFS,P,Q,ILSMF,S(10,10),NS,GA(20),SI(10,10),NSI,
     1             X2,NDFM,NRP,IE
      INTEGER P,Q
      COMMON /ITERL/MSD,MFP,EPSSD,EPSFP
      COMMON /WFUNCT/FMAX,NX,X(20),F,G(20),H(20,21),NH
      COMMON IOD
      COMMON /WMINR/EPS,MITER,IPNT,IERR,ITER,LOADO,NFCT
      DATA ILSMF,NS,NSI/2,10,10/
      DATA MSD,MFP,EPSSD,EPSFP/0,100,1.E-02,1.E-06/
      DATA IPNT/1/
C
C INITIALISATION
C
      IE=0
      NX=Q
      DO 30 I=1,Q
   30 X(I)=GA(I)
      IOD=-3
      CALL VAFN
      IF(F.NE.FMAX)GO TO 40
      IE=6
```

```
      GO TO 9999
   40 IOD=1
      CALL VAFN
      CALL IDENT
      IF(MSD.EQ.0)GO TO 80
C
C COMMENCE STEEPEST DESCENT ITERATIONS
C
      EPS=EPSSD
      MITER=MSD
      LOADO=4
      CALL FLEPO(VAFN)
      IE=IERR
   80 IF(MFP.EQ.0)GO TO 130
C
C COMMENCE FLETCHER-POWELL ITERATIONS
C
      IOD=-2
      CALL VAFN
      CALL CHOL(4,Q,Q,H,NH,DH)
      IOD=1
      EPS=EPSFP
      MITER=MFP
      LOADO=2
      CALL FLEPO(VAFN)
      IE=IERR
C
C FORM CHISQUARE TEST STATISTIC AND RETURN
C
  130 NDFM=P*(P+1)/2-Q
      X2=0.
      IF(NDFS.LE.0)GO TO 9999
      RN=NDFS
      X2=RN*F
 9999 RETURN
      END
```

****

```
      SUBROUTINE VAFN
C
C PURPOSE.  CALLED BY VARS, AND BY MINIMISATION ROUTINES TO EVALUATE
C F, G AND H FOR THE GENERAL MODEL SIGMA(GAMMA), USING THE FORMULAE OF
C SECTION 2.5 FOR F AND G, AND THE LIMIT AS AN APPROXIMATION TO H
C
C WHAT RETURNED IS DEPENDENT ON THE VALUE OF IOD
C 0 = JUST F,  1 = F AND G,  2 = F, G AND H,  -1 = JUST G,  -2 = JUST H
C -3 = INITIALISATION (MUST PRECEED ALL OTHER CALLS)
C IOD=-1 ASSUMES THE RESULT OF AN IOD=1 CALL
C IOD=-2 ASSUMES THE RESULT OF AN IOD=1 CALL
C
C NOTE.  METHOD USED IS DEPENDENT ON ILSMF,  IF ILSMF=2 OR 5 (ML OR GLS)
C THEN THE PARTICULAR FORMULAE OF TABLE 2 ARE USED.  OTHERWISE THE
C GENERAL FORMS OF SECTION 2.5 (WHICH REQUIRE EIGEN CALN.) ARE USED.
C
      COMMON IOD
      COMMON /ARGS/NDFS,P,Q,ILSMF,S(10,10),NS,GA(20),SI(10,10),NSI,
     1             X2,NDFM,NRP,IE
      INTEGER P,Q
      COMMON /WFUNCT/FMAX,NX,X(20),F,G(20),H(20,21),NH
      COMMON /WORK1/A(10,10),NA,B(10,10),NB,C(10)
      COMMON /WORK2/D2(55,20),ND2,E2(55)
      COMMON /WORK3/D(10,10),ND,E(10),U(10,10),NU
      DO 10  I=1,Q
   10 GA(I)=X(I)
      JOD=IOD+4
      GO TO (280 ,240 ,120,20 ,20 ,20 ),JOD
C
C CALCULATE F
C
   20 F=0.
      CALL SIGA(0)
      IF(F.EQ.FMAX)GO TO 9999
      GO TO (30,50 ,30,30,90 ,30),ILSMF
C
```

```
C GENERAL FUNCTION - DETERMINE THE EIGEN VALUES AND EIGEN VECTORS OF
C (S**-.5)*SI*(S**-.5) IN C AND A RESP. AND COMPUTE F
C
   30 CALL SMXM(P,P,SI,NSI,D,ND,U,NU)
      CALL SMXM(P,P,D,ND,U,NU,B,NB)
      CALL EISSOL(P,B,NB,C,A,NA,IERR)
      IF(IERR.NE.0)GO TO 360
      F=0.
      DO 40 I=1,P
   40 F=F+SMF(C(I),ILSMF)
      GO TO 110
C
C ML - SET A TO SI**-1, B TO I - S*SI**-1 AND COMPUTE F
C
   50 DO 60  I=1,P
      DO 60  J=I,P
   60 A(I,J)=A(J,I)=SI(I,J)
      CALL CHOL(4,P,P,A,NA,DSI)
      IF(DSI.LE.0)GO TO 360
      CALL SMXM(P,P,S,NS,A,NA,B,NB)
      TR=0.
      DO 80  I=1,P
      DO 70  J=1,P
   70 B(I,J)=-B(I,J)
      TR=TR-B(I,I)
   80 B(I,I)=B(I,I)+1.
      F=TR+ALOG(DSI)+FK
      GO TO 110
C
C GLS - SET B TO SI*S**-1 - I, AND COMPUTE F
C
   90 CALL SMXM(P,P,SI,NSI,A,NA,B,NB)
      TR=0.
      DO 100 I=1,P
      B(I,I)=B(I,I)-1.
      DO 100 J=1,P
  100 TR=TR+B(I,J)*B(J,I)
      F=.5*TR
  110 GO TO (9999,240,120,9999,120,120),JOD
C
C CALCULATE DF/DSI AND THENCE G
C
  120 GO TO (130,180,130,130,180,130),ILSMF
C
C GENERAL FUNCTION - U=(S**-.5)*A,  B=U*(DIAG(FD(C(I))))*UT
C WITH OFF-DIAGONALS MULTIPLIED BY 2.
C
  130 DO 140 I=1,P
  140 E(I)=SMFD(C(I),ILSMF)
      CALL SMXM(P,P,D,ND,A,NA,U,NU)
      DO 170 I=1,P
      DO 160 J=1,I
      T=0.
      DO 150 K=1,P
  150 T=T+U(I,K)*E(K)*U(J,K)
  160 B(I,J)=B(J,I)=2.*T
  170 B(I,I)=T
      GO TO 220
C
C ML - B = (SI**-1)*B WITH OFF-DIAGONALS MULTIPLIED BY 2.
C GLS - B = (S**-1)*B WITH OFF-DIAGONALS MULTIPLIED BY 2.
C
  180 DO 210 J=1,P
      DO 200 I=1,P
      T=0.
      DO 190 K=1,P
  190 T=T+A(I,K)*B(K,J)
      IF(I.NE.J)T=T*2.
  200 C(I)=T
      DO 210 I=1,P
  210 B(I,J)=C(I)
  220 CALL DFDGA(1)
```

```
  230 GO TO (9999,240,9999,9999,9999,240),JOD
C
C COMPUTE APPROXIMATION TO SECOND DERIVATIVES
C
  240 GO TO (250 ,270 ,250 ,250 ,270 ,250 ),ILSMF
  250 DO 260  I=1,P
      DO 260  J=1,I
  260 A(I,J)=A(J,I)=SI(I,J)
      CALL CHOL(4,P,P,A,NA,DA)
  270 CALL DFDGA(2)
      GO TO 9999
C
C INITIALISATION OF WORKING ARRAYS, CHECK S IS POSITIVE DEFINITE
C
  280 NA=NB=10
      ND2=55
      ND=NU=10
      F=0.
      DO 290 I=1,P
      DO 290 J=I,P
  290 A(I,J)=A(J,I)=S(I,J)
      CALL CHOL(1,P,P,A,NA,DS)
      IF(DS.LE.0)GO TO 360
      GO TO (300,340,300,300,350,300),ILSMF
C
C GENERAL FUNCTIONS D = (S**-.5)
C
  300 CALL EISSOL(P,S,NS,C,B,NB,IERR)
      IF(IERR.NE.0)GO TO 360
      DO 310 I=1,P
      IF(C(I).LE.0.)GO TO 360
  310 C(I)=1./SQRT(C(I))
      DO 330 I=1,P
      DO 330 J=1,I
      T=0.
      DO 320 K=1,P
  320 T=T+B(I,K)*C(K)*B(J,K)
  330 D(I,J)=D(J,I)=T
      GO TO 9999
C
C ML - DETERMINE FIXED PART OF F
C
  340 RP=P
      FK=(-ALOG(DS)-RP)
      GO TO 9999
C
C GLS - A=(S**-1)
C
  350 CALL CHOL(6,P,P,A,NA,DS)
      GO TO 9999
  360 F=FMAX
 9999 RETURN
      END

*****************************************************************************

      SUBROUTINE DFDGA(IOD)
C
C PURPOSE.  TO EVALUATE THE EXPRESSION FOR FIRST DERIVATIVES OF F
C WITH RESPECT TO THE ELEMENTS OF GAMMA (IOD=1), AND TO EVALUATE AN
C EXPRESSION TO APPROXIMATE THE MATRIX OF SECOND DERIVATIVES (IOD=2)
C
C IOD=1 ASSUMES THAT THE LOWER TRIANGLE OF B CONTAINS DF/DSI
C IOD=2 ASSUMES THAT A CONTAINS EITHER S**-1 OR SI**-1, AND THAT D
C CONTAINS DSI/DGA, I.E. USUALLY PRECEEDED BY AN IOD=1 CALL.
C
      COMMON /ARGS/NDFS,P,Q,ILSMF,S(10,10),NS,GA(20),SI(10,10),NSI,
     1             X2,NDFM,NRP,IE
      INTEGER P,Q
      COMMON /WFUNCT/FMAX,NX,X(20),F,G(20),H(20,21),NH
      COMMON /WORK1/A(10,10),NA,B(10,10),NB,C(10)
      COMMON /WORK2/D(1100),ND,E(55)
```

```
      INTEGER PC
C
      IF(IOD.EQ.2)GO TO 25
C
C CALCULATE FIRST DERIVATIVES
C
      PC=P*(P+1)/2
      CALL SIGA(1)
      DO 20 K=1,NX
      ID=ND*(K-1)
      T=0.
      DO 10 J=1,P
      DO 10 I=J,P
      ID=ID+1
   10 T=T+B(I,J)*D(ID)
   20 G(K)=T
      GO TO 9999
C
C APPROXIMATE SECOND DERIVATIVES
C
   25 DO 60 N=1,NX
      ID=ND*(N-1)
      DO 30 IE=1,PC
   30 E(IE)=0.
      DO 40 L=1,P
      DO 40 K=L,P
      ID=ID+1
      U=D(ID)
      IF(U.EQ.0)GO TO 40
      IE=0
      DO 35 J=1,P
      DO 35 I=J,P
      IE=IE+1
   35 E(IE)=E(IE)+RDPA(I,J,K,L)*U
   40 CONTINUE
      DO 60 M=N,NX
      ID=ND*(M-1)
      T=0.
      DO 50 I=1,PC
   50 T=T+D(ID+I)*E(I)
   60 H(M,N)=H(N,M)=T
 9999 RETURN
      END
```

****

```
      FUNCTION RDPA(I,J,K,L)
C
C PURPOSE.  TO RETURN THE IJ KL ELEMENT OF THE REDUCED DIRECT PRODUCT
C OF MATRIX A WITH ITSELF, WHERE A IS STORED IN THE ARRAY A.
C
      COMMON /WORK1/A(10,10),NA,B(10,10),NB,C(10)
      IF((I.EQ.J).OR.(K.EQ.L))GO TO 10
      RDPA=2.*(A(I,K)*A(J,L)+A(J,K)*A(I,L))
      GO TO 9999
   10 IF((I.EQ.J).AND.(K.EQ.L))GO TO 20
      RDPA=2.*A(I,K)*A(J,L)
      GO TO 9999
   20 RDPA=A(I,K)*A(I,K)
 9999 RETURN
      END
```

****

```
      SUBROUTINE IDENT
C
C PURPOSE.  TO PROVIDE A CHECK ON THE IDENTIFICATION OF THE PARAMETERS
C BY CHECKING THE RANK OF DSI/DGA.  ON RETURN NRP IS THE DIFFERENCE
C Q - RANK(DSI/DGA).  ASSUMES PRECEEDED BY A CALL OF SIGA(1)
C
C NOTE1.  DETERMINES NRP AS THE NUMBER OF ALMOST ZERO PIVOTS IN THE
C CHOLESKY DECOMPOSITION
```

```
C NOTE2.  USES DOT AND CHOL.   ALSO DESTROYS H OF /WFUNCT/
C
      COMMON /ARGS/NDFS,P,Q,ILSMF,S(10,10),NS,GA(20),SI(10,10),NSI,
     1               X2,NDFM,NRP,IE
      INTEGER P,Q
      COMMON /WFUNCT/FMAX,NX,X(20),F,G(20),H(20,21),NH
      COMMON /WORK2/D(55,20),ND,E(55)
C
      IPC=P*(P+1)/2
      DO 10 I=1,NX
      DO 10 J=1,I
   10 H(I,J)=H(J,I)=DOT(IPC,D(1,I),D(1,J))
      CALL CHOL(1,NX,NX,H,NH,DH)
      NRP=0
      IF(DH.GT.0.)GO TO 9999
      DO 20 I=1,NX
   20 IF(H(I,I).EQ.0.)NRP=NRP+1
 9999 RETURN
      END
*****************************************************************************

      FUNCTION SMF(T,ILSMF)
C
C PURPOSE.  CALLED BY VAFN TO EVALUATE F(THETA) AT THETA=T,
C WITH THE PARTICULAR F USED BEING DETERMINED BY ILSMF - SEE VARS LIST
C
      GO TO(10,20,30,40,50,60),ILSMF
   10 SMF=.5*(T-1.)*(T-1.)/(T*T)
      GO TO 9999
   20 SMF=1./T+ALOG(T)-1.
      GO TO 9999
   30 SMF=.5*ALOG(T)**2
      GO TO 9999
   40 SMF=.5*(T+1./T-2.)
      GO TO 9999
   50 SMF=.5*(T-1.)*(T-1.)
      GO TO 9999
   60 SMF=.5*(T*T-2.*T+1.)*EXP(T-1.)
 9999 RETURN
      END
*****************************************************************************

      FUNCTION SMFD(T,ILSMF)
C
C PURPOSE.  CALLED BY VAFN TO EVALUATE 1ST DERIV. OF F(THETA) AT T,
C WITH THE PARTICULAR F USED BEING DETERMINED BY ILSMF - SEE VARS LIST
C
      GO TO (10,20,30,40,50,60),ILSMF
   10 SMFD=(T-1.)/(T*T*T)
      GO TO 9999
   20 SMFD=(T-1.)/(T*T)
      GO TO 9999
   30 SMFD=ALOG(T)/T
      GO TO 9999
   40 SMFD=.5*(1.-1./(T*T))
      GO TO 9999
   50 SMFD=T-1.
      GO TO 9999
   60 SMFD=.5*(T*T-1.)*EXP(T-1.)
 9999 RETURN
      END
*****************************************************************************

      SUBROUTINE SMXM(M,N,A,NA,B,NB,C,NC)
C
C PURPOSE.  CALLED BY VAFN TO MULTIPLY  M X N MATRIX B BY THE  M X M
C SYMMETRIC MATRIX A  AND STORE RESULT IN C.  I.E. C = A*B - C MUST BE
C DIFFERENT FROM BOTH A AND B
C
```

```
      REAL A(NA,M),B(NB,N),C(NC,N)
      DO 10 I=1,M
      DO 10 J=1,N
   10 C(I,J)=DOT(M,A(1,I),B(1,J))
      RETURN
      END
```

****

```
      SUBROUTINE CHOL(IO,M,N,A,NA,DA)
C
C PURPOSE.  TO DETERMINE USING THE CHOLESKY METHOD (SQUARE ROOT METHOD)
C THE SOLUTION OF A SYSTEM OF LINEAR EQUATIONS  B*X = Y, WHERE B IS A
C SYMMETRIC, POSITIVE SEMIDEFINITE  M X M  MATRIX AND Y IS  M X (N-M).
C ALTERNATIVELY ON OPTION TO DETERMINE THE TRIANGULAR DECOMPOSITION
C B = L*LT, WHERE L IS LOWER TRIANGULAR WITH TRANSPOSE LT AND G-INVERSE
C LINV, AND/OR A GENERALISED INVERSE, BINV, FOR B.
C
C PROGRAMMER.  A.J.SWAIN  JAN 1975.
C
C ARGUEMENTS.  IO  OPTION
C                  1 ON CALL A = (B Y)          ON RETURN A = (LT LINV*Y)
C                  2 ON CALL A = (B Y)          ON RETURN A = (LT X)
C                  3 ON CALL A = (LT LINV*Y)    ON RETURN A = (LT X)
C                  4 ON CALL A = (B Y)          ON RETURN A = (BINV X)
C                  5 ON CALL A = (LT LINV*Y)    ON RETURN A = (BINV X)
C                  6 ON CALL A = (LT X)         ON RETURN A = (BINV X)
C              M   NUMBER OF ROWS OF B
C              N   NUMBER OF COLUMNS OF (B X) - IF N=M ONLY LT AND
C                  BINV ARE CALCULATED
C              A   AS ABOVE
C                  NA  ROW DIMENSION OF A
C              DB  DETERMINANT OF B - CALCULATED FOR IO=1,2 OR 4.
C                  RETURNED AS -1 IF B IS NOT POSITIVE SEMIDEFINITE
C
C NOTE1.  B IS TAKEN AS SINGULAR IF THE ABSOLUTE VALUE OF A DIAGONAL
C ELEMENT IN THE DECOMPOSITION IS LESS THAN EPS.  IN THIS CASE THERE IS
C NO CHECK ON THE CONSISTENCY OF THE EQUATIONS, SO THAT IF THE EQUATIONS
C ARE NOT CONSISTENT SO NO SOLUTION EXITS, NO INDICATION IS GIVEN OF
C THIS.
C NOTE2.  ONLY THE UPPER TRIANGLE OF B IS USED IN COMPUTATION (ALTHOUGH
C BINV IS MADE SYMMETRIC) AND ON RETURN FROM IO=1,2 OR 3 THE LOWER
C TRIANGLE OF B IS PRESERVED.
C
      REAL A(NA,N)
      DATA EPS/1.E-10/
C
      GO TO(9,9,80,9,80,130),IO
C
C FIRST PASS - FORM LT AND DO THE SAME OPERATIONS ON Y TO GIVE LINV*Y.
C IF A DIAGONAL ELEMENT IN THE DECOMPOSITION IS ALMOST ZERO, SET IT AND
C THE REST OF THE ROW TO ZERO.
C
    9 DA=1.
      DO 70 I=1,M
      IM1=I-1
      DO 40 J=I,N
      T=0.
      IF(I.EQ.1) GO TO 20
      DO 10 K=1,IM1
   10 T=T+A(K,I)*A(K,J)
   20 IF(J.NE.I)GO TO 30
      T=A(I,I)-T
      IF(ABS(T).LT.EPS) GO TO 50
      IF(T.GT.0.)GO TO 25
      DA=-1.
      GO TO 9999
   25 DA=DA*T
      A(I,I)=SQRT(T)
      GO TO 40
   30 A(I,J)=(A(I,J)-T)/A(I,I)
   40 CONTINUE
```

```
      GO TO 70
   50 DA=0.
      DO 60 J=I,N
   60 A(I,J)=0.
   70 CONTINUE
C
   80 GO TO(9999,90,90,90,90,130),IO
C
C COMPLETE THE SOLUTION OF THE EQUATIONS.
C
   90 IF(N.EQ.M) GO TO 130
      MP1=M+1
      DO 120 I1=1,M
      I=M-I1+1
      IP1=I+1
      IF(A(I,I).EQ.0.) GO TO 120
      DO 110 J=MP1,N
      T=0.
      IF(I.EQ.M) GO TO 110
      DO 100 K=IP1,M
  100 T=T+A(I,K)*A(K,J)
  110 A(I,J)=(A(I,J)-T)/A(I,I)
  120 CONTINUE
C
  130 GO TO(9999,9999,9999,140,140,140),IO
C
C SECOND PASS - INVERT LT.
C
  140 DO 160 I1=1,M
      I=M-I1+1
      IP1=I+1
      IF(A(I,I).EQ.0.) GO TO 160
      A(I,I)=1./A(I,I)
      DO 155 J1=I,M
      J=M+I-J1
      IF(J.EQ.I) GO TO 160
      T=0.
      DO 150 K=IP1,J
  150 T=T+A(I,K)*A(K,J)
  155 A(I,J)=(-T)*A(I,I)
  160 CONTINUE
C
C THIRD PASS - FORM BINV = LTINV*LINV.
C
      DO 180 I=1,M
      DO 180 J=I,M
      T=0.
      DO 170 K=J,M
  170 T=T+A(I,K)*A(J,K)
  180 A(I,J)=A(J,I)=T
 9999 RETURN
      END
```

```
*************************************************************************************

      SUBROUTINE FLEPO(FUNCT)
C
C PURPOSE.  TO MINIMISE A FUNCTION F OF N VARIABLES X(1) ... X(N)
C USING THE METHOD OF FLETCHER AND POWELL (1963) OR ON OPTION THE
C METHOD OF STEEPEST DESCENT.
C
C PROGRAMMER.  A.J.SWAIN, JAN 1975
C CODING ORIGINALLY BASED ON WELLS,M.(1965) ALGORITHM 251. COMM. ACM, 8,
C 169-170, BUT SINCE EXTENSIVELY MODIFIED.
C
C USAGE.  THE ARGUEMENT FUNCT OF THE CALL IS A USER SUPPLIED SUBROUTINE
C (DECLARED EXTERNAL IN THE ROUTINE CALLING FLEPO) WHICH WHEN CALLED
C EVALUATES F AND G (THE VECTOR OF FIRST DERIVATIVES) AT A GIVEN POINT
C X.  THE NECESSARY ARGUEMENTS ARE TRANSFERED BETWEEN THE CALLING
C ROUTINE, FLEPO AND FUNCT THROUGH THE COMMON BLOCK
C /WFUNCT/ FMAX          AN UNSATISFACTORY X CAN BE INDICATED TO FLEPO BY
C                        FUNCT BY RETURNING F=FMAX (LISTED VALUE 1.E+75)
```

```
C           N            NUMBER OF VARIABLES (.LE.MN,SAY - AS LISTED MN=20)
C           X(MN)        CURRENT X
C           F            VALUE OF F AT X
C           G(MN)        VALUE OF FIRST DERIVATIVES OF F AT X
C           H(MN,MN+1)   FOR A FLETCHER-POWELL MINIMISATION CONTAINS AN
C                        APPROXIMATION TO THE INVERSE OF THE MATRIX OF
C                        SECOND DERIVATIVES AT X
C           NH           ROW DIMENSION OF H (=MN)
C WHEN FLEPO CALLED, X IS TAKEN AS AN APPROXIMATION TO THE MINIMISING X,
C WITH F AND G GIVING THE CORRESPONDING VALUES, I.E. A CALL TO FUNCT
C MUST PRECEDE THE CALL TO FLEPO.
C
C OTHER ARGUEMENTS ARE TRANSFERED BETWEEN THE CALLING ROUTINE AND FLEPO
C THROUGH THE COMMON BLOCK
C /WMINR/ EPS    ITERATION IS TERMINATED WHEN MAX(ABS(G(I))).LE.EPS,
C                E.G. EPS=1.E-06
C         MITER  MAXIMUM NUMBER OF ITERATIONS
C         IPNT   PRINT OPTION  0=NO PRINTING, 1=PRINTING
C         IERR   ERROR INDICATOR -
C                0  SUCCESSFUL CONVERGENCE
C                1  FAILURE TO CONVERGE IN MITER ITERATIONS
C                2,3,4  FAILURE TO LOCATE MINIMUM IN CURRENT DIRECTION
C                5 FUNCTION IS NOT DECREASING IN NEXT DIRECTION
C         ITER   NUMBER OF ITERATIONS REQUIRED
C         LOADO  CALL OPTION ON H
C                1 SET H TO I
C                2 APPROXIMATION TO H PROVIDED BY CALLER
C                3 SINGLE MINIMISATION ALONG THE DIRECTION SPECIFIED BY
C                  THE FIRST COLUMN OF H REQUIRED
C                4 STEEPEST DESCENT METHOD REQUIRED - IN WHICH CASE H IS
C                  NOT USED
C
C NOT NORMALLY REQUIRING CHANGE, BUT ALSO ACCESSIBLE TO THE CALLING
C ROUTINE ARE THE PARAMETERS IN THE COMMON BLOCK
C /PARAFP/ PROPCH MAX.RELATIVE CHANGE IN ANY ELEMENT OF X AT THE START
C                 OF EACH ITERATION
C          STR    STEP REDUCTION PARAMETER
C          EPSS   MINIMUM VALUE FOR A MODIFIED STEP SIZE
C
C NOTE1.  REQUIRES FUNCTIONS DOT(N,A,B) AND GREAT(N,A) WHICH FOLLOW ON
C LISTING.
C NOTE2.  TO CHANGE MN IT IS NECESSARY TO MODIFY THE DIMENSIONS OF THE
C ARRAYS OF /WFUNCT/ AND OF THE WORKING ARRAYS E, SIGMA AND GAMMA.
C
      COMMON /WFUNCT/FMAX,N,X(20),F,G(20),H(20,21),NH
      COMMON /WMINR/EPS,MITER,IPNT,IERR,ITER,LOADO,NFCT
      COMMON /PARAFP/PROPCH,STR,EPSS
      REAL E(20),SIGMA(20),GAMMA(20)
      DATA FMAX,NH/1.E+75,20/
      DATA PROPCH,STR,EPSS/.3,.7,1.E-12/
      DATA ENTFP,ENTSD/10HFLEPO     ,10HDESCENT    /
C
C INITIALISATION
C
      IERR=0
      IF((MITER.LE.0).AND.(LOADO.NE.3))GO TO 9999
      NFC=NFCT=0
      IF(IPNT.NE.1)GO TO 5
      ENTT=ENTFP
      IF(LOADO.EQ.4)ENTT=ENTSD
      PRINT 1,ENTT
    1 FORMAT(4H0    ,A10/4X,14HITERN FN.CALLS,6X,1HF,12X,5HMAX G,12X,1HX)
    5 CONTINUE
      IF(LOADO.NE.1) GO TO 20
      DO 15 I=1,N
      DO 10 J=I,N
   10 H(I,J)=H(J,I)=0.
   15 H(I,I)=1.
   20 GX=GREAT(N,G)
C
C COMMENCE ITERATION
C
```

```
      DO 180 IFP=1,MITER
      ITER=IFP-1
      NFCT=NFCT+NFC
      IF(IPNT.EQ.1)PRINT 2,ITER,NFC,F,GX,(X(I),I=1,N)
    2 FORMAT(2I8,2E15.6,6G15.5/(46X,6G15.5))
      NFC=0
C
C DETERMINE DIRECTION FOR MINIMISATION
C
      GO TO (35,35,42,45),LOADO
   35 DO 40 I=1,N
      SIGMA(I)=X(I)
      GAMMA(I)=G(I)
   40 E(I)=(-DOT(N,H(1,I),G))
      GO TO 50
   42 DO 43 I=1,N
   43 E(I)=H(I,1)
      GO TO 50
   45 DO 46 I=1,N
   46 E(I)=(-G(I))
C
C CHECK THAT F IS DECREASING AT THIS POINT
C
   50 FB=F
      GB=DOT(N,G,E)
      IF(GB.LT.0.) GO TO 60
      IERR=5
      GO TO 9999
C
C DETERMINE AN INITIAL STEP SIZE
C
   60 STEP=1./PROPCH
      DO 63 I=1,N
      IF((X(I).EQ.0.).OR.(E(I).EQ.0.)) GO TO 63
      T=ABS(X(I)/E(I))
      IF(T.LT.STEP) STEP=T
   63 CONTINUE
      STEP=STEP*PROPCH
C
C CALL THIS POINT A AND COMPUTE NEW POINT B=A+STEP*E
C SUCH THAT MINIMUM LIES BETWEEN A AND B - MAY NEED
C EITHER A REDUCTION IN STEP SIZE (STATEMENT 86) TO MAKE B SENSIBLE
C OR AN INCREASE IN STEP SIZE, IN WHICH CASE USE LINEAR EXTRAPOLATION
C ON G TO DETERMINE APPROPIATE STEP
C
   70 FA=FB
      GA=GB
      DO 80 I=1,N
   80 X(I)=X(I)+STEP*E(I)
   85 CALL FUNCT
      NFC=NFC+1
      IF(F.GE.FMAX) GO TO 86
      FB=F
      GB=DOT(N,G,E)
      IF(GB.GE.0.) GO TO 90
      IF(FB.GT.FA)GO TO 86
      T=GB*STEP/(GA-GB)
      IF(T.GT.EPSS)STEP=1.25*T
      GO TO 70
   86 T=STR*STEP
      STEP=(1.-STR)*STEP
      IF(STEP.GE.EPSS) GO TO 87
      IERR=2
      GO TO 9999
   87 DO 88 I=1,N
   88 X(I)=X(I)-T*E(I)
      GO TO 85
C
C USE CUBIC INTERPOLATION TO LOCATE MINIMUM
C (QUADRATIC IF COEFF. OF X**3 TOO SMALL)
C
   90 CONTINUE
```

```
      DSS=2.*(FA-FB)/STEP+GA+GB
      IF(ABS(DSS).GT.1.E-8)GO TO 95
      T=(GB*STEP)/(GA-GB)
      GO TO 98
   95 CS=3.*(FA-FB)/STEP+GA+2.*GB
      W=SQRT(CS*CS-3.*GB*DSS)
      T=(-CS+W)/(3.*DSS)*STEP
   98 DO 100 I=1,N
  100 X(I)=X(I)+T*E(I)
      CALL FUNCT
      NFC=NFC+1
      IF((F.LE.FA).AND.(F.LE.FB)) GO TO 130
C
C INTERPOLATE UNSATISFACTORY - EITHER PUT A = X (STATEMENT 110)
C OR B = X (STATEMENT 120) OR START AGAIN FROM A WITH REDUCED STEP
C
      GT=DOT(N,G,E)
      IF((FA.GT.FB).AND.(GT.LT.0.)) GO TO 110
      IF((FA.LT.FB).AND.(GT.GT.0.)) GO TO 120
      STEP=STEP+T
      GO TO 86
  110 FA=F
      GA=GT
      STEP=(-T)
      IF(STEP.GE.EPSS) GO TO 114
      IERR=3
      GO TO 9999
  114 DO 115 I=1,N
  115 X(I)=X(I)+STEP*E(I)
      GO TO 90
  120 FB=F
      GB=GT
      STEP=STEP+T
      IF(STEP.GE.EPSS) GO TO 90
      IERR=4
      GO TO 9999
C
C MINIMUM HAS BEEN LOCATED
C
  130 IF(LOADO.EQ.3) GO TO 190
      IF(LOADO.EQ.4)GO TO 170
C
C UPDATE H FOR NEXT ITERATION
C
      DO 140 I=1,N
      SIGMA(I)=X(I)-SIGMA(I)
  140 GAMMA(I)=G(I)-GAMMA(I)
      SG=DOT(N,SIGMA,GAMMA)
      DO 150 I=1,N
  150 E(I)=DOT(N,H(1,I),GAMMA)
      GHG=DOT(N,E,GAMMA)
      IF((SG.EQ.0.).OR.(GHG.EQ.0.)) GO TO 170
      DO 160 I=1,N
      DO 160 J=I,N
  160 H(I,J)=H(J,I)=H(I,J)+SIGMA(I)*SIGMA(J)/SG-E(I)*E(J)/GHG
C
C CHECK FOR CONVERGENCE
C
  170 GX=GREAT(N,G)
      IF(GX.LE.EPS) GO TO 190
  180 CONTINUE
      IERR=1
C
C ITERATION COMPLETE - TIDY UP AND RETURN
C
  190 ITER=ITER+1
      NFCT=NFCT+NFC
      IF(IPNT.EQ.1)PRINT 2,ITER,NFC,F,GX,(X(I),I=1,N)
 9999 RETURN
      END
```

****

```
      FUNCTION DOT(N,A,B)
C
C USED BY FLEPO TO FORM THE SCALAR PRODUCT OF VECTORS A AND B
C
      REAL A(N),B(N)
C
      T=0
      DO 10 I=1,N
   10 T=T+A(I)*B(I)
      DOT=T
      RETURN
      END
```

****

```
      FUNCTION GREAT(N,A)
C
C USED BY THE MINIMISATION ROUTINES TO FIND THE MAXIMUM VALUE
C OF MODULUS(A(I))
C
      REAL A(N)
C
      T=0
      DO 10 I=1,N
      U=ABS(A(I))
   10 IF(U.GT.T) T=U
      GREAT=T
      RETURN
      END
```

****

```
C PURPOSE. SAMPLE MAIN PROGRAM FOR VARS
C
C USAGE. CARD 1 - TITLE CARD
C        CARD 2 - NDFS, P, Q, ILSMF (SYMBOLS AS FOR VARS)
C        CARD 3 AND FOLLOWING - LOWER TRIANGLE OF S PUNCHED ROW BY ROW
C        CARD 4 - INITIAL APPROXIMATION FOR GAMMA
C ANY NUMBER OF SETS OF DATA - TERMINATE RUN WITH BLANK CARD
C
C NOTE1. CALLS OUT TO PRINT MATRICES - SEE LISTING BELOW
C
C NOTE2. ALPHANUMERIC PART OF INPUT/OUTPUT AND THE TERMINATION TEST
C WOULD ALMOST CERTAINLY NEED CHANGE FOR COMPUTERS WITH DIFFERENT
C WORD SIZE.
C
      COMMON /ARGS/NDFS,P,Q,ILSMF,S(10,10),NS,GA(20),SI(10,10),NSI,
     1              X2,NDFM,NRP,IE
      INTEGER P,Q
      REAL TITLE(8)
      DATA BLANKS/10H          /
C
   10 READ 1,(TITLE(I),I=1,8)
    1 FORMAT(8A10)
      IF(TITLE(1).EQ.BLANKS)GO TO 9999
      PRINT 2,(TITLE(I),I=1,8)
    2 FORMAT(1H1,8A10)
C
      READ 3,NDFS,P,Q,ILSMF
    3 FORMAT(4I4)
      DO 20 I=1,P
      READ 4,(S(I,J),J=1,I)
    4 FORMAT(9F8)
      DO 20 J=1,I
   20 S(J,I)=S(I,J)
      CALL OUT(P,P,S,NS,1HS,1)
      READ 4,(GA(I),I=1,Q)
C
      CALL VARS
C
      IF(IE.NE.0)PRINT 5,IE
    5 FORMAT(10H0ERROR NO.,I4)
```

```
      IF(IERR.EQ.6)GO TO 10
      CALL OUT(1,Q,GA,1,5HGAMMA,1)
      CALL OUT(P,P,SI,NSI,5HSIGMA,1)
      IF(NRP.NE.0)PRINT 6,NRP
    6 FORMAT(18H0RANK DEFICIENCY =,I4)
      PRINT 7,NDFM,X2
    7 FORMAT(8H0CHISQ (,I4,4H ) =,F11.4)
      GO TO 10
C
 9999 STOP
      END
```

```
**********************************************************************

      SUBROUTINE SIGA(IOD)
C
C PURPOSE.  TO EVALUATE SIGMA(GAMMA) (FOR IOD=0) AND THE  P*(P+1)/2 X Q
C MATRIX OF FIRST DERIVATIVES (FOR IOD=1) FOR THE QUASI-SIMPLEX MODEL
C
      COMMON /ARGS/NDFS,P,Q,ILSMF,S(10,10),NS,GA(20),SI(10,10),NSI,
     1              X2,NDFM,NRP,IE
      INTEGER P,Q
      COMMON /WORK2/D(55,20),ND,E(55)
C
      IF(IOD.EQ.1)GO TO 40
C
C FORM SI = SIGMA(GAMMA)
C
      T=0.
      DO 30 J=1,P
      T=T+GA(J)
      DO 20 I=J,P
   20 SI(I,J)=SI(J,I)=T
   30 SI(J,J)=T+GA(Q)
      GO TO 9999
C
C FORM D = DSIGMA/DGAMMA
C
   40 IJ=0
      DO 60 J=1,P
      DO 60 I=J,P
      IJ=IJ+1
      DO 50 K=1,P
      D(IJ,K)=1.
   50 IF(K.GT.J)D(IJ,K)=0.
      D(IJ,Q)=0.
   60 IF(I.EQ.J)D(IJ,Q)=1.
 9999 RETURN
      END
```

```
**********************************************************************

      SUBROUTINE OUT(M,N,A,NA,L,K)
C
C PURPOSE.  TO PRINT THE  M X N  MATRIX A, WITH THE HEADING GIVEN IN L.
C FORMAT, AS LISTED, IS F11.5 WITH 12 NUMBERS PER LINE.  CAN BE CHANGED
C BY CHANGING THE VALUES OF NCL AND FORM.
C
      DIMENSION A(NA,1),L(K)
      COMMON /WOUT/NCL,FORM(8)
      DATA NCL,FORM/12,12H(1X,12F11.5)/
      PRINT 1,(L(I),I=1,K)
    1 FORMAT(/10X,12A10)
      NI=1-NCL
      NF=0
   10 NI=NI+NCL
      NF=NF+NCL
      IF(NF.GT.N)NF=N
      DO 20 I=1,M
   20 PRINT FORM,(A(I,J),J=NI,NF)
      IF(NF.EQ.N)GO TO 9999
      PRINT 2
```

```
    2 FORMAT(1H )
      GO TO 10
 9999 RETURN
      END
```

****

```
SAMPLE DATA FOR VARS.  QUASI-SIMPLEX MODEL  C.F. BROWNE(1974)
 151   6   7   2
    521.
    477.    576.
    484.    536.    601.
    510.    575.    593.    755.
    523.    580.    598.    718.    797.
    528.    584.    613.    722.    751.    802.
    475.     51.     13.     78.     24.      5.     46.
```

****

APPENDIX C. SUBROUTINES FOR UNRESTRICTED FACTOR ANALYSIS USING GENERAL ESTIMATION FUNCTIONS.

```
      SUBROUTINE GFFA
C
C PURPOSE.  TO DETERMINE BY MINIMISATION OF A GENERAL ESTIMATION
C FUNCTION THE ESTIMATES CORRESPONDING TO A  P X P  SAMPLE VARIANCE
C MATRIX S OF THE PARAMETERS LAMBDA AND PSI OF THE UNRESTRICTED
C FACTOR ANALYSIS MODEL.
C
C METHOD.  AN INITIAL APPROXIMATION TO THE MINIMISING PSI IS
C IMPROVED BY THE METHOD OF STEEPEST DESCENT, AND THE FINAL ESTIMATE
C OBTAINED BY NEWTON-RAPHSON ITERATION
C
C PROGRAMMER.  A.J.SWAIN.  JAN 1975.
C
C USAGE.  THE MAIN ARGUEMENTS ARE COMMUNICATED BETWEEN THE CALLING
C ROUTINE AND GFFA THROUGH THE COMMON BLOCK
C /ARGS/N       DEGREES OF FREEDOM OF S (0 IF UNKNOWN)
C       P       (INTEGER) SIZE OF S (AS LISTED .LE. 20)
C       K       NUMBER OF FACTORS (AS LISTED .LE. 14)
C       IOPT    0  GFFA TO GENERATE INITIAL APPROX. TO PSI
C               1  ON CALL V CONTAINS AN INITIAL APPROX. TO PSI
C               2  ON CALL V CONTAINS AN INITIAL APPROX. TO PSI**.5
C       ILSMF   DETERMINES THE ESTIMATION FUNCTION TO BE USED
C               1  TGLS - TRUE GENERALISED LEAST SQUARES
C               2  ML - MAXIMUM WISHART LIKELIHOOD
C               3  GD - GEODESIC DISTANCE
C               4  DIV - DIVERGENCE
C               5  GLS - GENERALISED LEAST SQUARES
C               6  GLSE - GENERALISED LEAST SQUARES EXTENDED
C               LOADED WITH VALUE 2 SO ML USED ON DEFAULT
C       S,NS    SAMPLE VARIANCE MATRIX (PRESERVED BY GFFA)
C       V       ON CALL AS DETERMINED BY IOPT
C               ON RETURN ESTIMATE OF PSI
C       L,NL    (REAL) ON RETURN ESTIMATE OF LAMBDA
C       SIGMA,NSIGMA  ON RETURN ESTIMATE OF SIGMA
C       X2      APPROX. CHISQUARE TEST STATISTIC FOR GOODNESS OF FIT
C       NDF     DEGREES OF FREEDOM OF MODEL (AND X2)
C       IE      ERROR INDICATOR
C               0  SUCCESSFUL CONVERGENCE TO ESTIMATES
C               1  FAILURE TO CONVERGE IN SPECIFIED MAX. NO. OF ITERNS.
C               2-5  FAILURE OF NEWTON-RAPHSON METHOD IN LOCATING
C                  MINIMUM (OR OF STEEPEST DESCENT IF NEWRA NOT CALLED).
C                  CODES CORRESPOND TO THOSE OF THE APPROPRIATE ROUTINE.
C               6  S NOT POSITIVE DEFINITE
C               7  FAILURE OF EIGENVALUE ROUTINE TO DETERMINE THE
C                  EIGENVALUES OF S
C
C PARAMETERS NOT NORMALLY REQUIRING CHANGE, BUT ACCESIBLE TO THE
C CALLING ROUTINE ARE CONTAINED IN THE COMMON BLOCK
C /ITERL/MSD     MAXIMUM NUMBER OF STEEPEST DESCENT ITERATIONS
C        MNR     MAXIMUM NUMBER OF NEWTON-RAPHSON ITERATIONS
C        EPSSD   CONVERGENCE PARAMETER FOR STEEPEST DESCENT
C        EPSNR   CONVERGENCE PARAMETER FOR NEWTON-RAPHSON
C
C OTHER ARGUEMENTS AVAILABLE ON RETURN ARE CONTAINED IN THE COMMON BLOCK
C /WFUNCT/FMAX  NOT USED
C         NP    =P
C         X     V**.5
C         F     VALUE OF F(SIGMA,S) AT X
C         G     VALUE OF DERIVATIVES OF F AT X
C         H,NH  VALUE OF SECOND DERIVATIVES OF F AT X - NOTE (2./N)**
C               H INVERSE ESTIMATES THE ASYMPTOTIC VARIANCE MATRIX OF X.
C
C THE COMMON BLOCK /WORKFA/ CONTAINS THE WORKING ARRAYS USED BY GFFA
C (AS LISTED 1244 LOCATIONS) AND COULD BE USED AS WORKING STORAGE
C OUTSIDE GFFA.
C
C THE COMMON BLOCK /WMINR/ IS USED FOR COMMUNICATION WITH THE MINIMIS-
C ATION ROUTINES.  PRINTING BY THESE ROUTINES MAY BE INHIBITED BY
```

```
C SETTING IPNT TO ZERO.
C
C NOTE1.  GFFA USES THE SUBROUTINES FAFN, FLEPO AND NEWRA, WHICH ARE
C LISTED BELOW OR WITH VARS, AND THENCE THE SUBROUTINES SMF, SMFD,
C SMFDD, DOT, GREAT AND CHOL WHICH THESE CALL.
C IT ALSO USES AN EIGENVALUE/EIGENVECTOR ROUTINE EISSOL
C WITH SPECIFICATIONS AS LISTED IN VARS.
C
C NOTE2.  TO EXTEND THE NUMBER OF VARIABLES PERMITTED IT IS NECESSARY TO
C ALTER THE COMMON BLOCKS
C /ARGS/ AND /WFAFN/ IN GFFA AND FAFN
C /WFUNCT/ IN GFFA, FAFN, FLEPO AND NEWRA,
C AND TO CHANGE THE ASSIGNMENT STATEMENT FOR NB,ND,NU IN GFFA
C
C NOTE3.  TO INCLUDE OTHER ESTIMATION FUNCTIONS IT IS ONLY NECESSARY TO
C INCLUDE THE EVALUATING FUNCTIONS IN SMF, SMFD AND SMFDD AND TO ALTER
C THE COMPUTED GO TO STATEMENTS OF THESE FUNCTIONS.
C
      EXTERNAL FAFN
      COMMON /ARGS/N,P,K,IOPT,ILSMF,S(20,20),NS,V(20),L(20,14),NL,
     1               SIGMA(20,20),NSIGMA,X2,NDF,IE
      COMMON /ITERL/MSD,MNR,EPSSD,EPSNR
      COMMON /WFUNCT/FMAX,NP,X(20),F,G(20),H(20,21),NH
      COMMON /WORKFA/K1,A(20),B(20,20),NB,D(20,20),ND,E(20),U(20,20),NU
      COMMON /WMINR/EPS,MITER,IPNT,IERR,ITER,LOADO,NFCT
      COMMON IOD
      INTEGER P
      REAL L
      INTEGER Q,R,IT(14)
      DATA ILSMF/2/
      DATA NS,NL,NSIGMA/20,20,20/
      DATA MSD,MNR,EPSSD,EPSNR/6,20,1.E-02,1.E-06/
      DATA IPNT/1/
      DATA EPS3,EPS4/1.E-05,1.E-20/
C
      NB=ND=NU=20
      IE=0
      NP=P
      RP=P
      RK=K
      K1=K+1
C
C GENERATE STARTING VARIANCES AS REQUIRED BY IOPT
C
      IF(IOPT-1) 10,60,80
   10 DO 20   I=1,P
      DO 20   J=1,P
   20 D(I,J)=S(I,J)
      CALL CHOL(4,P,P,D,ND,T)
      IF(T.GT.0)GO TO 40
   30 IE=6
      GO TO 9999
   40 T=1.-RK/(2.*RP)
      DO 50 I=1,P
      V(I)=T/D(I,I)
   50 X(I)=SQRT(V(I))
      GO TO 100
   60 DO 70 I=1,P
   70 X(I)=SQRT(V(I))
      GO TO 100
   80 DO 90 I=1,P
      X(I)=V(I)
   90 V(I)=X(I)*X(I)
  100 CONTINUE
C
C COMPUTE  D = S**-.5  VIA SPECTRAL REPRESENTATION
C
      CALL EISSOL(P,S,NS,A,B,NB,IERR)
      IF(IERR.EQ.0)GO TO 110
      IE=7
      GO TO 9999
  110 IF(A(1).LT.EPS4)GO TO 30
```

```
      DO 120 I=1,P
  120 A(I)=1./SQRT(A(I))
      DO 140 I=1,P
      DO 140 J=I,P
      T=0.
      DO 130 R=1,P
  130 T=T+B(I,R)*A(R)*B(J,R)
  140 D(I,J)=D(J,I)=T
C
C COMMENCE STEEPEST DESCENT ITERATIONS
C
      IOD=1
      CALL FAFN
      IF(MSD.EQ.0)GO TO 150
      EPS=EPSSD
      MITER=MSD
      LOADO=4
      CALL FLEPO(FAFN)
      IE=IERR
C
C COMMENCE NEWTON-RAPHSON ITERATIONS
C
  150 IOD=(-2)
      CALL FAFN
      IF(MNR.EQ.0)GO TO 160
      IOD=2
      EPS=EPSNR
      MITER=MNR
      CALL NEWRA(FAFN)
      IE=IERR
      IF(IE.EQ.3)GO TO 9999
C
C DETERMINE DIRECT ESTIMATE OF LAMBDA
C
  160 CALL CHOL(4,P,P,D,ND,T)
      DO 180 J=1,K
      T=SQRT(1.-E(J))
      DO 180 I=1,P
      Z=0.
      DO 170 R=1,P
  170 Z=Z+D(I,R)*SIGMA(R,J)
  180 L(I,J)=Z*T
C
C COUNT UP THE NUMBER OF V(I) WHICH ARE EFFECTIVELY ZERO.  IF GREATER
C THAN ONE, FORM THE PRINCIPAL COMPONENT SOLUTION FOR THE VARIATES WITH
C V(I)=0., SINCE IN THIS CASE THE CORRESPONDING PARTS OF S ARE
C REPRODUCED EXACTLY, AND THE CORRESPONDING COLUMNS OF LAMBDA ARE NOT
C DETERMINED UNIQUELY.
C
      M=0
      DO 190  I=1,P
  190 IF(V(I).LT.EPS3)M=M+1
      IF(M.LE.1)GO TO 260
      M1=M+1
      Q=0
      R=M
      DO 210  I=1,P
      IF(V(I).GT.EPS3)GO TO 200
      Q=Q+1
      IT(Q)=I
      GO TO 210
  200 R=R+1
      IT(R)=I
  210 CONTINUE
C
C SECONDLY EXTRACT THE APPROP, PARTS OF S, AND FORM THE EIGENVALUES ETC.
C
      DO 220  J=1,M
      ITJ=IT(J)
      DO 220  I=1,P
      ITI=IT(I)
  220 B(I,J)=S(ITI,ITJ)
```

```
      MM=-M
      CALL EISSOL(MM,B,NB,E,U,NU,IERR)
C
C THIRDLY CONSTRUCT THE APPROPRIATE COLUMNS OF L
C
      DO 250  J=1,M
      Y=SQRT(E(J))
      DO 230  I=1,M
      ITI=IT(I)
  230 L(ITI,J)=U(I,J)*Y
      Z=1./Y
      DO 250  I=M1,P
      ITI=IT(I)
      T=0.
      DO 240  R=1,M
  240 T=T+B(I,R)*U(R,J)
  250 L(ITI,J)=T*Z
C
C FORM THE ESTIMATE OF SIGMA
C
  260 DO 290 I=1,P
      DO 280 J=I,P
      T=0.
      DO 270 R=1,K
  270 T=T+L(I,R)*L(J,R)
  280 SIGMA(I,J)=SIGMA(J,I)=T
  290 SIGMA(I,I)=SIGMA(I,I)+V(I)
C
C COMPUTE DEGREES OF FREEDOM AND CHISQUARE VALUE, IF POSSIBLE
C
      NDF=((P-K)**2-(P+K))/2
      X2=0.
      IF(N.LE.0) GO TO 9999
      T=N
      X2=T*F
 9999 RETURN
      END
```

****

```
      SUBROUTINE FAFN
C
C PURPOSE.  CALLED BY GFFA, AND BY MINIMISATION ROUTINES, TO EVALUATE
C F, G AND H FOR FACTOR ANALYSIS MODEL, AT   X = PSI**.5.
C
C WHAT RETURNED IS DEPENDENT ON THE VALUE OF IOD
C 0 = JUST F,   1 = F AND G,   2 = F, G AND H,   -1 = JUST G,   -2 = JUST H
C IOD=0, 1 OR 2 ASSUMES THAT ON CALL D=S**-.5,AND RETURNS THE
C EIGENVALUES AND VECTORS OF  D*V*D  IN E AND SIGMA RESP.
C IOD=-1 ASSUMES THE RESULT OF A IOD=0 CALL
C IOD=-2 ASSUMES THE RESULTS OF A IOD=1 CALL
C
      COMMON IOD
      COMMON /ARGS/N,P,K,IOPT,ILSMF,S(20,20),NS,V(20),L(20,14),NL,
     1             SIGMA(20,20),NSIGMA,X2,NDF,IE
      COMMON /WFUNCT/FMAX,NP,X(20),F,G(20),H(20,21),NH
      COMMON /WORKFA/K1,A(20),B(20,20),NB,D(20,20),ND,E(20),U(20,20),NU
      INTEGER P
      REAL L
      INTEGER Q,R
C
C INITIALISATION
C
      DO 10 I=1,P
   10 V(I)=X(I)*X(I)
      I=IOD+3
      GO TO (120,50,20,20,20),I
C
C SET UP (S**-.5)*PSI*(S**-.5) AND DETERMINE ITS EIGENVALUES AND VECTORS
C IF NECESSARY ADJUST V SUCH THAT E(K).LE.1.
C
   20 DO 40 I=1,P
```

```
      DO 40 J=I,P
      T=0.
      DO 30 R=1,P
   30 T=T+D(I,R)*V(R)*D(R,J)
   40 B(I,J)=B(J,I)=T
      CALL EISSOL(P,B,NB,E,SIGMA,NSIGMA,IERR)
      IF(E(K).LE.1.)GO TO 50
      T=E(K)
      Y=SQRT(T)
      DO 45 I=1,P
      X(I)=X(I)/Y
      V(I)=V(I)/T
   45 E(I)=E(I)/T
C
C COMPUTE F
C
   50 F=0.
      DO 60 I=K1,P
   60 F=F+SMF(E(I),ILSMF)
      IF(IOD.EQ.0) GO TO 9999
C
C SET UP U AND ALPHA AND COMPUTE G
C
      DO 80 I=1,P
      DO 80 J=1,P
      T=0.
      DO 70 R=1,P
   70 T=T+D(I,R)*SIGMA(R,J)
   80 U(I,J)=T
      DO 90 I=K1,P
   90 A(I)=SMFD(E(I),ILSMF)
      DO 110 I=1,P
      T=0.
      DO 100 J=K1,P
  100 T=T+A(J)*U(I,J)**2
  110 G(I)=2.*X(I)*T
      IF((IOD.EQ.1).OR.(IOD.EQ.-1))GO TO 9999
C
C SET UP BETA AND COMPUTE H
C
  120 DO 130 I=K1,P
      B(I,I)=SMFDD(E(I),ILSMF)
      DO 130 J=1,P
      IF(J.EQ.I) GO TO 130
      B(I,J)=2.*A(I)/(E(I)-E(J))
  130 CONTINUE
      DO 170 I=1,P
      T=0.
      DO 140 J=K1,P
  140 T=T+A(J)*U(I,J)**2
      DO 160 J=I,P
      Y=0.
      DO 150 Q=K1,P
      Z=U(I,Q)*U(J,Q)
      DO 150 R=1,P
  150 Y=Y+B(Q,R)*U(I,R)*U(J,R)*Z
  160 H(I,J)=H(J,I)=4.*X(I)*X(J)*Y
  170 H(I,I)=H(I,I)+2.*T
 9999 RETURN
      END

***********************************************************************************

      FUNCTION SMF(T,ILSMF)
C
C PURPOSE.  CALLED BY FAFN TO EVALUATE F(THETA) AT THETA=T,
C WITH THE PARTICULAR F USED BEING DETERMINED BY ILSMF - SEE GFFA LIST
C
      GO TO(10,20,30,40,50,60),ILSMF
   10 SMF=.5*(T-1.)*(T-1.)/(T*T)
      GO TO 9999
   20 SMF=1./T+ALOG(T)-1.
```

```
      GO TO 9999
   30 SMF=.5*ALOG(T)**2
      GO TO 9999
   40 SMF=.5*(T+1./T-2.)
      GO TO 9999
   50 SMF=.5*(T-1.)*(T-1.)
      GO TO 9999
   60 SMF=.5*(T*T-2.*T+1.)*EXP(T-1.)
 9999 RETURN
      END

***********************************************************************

      FUNCTION SMFD(T,ILSMF)
C
C PURPOSE.  CALLED BY FAFN TO EVALUATE 1ST DERIV. OF F(THETA) AT T,
C WITH THE PARTICULAR F USED BEING DETERMINED BY ILSMF - SEE GFFA LIST
C
      GO TO (10,20,30,40,50,60),ILSMF
   10 SMFD=(T-1.)/(T*T*T)
      GO TO 9999
   20 SMFD=(T-1.)/(T*T)
      GO TO 9999
   30 SMFD=ALOG(T)/T
      GO TO 9999
   40 SMFD=.5*(1.-1./(T*T))
      GO TO 9999
   50 SMFD=T-1.
      GO TO 9999
   60 SMFD=.5*(T*T-1.)*EXP(T-1.)
 9999 RETURN
      END

***********************************************************************

      FUNCTION SMFDD(T,ILSMF)
C
C PURPOSE.  CALLED BY FAFN TO EVALUATE 2ND DERIV. OF F(THETA) AT T,
C WITH THE PARTICULAR F USED BEING DETERMINED BY ILSMF - SEE GFFA LIST
C
      GO TO (10,20,30,40,50,60),ILSMF
   10 SMFDD=(3.-2.*T)/(T*T*T*T)
      GO TO 9999
   20 SMFDD=(2.-T)/(T*T*T)
      GO TO 9999
   30 SMFDD=(1.-ALOG(T))/(T*T)
      GO TO 9999
   40 SMFDD=1./(T*T*T)
      GO TO 9999
   50 SMFDD=1.
      GO TO 9999
   60 SMFDD=.5*(T*T+2.*T-1.)*EXP(T-1.)
 9999 RETURN
      END

***********************************************************************

      SUBROUTINE NEWRA(FUNCT)
C
C PURPOSE.  TO MINIMISE A FUNCTION F OF N VARIABLES X(1) ... X(N)
C USING THE CLASSICAL NEWTON-RAPHSON METHOD
C
C PROGRAMMER.  A.J.SWAIN, JAN 1975.
C
C USAGE.  THE ARGUEMENT FUNCT OF THE CALL IS A USER SUPPLIED SUBROUTINE
C (DECLARED EXTERNAL IN THE ROUTINE CALLING NEWRA) WHICH WHEN CALLED
C EVALUATES F,G (THE VECTOR OF FIRST DERIVATIVES) AND H (THE MATRIX OF
C SECOND DERIVATIVES) AT A GIVEN POINT X.  THE NECESSARY ARGUEMENTS ARE
C TRANSFERED BETWEEN THE CALLING ROUTINE, NEWRA AND FUNCT THROUGH THE
C COMMON BLOCK
C /WFUNCT/ FMAX          AN UNSATISFACTORY X CAN BE INDICATED TO NEWRA BY
C                        FUNCT BY RETURNING F=FMAX (LISTED VALUE 1.E+75)
```

```
C           N             NUMBER OF VARIABLES (.LE.MN,SAY - LISTED MN=20)
C           X(MN)         CURRENT X
C           F             VALUE OF F AT X
C           G(MN)         VALUE OF FIRST DERIVATIVES AT X
C           H(MN,MN+1)    VALUE OF SECOND DERIVATIVES AT X
C           NH            ROW DIMENSION OF H (=MN)
C WHEN NEWRA CALLED, X IS TAKEN AS AN APPROXIMATION TO THE MINIMISING X,
C WITH F,G AND H GIVING THE CORRESPONDING VALUES, I.E. A CALL TO FUNCT
C MUST PRECEDE THE CALL TO NEWRA.
C
C OTHER ARGUMENTS ARE TRANSFERED BETWEEN THE CALLING ROUTINE AND NEWRA
C THROUGH THE COMMON BLOCK
C /WMINR/ EPS    ITERATION IS TERMINATED WHEN MAX(ABS(G(I))).LE.EPS,
C                E.G. EPS=1.E-06
C          MITER MAXIMUM NUMBER OF ITERATIONS
C          IPNT  PRINT OPTION  0=NO PRINTING, 1=PRINTING
C          IERR  ERROR INDICATOR
C                0 SUCCESSFUL CONVERGENCE
C                1 FAILURE TO CONVERGE IN MITER ITERATIONS
C                2 FUNCT RETURNED FMAX
C                3 H NOT POSITIVE SEMIDEFINITE
C          ITER  NUMBER OF ITERATIONS REQUIRED
C
C NOTE.  REQUIRES GREAT(N,A) WHICH IS LISTED WITH FLEPO,
C AND CHOL(IO,N,M,A,MA,DA) WHICH IS LISTED BELOW
C
      COMMON /WFUNCT/FMAX,N,X(20),F,G(20),H(20,21),NH
      COMMON /WMINR/EPS,MITER,IPNT,IERR,ITER
      DATA FMAX,NH/1.E+75,20/
C
C INITIALISATION
C
      IERR=0
      IF(MITER.EQ.0)GO TO 9999
      N1=N+1
      IF(IPNT.EQ.1)PRINT 1
    1 FORMAT(9H0   NEWRA/4X,9HITERATION,11X,1HF,12X,5HMAX G,12X,1HX)
      GX=GREAT(N,G)
C
C COMMENCE ITERATION
C
      DO 100 INR=1,MITER
      ITER=INR-1
      IF(IPNT.EQ.1)PRINT 2,ITER,F,GX,(X(I),I=1,N)
    2 FORMAT(I8,8X,2E15.6,6G15.5/(46X,6G15.5))
C
C DETERMINE THE NEW X,F,G, AND H
C
      DO 10 I=1,N
   10 H(I,N1)=G(I)
      CALL CHOL(2,N,N1,H,NH,DH)
      IF(DH.GE.0)GO TO 20
      IERR=3
      GO TO 9999
   20 DO 25 I=1,N
   25 X(I)=X(I)-H(I,N1)
      CALL FUNCT
      IF(F.LT.FMAX)GO TO 30
      IERR=2
      GO TO 9999
C
C CHECK FOR CONVERGENCE
C
   30 GX=GREAT(N,G)
      IF(GX.LT.EPS)GO TO 110
  100 CONTINUE
      IERR=1
C
C ITERATION COMPLETE - TIDY UP AND RETURN
C
  110 ITER=ITER+1
      IF(IPNT.EQ.1)PRINT 2,ITER,F,GX,(X(I),I=1,N)
```

```
 9999 RETURN
      END
```

```
*************************************************************************
```

```
C PURPOSE.  SAMPLE MAIN PROGRAM FOR GFFA.
C
C USAGE.  CARD 1 - TITLE CARD
C         CARD 2 - N, P, K, IOPT, ILSMF (SYMBOLS SAME AS FOR GFFA)
C         CARD 3 AND FOLLOWING - LOWER TRIANGLE OF S PUNCHED ROW BY ROW
C        CARD 4 - INITIAL APPROXIMATION TO PSI (DEPENDING ON IOPT)
C ANY NUMBER OF SETS OF DATA - TERMINATE RUN BY BLANK CARD
C
C NOTE1.  CALLS OUT TO PRINT MATRICES - SEE LISTING OF VARS
C
C NOTE2.  ALPHANUMERIC PART OF INPUT/OUTPUT AND THE TERMINATION TEST
C WOULD ALMOST CERTAINLY NEED CHANGE FOR COMPUTERS WITH DIFFERENT
C WORD SIZE.
C
      COMMON /ARGS/N,P,K,IOPT,ILSMF,S(20,20),NS,V(20),L(20,14),NL,
     1              SIGMA(20,20),NSIGMA,X2,NDF,IE
      REAL TITLE(8)
      INTEGER P
      REAL L
      DATA BLANKS/10H          /
C
   10 READ 1,(TITLE(I),I=1,8)
    1 FORMAT(8A10)
      IF(TITLE(1).EQ.BLANKS)GO TO 9999
      PRINT 2,(TITLE(I),I=1,8)
    2 FORMAT(1H1,8A10)
C
      READ 3,N,P,K,IOPT,ILSMF
    3 FORMAT(5I4)
      DO 20 I=1,P
      READ 4,(S(I,J),J=1,I)
    4 FORMAT(9F8)
      DO 20 J=1,I
   20 S(J,I)=S(I,J)
      CALL OUT(P,P,S,NS,1HS,1)
      IF(IOPT.GT.0)READ 4,(V(I),I=1,P)
C
      CALL GFFA
C
      IF(IE.NE.0)PRINT 5,IE
    5 FORMAT(10H0ERROR NO.,I4)
      CALL OUT(P,K,L,NL,6HLAMBDA,1)
      CALL OUT(1,P,V,1,3HPSI,1)
      CALL OUT(P,P,SIGMA,NSIGMA,5HSIGMA,1)
      PRINT 6,NDF,X2
    6 FORMAT(7H0CHISQ(,I4,4H ) =,F11.4)
      GO TO 10
C
 9999 STOP
      END
```

```
*************************************************************************
```

```
GLS SOLN. EMMETT DATA  C.F. JORESKOG AND GOLDBERGER P.256
 210   9   3   0   5
    1.00
    .523    1.00
    .395    .479    1.00
    .471    .506    .355    1.00
    .346    .418    .270    .691    1.00
    .426    .462    .254    .791    .679    1.00
    .576    .547    .452    .443    .383    .372    1.00
    .434    .283    .219    .285    .149    .314    .385    1.00
    .639    .645    .504    .505    .409    .472    .680    .470    1.00
```

```
*************************************************************************
```

APPENDIX D. SUBROUTINES FOR FITTING THE DIRECT PRODUCT MODEL USING MAXIMUM LIKELIHOOD.

```
      SUBROUTINE DTDP(NDFS,P,P1,P2,S,NS,SI1,NSI1,SI2,NSI2,SI,NSI,
     1                X2,NDFM,IERR)
C
C PURPOSE.  TO DETERMINE FOR A SAMPLE VARIANCE MATRIX S THE MAXIMUM
C LIKELIHOOD ESTIMATES OF THE PARAMETERS SIGMA1 AND SIGMA2 OF THE
C DIRECT PRODUCT MODEL, WITH THE IDENTIFICATION CONDITION SIGMA1(1,1)=1.
C
C METHOD.  THE ITERATIVE METHOD DESCRIBED IN SECTION 5.3 FOR SOLUTION
C OF THE LIKELIHOOD EQUATIONS IS USED.
C
C PROGRAMMER.  A.J.SWAIN  JAN 1975.
C
C ARGUEMENTS.  NDFS      DEGREES OF FREEDOM OF S
C              P         SIZE OF S
C              P1        SIZE OF SIGMA1 (AS LISTED  .LE.5)
C              P2        SIZE OF SIGMA2 (AS LISTED  .LE.12)
C              S,NS      SAMPLE VARIANCE MATRIX (PRESERVED BY DTDP)
C              SI1,NSI1  ON RETURN ESTIMATE OF SIGMA1
C              SI2,NSI2  ON RETURN ESTIMATE OF SIGMA2
C              SI,NSI    ON RETURN ESTIMATE OF SIGMA
C              X2        APPROX. CHISQUARE TEST STATISTIC FOR GOODNESS
C                        OF FIT
C              NDFM      DEGREES OF FREEDOM OF MODEL (AND X2)
C              IERR      ERROR INDICATOR
C                        0 NO ERROR
C                        1 FAILURE TO CONVERGE IN MITER ITERATIONS
C                        2 S NOT POSITIVE DEFINITE
C
C ALSO AVAILABLE TO THE CALLING ROUTINE ARE THE PARAMETERS AND RESULT
C IN THE COMMON BLOCK
C /WDTDP/EPS    CONVERGENCE PARAMETER
C        MITER  MAXIMUM NUMBER OF ITERATIONS
C        IPNT   =1 PRINTING REQUIRED, OTHERWISE NO PRINTING
C               (LOADED WITH THE VALUE 1)
C        ITER   ON RETURN CONTAINS THE NUMBER OF ITERATIONS REQUIRED
C
C THE COMMON BLOCK /WORKDP/ CONTAINS THE WORKING ARRAYS FOR DTDP
C (AS LISTED 2543 LOCATIONS) AND COULD BE USED AS WORKING STORAGE
C OUTSIDE DTDP
C
C NOTE1.  USES SUBROUTINES PACK, UNPACK, CHINV, MPYV, SCV AND DP
C AS LISTED BELOW
C
C NOTE2.  TO EXTEND THE SIZES FOR P1 OR P2 IT IS NECESSARY TO ALTER THE
C DIMENSIONS IN /WORKDP/ AND THE ASSIGNMENT STATEMENTS FOR NG1 AND NG2.
C
      INTEGER P,P1,P2,P1T,P2T
      REAL S(NS,P),SI1(NSI1,P1),SI2(NSI2,P2),SI(NSI,P)
      COMMON /WDTDP/EPS,MITER,IPNT,ITER
      COMMON /WORKDP/G1(15,78),NG1,G2(78,15),NG2,U(78),V(78),W(15),X(15)
     1               ,W0(15)
      DATA EPS,MITER,IPNT/1.E-09,30,1/
C
C INITIALISATION
C
      NT(I)=I*(I+1)/2
      NG1=15
      NG2=78
      IERR=0
      RP=P
      RP1=P1
      RP2=P2
      IF(IPNT.EQ.1)PRINT 1
    1 FORMAT(8H0   DTDP)
C
C CHECK THAT S IS POSITIVE DEFINITE AND FORM CONSTANT PART OF
C LIKELIHOOD F
C
```

```
      CALL PACK(P,S,NS,SI)
      CALL CHINV(1,P,SI,SI,DS)
      IF(DS.GT.0)GO TO 10
      IERR=2
      GO TO 9999
   10 RL0=(-ALOG(DS)-RP)
C
C SET UP MATRICES FOR ITERATION
C
      IJ=0
      DO 20 J=1,P2
      DO 20 I=J,P2
      IJ=IJ+1
      KL=0
      DO 20 L=1,P1
      DO 20 K=L,P1
      KL=KL+1
      G2(IJ,KL)=G1(KL,IJ)=S(P2*(K-1)+I,P2*(L-1)+J)
      IF(I.NE.J)G1(KL,IJ)=G1(KL,IJ)+S(P2*(K-1)+J,P2*(L-1)+I)
   20 IF(K.NE.L)G2(IJ,KL)=G2(IJ,KL)+S(P2*(L-1)+I,P2*(K-1)+J)
      SP1=1./RP1
      SP2=1./RP2
C
C SET UP INITIAL APPROXIMATION
C
      CALL PACK(P2,S,NS,U)
      CALL CHINV(2,P2,U,V,DU)
      P1T=NT(P1)
      P2T=NT(P2)
      DO 30 I=1,P1T
   30 W0(I)=0.
      IF(IPNT.EQ.1)PRINT 3
    3 FORMAT(4X,5HITERN,7X,1HF,10X,10HMAX CHANGE)
C
C COMMENCE ITERATION
C
      DO 60 ITE=1,MITER
      ITER=ITE
C
C FORM NEW APPROXIMATION
C
      CALL MPYV(P1T,P2T,G1,NG1,V,W)
      CALL SCV(SP2,W,P1T)
      TS=1./W(1)
      AC=0.
      DO 40 I=1,P1T
      T=ABS(W0(I)-W(I))
      IF(T.GT.AC)AC=T
   40 W0(I)=W(I)=W(I)*TS
      CALL CHINV(2,P1,W,X,DW)
      CALL MPYV(P2T,P1T,G2,NG2,X,U)
      CALL SCV(SP1,U,P2T)
      CALL CHINV(2,P2,U,V,DU)
      IF((IPNT.NE.1).AND.(AC.GT.EPS).AND.(ITER.LT.MITER))GO TO 60
C
C CALCULATE LIKELIHOOD F
C
      CALL UNPACK(P1,SI1,NSI1,X)
      CALL UNPACK(P2,SI2,NSI2,V)
      CALL DP(P1,P1,P2,P2,SI1,NSI1,SI2,NSI2,SI,NSI)
      T=0.
      DO 50 I=1,P
      DO 50 J=1,P
   50 T=T+S(I,J)*SI(J,I)
      RL=RL0+RP1*ALOG(DU)+RP2*ALOG(DW)+T
      IF(IPNT.EQ.1)PRINT 4,ITER,RL,AC
    4 FORMAT(I8,2E15.6)
      IF(AC.LE.EPS)GO TO 70
   60 CONTINUE
      IERR=1
C
C ITERATION COMPLETE - UNSCRAMBLE ESTIMATES, FORM TEST STATISTIC,
```

```
C IF POSSIBLE, AND RETURN
C
   70 CALL UNPACK(P1,SI1,NSI1,W)
      CALL UNPACK(P2,SI2,NSI2,U)
      CALL DP(P1,P1,P2,P2,SI1,NSI1,SI2,NSI2,SI,NSI)
      IQ=P1T+P2T-1
      NDFM=NT(P)-IQ
      X2=0.
      IF(NDFS.LE.0)GO TO 9999
      RN=NDFS
      Q=IQ
      Y=.5*(-1.+SQRT(1.+8.*Q))
      D=NDFM
      RHO4=1.-(RP*((2.*RP+3.)*RP-1.)-Y*((2.*Y+3.)*Y-1.))/(12.*D*RN)
      X2=RHO4*RN*RL
 9999 RETURN
      END

*************************************************************************

      SUBROUTINE DPVAR(N,P1,P2,SI1,NSI1,SI2,NSI2,V,NV)
C
C PURPOSE.  TO CALCULATE THE LARGE-SAMPLE VARIANCE MATRIX FOR THE
C PARAMETER ESTIMATES FOR THE DIRECT PRODUCT MODEL, USING THE FORMULAE
C OF SECTION 5.5 (IDENTIFICATION CONDITION SIGMA1(1,1)=1)
C
C PROGRAMMER.  A.J.SWAIN  JAN 1975.
C
C ARGUEMENTS.  N          DEGREES OF FREEDOM OF S
C              P1         (INTEGER) SIZE OF SIGMA1
C              P2         (INTEGER) SIZE OF SIGMA2
C              SI1,NSI1   SIGMA1 (PRESERVED)
C              SI2,NSI2   SIGMA2 (PRESERVED)
C              V,NV       ON RETURN CONTAINS THE VARIANCE MATRIX
C
      INTEGER P1,P2,PV
      REAL SI1(NSI1,1),SI2(NSI2,1),V(NV,1)
C
C INITIALISATION
C
      I1=P1*(P1+1)/2-1
      PV=I1+P2*(P2+1)/2
      RP1=P1
      RP2=P2
      F1=1./RP2
      F2=1./RP1
      F3=2.*F1
      F4=F3*(1.-F2)
C
C CALCULATE VARIANCE MATRIX FOR ELEMENTS OF SIGMA1(EXCLUDING SIGMA1(1,1)
C
      IJ=0
      DO 15 J=1,P1
      DO 15 I=J,P1
      IF((J.EQ.1).AND.(I.EQ.1))GO TO 15
      IJ=IJ+1
      KL=0
      DO 10 L=1,P1
      DO 10 K=L,P1
      IF((L.EQ.1).AND.(K.EQ.1))GO TO 10
      KL=KL+1
      V(IJ,KL)=F1*(SI1(J,L)*SI1(I,K)+SI1(I,L)*SI1(J,K)+2.*
     1(SI1(I,J)*SI1(K,L)-SI1(I,J)*SI1(K,1)*SI1(L,1)-SI1(I,1)*SI1(J,1)*
     2SI1(K,L)))
   10 CONTINUE
   15 CONTINUE
C
C CALCULATE VARIANCE MATRIX FOR ELEMENTS OF SIGMA2
C
      IJ=I1
      DO 20 J=1,P2
      DO 20 I=J,P2
```

```
      IJ=IJ+1
      KL=I1
      DO 20 L=1,P2
      DO 20 K=L,P2
      KL=KL+1
      V(IJ,KL)=F2*(SI2(J,L)*SI2(I,K)+SI2(I,L)*SI2(J,K))
     1+F4*SI2(I,J)*SI2(K,L)
   20 CONTINUE
C
C CALCULATE COVARIANCE MATRIX BETWEEN ELEMENTS OF SIGMA1 AND OF SIGMA2
C
      IJ=0
      DO 35 J=1,P1
      DO 35 I=J,P1
      IF((J.EQ.1).AND.(I.EQ.1))GO TO 35
      IJ=IJ+1
      KL=I1
      T=F3*(SI1(I,1)*SI1(J,1)-SI1(I,J))
      DO 30 L=1,P2
      DO 30 K=L,P2
      KL=KL+1
      V(IJ,KL)=V(KL,IJ)=T*SI2(K,L)
   30 CONTINUE
   35 CONTINUE
C
C DIVIDE BY N IF GREATER THAN ZERO, AND MAKE V SYMMETRIC
C
      T=1.
      RN=N
      IF(N.NE.0)T=T/RN
      DO 40 I=1,PV
      DO 40 J=I,PV
   40 V(I,J)=V(J,I)=V(I,J)*T
      RETURN
      END
```

****

```
      SUBROUTINE DP(M,N,P,Q,A,NA,B,NB,C,NC)
C
C PURPOSE.  TO FORM THE DIRECT PRODUCT OF M X N MATRIX A AND P X Q
C MATRIX B IN THE ARRAY C,  I.E.  C = A DP B
C
      INTEGER P,Q
      REAL A(NA,1),B(NB,1),C(NC,1)
      IC=0
      DO 10 L=1,N
      DO 10 K=1,Q
      IC=IC+1
      IR=0
      DO 10 J=1,M
      DO 10 I=1,P
      IR=IR+1
   10 C(IR,IC)=A(J,L)*B(I,K)
      RETURN
      END
```

****

```
      SUBROUTINE SCV(F,V,NV)
C
C PURPOSE.  CALLED BY DTDP TO MULTIPLY THE NV X 1 VECTOR V BY F
C
      REAL V(NV)
      DO 10 I=1,NV
   10 V(I)=V(I)*F
      RETURN
      END
```

****

```
      SUBROUTINE MPYV(M,N,A,NA,B,C)
C
C PURPOSE.  CALLED BY OTOP TO MULTIPLY THE N X 1 VECTOR B BY THE M X N
C MATRIX A, RETURNING THE RESULT IN C - WHICH MUST BE DISTINCT FROM B.
C
      REAL A(NA,N),B(N),C(M)
      DO 20 I=1,M
      T=0
      DO 10 J=1,N
   10 T=T+A(I,J)*B(J)
   20 C(I)=T
      RETURN
      END
```

```
*************************************************************************
```

```
      SUBROUTINE PACK(M,A,NA,B)
C
C PURPOSE.  TO STORE BY COLUMNS THE LOWER TRIANGLE OF THE  M X M
C SYMMETRIC MATRIX A IN CONTINUOUS LOCATIONS OF THE 1-DIMENSIONAL ARRAY
C B, WHICH MAY BE THE SAME AS A. OPERATION REVERSED BY SUBROUTINE UNPACK
C
      REAL A(NA,1),B(1)
      K=0
      DO 10 J=1,M
      DO 10 I=J,M
      K=K+1
   10 B(K)=A(I,J)
      RETURN
      END
```

```
*************************************************************************
```

```
      SUBROUTINE UNPACK(M,A,NA,B)
C
C PURPOSE.  TO REVERSE THE OPERATION OF SUBROUTINE PACK.  THE  M X M
C SYMMETRIC MATRIX A IS RECREATED FROM ITS LOWER TRIANGLE WHICH IS
C STORED BY COLUMNS IN CONTINUOUS LOCATIONS IN THE 1-DIMENSIONAL ARRAY
C B, WHICH MAY BE THE SAME AS A.
C
      REAL A(NA,1),B(1)
      K=M*(M+1)/2+1
      DO 10 J1=1,M
      J=M-J1+1
      DO 10 I1=1,J1
      I=M-I1+1
      K=K-1
   10 A(I,J)=A(J,I)=B(K)
      RETURN
      END
```

```
*************************************************************************
```

```
      SUBROUTINE CHINV(IO,N,A,B,DA)
C
C PURPOSE.  TO DETERMINE USING THE CHOLESKY METHOD (SQUARE ROOT METHOD)
C A G-INVERSE, AINV, FOR A SYMMETRIC, POSITIVE SEMIDEFINITE  N X N
C MATRIX A.  ON OPTION TO DETERMINE THE TRIANGULAR DECOMPOSITION
C A = L*LT, WHERE L IS LOWER TRIANGULAR WITH TRANSPOSE LT.
C
C PROGRAMMER.  A.J.SWAIN  JAN 1975.
C
C ARGUEMENTS.  IO  OPTION
C                  1 ON RETURN B = L
C                  2 ON RETURN B = AINV
C                  3 ON CALL B = L,  ON RETURN B = AINV
C              N   SIZE OF MATRICES
C              A   ARRAY WITH N*(N+1)/2 ELEMENTS CONTAINING THE LOWER
C                  TRIANGLE OF THE MATRIX A STORED BY COLUMNS
C              B   ARRAY IN WHICH RESULTS ARE RETURNED.  MAY BE SAME AS
C                  A.  IF NOT, A IS PRESERVED.
C              DA  DETERMINANT OF A - CALCULATED FOR IO=1 OR 2.
```

```
C                     RETURNED AS -1 IF A IS NOT POSITIVE SEMIDEFINITE
C
C NOTE.  A IS TAKEN AS SINGULAR IF THE ABSOLUTE VALUE OF A DIAGONAL
C ELEMENT IN THE DECOMPOSITION IS LESS THAN EPS.
C
      REAL A(1),B(1)
      DATA EPS/1.E-10/
      IND(I,J)=(I-1)*(N2-I)/2+J
C
      N2=2*N
      GO TO (9,9,130),IO
C
C FIRST PASS - FORM L. IF A DIAGONAL ELEMENT IN THE DECOMPOSITION
C IS ALMOST ZERO, SET IT AND THE REST OF THE COLUMN TO ZERO.
C
    9 DA=1.
      DO 70 I=1,N
      IM1=I-1
      II=IND(I,I)
      DO 40 J=I,N
      T=0.
      IF(I.EQ.1)GO TO 20
      DO 10 K=1,IM1
      KI=IND(K,I)
      KJ=IND(K,J)
   10 T=T+B(KI)*B(KJ)
   20 IF(J.NE.I)GO TO 30
      T=A(II)-T
      IF(ABS(T).LT.EPS)GO TO 50
      IF(T.GT.0)GO TO 25
      DA=-1.
      GO TO 9999
   25 DA=DA*T
      B(II)=SQRT(T)
      GO TO 40
   30 IJ=IND(I,J)
      B(IJ)=(A(IJ)-T)/B(II)
   40 CONTINUE
      GO TO 70
   50 DA=0
      DO 60 J=I,N
      IJ=IND(I,J)
      B(IJ)=0
   60 CONTINUE
   70 CONTINUE
  130 GO TO (9999,140,140),IO
C
C SECOND PASS - INVERT L.
C
  140 DO 160 I1=1,N
      I=N-I1+1
      IP1=I+1
      II=IND(I,I)
      IF(B(II).EQ.0)GO TO 160
      B(II)=1./B(II)
      DO 155 J1=I,N
      J=N+I-J1
      IF(J.EQ.I)GO TO 160
      T=0
      DO 150 K=IP1,J
      IK=IND(I,K)
      KJ=IND(K,J)
  150 T=T+B(IK)*B(KJ)
      IJ=IND(I,J)
      B(IJ)=(-T)*B(II)
  155 CONTINUE
  160 CONTINUE
C
C THIRD PASS - FORM AINV = LTINV*LINV.
C
      DO 180 I=1,N
      DO 180 J=I,N
```

```
      T=0
      DO 170 K=J,N
      IK=IND(I,K)
      JK=IND(J,K)
  170 T=T+B(IK)*B(JK)
      IJ=IND(I,J)
      B(IJ)=T
  180 CONTINUE
 9999 RETURN
      END
```

```
************************************************************************

C PURPOSE.  SAMPLE MAIN PROGRAM FOR DTDP.
C
C USAGE.  CARD 1 - TITLE CARD
C         CARD 2 - NDFS, P, P1, P2 (SYMBOLS SAME AS FOR DTDP)
C         CARD 3 AND FOLLOWING - LOWER TRIANGLE OF S PUNCHED ROW BY ROW
C ANY NUMBER OF SETS OF DATA - TERMINATE RUN BY BLANK CARD
C
C NOTE1.  CALLS OUT TO PRINT MATRICES - SEE LISTING OF VARS
C
C NOTE2.  ALPHANUMERIC PART OF INPUT/OUTPUT AND THE TERMINATION TEST
C WOULD ALMOST CERTAINLY NEED CHANGE FOR COMPUTERS WITH DIFFERENT
C WORD SIZE.
C
      INTEGER P,P1,P2
      REAL S(25,25),SI1(5,5),SI2(12,12),SI(25,25)
      COMMON /WORKDP/V(50,50),NV,SPARE(42)
      REAL TITLE(8)
      DATA NS,NSI1,NSI2,NSI/25,5,12,25/
      DATA BLANKS/10H          /
C
   10 READ 1,(TITLE(I),I=1,8)
    1 FORMAT(8A10)
      IF(TITLE(1).EQ.BLANKS)GO TO 9999
      PRINT 2,(TITLE(I),I=1,8)
    2 FORMAT(1H1,8A10)
C
      READ 3,N,P,P1,P2
    3 FORMAT(4I4)
      DO 20 I=1,P
      READ 4,(S(I,J),J=1,I)
    4 FORMAT(9F8)
      DO 20 J=1,I
   20 S(J,I)=S(I,J)
      CALL OUT(P,P,S,NS,1HS,1)
C
      CALL DTDP(N,P,P1,P2,S,NS,SI1,NSI1,SI2,NSI2,SI,NSI,X2,NDFM,IERR)
C
      IF(IERR.NE.0)PRINT 5,IERR
    5 FORMAT(10H0ERROR NO.,I4)
      IF(IERR.EQ.2)GO TO 10
      CALL OUT(P1,P1,SI1,NSI1,6HSIGMA1,1)
      CALL OUT(P2,P2,SI2,NSI2,6HSIGMA2,1)
      CALL OUT(P,P,SI,NSI,5HSIGMA,1)
      PRINT 6,NDFM,X2
    6 FORMAT(8H0CHISQ (,I4,4H ) =,F11.4)
C
      NV=50
      IQ=P*(P+1)/2-NDFM
      IF(IQ.GT.NV)GO TO 10
C
      CALL DPVAR(N,P1,P2,SI1,NSI1,SI2,NSI2,V,NV)
C
      CALL OUT(IQ,IQ,V,NV,1HV,1)
      GO TO 10
C
 9999 STOP
      END

************************************************************************
```

```
SAMPLE DATA FOR DTDP - G.TUDOR  (DATA 2 - MUSCLE E EXCL.)
  38   8   2   4
  1.0000
   .5989  1.0000
   .6112   .4797  1.0000
   .5811   .5130   .6006  1.0000
  1.1091   .6636   .7905   .7181  1.4262
   .6866   .9765   .5406   .6193   .8473  1.1288
   .6390   .4458   .9710   .6319   .8782   .5727  1.0524
   .6620   .6221   .6301   .9494   .8334   .7649   .7049  1.0611

*****************************************************************************
```

BIBLIOGRAPHY

AITCHISON, J. & SILVEY, S.D. (1960). Maximum likelihood estimation procedures and associated tests of significance.
J. Roy. Statist. Soc. B, 22, 154-171.

ANDERSON, T.W. (1958). An introduction to multivariate statistical analysis.
Wiley, New York.

ANDERSON, T.W. (1963). Asymptotic theory for principal component analysis.
Ann. Math. Statist. 34, 122-148.

ANDERSON, T.W. (1969). Statistical inference for covariance matrices with linear structure. In: Krishnaiah, P.R. (Ed.) Multivariate analysis II.
Academic Press, New York, 55-66.

ANDERSON, T.W. (1970). Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices. In: Bose, R.C. et al. (Eds.) Essays in probability and statistics.
University of North Carolina Press, Chapel Hill, 1-24.

ANDERSON, T.W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure.
Ann. Statist. 1, 135-141.

ANDERSON, T.W. & RUBIN, H. (1956). Statistical inference in factor analysis. In: Neyman, J. (Ed.) Proceedings of the third Berkeley symposium on mathematical statistics and probability. University of California Press, Berkeley. Vol. 5, 111-150.

APOSTOL, T.M. (1957). Mathematic al analysis. Addison - Wesley, Reading, Massachusetts.

BARTLETT, M.S. (1954). A note on the multiplying factors for various χ^2 approximations. J. Roy. Statist. Soc. B, 16, 296-298.

BELLMAN, R.E. (1960). Introduction to matrix analysis. McGraw-Hill, New York.

BILODEAU, E.A. (1957). The relationship between a relatively complex motor skill and its components. Am. J. Psychol., 70, 49-55.

BOCK, R.D. & BARGMANN, R.E. (1966). Analysis of covariance structures. Psychometrika, 31, 507-534.

BORUCH, R.F. (1970). Extensions of a multitrait - multimethod model to experimental psychology. Multivariate Behavioural Research. 5, 351-368.

BOX, G.E.P. (1949). A general distribution theory for a class of likelihood criteria. Biometrika, 36, 317-346.

BROWNE, M.W. (1968). A comparison of factor analytic techniques.
Psychometrika, 33, 267-334.

BROWNE, M.W. (1974). Generalised least squares estimators in the analysis of covariance structures.
S. Afr. Statist. J., 8, 1-24.

CAMPBELL, D.T. & FISKE, D.W. (1959). Convergent and discriminant validation by the multitrait - multi-method matrix.
Psychological Bulletin, 56, 81-105.

CAMPBELL, D.T. & O'CONNELL, E.J. (1967). Method factors in multitrait - multimethod matrices: multiplicative rather than additive? Multivariate Behavioural Research.
2, 409-426.

CHERNOFF, H. (1956). Large - sample theory: parametric case.
Ann. Math. Statist., 27, 1-22.

CLARKE, M.R.B. (1970). A rapidly convergent method for maximum - likelihood factor analysis.
Br. J. math. statist. Psychol. 23, 43-52.

CONSTANTINE, A.C. (1963). Some non-central distribution problems in multivariate analysis.
Ann. math. Statist. 34, 1270-1285.

CORBALLIS, M.C. (1973). A factor model for analysing change.
Br. J. math. statist. Psychol. 26, 90-97.

DAS GUPTA, S. (1968). Some aspects of discrimination function coefficients.
Sankhyā A, 30, 387-400.

DAVIS, A.W. (1971). Percentile approximations for a class of likelihood ratio criteria.
Biometrika, 58, 349-356.

DEMPSTER, A.P. (1972). Covariance selection.
Biometrics, 28, 157-175.

ESCOUFIER, Y. (1973). Le traitement des variables vectorielles.
Biometrics, 29, 751-760.

EVANS, G.T. (1967). Factor analytic treatment of growth data. Multivariate Behavioural Research, 2, 109-134.

FERGUSON, T.S. (1958) A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities.
Ann. Math. Statist., 29, 1049-1062.

FISHER, R.A. (1925). Theory of statistical estimation.
Proc. Camb. Phil. Soc., 22, 700-725.

FISKE, D.W. (1949). Consistency of the factorial structures of personality ratings from different sources.
J. abnorm. soc. Psychol. 44, 329-344.

FLETCHER, R. & POWELL, M.J.D. (1963). A rapidly convergent descent method for minimisation.
Computer J. 2, 163-168.

FUJIKOSHI, Y. (1973). Asymptotic formula for the distribution of three statistics for multivariate linear hypothesis.
Ann. Inst. Statist. Math., 25, 423-437.

HAKSTIAN, A.R. (1973). Procedures for the factor analytic treatment of measures obtained on different occasions.
Br. J. math. statist. Psychol. 26, 219-239.

JAMES, A.T. (1964). Distributions of matrix variates and latent roots derived from normal samples.
Ann. Math. Statist. 35, 475-501.

JAMES, A.T. (1973). The variance information manifold and the functions on it. In: Krishnaiah, P.R. (Ed.) Multivariate analysis III.
Academic Press, Yew York, 157-169.

JENNRICH, R.I. & ROBINSON, S.M. (1969). A Newton-Raphson algorithm for maximum likelihood factor analysis.
Psychometrika 34, 111-123.

JENNRICH, R.I. & THAYER, D.T. (1973). A note on Lawley's formulas for standard errors in maximum likelihood factor analysis.
Psychometrika 38, 571-580.

JÖRESKOG, K.G. (1967). Some contributions to maximum likelihood factor analysis.
Psychometrika, 32, 443-482.

JÖRESKOG, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis.
Psychometrika 34, 183-202.

JÖRESKOG, K.G. (1970a). A general method for analysis of covariance structures.
Biometrika, 57, 239-251.

JÖRESKOG, K.G. (1970b). Estimation and testing of simplex models.
Br. J. math. statist. Psychol. 23, 121-145.

JÖRESKOG, K.G. (1971). Statistical analysis of sets of congeneric tests.
Psychometrika, 36, 109-133.

JÖRESKOG, K.G. (1972). Analysis of covariance structures. In: Krishnaiah, P.R. (Ed.) Multivariate analysis III.
Academic Press, New York, 263-285.

JÖRESKOG, K.G. & GOLDBERGER, A.S. (1972). Factor analysis by generalised least squares.
Psychometrika, 37, 243-260.

KELLEY, E.L. & FISKE, D.W. (1951). The prediction of performance in clinical psychology.
Univ. Michigan Press. Ann Arbor.

KOOPMANS, T.C. & REIERSØL, O. (1950). The identification of structural characteristics.
Ann. Math. Statist., 21, 165-181.

KORIN, B.P. (1968). On the distribution of a statistic used for testing a covariance matrix.
Brometrika, 55, 171-178.

KSHIRSAGAR, A.M. (1972). Multivariate analysis.
Marcel Dekker, New York.

KULLBACK, S. (1959). Information theory and statistics.
Wiley, New York.

LAWLEY, D.N. (1940). The estimation of factor loadings by the method of maximum likelihood.
Proc. Roy. Soc. Edin. A, 60, 64-82.

LAWLEY, D.N. (1941). Further investigations in factor estimation.
Proc. Roy. Soc. Edin. A, 61, 176-185.

LAWLEY, D.N. (1956). A general method for approximating to the distribution of likelihood ratio criteria.
Biometrika, 43, 295-303.

LAWLEY, D.N. & MAXWELL, A.E. (1971). Factor analysis as a statistical method. 2nd Edn.
Butterworth, London.

McDONALD, R.P. (1969). A generalised common factor analysis based on residual covariance matrices of prescribed structure.
Br. J. math. statist. Psychol., 22, 149-163.

McDONALD, R.P. (1970). Three common factor models for groups of variables.
Psychometrika, 35, 111-128.

McDONALD, R.P. (1974). Testing pattern hypotheses for covariance matrices.
Psychometrika, 39, 189-201.

MacRAE, E.C. (1974). Matrix derivatives with an application to adaptive linear decision problem.
Ann. Statist., 2, 337-346.

MANN, H.B. & WALD, A. (1943). On stochastic limit and order relationships.
Ann. Math. Statist., 14, 217-226.

MAUCHLY, J.W. (1940). Significance test for sphericity of a normal n-variate distribution.
Ann. Math. Statist., 11, 204-209.

MUKHERJEE, B.N. (1970). Likelihood ratio tests of statistical hypotheses associated with patterned covariance matrices in psychology.
Br. J. math. statist. Psychol., 23, 89-120.

MUKHERJEE, B.N. (1973). Analysis of covariance structures and exploratory factor analysis.
Br. J. math. statist. Psychol., 26, 125-154.

NAGAO, H. (1973). On some test criteria for covariance matrix.
Ann. Statist., 1, 700-709.

NEUDECKER, H. (1969). Some theorems on matrix differentiation with special reference to Kronecker matrix products.
J. Am. Statist. Assoc., 64, 953 963.

NEWTON TURNER, H., and YOUNG, S.S.Y. (1969). Quantitative genetics in sheep breeding.
MacMillan, Melbourne.

NEYMAN, J. (1949). Contributions to the theory of the χ^2 test.
In: Neyman, J. (Ed.) Proceedings of Berkeley Symposium on mathematical statistics and probability.
University of California Press, Berkeley, 239-273.

OLKIN, I. (1973). Testing and estimation for structures which are circularly symmetric in blocks.
In: Kabe, D.G., and Gupta, R.P. (Eds.) Multivariate statistical inference.
North Holland Publishing Co., Amsterdam.

RAO, C.R. (1955) Theory of the method of estimation by minimum chi-square.
Bull. Inst. Inter. Statist., 35, 25-32.

RAO, C.R. (1963). Criteria of estimation in large samples.
Sankya A, 25, 189-206.

RAO, C.R. (1973). Linear statistical inference and its applications. 2nd Edn.
Wiley, New York.

ROTHENBERG, T.J. (1971). Identification in parametric models.
Econometrika, 39, 577-591.

SILVEY, S.D. (1959). The Lagrangian multiplier test.
Ann. Math. Statist., 30, 389-407.

SWAIN, A.J. (1975). A class of factor analysis estimation procedures with common asymptotic sampling properties.
Psychometrika. (to appear).

TAYLOR, W.F. (1953). Distance functions and regular best asymptotically normal estimates.
Ann. Math. Statist., 24, 85-92.

TUCKER, L.R. (1963) Implications of factor analysis of three-way matrices for measurement of change.
In: HARRIS, C.W. (Ed.) Problems in measuring change.
The University of Wisconsin Press, Madison, Wisconsin. 122-137.

TUCKER, L.R. (1966). Some mathematical notes on three-mode factor analysis.
Psychometrika, 31, 279-311.

TUDOR, G.D. (1971). The effect of pre-natal and early post-natal nutrition on the subsequent growth and body composition of beef cattle.
Unpublished M. Ag. Sc. thesis, University of Q'ld.

VOTAW, D.F. (1948). Testing compound symmetry in a normal multivariate distribution.
Ann. Math. Statist., 19, 447-473.

WILKS, S.S. (1946). Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution.
Ann. Math. Statist., 17, 257-281.

ZACKS, S. (1971). The theory of statistical inference.
Wiley, New York.