ON SOME PROBLEMS IN ANALYSIS OF COVARIANCE STRUCTURES

BY

SADHAN SAMAR MAITI

A THESIS SUBMITTED TO THE
INDIAN STATISTICAL INSTITUTE
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE OF
DOCTOR OF PHILOSOPHY

CALCUTTA
JULY, 1990

See Errata at end.

ACKNOWLEDGEMENTS

I express my deep sense of gratitude to my thesis advisor Prof. B.N. Mukherjee for his guidance, encouragement and patience. His keen interest in the subject is contagious and I have learned much from him.

I owe my sincere thanks to the Computer Science Unit of Indian Statistical Institute, Calcutta for providing me facilities for carrying out my work. I am indebted to the University of Kalyani for sanctioning me leave for a period of two years to complete my research work. I also take this opportunity to express my grateful appreciation to the teaching staff of the Dept. of Statistics, University of Kalyani for many helpful academic discussions. I am indebted to Prof. G.S. Sanyal of the Dept. of Chemistry, University of Kalyani for his constant encouragement and untiring help in many directions.

My appreciation goes to Mr. Subhas Datta for carefully typing the thesis.

Finally, I would like to thank my wife for her understanding and inspiration which helped me to complete the thesis.

Sadhan Samar Maiti

CONTENTS

Chapter 1

INTRODUCTION

1.1 AN INTRODUCTION TO ANALYSIS OF COVARIANCE STRUCTURES

In recent years, the techniques of structural analysis of covariance and correlation matrices have frequently been employed especially in the social and behavioural sciences for analyzing multivariate data. Analysis of covariance structures (ACOVS) is a " generic term describing a variety of statistical procedures for testing and measuring the goodness-of-fit of certain types of structures postulated a priori for the covariance matrix by placing alternative restrictions on the parameter matrices of the general model" [Mukherjee, 1976, p. 132].

The acronym 'ACOVS' stands for " analysis of covariance structures" and was first proposed by Bock (1960) as a method for studying the structural and discriminal power of psychological tests which are designed on the basis of certain theories or assumptions. But its techniques are not necessarily the same as those involved in the study of structural relationships (cf. Kendall and Stuart, 1967, p. 374). ACOVS is a kind of pattern analysis of covariance matrices rather than the patterns of responses in the raw data which are generally taken into consideration for configurational scoring (McQuitty, 1956).

Analysis of covariance structures (ACOVS) is mainly aimed at estimating and testing a set of hypotheses of a specific

pattern of relations among a set of measured or manifest variables(MVs) and latent variables (LVs), both dependent and independent. In ACOVS, given an hypothesized model defined in terms of fixed and free parameters, and given a sample covariance matrix, $\underset{\sim}{S}$, for the MVs, one can solve for estimates of the free parameters of the model. The most common approach for fitting the hypothesized model to data is to obtain maximum likelihood (ML) estimates of parameters and an accompanying likelihood ratio chi-square test of the null hypothesis. A variety of other information can be obtained regarding goodness-of-fit of the model (Bentler and Bonett, 1980; Jörskog and Sörbom, 1988; Mulaik, James, Alstine, Bennett, Lind and Stilwell, 1989). The entire process of formulating the hypothesized structural model,the specification of the free and fixed parameters of the deduced covariance matrix, the estimation and testing of the model as well as an operation of sequentially modifying the model so as to improve its fit and/or parsimony is called analysis of covariance structures (ACOVS).

ACOVS has been quite useful in dealing with simultaneous equation models which arise very frequently in social and economic sciences. A set of regression equations is said to comprise a simultaneous equation model if one or more of the explanatory variables in one or more of the equations is itself the dependent (endogenous) variable associated with another equation in the full system. Such models have also been referred to as structural equation systems, dependence analysis, path or causal

analysis with latent variables (Everitt, 1984). These models are employed to analyze phenomena of diverse interests usually in terms of cause and effect variables in behavioural sciences and social research.ACOVS as a generic method "provides a framework for translating general ideas about behaviour into a more explicit quantitative description and encourages comparisons of the relative performance of two or more structures with empirical data" (Cudeck, 1989, p. 317).

A serious limitation of the causal models used till the middle of sixties was that they ignored the effects of errors of measurement which might have existed in substantial amount in the variables analyzed. As is now known, errors of measurement can lead to severe bias in the conclusions which might be drawn from structural equation models that do not take such errors into account (Goldberger, 1973, 1974). It is well-known that errors of measurement in variables attenuate correlations.

Social scientists are also becoming aware of the fact that certain variables involved in theoretical systems in general are beyond explicit and exact operationalization. For example, concepts such as quality of life, motivation, values, economic expectation, economic growth, cultural status, national integrity etc. may not be readily defined in exact operational terms. There is also an increasing realization that unless proper care

is taken to improve the quality of data collection as well as procedures for measuring errors in variables such as age, school achievement, occupational mobility, population growth, income etc., the conclusions would be misleading, if these errors are ignored. In ignoring these errors, the researcher runs the risk of obtaining biased estimates of the parameters.

In order to circumvent the afore-said difficulties it has been suggested that the traditional simultaneous equation model be replaced by new structural models. Structural Equation Modelling (SEM) is currently receiving a lot of attention in the whole field of social and behavioural sciences. In 1983, a special issue on structural equation modelling was published in Journal of Econometrics to discuss the need for integrating into a single comprehensive system the basic models of path analysis (from biology and sociology), factor analysis (from psychology) and simulteneous equation system (from economics). Such integrated models with latent variables are now being handled in various applied sciences by the general procedure of analysis of covariance structures (ACOVS). Bentler (1986) has exhaustively reviewed some major conceptual achievements involving general covariance structures representations, multiple population models and moment structures in the context of changing trends of psychometric research as evident from the five decades of publications of Psychometrika. The second and third issues of Child Development journal (Volume 58) of 1987 devoted to a special section on SEM

aimed to provide a diverse sampling of conceptual, empirical and theoretical work in the general field of ACOVS.

We acknowledge in this connection the notable contribution of Jöreskog and his collaborators who extended the ACOVS procedure to the general linear structural relations (LISREL) models expressing both structural and measurement inferences within a simple comprehensive model framework and thus solving the dual inference problems mentioned above [Jöreskog, 1973a; 1973b, 1974, 1977, 1978, 1981, 1982; Jöreskog and Sörbom, 1981, 1982]. The LISREL approach is now readily implemented in its latest version LISREL VII (Jöreskog and Sörbom, 1988) computer package. This package can handle a wide range of specific models such as confirmatory factor analysis models, covariance structure models, path analysis models, recursive as well as non-recursive models for cross-sectional and longitudinal data.

In all the above-mentioned cases, the covariance matrix has a pattern (or structure). The pattern is usually produced by specifying certain elements of the covariance matrix to zero or to any known constants and/or by imposing restriction of expressing some elements of the covariance matrix as function of the other elements. Thus, a matrix is said to have a pattern or structure if one or more of the following conditions is true :

(a) some elements of the matrix are constant,

(b) some elements are repeated, possibly with a change of sign, and

(c) some elements are functions of the other elements.

The integration as accomplished by LISREL which we discussed above, has brought new problems of model identification, parameter estimation, significance testing and development of new measures of goodness-of-fit for different types of patterned matrices involved in ACOVS. This thesis aims at addressing some of these problems. Covariance structure analysis associated with linear structural equation models is the primary focus of this dissertation.

We discuss in this chapter certain basic models of covariance structures and the main problems involved in the analysis of covariance structures (ACOVS). An attempt has been also made to present a brief review of some of the major applications of ACOVS. After presenting the role of ACOVS in Statistics and a discussion of how it can be regarded as a general statistical method, we delineate the main objectives of the present study. The structure of the thesis is then outlined in terms of organization of various chapters.

1.2 SOME SELECTED MODELS OF COVARIANCE STRUCTURES

Starting with Guttman's (1954) simplex and circumplex models, various structural models including the second order factor model, Tucker's three-mode factor analysis model, the MIMIC model, the Kronecker product models, the models of path analysis and seemingly unrelated regression equations have been

proposed from time to time many of which can be analysed by the general procedure of ACOVS and moment structures. These models are discussed in various sources (e.g.. Bentler, 1976, 1980, 1983; Bock and Bargmann, 1966; Browne, 1982; Jöreskog, 1970, 1973a; 1977; 1978; Krishnaiah and Lee, 1974; Long, 1983a, 1983b;McDonald, 1978, Mukherjee, 1970; 1976; Wiley, Schmidt and Bramble, 1973). As such, our discussion of the models of ACOVS is limited to a very few general class of linear models of covariance structures.

1.2.1 A Model of Congeneric Tests

Let us consider p-variate observational vector linearly composed as

$$\underline{X} = \underline{\mu} + \underset{\sim}{A}\,\underline{f} + \underline{e} \qquad (1.1)$$

where $\underline{\mu}$ is the population mean vector, $\underline{f}$ is m-component vector of unknowns (interpreted as a set of m latent variables) with $E(\underline{f}) = \underline{0}$ and $E(\underline{f}\,\underline{f}') = \underset{\sim}{\Delta} = \text{diag.}\,(\delta_1^2, \delta_2^2, \cdots, \delta_m^2)$, $\underset{\sim}{A} = (\underline{a}_1, \underline{a}_2, \cdots, \underline{a}_m)$ is p x m matrix of unknown constants. The symbol $\underline{e}$ stands for error vector with $E(\underline{e}) = 0$, $E(\underline{e}\,\underline{e}') = \underset{\sim}{\Psi}^2 = \text{diag.}\,(\Psi_1^2, \cdots, \Psi_p^2)$ and $E(\underline{f}\,\underline{e}') = \underset{\sim}{0}$. Therefore, the pxp order covariance matrix of $\underline{X}$ is given by

$$\underset{\sim}{\Sigma} = \underset{\sim}{A}\,\underset{\sim}{\Delta}\,\underset{\sim}{A}' + \underset{\sim}{\Psi}^2 = \sum_{i=1}^{m} \delta_i^2\,\underline{a}_i\,\underline{a}_i' + \underset{\sim}{\Psi}^2 . \qquad (1.2)$$

The above model of covariance structure shows a pattern in the p(p+1)/2 nonduplicated elements involving (m+p) unknown

constants, $m < p(p+1)/2$. In particular, when $\underset{\sim}{A}$ is a lower triangular matrix of unity, equation (1.2) reduces to a patterned covariance matrix showing Guttman's (1954) quasi-simplex structure of the form

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \alpha_1^2 + \Psi_1^2 & & & & \\ \alpha_1^2 & \alpha_2^2 + \Psi_2^2 & & & \\ \alpha_1^2 & \alpha_2^2 & \alpha_3^2 + \Psi_3^2 & \text{sym.} & \\ \vdots & \vdots & \vdots & & \\ \alpha_1^2 & \alpha_2^2 & \alpha_3^2 & \cdots\cdots & \alpha_p^2 + \Psi_p^2 \end{bmatrix} \qquad (1.3)$$

where $\alpha_i^2 = \sum_{j=1}^{i} \delta_j^2$, $i = 1,2, \cdots, p$. The structure as shown in (1.3) has been intensively studied by Mukherjee (1963, 1966, 1969b) and Jöreskog (1970b).

In case the elements of $\underset{\sim}{A}$ are unknown, the total number of unknown parameters in (1.2) would be increased and the resulting pattern would be somewhat different. As for example, for $m = 1$ and $\underset{\sim}{A}(p \times 1) = \underline{\beta} = (\beta_1, \cdots, \beta_p)'$, $\underset{\sim}{\Delta}(1 \times 1) = 1$ in equation (1.2), we get Spearman's unifactor structure as

$$\underset{\sim}{\Sigma} = \underline{\beta}\,\underline{\beta}' + \underset{\sim}{\Psi}^2 = \begin{bmatrix} \beta_1^2 + \Psi_1^2 & & & & \\ \beta_1\beta_2 & \beta_2^2 + \Psi_2^2 & & & \\ \beta_1\beta_3 & \beta_2\beta_3 & \beta_3^2 + \Psi_3^2 & \text{Sym.} & \\ \vdots & \vdots & \vdots & & \\ \beta_1\beta_p & \beta_2\beta_p & \beta_3\beta_p & \cdots & \beta_p^2 + \Psi_p^2 \end{bmatrix} \qquad (1.4)$$

The model (1.4) has $p(p+1)/2$ nonduplicated elements which are functions of $2p$ parameters.

1.2.2 Causal Models

In some causal models, the observed variables, mostly with measurement errors, appear only as effects of certain latent or hypothetical constructs, while in others, the observed variables may appear as causes or as both causes and effects of latent variables (Blalock, 1985). It becomes a straightforward matter to express the causal linkages and the relationships between the variables in equation form once the latent and observed variables for the analysis are selected and defined. These relationships can be formulated in two stages :

(a) relations of exogenous and endogenous variables to hypothetical or latent variables as well as to errors of measurement (a measurement model) and (b) the relations among the hypothetical variables (a structural model). Specification of these two sets of relations provides the information that allows the researcher to deduce the expected form of the covariance matrix

in the population, say $\underset{\sim}{\Sigma}$. This deduced $\underset{\sim}{\Sigma}$ matrix will usually have a pattern and once the structural parameters are identified, the estimable parameters can be always estimated by some known standard procedures.

Consider a causal model having p number of exogenous X variables and q number of endogenous Y variables written in a compact form as follows :

$$\begin{bmatrix} \underline{Y} \\ \underline{X} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{\Lambda}_y & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{\Lambda}_x \end{bmatrix} \begin{bmatrix} \underline{\eta} \\ \underline{\xi} \end{bmatrix} + \begin{bmatrix} \underline{e} \\ \underline{\delta} \end{bmatrix} \quad . \qquad (1.5)$$

In the above measurement model, the variables are expressed in terms of latent dependent ($\underline{\eta}$) and latent independent ($\underline{\xi}$) variables as well as by the error vectors $\underline{e}$ and $\underline{\delta}$. The matrices $\underset{\sim}{\Lambda}_y$ (q x m) and $\underset{\sim}{\Lambda}_x$ (p x n) are regression matrices of $\underline{Y}$ on $\underline{\eta}$ (m x 1) and $\underline{X}$ on $\underline{\xi}$ (n x 1) respectively. The structural model linking the two sets of latent variables may be expressed as

$$\underset{\sim}{B}\underline{\eta} = \underset{\sim}{\Gamma}\underline{\xi} + \underline{\zeta} \qquad (1.6)$$

where $\underset{\sim}{B}$(m x m) and $\underset{\sim}{\Gamma}$ (m x n) are coefficient matrices, $E(\underline{\zeta}) = \underline{0}$. Under the assumption of nonsingular $\underset{\sim}{B}$,

$$\underline{\eta} = \underset{\sim}{B}^{-1}\underset{\sim}{\Gamma}\underline{\xi} + B^{-1}\underline{\zeta} \quad . \qquad (1.7)$$

It is assumed that the residual vector $\underline{\zeta}$ is uncorrelated with $\underline{\xi}$ and the error vectors $\underline{e}$ and $\underline{\delta}$ are uncorrelated with $\underline{\eta}$ and $\underline{\xi}$ and $\underline{\zeta}$ but may be correlated among themselves within each set. Let

us denote the covariance matrices of $\underset{\sim}{\xi}$ and $\underset{\sim}{\zeta}$ respectively as $\underset{\sim}{\Phi}$ (n x n) and $\underset{\sim}{\Psi}$(m x m). Let $\underset{\sim}{\epsilon}_x$ and $\underset{\sim}{\epsilon}_y$ be the covariance matrices of $\underset{\sim}{e}$ and $\underset{\sim}{\delta}$ respectively. Under these assumptions, the covariance matrix of (p+q)-component vector (1.5) is given by

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \underset{\sim}{\Lambda}_y \underset{\sim}{B}^{-1} \underset{\sim}{\Gamma} \underset{\sim}{\Phi} \underset{\sim}{\Gamma}' \underset{\sim}{B}^{-1} \underset{\sim}{\Lambda}_y' + \underset{\sim}{\epsilon}_y & \underset{\sim}{\Lambda}_y \underset{\sim}{B}^{-1} \underset{\sim}{\Gamma} \underset{\sim}{\Phi} \underset{\sim}{\Lambda}_x' \\ \underset{\sim}{\Lambda}_x \underset{\sim}{\Phi} \underset{\sim}{\Gamma}' \underset{\sim}{B}^{-1} \underset{\sim}{\Lambda}_y' & \underset{\sim}{\Lambda}_x \underset{\sim}{\Phi} \underset{\sim}{\Lambda}_x' + \underset{\sim}{\epsilon}_x \end{bmatrix} \qquad (1.8)$$

For the purpose of illustration, let us consider the following causal model of school achievement (ACH) defined in terms of two dependent measures Y_1 and Y_2 having errors of measurement e_1 and e_2 respectively. ACH is conceptualized as a dependent latent variable (η) which is theoretically influenced by two causal latent variables (LVs), namely, socio-economic status (SES) (ξ_1) and ability (ξ_2). These latent variables (LVs) are measured by two sets of independent variables, namely, X_1 and X_2 as well as X_3 and X_4 respectively. Let δ_k be the error of measurement of endogenous variable X_k , k = 1,2,3 and 4. The factor loadings relating latent variables to (observed) manifest variables (MVs) are represented by λ_j (j = 1,2, ... , 6). The path diagram corresponding to this example is shown in Fig. 1.1.

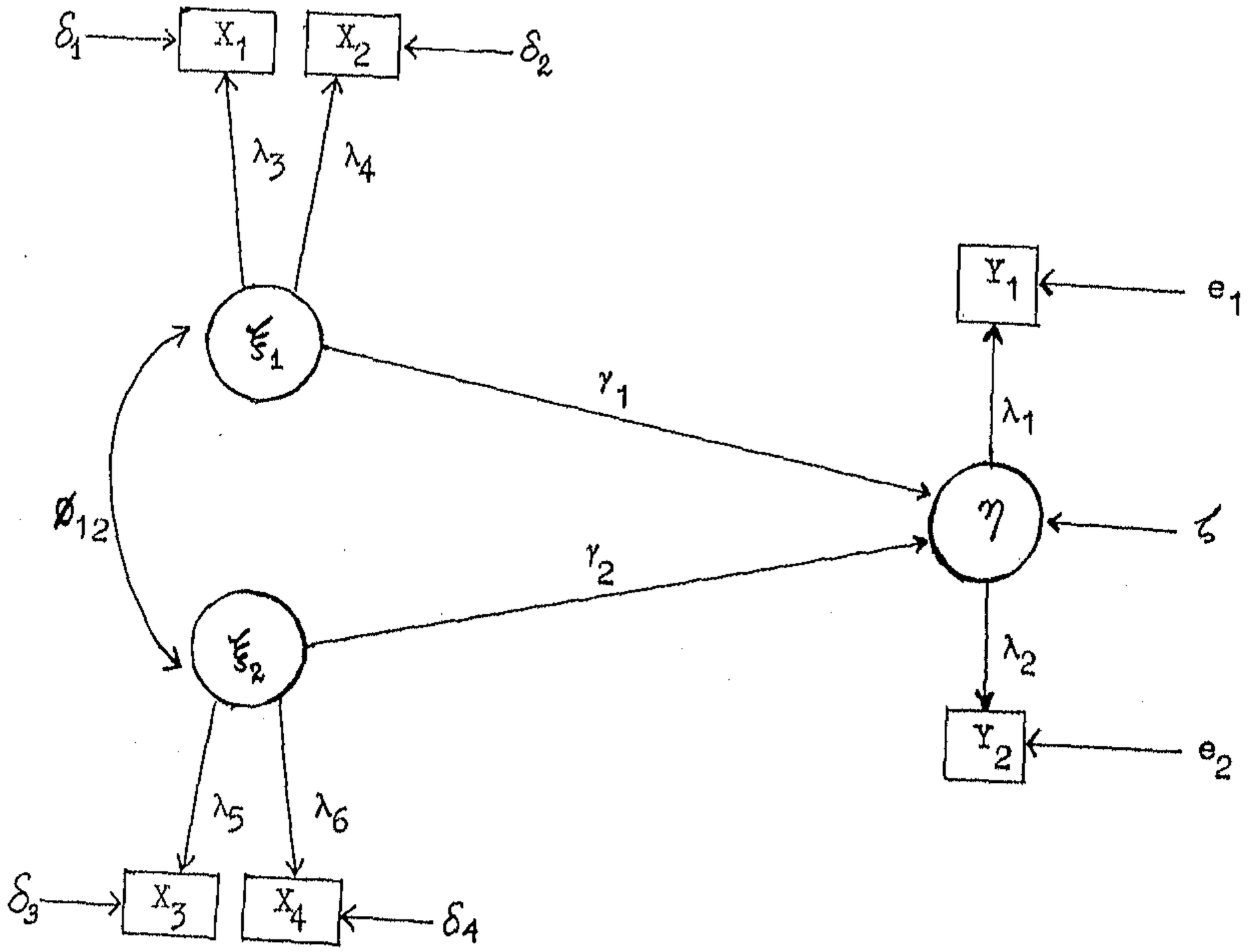

Fig. 1.1 A Causal Model Linking School Achievement with SES and Ability.

In the afore-said example, the measurement model can be written as

$$\begin{bmatrix} y_1 \\ y_2 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ \lambda_2 & 0 & 0 \\ 0 & \lambda_3 & 0 \\ 0 & \lambda_4 & 0 \\ 0 & 0 & \lambda_5 \\ 0 & 0 & \lambda_6 \end{bmatrix} \begin{bmatrix} \eta \\ \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix} \qquad (1.9)$$

The structural equations of the causal model defining the relationship among the two exogenous variables, namely, SES(ξ_1) and ability (ξ_2) and one endogenous variable , namely, School Achievement (η) can be written as :

$$\underset{\sim}{\Phi} = E(\underline{\xi}\,\underline{\xi}') = \begin{bmatrix} 1 & \phi_{12} \\ \phi_{12} & 1 \end{bmatrix} \qquad (1.9a)$$

and $$\eta = \gamma_1 \xi_1 + \gamma_2 \xi_2 + \zeta \qquad (1.9b)$$

where $E(\zeta) = 0$ and $Var(\zeta) = \sigma^2_\zeta$. The variances of ξ_1 and ξ_2 are set to unity, i.e. correlation with themselves. With the assumption of uncorrelated errors, $\underset{\sim}{\Sigma}$ matrix for the six observed variables can be written as

$$\underset{\sim}{\Sigma} = \begin{bmatrix}
\lambda_1^2(\gamma_1^2+\gamma_2^2)+\sigma^2_\zeta+\sigma^2_{e_1} & & & & & \\
\lambda_1\lambda_2\gamma_1^2+\gamma_1\gamma_2\phi_{12}(\lambda_1+\lambda_2) & \lambda_2^2(\gamma_1^2+\gamma_2^2) & & & & \\
+\sigma^2_\zeta & +\sigma^2_\zeta+\sigma^2_{e_2} & & & & \\
\lambda_1\lambda_3\gamma_1 & \lambda_2\lambda_3\gamma_1 & \lambda_3^2+\sigma^2_{\delta_1} & \text{Symmetric} & & \\
\lambda_1\lambda_4\gamma_1 & \lambda_2\lambda_4\gamma_1 & \lambda_3\lambda_4 & \lambda_4^2+\sigma^2_{\delta_2} & & \\
\lambda_1\lambda_5\phi_{12} & \lambda_2\lambda_5\phi_{12} & \lambda_3\lambda_5 & \lambda_4\lambda_5\phi_{12} & \lambda_5^2+\sigma^2_{\delta_3} & \\
\lambda_1\lambda_6\phi_{12} & \lambda_2\lambda_6\phi_{12} & \lambda_3\lambda_6 & \lambda_4\lambda_6\phi_{12} & \lambda_5\lambda_6 & \lambda_6^2+\sigma^2_{\delta_4}
\end{bmatrix}$$

If for computational convenience we set $\lambda_1 = \lambda_2 = 1$, $\lambda_3 = \lambda_4 = \lambda_5 = \lambda_6 = \lambda$ and $\gamma_1 = \gamma_2 = \gamma$, then the covariance matrix can be expressed as

$$\underset{\sim}{\Sigma} = \begin{bmatrix} 2\gamma^2+\sigma_\zeta^2+\sigma_{e_1}^2 & & & & & & \\ \gamma^2(1+2\phi_{12})+\sigma_\zeta^2 & 2\gamma^2+\sigma_\zeta^2+\sigma_{e_2}^2 & & & & & \\ \lambda\gamma(1+\phi_{12}) & \lambda\gamma(1+\phi_{12}) & \lambda^2+\sigma_{\delta_1}^2 & \text{symmetric} & & & \\ \lambda\gamma(1+\phi_{12}) & \lambda\gamma(1+\phi_{12}) & \lambda^2 & \lambda^2+\sigma_{\delta_2}^2 & & & \\ \lambda\gamma(1+\phi_{12}) & \lambda\gamma(1+\phi_{12}) & \lambda^2\phi_{12} & \lambda^2\phi_{12} & \lambda^2+\sigma_{\delta_3}^2 & & \\ \lambda\gamma(1+\phi_{12}) & \lambda\gamma(1+\phi_{12}) & \lambda^2\phi_{12} & \lambda^2\phi_{12} & \lambda^2\phi_{12} & \lambda^2+\sigma_{\delta_4}^2 & \end{bmatrix} \quad (1.10)$$

In the above matrix, there are 6 error variances in addition to σ_ζ^2 , γ, ϕ_{12} and λ structural parameters. These 10 parameters are overidentified from 21 equations thus giving 11 degrees of freedom for testing the fit of the above covariance structure model to empirical sample data.

The causal model described by the equations (1.5), (1.7) and (1.8) represents a general class of linear structural relations (LISREL) models as discussed in several sources (Long, 1983b). It is a generalization of similar previous models such as the longitudinal factor analysis model of Corballis and Traub (1970), Bentler (1973) as well as the model of Keesling (1972) and Jöreskog (1973a). This general model is readily implemented in LISREL VII (latest version) computer packages of Jöreskog and Sörbom (1988). The LISREL procedure can be suitably modified for the case of non-metric variables by

using tetrachoric, polychoric and polyserial correlations in stead of variances and covariances and by estimating the parameters by means of ordinary least squares (OLS) method.

1.2.3 Other Models of Covariance Structures

It may be noted that the original model proposed by Jöreskog (1970a) for ACOVS is equivalent to a second-order oblique factor model in which any element of any matrix can be prescribed to be constant or constrained to be equal to one or more other elements. McDonald and Swaminathan (1972) combined Jöreskog's (1970a) model with that of McDonald's (1969) generalized common factor model to yield a second-order common factor model with residual covariance matrices of prescribed structures.

Subsequently, Wiley, Schmidt and Bramble (1973) described a class of eight covariance structure models which can be regarded as a special case of the general covariance structure model of Jöreskog (1977) of the form

$$\underset{\sim}{\Sigma} = \underset{\sim}{B}(\underset{\sim}{\Lambda}\,\underset{\sim}{\Phi}\,\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi}^2)\,\underset{\sim}{B}' + \underset{\sim}{\epsilon}^2 \,. \qquad (1.11)$$

According to Bentler and Weeks (1982), " this model is extremely general, containing such special cases as the models of Bock and Bargmann (1966) and Wiley et al. (1973), as well as the MIMIC model of Jöreskog and Goldberger (1975) and numerous other models such as MANOVA or patterned covariance matrices (Browne, 1977). This model introduces the idea of higher order

LVs which was first exploited by Thurstone (1947). In such a model, some LVs have no 'direct' effect on MVs" (p. 752). The model has received extensive circulation through the ACOVS program. McDonald (1978) also proposed a simple comprehensive model which contains as special cases a large number of models for the analysis of covariance structures. Krishnaiah and Lee (1976) proposed a general model for ACOVS which can be written as

$$\underset{gxg}{\Sigma} = \underset{\sim}{G}_1 \otimes \underset{\sim}{\Sigma}_1 + \underset{\sim}{G}_2 \otimes \underset{\sim}{\Sigma}_2 + \cdots + \underset{\sim}{G}_k \otimes \underset{\sim}{\Sigma}_k \qquad (1.12)$$

where $\underset{\sim}{\Sigma}_1, \underset{\sim}{\Sigma}_2, \cdots, \underset{\sim}{\Sigma}_k$ are unknown pxp matrices and $\otimes$ denotes the Kronecker product. This model is useful for testing the block versions of testing the hypotheses for (i) sphericity, (ii) intraclass structure and (iii) circular symmetry or circumplex structure.

The general moment structure models as proposed by Bentler and Weeks (1979) can be written as

$$\underset{\sim}{\Sigma} = \underset{\sim}{G} (\underset{\sim}{I} - \underset{\sim}{\Lambda})^{-1} \underset{\sim}{\Phi} (\underset{\sim}{I} - \underset{\sim}{\Lambda})'^{-1} \underset{\sim}{G}' \qquad (1.13)$$

where $g = (k+1)p$, $\underset{\sim}{G} = [\underset{\sim}{O}, \cdots, \underset{\sim}{I}]$ is a known supermatrix of order gxg with k null block elements and a pxp identity matrix $\underset{\sim}{I}_p$. The matrix $\underset{\sim}{\Phi}$ represents the covariance matrix of the latent variable $\underline{\xi}$ where the supervector $\underline{\xi}$ of all vectors is arranged as $\underline{\xi}' = [\underline{\xi}_k', \cdots, \underline{\xi}_1', \underline{\xi}_o']$ with $\underline{\xi}_o$ being a null vector of p dimension. The conforming supermatrix $\underset{\sim}{\Lambda}$ is null except for matrix elements $\Lambda_k, \cdots, \Lambda_1$ in blocks immediately below the diagonal

blocks, i.e. with (k+1) blocks sequenced by the subscripts in $\underset{\sim}{\xi}$, the (i-1,i) block contains Λ_i.

Extending the work of Swain (1975), Browne (1984) proposed a class of direct product models of multitrait-multimethod data. The form of the complete direct-product model is

$$\underset{\sim}{\Sigma} = \underset{\sim}{D}_{\xi} \left(\underset{\sim}{P}_m \circledast \underset{\sim}{P}_t + \underset{\sim}{D}^2_{\eta} \right) \underset{\sim}{D}_{\xi} \qquad (1.14)$$

where $\underset{\sim}{P}_m$ and $\underset{\sim}{P}_t$ are correlation matrices among methods and trait common scores respectively. The diagonal matrix $\underset{\sim}{D}_{\xi}$ represents the common-score standard deviations, while the diagonal matrix $\underset{\sim}{D}^2_{\eta}$ represents the ratios of unique-score variances to common-score variances. The direct product, $\circledast$, between the component correlation matrices $\underset{\sim}{P}_m \circledast \underset{\sim}{P}_t$, itself yields a correlation matrix so that diagonal $(\underset{\sim}{P}_m \circledast \underset{\sim}{P}_t) = \underset{\sim}{I}$. Elements of $\underset{\sim}{D}^2_{\eta}$ are non-negative. Therefore, diagonal elements of $(\underset{\sim}{P}_m \circledast \underset{\sim}{P}_t + \underset{\sim}{D}^2_{\eta})$ in general are greater than unity. The order of $\underset{\sim}{\Sigma}$ is pq x pq since there are p number of traits and q number of methods for assessing these p traits. The above model is scale invariant, and the parameters in $\underset{\sim}{P}_m$, $\underset{\sim}{P}_t$ and $\underset{\sim}{D}_{\eta}$ are scale-free, so that a standard algorithm can be used to solve for the parameters (Cudeck, 1989). The rationale for this type of general model in the context of multitrait multimethod covariance matrix has been described and illustrated with several sets of empirical data by Cudeck (1988).

1.3 MAIN PROBLEMS IN ACOVS

The main statistical problems involved in ACOVS can be broadly discussed under (a) the estimation of the parameters in the hypothesized covariance structure arising from a set of structured variates after the model is properly identified, and (b) derivation of different tests of statistical significance and measures of the goodness-of-fit of the specified covariance structure model with the corresponding sample covariance matrix. Estimation of the magnitude and sign of the model parameters has been treated here essentially as an optimization procedure for determining parameter values of the model in such a manner that the theoretical variance-covariance matrix $\underset{\sim}{\Sigma}$ imposed by the model (null hypothesis) fits the sample variance - covariance matrix $\underset{\sim}{S}$ as closely as possible. When the covariance model can not be written in linear form, the problem assumes greater complexity. These problems are discussed in the subsections to follow.

1.3.1 Problem of Estimation and Testing

General problems of ACOVS have been considered by many authors, e.g. Wilks (1946), Anderson (1969, 1970), Bock and Bargmann (1966), Srivastava (1966), Srivastava and Maik (1967), Mukherjee (1970, 1976), Styan (1969), Jöreskog (1969, 1970a, 1973a, 1973b, 1977), Krishnaiah and Lee (1976) as well as McDonald (1974, 1978, 1980). Quite a few likelihood ratio tests have been developed for certain specific covariance structures,

e.g. by J. Roy (1951, 1954), Srivastava (1965), Olkin and Press (1969), Mukherjee (1966, 1970, 1973) and Jöreskog (1970b). Mukherjee (1970) dealt with a broad class of patterned covariance matrices which are either 'reducible' (spectrally decomposable) or 'non-reducible' but transformable to tri-diagonal form by a known matrix. Mukherjee (1970, 1976) gives a comprehensive review of linear covariance structures where maximum likelihood (ML) estimates can be expressed in closedform.

Excepting a few class of structures (Mukherjee, 1981; Rogers and Young, 1977), ML estimates of the parameters of $\underset{\sim}{\Sigma}$ are obtainable from the closed-form expressions of the likelihood equations (Anderson, 1969, 1970; Styan, 1969). Burg, Luenberger and Wenger (1982) proposed an iterative algorithm that solves the necessary gradient equations for ML estimation. Quasi-Newton method formulated by Fletcher and Powell (1963) has been used by Jöreskog (1970a) and Wiley et al. (1973) for obtaining the ML solution for a class of covariance structures.

In contrast to dealing with a broad class of models, another type of structures has been discoussed without reference to any specific models by Anderson (1973), McDonald (1974) and Browne (1974). In this model-free approach to analysis of covariance structures, the elements of $\underset{\sim}{\Sigma}$ are assumed to be generated as a function of a vector of more basic parameters ($\underline{\theta}$ ($k \times 1$)), i.e. $\underset{\sim}{\Sigma} = \underset{\sim}{\Sigma}(\underline{\theta})$. Jöreskog (1978), Lee and Jennrich (1979), Lee and Bentler (1980) and others adopted this approach in various contexts.

While estimating the parameters in ACOVS most of the above-mentioned authors employ the maximum likelihood (ML) procedure under the usual multinormal assumption. A special advantage of the ML procedure is that it readily provides the data necessary for the likelihood-ratio (LR) test in order to test the statistical goodness-of-fit of the hypothesized covariance structure model.

Browne (1974, 1982) provides the generalized least squares (GLS) estimators without multinormal assumption. He also shows that the estimates obtained from the GLS procedure and maximum Wishart likelihood (MWL) procedure under the assumption of Wishart distribution for sample covariance matrix ($\underset{\sim}{S}$), cover almost the same values. But GLS estimation requires less computations specially when the weight matrix is taken as $\underset{\sim}{S}^{-1}$. The GLS estimate so obtained enjoys most of the nice asymptotic properties of the ML estimate such as consistency, asymptotic normality and efficiency [Browne, 1974, 1977].

Browne (1982, 1984) developed an asymptotically distribution free (ADF) estimator for covariance structure models belonging to the family of GLS estimators. This development emphasizes the role of kurtosis of non-normally distributed data in ACOVS. Tanaka (1984) overcame some objections and shortcomings of ADF estimator of Browne and of iteratively reweighted least squares estimate of Lee and Jennrich (1979). He presented

a quasi-likelihood approach which is an extension of the results reported by Browne (1982, 1984) as well as Lee and Jennrich (1979) in the normal theory case to the case of ADF estimation. Further details of these results are given in Tanaka's (1984) unpublished doctoral dissertation.

Lee (1977) studied the comparative efficiency of five different algorithms, namely, the Gauss-Newton, Newton-Raphson, Scoring, Fletcher-Powell and Fletcher-Reeves algorithms in obtaining the weighted least squares and maximum likelihood (ML) estimator $\underline{\theta}$, which is the vector of unknown parameters in a general covariance structure model, $\underset{\sim}{\Sigma}(\underline{\theta})$. From the comparisons, the first algorithm appeared to be more efficient in certain respects. The standard error estimates obtained from the Fletcher-Powell as well as Fletcher-Reeves algorithms seemed questionable.

1.3.2 The Problem of Identifiability

One of the vexing problems of ACOVS is the question of identifiability of the model parameters. Identifiability of the parameters ($\underline{\theta}$) involved in $\underset{\sim}{\Sigma}(\underline{\theta})$ must be taken into consideration before estimation of the unknown parameters of $\underset{\sim}{\Sigma}$. Hence, we need to examine the identification status of the covariance structure model before we proceed to estimate its parameters. If identification conditions are satisfied (see Fisher, 1966 for an exhaustive analysis), the structural equation models can be estimated by a variety of consistent methods, the most common being the ML procedure.

Identifiability (or identification) of the parameters depends essentially upon the unique specification of the parameters (Jöreskog, 1970a; Wert, Jöreskog and Linn, 1973). In fact, attempts to estimate models that are not identified results in arbitrary estimates of the parameters and consequently meaningless interpretations [Long, 1983a, p. 35]. If a particular model is underidentified, it means that some of the parameters in the model have no unique solution and therefore additional constraints must be added. Otherwise, the researcher will be compelled to abandon the model as untestable.

Exactly identified models certainly possess a unique set of parameters consistent with the data at hand, but there is no information left over for statistically testing the fit of the model since the degrees of freedom in such cases will be exactly zero. From the point of view of theory testing, only overidentified models are useful, since they have more than enough information to determine the parameters. Thus, in the case of underidentified or just identified models, it is necessary to re-specify the underlying theory so as to make the model interpretation overidentified. This can be done, for example, by fixing and/or by constraining further parameters in agreement with a priori theoretical deduction.

Much work on the problem of identification in the field of econometrics (Fisher, 1966) has been done and various terms have been coined to indicate the role which different variables play in achieving

identification of any systems of linear structural equations. Thus, exogenous variables are those which are known to be uninfluenced by other variables in the system and which are included for the part they are supposed to play in determining the dependent or endogenous variables. Each exogenous variable thus allows at least one element of the structural equations to be specified. This is, of course, of no help in achieving identification of the endogenous variables, and generally, what is required for this purpose are exogenous variables which are known not to appear in the equations of some of the endogenous variables. In particular, if it is possible to choose for each endogenous variable a corresponding exogenous variable which occurs only once in the equation of that variable, complete identification of the model parameters is then possible.

One very simple condition for identifiability is available in terms of how the number of observed variances and covariances among p exogenous and q endogenous variables exceeds k number of distinct parameters :

$$k \leq \frac{1}{2}(p + q)(p + q - 1).$$

With the exception of this condition, only few guidelines can be found in the literature for examining the identification status of covariance structure models.

It can be readily seen that every vector θ in the admissible parameter space $(H)_k$ generates a $\underset{\sim}{\Sigma}(\underline{\theta})$ but two or more $\underline{\theta}$

may have a possibility to generate the same $\underset{\sim}{\Sigma}(\underline{\theta})$. The structure is said to be totally identifiable if for any two vectors $\underline{\theta}^*$ and $\underline{\theta}^{**}$ in $(H)_k$, locally or globally, $\underline{\theta}^* \neq \underline{\theta}^{**}$ implies that $\underset{\sim}{\Sigma}(\underline{\theta}^*) \neq \underset{\sim}{\Sigma}(\underline{\theta}^{**})$, i.e. $\underset{\sim}{\Sigma}(\underline{\theta})$ is generated by one and only one $\underline{\theta}$. Then only one can be certain that the parameters are properly identified. If the necessary and sufficient conditions for identification of structure parameters are satisfied, no ambiguity arises in the way of interpreting the empirical estimates and in obtaining their unique solutions. If the structure is identified, then the information matrix of the parameters is almost certainly positive definite (Jöreskog, 1978). Therefore, a positive definite information matrix, according to Jöreskog and Sörbom (1981) almost implies identification. McDonald and Krane (1977) believe that parameters are unambiguously locally identified if the Hessian matrix is non-singular. However, subsequent Monte Carlo studies by McDonald and Krane (1979) who now retract their earlier claims on unambiguous identifiability showed that the data-based evaluation of identification may be incorrect.

Following Mukherjee (1963) or Browne and Shapiro (1988), if $\underset{\sim}{\Sigma}(\underline{\theta}^*) = \underset{\sim}{\Sigma}(\underline{\theta}^{**}) \Longleftrightarrow \theta_i^* = \theta_i^{**}$, $i = 1,2, \ldots , k$, then the parameters in $\underset{\sim}{\Sigma}(\underline{\theta})$ are called identifiable and consequently the covariance structure is said to be identifiable with reference to the parameters.

An identifiable $\underset{\sim}{\Sigma}(\underline{\theta})$ is called <u>linearly structured</u> when it is expressed as

$$\underset{\sim}{\Sigma}(\underline{\theta}) = \sum_{i=1}^{k} \underset{\sim}{H}_i \, \theta_i \qquad (1.15)$$

where $\underset{\sim}{H}_i$'s (called design matrices) are known, symmetric and linearly independent in the sense that the matrix $\underset{\sim}{H}(p^2 \times k)$ having columns as the vector configuration of $\underset{\sim}{H}_i$'s, would have full column rank. Without any loss, through suitable reparametrization, if necessary, such linear independence may be ensured by $tr(\underset{\sim}{H}_i \, \underset{\sim}{H}_j) = 0$ for all $i \neq j$ due to following reasons.

Since there exists a kxk nonsingular matrix F such that the columns of the matrix $\underset{\sim}{G} = \underset{\sim}{H}\underset{\sim}{F}$ are orthogonal and vector configuration of k suitably chosen symmetric matrices $\underset{\sim}{G}_1$, $\underset{\sim}{G}_2$, ... , $\underset{\sim}{G}_k$, through reparametrization $\underline{\gamma} = \underset{\sim}{F}^{-1}\underline{\theta}$, (1.15)may be reexpressed as $\underset{\sim}{\Sigma}(\underline{\gamma}) = \sum_{i=1}^{k} \underset{\sim}{G}_i \, \gamma_i$ such that $tr(\underset{\sim}{G}_i \, \underset{\sim}{G}_j) = 0$ for all $i \neq j$.

It may happen in some practical situations that although the parameters are identifiable and expressible as (1.15), available $\underset{\sim}{H}_i$'s are not all linearly independent. In such " seemingly non-linear" cases, it is very often possible to get a different set of linearly independent $\underset{\sim}{H}_i$'s through a suitable reparametrization. As for example, we may recall equ. (1.3). Once we reparametrize its last diagonal element (as a simple parameter), all the (2p-1) parameters are identifiable but the relevant design matrices (when expressed as (1.15)) are not all linearly independent. Such " seemingly non-linear" structure may be converted to a linear structure simply by further reparametrizing the other diagonal elements.

The spherical structure, e.g. $\underset{\sim}{\Sigma} = \sigma^2 \underset{\sim}{I}_p$ is the simplest linear structure and in a sense most restrictive one also. When along with equal diagonal elements, all off-diagonal elements are set equal, we get the uniform (or intraclass) structure (Geisser, 1964) which very often arises in the analysis of data related to repeated measurements and/or involving exchangeable variables (Srivastava, 1965). There are various other linear structures involved in many occasions, e.g. simplex, quasi-simplex, circumplex (Guttman, 1954, 1957; Anderson, 1960; Mukherjee, 1966; Cudeck, 1986), compound symmetry (Votaw, 1948), equipredictability (Bargmann, 1957), tri-diagonal Jacobi (Mukherjee, 1970), democratic (Jones, 1960), aristocratic structures (Jones, 1960) etc. Most of these covariance structures are overly identified.

1.3.3 Problems with Non-linear Covariance Structures

In contrast to linear structures, we shall say an identifiable $\underset{\sim}{\Sigma}(\underline{\theta})$ involving k unknown parameters to have a non-linear structure when either $\underset{\sim}{\Sigma}(\underline{\theta})$ could not be expressed entirely as (1.15), or when so expressed, $\underset{\sim}{H}_i$'s would involve some parameters in addition to known values. As for example, reparametrizing the diagonal elements, equation (1.4) may be expressed as (1.15) but here as many as (p-1) out of (2p-1) number of design matrices would involve parameters.

We may come across in the literature a variant to structured covariance matrix when $\underset{\sim}{\Sigma}$ is expressed as

$$\underset{\sim}{\Sigma} = \underset{\sim}{D}_{\sigma} \underset{\sim}{\rho} \underset{\sim}{D}_{\sigma} \qquad (1.16)$$

where $\underset{\sim}{D}_{\sigma}$ = diag. $(\sigma_1, \sigma_2, \cdots, \sigma_p)$ and $\underset{\sim}{\rho}$ is a structured correlation matrix. The diagonal matrix $\underset{\sim}{D}_{\sigma}$ may be structured in the sense that σ_i's may be functionally dependent among themselves but not so upon the elements of $\underset{\sim}{\rho}$ matrix. Although (1.16) is a case of non-linear covariance structure, the main attention is focussed on the structure of correlation matrix which may be linear or non-linear. As for example, a Toeplitz (or Laurent) correlation structure given by

$$\underset{\sim}{\rho} = \begin{bmatrix} 1 & & & & \\ \rho_1 & 1 & & & \\ \rho_2 & \rho_1 & 1 & \text{Sym.} & \\ \vdots & \vdots & \vdots & & \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} \cdots\cdots & \rho_1 & 1 \end{bmatrix} \qquad (1.17)$$

and an equally spaced simplex structure given by

$$\underset{\sim}{\rho} = \begin{bmatrix} 1 & & & & \\ \rho & 1 & & & \\ \rho^2 & \rho & 1 & \text{Sym.} & \\ \vdots & \vdots & \vdots & & \\ \rho^{p-1} & \rho^{p-2} & \rho^{p-3} \cdots\cdots & \rho & 1 \end{bmatrix} \qquad (1.18)$$

are respectively the linear and non-linear correlation structures. Some of the non-linear structures may be handled by ACOVS after necessary reparametrization or linearlization.

1.3.4 Problem of Developing Measures of Goodness-of-Fit

Once the parameters of the covariance structure model are estimated, researchers are next concerned with the issue of how to best ascertain the model fit. The most popular method of assessment of fit is the chi-square goodness-of-fit statistic obtained under ML and GLS methods of estimation (Jöreskog and Sörbom, 1984). Jöreskog and Sörbom (1985) also provided fit indices for covariance structure models under ML and OLS (ordinary least squares) methods of estimation. Tucker and Lewis (1973) proposed a measure of goodness-of-fit based on the criterion of percentage of variance explained by the model. Bentler and Bonett (1980) suggested normed and non-normed fit indices based on null-model logic. Tanaka and Huba (1985) were successful in obtaining the statistical origin of these indices.

The above-mentioned chi-square goodness-of-fit statistic is, however, dependent on sample size and sensitive to data non-normality (Bentler and Bonett, 1980; Tanaka and Huba, 1987). Unlike this statistic, sampling distributions of no other fit indices are available. Their behaviours are still quite unknown with regard to sample size, normality or the complexity of the structure. Thus, the procedure of goodness-of-fit assessment is

still influenced by sample size, method of estimation of parameters and specification of the structures, (La Du and Tanaka, 1989).

Specifically, the application of the likelihood ratio chi-square test is unjustified in many practical situations because we must assume that (a) the observed variables have multivariate normal distribution, (b) the analysis is based on a sample covariance matrix $\underset{\sim}{S}$, rather than a sample correlation matrix $\underset{\sim}{R}$, and (c) the sample size is large enough to justify the asymptotic properties of the chi-square test. In practical situations, there is a high probability that at least one of the above-mentioned assumptions are not justified.

We also note in this context the inappropriateness of the classical hypothesis testing approach in assessing how well a given covariance structure model fits a particular data set since the focus of model testing is not either the absolute fit or the rejection of any particular null hypothesis but the comparative plausability of a set of competing models (Jöreskog, 1974; Steiger, 1989). Rejection of a specified covariance structure hypothesis does not give in general any indications as to why the model is inadquate. Similarly, acceptance of a particular covariance structure model does not necessarily imply that it is the correct model for the data. Specifically a good model fit does not imply that all the hypothesized relationships

imposed by the model are statistically significant. Acceptance of the model only implies that the model represents one of the plausible representations for the population variance-covariance matrix. There may be many other untested models that may provide a better or at least comparable fit to the observed data (Mukherjee, 1970, p. 116).

Population fit for a covariance structure model with positive degrees of freedom is never really perfect. Testing whether it is perfect makes little sense. It is what statisticians sometimes call an " accept-support" hypothesis test, because accepting the null hypothesis supports what is generally the experimenter's point of view, i.e., that the model does fit. Accept-support hypothesis tests are subject to a host of problems" (Steiger, 1989, p. 76-77).

We therefore require other procedures of model evaluation in ACOVS. In addition to statistical evaluation of the goodness-of-fit of a hypothesized model, certain non-statistical assessments should simultaneously be undertaken. From the point of view of parsimony, adequacy of a structure is sometimes evaluated not only by degree of model fit but also by the number of parameters required to be estimated to reach that degree of fit (cf. Mulaik et al., 1989).

Akaike (1974) and Schwartz (1978) propose fit indices that incorporate penalty functions based on the number of

parameters that are estimated. Cudeck and Browne (1983), Browne and Cudeok (1989) propose rescaled versions of these indices expressed in terms of chi-square goodness-of-fit statistic to eliminate the effect of sample size by considering covariance structure as a nonlinear fixed predictor regression model. Sörbom (1989) has suggested a "modification index" to detect a "better" model. This modification index is equal to the expected decrease in the chi-square value if a single constraint in the model specification is relaxed. In using the modification index, only one parameter is usually relaxed at a time since freeing one parameter may reduce or eliminate the improvement in fit possible by freeing a second parameter. However, there is no guarantee that the models suggested by the modification indices are identifiable.

1.3.5 Availability of Computer Soft-ware

There are many soft-ware packages useful for estimating the parameters and evaluating the fit of the structured covariance matrices to sample covariance matrix. Most of these programmes were developed from the confirmatory factor analysis point of view. LISREL (version I through VII) of Jöreskog and Sörbom (1988), EQS (structural equation programme, Bentler, 1989) and MILS (multiple indicator linear structure models, Schoenberg, 1982) are most commonly available programmes for ACOVS. Just in the way the complete programme ACOVSM of Jöreskog, van Thillo

and Gruvaeus (1971) was a forerunner of LISREL, similarly TESPAT of McDonald (1974) was a forerunner to COSAN.

Among others, standard programmes are COSAN (covariance structure analysis, McDonald, 1978; Fraser, 1982), BENWEE (version 2) (computer programme for path analysis with latent variables, Browne and Cudeck, 1983), and LISCOMP (analysis of linear structural equations using a comprehensive measurement model, Muthén, 1987). Recently, Steiger (1989) has released a computer programme called EzPATH. This programme distributed internationally by SYSTAT Corporation, includes several new techniques for entering and modifying causal models and also implements several new indices for goodness-of-fit.

The most widely used programme for analysis of covariance structures, LISREL (Jöreskog and Sörbom, 1988), however, can not be set up to avoid negative values of variance components. Such values imply an improper and therefore inadmissible solution to the model parameters. The COSAN programme is based on a comprehensive model of covariance structure proposed by McDonald (1978) and it avoids any possibility of negative variance estimates. In the case of LISREL Programme, the user must be conversant with the equation system and its parameter matrices, so as to be able to specify the particular covariance structure model. In the case of EQS programme of Bentler (1989), even this step can be avoided. The user simply has to write out a series of measurement and structural equations, variance and covariance

specifications and the programme generates the relevant matrices internally. Browne's BENWEE programme (Browne and du Toit, 1987) utilizes the Bentler-Weeks covariance structure model (Bentler and Weeks, 1979) of the form (1.13).

1.4 ACOVS AS A GENERAL STATISTICAL METHOD

The ACOVS method has expanded so voluminously that it encompasses now not only the simple extension of factor analysis to arbitrary covariance structure models [Jöreskog, 1973b; Browne, 1974; Mukherjee, 1970; Bentler, 1976; McDonald, 1978] but also such diverse techniques as causal or path analysis [Blalock, 1971; Bagozzi, 1980; James, Mulaik and Brett, 1982; Singh, 1975; Tukey, 1964; Wright, 1934, 1960a, 1960b], simulteneous equation models [Geraci, 1977; Hausman, 1977 and Hsiao, 1976, Robinson, 1977], structural equation models [Goldberger and Duncan, 1973; Duncan, 1975; Jöreskog, 1977; Bentler and Weeks, 1980], errors in variable models, including multiple regression [Bhargava, 1977; Robinson, 1974] and canonical correlation [Bagozzi, Fornell and Larcker, 1981]. Most of the above-mentioned techniques have been shown to be special cases of ACOVS.

There are quite a few advantages gained by treating the above techniques as special cases of ACOVS. First, ML estimates of the identifiable model parameters called structural parameters, can be readily obtained. Structural parameters often represent rela-

tively invariant parameters of a causal process and are, thus, theoretically more meaningful than ordinary predictive regression weights.

Secondly, the use of ACOVS enables the researcher not only to study the effect of measurement errors on the statistical analysis of observed variables but also to assess the relationship between the hypothesized latent variables which is error-free and the indicators used for defining the latent variable. By using multiple flawed indicators to triangulate on the true scores through employment of structural coefficients relating the latent variable which can be estimated without distortion (bias), even though the true scores can never be observed, we can study the true relationship between a set of observed indicator and a latent variable. This relationship is called the validity of the indicator. Therefore, ACOVS plays an important role in the study of validity of different types of indicators and flawed measurements.

Thirdly, the problem resulting from an absence of statistical test for factor loadings, canonical weights, path coefficients are easily resolved in ACOVS. Under the usual multinormal assumption, the standard errors for the ML estimates of the parameters can be readily obtained from the inverse of the Fisher information matrix. The statistical significance of above-mentioned coefficients and loadings can be evaluated by calculating the large sample critical ratio (dividing the

parameter estimate by its standard error). Bagozzi *et al.* (1981) have shown how the problem of determining statistical significance for canonical weights and loadings in canonical correlation is overcome when the analysis is accomplished via structural relations modelling based on ACOVS. They have also shown how " the structural relations model is a more flexible tool for data analysis compared to canonical correlation. It is possible that a canonical model may fail to reject the null hypothesis regarding the relationship between two variates, although this relationship is significant in a less restrictive structural relations models" (Bagozzi, *et al.* 1981, p. 453).

The aim of ACOVS is to help the researcher in correctly specifying and translating the theory of the substantive discipline (e.g. Psychology, Sociology, Economics etc.) into empirical testing and evaluation. It attempts to attain this goal by forcing the researcher to explicitly state the assumptions about the constructs of interest, their assessment and their interrelations.

As stated already, ACOVS arising from different structural equation models provides a mechanism for investigating the hypothesized relationships between and among latent constructs as well as manifest variables. While the manifest variables can be observed, the latent variables in psychology are a set of postulated internal states such as attitudes, values etc. which can be only inferred.

The ACOVS also provides a means for assessing the adequacy

with which the latent constructs have been measured. The technique permits the construction of causal models that incorporate and correct for the effects of measurement errors and are, therefore fundamental in adequately estimating the parameters of causal models and assessing the goodness of the fit of the model to the data.

Thus, the ACOVS may be roughly viewed as a general procedure which combines the logic of 'path or causal analysis with that of factor analysis' (Bentler and Speckart, 1981). As such, the procedure has found wide applicability for studies aiming at indirect causal inference from non-experimental data in various fields of social and biological sciences. [Aigner and Goldberger, 1977; Bielby and Hauser, 1977; Goldberger and Duncan, 1973; Intriligator, 1978; Li, 1975]. However, the procedure is potentially more useful within an experimental context as has been demonstrated by Bagozzi (1980) as well as Fiske, Kenny and Taylor (1982). In particular, latent-variable causal modelling based on ACOVS seems to be ideally suited to the experimental testing of mediational models, as in process analysis (Kenny, 1979).

Justifications for the above optimism come from four observations. Firstly, the ACOVS procedure affords an explicit test of the consistency of any postulated links between mediating and dependent variables with the data. Secondly, because causal paths are estimated among latent rather than manifest variables,

the path estimates are free from the unreliabilities in the manifest variables by which one measures these constructs. Thirdly, latent-variable causal modelling permits the testing of all the links in a mediational model simultaneously, rather than in the typical piecemeal fashion as in the univariate analysis of variance. Lastly, structural modelling with latent variables affords the possibility of clarifying controversial or ambiguous aspects of a model, as well as elaborating and refining the model, by contrasting it with competing alternatives (rival hypotheses).

The value of general structural representation by SEM for practical applications can not be overemphasized. Various specific models such as the MIMIC model of Jöreskog and Goldberger (1975) and other models provided in Aigner and Goldberger (1977) have been surpassed by the development of the general structural models with latent variables. Most of these specialized models can now be formulated into the general random vector and parametric framework of Structural Equation models and then by using any ACOVS programmes, the estimation and testing problems can be handled.

Because of these reasons, structural equation modelling (SEM) or causal modelling techniques have become increasingly important due to their usefulness as a general tool for analyzing a family of linear statistical models that include analysis of variance (ANOVA), multiple regression, and other common multivariate statistics. This is an extension of a general linear

model framework which may involve both observed and latent variables so that ANOVA models may be treated as its special case.

The most popular technique of path analysis (Wright, 1934; Hauser and Goldberger, 1971) resembles the SEM as it requires the estimation of the parameters of the linear structural equations representing the cause and effect relationships. This point has been discussed earlier.

Anderson and Gerbing (1988) reviewed and provided some guidelines on the use of SEM in practice. The topics of LISREL or SEM are sometimes discussed interchangeably including latent variable modelling, confirmatory factor analytic simulteneous modelling, covariance (or moment) structure modelling etc. Almost up-to-date lists of references of work in this area are available in Tanaka, Panter, Winborne and Huba (1989), MacCallum (1986) as well as in Coovert, Penner, and MacCallum (1990).

1.5 A BRIEF REVIEW OF SELECTED APPLICATIONS OF ACOVS

The importance of analysis of covariance structures (ACOVS) was first realized in 1960 when Bock (1960) used it as a multivariate extension of the component of variance analysis, especially for assessing the effects on the test scores of the ways of classification in the design of psychological tests. Although the idea of developing psychological

test items according to some designs originated with Burt (1940), Wilks (1946) and Votaw (1948), psychologists such as Guttman (1954), Foa (1965) as well as Guilford (1956) have fully made use of this rational strategy in test construction. Conventionally, the analysis of variance technique has been used for estimating the variance components. Following Burt's (1947) logic, Bock (1960) for the first time showed that the scores of N subjects obtained on a set of psychological tests classified in a 2^n factorial design may be treated as data from an N x 2^n experimental design. As such, the conventional mixed-model analysis of variance for this type of design provides useful information about the psychometric properties of the tests. The presence of nonzero variance components for the random errors inherent in test data along with the components associated with the fixed and interaction effects may provide a rational way of determining the number of dimensions in which the tests in question are able to discriminate among testees. The relative magnitude of these components measures the discriminative power of the test along the respective dimension.

Bock, Dickens and Van Pelt (1969) also employed the ACOVS approach in estimating the various components of variance for assessing the content-acquiescence correlation in the Minnesota Multiphasic Personality Inventory (MMPI). Bramble and Wiley (1974) used ACOVS for investigating content variance, variance due to non-content characteristics of items, and the

covariances of content and different item characteristics using a set of covariance structure models which is more general than those of Bock (1960) and their extensions (Bock and Bargmann, 1966). These applications of ACOVS as well as those reviewed subsequently in this section show promise of ACOVS " as a powerful technique for investigating the structure of psychological test. If the test items are constructed systematically to represent the construct(s) being measured and various item characteristics of interest, the variances and covariances due to these sources may be estimated and assessed. For example, a test may be reformulated or constructed by each source, as well as the covariation among sources could be assessed using the covariance structure procedures" (Bramble and Wiley, 1974, p. 189). In the field of psychometrics, ACOVS has been found useful (a) as a measurement tool for developing reliable scales, indicators and test battery, (b) as a procedure for examiningmany forms of validity of the tools especially the ones devised from the point of view of convergent and discriminant validity, (c) as a methodology for testing structural hypotheses and evaluation of different types of covariance structure models and (d) as a statistical technique for making causal analysis with recursive or non-recursive equations.

Structural equation modelling (SEM) based on the general procedure of ACOVS has been used successfully in tackling models of macroeconomic policy formulation, inter-

generation occupational mobility, racial discrimination in employment, housing and earning, drug use, scholastic achievement, social development programmes (Sörbom and Jöreskog, 1982), individual differences in various aspects of personality, child development etc. Almost every issue of Journal of Personality and Social Psychology carries one or two applications of SEM. Judd, Jessor and Donovan (1986) have discussed the importance of SEM in personality research. Mukherjee (1969) examined the usefulness of ACOVS in psychological research.

In the subsections to follow, we present a brief review of the applications of ACOVS in psychometry and environmental sciences. The review is not meant to be exhaustive since the main purpose is to acquaint the readers with some of the major utilizations of ACOVS in applied sciences.

1.5.1 Use of ACOVS for Estimating Reliability of Psychological Tests

Analysis of covariance structures (ACOVS) has been extensively used for the purpose of estimating the reliabilities of psychological tests which are inherently subject to measurement errors. As a matter of fact, in any ACOVS, the estimation of reliability goes hand in hand with the estimation of error variance and other components. Once a structural model is hypothesized for the psychological test composed of p items each of which can be assumed to have an error component, the population covariance matrix can be written in the form :

$\underset{\sim}{\Sigma} = \underset{\sim}{B}(\underset{\sim}{\Lambda}\,\underset{\sim}{\Phi}\,\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi}^2)\underset{\sim}{B}' + \underset{\sim}{\epsilon}^2$ where the matrices $\underset{\sim}{B}(p \times q), \underset{\sim}{\Lambda}(q \times r)$

are often known or can be specified in advance and the symmetric matrix $\underset{\sim}{\Phi}$ (r x r) and the diagonal matrices $\underset{\sim}{\Psi}^2$(q x q) and $\underset{\sim}{\epsilon}^2$(p x p) are parameter matrices. Under the assumption of multinormality, the ML estimates of the structural parameters can be obtained including the error variance component of each individual test item. It is possible to constrain some parameters to be equal. Other parameters can be allowed to be free. Hence, they have to be estimated by the data.

For example, assuming that every pair of error-free (true) scores τ_j and τ_k have unit correlation for all j, k = 1,2, ... , p and the error scores (e_j) are mutually independent and orthogonal to the true scores, the population covariance matrix for the p number of test or item scores (observed) can be written as the Spearman unifactor structure already shown in (1.4) where the items are supposed to be factorially homogeneous in one dimension. This is also the general model of a set of congeneric test items or tests (Jöreskog, 1971) [vide subsection 1.2.1].

In the classical test theory of psychometrics (cf. Gulliksen, 1950), the reliability of any j-th test is measured by the ratio of true (error-free) score variance to observed score variance, e.g.

$$\rho_{jj} = \frac{\beta_j^2}{\beta_j^2 + \Psi_j^2} \qquad (1.19)$$

for all j = 1,2, ... p. Under this set-up the reliability of a composite test consisting of p tests is given by

$$\rho^* = \frac{(\sum_j \beta_j)^2}{(\sum_j \beta_j)^2 + \sum_j \Psi_j^2} \qquad (1.20)$$

If in stead of the unidimensional and factorial homogeneity assumptions, items are " essentially tau equivalent" (Lord and Novick, 1968, p. 135), then with unequal error variances, $\underset{\sim}{\Sigma}$ of (1.4) reduces to

$$\underset{\sim}{\Sigma} = \beta^2 \, \underline{1} \, \underline{1}' + \underset{\sim}{\Psi}^2 \qquad (1.21)$$

which simplifies the reliability (1.20) to

$$\rho_1^* = \frac{p \, \bar{\rho}}{1 + (p-1) \, \bar{\rho}} \qquad (1.22)$$

where $\bar{\rho}$ is defined as the index of average correlation or the reliability of each item and hence equal to $\beta^2/(\beta^2 + \sum_j \Psi_j^2/p)$. We note that formula (1.22) is obtained from the covariance matrix $\underset{\sim}{\Sigma}$ (1.21) which has, in fact, an intraclass structure. In the case of equal error variances, formula (1.22) further reduces to

$$\rho_2^* = \frac{p \, \rho}{1 + (p-1)\rho} \qquad (1.23)$$

where $\rho = \beta^2/(\beta^2 + \Psi^2)$.

The problem of estimation of reliability of the composite test using the equations (1.20), (1.22) and (1.23) can be easily handled, once the ML estimates of β's and Ψ's are obtained and the postulated structural model is found to be consistent with the data. Werts, Rock, Linn and Jöreskog

(1978) used the ACOVS procedure for estimating the reliability of a factorially complex composite. Earlier, Jöreskog (1971) provided a formula for computing the reliability of a composite when the components are factorially univocal. Werts, Linn and Jöreskog (1974) also obtained formulae for estimating intra-class correlation reliability based on the assumption that the various measures are parallel in the sense of Lord and Novick (1968, p. 48). By considering different statistical models for test-retest situations, Jöreskog and Sörbom (1977) also illustrated the use of ACOVS for estimating the test-retest reliability.

The assumption of independent errors of measurement in the above formulation may be questionable under certain conditions. Using Jöreskog's LISREL procedure, a model could be used which would allow for the appropriate pairs of errors to be correlated, as illustrated by Werts and Linn (1972).

When there is some instability of the true scores in the interval between the first and second administration of the same test and/or when the measurement errors are themselves correlated, simple test-retest correlations can lead to biased reliability estimates. Using ACOVS, Werts, Breland, Grandy and Rock (1980) demonstrated on the basis of real data collected over three occasions on the Test of Standard Written English (TSWE) and essay ratings (ER) how to separate true score

instability and correlated errors.

Defining τ_1, τ_2 and τ_3 as the true scores underlying TSWE at each of the three testing periods, we can assume for the purpose of illustration that these true scores are perfectly correlated and their error scores have identical variances, are independent and mutually orthogonal to the test scores, as in the classical test theory. With these assumptions, we can write the six scores as below :

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \hline y_1 \\ y_2 \\ y_3 \end{bmatrix} = \left[\begin{array}{ccc} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \\ \hline b & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & b \end{array}\right] \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} + \left[\begin{array}{c} e_1 \\ e_2 \\ e_3 \\ \hline e_4 \\ e_5 \\ e_6 \end{array}\right]$$

or, $\underline{X} = \underset{\sim}{\Lambda}\, \underline{\tau} + \underline{e}$, say (1.24)

Assuming further that the error scores associated with ER are identically distributed with a common variance (γ_2^2) as well as common covariance (ζ) among each other and independent of the true scores, the covariance matrix arising from such structural model(1.24)can be written as

$$\underset{\sim}{\Sigma} = \underset{\sim}{\Lambda}\underline{1}\,\underline{1}'\underset{\sim}{\Lambda}' + \underset{\sim}{\Psi}^2$$

$$= \left[\begin{array}{ccc|ccc} a^2+\gamma_1^2 & & & & & \\ a^2 & a^2+\gamma_1^2 & & & & \\ a^2 & a^2 & a^2+\gamma_1^2 & \text{Symm.} & & \\ \hline ab & ab & ab & b^2+\gamma_2^2 & & \\ ab & ab & ab & b^2+\zeta & b^2+\gamma_2^2 & \\ ab & ab & ab & b^2+\zeta & b^2+\zeta & b^2+\gamma_2^2 \end{array}\right] \qquad (1.25)$$

The structure in the covariance matrix (1.25) is known as Votaw's (1948) compound symmetry structure. We can obtain explicit ML estimates of the five parameters in this case and proceed first with the goodness-of-fit of the model to empirical data. If the fit is satisfactory, then the expected covariances among the three occasions for TSWE will be given by the ML estimate of a^2. Similarly, the expected reliability for ER will require the ML estimate of b^2 plus that of ζ. The expected covariance between TSWE and ER will be given by the product of the ML estimates of a and b .

When the data are longitudinal in nature with respect to repeated observations on the same sample for the same measurement and the correlations between the first and the subsequent measurements show a " quasi-Markov simplex" covariance structure, then following Werts, Linn and Jöreskog (1978), the reliability (ρ_{jj}) of an observed X_j is estimated by

$$\hat{\rho}_{jj} = \frac{r_{jn}\ r_{jm}}{r_{nm}} \qquad (1.26)$$

where $n < j < m$ and the notation r_{vt} denotes the observed correlation between X_v and X_t . In the above structural model, although observed scores (X_j) are assumed to obey the classical test theory, the true scores (τ_j) are related among themselves by linear regression equation given by

$$\tau_{j+1} = \beta_j \tau_j + d_j \qquad (1.27)$$

where the d_j residuals are assumed to be independent of each other. If there is more than one observed score prior to or following X_j, then there will be more than one possible estimate of ρ_{jj} following eqn. (1.26). The possible estimates of ρ_{jj} will be equal within the limits of sampling error provided the simplex model fits the data. The difficulty with this model is that reliabilities can not be estimated either for the first or the last observed variable.

For a set of variables obeying the quasi-Weiner simplex covariance structure (Mukherjee, 1966) of the form already shown in (1.3), the reliabilities are given by

$$\rho_{jj} = \frac{\alpha_j^2}{\alpha_j^2 + \Psi_j^2} \qquad (1.28)$$

for all $j = 1,2, \cdots , (p-1)$.

1.5.2 Selected Applications of ACOVS in Life and Environmental Sciences

In biology and biometric research, path analysis technique (Wright, 1934 ; Li, 1975) has been used to estimate the direct causal influence of one variable to another based on non-experimental data. The problem, in general, is that of estimating the structural parameters of a set of simultaneous equations models representing the cause and effect relationships. Estimating structural parameters requires the application of covariance structure models and methodology. As such, ACOVS has been quite useful in the field of biological research.

For example, ACOVS has been applied by Eaves, Martin and Eysenck (1977) to the genotype environmental analysis of covariation of four aspects of impulsiveness in male and female monozygotic and dizygotic twins. These authors demonstrated how the ACOVS can be exploited to combine in the same analysis a model for the psychological relationships with a causal genotype-environmental model. The use of this technique helps in shifting the emphasis away from a nature versus nurture approach toward consideration of the structure of both genetic and environmental variation.

For genetical analysis of covariance structures, Martin and Eaves (1977) presented a general structural equation model

for k populations. They used likelihood ratio chi-square statistic for testing more restrictive structural equation model against less restrictive one.

Many of the multivariate procedures used in life sciences have the theoretical requirement of homogeneity in the covariance matrices. Essentially, the assumption calls for a test of equal variances and equal covariances i.e. the intraclass or the uniform covariance structure (Mukherjee, 1982).

Confirmatory factor analysis is a special case of ACOVS and can be gainfully applied to problems of morphometrics. For example, size variation in thrips, the identification of shells deriving from the life cycle of fossil foraminifers, variation in scale insects and the morphological variants of the common salamander (Blackith and Reyment, 1971) can be meaningfully subjected to the procedure of ACOVS. Jöreskog, Klovan and Reyment (1976) have mentioned the application of factor analysis procedure to various problems of geology. Folmer (1980, 1981, 1983) utilized the structural modelling approach for measuring effects of regional policy. The nature, detection and structural specifications of spatio-temporal (auto - and cross-) correlation are extensively studied in Folmer (1983) as well as in Folmer and Nijkamp (1983).

1.6 THE ROLE OF ACOVS

While analyzing behavioural or biological data, a number of situations calls for testing certain hypotheses associated with the structure of covariance matrix before certain conventional statistical tests are applied. The standard techniques, such as the principal component analysis, canonical correlation analysis, multiple discriminant analysis etc. strongly involve the covariance structure of the population (Dempster, 1966, p. 316). Even in the case of analysing univariate data arising from latin squares, one-factor repeated design etc., search of patterns in the covariance matrix is important (Graybill, 1961; Srivastava and Malk, 1967; Ghosh, 1987).

Book (1963) has shown that one can proceed to use the transformed sum of products for testing hypotheses on the polynomial representation of group differences only after verifying that the structure of covariance matrix is consistent either with the " compound symmetry" hypothesis or the hypothesis of independence among variates. In case the structure of $\underset{\sim}{\Sigma}$ is unknown, the degree of the polynomial representation for repeated measurements can not be investigated directly. Geisser (1963) has shown that when the relevant $\underset{\sim}{\Sigma}$ matrix in the repeated measurement model has " uniform" or " intraclass" structure, usual F-test, as is used under analysis of variance (ANOVA) assumption, remains valid.

As mentioned in Section 1.4, ACOVS may be viewed as a generalization of the ANOVA of random effects in the mixed model for the experimental designs with one random way of classification [Bock, 1975, p. 450; Bock and Bergmann, 1966].

Patterned covariance matrix plays an important role in the stochastic process and time series analysis (Kendall and Stuart, 1966, p. 472-503). The need for their study in response surface fitting and other statistics has been discussed in Greenberg and Sarhan (1959, 1960), Roy and Sarhan (1956) and Roy, Greenberg and Sarhan (1960).

In psychological literature at least, there are some specific instances where the data are unfortunately subjected to a factor analysis even when the intercorrelation matrices are not significantly different from the identity matrix. Such matrices reflect nothing other than the interrelations among random errors (white noise) in the data (Armstrong and Soelberg, 1968). So before actually conducting a factor analysis, a wise investigator needs to ascertain the untenability of the sphericity hypothesis, e.g. $\underset{\sim}{\Sigma} = \sigma^2 \underset{\sim}{I}_p$. Only when this sphericity hypothesis is rejected, we can subject our data to factor analysis.

Generally in factor analysis the pattern of factor loadings conforms as closely as possible with the correlational pattern among the variables. When such pattern is known <u>a priori</u>, the transformation of the factor matrix may

be anticipated (Howe, 1955; Lawley and Maxwell, 1971). As a special case of ACOVS, McDonald (1969) suggested a generalized factor analysis procedure based on residual covariance matrices of prescribed structure relaxing the assumption of orthogonality in unique factors.

Following the analysis of structural effects in experimental design (Mallios, 1970), the study of covariance structures may also contribute to a more powerful test than the conventional tests used in multivariate analysis of variance (MANOVA).

Whenever the population covariance matrix is patterned, it implies a 'reduced parameter situation' which has a great bearing upon the power of any statistical test (Geisser, 1963; Dempster, 1966, p. 315) resulting in more powerful tests in general as compared to the corresponding unrestricted tests (Mukherjee, 1976). Matsueda and Bielby (1986) have discussed the question of power with regard to SEM in general. There are distinct advantages in the computation of the estimates of the structured covariance matrices. Mukherjee (1969b, 1976, 1982) has shown certain computational short-cuts in the study of Guttman quasi-simplex structure and intraclass structure. Various multivariate measures such as multiple correlation, partial correlation, canonical correlation etc. may be sometimes calculated and handled efficiently avoiding difficult computations.

Geisser (1964) has proved that for uniform covariance

structure, the confidence region as obtained by fiducial and Bayesian methods will yield the same confidence region for the mean vector and the same confidence interval for the common correlation coefficient. Thus, it seems likely that various methods of obtaining interval estimation of the parameters of the multinormal distribution would yield the same results when the covariance matrix has a spectrally decomposable form (Mukherjee, 1981, p. 445).

1.7 OBJECTIVES OF THE PRESENT STUDY

Keeping in view the main statistical problems in ACOVS as described in section 1.3, we list below the major objectives of the present study.

(a) Since the ML estimates of the parameters involved in structured covariance matrix requires rigorous computational technique, we propose to develop a new technique of computation along with its mathematical foundation when the covariance structure is linear.

(b) Testing of structured covariance hypothesis is no doubt a long standing problem in multivariate analysis. In ACOVS, most of the authors developed testing procedure based on likelihood ratio criterion. There are other large sample test criteria, e.g. Wald's quadratic form criterion, Rao's efficient score criterion etc. which need to be examined in the context of ACOVS. Simplified forms of corresponding

test statistics are obtained in certain structural subclass which allows estimation technique developed in (a). Thus, an attempt will be made to develop test procedures for ACOVS based on criterion other than the Wilks' lamda criterion.

(c) We also plan to develop new goodness-of-fit indices employing the concept of structural closeness. These indices are based on simulteneous use of both ML and GLS estimates of the parameters of the covariance matrix which is linearly structured. This type of approach in developing a fit index has never been used before.

(d) Encouraged by certain computational ease and simplifications in statistical measures such as multiple, partial correlations etc. and various heuristic tests in the case of structured covariance matrices such as Guttman quasi-simplex (Mukherjee, 1969a), equipredictability (Bargmann, 1957), circulant (Olkin and Press, 1969) structures, we plan to study the algebraic and statistical aspects of positive definite Toeplitz matrix which very often appears as a covariance matrix in wide-sense stationary stochastic process. A covariance matrix having such structure has the elements depending only upon the difference of row and column specifications. Specific test procedure for testing such types of structures will be developed.

(e) In addition to the estimation of Toeplitz covariance structure, a simplified algorithm for the ML estimation of the

Toeplitz correlation structure [vide eqn. (1.8)] is also proposed.

In the present study, the problem involving structured mean vector with or without structured covariance matrix are not considered. Problems of ACOVS across different populations are also avoided. The use of ACOVS for comparing multiple variables across different populations has been discussed by Rock, Wert and Flaugher (1978). Models of covariance structures arising from qualitative data are also omitted from consideration, as are general nonlinear models that can not be specified a priori in terms of limited number of structural parameters and their interrelations. We will consider mainly those ACOVS models which are linear in both the variables and parameters and in which the errors of measurement (disturbances) are additive in nature.

1.8 ORGANIZATION OF THE DISSERTATION

This thesis is presented in two broad parts. The first part deals with the problem of estimation, testing and goodness-of-fit aspects of ACOVS in general. Chapters 2 and 3 aim at addressing to these problems in detail. The second part deals with an indepth study of special covariance structure, namely, the Toeplitz or Laurent structure.

Chapter 4 of the thesis covers the algebraic and statistical aspects related to such covariance structure

while chapter 5 deals only with the ML estimation of the parameters of a Toeplitz correlation structure. In the last chapter (Chapter 6), suggestions regarding future lines of work in ACOVS are briefly touched upon. Below we present the details of the contents of various chapters.

In Chapter 1, the main problems of covariance structure analysis (ACOVS) have been discussed. Some selected models of covariance structures and a brief review of the literature connected with the applications and role of ACOVS are presented. The chapter concludes with the objectives of the thesis and its organization.

In Chapter 2, a new iterative procedure based on "inverse residual matrix" defined as $\underset{\sim}{Q} = \underset{\sim}{\Sigma}^{-1} - \underset{\sim}{\Sigma}^{-1} \underset{\sim}{S} \underset{\sim}{\Sigma}^{-1}$ is proposed to solve the likelihood equations arising in computing the ML estimates of the parameters of linearly structured covariance matrices, $\underset{\sim}{\Sigma}$. This is done without inverting $\underset{\sim}{\Sigma}$ or $\underset{\sim}{S}$, the latter matrix being the sample covariance matrix based on a random sample drawn from $N_p(\underline{X} ; \underline{\mu} , \underset{\sim}{\Sigma})$. Some algebraic aspects and certain points of superiority of this Q-procedure over Anderson's (1973) algorithm have been discussed. The proposed algorithm seems to have the advantage of rapid covergence, of robustness against bad starting value and of not requiring the computation of either $\underset{\sim}{\Sigma}^{-1}$ or the second order derivatives of likelihood function.

A checking formula to verify the computational consistency in the Hessian matrix of likelihood at the ML estimate is also suggested. In lieu of checking the negative definiteness of the Hessian matrix at the likelihood-solution, we may alternatively consider this checking formula to ensure the correctness of the ML estimates of the parameters.

As an extension to Q-procedure, Q^*-procedure is developed in the case of linearly structured correlation matrix ($\underset{\sim}{\rho}$) involved in $\underset{\sim}{\Sigma}$-matrix when decomposed as $\underset{\sim}{D}_\sigma \underset{\sim}{\rho} \underset{\sim}{D}_\sigma$.

Lastly, the generalized least squares estimation procedure with $\underset{\sim}{S}^{-1}$ as the weight matrix is viewed through what is called R-procedure where $\underset{\sim}{R} = \underset{\sim}{S}^{-1} - \underset{\sim}{S}^{-1} \underset{\sim}{\Sigma} \underset{\sim}{S}^{-1}$. Q - , Q^*- and R-procedures are illustrated by numerical examples. A major part of this chapter is based on Mukherjee and Maiti (1988a, 1988b).

In chapter 3, testing of linearly structured covariance matrix is considered from the point of view of estimated " inverse residual matrix" ($\underset{\sim}{\hat{Q}}$) which appears as a by product while we use the Q-procedure for ML estimation. Rao's efficient score criterion and Wald's (1943) quadratic form criterion are utilized to obtain two new large-sample test statistics. Numerical examples and the results of a Monte Carlo study are discussed with regard to the power of the proposed test statistics for testing the sphericity versus uniform structure

assumed for the covariance matrix of 8-variate normal population.

We next discuss the large-sample distributional properties of the most frequently employed Joreskog and Sorbom's (1981) goodness-of-fit index. We then propose two new fit indices based simulteneously on ML and GLS estimates of parameters of linearly structured covariance matrix introducing "structural closeness" concept. Life example and results of a Monte Carlo study under the same set-up as mentioned in the last paragraph is reported and the performances of the proposed indices are discussed. This chapter is based on Maiti and Mukherjee (1989a, 1989c, 1990).

In Chapter 4, the involvement of Toeplitz or "stripe" matrix in various fields of applied sciences is surveyed from the statistical point of view. Algebraic properties of positive definite Toeplitz matrix is discussed with the emphasis on inverse, determinant and eigenvalue problems.

Statistical features or rather the specific features when we apply various statistical measures such as multiple, partial, bi-partial or canonical correlations etc. upon the anticipated Toeplitz covariance matrix, are studied. An ad hoc procedure for the detection of Toeplitz covariance structure is also suggested via principal component analysis.

Formal large-sample two-stage test procedures for

testing the significance of Toeplitz covariance structure are then proposed. Numerical illustrations of the statistical measures and that of two-stage tests are provided with life data. The algebraic aspect of this chapter is based on Mukherjee and Maiti (1988c) while the large-sample tests are based on Maiti and Mukherjee (1989c).

In Chapter 5, an iterative algorithm is developed for obtaining the ML estimate of the parameters of the covariance matrix given by $\underset{\sim}{\Sigma} = \underset{\sim}{D}_{\sigma} \underset{\sim}{\rho} \underset{\sim}{D}_{\sigma}$ with the correlation matrix $\underset{\sim}{\rho}$ having Toeplitz structure and $\underset{\sim}{D}_{\sigma} = \text{diag.}\ (\sigma_1, \sigma_2, \cdots, \sigma_p)$. The p number of scale parameters and p Toeplitz correlation parameters may be estimated in a general way by Q^*-procedure as suggested in Chapter 2. But in this chapter, the specific features of Toeplitz correlation matrix is exploited and a new sum (called " systematic sum") is introduced. This algorithm is applied upon a set of psychometric data and the plausibility of Toeplitz correlation structure is discussed. This chapter is based mainly on Mukherjee and Maiti (1988b).

In the last chapter, various possibilities of extensions, limitation of our investigation and the suggestions for future research work are discussed.

1.9 SUMMARY

After presenting the generic technique of analysis of covariance structures (ACOVS) as a statistical method, certain

basic models of ACOVS and the main problems involved in it are discussed. A brief review of some of the major applications of ACOVS in both biological and social sciences is also presented. An attempt is then made to delineate the basic objectives of the present study. The structure of the thesis is then outlined in terms of its various chapters.

CHAPTER 2

MAXIMUM LIKELIHOOD ESTIMATION OF PARAMETERS OF LINEARLY STRUCTURED COVARIANCE MATRICES

2.1 INTRODUCTION

In the analysis of linear covariance structures, we are interested in the estimation and testing of the unknown population covariance matrix $\underset{\sim}{\Sigma}$ of order pxp, which is structured as

$$\underset{\sim}{\Sigma}(\underline{\theta}) = \sum_{t=1}^{k} \theta_t \underset{\sim}{H}_t \qquad (2.1.1)$$

where $\underset{\sim}{H}_t$'s are fixed and <u>known</u> linearly independent symmetric matrices, called design matrices satisfying the relationship $tr(\underset{\sim}{H}_i \underset{\sim}{H}_j) = 0$ whenever $i \neq j$. The θ_t's are unknown but functionally independent parameters $\in \Theta_k$, $(1 \leq k < \binom{p+1}{2})$ so restricted that $\underset{\sim}{\Sigma}(\underline{\theta})$ is a positive definite (p.d.) matrix, as stated in Chapter 1 .

Given a random sample $\{\underline{X}_\alpha , \alpha = 1(1)N\}$ drawn from a multinormal population $N_p(\underline{X} ; \underline{\mu} , \underset{\sim}{\Sigma})$ having $\underset{\sim}{\Sigma}$ structured as (2.1.1), different problems on estimation, particularly emerging from the use of maximum likelihood (ML) procedure in estimating the unknown parameters θ_t's, have been studied by many authors [e.g. Anderson (1969); Bock and Bergmann (1966); Browne (1977); Burg, Luenberger and Wenger (1982), Jöreskog (1970), McDonald (1974), Mukherjee (1970, 1976), Szatrowski (1980) etc.]. But the solution of the associated likelihood equations still poses a general computational problem, although in some specific situations explicit solutions have been obtained [Rogers and Young

(1977), Szatrowski (1980), Mukherjee (1981)].

Several methods for the iterative solution of likelihood equations have been considered in the literature [e.g. Barnett (1966), Kale (1962), Brown (1966)]. Iterative schemes based on Newton-Raphson method, Fisher's method of scoring [Lee and Jennrich (1979)], Fletcher-Powell (1963) algorithm as used by Jöreskog (1970, 1978) for his LISREL programme and EM algorithm of Dempster, Laird and Rubin (1977) as used by Rubin and Szatrowski (1982) have been developed for ACOVS. Among these iterative schemes, the most commonly used one is that which was suggested by Anderson (1973) and is based on the Newton-Raphson method. This scheme is believed to be quadratically convergent (Brown, 1966).

All the above-mentioned iterative schemes, however, do not avoid the calculation of the inverse of $\underset{\sim}{\Sigma}$ matrix in the course of iteration. Keeping this in view, we employ a procedure first proposed by Mukherjee (1984), called Q-procedure which circumvents the problem of inverting the $\underset{\sim}{\Sigma}$-matrix. It is based on " inverse residual matrix'' defined as

$$\underset{\sim}{Q} = \underset{\sim}{\Sigma}^{-1} - \underset{\sim}{\Sigma}^{-1}\ \underset{\sim}{S}\ \underset{\sim}{\Sigma}^{-1} \qquad (2.1.2)$$

where $\underset{\sim}{S}$ is the sample covariance matrix.

This chapter deals with the algorithm of the Q-procedure when $\underset{\sim}{\Sigma}$ is linearly structured as (2.1.1). In section 2.2, we present the development and some algebraic aspects of the algorithm derived on the basis of Q-procedure. Some well-known structures are considered to illustrate this procedure. Anderson's (1973) algorithm is compared in section 2.3 with the iterative algorithm of

Q-procedure. This exercise is followed by a numerical illustration.

Usually the Hessian matrix of log-likelihood (when $\underset{\sim}{\theta}$ is replaced by its maximum likelihood estimate (MLE)) is negative definite. Such checking requires a lot of computational labour. A new simple checking formula is established in section 2.4.

In section 2.5, we extend the Q-procedure to the covariance matrices decomposable as

$$\underset{\sim}{\Sigma} = \underset{\sim}{D}_{\sigma} \underset{\sim}{\rho} \underset{\sim}{D}_{\sigma} \qquad (2.1.3)$$

where $\underset{\sim}{D}_{\sigma} = \text{diag.}\ (\sigma_1 , \sigma_2 , \cdots , \sigma_p)$ and $\underset{\sim}{\rho}$ (correlation matrix) is linearly structured as

$$\underset{\sim}{\rho} = \underset{\sim}{I}_p + \sum_{t=1}^{k} \rho_t \underset{\sim}{H}_t \ . \qquad (2.1.4)$$

The MLE of the scale parameters σ_i, $i = 1(1)p$ and the correlation parameters ρ_t, $t = 1(1)k$ are iteratively solved by what is called Q^*-procedure where

$$\underset{\sim}{Q}^* = \underset{\sim}{D}_{\sigma} \underset{\sim}{Q} \underset{\sim}{D}_{\sigma} \ . \qquad (2.1.5)$$

In section 2.6, we consider the generalized least squares estimate (GLSE) of linearly structured $\underset{\sim}{\Sigma}$. Analogous to Q-procedure, we propose here R-procedure defining

$$\underset{\sim}{R} = \underset{\sim}{S}^{-1} \underset{\sim}{S}^{-1} \underset{\sim}{\Sigma} \underset{\sim}{S}^{-1} \qquad (2.1.6)$$

to obtain GLSE with $\underset{\sim}{S}^{-1}$ as weight matrix. Using the Kodak data (Jackson and Morris, 1957), we illustrate the R-procedure. Finally, we make some concluding remarks in section 2.7.

2.2 THE Q-PROCEDURE

Assuming the population mean vector $\underline{\mu}$ perfectly arbitrary, its MLE is the sample mean vector $\underline{\bar{X}}$. The log-likelihood after maximization with respect to $\underline{\mu}$ is given by

$$\log f = -\frac{N}{2}[p \log(2\pi) + \log \det \underset{\sim}{\Sigma} + \mathrm{tr}(\underset{\sim}{\Sigma}^{-1}\underset{\sim}{S})]. \qquad (2.2.1)$$

To obtain the MLE of θ_t , t = 1(1)k when $\underset{\sim}{\Sigma}$ is expressed as (2.1.1), log f is to be maximized. This is equivalent to minimization of

$$h = \log \det \underset{\sim}{\Sigma} + \mathrm{tr}(\underset{\sim}{\Sigma}^{-1}\underset{\sim}{S}). \qquad (2.2.2)$$

Differentiating (2.2.2) with respect to θ_t and equating this to zero, we get what are called likelihood equations (also sometimes called normal equations with reference to h) of the following form :

$$\mathrm{tr}[(\underset{\sim}{\Sigma}^{-1} - \underset{\sim}{\Sigma}^{-1}\underset{\sim}{S}\underset{\sim}{\Sigma}^{-1})(\partial\underset{\sim}{\Sigma}/\partial\theta_t)] = 0$$

or

$$\mathrm{tr}(\underset{\sim}{H}_t \underset{\sim}{Q}) = 0, \quad t = 1(1)k \qquad (2.2.3)$$

where $\underset{\sim}{Q}$ is defined as in (2.1.2). We shall now prove a lemma for the purpose of developing Q-procedure.

Lemma 2.2.1 : Under normal equations (2.2.3), the matrix $\underset{\sim}{Q}$ may be expressed as

$$\underset{\sim}{Q} = \sum_{\alpha=1}^{\upsilon} \underset{\sim}{M}_\alpha \, q_\alpha \qquad (2.2.4)$$

where q_α's are some υ distinct elements and $\underset{\sim}{M}_\alpha$'s are suitable symmetric matrices and $\upsilon = \binom{p+1}{2} - k$.

Furthermore, $\operatorname{tr}(\underset{\sim}{M}_\alpha \underset{\sim}{H}_t) = 0$ (2.2.5)

$\forall\ \alpha = 1(1)\upsilon$ and $\forall\ t = 1(1)k$ so that

$$\operatorname{tr}(\underset{\sim}{M}_\alpha \underset{\sim}{\Sigma}) = 0 \qquad (2.2.6)$$

$\forall\ \alpha = 1(1)\upsilon$.

Proof : Under normal equations (2.2.3), $\binom{p+1}{2}$ unknown elements of symmetric $\underset{\sim}{Q}$ are restricted by a system of k linearly independent homogeneous equations formed by them. Clearly, some υ ($= \binom{p+1}{2} - k$)elements of $\underset{\sim}{Q}$, say q_α , $\alpha = 1(1)\upsilon$ would remain arbitrary while the other k elements, q_α , $\alpha = (\upsilon+1)\ (1)\binom{p+1}{2}$ would be their linear combinations.

In effect, $\underset{\sim}{Q}$ may be expressed as (2.2.4). Note that unlike $\underset{\sim}{H}_t$'s, $\underset{\sim}{M}_\alpha$'s are not necessarily linearly independent matrices.

Using (2.2.4), we now express the normal equations (2.2.3) as

$$\sum_{\alpha=1}^{\upsilon} \operatorname{tr}(\underset{\sim}{H}_t \underset{\sim}{M}_\alpha)\ q_\alpha = 0 \qquad \forall\ t = 1(1)k \ . \qquad (2.2.7)$$

Due to arbitrariness of q_α's , (2.2.7) holds only when

$$\operatorname{tr}(\underset{\sim}{H}_t \underset{\sim}{M}_\alpha) = 0 \qquad \forall t \text{ and } \forall \alpha = 1(1)\upsilon \ . \qquad (2.2.8)$$

As $\underset{\sim}{\Sigma} = \sum_{t=1}^{k} \theta_t \underset{\sim}{H}_t$, we multiply (2.2.8) by θ_t and sum over $t = 1(1)k$ to get

$$\operatorname{tr}(\underset{\sim}{M}_\alpha \underset{\sim}{\Sigma}) = 0 \qquad \forall \alpha = 1(1)\upsilon \ . \qquad \blacksquare$$

We now put forward a matrix identity from (2.1.2) as

$$\underset{\sim}{\Sigma} = \underset{\sim}{S} + \underset{\sim}{\Sigma}\ \underset{\sim}{Q}\ \underset{\sim}{\Sigma}\ . \qquad (2.2.9)$$

This would help to solve υ number of elements (called "auxiliary parameters") of $\underset{\sim}{Q}$-matrix, as assured by lemma 2.2.1,and k number of θ_t's simulteneously through an iterative process.

Let $\underset{\sim}{\Sigma}_{(i)} = \underset{\sim}{\Sigma}(\underline{\theta}_{(i)})$ be some i-th trial estimate of $\underset{\sim}{\Sigma}$. Then recalling (2.2.9), we set

$$\underset{\sim}{\Sigma}_{(i+1)} = \underset{\sim}{S} + \underset{\sim}{\Sigma}_{(i)}\ \underset{\sim}{Q}_{(i+1)}\ \underset{\sim}{\Sigma}_{(i)} \qquad (2.2.10)$$

to obtain (i+1)th trial estimate of $\underset{\sim}{\Sigma}$, $i = 0(1)\ \infty$. $\underline{\theta}_{(o)}$ is an initial solution vector of $\underline{\theta} \in \textcircled{H}_k$ and $\underset{\sim}{Q}_{(i+1)}$ is the estimate of $\underset{\sim}{Q}$ based on $\underset{\sim}{\Sigma}_{(i)}$.

It is observed that in most practical problems, element-wise collection of $\upsilon+k$ [$= \binom{p+1}{2}$] equations from (2.2.9) would show the entrance of q_α's linearly and θ_t's at most quadratically. Various choices of υ and k equations from $\binom{p+1}{2}$ equations to solve for q_α's and θ_t's respectively may lead either to divergent iterations or to non-unique convergent iterations even for the same initial solution vector $\underline{\theta}_{(o)}$.

Remark 2.1 The afore-said problem has been studied by Mukherjee and Maiti (1986). The difficulties are circumvented by considering " trace operation" discussed below.

2.2.1 Trace Operation

Premultiplying both sides of (2.2.9) by $\underset{\sim}{H}_t$ and $\underset{\sim}{M}_\alpha$ and taking the trace would lead respectively to

$$tr(\underset{\sim}{H}_t\underset{\sim}{\Sigma}) = tr(\underset{\sim}{H}_t\ \underset{\sim}{S}) + tr(\underset{\sim}{H}_t\ \underset{\sim}{\Sigma}\ \underset{\sim}{Q}\ \underset{\sim}{\Sigma}) \qquad (2.2.11)$$

and

$$tr(\underset{\sim}{M}_\alpha\underset{\sim}{\Sigma}) = tr(\underset{\sim}{M}_\alpha\ \underset{\sim}{S}) + tr(\underset{\sim}{M}_\alpha\ \underset{\sim}{\Sigma}\ \underset{\sim}{Q}\ \underset{\sim}{\Sigma}), \qquad (2.2.12)$$

$t = 1(1)\ k,\ \alpha = 1(1)\upsilon$.

Recalling the linear structure (2.1.1) of $\underset{\sim}{\Sigma}$, we note that

$$tr(\underset{\sim}{H}_t\underset{\sim}{\Sigma}) = tr(\underset{\sim}{H}_t^2)\theta_t = tr(\underset{\sim}{H}_t'\ \underset{\sim}{H}_t)\theta_t\ . \qquad (2.2.13)$$

Defining Euclidean or Frobenius norm of $\underset{\sim}{H}_t$ as $\|\underset{\sim}{H}_t\|_F = [\,tr(\underset{\sim}{H}_t'\ \underset{\sim}{H}_t)]^{1/2}$ and writing $\bar{\underset{\sim}{H}}_t = \underset{\sim}{H}_t / \|\underset{\sim}{H}_t\|_F$, we have from (2.2.13)

$$tr(\bar{\underset{\sim}{H}}_t\underset{\sim}{\Sigma}) = \theta_t\ ,\ t = 1(1)k\ . \qquad (2.2.14)$$

It may be noted here that in the event of $\underset{\sim}{H}_t$ being a sparse matrix, $\|\underset{\sim}{H}_t\|_F^2$ will be merely the frequency of θ_t occurring in $\underset{\sim}{\Sigma}$.

Applying (2.2.6) of lemma 2.2.1 and (2.2.14) upon (2.2.12) and (2.2.11) respectively, we get

$$\theta_t = tr(\bar{\underset{\sim}{H}}_t\ \underset{\sim}{S}) + \sum_{\alpha=1}^{\upsilon} tr(\bar{\underset{\sim}{H}}_t\ \underset{\sim}{\Sigma}\ \underset{\sim}{M}_\alpha\ \underset{\sim}{\Sigma})q_\alpha\ , \qquad (2.2.15)$$

and

$$\sum_{\beta=1}^{\upsilon} tr(\underset{\sim}{M}_\alpha\ \underset{\sim}{\Sigma}\ \underset{\sim}{M}_\beta\ \underset{\sim}{\Sigma})q_\beta = -tr(\underset{\sim}{M}_\alpha\ \underset{\sim}{S}), \qquad (2.2.16)$$

$t = 1(1)k,\ \alpha = 1(1)\upsilon$.

By combining them in matrix form, we obtain

$$\underline{\theta} = \underline{\hat{\theta}}^{\ell} + \underset{\sim}{B}\,\underline{q} = \underline{\hat{\theta}}^{\ell} + \underset{\sim}{B}\underset{\sim}{A}^{-1}\,\underline{b} \qquad (2.2.17)$$

provided $\underset{\sim}{A}^{-1}$ exists where

$$\underset{\sim}{B}(k \times \upsilon) = ((b_{t\alpha})),\ b_{t\alpha} = tr(\underset{\sim}{\bar{H}}_t\ \underset{\sim}{\Sigma}\ \underset{\sim}{M}_\alpha\ \underset{\sim}{\Sigma})$$

$$\underline{g}(\upsilon \times 1) = ((g_\alpha))$$

$$\underset{\sim}{A}(\upsilon \times \upsilon) = ((a_{\alpha\beta})),\ a_{\alpha\beta} = tr(\underset{\sim}{M}_\alpha\ \underset{\sim}{\Sigma}\ \underset{\sim}{M}_\beta\ \underset{\sim}{\Sigma})$$

$$\underline{b}(\upsilon \times 1) = ((b_\alpha)),\quad b_\alpha = -\ tr(\underset{\sim}{M}_\alpha \underset{\sim}{S})$$

$$\underline{\hat{\theta}}^{\ell}(k \times 1) = ((\hat{\theta}^{\ell}_t)) = \text{least squares estimate (LSE) of } \underline{\theta}$$

$$\hat{\theta}^{\ell}_t = tr(\underset{\sim}{\bar{H}}_t\ \underset{\sim}{S}).$$

For iterative purposes, the identity (2.2.10) may lead to the following equivalent process:

$$\underline{\theta}_{(i+1)} = \underline{\hat{\theta}}^{\ell} + \underset{\sim}{B}_{(i)}\ \underset{\sim}{A}^{-1}_{(i)}\ \underline{b}\ ,\ i = 0(1)\ \infty\ . \qquad (2.2.18)$$

The matrices $\underset{\sim}{B}_{(i)}$ and $\underset{\sim}{A}_{(i)}$ refer to the i-th initial estimate of $\underset{\sim}{B}$ and $\underset{\sim}{A}$ based on $\underset{\sim}{\Sigma}_{(i)}$. The column vectors $\underline{\hat{\theta}}^{\ell}$ and $\underline{b}$ remain unchanged during any iterations since they are functions of the elements of $\underset{\sim}{S}$.

It may be noted that the estimate of the "auxiliary parameter vector" $\underline{g}(\upsilon \times 1)$ at each iterative step is given by $\underline{g}_{(i+1)} = \underset{\sim}{A}^{-1}_{(i)}\ \underline{b}$ which is, in fact, based upon the i-th iterate of $\underline{\theta}$. The choice of the initial trial estimate $\underline{\theta}_{(o)}$ is, of course, a problem. However, one may choose $\underline{\hat{\theta}}^{\ell}$ as the initial try-out value of $\underline{\theta}$ because of its closeness to ML solution in large samples (McDonald, 1974, p. 190). By using $\underline{\hat{\theta}}^{\ell}$ or any consistent estimate of $\underline{\theta}$, the sequence of vectors $\underline{\theta}_{(1)}$, $\underline{\theta}_{(2)}$, ... is likely to converge to $\underline{\theta}$ (Kale, 1962) in the sense that for any vector norm $\|\cdot\|$

$$\underset{n \longrightarrow \infty}{\text{limit}} \left\| \underline{\theta}_{(n)} - \underline{\theta} \right\| = 0\ .$$

2.2.2 Two Important Implications for the Iterative Process

Equation (2.2.17) has the following two implications for the iterative process:

(1) When $\underset{\sim}{\Sigma}$ of (2.1.1) would have all elements distinct (except symmetry), we shall call $\underset{\sim}{\Sigma}$ unstructured. In that case, $k = \binom{p+1}{2}$ and under normal equations $\underset{\sim}{Q} = \underset{\sim}{O}$ so that from the identity (2.2.9), the MLE of $\underset{\sim}{\Sigma}$ is $\underset{\sim}{S}$. Thus, when no constraint is imposed on the structure of $\underset{\sim}{\Sigma}$, the MLE of $\underset{\sim}{\Sigma}$ is obtained explicitly as $\underset{\sim}{S}$.

(2) The chance of getting the null solution of the system (2.2.17) is very very small since the coefficient matrix $\underset{\sim}{A}$ is positive definite and $\underline{b}$ to be a null vector depends upon whether or not the sample covariance matrix $\underset{\sim}{S}$ is structured identically as $\underset{\sim}{\Sigma}$. Irrespective of any $\underset{\sim}{S}$, the second term of r.h.s. of (2.2.15) would algebraically vanish if all the coefficients $tr(\underset{\sim}{H}_t \underset{\sim}{\Sigma} \underset{\sim}{M}_\alpha \underset{\sim}{\Sigma})$, $\alpha = 1(1)\upsilon$ are equal to zero. The MLE of certain parameter θ_t of (2.2.17) would thus be identical to its LSE when the t-th column of $\underset{\sim}{B}$ would be null. In general, if some k_o ($\leq k$) rows of $\underset{\sim}{B}$ are null vectors, then by reshuffling and partitioning the columns of (2.2.17), we may get

$$\begin{bmatrix} \underline{\theta}_1 \\ k_o x 1 \\ \underline{\theta}_2 \\ (k-k_o)x1 \end{bmatrix} = \begin{bmatrix} \underline{\hat{\theta}}_1^{\ell} \\ k_o x 1 \\ \underline{\hat{\theta}}_2^{\ell} \\ (k-k_o)x1 \end{bmatrix} + \begin{bmatrix} \underline{0} \\ k_o x 1 \\ B_2 A^{-1} \underline{b} \\ (k-k_o)x1 \end{bmatrix} \qquad (2.2.19)$$

This implies that the MLE of $\underline{\theta}_1 = \underline{\hat{\theta}}_1^{\ell}$ whereas that of $\underline{\theta}_2$ is

obtained by performing iteration upon the reduced iterative process given by

$$\underset{\sim}{\theta}_2 = \underset{\sim}{\hat{\theta}}_2^{(\ell)} + \underset{\approx}{B}_2 \; \underset{\approx}{A}^{-1} \; \underset{\sim}{b} \; . \qquad (2.2.20)$$

Hence, such reduction in the number of iterated parameters is likely to save some computer space and computer time.

2.2.3 Examples of Some Simple Structures

The ML estimation method using the Q-procedure will be illustrated using four different examples of linearly structured covariance matrices. These examples are presented below.

(a) Intraclass Covariance Structure

Let $\underset{\sim}{\Sigma} = ((\sigma_{ij}))_{pxp}$ where $\sigma_{ij} = a$ for $i = j$, and $= b$ for $i \neq j$. Then $\underset{\sim}{\Sigma} = a \; \underset{\approx}{H}_1 + b \; \underset{\approx}{H}_2$ is an alternative expression where $\underset{\approx}{H}_1 = \underset{\sim}{I}_p$, an identity matrix of order p, and $\underset{\approx}{H}_2 = \underset{\sim}{E}_p - \underset{\sim}{I}_p$. $\underset{\sim}{E}_p$ is a p x p matrix of unity. Relevant likelihood equations are given by

$$\text{tr } \underset{\sim}{Q} = 0 \qquad (2.2.21)$$

and

$$\text{tr}(\underset{\sim}{Q} \; \underset{\sim}{E}_p) = 0 \; . \qquad (2.2.22)$$

We next derive the following :

$$\text{tr}(\underset{\sim}{\Sigma} \underset{\sim}{Q} \underset{\sim}{\Sigma}) = (a-b)^2 \; \text{tr } \underset{\sim}{Q} + [\,2b(a-b)+pb^2\,] \; \text{tr}(\underset{\sim}{Q} \underset{\sim}{E}_p)$$

and

$$\text{tr}(\underset{\sim}{\Sigma} \underset{\sim}{Q} \underset{\sim}{\Sigma} \underset{\sim}{E}_p) = (a-b)^2[\,\text{tr}(\underset{\sim}{Q} \underset{\sim}{E}_p) - \text{tr} \underset{\sim}{Q}\,] + (p-1)[\,2b(a-b)+pb^2\,] \, \text{tr}(\underset{\sim}{Q} \underset{\sim}{E}_p)$$

which, on application of (2.2.21) and (2.2.22), reduces to

$$\text{tr}(\underset{\sim}{\Sigma} \underset{\sim}{Q} \underset{\sim}{\Sigma}) = 0 \qquad (2.2.23)$$

and

$$\text{tr}(\underset{\sim}{\Sigma} \underset{\sim}{Q} \underset{\sim}{\Sigma} \underset{\sim}{E}_p) = 0 \; . \qquad (2.2.24)$$

Therefore, from the identity $\underset{\sim}{\Sigma} = \underset{\sim}{S} + \underset{\sim}{\Sigma}\ \underset{\sim}{Q}\ \underset{\sim}{\Sigma}$, we get

$$\text{tr}\ \underset{\sim}{\Sigma} = \text{tr}\ \underset{\sim}{S} \qquad (2.2.25)$$

and

$$\text{tr}(\underset{\sim}{\Sigma}\ \underset{\sim}{E}_p) = \text{tr}(\underset{\sim}{S}\underset{\sim}{E}_p)\ . \qquad (2.2.26)$$

From (2.2.25) and (2.2.26), we obtain the MLE of a and b as

$$\hat{a} = \text{tr}\underset{\sim}{S}/p = \sum_{i=1}^{p} s_{ii}/p$$

and

$$\hat{b} = [\ \text{tr}(\underset{\sim}{S}\underset{\sim}{E}_p) - \text{tr}\underset{\sim}{S}]/p(p-1) = \sum_{i\neq j}\sum s_{ij}/p(p-1).$$

Thus, the MLE of the two unknown parameters of the intraclass covariance matrix is obtained explicitly.

(b) <u>Covariance Matrix having zero elements in the reverse diagonal</u>

Let $\underset{\sim}{\Sigma} = (\sigma_{ij}))_{pxp}$ whose elements are all distinct (apart from symmetry) but its reverse diagonal elements e.g. $\sigma_{i,p-i+1}$, $i = 1(1)p$ are all zeroes. This means that $\underset{\sim}{\Sigma}$ is linearly structured as

$$\underset{\sim}{\Sigma} = \sum_{i=1}^{p}\sum_{\substack{j=i \\ j\neq p-i+1}}^{p} \underset{\sim}{H}_{ij}\ \sigma_{ij} \qquad (2.2.27)$$

where $\underset{\sim}{H}_{ij}$ is a pxp sparse matrix with only (i, j) and (j,i) elements unity.

We observe that under likelihood equations, all the elements of $\underset{\sim}{Q}$-matrix would vanish except its reverse diagonal elements. Hence, $\underset{\sim}{Q}$ may be expressed as

$$\underset{\sim}{Q} = \sum_{\alpha=1}^{\upsilon} \underset{\sim}{M}_{\alpha,p-\alpha+1}\ q_{\alpha,p-\alpha+1} \qquad (2.2.28)$$

where $\underset{\sim}{M}_{\alpha,p-\alpha+1}$ is a pxp sparse matrix with only $(\alpha,p-\alpha+1)$ and $(p-\alpha+1,\alpha)$ elements unity and υ is the number of "auxiliary parameters" equal to $[\frac{p+1}{2}]$. We next derive the following results employing (2.2.27) and (2.2.28).

Keeping in view that the notation $\underset{\sim}{\bar{H}}_{ij} = \underset{\sim}{H}_{ij} / \| \underset{\sim}{H}_{ij} \|_F$,

$$\operatorname{tr}(\underset{\sim}{\bar{H}}_{ij} \underset{\sim}{\Sigma} \underset{\sim}{M}_{\alpha,p-\alpha+1} \underset{\sim}{\Sigma}) = \begin{cases} \sigma_{i\alpha}\ \sigma_{p-\alpha+1,j} + \sigma_{i,p-\alpha+1}\ \sigma_{\alpha j} & \text{for } \alpha \neq p-\alpha+1 \\ \sigma_{i\alpha}\ \sigma_{\alpha j} & \text{for } \alpha = p-\alpha+1 \end{cases}$$

$$\forall i = 1(1)p\ ,\ j = i(1)p,\ j \neq p-i+1\ . \qquad (2.2.29)$$

$$\operatorname{tr}(\underset{\sim}{M}_{\beta,p-\beta+1} \underset{\sim}{\Sigma} \underset{\sim}{M}_{\alpha,p-\alpha+1} \underset{\sim}{\Sigma}) = \begin{cases} 2(\sigma_{\beta\alpha}\sigma_{p-\alpha+1,p-\beta+1} + \sigma_{\beta,p-\alpha+1}\ \sigma_{\alpha,p-\beta+1}), \\ \qquad \text{for } \beta \neq p-\beta+1, \alpha \neq p-\alpha+1 \\ 2(\sigma_{\beta\alpha}\ \sigma_{\alpha,p-\beta+1} \quad \text{for } \beta \neq p-\beta+1, \alpha = p-\alpha+1 \\ 2\ \sigma_{\beta\alpha}\ \sigma_{p-\alpha+1,\beta} \quad \text{for } \beta = p-\beta+1, \alpha \neq p-\alpha+1 \\ 0 \qquad \text{for } \beta = p-\beta+1, \alpha = p-\alpha+1. \end{cases}$$

(2.2.30)

$\forall\ \alpha,\beta = 1(1)\upsilon$, and

$$\operatorname{tr}(\underset{\sim}{M}_{\beta,p-\beta+1} \underset{\sim}{S}) = \begin{cases} 2\ s_{\beta,p-\beta+1} & \text{for } \beta \neq p-\beta+1 \\ s_{\beta,\beta} & \text{for } \beta = p-\beta+1 \end{cases} \qquad (2.2.31)$$

$\forall\ \beta = 1(1)\upsilon$.

The expressions (2.2.29) to (2.2.31) will help in programming the iterative scheme (2.2.18) to solve for the parameters $\underline{\theta} = (\sigma_{ij}\ ,\ i = 1(1)p,\ j = i(1)p,\ j \neq p-i+1)$.

(c) Tri-diagonal (Jacobi) Structure

$$\underset{\sim}{\Sigma} = \begin{bmatrix} a_1 & & & & \\ b_1 & a_2 & & & \\ 0 & b_2 & a_3 & \text{Sym.} & \\ \vdots & & & & \\ 0 & 0 & 0 & b_{p-1} & a_p \end{bmatrix}$$

$$= \sum_{i=1}^{p} \underset{\sim}{H}_i \, a_i + \sum_{j=1}^{p-1} \underset{\sim}{H}_{p+j} \, b_j \qquad (2\cdot2\cdot32)$$

where $\underset{\sim}{H}_i$ is a (p x p) matrix whose (i,i) element is 1, and 0 elsewhere $\forall$ i = 1(1)p, and $\underset{\sim}{H}_{p+j}$ is a (p x p) matrix whose (j+1, j) and (j, j+1) elements are 1, and 0 elsewhere $\forall$ j = 1(1)(p-1).

Relevant likelihood equations are given by

$$\text{tr}(\underset{\sim}{Q} \underset{\sim}{H}_i) = \text{tr}(\underset{\sim}{Q} \underset{\sim}{H}_{p+j}) = 0 \;\forall\; i = 1(1)p, \;\forall\; j = 1(1)(p-1).$$

From these, we easily derive

$$\underset{\sim}{Q} = \begin{bmatrix} 0 & & & & \\ 0 & 0 & & & \\ q_{31} & 0 & 0 & \text{Sym.} & \\ \vdots & \vdots & & & \\ q_{p1} & q_{p2} & q_{p3} \cdots 0 & & 0 \end{bmatrix} = \sum_{\alpha=1}^{\upsilon} \underset{\sim}{M}_\alpha \, q_\alpha \qquad (2\cdot2\cdot33)$$

where $\upsilon = \binom{p+1}{2} - (2p-1)$ and $\underset{\sim}{M}_\alpha$'s are appropriately chosen symmetric matrices. We note that

$$\text{diag}\cdot(\underset{\sim}{Q}) = 0 \;,\; \text{diag}(\underset{\sim}{Q}\underset{\sim}{\Sigma}) = \underline{0} \;,\; \text{diag}\cdot(\underset{\sim}{\Sigma}^{-1}\underset{\sim}{S}) = \text{diag}(\underset{\sim}{I}_p)$$

and (1,2) and (p-1,p) element of $Q\Sigma$ are equal to 0 . Moreover, $\Sigma\ Q\ \Sigma$ would be found to have its (1,1) and (p,p) element equal to 0, as a result of which we get [vide (2.2.19) and (2.2.20)]

$$\hat{\underline{\theta}}_1 = (\hat{a}_1\ \hat{a}_p)' = \hat{\underline{\theta}}_1^{(\ell)} = (s_{11}\ s_{pp})' \qquad (2.2.34)$$

and

$$\underline{\theta}_2 = (a_2, a_3, \cdots a_{p-1};\ b_1, b_2, \cdots, b_{p-1})'$$

$$= \underset{\sim}{B}_2\ \underset{\sim}{A}^{-1}\underline{b} + \hat{\underline{\theta}}_2^{(\ell)} \qquad (2.2.35)$$

where the column vector $\hat{\underline{\theta}}_2^{(\ell)} = (s_{22}, s_{33}, \cdots, s_{p-1,p-1}, s_{23}, \cdots, s_{p-1,p})'$ and $\underset{\sim}{B}_2$ is the matrix $\underset{\sim}{B}$ of (2.2.17) whose first and last row are deleted, $\underset{\sim}{A}$ and $\underline{b}$ being the same as in (2.2.17).

As MLE of $\underline{\theta}_1$ is found in explicit form (equal to its LSE), the MLE of $\underline{\theta}_2$ is obtained by iteratively solving the equation (2.2.35).

In particular, for p = 3, components of $\underline{\theta}_2$ may be solved explicitly as follows :

$$\hat{b}_2 = s_{23} - \left[\frac{s_{12}s_{33} - s_{23}s_{13}}{s_{11}s_{33} - s_{13}^2} \right] s_{13} ,$$

$$\hat{b}_1 = s_{12} - (s_{13}/s_{33})\hat{b}_2 ,$$

$$\hat{a}_2 = s_{22} - \left[\frac{2s_{13}}{s_{11}\ s_{33}} \right] \hat{b}_1\ \hat{b}_2 . \qquad (2.2.36)$$

(d) Centrosymmetric Covariance Structure

A pxp $\underset{\sim}{\Sigma}$-matrix will be called a centrosymmetric matrix when, apart from symmetry, its elements (σ_{ij}) obey the relationship

$$\sigma_{ij} = \sigma_{p-i+1,p-j+1} \qquad (2.2.37)$$

$\forall$ i,j = 1(1)p. The number (k) of estimable parameters in the case of a centrosymmetric covariance matrix will be $[(p+1)/2][(p+2)/2]$ where the symbol [m] stands for the greatest integer contained in m .

Relevant likelihood equations will also establish a relationship in the elements (q_{ij}) of $\underset{\sim}{Q}$-matrix as

$$q_{ij} = - q_{p-i+1,\ p-j+1} \qquad (2.2.38)$$

$\forall$ i,j = 1(1)p. We next establish a relationship in the elements $(\underset{\sim}{\Sigma}\ \underset{\sim}{Q}\ \underset{\sim}{\Sigma})_{ij}$ of $\underset{\sim}{\Sigma}\ \underset{\sim}{Q}\ \underset{\sim}{\Sigma}$ matrix as follows :

$$(\underset{\sim}{\Sigma}\ \underset{\sim}{Q}\ \underset{\sim}{\Sigma})_{ij} = \sum_{u=1}^{p}\sum_{v=1}^{p} \sigma_{iu}q_{uv}\sigma_{vj}$$

$$= \sum_{u=1}^{p}\sum_{v=1}^{p} \sigma_{p-i+1,p-u+1}(-q_{p-u+1,p-v+1})\sigma_{p-v+1,p-j+1}$$

$$= \sum_{u'=1}^{p}\sum_{v'=1}^{p} \sigma_{p-i+1,u'}\ q_{u'v'}\ \sigma_{v',p-j+1}$$

$$= - (\underset{\sim}{\Sigma}\ \underset{\sim}{Q}\ \underset{\sim}{\Sigma})_{p-i+1,p-j+1} \qquad (2.2.39)$$

$\forall$ i,j = 1(1)p . We may apply this relationship directly upon the identity $\underset{\sim}{\Sigma} = \underset{\sim}{S} + \underset{\sim}{\Sigma}\ \underset{\sim}{Q}\ \underset{\sim}{\Sigma}$ to find out MLE of σ_{ij} as follows :

$$tr(\underset{\sim}{H}_{ij}\ \underset{\sim}{\Sigma}) = tr(\underset{\sim}{H}_{ij}\underset{\sim}{S}) + tr(\underset{\sim}{H}_{ij}\ \underset{\sim}{\Sigma}\underset{\sim}{Q}\underset{\sim}{\Sigma}).$$

$\underset{\sim}{H}_{ij}$ is the sparse matrix with (i,j), (j,i), (p-i+1,p-j+1) and (p-j+1, p-i+1) elements being unity. Hence

$$\sigma_{ij} + \sigma_{p-i+1,p-j+1} = (s_{ij} + s_{p-i+1,p-j+1}) + (\underset{\sim}{\Sigma Q \Sigma})_{ij} + (\underset{\sim}{\Sigma Q \Sigma})_{p-i+1,p-j+1} \cdot$$

Using (2.2.37) and (2.2.39), we obtain the desired result as follows :

$$\hat{\sigma}_{ij} = (s_{ij} + s_{p-i+1,p-j+1})/2 \qquad (2.2.40)$$

$\forall$ i,j = 1(1)p. Thus the MLE of [(p+1)/2] [(p+2)/2] number of unknown parameters of the centrosymmetric covariance matrix is obtained explicitly.

2.2.4 Alternative Expressions of the Iterative Process

A number of alternative expressions of the iterative scheme based on (2.2.18) can be derived from the basic identity (2.2.9). We shall present here only three of them.

(a) Mapping from Vec $\underset{\sim}{S}$ onto $\underline{\theta}_{(i+1)}$

Using 'Vec' operator upon matrices [cf. Graham (1981)], we may reexpress the elements of $\underset{\sim}{B}$, $\underset{\sim}{A}$ and $\underline{b}$ of (2.2.17) as follows :

$$b_{t\alpha} = (\text{Vec } \underset{\sim}{M}_{\alpha})' (\underset{\sim}{\Sigma} \otimes \underset{\sim}{\Sigma})(\text{Vec } \underset{\sim}{\bar{H}}_{t}), t = 1(1)k, \alpha = 1(1)\upsilon$$

$$a_{\alpha\beta} = (\text{Vec } \underset{\sim}{M}_{\beta})' (\underset{\sim}{\Sigma} \otimes \underset{\sim}{\Sigma})(\text{Vec } \underset{\sim}{M}_{\alpha}), \alpha,\beta = 1(1)\upsilon ,$$

$$b_{\alpha} = - (\text{Vec } \underset{\sim}{M}_{\alpha})' (\text{Vec } \underset{\sim}{S}), \alpha = 1(1)\upsilon .$$

Let us now define

$$\underset{\sim}{\Lambda} = (\text{Vec } \underset{\sim}{M}_1, \text{Vec } \underset{\sim}{M}_2 , \ldots, \text{Vec } \underset{\sim}{M}_{\upsilon})_{p^2 x \upsilon} ,$$

$$\underset{\sim}{\Gamma} = (\text{Vec } \bar{\underset{\sim}{H}}_1 , \text{Vec } \bar{\underset{\sim}{H}}_2 , \cdots , \text{Vec } \bar{\underset{\sim}{H}}_k)_{p^2 x k} ,$$

$$\tilde{\underline{S}} = (\text{Vec } \underset{\sim}{S})_{p^2 x 1} \quad \text{and} \quad \underset{\sim}{L} = \underset{\sim}{\Sigma} \otimes \underset{\sim}{\Sigma}$$

to get

$$\underset{\sim}{A} = \underset{\sim}{\Lambda}' \underset{\sim}{L} \underset{\sim}{\Lambda}, \quad B = \underset{\sim}{\Gamma}' \underset{\sim}{L} \underset{\sim}{\Lambda}, \quad \underline{b} = \underset{\sim}{\Lambda}' \tilde{\underline{S}} , \quad \hat{\underline{\theta}}' = \underset{\sim}{\Gamma}' \tilde{\underline{S}}$$

so that (2.2.15) and (2.2.16) may be reformulated as

$$\underline{\theta} = \hat{\underline{\theta}}' - \underset{\sim}{\Gamma}' \underset{\sim}{L} \underset{\sim}{\Lambda} (\underset{\sim}{\Lambda}' \underset{\sim}{L} \underset{\sim}{\Lambda})^{-1} \underset{\sim}{\Lambda}' \tilde{\underline{S}}$$

$$= \underset{\sim}{\Gamma}' \underset{\sim}{P}^* \tilde{\underline{S}} , \text{ say} \qquad (2.2.41)$$

and $$\underline{\theta}(i+1) = \underset{\sim}{\Gamma}' \underset{\sim}{P}^*_{(i+1)} \tilde{\underline{S}} , \quad i = 0(1) \infty \qquad (2.2.42)$$

where $\underset{\sim}{P}^* = [\underset{\sim p^2}{I} - \underset{\sim}{L} \underset{\sim}{\Lambda} (\underset{\sim}{\Lambda}' \underset{\sim}{L} \underset{\sim}{\Lambda})^{-1} \underset{\sim}{\Lambda}']$ and $\underset{\sim}{P}^*_{(i)}$ is the i-th trial estimate of $\underset{\sim}{P}^*$ based on $\underline{\theta}_{(i)}$. The matrix $\underset{\sim}{\Gamma}' \underset{\sim}{P}^*_{(i)}$ in (2.2.28) maps $\tilde{\underline{S}}$ (or Vec$\underset{\sim}{S}$) onto $\underline{\theta}_{(i+1)}$.

(b) Mapping from $\underline{\theta}_{(i)}$ onto $\underline{\theta}_{(i+1)}$

Noting that Vec $\underset{\sim}{\Sigma} = \underset{\sim}{\Gamma}^* \underline{\theta}$ where $\underset{\sim}{\Gamma}^* = (\text{Vec } \underset{\sim}{H}_1 , \cdots , \text{Vec } \underset{\sim}{H}_k)_{p^2 x k}$, we have from the identity (2.2.9),

$$\text{Vec } \underset{\sim}{\Sigma} = (\underset{\sim p}{I} \otimes \underset{\sim}{\Sigma Q}) \text{Vec} \underset{\sim}{\Sigma} + \text{Vec } \underset{\sim}{S} ,$$

$$\text{or} \quad \underset{\sim}{\Gamma}^* \underline{\theta} = (\underset{\sim p}{I} \otimes \underset{\sim}{\Sigma Q}) \underset{\sim}{\Gamma}^* \underline{\theta} + \tilde{\underline{S}} .$$

Premultiplying both sides by $(\underset{\sim}{\Gamma}^{*\prime} \underset{\sim}{\Gamma}^*)^{-1} \underset{\sim}{\Gamma}^{*\prime}$ and noting that

$$\underset{\sim}{\Gamma}^{*\prime} \underset{\sim}{\Gamma}^* = \text{diag}(\| \underset{\sim}{H}_t \|_F^2 , \ t = 1(1)k)$$

and $(\underset{\sim}{\Gamma}^{*\prime} \underset{\sim}{\Gamma}^*)^{-1} \underset{\sim}{\Gamma}^{*\prime} = \underset{\sim}{\Gamma}'$,

we have

$$\underset{\sim}{\theta} = \underset{\sim}{\Gamma}' (\underset{\sim}{I}_p \circledast \underset{\sim}{\Sigma} \underset{\sim}{Q}) \underset{\sim}{\Gamma}^* \underset{\sim}{\theta} + \hat{\underset{\sim}{\theta}}^{\ell} \quad . \qquad (2.2.43)$$

For iterative purposes, we may write

$$\underset{\sim}{\theta}_{(i+1)} = \underset{\sim}{\Gamma}' (\underset{\sim}{I}_p \circledast \underset{\sim}{\Sigma}_{(i)} \underset{\sim}{Q}_{(i+1)}) \underset{\sim}{\Gamma}^* \underset{\sim}{\theta}_{(i)} + \hat{\underset{\sim}{\theta}}^{\ell}$$

$$= \underset{\sim}{T}_{(i)} \underset{\sim}{\theta}_{(i)} + \hat{\underset{\sim}{\theta}}^{\ell} \quad , \text{ say } i = O(1) \infty \; . \qquad (2.2.44)$$

This expression shows a mapping from $\underset{\sim}{\theta}_{(i)}$ onto $\underset{\sim}{\theta}_{(i+1)}$. It may be noted that the matrix $\underset{\sim}{Q}_{(i+1)}$ required for (2.2.44) is, however, obtained by recalling (2.2.6) and (2.2.12) as

$$\underset{\sim}{Q}_{(i+1)} = \sum_{\alpha=1}^{\upsilon} \underset{\sim}{M}_{\alpha} \; q_{\alpha(i+1)} \qquad (2.2.45)$$

from which

$$\text{Vec } \underset{\sim}{Q}_{(i+1)} = \underset{\sim}{\Lambda} \underset{\sim}{q}_{(i+1)} \qquad (2.2.46)$$

and

$$\underset{\sim}{q}_{(i+1)} = \underset{\sim}{A}^{-1}_{(i)} \; \underset{\sim}{b} = - (\underset{\sim}{\Lambda}' \underset{\sim}{L}_{(i)} \underset{\sim}{\Lambda})^{-1} \underset{\sim}{\Lambda}' \underset{\sim}{\tilde{S}} \; . \qquad (2.2.47)$$

(c) Mapping from $\hat{\underset{\sim}{\theta}}^{\ell}$ onto $\underset{\sim}{\theta}_{(i+1)}$

Reformulating (2.2.43) and (2.2.44) respectively as

$$\underset{\sim}{\theta} = [\underset{\sim}{I}_k - \underset{\sim}{\Gamma}' (\underset{\sim}{I}_p \circledast \underset{\sim}{\Sigma} \underset{\sim}{Q}) \underset{\sim}{\Gamma}^*]^{-1} \; \hat{\underset{\sim}{\theta}}^{\ell} \qquad (2.2.48)$$

and $$\underset{\sim}{\theta}_{(i+1)} = [\underset{\sim}{I}_k - \underset{\sim}{\Gamma}' (\underset{\sim}{I}_p \circledast \underset{\sim}{\Sigma}_{(i)} \underset{\sim}{Q}_{(i+1)}) \underset{\sim}{\Gamma}^*]^{-1} \; \hat{\underset{\sim}{\theta}}^{\ell} \qquad (2.2.49)$$

$i = O(1)\infty$, we find that equation (2.2.49) may be treated as a mapping from $\hat{\underset{\sim}{\theta}}^{\ell}$ onto $\underset{\sim}{\theta}_{(i+1)}$. Here, $Q_{(i+1)}$ is to be replaced by the expression (2.2.45).

Thus, in order to solve the normal equations based on the Q-procedure, we get three alternative algorithms expressed as (2.2.41) or (2.2.17), (2.2.43) and (2.2.48), all of which have built-in iterative processes given respectively by (2.2.42) or (2.2.18), (2.2.44) and (2.2.49). However, in this chapter, we confine ourselves to the iterative process (2.2.18) or (2.2.42) due to its simplicity.

2.2.5 Q-procedure ; An Application of Projection Operator

It may be noted that $\underset{\sim}{P}^{*} = \underset{\sim}{I}_{p^2} - \underset{\sim}{P}$ where $\underset{\sim}{P} = \underset{\sim}{L}\underset{\sim}{\Lambda}(\underset{\sim}{\Lambda}'\underset{\sim}{L}\underset{\sim}{\Lambda})^{-1}\underset{\sim}{\Lambda}'$, involved in the iterative process (2.2.41) is a projection operator in $\textcircled{H}_{p^2}^{1} \subset \mathbb{E}^{p^2}$, where $\underset{\sim}{\widetilde{S}} \in \textcircled{H}_{p^2}$. This is a continuous linear mapping $\underset{\sim}{P} : \textcircled{H}_{p^2} \longrightarrow \textcircled{H}_{p^2}$ with $\underset{\sim}{P}^2 = \underset{\sim}{P}$. This means that if $\textcircled{H}_{p^2}^{0}$ is the range of $\underset{\sim}{P}$ and $\textcircled{H}_{p^2}^{1}$ the null space of $\underset{\sim}{P}$, then $\textcircled{H}_{p^2} = \textcircled{H}_{p^2}^{0} + \textcircled{H}_{p^2}^{1}$ (direct sum) and dim. $(\textcircled{H}_{p^2}^{0})$ = rank of $\underset{\sim}{P} = \upsilon$ and dim. $(\textcircled{H}_{p^2}^{1})$ = rank $\underset{\sim}{P}^{*} = p^2 - \upsilon$, and every $\underset{\sim}{\widetilde{S}} \in \textcircled{H}_{p^2}$ has unique decomposition

$$\underset{\sim}{\widetilde{S}} = \underset{\sim}{P}\,\underset{\sim}{\widetilde{S}} + \underset{\sim}{P}^{*}\underset{\sim}{\widetilde{S}} . \qquad (2.2.50)$$

$\underset{\sim}{I}_{p^2}$ is the identity operator in $\textcircled{H}_{p^2}$.

Thus, equation (2.2.50) shows that although $\underset{\sim}{P}$ and $\underset{\sim}{P}^{*}$ are functions of $\underline{\Theta}$ (through $\underset{\sim}{L}$), $\underset{\sim}{P}\underset{\sim}{\widetilde{S}}$ and $\underset{\sim}{P}^{*}\,\underset{\sim}{\widetilde{S}}$ add to a fixed vector $\underset{\sim}{\widetilde{S}}$. This fact is very important in the iterative process (2.2.41) in the sense that everytime an unique component (depending upon the earlier iterate) from $\underset{\sim}{\widetilde{S}}$ is utilized to calculate the next iterate.

From the linear structure (2.1.1), Vec $\underset{\sim}{\Sigma} = \underset{\sim}{\Gamma}^{*}\,\underline{\Theta}$

where $\underset{\sim}{\Gamma}^* = (\text{Vec } \underset{\sim}{H}_1 , \cdots \text{ Vec } \underset{\sim}{H}_k)_{p^2 xk}$. Vec $\underset{\sim}{\Sigma} \in \mathcal{M}(\underset{\sim}{\Gamma}^*)$, a subspace formed by the columns of $\underset{\sim}{\Gamma}^*$ — representing the image of linear mapping $\underset{\sim}{\Gamma}^* : \textcircled{H}_k \longrightarrow \mathcal{M}(\underset{\sim}{\Gamma}^*)$. By denoting $\underset{\sim}{S}^* = (\text{Vec}\underset{\sim}{S}_1 , \text{Vec}\underset{\sim}{S}_2 , \cdots , \text{Vec}\underset{\sim}{S}_k)_{p^2 xk}$ with $\text{Vec}\underset{\sim}{S}_t = (\text{Vec}\underset{\sim}{S})^* (\text{Vec}\underset{\sim}{H}_t)$ (Hadamard product), $\forall t = 1(1)k$, $\underset{\sim}{\widetilde{S}}$ may be expressed as

$$\underset{\sim}{\widetilde{S}} = \underset{\sim}{S}^* \underline{1}_k \qquad (2.2.51)$$

where $\underline{1}_k = (1, \cdots , 1)'_{1xk}$. In general, $\underset{\sim}{S}^* \underline{1}_k \notin \mathcal{M}(\underset{\sim}{\Gamma}^*)$, but $\underset{\sim}{S}^* \underline{1}_k \in \textcircled{H}_{p^2}$. Equation (2.2.41) is an iterative process of approximating $\underset{\sim}{S}^* \underline{1}_k$ by certain $\underset{\sim}{\Gamma}^* \underline{\theta} \in \mathcal{M}(\underset{\sim}{\Gamma}^*)$, through the application of the projection operator.

2.2.6 Q-procedure in terms of Orthogonal Projector

We note that $\underset{\sim}{P}^*$ involved in the iterative process (2.2.41) is not an orthogonal projection operator in the Euclidean norm. Nevertheless, it may be treated so in a special innerproduct space defined under $\underset{\sim}{L}^{-1}$-semi inner product [e.g. $(\underline{X} , \underline{Y}) = \underline{Y}' \underset{\sim}{L}^{-1} \underline{X}$] and $\underset{\sim}{L}^{-1}$-semi norm [e.g. $\|\underline{X}\| = \underline{X}' \underset{\sim}{L}^{-1} \underline{X}$] where $\underset{\sim}{L} =$ p.d. matrix $\underset{\sim}{\Sigma} \textcircled{x} \underset{\sim}{\Sigma}$. [cf. Rao and Mitra, 1971, p. 45]. Let us define

$$\underset{\sim}{Z} = \underset{\sim}{L}^{1/2} \underset{\sim}{\Gamma} \quad , \quad \underset{\sim}{W} = \underset{\sim}{L}^{1/2} \underset{\sim}{\Lambda} , \quad \underline{s}^* = \underset{\sim}{L}^{-1/2} \underset{\sim}{\widetilde{S}} \qquad (2.2.52)$$

where $\underset{\sim}{L}^{1/2} = \underset{\sim}{\Sigma}^{1/2} \textcircled{x} \underset{\sim}{\Sigma}^{1/2}$ and $\underset{\sim}{\Sigma}^{1/2}$ is the square root matrix of $\underset{\sim}{\Sigma}$. Then (2.2.41) may be rewritten as

$$\underline{\theta} = \underset{\sim}{Z}' \underset{\sim}{R}^* \underline{s}^* \qquad (2.2.53)$$

where $\underset{\sim}{R}^* = \underset{\sim}{I}_{p^2} - \underset{\sim}{P}^{**}$ and $\underset{\sim}{P}^{**} = \underset{\sim}{W}(\underset{\sim}{W}'\underset{\sim}{W})^{-1}\underset{\sim}{W}'$. The factors $\underset{\sim}{Z}$, $\underset{\sim}{R}^*$, $\underline{s}^*$ are all functions of $\underline{\theta}$ by their very expressions. The matrix

$\underset{\sim}{R}^{*}$ is an orthogonal projector onto range $\mathcal{M}(\underset{\sim}{R}^{*})$ along the null space $\mathcal{M}(\underset{\sim}{P}^{**})$, the inner product space being $\textcircled{H}_{p}^{*}{}_{2} = \mathcal{M}(P^{**}) \oplus \mathcal{M}(R^{*})$ and by construction

$$\textcircled{H}_{p}^{*}{}_{2} = \left\{ \underline{s}^{*} / \underline{s}^{*} = \underset{\sim}{L}^{-1/2} \underline{\tilde{S}} \,,\ \underline{\tilde{S}} = \mathrm{Vec}\underset{\sim}{S} \in \textcircled{H}_{p}{}_{2} \right\}.$$

Hence, $\textcircled{H}_{p}{}_{2} \longrightarrow \textcircled{H}_{p}^{*}{}_{2}$ is one-to-one and onto for a given $\underset{\sim}{\Sigma}$.

2.2.7 Q-procedure : An Application of Constrained Inverse

Iterative process (2.2.41) may be observed from the standpoint of constrained inverse (Rao and Mitra, 1971, p. 102) of $\underset{\sim}{L}$. To show this, we shall reproduce two theorems from Rao and Mitra (1971) (without proof) in our version.

Theorem 2.1 : Let $\underset{\sim}{\Lambda}$ be a given matrix and $\mathcal{M}(\underset{\sim}{\Lambda})$ be a linear subspace formed by the columns of $\underset{\sim}{\Lambda}$. Then, $\underset{\sim}{T}$ may be called constrained inverse of $\underset{\sim}{L}$ and will be of the form

$$\underset{\sim}{T} = \underset{\sim}{\Lambda}(\underset{\sim}{\Lambda}' \underset{\sim}{L} \underset{\sim}{\Lambda})^{-1} \underset{\sim}{\Lambda}' \qquad (2.2.54)$$

if and only if rank $(\underset{\sim}{\Lambda}' \underset{\sim}{L} \underset{\sim}{\Lambda})$ = rank $(\underset{\sim}{\Lambda})$. This $\underset{\sim}{T}$ is independent of the choice of $\underset{\sim}{\Lambda}$, but the chosen matrix must belong to $\mathcal{M}(\underset{\sim}{\Lambda})$. ■

Theorem 2.2: Given $\underset{\sim}{L}, \underset{\sim}{\Lambda}, \underset{\sim}{T}$ and $\mathcal{M}(\underset{\sim}{\Lambda})$, as above, any arbitrary vector $\underline{\tilde{S}}$ admits a unique decomposition

$$\underline{\tilde{S}} = \underset{\sim}{L}\,\underline{u} + \underline{w} \qquad (2.2.55)$$

where $\underline{u} \in \mathcal{M}(\underset{\sim}{\Lambda})$, $\underline{w} \in \mathcal{M}^{\perp}(\underset{\sim}{\Lambda})$ and $\underline{u} = \underset{\sim}{T}\,\underline{\tilde{S}}$. ■

Clearly, another way of writing (2.2.41) and (2.2.42) is that

$$\underset{\sim}{\Theta} = \underset{\sim}{\Gamma}' \underline{w} \qquad (2.2.56a)$$

and

$$\underset{\sim}{\Theta}_{(i+1)} = \underset{\sim}{\Gamma}' \underline{w}(\underset{\sim}{\Theta}_{(i)}) , \qquad (2.2.56b)$$

$i = 0(1) \infty$.

For an interpretation of (2.2.56a), we note that every new iteration (i+1) starts with the old iterated vector $\underset{\sim}{\Theta}_{(i)}$ to generate the sequence

$$\underset{\sim}{L}(\underset{\sim}{\Theta}_{(i)}) = \underset{\sim}{\Sigma}(\underset{\sim}{\Theta}_{(i)}) \otimes \underset{\sim}{\Sigma}(\underset{\sim}{\Theta}_{(i)}),$$

and

$$\underset{\sim}{T}(\underset{\sim}{\Theta}_{(i)}) = \underset{\sim}{\Lambda}[\underset{\sim}{\Lambda}' \underset{\sim}{L}(\underset{\sim}{\Theta}_{(i)}) \underset{\sim}{\Lambda}]^{-1} \underset{\sim}{\Lambda}'$$

which split the given vector $\underset{\sim}{\widetilde{S}}$ uniquely as

$$\underset{\sim}{\widetilde{S}} = \underset{\sim}{L}(\underset{\sim}{\Theta}_{(i)}) \underset{\sim}{T}(\underset{\sim}{\Theta}_{(i)}) \underset{\sim}{\widetilde{S}} + \underline{w}(\underset{\sim}{\Theta}_{(i)}) \qquad (2.2.57)$$

so that $\underline{w}(\underset{\sim}{\Theta}_{(i)})$ may be utilised to calculate the new iterate $\underset{\sim}{\Theta}_{(i+1)}$ by (2.2.56b).

We may count two extreme situations without making any iterations. One, when $\underset{\sim}{S}$ is perfectly in tune with the structure of $\underset{\sim}{\Sigma}$ i.e., when $\underset{\sim}{S} = \sum_{t=1}^{k} \underset{\sim}{H}_t s_t$, say, then $\underset{\sim}{L}\underset{\sim}{T}\,\underset{\sim}{\widetilde{S}} = 0$ so that $\underline{w} = \underset{\sim}{\widetilde{S}}$ and $\underline{\hat{\theta}} = \underline{\hat{\theta}}^{(k)} = (s_1, s_2, \cdots, s_k)'$. Secondly, when $\underset{\sim}{\Sigma}$ has got no constraints i.e., no structure at all (except symmetry), then under likelihood equations, $\underset{\sim}{\Lambda} = \underset{\sim}{0}$ [vide subsec. 2.2.2(1)]. This indicates $\underset{\sim}{\Lambda}$ to be $\underset{\sim}{0}$ so that $\underline{\hat{\theta}} = \underset{\sim}{\Gamma}' \underset{\sim}{\widetilde{S}} = \underline{\hat{\theta}}^{(k)}$.

2.2.8 Boundedness of the Proposed Iterative Process

In order to establish boundedness of the proposed iterative process (2.2.17), we shall make use of the Euclidean norm to apply upon (2.2.53).

$$\|\underset{\sim}{Z}\|_F^2 = \mathrm{tr}(\underset{\sim}{Z}'\underset{\sim}{Z}) = \sum_{i=1}^{k} \mathrm{tr}(\underset{\sim}{\bar{H}}_t\underset{\sim}{\Sigma})^2$$

$$\leq \sum_{t=1}^{k} \mathrm{tr}(\underset{\sim}{\bar{H}}_t\,\underset{\sim}{\Sigma})'(\underset{\sim}{\bar{H}}_t\,\underset{\sim}{\Sigma}) \qquad [\text{cf. Graybill, 1969, p.235}]$$

$$= \sum_{t=1}^{k} \|\underset{\sim}{\bar{H}}_t\,\underset{\sim}{\Sigma}\|_F^2 \leq \sum_{t=1}^{k} \|\underset{\sim}{\bar{H}}_t\|_F^2 \|\underset{\sim}{\Sigma}\|_F^2 = \sum_{t=1}^{k} \mathrm{tr}(\underset{\sim}{\bar{H}}_t'\,\underset{\sim}{\bar{H}}_t)\|\underset{\sim}{\Sigma}\|_F^2$$

$$= \Big[\sum_{t=1}^{k} 1/\|\underset{\sim}{H}_t\|_F^2\Big]\|\underset{\sim}{\Sigma}\|_F^2 . \qquad (2.2.58)$$

As $\underset{\sim}{R}^*$ is a symmetric idempotent matrix of rank $(p^2-\upsilon)$,

$$\|\underset{\sim}{R}^*\|_F^2 = \mathrm{tr}(\underset{\sim}{R}^{*\prime}\underset{\sim}{R}^*) = (p^2-\upsilon) \qquad (2.2.59)$$

and

$$\|\underline{s}^*\|_F^2 = \mathrm{tr}(\underset{\sim}{S}\underset{\sim}{\Sigma}^{-1})^2 \leq \|\underset{\sim}{S}\|_F^2 \|\underset{\sim}{\Sigma}^{-1}\|_F^2 . \qquad (2.2.60)$$

Using (2.2.58)-(2.2.60), we obtain

$$\|\underset{\sim}{Z}'\underset{\sim}{R}^*\underline{s}^*\|_F \leq \Big[\sum_{t=1}^{k} 1/\|\underset{\sim}{H}_t\|_F^2\Big]^{1/2}(p^2-\upsilon)^{1/2}\big[\|\underset{\sim}{\Sigma}\|_F \|\underset{\sim}{\Sigma}^{-1}\|_F\big]\|\underset{\sim}{S}\|_F . \qquad (2.2.61)$$

For any norm $\|\cdot\|$, $\|\underset{\sim}{\Sigma}\|\,\|\underset{\sim}{\Sigma}^{-1}\|$ is termed as the condition number $w(\underset{\sim}{\Sigma})$ of a nonsingular matrix $\underset{\sim}{\Sigma}$ to indicate the sensitivity of $\underset{\sim}{\Sigma}$ to small change. This concept is important in perturbation problem in the area of error analysis [cf. Aitkinson, 1978]. The condition number $w(\underset{\sim}{\Sigma})$ would have lower bound unity but may have unlimited upper bound. Let us assume that

$\|\underset{\sim}{\Sigma}\|_F \|\underset{\sim}{\Sigma}^{-1}\|_F$ i.e., $w_F(\underset{\sim}{\Sigma})$ of (2.2.61) is finite.

Defining spectral condition number $w_\lambda(\underset{\sim}{\Sigma})$ as the ratio of the largest and the smallest eigen value of $\underset{\sim}{\Sigma}$, it may be shown that [cf. Ikramov, 1983, p. 237]

$$w_F(\underset{\sim}{\Sigma}) \leq p \; w_\lambda(\underset{\sim}{\Sigma}) \qquad (2.2.62)$$

so that from (2.2.61)

$$\| \underset{\sim}{Z}' \; \underset{\sim}{R}^* \; \underline{s}^* \|_F \leq p \left[\sum_{t=1}^{k} 1/ \| \underset{\sim}{H}_t \|_F^2 \right]^{1/2} (p^2-\upsilon)^{1/2} \cdot w_\lambda(\underset{\sim}{\Sigma}) \cdot \| \underset{\sim}{S} \|_F \cdot \qquad (2.2.63)$$

We note that r.h.s. of the inequality (2.2.63) is finite so as to mean that $\| \underset{\sim}{Z}' \; \underset{\sim}{R}^* \; \underline{s}^* \|_F$ is bounded. Any norm $\| \underset{\sim}{Z}' \; \underset{\sim}{R}^* \; \underline{s}^* \|$ may then be claimed to be bounded from the equivalence of norms (cf. Aitkinson, 1978, p. 415), once $\| \underset{\sim}{Z}' \; \underset{\sim}{R}^* \; \underline{s}^* \|_F$ is bounded.

2.2.9 Local Convergence

It is difficult in general to establish condition(s) under which an iterative process would converge globally. If, however, the initial estimate is close to the true solution, we can hope to obtain at least a locally converged solution.

We have proved the bounded nature of the iterative process (2.2.17) in subsection 2.2.8. This property, however, does not ensure its convergence. Nevertheless, once the boundedness is assured, we may look for some other conditions sufficient for local convergence. To achieve a local convergence with reference to some initial vector, we shall now state a theorem whose complete proof is available in Wendroff (1966, p. 166).

Theorem 2.3 : Suppose that $D_o \subset R^n$ and $\underline{g}(\underline{X}) \in D_o \; \forall \; \underline{X} \in D_o$.

Define a sequence

$$\underline{X}_{r+1} = \underline{g}(\underline{X}_r), \; r = 0(1) \; \infty \; . \qquad (2.2.64)$$

Let $\underset{\sim}{\mathcal{J}}(\underline{X}) = \dfrac{\partial \underline{g}(\underline{X})}{\partial \; \underline{X}}$ be the Jacobian matrix of $\underline{g}(\underline{X})$. If $\underset{\sim}{\mathcal{J}}(\underline{X})$ be

continuous in the neighborhood of $\underline{X}_0 \in D_0$ and $\| \underset{\sim}{J}(\underline{X}_0) \| < 1$, then the sequence (2.2.64) converges uniquely in the neighborhood of $\underline{X}_0$, provided $\| \underline{X}_0 - \underline{g}(\underline{X}_0) \|$ is sufficiently small. ■

This theorem seems to be quite useful to get locally unique solution of (2.2.17) in the neighborhood of an initial vector $\underline{\theta}_{(o)} \in \textcircled{H}_k$.

Writing $\underline{G}(\underline{\theta})$ for the l.h.s. of (2.2.17), $\underline{G}(\underline{\theta})$ is a scale invariant function of $\underline{\theta}$ in the sense that

$$\underline{G}(\alpha \ \underline{\theta}) = G(\underline{\theta}) \qquad (2.2.65)$$

for any non-zero scalar α . Such invariance property implies the proportionate Jacobian matrix meaning that

$$\underset{\sim}{J}(\alpha \ \underline{\theta}) = \frac{1}{\alpha} \underset{\sim}{J}(\underline{\theta}) \qquad (2.2.66)$$

since $$\underset{\sim}{J}(\alpha \ \underline{\theta}) = \frac{\partial \underline{G}(\alpha \ \underline{\theta})}{\partial \alpha \ \underline{\theta}} = \frac{\partial \underline{G}(\alpha \ \underline{\theta})}{\partial \underline{\theta}} \cdot \frac{\partial \underline{\theta}}{\partial \alpha \ \underline{\theta}} = \frac{1}{\alpha} \frac{\partial \underline{G}(\underline{\theta})}{\partial \underline{\theta}} .$$

This property of scale invariance makes the problem of choosing initial estimate $\underline{\theta}_{(o)}$ easier because it is sure that $\theta_{(o)} \in \textcircled{H}_k$ and any of its scalar multiple $\alpha \underline{\theta}_{(o)} \in \textcircled{H}_k$ would produce the same sequence of iterates. In order to locate the neighborhood of convergence (which is denoted by D_o in Theorem 2.3), it is, of course, necessary to choose such an α as will ensure that (i) $\| \alpha \underline{\theta}_{(o)} - \underline{G}(\underline{\theta}_{(o)}) \|$ is sufficiently small (smallness being quantified *a priori* by ϵ , say) and (ii) $\| \underset{\sim}{J}(\alpha \underline{\theta}_{(o)}) \|$ being equal to $\| \underset{\sim}{J}(\underline{\theta}_{(o)}) \| / |\alpha|$, is less than unity. In the case of Euclidean norm, we observe that

$$\underset{\alpha}{\text{Min}} \left\| \alpha \underset{\sim}{\Theta}_{(o)} - \underset{\sim}{G}(\underset{\sim}{\Theta}_{(o)}) \right\|_F^2 = \left\| \underset{\sim}{G}(\underset{\sim}{\Theta}_{(o)}) \right\|_F^2 - \frac{[\underset{\sim}{\Theta}'_{(o)} \underset{\sim}{G}(\underset{\sim}{\Theta}_{(o)})]^2}{\left\| \underset{\sim}{\Theta}_{(o)} \right\|_F^2} . \tag{2.2.67}$$

In view of the afore-said result, the choice of $\underset{\sim}{\Theta}_{(o)}$ primarily should be such that r.h.s. of (2.2.67) must not exceed ϵ^2. The choice of suitable α to satisfy both the above-mentioned conditions (i) and (ii) is undertaken at the next stage. The smaller the two norms in (i) and (ii), the lesser is the number of iterations required to get the converged solution to a desired degree of accuracy.

2.3 COMPARISON OF Q-PROCEDURE AND ANDERSON'S PROCEDURE

The iterative scheme based on Anderson's (1973, formula 9) procedure for obtaining the solution of ML equations is derived directly from (2.2.3) as follows :

$$\sum_{u=1}^{k} \text{tr}(\underset{\sim}{\Sigma}^{-1} \underset{\sim}{H}_t \underset{\sim}{\Sigma}^{-1} \underset{\sim}{H}_u) \, \theta_u = \text{tr}(\underset{\sim}{\Sigma}^{-1} \underset{\sim}{H}_u \underset{\sim}{\Sigma}^{-1} \underset{\sim}{S}), \quad t = 1(1)k. \tag{2.3.1}$$

At the (i+1)th stage, we slove

$$\sum_{u=1}^{k} \text{tr}(\underset{\sim}{\Sigma}_{(i)}^{-1} \underset{\sim}{H}_t \underset{\sim}{\Sigma}_{(i)}^{-1} \underset{\sim}{H}_u) \, \theta_{u(i+1)} = \text{tr}(\underset{\sim}{\Sigma}_{(i)}^{-1} \underset{\sim}{H}_u \underset{\sim}{\Sigma}_{(i)}^{-1} \underset{\sim}{S}), \tag{2.3.2}$$

$t = 1(1)k$, $i = 0(1) \infty$. This is equivalent to Browne's (1977) algorithm for obtaining generalized least squares estimate (GLSE) of the parameters of $\underset{\sim}{\Sigma}$ of (2.1.1) using weight matrix $\underset{\sim}{\Sigma}^{-1}$. Following Browne (1977), the equation (2.3.1) may be expressed as

$$\underset{\sim}{\theta} = (\underset{\sim}{\Gamma}^{*\prime}\ \underset{\sim}{L}^{-1}\ \underset{\sim}{\Gamma}^{*})^{-1}\ \underset{\sim}{\Gamma}^{*\prime}\ \underset{\sim}{L}^{-1}\ \underset{\sim}{S}$$

$$= \underset{\sim}{U}(\underset{\sim}{\theta})\ \underset{\sim}{S}\ , \quad \text{say} \qquad (2.3.3)$$

where $\underset{\sim}{\Gamma}^{*} = (\text{Vec } H_1, \text{ Vec } H_2, \cdots, \text{ Vec } H_k)$. Other notations have already been defined.

The estimates of $\underline{\theta}$ in (2.3.3) cannot be solved algebraically in closed form, so that some non-linear optimization procedure such as the Newton-Raphson algorithm or the Gauss-Newton algorithm or its modification is required. Here, the matrix $\underset{\sim}{\Gamma}^{*\prime}\ \underset{\sim}{L}^{-1}\ \underset{\sim}{\Gamma}^{*} = \underset{\sim}{H}\ (\underline{\theta})$, say, is an appropriate p.d. matrix that approximates minus two times the Hessian matrix, i.e., the matrix of second derivative of log-likelihood function, at the point of convergence.

Since the matrix $\underset{\sim}{H}^{-1}(\underline{\theta})$ is readily available at the last iteration, the Newton-Rephson procedure as applied to Anderson's algorithm not only produces the MLE of $\underline{\theta}$ but also its standard error (SE) as well. This 'by product' feature is not available in the proposed Q-procedure because it does not require the second derivative of (2.2.1).

Hereafter, we refer to Anderson's (1973) algorithm (2.3.3) as U-scheme. This U-scheme is quite competitive to the G-scheme, say, (2.2.41) or (2.2.17) with respect to its performance of obtaining the ML solutions. As in the case of the G-scheme, the U-scheme is also scale invariant and has a proportionate Jacobian matrix. We are, however, more optimistic about the G-scheme in the context of covergence conditions as given

in Theorem 2.3 and also in terms of practicability due to the following reasons :

(a) Fulfilment of the conditions under Theorem 2.1

We may derive (i) the upper bound of the iterates, (ii) upper bound of the Jacobian matrix and (iii) upper bound of the difference between the first iterate and the initial estimate for both the U-scheme and G-scheme. We may observe that all these upper bounds for the G-scheme are less than the corresponding one of those in the U-scheme for the same initial start. [Vide Appendix 1 of Mukherjee and Maiti (1988a)].

Clearly, to apply Theorem 2.3, its relevant conditions are expected to be more often satisfied by the G-scheme than by the U-scheme for the same initial estimate.

(b) Operational efficiency

(i) The U-scheme requires the computation of inverses of two matrices of order (p x p) and (k x k) whereas the G-scheme requires a single inverse of order (υ x υ). Furthermore, as the number of estimable parameters (k) increases for a given $\underset{\sim}{\Sigma}$ matrix, the number of " auxiliary parameters" (υ) becomes smaller and smaller. Thus, the G-scheme greatly saves computing time and computer space.

(ii) Computational error generally becomes considerable in the U-scheme as it undertakes some sort of inverse of another inverse matrix contrary to a single inverse in the G-scheme.

Such computational error becomes severe in the next

iterate, particularly when the preceding iterate makes $\underset{\sim}{\Sigma}$ " ill-conditioned" . As a result, invariance property of the U-scheme breakes down. But such a property is always preserved in the G-scheme.

(iii) With reference to any matrix norm, we obtain from the identity (2.2.8) the following result :

$$\| \underset{\sim}{\Sigma}_{(i+1)} \| \leq \| \underset{\sim}{S} \| + \| \underset{\sim}{\Sigma}_{(i)} \|^2 \| \underset{\sim}{Q}_{(i+1)} \| \quad , \; i = 0(1)\infty \; . \qquad (2.3.4)$$

If initially $\underset{\sim}{Q}_{(1)}$ is bounded, then at each iterative step in the G-scheme, the resulting $\underset{\sim}{\Sigma}$ will also be bounded. It is to be noted here that the boundedness of $\underset{\sim}{Q}_{(i+1)}$ is guaranteed by the non-singularity of $\underset{\sim}{A}_{(i)}$ of (2.2.18). However, in the U-scheme such a control on boundedness of $\underset{\sim}{\Sigma}_{(i)}$ is not readily available.

(iv) Gradual rectification in the successive iterates in the G-scheme is possible even in the case of inadmissible initial trial solution as every iterate is the sum of a constant " rectifier" vector $\underline{\hat{\theta}}$ and some contribution due to preceding iterate. Such contribution is rarely unbounded if $(\underset{\sim}{\Lambda}' \underset{\sim}{L} \underset{\sim}{\Lambda})^{-1}$ exists [vide subsection 2.2.8]. The convergence, however, may be delayed depending upon the fulfilment of requirements of the Theorem 2.3.

In the U-scheme, such rectification at each iterative step is lacking and, as a result, inadmissible solutions during any iteration or even after convergence cannot be altogether ruled out.

(v) Unlike the U-scheme, sometimes reduction of the number

of iterated variable in the G-scheme is possible by careful observations on the rows of $\underset{\sim}{B}$-matrix which, of course, depends upon the specific structure of $\underset{\sim}{\Sigma}$. As mentioned earlier, such reduction greatly saves computational labour and space. This type of saving is not seen in the U-scheme.

2.3.1 A Numerical Illustration

For the purpose of numerical illustration, let us consider a sample covariance matrix

$$\underset{\sim}{S} = \begin{bmatrix} 8 & 6 & 3 \\ 6 & 10 & 4 \\ 3 & 4 & 9 \end{bmatrix}$$

obtained from an artificial sample data from a $N_3(\underline{X} ; \underline{\mu} , \underset{\sim}{\Sigma})$ where $\underset{\sim}{\Sigma}$ has a Toeplitz or Laurent structure as follows :

$$\underset{\sim}{\Sigma} = \begin{bmatrix} a & b & c \\ b & a & b \\ c & b & a \end{bmatrix} = a\underset{\sim}{H}_1 + b\underset{\sim}{H}_2 + c\underset{\sim}{H}_3$$

where $\underset{\sim}{H}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $\underset{\sim}{H}_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ and $\underset{\sim}{H}_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$.

The normal equations in terms of the elements of $\underset{\sim}{Q}$-matrix are

$$q_{11} + q_{22} + q_{33} = 0,\ q_{12} + q_{23} = 0 \quad \text{and} \quad q_{13} = 0$$

Under normal equations,

$$\underset{\sim}{Q} = \underset{\sim}{M}_{11}q_{11} + \underset{\sim}{M}_{12}q_{12} + \underset{\sim}{M}_{22}q_{22}$$

where

$$\underset{\sim}{M}_{11} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad \underset{\sim}{M}_{12} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix} \text{ and } \underset{\sim}{M}_{22} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

We also define

$$\underset{\sim}{\Gamma} = (\text{Vec } \underset{\sim}{H}_1/3, \quad \text{Vec } \underset{\sim}{H}_2/4, \quad \text{Vec } \underset{\sim}{H}_3/2)_{9x3}$$

$$\underset{\sim}{\Lambda} = (\text{Vec } \underset{\sim}{M}_1, \quad \text{Vec } \underset{\sim}{M}_2, \quad \text{Vec } \underset{\sim}{M}_3)_{9x3} .$$

To apply (2.2.17) or (2.2.41) we observe that

$$\underset{\sim}{A} = \underset{\sim}{\Lambda}' \underset{\sim}{L} \underset{\sim}{\Lambda} = \begin{bmatrix} 2(a^2-c^2) & 4b(a-c) & a^2-c^2 \\ 4b(a-c) & 4a(a-c) & 2b(a-c) \\ a^2-c^2 & 2b(a-c) & 2(a^2-b^2) \end{bmatrix}$$

$$\underset{\sim}{B} = \underset{\sim}{\Gamma}' \underset{\sim}{L} \underset{\sim}{\Lambda} = \begin{bmatrix} 0 & 0 & (b^2-c^2)/3 \\ 0 & 0 & b(a-c)/2 \\ 0 & 0 & b^2-ac \end{bmatrix},$$

$$\underline{b} = -\underset{\sim}{\Lambda}' \underset{\sim}{\tilde{S}} = [s_{11}-s_{33} \;\; 2(s_{12}-s_{23}) \;\; s_{22}-s_{33}]'$$

and

$$\underline{\hat{\theta}}^{\ell} = \underset{\sim}{\Gamma}' \underline{S} = [(s_{11}+s_{22}+s_{33})/3 \;\; (s_{12}+s_{23})/2 \;\; s_{13}]'.$$

Employing the G-scheme (2.2.18) or (2.2.42), we obtain the successive iterates which we report in Table 2.1 . Here, the ML solution for a, b and c (and the " auxiliary parameters" q_{11}, q_{12} and q_{22}) is obtained with LSE as the initial try-out value. Solution is found to be correct to fourth place of decimals when convergence is attained at sixth iteration only.

It may be noted that although q_{11}, q_{12} and q_{22} are estimated at each iteration, only the value of q_{22} is eventually used in calculating a, b and c due to two null columns of $\underset{\sim}{B}$.

Table 2.1

Iterated values of a, b and c with LSE as initial try-out value

Iteration	q_{11}	q_{12}	q_{22}	a	b	c
0	-	-	-	9.0	5.0	3.0
1	.051535	-.041667	-.019737	8.894737	4.703947	3.039474
2	.047718	-.039418	-.018978	8.918466	4.738648	3.093147
3	.047925	-.039680	-.018943	8.918624	4.738540	3.097206
4	.047923	-.039691	-.018939	8.918808	4.738783	3.097897
5	.047924	-.039694	-.018939	8.918823	4.738801	3.097977
6	.047924	-.039694	-.018938	8.918825	4.738804	3.097988

In order to observe the behaviour of the 1st iterate for different initial starts, we present in Table 2.2 the actual values obtained at the 1st iteration. Preserving positive definiteness of $\underset{\sim}{\Sigma}_{(o)}$, admissible values of $a_{(o)}$, $b_{(o)}$ and $c_{(o)}$ are chosen satisfying the restictions $a_{(o)} > b_{(o)}$, $a_{(o)} > c_{(o)}$ and $b_{(o)} < [a_{(o)}(a_{(o)} + c_{(o)})/2]^{1/2}$. Fixing $a_{(o)}$ at 8.9188 which is the MLE correct to 4 places of decimal (vide Table 2.1), we choose various values of $b_{(o)}$ and $c_{(o)}$ keeping the spectral condition number $w_{\lambda}(\underset{\sim}{\Sigma}_{(o)})$ increasing. We shall call $\underset{\sim}{\Sigma}_{(o)}$ "ill-conditioned" if its condition number $w_{\lambda}(\underset{\sim}{\Sigma}_{(o)})$ is

Table 2.2

Iterated values as obtained from U-Scheme where Initial values Condition Number and Alpha vary from set to set

Initial Try-out Values			Condition number	Scalar	1st Iterated Values			Minimum**	Maximum**	No.of Iterations
$a_{(o)}$	$b_{(o)}$	$c_{(o)}$	$w_{\lambda}(\Sigma_{(o)})$	α	$a_{(1)}$	$b_{(1)}$	$c_{(1)}$			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
8.9188	0.0991	0.8919	1.22	1	9.003282	4.995016	3.099563	0	0	11
				10^{2}	9.003282	4.995016	3.099563	0	0	11
				10^{-2}	9.003282	4.995016	3.099563	0	0	11
				-	(9.003282)*	(4.995016)	(3.099563)	-	-	(13)
8.9188	6.6891	1.3378	182.89	1	8.300969	3.762137	1.398048	0	0	12
				10^{2}	8.300969	3.762132	1.398048	0	5×10^{-8}	12
				10^{-2}	8.300978	3.762131	1.398048	0	88×10^{-8}	12
				-	(8.300969)	(3.762132)	(1.398048)	-	-	(12)
8.9188	1.7415	8.9178	18471.56	1	9.247801	4.999221	3.748601	0	.0001	11
				10^{2}	9.254630	5.000934	3.754830	17×10^{-8}	.0002	11
				10^{-2}	9.253131	5.001085	3.463631	0	.0001	11
				-	(9.249956)	(4.999991)	(3.749956)	-	-	(11)
8.9188	1.7403	8.9187	184715.86	1	9.183190	5.039243	3.793190	0	.0100	11
				10^{2}	9.183183	5.031241	3.853183	177×10^{-8}	.0100	11
				10^{-2}	9.261570	4.955446	-7.568429	0	.0060	13
				-	(9.249996)	(4.999999)	(3.749996)	-	-	(11)
8.9188	8.9187	8.9187	267562.00	1	9.066000	4.756000	3.826000	.0110	.0300	11
				10^{2}	9.450000	5.000000	4.040000	0	.0210	11
				10^{-2}	8.458000	4.528000	-15.142000	0	.5110	13
				-	(9.000000)	(4.750000)	(4.040000)	-	-	11
8.918825	8.918824	8.918824	26756473.00	1	6.500000	1.900000	- 0.100000	0	.2000	11
				10^{2}	11.400000	6.700000	2.000000	0	.2000	12
				10^{-2}	-8.400000	-7.600000	-3635.300000	.3	95.4000	12
				-	(9.000000)	(4.750000)	(3.500000)	-	-	(11)

* Figures appearing in the parantheses are related to G-Scheme.

** Minimum or Maximum absolute value obtained in the entries of $U(\underline{\theta}_{(o)})$ matrix where zero is expected.

very very large where $w_\lambda(\underset{\sim}{\Sigma}_{(o)})$ is the ratio of the largest and the smallest singular value of $\underset{\sim}{\Sigma}_{(o)}$. The three singular values (equal to the corresponding eigenvalues) of $\Sigma_{(o)}$ (for this example) are equal to

$$a_{(o)} - c_{(o)} \ , \ a_{(o)} + (c_{(o)}/2) \pm [\ (c^2_{(o)}/4) + 2b^2_{(o)}]^{1/2} .$$

To verify the scale invariance property of the U-and G-schemes, we take two additional scalar multiples of every initial vector, scalar (α) being equal to 10^2 and 10^{-2}. Here, $\alpha = 1$ refers to the original vector. The respective values obtained at the 1st iteration by the U-scheme are given in column 6, 7 and 8. For a given $(a_{(o)}, b_{(o)}, c_{(o)})$, the entries $(a_{(1)}, b_{(1)}, c_{(1)})$ for the three α-values are not found identical everywhere for the U-scheme. Variations among them become prominent for increasing condition number. But in the G-scheme no appreciable variations are found even upto 10 places of decimal. We therefore record only the common value of $(a_{(1)}, b_{(1)}, c_{(1)})$ obtained by the G-scheme in the parentheses in columns 6,7 and 8 of Table 2.2.

The main reason behind such variations resulting in the U-scheme is due to the fact that the entries of $\underset{\sim}{U}$-matrix [vide(2.3.3)]are to some extent affected by the computational errors resulting from the computation of two inverses, e.g.

$$\underset{\sim}{\Sigma}^{-1}_{(o)} \text{ (to get } \underset{\sim}{L}^{-1}_{(o)}) \text{ and } (\underset{\sim}{\Gamma}^{*\prime}\ \underset{\sim}{L}^{-1}_{(o)}\ \underset{\sim}{\Gamma}^{*})^{-1} .$$

In order to substantiate the above observation, we carefully note those entries of $\underset{\sim}{U}(3 \times 9)$ which, according to our algebraic

derivation, should be zero. Specifically, (1,2), (1,3), (1,4), (1,6), (1,7) and (1,8) elements of the 1st row, (2,2) and (2,7) elements in the 2nd row and (3,2), (3,4), (3,6) and (3,8) elements in the 3rd row of $\underset{\sim}{U}$-matrix should be exactly zero. To get an idea of the effect, we only record the maximum absolute and minimum absolute values among these 12 entries in the columns 9 and 10 of Table 2.2 which are supposed to be exactly zero. For easy interpretation, the absolute values less than 10^{-8} are treated as zero. In the G-scheme, no such ill effects can be claimed. As such, the number of iterations required to get the converged solution (if it converges) remains invariant irrespective of what α-value is chosen. Contrarily, in the U-scheme, the number of iterations might vary a little bit because of computational error. This may be observed from column 11 of the Table 2.2.

2.4 A CHECKING FORMULA FOR THE HESSIAN MATRIX

Let us consider log-likelihood (2.2.1) with linearly structured $\underset{\sim}{\Sigma}$ as (2.1.1). In a straight forward manner we obtain

$$\frac{\partial \log f}{\partial \theta_u} = -\frac{N}{2}\,\mathrm{tr}(\underset{\sim}{Q}\,\underset{\sim}{H}_u), \quad u = 1(1)k \tag{2.4.1}$$

$$\frac{\partial Q}{\partial \theta_t} = \underset{\sim}{\Sigma}^{-1}\underset{\sim}{H}_t\underset{\sim}{\Sigma}^{-1} - \underset{\sim}{\Sigma}^{-1}\,\underset{\sim}{H}_t\,\underset{\sim}{Q} - \underset{\sim}{Q}\,\underset{\sim}{H}_t\underset{\sim}{\Sigma}^{-1}, \quad t = 1(1)k$$

$$\frac{\partial^2 \log f}{\partial\theta_t\,\partial\theta_u} = \frac{N}{2}\left[2\,\mathrm{tr}(\underset{\sim}{Q}\,\underset{\sim}{H}_t\underset{\sim}{\Sigma}^{-1}\,\underset{\sim}{H}_u) - \mathrm{tr}(\underset{\sim}{H}_t\underset{\sim}{\Sigma}^{-1}\,\underset{\sim}{H}_u\underset{\sim}{\Sigma}^{-1})\right], \tag{2.4.2}$$

$t, u = 1(1)k$. Let us denote $\underline{h}(\underline{\theta}) = \left(\frac{\partial \log f}{\partial \theta_i},\ i = 1, \ldots, k\right)'$ and $\underset{\sim}{J}(\underline{\theta}) = \left(\left(\frac{\partial^2 \log f}{\partial\theta_t\,\partial\theta_u}\right)\right)_{kxk}$ (Hessian matrix of log-likelihood).

lt 2.4.1 $\underline{\theta}'\underset{\sim}{J}\underline{\theta} + 2\underline{\theta}'\underline{h} + \frac{Np}{2} = 0 \quad \forall \underline{\theta} \in \textcircled{H}_k$.

: From (2.4.1) $\underline{\theta}'\underline{h} = -\frac{N}{2}\sum_{i=1}^{k}\mathrm{tr}(\underset{\sim}{Q}\,\underset{\sim}{H}_i)\,\theta_i = -\frac{N}{2}\,\mathrm{tr}(\underset{\sim}{Q}\underset{\sim}{\Sigma})$ (2.4.3)

From (2.4.2) $\underline{\theta}' \underset{\sim}{J} \underline{\theta} = \sum_i \sum_j \frac{\partial^2 \log f}{\partial\theta_i \partial\theta_j} \theta_i \theta_j = N tr(\underset{\sim}{Q}\underset{\sim}{\Sigma}) - \frac{Np}{2}$. (2.4.4)

Hence, $\underline{\theta}' \underset{\sim}{J} \underline{\theta} + 2\underline{\theta}' \underline{h} = -\frac{Np}{2}$.

As $E(\underset{\sim}{Q}) \approx \underset{\sim}{Q}$, the large sample expectations, variances and covariance of $\underline{\theta}' \underline{h}$ and $\underline{\theta}' \underset{\sim}{J} \underline{\theta}$ may easily be derived via Result 2.4.1. These are assembled below.

Result 2.4.2 In large samples, $\forall \underline{\theta} \in (H)_k$

$$(i)\ E(\underline{\theta}' \underline{h}) = 0,\ (ii)\ Var(\underline{\theta}' \underline{h}) = Np/2 ,$$

$$(iii) E(\underline{\theta}' \underset{\sim}{J} \underline{\theta}) = -\frac{Np}{2} ,\ (iv)\ Var(\underline{\theta}' \underset{\sim}{J}\ \underline{\theta}) = 2Np$$

and (v) $Cov(\underline{\theta}' \underline{h} ,\ \underline{\theta}' \underset{\sim}{J}\ \underline{\theta}) = -Np$.

Since MLE ($\hat{\underline{\theta}}$) of $\underline{\theta}$ is a solution to $\underline{h} = \underline{0}$, it is obvious that $\hat{\underline{\theta}}' \underline{h} = 0$. Hence, due to Result 2.4.1

$$\hat{\underline{\theta}}' \underset{\sim}{\hat{J}}\ \hat{\underline{\theta}} = -\frac{Np}{2} . \qquad (2.4.5)$$

We thus observe that although $\underset{\sim}{J}$ is a function of $\underset{\sim}{S}$ and $\underline{\theta}$, the expression $\underline{\theta}' \underset{\sim}{J}\ \underline{\theta}$, when $\underline{\theta}$ is replaced by its MLE $\hat{\underline{\theta}}$, is numerically equal to a constant quantity $-\frac{Np}{2}$. From the computational point of view, $\hat{\underline{\theta}}' \underset{\sim}{\hat{J}}\ \hat{\underline{\theta}}$ would, on the reverse, ensure that $\hat{\underline{\theta}}$ used would be MLE of $\underline{\theta}$ when its value would be exactly $-\frac{Np}{2}$. This is, no doubt, a sufficient condition for the estimate $\hat{\underline{\theta}}$ to be MLE of $\underline{\theta}$.

Usually, when $\hat{\underline{\theta}}$ is an MLE, the Hessian matrix $\underset{\sim}{J}$ at $\hat{\underline{\theta}}$ is negative definite. Checking of such sufficient condition requires considerable amount of computational labour even when k is as small as 4 or 5 unless we take help of an electronic computer. As an alternative to this, we may employ computation of $\hat{\underline{\theta}}' \underset{\sim}{\hat{J}}\ \hat{\underline{\theta}}$ and verify whether its value coincides with $-\frac{Np}{2}$.

Such checking (although mathematically less appealing) ensures the correctness not only of the MLE but also of the Hessian matrix based on the MLE.

As MLE $\underset{\sim}{\hat{\theta}}$ is asymptotically efficient with asymptotic dispersion matrix - $[E\underset{\sim}{J}(\theta)]^{-1}$, as estimate ($\underset{\sim}{E}_1$) of disp.($\underset{\sim}{\hat{\theta}}$) may be given as - $[\underset{\sim}{J}(\hat{\theta})]^{-1}$.

Let us denote $\underset{\sim}{J}_Q(\underset{\sim}{\theta}) = N \underset{\sim}{\Gamma}^{*'}[\underset{\sim}{\Sigma}^{-1} \otimes \underset{\sim}{Q}]\underset{\sim}{\Gamma}^{*}$ and $\underset{\sim}{\mathcal{J}}(\underset{\sim}{\theta}) = -\frac{N}{2}(\underset{\sim}{\Gamma}^{*'}\ \underset{\sim}{L}^{-1}\underset{\sim}{\Gamma}^{*})$ to express $\underset{\sim}{J}(\underset{\sim}{\theta})$ formed by (2.4.2) as

$$\underset{\sim}{J}(\underset{\sim}{\theta}) = \underset{\sim}{J}_Q(\underset{\sim}{\theta}) + \underset{\sim}{\mathcal{J}}(\underset{\sim}{\theta}) \ . \qquad (2.4.6)$$

When the sample size (N) is so large that we are ready to approximate $E(\underset{\sim}{Q})$ by $\underset{\sim}{Q}$, then in (2.4.7), $E\ \underset{\sim}{J}_Q(\underset{\sim}{\theta}) \approx \underset{\sim}{Q}$ so that another estimate ($\underset{\sim}{E}_2$) of disp. ($\underset{\sim}{\hat{\theta}}$) may be obtained as $-\underset{\sim}{\mathcal{J}}^{-1}(\underset{\sim}{\hat{\theta}})$. $\underset{\sim}{\mathcal{J}}(\underset{\sim}{\theta})$ denotes the well-known Fisher's information matrix. We would mention here that $(\underset{\sim}{\Gamma}^{*'}\ \underset{\sim}{L}^{-1}\ \underset{\sim}{\Gamma}^{*})^{-1}$ is a by-product available at the last iteration of Anderson's scheme (vide subsec. 2.3). Estimate of standard errors (SE) of the components of $\underset{\sim}{\hat{\theta}}$ may be obtained as the square root of the diagonal elements of either $\underset{\sim}{E}_1$ or $\underset{\sim}{E}_2$. In practice, however, $\underset{\sim}{E}_2$ is made useful to calculate SE of $\underset{\sim}{\hat{\theta}}$.

For the example cited in 2.3.1, we calculate

$$\underset{\sim}{J}_Q(\underset{\sim}{\hat{\theta}})/N = ((\mathrm{tr}\ \underset{\sim}{\hat{Q}}\underset{\sim}{H}_t\underset{\sim}{\hat{\Sigma}}^{-1}\underset{\sim}{H}_u)) = \begin{bmatrix} -.000672 & .001441 & -.000270 \\ .001441 & -.001770 & -.001441 \\ -.000270 & -.001441 & .002983 \end{bmatrix}$$

$$\underset{\sim}{\mathcal{J}}(\hat{\underline{\theta}})/N = -\frac{1}{2}((\operatorname{tr}\hat{\underset{\sim}{\Sigma}}^{-1}H_t\hat{\underset{\sim}{\Sigma}}^{-1}H_u)) = -\begin{bmatrix} .055226 & -.051182 & .001295 \\ -.051182 & .078462 & -.021804 \\ .001295 & -.021804 & .025016 \end{bmatrix}$$

$\hat{\underline{\theta}}'\ \underset{\sim}{J}_Q(\hat{\underline{\theta}})\ \hat{\underline{\theta}} = .000003N$ and $\hat{\underline{\theta}}'\ \underset{\sim}{\mathcal{J}}(\hat{\underline{\theta}})\ \hat{\underline{\theta}} = -1.500005N$.

Hence $\hat{\underline{\theta}}'\ \underset{\sim}{J}(\hat{\underline{\theta}})\ \hat{\underline{\theta}} = -1.500002N$, satisfying (2.4.6).

We next calculate

$$\underset{\sim}{E}_1 = \begin{bmatrix} 76.0895 & 63.4032 & 53.1969 \\ 63.4032 & 69.1166 & 59.3713 \\ 53.1969 & 59.3713 & 96.4742 \end{bmatrix},$$

and

$$\underset{\sim}{E}_2 = \begin{bmatrix} 76.7934 & 64.6476 & 52.3715 \\ 64.6476 & 71.2420 & 58.7477 \\ 52.3715 & 58.7477 & 88.4681 \end{bmatrix}$$

from which the standard error (SE) of the components of $\hat{\underline{\theta}}$ are calculated as

$$SE(\hat{\underline{\theta}}) = \frac{1}{\sqrt{N}}[8.7229\ ,\ 8.3136\ ,\ 9.8221]\ ,\ \text{if } \underset{\sim}{E}_1 \text{ is used.}$$

$$= \frac{1}{\sqrt{N}}[8.7632,\ 8.4405\ ,\ 9.4057],\ \text{if } \underset{\sim}{E}_2 \text{ is used.}$$

2.5 Q^*-PROCEDURE FOR LINEAR CORRELATION STRUCTURES

In stead of structured population covariance matrix ($\underset{\sim}{\Sigma}$), we may very often come across structured population correlation matrix ($\underset{\sim}{\rho}$) with or without structured variances σ_i^2 , $i = 1,\ldots,p$. Let $\underset{\sim}{\Sigma}$ be expressed as $\underset{\sim}{\Sigma} = \underset{\sim}{D}_\sigma \underset{\sim}{\rho}\ \underset{\sim}{D}_\sigma$ such that only $(p+k)$ number of functionally independent parameters, namely, $\sigma_i(>0)$ and $\rho_t(\neq \pm 1)$

as involved in $\underset{\sim}{D}_\sigma = \text{diag.}(\sigma_i,\ i = 1(1)p)$ and $\underset{\sim}{\rho} = \underset{\sim}{I}_p + \sum_{t=1}^{k} \rho_t \underset{\sim}{H}_t$, [$\underset{\sim}{H}_t$'s being linearly independent design matrices, $t = 1(1)k$, $k < \frac{p(p-1)}{2}$] are to be estimated by solving their likelihood equations being expressed as follows :

$$\text{tr}\left(\underset{\sim}{Q} \frac{\partial \Sigma}{\partial \sigma_i}\right) = 0 \ , \ i = 1(1)p \qquad (2.5.1)$$

and

$$\text{tr}\left(\underset{\sim}{Q} \frac{\partial \Sigma}{\partial \rho_t}\right) = 0 \ , \ t = 1(1)k \qquad (2.5.2)$$

where $\underset{\sim}{Q}$ is defined as in (2.1.2). Let us define

$$\underset{\sim}{Q}^* = \underset{\sim}{D}_\sigma \ \underset{\sim}{Q} \ \underset{\sim}{D}_\sigma \ . \qquad (2.5.3)$$

Then, in terms of $\underset{\sim}{Q}^*$, likelihood equations reduce to

$$\text{diag.}(\underset{\sim}{Q}^* \underset{\sim}{\rho}) = \underline{0} \qquad (2.5.4)$$

and

$$\text{tr}(\underset{\sim}{Q}^* \underset{\sim}{H}_t) = 0, \ t = 1(1)k \ . \qquad (2.5.5)$$

Since the diagonal elements of $\underset{\sim}{\rho}$ and $\underset{\sim}{H}_t$ are 1 and 0 respectively, from (2.5.4) and (2.5.5),

$$q^*_{ii} = -\sum_{j=1}^{i-1} \rho_{ij}\ q^*_{ij} - \sum_{j=i+1}^{p} \rho_{ij}\ q^*_{ij} \ , \ \forall i = 1(1)p \qquad (2.5.6)$$

and

$$\sum_{i=2}^{p} \sum_{j=1}^{i-1} h_{t,ij}\ q^*_{ij} = 0 \ , \ \forall t = 1(1)k \qquad (2.5.7)$$

where ρ_{ij} , $h_{t,ij}$ and q^*_{ij} are the typical off-diagonal elements of $\underset{\sim}{\rho}$, $\underset{\sim}{H}_t$ and $\underset{\sim}{Q}^*$ respectively. Similarly, q^*_{ii} is the typical

diagonal element of $\underset{\sim}{Q}^*$. Now, (2.5.6) and (2.5.7) may be treated as the (p+k) linear restrictions upon p(p+1)/2 unknown elements of the symmetric matrix $\underset{\sim}{Q}^*$. Replacing p number of q^*_{ii}'s by (2.5.6) and some k number of q^*_{ij}'s, i > j satisfying (2.5.7), every element of $\underset{\sim}{Q}^*$ may thus be a linear combination of the remaining q^*_{ij}'s (say, q^*_α , $\alpha = 1(1)\upsilon$, where $\upsilon = p(p-1)/2-k$), so that $\underset{\sim}{Q}^*$, under the likelihood equations would be expressed as

$$\underset{\sim}{Q}^* = \sum_{\alpha=1}^{\upsilon} \underset{\sim\alpha}{M}^*_\alpha \, q^*_\alpha \qquad (2.5.8)$$

where q^*_α's are unknown scalars (called " auxiliary parameters"). $\underset{\sim\alpha}{M}^*_\alpha$'s are suitable symmetric matrices whose diagonal elements only may involve correlation parameters (ρ_t , t = 1(1)k), as indicated by (2.5.6).

Following lemma 2.2.1 of subsection 2.2, it is easy to show that $tr(\underset{\sim\alpha}{M}^*_\alpha \underset{\sim}{\rho}) = 0$, $\alpha = 1(1)\upsilon$. From the definition of $\underset{\sim}{Q}^*$, we set the matrix identity

$$\underset{\sim}{\rho} = \underset{\sim}{S}^* + \underset{\sim}{\rho}\,\underset{\sim}{Q}^*\,\underset{\sim}{\rho} \qquad (2.5.9)$$

where $\underset{\sim}{S}^* = \underset{\sim\sigma}{D}^{-1}_\sigma \, \underset{\sim}{S} \, \underset{\sim\sigma}{D}^{-1}_\sigma$. Next we solve for q^*_α's, σ_i's and ρ_t's simulteneously by using trace operations upon (2.5.9) keeping the following obvious trace results in mind.

$$tr(\underset{\sim\alpha}{M}^*_\alpha \, \underset{\sim}{\rho}) = 0, \; tr(\underset{\sim i}{Z}_i \underset{\sim}{\rho}) = 1, \; tr(\underset{\sim t}{H}_t \, \underset{\sim}{\rho}) = \rho_t \, tr(\underset{\sim t}{H}^2_t) \qquad (2.5.10)$$

$\forall \alpha = 1(1)\upsilon$, i = 1(1)p , t = 1(1)k, where $\underset{\sim i}{Z}_i$ denotes a pxp sparse matrix with only i-th diagonal element unity. Inherent iterative scheme is thus

$$(i)\ \underset{\sim}{q}^* = \underset{\sim}{A}^{*-1}\ \underset{\sim}{b}^* \quad (ii)\ \underset{\sim}{D}_{\sigma^2} = \underset{\sim}{D}_{s^2}[\underset{\sim}{I}_p - \mathrm{diag}(\underset{\sim}{\rho}\ \underset{\sim}{Q}^*\underset{\sim}{\rho})]^{-1}$$

$$(iii)\ \underset{\sim}{\rho} = \underset{\sim}{\rho}^{\ell *} + \underset{\sim}{B}^*\ \underset{\sim}{q}^* \qquad (2.5.11)$$

where $\underset{\sim}{q}^* = ((q_\alpha^*))_{\upsilon x 1}$, $\underset{\sim}{D}_{\sigma^2} = \mathrm{diag}(\sigma_1^2 , \sigma_2^2 , \cdots , \sigma_p^2)$,

$\underset{\sim}{D}_{s^2} = \mathrm{diag}(s_{11}, s_{22}, \cdots , s_{pp})$, $\underset{\sim}{\rho} = ((\rho_t))_{kx1}$, $\underset{\sim}{b}^* = ((b_\beta^*))_{\upsilon x 1}$,

$b_\beta^* = -\ \mathrm{tr}(\underset{\sim}{M}_\beta^*\ \underset{\sim}{S}^*) = -\ \mathrm{tr}(\underset{\sim}{D}_\sigma^{-1}\ \underset{\sim}{M}_\beta^*\ \underset{\sim}{D}_\sigma^{-1}\underset{\sim}{S})$, $\underset{\sim}{A}^* = ((a_{\alpha\beta}^*))_{\upsilon x \upsilon}$, $a_{\alpha\beta}^* =$

$\mathrm{tr}(\underset{\sim}{M}_\alpha^*\ \underset{\sim}{\rho}\ \underset{\sim}{M}_\beta^*\ \underset{\sim}{\rho})$, $\underset{\sim}{\rho}^{\ell *} = ((\rho_t^{\ell *}))_{kx1}$, $\rho_t^{\ell *} = \mathrm{tr}(\underset{\sim}{H}_t\underset{\sim}{S}^*)/\mathrm{tr}(\underset{\sim}{H}_t^2) =$

$\mathrm{tr}(\underset{\sim}{D}_\sigma^{-1}\ \underset{\sim}{H}_t\ \underset{\sim}{D}_\sigma^{-1}\ \underset{\sim}{S})/\mathrm{tr}(\underset{\sim}{H}_t^2)$, $\underset{\sim}{B}^* = ((b_{t\alpha}^*))_{kx\upsilon}$, $b_{t\alpha}^* = \mathrm{tr}(\underset{\sim}{H}_t\rho\underset{\sim}{M}_\alpha^*\ \underset{\sim}{\rho})$ and

the notation $\mathrm{diag}(\underset{\sim}{A})$ means the column vector $(a_{11}, a_{22}, \cdots, a_{pp})'$.

Convergence of this iterative scheme (called Q^*-procedure due to involvement of $\underset{\sim}{Q}^*$) depends upon the suitable choice of the initial estimates of the parameters. One workable choice is obvious from the very scheme (2.5.11) which yields the following trial values :

$$\sigma_{i(o)}^2 = s_{ii} \ , \ \rho_{t(o)} = \mathrm{tr}(\underset{\sim}{H}_t\ \underset{\sim}{R})/\mathrm{tr}(\underset{\sim}{H}_t^2) \qquad (2.5.12)$$

$i = 1(1)p$, $t = 1(1)k$, where $\underset{\sim}{R}$ is the sample correlation matrix. [This $\underset{\sim}{R}$ should not be confused with the $\underset{\sim}{R}$ appearing in (2.1.6).]

Iterative scheme may be similarly modified when, in addition, $\underset{\sim}{D}_\sigma$ is structured i.e., when σ_i's are linearly related.

In particular, when $\underset{\sim}{D}_\sigma = \sigma\underset{\sim}{I}_p$ (a scalar matrix), $\sigma > 0$, $\underset{\sim}{\Sigma}$ would have the same structure as $\underset{\sim}{\rho}$ so that Q^*-procedure, in fact

coincides with the Q-procedure.

However, in the present chapter, we are interested only in the situation where the elements of $\underset{\sim}{D}_\sigma$ are functionally independent. In such cases, when additionally the $\underset{\sim}{\rho}$-matrix involves functionally independent non-zero off-diagonal elements, the ML estimates are simple, e.g. $\hat{\underset{\sim}{D}}_\sigma = \underset{\sim}{D}_s = \text{diag}(\sqrt{s_{11}}, \sqrt{s_{22}}, \cdots, \sqrt{s_{pp}})$, and $\hat{\underset{\sim}{\rho}} = \underset{\sim}{R}$. Further, if some of the off-diagonal elements are zero, relevant Q^*-procedure is equivalent to Q-procedure (although comparative iterative behaviour of the two procedures is not known). For example, a correlation matrix having Tri-diagonal Jacobi structure may be iteratively solved by Q-procedure as well as by Q^*-procedure with the same result (vide subsection 2.2.3(c)).

Remark 2.2. In practical situations, scale parameters σ_i, $i = 1(1)p$ are very often replaced by the sample standard deviations $\sqrt{s_{ii}}$, $i = 1(1)p$. In such cases, relevant MLE of the correlation parameters ρ_t, $t = 1(1)k$ may be obtained by using (2.5.11(i)) and (2.5.11(iii)) with considerably less amount of labour.

2.5.1 Some Simple Results based on Q^*-matrix under Likelihood Equations

We shall present here some simple results based on $\underset{\sim}{Q}^*$-matrix when the covariance matrix $\underset{\sim}{\Sigma}$ is expressed as $\underset{\sim}{D}_\sigma \underset{\sim}{\rho} \underset{\sim}{D}_\sigma$ when $\underset{\sim}{\rho}$ is linearly structured but the elements of $\underset{\sim}{D}_\sigma$ are functionally independent.

Proposition 2.5.1 : Under normal equations (2.5.1),

$$\text{(i)}\quad \underline{\text{diag.}(\underset{\sim}{Q}\underset{\sim}{\Sigma}) = \text{diag.}\ (\underset{\sim}{Q}^{*}\ \underset{\sim}{\rho}) = \underline{0}}$$

$$\text{(ii)}\quad \underline{\text{diag.}(\underset{\sim}{\rho}^{-1}\underset{\sim}{S}^{*}) = \text{diag.}(\underset{\sim}{\Sigma}^{-1}\underset{\sim}{S}) = \text{diag}(\underset{\sim}{I}_{p})}.$$

Proof : (i) Pre - and post- multiplying (2.5.4) by non-null diagonal matrix $\underset{\sim}{D}_{\sigma}$, the result may be obtained.

(ii) By definition of $\underset{\sim}{Q}$- matrix (2.1.2),

$$\text{diag.}\ (\underset{\sim}{Q}\underset{\sim}{\Sigma}) = \text{diag.}\ (\underset{\sim}{I}_{p} - \underset{\sim}{\Sigma}^{-1}\underset{\sim}{S})$$

$$= \text{diag.}\ [\underset{\sim}{I}_{p} - \underset{\sim}{D}_{\sigma}^{-1}\ \underset{\sim}{\rho}^{-1}\ \underset{\sim}{D}_{\sigma}^{-1}\ \underset{\sim}{S}]$$

$$= \text{diag.}\ (\underset{\sim}{I}_{p} - \underset{\sim}{\rho}^{-1}\ \underset{\sim}{S}^{*}).$$

Hence, by applying (i) upon it, the result follows.

Proposition 2.5.2 : Under normal equations (2.5.1) and (2.5.2),

$$\text{(i)}\quad \underline{\text{tr}\ \underset{\sim}{Q}^{*} = 0}$$

$$\text{(ii)}\quad \underline{\text{tr}(\underset{\sim}{\rho}^{-1} - \underset{\sim}{\rho}^{-1}\ \underset{\sim}{S}^{*}\ \underset{\sim}{\rho}^{-1})\underset{\sim}{H}_{t} = 0,\ t = 1(1)k.}$$

Proof : (i) Multiplying both sides of (2.5.5) by the scalar ρ_t and adding over t = 1(1)k,

$$\text{tr}\ [\underset{\sim}{Q}\underset{\sim}{D}_{\sigma}(\underset{\sim}{\rho} - \underset{\sim}{I}_{p})\underset{\sim}{D}_{\sigma}] = 0$$

$$\text{i.e.}\quad \text{tr}\ (\underset{\sim}{Q}^{*}) = \text{tr}(\underset{\sim}{Q}^{*}\ \underset{\sim}{\rho})$$

Hence due to Proposition 2.5.1, the desired result follows.

(ii) Since $\underset{\sim}{Q}^{*}$ has another version $\underset{\sim}{\rho}^{-1} - \underset{\sim}{\rho}^{-1}\ \underset{\sim}{S}^{*}\underset{\sim}{\rho}^{-1}$, the result follows from (2.5.5).

Proposition 2.5.3 (i) MLE of $\underset{\sim}{D}_\sigma = \underset{\sim}{D}_s$ when either $\underset{\sim}{\rho} = \underset{\sim}{I}_p$ or ρ has no specific structure as such, i.e., the off-diagonal elements of ρ are arbitrary except symmetry.

(ii) When MLE of $\underset{\sim}{D}_\sigma = \underset{\sim}{D}_s$ holds, $\text{diag}(\underset{\sim}{\rho}^{-1}\ \underset{\sim}{R}) = \text{diag}(\underset{\sim}{I}_p)$ and $\text{tr}(\underset{\sim}{\rho}^{-1} - \underset{\sim}{\rho}^{-1}\underset{\sim}{R}\ \underset{\sim}{\rho}^{-1}) = 0$ where $\underset{\sim}{R}$ is the sample correlation matrix.

Proof : The results are obvious.

Remark 2.3 (i) The above-mentioned propositions are particularly useful for checking purposes when MLE of the parameters are obtained by Q-and Q^*- or any other existing procedure.

(ii) It is obvious that when the elements of $\underset{\sim}{D}_\sigma$ are either primarily known or functionally not dependent, the above mentioned propositions will not hold. Consequently, $\text{tr}(\underset{\sim}{Q}\underset{\sim}{L})$ or $\text{tr}(\underset{\sim}{Q}^*\ \underset{\sim}{\rho})$ will not vanish.

2.5.2 Hessian Matrix and Estimate of SE in Q^*-procedure

Following the steps shown in subsection 2.4.1, we may derive the Hessian matrix in a partitioned form as follows :

$$\underline{\theta} = \begin{bmatrix} \underline{\sigma} \\ \underline{\rho} \end{bmatrix} \quad \text{where} \quad \underline{\sigma} = (\sigma_1\ \sigma_2\ \cdots\ \sigma_p)' \quad \text{and}$$

$\underline{\rho} = (\rho_1\ \rho_2\ \cdots\ \rho_k)'$.

$$\underset{\sim}{J}(\underline{\theta}) = \underset{\sim}{J}_{Q^*}(\underline{\theta}) + \underset{\sim}{\mathcal{J}}(\underline{\theta}) \qquad (2.5.13)$$

where

$$\underset{\sim}{J}_{Q^*}(\underline{\theta}) = N \begin{bmatrix} \underset{\sim}{D}_\sigma^{-1} & \\ & \underset{\sim}{I}_k \end{bmatrix} \begin{bmatrix} \underset{\sim}{V}_{11}(\underset{\sim}{Q}^*) & \underset{\sim}{V}_{12}(\underset{\sim}{Q}^*) \\ \underset{\sim}{V}_{21}(\underset{\sim}{Q}^*) & \underset{\sim}{V}_{22}(\underset{\sim}{Q}^*) \end{bmatrix} \begin{bmatrix} \underset{\sim}{D}_\sigma^{-1} & \\ & \underset{\sim}{I}_k \end{bmatrix} \qquad (2.5.14)$$

$$\underset{\sim}{V}_{11}(\underset{\sim}{Q}^{*}) = 2\ \underset{\sim}{D}_{W}(\underset{\sim}{Q}^{*}) + \underset{\sim}{\rho} * \underset{\sim}{Q}^{*} + \underset{\sim}{\rho}^{-1} *\ (\underset{\sim}{\rho}\underset{\sim}{Q}^{*}\underset{\sim}{\rho})$$

$$\underset{\sim}{V}_{12}(\underset{\sim}{Q}^{*}) = \underset{\sim}{V}'_{21}(\underset{\sim}{Q}^{*}) = [\text{diag}\bullet(\underset{\sim}{H}_{t}\underset{\sim}{Q}^{*}) + \text{diag}\ (\underset{\sim}{\rho}^{-1}\underset{\sim}{H}_{t}\underset{\sim}{Q}^{*}\underset{\sim}{\rho})]_{t=1(1)k}$$

$$\underset{\sim}{V}_{22}(\underset{\sim}{Q}^{*}) = [\ tr(\underset{\sim}{H}_{t}\ \underset{\sim}{\rho}^{-1}\underset{\sim}{H}_{u}\underset{\sim}{Q}^{*})]_{t,u\ =\ 1(1)k} \qquad \text{and}$$

$$\underset{\sim}{D}_{W}\ (\underset{\sim}{Q}^{*}) = \text{diag}\bullet[\ (\underset{\sim}{\rho}\underset{\sim}{Q}^{*})_{i,i}\ ,\ i = 1(1)\ p\ ].$$

Expression of $\underset{\sim}{\mathcal{J}}(\underline{\theta})$ is found notationally similar to $-\frac{1}{2}$ times $\underset{\sim}{J}_{Q^{*}}(\underline{\theta})$ (equ. (2.5.13)) in which $\underset{\sim}{Q}^{*}$ is to be replaced by $\underset{\sim}{\rho}^{-1}$.

We may derive with some algebraic effort the following result :

$$\underline{\theta}'\ \underset{\sim}{J}\ \underline{\theta} = N[9\ tr(\underset{\sim}{Q}^{*}\ \underset{\sim}{\rho}) - 6\ tr\ \underset{\sim}{Q}^{*} + tr(\underset{\sim}{Q}^{*}\ \underset{\sim}{\rho}^{-1})]$$

$$- \frac{N}{2}\ [9p - 6\ tr(\underset{\sim}{\rho}^{-1}) + tr(\underset{\sim}{\rho}^{-2})]\ . \qquad (2.5.15)$$

Under MLE $(\hat{\underline{\theta}})$, $tr(\hat{\underset{\sim}{Q}}^{*}\hat{\underset{\sim}{\rho}}) = tr(\hat{\underset{\sim}{Q}}^{*}) = 0$ so that

$$\hat{\underline{\theta}}'\ \hat{\underset{\sim}{J}}\hat{\underline{\theta}} = N\ tr(\hat{\underset{\sim}{Q}}^{*}\ \hat{\underset{\sim}{\rho}}^{-1}) - \frac{N}{2}\ [9p - 6\ tr(\hat{\underset{\sim}{\rho}}^{-1}) + tr(\hat{\underset{\sim}{\rho}}^{-2})]. \qquad (2.5.16)$$

As stated earlier, before estimating $\text{disp}.(\hat{\underline{\theta}})$, the checking of the Hessian matrix may be done by using

$$\hat{\underline{\theta}}'\ \hat{\underset{\sim}{J}}_{Q}\hat{\underline{\theta}} = N\ tr(\hat{\underset{\sim}{Q}}^{*}\ \hat{\underset{\sim}{\rho}}^{-1}) \qquad (2.5.17)$$

and

$$\hat{\underline{\theta}}'\ \hat{\underset{\sim}{\mathcal{J}}}\ \hat{\underline{\theta}} = -\frac{N}{2}\ [9p - 6\ tr(\hat{\underset{\sim}{\rho}}^{-1}) + tr(\hat{\underset{\sim}{\rho}}^{-2})]. \qquad (2.5.18)$$

$\text{Disp}.(\hat{\underline{\theta}})$ may be estimated either by $\underset{\sim}{E}_{1} = -\ [\underset{\sim}{J}(\hat{\underline{\theta}})]^{-1}$. The standard

error (SE) of $\underset{\sim}{\hat{\theta}}$ is obtained as the square root of the diagonal elements of the estimated disp. ($\underset{\sim}{\hat{\theta}}$).

2.5.3 Illustration of Q^*-procedure Applied to Intraclass Correlation Structure

Let $\underset{\sim}{\rho}$ -matrix involved in $\underset{\sim}{\Sigma} = \underset{\sim}{D}_\sigma \underset{\sim}{\rho} \underset{\sim}{D}_\sigma$, be an intraclass matrix expressed as

$$\underset{\sim}{\rho} = \underset{\sim}{I}_p + \rho(\underset{\sim}{E}_p - \underset{\sim}{I}_p) \ . \qquad (2.5.19)$$

We assume that the elements of $\underset{\sim}{D}_\sigma$, e.g. the scale parameters σ_i , $i = 1(1)p$ are all functionally independent. Here, correlation parameter is only one, namely ρ . Thus in total only (p+1) number of parameters is to be estimated by the proposed Q^*-procedure.

Relevant likelihood equations imply that

$$q^*_{ii} = - \rho \sum_{\substack{j=1 \\ j\neq i}}^{p} q^*_{ij} \ , \ i = 1(1)p \qquad (2.5.20)$$

and

$$\sum_{i=1}^{p-1} \sum_{j=i+1}^{p} q^*_{ij} = 0 \ . \qquad (2.5.21)$$

In effect, $\sum_{i=1}^{p} q^*_{ii} = 0$. (2.5.22)

For definiteness, let us choose $q^*_{p-1,p}$ to be expressed from (2.5.21) as

$$q^*_{p-1,p} = - \sum_{i=1}^{p-2} \sum_{j=i+1}^{p-1} q^*_{ij} \ . \qquad (2.5.23)$$

The Q^*-matrix under the likelihood equations (2.5.20) and (2.5.23) may be expressed similar to (2.5.8) as follows :

$$\underset{\sim}{Q}^* = \sum_{i=1}^{p-2} \sum_{j=i+1}^{p} \underset{\sim}{M}^*_{ij}\, q^*_{ij} \quad , \qquad (2.5.24)$$

where the non-zero elements of the symmetric $\underset{\sim}{M}^*_{ij}$ matrix will follow the following rule :

Cell	(i,i)	(j ,j)	(i,j)	(p-1,p-1)	(p,p)	(p-1,p)
element	$-\rho$	$-\rho$	1	ρ	ρ	-1

In case the subscript (ij) of $\underset{\sim}{M}^*_{ij}$-matrix be such that certain non-zero cell of $\underset{\sim}{M}^*_{ij}$ by the above rule happens to contain both ρ and $-\rho$ simulteneously, then that cell would be filled up by 0 . For example, $\underset{\sim}{M}^*_{1,p-1}$ would have the (p-1,p-1) the element equal to 0 so that non-zero elements would be available only at four cells namely (1,1), (1,p-1), (p,p) and (p-1,p) and the corresponding elements will be $-\rho$,1, ρ and -1 respectively, apart from the symmetric property of $\underset{\sim}{M}^*_{ij}$-matrix.

From such a pattern of the elements of $\underset{\sim}{M}^*_{ij}$-matrices, it is evident that for any matrix $\underset{\sim}{P}$ having intraclass pattern, $tr(\underset{\sim}{P}\underset{\sim}{M}^*_{ij}) = 0$. Since $\underset{\sim}{\rho}\underset{\sim}{E}_p\underset{\sim}{\rho}$ and $\underset{\sim}{\rho}^2$ would have intraclass pattern, $tr(\underset{\sim}{E}_p\underset{\sim}{\rho}\underset{\sim}{M}^*_{ij}\underset{\sim}{\rho}) = tr(\underset{\sim}{\rho}\ \underset{\sim}{M}^*_{ij}\ \underset{\sim}{\rho}) = 0$. Thus, for the only design matrix $H_1 (= \underset{\sim}{E}_p - \underset{\sim}{I}_p)$ in connection with the single correlation parameter ρ , $tr(\underset{\sim}{H}_1\underset{\sim}{\rho}\ \underset{\sim}{M}^*_{ij}\underset{\sim}{\rho}) = 0$. Hence, all the elements of $\underset{\sim}{B}^*$-matrix [which is here a $\upsilon \times 1$ column vector, $\upsilon = \frac{p(p-1)}{2} - 1$,

vide equation (2.5.11)] would vanish. We therefore obtain the desired result

$$\rho = \rho^{\ell *} = \sum\sum_{i \neq j} s^*_{ij}/p(p-1) \ . \qquad (2.5.25)$$

By recalling (ii) of (2.5.11), we then find that

$$(\underset{\sim}{\rho}\underset{\sim}{Q}^*\underset{\sim}{\rho})_{ii} = -(1-\rho)^2 \ q^*_{ii} \qquad (2.5.26)$$

so that

$$\sigma_i^2 = s_{ii}/[1-(1-\rho)^2 \ q^*_{ii}] \ , \qquad (2.5.27)$$

$i = 1(1)p$. Once q^*_{ij}'s are obtained by using (1) of (2.5.11), q^*_{ii} involved in (2.5.27) may be calculated using (2.5.20). Thus, iterative scheme would be based on (2.5.25), (2.5.11) and (2.5.27) to obtain ρ, q^*_{ij}'s and σ_i's in sequence. For convenience, initial estimates are taken as $\rho_{(o)} = \bar{h} = \sum\sum_{i \neq j} r_{ij}/p(p-1)$ and $\sigma_{i(o)} = \sqrt{s_{ii}}$, $i = 1(1)p$ where $r_{ij} = s_{ij}/\sqrt{s_{ii}s_{jj}}$, $i \neq j$.

In case $\underset{\sim}{D}_\sigma$ is known, the 2nd set of equations of (2.5.11) does not arise. MLE of only parameter ρ is to be solved iteratively from

$$\rho = \rho^*_{\ell} + \frac{\rho\{2+(p-2)\rho\}}{1+(p-1)\rho^2}(1-s^*_{\ell}) \ , \qquad (2.5.28)$$

where $\rho^*_{\ell} = \sum\sum_{i \neq j} s^*_{ij}/p(p-1)$, $s^*_{\ell} = \sum_{i=1}^{p} s^*_{ii}/p$. We may observe in this case that

$$tr(\underset{\sim}{\hat{Q}}\ \underset{\sim}{\hat{\Sigma}}) = tr(\underset{\sim}{\hat{Q}}^*\ \underset{\sim}{\hat{\rho}}) = tr(\underset{\sim}{\hat{Q}}^*) = \frac{p(1-s^*_{\ell})}{1+(p-1)\hat{\rho}^2}$$

which does not vanish unless $s^*_{\ell} = 1$.

For the purpose of numerical illustration, let us consider the case p = 3. Data are taken from Jolicoeur and Mosimann (1960) who investigated the principal components of carapace length, width and height of painted turtles in an effort to give meanings to the concept of size and shape. The covariance (with correlation) matrix ($\underset{\sim}{S}$) based on the length, width and height measurements in millimeters collected from the carapaces of 24 female turtles is reproduced below.

$$\underset{\sim}{S} = \begin{bmatrix} 451.39 & .973963 & .972611 \\ 271.17 & 171.73 & .965462 \\ 168.70 & 103.29 & 66.65 \end{bmatrix} .$$

For $p = 3$, we have

$$\underset{\sim}{Q}^{*} = \underset{\sim}{M}^{*}_{12}\, q^{*}_{12} + \underset{\sim}{M}^{*}_{13}\, q^{*}_{13}$$

where $\underset{\sim}{M}^{*}_{12} = \begin{bmatrix} -\rho & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & \rho \end{bmatrix}$ and $\underset{\sim}{M}^{*}_{13} = \begin{bmatrix} -\rho & 0 & 1 \\ 0 & \rho & -1 \\ 1 & -1 & 0 \end{bmatrix}$.

$$\underset{\sim}{A}^{*} = \begin{bmatrix} tr(\underset{\sim}{M}^{*}_{12}\underset{\sim}{\rho}\underset{\sim}{M}^{*}_{12}\underset{\sim}{\rho}) & tr(\underset{\sim}{M}^{*}_{13}\underset{\sim}{\rho}\underset{\sim}{M}^{*}_{12}\underset{\sim}{\rho}) \\ tr(\underset{\sim}{M}^{*}_{12}\underset{\sim}{\rho}\underset{\sim}{M}^{*}_{13}\underset{\sim}{\rho}) & tr(\underset{\sim}{M}^{*}_{13}\underset{\sim}{\rho}\underset{\sim}{M}^{*}_{13}\underset{\sim}{\rho}) \end{bmatrix}$$

$$= (1-\rho)^2 \left\{3 - (1-\rho)^2\right\} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} ,$$

$$\underline{b}^{*} = -[\, tr(\underset{\sim}{M}^{*}_{12}\underset{\sim}{S}^{*}),\ \ tr(\underset{\sim}{M}^{*}_{13}\underset{\sim}{S}^{*})]'$$

$$= [\rho(s^{*}_{11}-s^{*}_{33})-2(s^{*}_{12}-s^{*}_{23}), \rho(s^{*}_{11}-s^{*}_{22})-2(s^{*}_{13}-s^{*}_{23})]' ,$$

and

$$\underset{\sim}{q}^{*} = (q_{12}, q_{13})' = \underset{\sim}{A}^{*-1}\ \underset{\sim}{b}^{*} .$$

Applying the iterative scheme (2.5.11), we obtain the MLE of the parameters correct to 5 places of decimal by only 3 iterations as follows :

$$\hat{\sigma}_1 = 21.210203,\ \hat{\sigma}_2 = 13.112760,\ \hat{\sigma}_3 = 8.172634,$$

$$\hat{\rho} = 0.970681 \text{ as well as } \hat{q}^{*}_{12} = -2.547032 \text{ and}$$

$$\hat{q}^{*}_{13} = -1.495083.$$

For verification purpose, we calculate the following :

$$\text{diag.}(\hat{\underset{\sim}{Q}}^{*}\hat{\underset{\sim}{\rho}}) = [.000009,\ -.000004,\ -.000009]' \approx [0,0,0]'$$

which satisfies Proposition 2.5.1(i) where Q^{*} is calculated as

$$\hat{\underset{\sim}{Q}}^{*} = \hat{\underset{\sim}{\rho}}^{-1} - \hat{\underset{\sim}{\rho}}^{-1}\underset{\sim}{S}^{*}\hat{\underset{\sim}{\rho}}^{-1} = \begin{bmatrix} 3.923563 & & \\ -2.547032 & -1.451244 & \text{Sym.} \\ -1.495031 & 4.042111 & -2.472404 \end{bmatrix}$$

with $\text{tr}\ \hat{\underset{\sim}{Q}}^{*} = -.000085 \approx 0$, and $\sum\limits_{i \neq j}\sum \hat{q}^{*}_{ij} = .000096 \approx 0$

satisfying (2.5.21) and (2.5.22).

Since we need the estimated SE of $\hat{\underline{\theta}}$, we have to calculate the Hessian matrix (at $\underline{\theta} = \hat{\underline{\theta}}$) with the following pair of matrices :

$$\underset{\sim}{J}_{Q*}(\underline{\hat{\theta}}) = N\begin{bmatrix} .0086 & & & \\ -.0087 & -.0083 & \text{Sym.} & \\ -.0082 & .0358 & -.0363 & \\ -.1925 & .1151 & .3148 & -.0030 \end{bmatrix}$$

and

$$\underset{\sim}{\mathcal{J}}(\underline{\hat{\theta}}) = -\frac{N}{2}\begin{bmatrix} .0508 & & & \\ -.0393 & .1329 & \text{Sym.} & \\ -.0630 & -.1019 & .3421 & \\ -2.1227 & -3.4336 & -5.5090 & 2326.6531 \end{bmatrix}.$$

As a step in the computational checking of these two matrices, we calculate

$$\underline{\hat{\theta}}' \underset{\sim}{J}_{Q*}(\underline{\hat{\theta}})\underline{\hat{\theta}} = N(-.0026), \quad \underline{\hat{\theta}}' \underset{\sim}{\mathcal{J}}(\underline{\hat{\theta}})\underline{\hat{\theta}} = -\frac{N}{2}(1933.2201).$$

By (2.5.17), $\underline{\hat{\theta}}' \underset{\sim}{J}_{Q*}(\hat{\theta})\underline{\hat{\theta}}$ is found out numerically as $N(-.0030)$ (although algebraically it should be exactly zero for the intraclass matrix $\underset{\sim}{\hat{\rho}}^{-1}$). Similarly, by (2.5.18), $\underline{\hat{\theta}}'\underset{\sim}{\mathcal{J}}(\underline{\hat{\theta}})\underline{\hat{\theta}}$ is calculated as $-\frac{N}{2}(1942.3223)$. Granting that slight variations so far found in the calculated figures are purely computational (probably due to an effect of the estimated $\hat{\rho}$ close to unity), we have the estimated Hessian matrix (as the sum of $\underset{\sim}{J}_{Q*}(\underline{\hat{\theta}})$ and $\underset{\sim}{\mathcal{J}}(\underline{\hat{\theta}})$). Using these matrices, the SEs of the estimated parameters $\hat{\sigma}_1$, $\hat{\sigma}_2$, $\hat{\sigma}_3$ and $\hat{\rho}$ are calculated as follows :

$$SE(\underline{\hat{\theta}}) = (2.4425, 1.6056, 1.0073, 0.0046) \text{ by using } \underset{\sim}{E}_1 .$$

$$= (3.0341, 1.8759, 1.1688, 0.0054) \text{ by using } \underset{\sim}{E}_2 .$$

2.6 GENERALIZED LEAST SQUARES ESTIMATE

To obtain the generalized least squares estimate (GLSE) of $\underline{\theta}$ involved in $\Sigma(\underline{\theta})$, the procedure is to minimize the (weighted) residual quadratic form

$$G(\underset{\sim}{S},\underset{\sim}{\Sigma}) = \frac{1}{2}\ tr\left[(\underset{\sim}{\Sigma}-\underset{\sim}{S})\underset{\sim}{V}\right]^2 \qquad (2.6.1)$$

with respect to $\underline{\theta}$, where $\underset{\sim}{V}$ is a suitable p.d. matrix (called weight matrix). Once we are interested in the linearly structured covariance matrix (2.1.1), relevant normal equations would have simple expressions as

$$tr[(\underset{\sim}{\Sigma}-\underset{\sim}{S})\underset{\sim}{V}\underset{\sim}{H}_i\underset{\sim}{V}] = 0,\ i = 1(1)k\ . \qquad (2.6.2)$$

The choice of $\underset{\sim}{V}$ is important since the relevant estimate depends upon $\underset{\sim}{V}$. Following Browne (1974), $\underset{\sim}{V}$ is either a p.d. constant matrix ($\bar{\underset{\sim}{V}}$, say) or a stochastic matrix converging in probability to p.d. matrix $\bar{\underset{\sim}{V}}$ as $N \longrightarrow \infty$. In such cases, GLS estimate ($\hat{\underline{\theta}}$) is consistent and asymptotically normal with $E(\hat{\underline{\theta}}) = \underline{\theta}_0$ (true value of $\underline{\theta}$) and the covariance matrix

$$Disp.(\hat{\underline{\theta}}) = \frac{2}{N}\left[\underset{\sim}{\Gamma}(\bar{\underset{\sim}{V}})\right]^{-1}\left[\underset{\sim}{\Gamma}(\bar{\underset{\sim}{V}}\ \underset{\sim}{\Sigma}_0\bar{\underset{\sim}{V}})\right]\left[\underset{\sim}{\Gamma}(\bar{\underset{\sim}{V}})\right]^{-1} \qquad (2.6.3)$$

where we use the notation $\underset{\sim}{\Gamma}(\bar{\underset{\sim}{V}}) = \underset{\sim}{\Lambda}'(\bar{\underset{\sim}{V}} \otimes \bar{\underset{\sim}{V}})\underset{\sim}{\Lambda}$, $\underset{\sim}{\Sigma}_0 = \underset{\sim}{\Sigma}(\underline{\theta}_0)$ and $\underset{\sim}{\Lambda}(p^2 \times k) = (Vec\ \underset{\sim}{H}_1\ ,\ Vec\underset{\sim}{H}_2,\ \cdots\ ,\ Vec\underset{\sim}{H}_k)$. Furthermore, in the Loewner sense (i.e., $\underset{\sim}{A} \geq \underset{\sim}{B}$ if $\underset{\sim}{A}-\underset{\sim}{B}$ is positive semi-definite), $\hat{\underline{\theta}}$ would be called the Best-GLSE(BGLSE) of $\underline{\theta}$ attaining lower bound dispersion matrix $\frac{1}{N}\left[\underset{\sim}{\Gamma}(\underset{\sim}{\Sigma}_0^{-1})\right]^{-1}$ asymptotically when $\bar{\underset{\sim}{V}}$ is a scalar multiple of $\underset{\sim}{\Sigma}_0^{-1}$ (Browne, 1974).

Lee (1977) showed that the ML estimator and GLS estimator are asymptotically equal. Fixing $\underset{\sim}{V} = \underset{\sim}{\Sigma}_o^{-1}$, $\underset{\sim}{V} = \underset{\sim}{S}^{-1}$ converges in probability to $\underset{\sim}{\Sigma}_o^{-1}$ as $N \longrightarrow \infty$. Thus by choosing $\underset{\sim}{V} = \underset{\sim}{S}^{-1}$ or $\underset{\sim}{V} = \underset{\sim}{\Sigma}^{-1}$, we may be able to get BGLSE of $\underline{\theta}$.

For $\underset{\sim}{V} = \underset{\sim}{\Sigma}^{-1}$, from (2.6.2) we get

$$tr(\underset{\sim}{Q}\ \underset{\sim}{H}_i) = 0,\ i = 1(1)k \qquad (2.6.4)$$

with $\underset{\sim}{Q} = \underset{\sim}{\Sigma}^{-1} - \underset{\sim}{\Sigma}^{-1}\underset{\sim}{S}\underset{\sim}{\Sigma}^{-1}$.

These are, indeed, equivalent to likelihood equations under multinormal set-up (vide section 2.2) and hence we use Q-procedure to solve (2.6.4).

For $V = S^{-1}$, from (2.6.2)

$$tr(\underset{\sim}{R}\ \underset{\sim}{H}_i) = 0,\ \ i = 1(1)k \qquad (2.6.5)$$

with $\underset{\sim}{R} = \underset{\sim}{S}^{-1} - \underset{\sim}{S}^{-1}\underset{\sim}{\Sigma}\underset{\sim}{S}^{-1}$. (2.6.6)

Solution of $\underline{\theta}$ to the system (2.6.5) is known as <u>special GLSE</u> of $\underline{\theta}$. Involvement of $\underset{\sim}{R}$-matrix (2.6.6) to obtain such special GLSE may be henceforth called "R-procedure" analogous to Q-procedure.

Since $\underset{\sim}{\Sigma}(\underline{\theta})$ has a linear structure, special GLS solution ($\underline{\hat{\theta}}_R$, say) may be obtained by a single step from (2.6.5) as follows :

$$\underline{\hat{\theta}}_R = \underset{\sim}{A}_R^{-1}\ \underline{b}_R \qquad (2.6.7)$$

where (i,j) element of $\underset{\sim}{A}_R$ (k x k) = $tr(\underset{\sim}{S}^{-1}\underset{\sim}{H}_i\underset{\sim}{S}^{-1}\underset{\sim}{H}_j)$ =

$(\text{Vec}\underset{\sim}{H}_i)'(\underset{\sim}{S}^{-1} \otimes \underset{\sim}{S}^{-1})(\text{Vec}\underset{\sim}{H}_j)$ and i-th element of $\underline{b}_R$ (k x 1)

$= \text{tr}(\underset{\sim}{S}^{-1}\underset{\sim}{H}_i) = (\text{Vec}\underset{\sim}{H}_i)'(\underset{\sim}{S}^{-1} \otimes \underset{\sim}{S}^{-1})(\text{Vec}\underset{\sim}{S})$, i,j = 1, ..., k .

Due to such simplicity in obtaining GLS solution by considering $\underset{\sim}{V} = \underset{\sim}{S}^{-1}$ (which is also Best-GLSE), we shall hence-forth call them simply the GLSE obtained by R-method.

We further note that for linear structure (2.1.1),

$$\frac{\partial^2 G}{\partial\theta_i \partial\theta_j} = \text{tr}(\underset{\sim}{V}\underset{\sim}{H}_i \underset{\sim}{V}\underset{\sim}{H}_j), \quad i,j = 1(1)k. \qquad (2.6.8)$$

In GLS method, $\frac{2}{N}$ times the inverse of the matrix formed by (2.6.8) provides an estimate of the asymptotic dis-persion matrix $\frac{2}{N}[\underset{\sim}{\Gamma}(\underset{\sim}{\Sigma}_o^{-1})]^{-1}$. The square root of the diagonal elements so obtained would be the large sample estimate of the SE of the estimated parameters. In R-method, the elements of (2.6.8) are the elements of $\underset{\sim}{A}_R$ as defined in (2.6.7) so that the estimate of SE of such GLSE is simply the square root of $\frac{2}{N}$ times the diagonal element of $\underset{\sim}{A}_R$.

2.6.1 Q-procedure in GLS Estimation

It may be thought from (2.6.4) that in the case of linear structure of $\underset{\sim}{\Sigma}(\underline{\theta})$, what is MLE of $\underline{\theta}$ in multinormal set-up, is equivalent to GLSE with $\underset{\sim}{V} = \underset{\sim}{\Sigma}^{-1}$. We shall prove the following in this context.

Result 2.6.1 For a $\underset{\sim}{\Sigma}(\underline{\theta})$ obeying linear structure (2.1.1), its GLSE with $\underset{\sim}{V} = \underset{\sim}{\hat{\Sigma}}_Q^{-1}$ is the same as its GLSE with $\underset{\sim}{V} = \underset{\sim}{\Sigma}^{-1}$, the common value being $\underset{\sim}{\hat{\Sigma}}_Q$ where $\underset{\sim}{\hat{\Sigma}}_Q$ is the MLE of $\underset{\sim}{\Sigma}$.

[subscript 'Q' is attached to $\hat{\Sigma}$ to mean the solution of normal equations involving $\underset{\sim}{Q}$-matrix].

Proof : It is sufficient to show by using (2.6.2) that $\underset{\sim}{\Sigma} = \underset{\sim}{\Sigma}_Q$ is the solution to the system

$$tr[\hat{\underset{\sim}{\Sigma}}_Q^{-1} \underset{\sim}{H}_i \hat{\underset{\sim}{\Sigma}}_Q^{-1}(\underset{\sim}{\Sigma}-\underset{\sim}{S})] = 0, \; i = 1(1)k \qquad (2.6.9)$$

whenever $\hat{\underset{\sim}{\Sigma}}_Q$ is so to the system (2.2.3), i.e., whenever

$$tr[\hat{\underset{\sim}{\Sigma}}_Q^{-1} \underset{\sim}{H}_i \hat{\underset{\sim}{\Sigma}}_Q^{-1} (\hat{\underset{\sim}{\Sigma}}_Q-\underset{\sim}{S})] = 0, \; i = 1(1)k. \qquad (2.6.10)$$

Subtracting (2.6.9) from (2.6.10), we obtain

$$tr[\hat{\underset{\sim}{\Sigma}}_Q^{-1} \underset{\sim}{H}_i \hat{\underset{\sim}{\Sigma}}_Q^{-1}(\hat{\underset{\sim}{\Sigma}}_Q - \underset{\sim}{\Sigma})] = 0, \; i = 1(1)k. \qquad (2.6.11)$$

Multiplying (2.6.11) by θ_i and $\hat{\theta}_{Qi}$ separately and adding over $i = 1, \ldots, k$, we get

$$tr(\hat{\underset{\sim}{\Sigma}}_Q^{-1} \underset{\sim}{\Sigma})^2 = tr(\hat{\underset{\sim}{\Sigma}}_Q^{-1} \underset{\sim}{\Sigma}) = p \; . \qquad (2.6.12)$$

Defining $\hat{\underset{\sim}{\Sigma}}_Q^{-1} = \underset{\sim}{A}^2$ and $\underset{\sim}{\Sigma} = \underset{\sim}{B}^2$ where $\underset{\sim}{A}$ and $\underset{\sim}{B}$ are square root matrices of $\hat{\underset{\sim}{\Sigma}}_Q^{-1}$ and $\underset{\sim}{\Sigma}$ respectively, we find that $tr(\hat{\underset{\sim}{\Sigma}}_Q^{-1}\underset{\sim}{\Sigma})=tr(\underset{\sim}{C}\underset{\sim}{C}')$ and $tr(\hat{\underset{\sim}{\Sigma}}_Q^{-1}\underset{\sim}{\Sigma})^2 = tr(\underset{\sim}{C}\underset{\sim}{C}')^2$ where $\underset{\sim}{C} = \underset{\sim}{A}\underset{\sim}{B}$. Hence due to (2.6.12), $\underset{\sim}{C}\underset{\sim}{C}'$ has all its eigenvalues equal to unity. Eigenvalue unity has thus a multiplicity of p giving rise to p linearly independent eigenvectors creating a nonsingular matrix $\underset{\sim}{P}$, say [cf. Bellman (1974), p. 54] so that from $\underset{\sim}{C}\underset{\sim}{C}'\underset{\sim}{P} = \underset{\sim}{P}$, we easily obtain $\underset{\sim}{\Sigma} = \hat{\underset{\sim}{\Sigma}}_Q$.

Remark 2.4 It is technically not possible to get linearly

independent eigenvectors for any nonsymmetric matrix corresponding to its multiple eigenvalues [cf. Bellman, 1974, p.199]. ■

In non-normal set-up of the population, or more specifically, when the population distribution of the vector variate $\underset{\sim}{X}(p \times 1)$ is unknown, parameters of its linearly structured dispersion matrix $\underset{\sim}{\Sigma}(\underline{\theta})$ may be estimated by GLSE with $\underset{\sim}{\Sigma}^{-1}(\underline{\theta})$ as weight matrix with the help of Q-procedure. Large sample standard error may be obtained in the same way as mentioned earlier in the paragraph following equation (2.6.8).

We shall mention here that for $\underset{\sim}{V} = \underset{\sim}{I}_p$, GLS solution leads to what is called OLSE or LSE (ordinary or unweighted least squares estimate) of $\underline{\theta}$. The estimates obtained by Q- and R-procedure are scale-free in the sense that relevant $\underset{\sim}{G}$'s [vide (2.6.1)], say $G_Q(\underset{\sim}{S},\underset{\sim}{\Sigma})$ and $G_R(\underset{\sim}{S},\underset{\sim}{\Sigma})$, equal $G_Q(\underset{\sim}{D}\underset{\sim}{S}\underset{\sim}{D},\underset{\sim}{D}\underset{\sim}{\Sigma}\underset{\sim}{D})$ and $G_R(\underset{\sim}{D}\underset{\sim}{S}\underset{\sim}{D},\underset{\sim}{D}\underset{\sim}{\Sigma}\underset{\sim}{D})$, for any diagonal matrix $\underset{\sim}{D}$ of positive scale factors. OLSE does not have this property. Since variances and covariances in S are generally correlated and have unequal variances, it would seem that OLSE uses wrong metric in measuring deviations between $\underset{\sim}{S}$ and $\underset{\sim}{\Sigma}$. Nevertheless, OLS procedure produces consistent estimators under more general assumptions than those assumed in GLS and ML procedures. No statistical theory, however, is available for computing standard errors for OLS estimators [Jöreskog, 1978].

Mukherjee and Maiti (1989d) have obtained the general condition upon the weight matrix ($\underset{\sim}{V}$) for making GLSE equivalent to OLSE. It is also confirmed that in spectral class of linearly structured $\underset{\sim}{\Sigma}$, the estimate obtained by Q-procedure (i.e., GLSE with $\underset{\sim}{V} = \underset{\sim}{\Sigma}^{-1}$ or MLE) is equivalent to OLSE.

2.6.2 R-procedure to Estimate Toeplitz Parameters from Kodak Data

For the analysis of Kodak data of Jackson and Morris (1957), we wish to get the GLSE of the Toeplitz covariance matrix on the basis of the following sample covariance matrix based on 108 observations.

$$\underset{\sim}{S} = \begin{bmatrix} 158 & 102 & 42 \\ 102 & 137 & 96 \\ 42 & 96 & 128 \end{bmatrix}$$

Let us recall Toeplitz covariance matrix as expressed in the subsection 2.4. According to the R-procedure, we need to solve a,b and c from the normal equations $tr(\underset{\sim}{H}_i \underset{\sim}{R}) = 0$, $i = 1,2,3$. We recall (2.6.7) and calculate

$$\underset{\sim}{A}_R = \begin{bmatrix} .0030486 & -.0040616 & .0011153 \\ -.0040616 & .0057055 & -.0017143 \\ .0011153 & -.0017143 & .0006913 \end{bmatrix}$$

and $\underline{b}_R = (.0674497, -.0706320, .0143264)'$.

$\hat{\underline{\theta}}_R = (\hat{a}_R, \hat{b}_R, \hat{c}_R)' = (138.0116, 98.8329, 43.1519)'$.

An estimate of the asymp. disp. matrix of $\hat{\underline{\theta}}_R$ is simply $\frac{2}{N} \underset{\sim}{A}_R^{-1}$ from which

$$SE\,[\hat{a}_R,\ \hat{b}_R,\ \hat{c}_R] = (13.99,\ 12.97,\ 13.22).$$

Additionally, we may apply Q-procedure to get GLSE with $\underset{\sim}{V} = \underset{\sim}{\Sigma}^{-1}$ as

$$\underline{\hat{\theta}}_Q = (\hat{a}_Q,\ \hat{b}_Q,\ \hat{c}_Q)' = (142.5646,\ 101.7946,\ 44.2632)'.$$

We calculate $\underset{\sim}{A}_Q = ((tr(\underset{\sim}{H}_i\hat{\underset{\sim}{\Sigma}}_Q^{-1}\ \underset{\sim}{H}_j\hat{\underset{\sim}{\Sigma}}_Q^{-1}))_{3x3}$ as

$$\underset{\sim}{A}_Q = \begin{bmatrix} .0028707 & -.0038390 & .0010715 \\ -.0038390 & .0054223 & -.0016616 \\ .0010715 & -.0016616 & .0006852 \end{bmatrix}$$

so as to get an estimate of the disp. matrix of $\hat{\underline{\theta}}_Q$ given by $\frac{2}{N}\underset{\sim}{A}_Q^{-1}$ and hence

$$SE\,[\hat{a}_Q,\ \hat{b}_Q,\ \hat{c}_Q] = (14.33,\ 13.27,\ 13.34)$$

We note here that such quantities may be obtained otherwise as usual from the expression of $\underset{\sim}{E}_2$ involved in subsection 2.4.1.

2.7 DISCUSSION

This chapter presents analytically a simple iterative procedure, called Q-procedure, for obtaining the MLE of any covariance matrix having a linear structure. The main difference of this method with the one suggested by Anderson (1973) results from a reformulation of the likelihood equation using the linear structure of Q-matrix so as not to use $\underset{\sim}{\Sigma}^{-1}$ in computation. Anderson's algorithm calls for inversion of two matrices, one (p x p) and another (k x k) whereas Q-procedure

requires only one inversion of a (υ x υ) matrix, where $\upsilon = \binom{p+1}{2} - k$. Thus, as k increases, the computational efficiency of Q-procedure becomes more and more obvious. Anderson's procedure has apparently a tendency of slower rate of convergence.

We have obtained the MLE of a quasi-simplex structure hypothesized for a sample covariance matrix obtained by Bilodeau (1961). This requires only 10 iterations by Q-procedure to get the converged solution keeping the approximation error equal to 10^{-7}, while Anderson's procedure does not give converged solution even after 55 iterations keeping the same stopping rule.

The Q-procedure seems to be quite efficient in locating the regions of convergence very quickly. Although there is a theoretical possibility of obtaining several sets of solution, with empirical data, we still have not found a single example of more than one set of solutions for any linearly structured $\underset{\sim}{\Sigma}$. Further, the Q-procedure does not necessarily guarantee that a maximum of log-likelihood (2.2.1) will be always found for any p.d. $\underset{\sim}{S}$. But still, we have yet to find out a ML solution that would be outside the positive definite region and thereby inadmissible.

The Q^{*}-and R-procedures are two variants of Q-procedure which are useful respectively in the case of linearly structured correlation matrix, and GLS procedure with weight matrix $\underset{\sim}{S}^{-1}$ for

estimating the parameters of linearly structured covariance matrix. Compared to OLS and GLS methods of estimation, ML procedure is very often preferred mainly due to asymptotic efficiency of the estimates (so obtained) and the associated likelihood ratio test. A GLS estimator converges stochastically to the ML estimators and has the same asymptotic properties as possessed by ML estimators [Anderson, 1973].

Jöreskog and Goldberger (1972) show that the weighted residual quadratic form (2.6.1) when $\underset{\sim}{V} = \underset{\sim}{S}^{-1}$ and $\underset{\sim}{\Sigma}$ is replaced by $\hat{\underset{\sim}{\Sigma}}_R$ converges stochastically to the likelihood ratio statistic. Browne (1977) has also provided GLSE of population parameters and an asymptotic chi-square test for the situation where sets of elements of a population correlation matrix are equal and have specific values.

When nothing is known about the distribution of $\underset{\sim}{S}$, the GLS estimate has a natural appeal. But when $\underset{\sim}{S}$ can be assumed to follow a Wishart distribution, maximum likelihood is a logical approach (Swain, 1975; Browne, 1977, p. 124). Although even for the MLE nothing is known about the small-scale properties, we know the theory that if an efficient estimator for small sample exists, which has the minimum variance, the MLE adjusted for bias, if necessary, will be efficient. Q-procedure may, as a consequence, be found useful whenever we need the MLE in preference to R-procedure employed for GLSE.

2.8 Summary

For obtaining the MLE of unknown parameters of any linearly structured covariance, a new procedure, called Q-procedure, which is iterative in nature has been developed. This procedure seems to have the advantage of rapid convergence, of robustness against bad starting values and of not requiring the computation of either $\underset{\sim}{\Sigma}^{-1}$ or the Hessian matrix. The algorithm involved in it has been compared to Anderson's (1973) algorithm, both algebraically and numerically. The Q-procedure is extended to the case of linearly structured correlation matrix ($\underset{\sim}{\rho}$) involved in the covariance matrix which is decomposable as $\underset{\sim}{\Sigma} = \underset{\sim}{D}_{\sigma} \underset{\sim}{\rho} \underset{\sim}{D}_{\sigma}$. Analogous to Q-procedure, the R-procedure has been developed here and illustrated to obtain the GLSE of a linearly structured $\underset{\sim}{\Sigma}$.

CHAPTER 3

LARGE SAMPLE TESTS OF SIGNIFICANCE AND GOODNESS-OF-FIT MEASURES IN ACOVS

3.1 INTRODUCTION

Apart from the problem of parameter estimation, analysis of covariance structures (ACOVS) deals with the problem of developing different tests of significance and measures of goodness-of-fit of the specified covariance matrix ($\underset{\sim}{\Sigma}$) with the corresponding sample covariance matrix ($\underset{\sim}{S}$) .

Statistical techniques which are primarily concerned with testing of hypotheses regarding the pattern of covariance matrix, have been considered by many authors such as those whose names appeared in the second section of Chap. 2. To evaluate the empirical support for a structural hypothesis on $\underset{\sim}{\Sigma}$ via observed covariance matrix ($\underset{\sim}{S}$) is still a basic problem to researchers in ACOVS. In applied work any "well-specified" model still contains specification errors which, despite being small or irrelevant with respect to a substantive theory, can significantly contribute to the value of the test statistic when the sample size is quite large. On the otherhand, if the sample size is very small, large specification errors (relevant to the substantive theory) may not be detected by the test statistic (Saris, Pijper and

Zegwaart, 1979). This leads to the empirical fact that there is a tendency to reject well-specified models if the sample size is very large and to accept wrongly specified models when the sample size is very small.

The afore-said problem has led several authors to develop various measures of goodness-of-fit that are explicitly not dependent upon the sample size (Tanaka and Lewis, 1973; Bentler and Bonett, 1980; Jöreskog and Sörbom, 1981, 1984). Most of these measures are discussed in section 3.3.

In this chapter, we shall summarize some matrix results following Anderson (1969) and report some simple results in connection with the "inverse residual matrix" ($Q = \Sigma^{-1} - \Sigma^{-1} S \Sigma^{-1}$) as defined in Chap. 2. In addition to the likelihood ratio test (LR) statistic, we then propose two new test statistics— one of which may be treated as the ACOVS version of Rao's (1947) efficient score criterion simplified under certain assumptions while the other is based on Wald's (1943) quadratic form criterion. A numerical example based on Basic Airmen's data (Bilodeau, 1960) is provided to use the test statistic for testing a tridiagonal structure of Σ. We then report the results of a Monte Carlo study aimed at comparing the power of the proposed test statistics for testing the null hypothesis that $\Sigma = I_p$ against an alternative hypothesis of intraclass structure with a

common variance unity and the common correlation ρ having different admissible values.

Covering all the above-mentioned results in section 3.2, we review, in brief, the existing goodness-of-fit measures in section 3.3 and discuss some of the distributional properties of a very frequently employed goodness-of-fit index, namely, the Jöreskog-Sörbom fit index (GFIQ). We then propose two new indices based on the simultaneous use of ML and GLS estimates. One illustrative example based on GRE repeaters' data (Werts et al., 1981) is then cited. The results of a Monte-Carlo study under the same set-up as described in section 3.2, are reported to show the relative performance of GFIQ, Tanaka and Huba's (1985) fit index (GFIR) and the proposed indices. Finally, in section 3.4 some concluding remarks are given with respect to specification search, that is, a process of sequentially modifying a covariance structure model so as to improve its fit and/or parsimony. Such specification search is very important (MacCallum, 1986) because many a times, the decision based both in the results of tests of statistical significance and goodness-of-fit might encourage the rejection of a well formulated hypothesized model. Thus, there must be some further efforts in improving and reformulating the model. Sörbom (1989) has suggested a " modification index'' for this purpose which would serve as a guide in the search for

a "better" model. In the final section, some comments are also addressed to the use of such indices for model modification.

3.2 TESTING OF HYPOTHESIZED STRUCTURE OF $\underset{\sim}{\Sigma}$

Before we discuss various tests of significance connected with the testing of structural hypotheses, a number of definitions and results useful in ACOVS will be presented below.

3.2.1 Matrix Preliminaries

Definition 3.2.1 Let $\underset{\sim}{A}$ be a p x p symmetric matrix. Following Anderson (1969), the symbol $\langle \underset{\sim}{A} \rangle$ denotes a $p(p+1)/2$ component column vector consisting the upper triangle elements of $\underset{\sim}{A}$.

$$\langle \underset{\sim}{A} \rangle = (a_{11}, a_{22}, a_{33}, \cdots, a_{pp}; a_{12}, a_{13}, \cdots, a_{1p}, a_{23}, \cdots, a_{p-1,p})'. \qquad (3.2.1)$$

Using this definition, for a symmetric matrix $\underset{\sim}{\Sigma} = \sum_{i=1}^{k} \underset{\sim}{H}_i \theta_i$ [for example, Ch. 2, equ. (2.1.1)], we may define $\underset{\sim}{H} = [\langle \underset{\sim}{H}_1 \rangle, \langle \underset{\sim}{H}_2 \rangle, \cdots, \langle \underset{\sim}{H}_k \rangle]$ and $\underline{\theta} = (\theta_1, \theta_2, \cdots, \theta_k)'$ so that $\langle \underset{\sim}{M} \rangle = \underset{\sim}{H}\underline{\theta}$. For identity matrix $(\underset{\sim}{I}_p)$ of order p, $\langle \underset{\sim}{I}_p \rangle = (1, 1, \cdots, 1; 0, 0, \cdots, 0)$.

Definition 3.2.2 For a nonsingular (n.s) matrix $\underset{\sim}{B}$, $\underset{\sim}{\Phi}(\underset{\sim}{B})$

be defined to be a $(p+1)/2$ x $(p+1)/2$ symmetric matrix with elements $\underset{\sim}{\Phi}(\underset{\sim}{B}) = ((\varphi_{ij,gh}))$ where

$$\varphi_{ij,gh} = b_{ig}\, b_{ih} + b_{ih}\, b_{jg} \qquad (3.2.2)$$

$i \leq j$, $g \leq h$ where b_{ij} denotes the ij-th element of $\underset{\sim}{B}$. The notation $\varphi_{ij,gh}$ represents the elements of $\underset{\sim}{\Phi}$ with row in the same position as the element a_{ij} in $\langle \underset{\sim}{A} \rangle$ where $\underset{\sim}{A}$ is pxp symmetric, and column in the same position as a_{gh} in $\langle \underset{\sim}{A} \rangle'$. Notationally, $\underset{\sim}{\Phi}(\underset{\sim}{I}_p) = \text{diag.}(2\underset{\sim}{I}_p, \underset{\sim}{I}_{p(p-1)/2})$.

From the above two definitions, Anderson (1969) obtained the following results.

Lemma 3.1 If $\underset{\sim}{B}$ is p.d., so is $\underset{\sim}{\Phi}(\underset{\sim}{B})$.

Lemma 3.2 For two symmetric matrices $\underset{\sim}{A}$ and $\underset{\sim}{C}$ both of order pxp and any p.d. matrix $\underset{\sim}{B}(Pxp)$,

$$\langle \underset{\sim}{A} \rangle' \, \underset{\sim}{\Phi}^{-1}(\underset{\sim}{B}) \, \langle \underset{\sim}{C} \rangle = \frac{1}{2} \text{ tr } (\underset{\sim}{B}^{-1} \, \underset{\sim}{A}\underset{\sim}{B}^{-1}\underset{\sim}{C}). \qquad (3.2.3)$$

Notationally, $\langle \underset{\sim}{I}_p \rangle' \, \underset{\sim}{\Phi}(\underset{\sim}{I}_p) \langle \underset{\sim}{I}_p \rangle = 2p$ and $\langle \underset{\sim}{I}_p \rangle' \underset{\sim}{\Phi}^{-1}(\underset{\sim}{I}_p) \langle \underset{\sim}{I}_p \rangle = p/2$.

Szatrowski (1979) added the following two new results in this connection.

Lemma 3.3 If $\underset{\sim}{R}$ is a n.s. pxp matrix, then there exists a n.s. matrix $\underset{\sim}{B}$ such that

$$\langle \underset{\sim}{R}\underset{\sim}{S}\underset{\sim}{R}' \rangle = \underset{\sim}{B}\langle \underset{\sim}{S} \rangle \qquad (3.2.4)$$

for any pxp symmetric matrix $\underset{\sim}{S}$. If in addition, $\underset{\sim}{S}$ is p.d.,

$$\underset{\sim}{\Phi}(\underset{\sim}{R}\underset{\sim}{S}\underset{\sim}{R}') = \underset{\sim}{B}\,\underset{\sim}{\Phi}(\underset{\sim}{S})\underset{\sim}{B}' \qquad (3.2.5)$$

where the elements of $\underset{\sim}{\Phi}$ are defined as in (3.2.2).

Remarks 3.1 We observe that for $\underset{\sim}{\Sigma} = \underset{\sim}{D}_\sigma \underset{\sim}{\rho} \underset{\sim}{D}_\sigma$ where $\underset{\sim}{D}_\sigma = \text{diag}(\sigma_1, \sigma_2, \cdots, \sigma_p)$,

$$\langle \underset{\sim}{\Sigma} \rangle = \underset{\sim}{D}_1 \langle \underset{\sim}{\rho} \rangle \qquad (3.2.6)$$

and

$$\underset{\sim}{\Phi}(\underset{\sim}{\Sigma}) = \underset{\sim}{D}_1 \underset{\sim}{\Phi}(\underset{\sim}{\rho})\underset{\sim}{D}_1 \qquad (3.2.7)$$

where $\underset{\sim}{D}_1 = \text{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_p^2; \sigma_1\sigma_2, \cdots \sigma_1\sigma_p; \sigma_2\sigma_3, \cdots, \sigma_{p-1}\sigma_p)$.

Lemma 3.4 : If $\underset{\sim}{\Sigma}_1$, $\underset{\sim}{\Sigma}_2$ and $\underset{\sim}{\Sigma}_3$ be pxp p.d. covariance matrices with $\underset{\sim}{\Phi}_i = \underset{\sim}{\Phi}(\underset{\sim}{\Sigma}_i)$, i = 1,2,3,

$$\langle \underset{\sim}{\Sigma}_1 \rangle' \underset{\sim}{\Phi}_1^{-1} \underset{\sim}{\Phi}_3 \underset{\sim}{\Phi}_2^{-1} \langle \underset{\sim}{\Sigma}_2 \rangle = \frac{1}{2} \text{tr}(\underset{\sim}{\Sigma}_1^{-1}\underset{\sim}{\Sigma}_3\underset{\sim}{\Sigma}_2^{-1}\underset{\sim}{\Sigma}_3). \qquad (3.2.8)$$

■

3.2.2 Some Results on $\underset{\sim}{Q}$-matrix

Let $\underline{X}_\alpha$, $\alpha = 1(1)N$ be a random sample drawn from $N_p(\underline{X}; \underline{\mu}, \underset{\sim}{\Sigma})$ where $\underline{\mu}$ (px1) is arbitrary and unknown, but the elements of $\underset{\sim}{\Sigma}$ are functions of basic parameter vector $\underline{\theta} = (\theta_1, \theta_2, \cdots, \theta_k)'$ such that $\underline{\theta} \in \textcircled{H}_k$, $\textcircled{H}_k = \{\underline{\theta} \in R^k / \underset{\sim}{\Sigma}(\underline{\theta})$ is p.d.$\}$. Let us define

$$\underset{\sim}{Q} = \underset{\sim}{\Sigma}^{-1} - \underset{\sim}{\Sigma}^{-1}\,\underset{\sim}{S}\underset{\sim}{\Sigma}^{-1} \qquad (3.2.9)$$

where $\underset{\sim}{S}$ is the sample covariance matrix.

Result 3.2.1 $\quad E\langle\underset{\sim}{Q}\rangle \approx \langle\underset{\sim}{Q}\rangle$, disp. $\langle\underset{\sim}{Q}\rangle \approx \frac{1}{N}\underset{\sim}{\Phi}(\underset{\sim}{\Sigma}^{-1})$.

Proof Since by definition of $\underset{\sim}{Q}$

$$\underset{\sim}{\Sigma} = \underset{\sim}{S} + \underset{\sim}{\Sigma}\,\underset{\sim}{Q}\,\underset{\sim}{\Sigma} \qquad (3.2.10)$$

by using lemma 3.3, we obtain

$$\langle\underset{\sim}{\Sigma}\rangle = \langle\underset{\sim}{S}\rangle + \underset{\sim}{B}^{*}\langle\underset{\sim}{Q}\rangle \qquad (3.2.11)$$

where $\underset{\sim}{B}^{*}$ is a $p(p+1)/2 \times p(p+1)/2$ n.s. matrix defined by the set of equations

$$\langle\underset{\sim}{\Sigma}\,\underset{\sim}{H}_{ij}\,\underset{\sim}{\Sigma}\rangle = \underset{\sim}{B}^{*}\langle\underset{\sim}{H}_{ij}\rangle \qquad (3.2.12)$$

$i \leq j = 1(1)p$, where $\underset{\sim}{H}_{ij}$ is a pxp symmetric matrix of zeroes with a one in (i,j) and (j,i) positions. Any pxp symmetric matrix can be expressed uniquely as a linear combination of $p(p+1)/2$ symmetric matrices $\underset{\sim}{H}_{ij}$, $i \leq j = 1(1)p$. For example, the symmetric $\underset{\sim}{\Sigma}^{-1}$ with elements σ^{ij} may be expressed as $\sum_{i\leq j} \sigma^{ij}\,\underset{\sim}{H}_{ij}$ [vide Szatrowski, 1979, proof of lemma 1].

From (3.2.12), we observe that

$$\langle\underset{\sim}{\Sigma}\Big(\sum_{i\leq j}\sigma^{ij}\,\underset{\sim}{H}_{ij}\Big)\,\underset{\sim}{\Sigma}\rangle = \underset{\sim}{B}^{*}\Big\langle\sum_{i\leq j}\sigma^{ij}\,\underset{\sim}{H}_{ij}\Big\rangle$$

i,e, $\langle\underset{\sim}{\Sigma}\rangle = \underset{\sim}{B}^{*}\langle\underset{\sim}{\Sigma}^{-1}\rangle$ implying that

$$\underset{\sim}{B}^{*-1}\langle\underset{\sim}{\Sigma}\rangle = \langle\underset{\sim}{\Sigma}^{-1}\rangle\,. \qquad (3.2.13)$$

Now, from (3.2.11), $\langle\underset{\sim}{Q}\rangle = \underset{\sim}{B}^{*-1}\langle\underset{\sim}{\Sigma}\rangle - \langle\underset{\sim}{S}\rangle$ so that in large

samples

$$E\,\langle \underset{\sim}{Q}\rangle = \underset{\sim}{B}^{*-1}\,(\langle \underset{\sim}{\Sigma}\rangle - \langle E\underset{\sim}{S}\rangle) \approx \langle \underset{\sim}{Q}\rangle \ .$$

Again since disp. $\langle \underset{\sim}{S}\rangle = \frac{1}{N}\underset{\sim}{\Phi}(\underset{\sim}{\Sigma})$ [cf. Szatrowski, 1979, Theo.2] by applying lemma 3.1 and equation (3.2.13) we find that

$$\begin{aligned} \text{disp. } \langle \underset{\sim}{Q}\rangle &\approx B^{*-1}\ \text{disp. } \langle \underset{\sim}{S}\rangle\ \underset{\sim}{B}^{*\prime-1} \\ &\approx \frac{1}{N}\ \underset{\sim}{B}^{*-1}\ \underset{\sim}{\Phi}(\underset{\sim}{\Sigma})\ \underset{\sim}{B}^{*\prime-1} \\ &= \frac{1}{N}\ \underset{\sim}{\Phi}\ (\underset{\sim}{\Sigma}^{-1})\,. \end{aligned} \qquad (3.2.14)$$

While the asymptotic distribution of $\langle \underset{\sim}{S}\rangle$ is multinormal with $(\langle \underset{\sim}{\Sigma}\rangle\ ,\ \frac{1}{N}\ \underset{\sim}{\Phi}(\underset{\sim}{\Sigma}))$, that of $\langle \underset{\sim}{Q}\rangle$ is also multinormal with $(\langle \underset{\sim}{Q}\rangle\ ,\ \frac{1}{N}\ \underset{\sim}{\Phi}(\underset{\sim}{\Sigma}^{-1}))$. [Through an explicit treatment, Jöreskog (1973, sec. 2.6) obtained the same distribution for the individual element of $\underset{\sim}{Q}$]. From this it is easy to obtain the result which shows that the quadratic forms $N\langle \underset{\sim}{S}-\underset{\sim}{\Sigma}\rangle'\ \underset{\sim}{\Phi}^{-1}(\underset{\sim}{\Sigma})\langle \underset{\sim}{S}-\underset{\sim}{\Sigma}\rangle$ and $N\langle \underset{\sim}{Q}\rangle'\ \underset{\sim}{\Phi}\ (\underset{\sim}{\Sigma}^{-1})\ \langle \underset{\sim}{Q}\rangle$ would follow asymptotically a chi-square distribution with $p(p+1)/2$ degrees of freedom (d.f.) each. Since algebraically $N\ \langle\ \underset{\sim}{S}-\underset{\sim}{\Sigma}\rangle'\ \underset{\sim}{\Phi}^{-1}(\underset{\sim}{\Sigma})\ \langle \underset{\sim}{S}-\underset{\sim}{\Sigma}\rangle = N\langle \underset{\sim}{Q}\rangle'\ \underset{\sim}{\Phi}^{-1}(\underset{\sim}{\Sigma}^{-1})\langle \underset{\sim}{Q}\rangle$ $= \frac{N}{2}\ \text{tr}(\underset{\sim}{Q}\underset{\sim}{\Sigma} - \underset{\sim}{Q}\underset{\sim}{S})$, we have the following result.

<u>Result 3.2.2</u> The statistic $\frac{N}{2}\ \text{tr}(\underset{\sim}{Q}\underset{\sim}{\Sigma}-\underset{\sim}{Q}\underset{\sim}{S})$ would asymptotically follow a chi-square distribution with $p(p+1)/2$ d.f. ■

3.2.3 <u>Large Sample Tests of a Specified Covariance Structure</u>

For getting some insight into the three large-scale tests useful in ACOVS, namely, (a) the conventional likelihood

ratio test based on Wilks (1946) lambda statistic. (b) a new test procedure obtained from Rao's (1947) criterion based on efficient scores and (c) another new test based on Wald's (1943) quadratic criterion, we shall present in this section the essential part of the procedure connected with each.

The second test is famous in econometrics in the name of 'Lagrange Multiplier' test (Aitchison and Silvey, 1958). The third test based on Wald's (1943) criterion is also quite popular in the same field. The asymptotic theory of all these large-sample restricted parametric tests has very recently been discussed by Satorra (1989).

(i) Likelihood Ratio Criterion

To test the null hypothesis of restricted $\underset{\sim}{\Sigma}$ namely $H_o : \underset{\sim}{\Sigma} = \underset{\sim}{\Sigma}(\underline{\theta})$ against unrestricted alternative, the most popular large sample test is based on Neyman-Pearson's likelihood ratio criterion (cf. Rao, 1974, sec. 6, p. 417) and is given by

$$\lambda = (\det \underset{\sim}{S}/\det \hat{\underset{\sim}{\Sigma}})^{-N/2} \qquad (3.2.15)$$

where $\hat{\underset{\sim}{\Sigma}} = \underset{\sim}{\Sigma}(\hat{\underline{\theta}})$ is the restricted MLE of $\underset{\sim}{\Sigma}(\underline{\theta})$ under H_o. When no restriction is imposed, the unrestricted MLE of $\underset{\sim}{\Sigma}$ is simply $\underset{\sim}{S}$, the sample covariance matrix. The statistic

$$T_1 = - N \log_e \lambda \qquad (3.2.16)$$

is distributed asymptotically under H_o as a χ^2 with υ d.f. where $\upsilon = p(p+1)/2 - k$ (Wald, 1943).

A level-α test rule is to reject H_o if T_1 exceeds $\chi^2_\alpha(\upsilon)$, the critical value of χ^2 with υ d.f. for α level of significance. The LR chisquare test is sensitive to data nonnormality (Brown, 1982, 1984; Tanaka and Huba, 1987).

(ii) Rao's Efficient Score Criterion

Given the log-likelihood function log $L(\underset{\sim}{\varphi})$ of the parameter vector $\underset{\sim}{\varphi}(rx1)$ based on random sample $\{\underset{\sim}{X}_\alpha, \alpha = 1(1)N\}$ from a population $\underset{\sim}{X} \sim f(\underset{\sim}{X}, \underset{\sim}{\varphi})$, Rao introduced a vector of 'efficient scores' defined by

$$\underset{\sim}{v}(\underset{\sim}{\varphi}) = \frac{1}{\sqrt{N}} \left(\frac{\partial \log L}{\partial \varphi_1}, \cdots, \frac{\partial \log L}{\partial \varphi_r} \right)' \qquad (3.2.18)$$

and showed that the asymptotic distribution of $\underset{\sim}{v}(\underset{\sim}{\varphi})$ is r-variate normal with null mean vector and dispersion matrix

$$\underset{\sim}{Z}(\underset{\sim}{\varphi}) = - \left(\left(E \frac{\partial^2 \log f}{\partial \varphi_i \partial \varphi_j} \right) \right)_{rxr} \qquad (3.2.19)$$

For the restricted null hypothesis involving t number of restrictions upon the elements of $\underset{\sim}{\varphi}$, say, $H_o[R_i(\underset{\sim}{\varphi}) = 0, i = 1,2, \cdots, t < r]$, Rao (1947) showed that the asymptotic distribution of

$$S_o = \underset{\sim}{v}'(\hat{\underset{\sim}{\varphi}}) \; \underset{\sim}{Z}^{-1}(\hat{\underset{\sim}{\varphi}}) \underset{\sim}{v} (\hat{\underset{\sim}{\varphi}}) \qquad (3.2.20)$$

is a χ^2 variate with t d.f. where $\hat{\underset{\sim}{\varphi}}$ is the MLE of $\underset{\sim}{\varphi}$ subject to t restrictions as specified under the null hypothesis.

Keeping the above results in view for ACOVS, we write

below the log-likelihood function under multinormality assumption as

$$\log L = -\frac{N}{2}[\log(2\pi) + \log\det \underset{\sim}{\Sigma} + \mathrm{tr}(\underset{\sim}{\Sigma}^{-1}\underset{\sim}{S})]. \qquad (3.2.21)$$

Here the relevant parameter vector is composed of $p(p+1)/2$ nonduplicated elements of $\underset{\sim}{\Sigma}$. We find that

$$\frac{\partial \log L}{\partial \sigma_{ij}} = -\frac{N}{2}\,\mathrm{tr}\left(\underset{\sim}{Q}\,\frac{\partial \Sigma}{\partial \sigma_{ij}}\right) = N \langle \underset{\sim}{H}_{ij}\rangle' \langle \underset{\sim}{Q}\rangle \qquad (3.2.22)$$

where $\underset{\sim}{H}_{ij}$ is the sparse matrix of order $p \times p$ with (i,j) and (j,i) elements unity, $i \leq j = 1,2,\cdots,p$.

$$\frac{\partial^2 \log L}{\partial \sigma_{ij}\,\partial \sigma_{i'j'}} = \frac{N}{2}\left[2\,\mathrm{tr}(\underset{\sim}{Q}\underset{\sim}{H}_{ij}\underset{\sim}{\Sigma}^{-1}\underset{\sim}{H}_{i'j'}) - \mathrm{tr}(\underset{\sim}{H}_{ij}\underset{\sim}{\Sigma}^{-1}\underset{\sim}{H}_{i'j'}\underset{\sim}{\Sigma}^{-1})\right] \qquad (3.2.23)$$

Since $E(\underset{\sim}{Q}) \approx \underset{\sim}{Q}$ due to Result 3.2.1,

$$E\left(\frac{\partial^2 \log L}{\partial \sigma_{ij}\,\partial \sigma_{i'j'}}\right) \approx -\frac{N}{2}\,\mathrm{tr}(\underset{\sim}{H}_{ij}\underset{\sim}{\Sigma}^{-1}\underset{\sim}{H}_{i'j'}\underset{\sim}{\Sigma}^{-1})$$

$$= -N \langle \underset{\sim}{H}_{ij}\rangle' \underset{\sim}{\Phi}^{-1}(\underset{\sim}{\Sigma}) \langle \underset{\sim}{H}_{i'j'}\rangle. \qquad (3.2.24)$$

Since the matrix $(\langle \underset{\sim}{H}_{11}\rangle \langle \underset{\sim}{H}_{22}\rangle \cdots \langle \underset{\sim}{H}_{pp}\rangle \langle \underset{\sim}{H}_{12}\rangle \langle \underset{\sim}{H}_{13}\rangle \cdots \langle \underset{\sim}{H}_{p-1,p}\rangle)$ is simply $\underset{\sim}{I}_{p(p+1)/2}$, the vector of efficient scores is given by

$$\underline{v}\,(\langle \underset{\sim}{\Sigma}\rangle) = -\sqrt{N}\,\langle \underset{\sim}{Q}\rangle \qquad (3.2.25)$$

having the dispersion matrix

$$\underset{\sim}{Z}(\langle \underset{\sim}{\Sigma}\rangle) = \underset{\sim}{\Phi}(\underset{\sim}{\Sigma}^{-1}). \qquad (3.2.26)$$

The asymptotic distribution of $\underset{\sim}{v}'(<\underset{\sim}{\Sigma}>)\bar{\underset{\sim}{\Phi}}^{-1}(\bar{\underset{\sim}{\Sigma}}^{-1})\underset{\sim}{v}(<\underset{\sim}{\Sigma}>)$ is χ^2 with $p(p+1)/2$ d.f. . This distribution may otherwise be confirmed from result 3.2.2 since the expression $\underset{\sim}{v}'(<\underset{\sim}{\Sigma}>)\ \bar{\underset{\sim}{\Phi}}^{-1}(\bar{\underset{\sim}{\Sigma}}^{-1})\underset{\sim}{v}(<\underset{\sim}{\Sigma}>)$ may be simplified to $\frac{N}{2}\ \mathrm{tr}(\underset{\sim}{Q}\underset{\sim}{\Sigma} - \underset{\sim}{Q}\underset{\sim}{S})$.

If we now wish to test $H_o : \underset{\sim}{\Sigma} = \underset{\sim}{\Sigma}(\underline{\theta})$, it is readily seen that the statistic

$$T_2 = \underset{\sim}{v}'(<\hat{\underset{\sim}{\Sigma}}>)\ \bar{\underset{\sim}{\Phi}}^{-1}(\hat{\bar{\underset{\sim}{\Sigma}}}^{-1})\ \underset{\sim}{v}(<\hat{\underset{\sim}{\Sigma}}>)$$

$$= \frac{N}{2}\ \mathrm{tr}(\hat{\underset{\sim}{Q}}\ \hat{\underset{\sim}{\Sigma}} - \hat{\underset{\sim}{Q}}\underset{\sim}{S}) \qquad (3.2.27)$$

follows asymptotically χ^2 distribution with $p(p+1)/2 - k = \upsilon$ d.f. under H_o , $\hat{\underset{\sim}{\Sigma}}$ being the restricted MLE of $\underset{\sim}{\Sigma}$ under H_o. This statistic (3.2.27) may be then treated to be the ACOVS-version of Rao's efficient score criterion in the multinormal set-up.

Remark 3.2 There are many structures of $\underset{\sim}{\Sigma}\ (\underline{\theta})$ for which the MLE $\underline{\theta}$ may be found to satisfy the equation

$$\mathrm{tr}(\hat{\underset{\sim}{Q}}\ \hat{\underset{\sim}{\Sigma}}) = 0 . \qquad (3.2.28)$$

Let us denote such structural class of $\underset{\sim}{\Sigma}$ as $\mathcal{C}$. For example, when $\underset{\sim}{\Sigma}(\underline{\theta})$ is linearly structured, e.g. $\underset{\sim}{\Sigma}(\underline{\theta}) = \sum_{i=1}^{k} \underset{\sim}{H}_i \theta_i$ where $\underset{\sim}{H}_i$'s are linearly independent design matrices [vide ch. 2, equ. (2.1.1)], $\mathrm{tr}(\hat{\underset{\sim}{Q}}\ \hat{\underset{\sim}{\Sigma}}) = 0$. Thus in the case of $\underset{\sim}{\Sigma} \in \mathcal{C}$. (3.2.27) simplifies to

$$T_{2c} = -\frac{N}{2}\ \mathrm{tr}(\hat{\underset{\sim}{Q}}\ \underset{\sim}{S}) \qquad (3.2.29)$$

following asymptotically a χ^2 distribution with υ d.f.

The statistic T_{2c} will be also found useful in deriving the asymptotic distribution of the Jöreskog-Sörbom fit index which we discuss in sec. 3.3.1 in detail.

Remark 3.3 In the case of nonlinear structures as well as in some other special conditions, $tr(\Sigma\ Q)$ will not be necessarily equal to 0 . For example, when testing the H_0 that $\underset{\sim}{\Sigma} = \underset{\sim}{I}_p$, we may find that $tr(\hat{\underset{\sim}{\Sigma}}\ \hat{\underset{\sim}{Q}}) \neq 0$. Hence, for testing this H_0 against the unrestricted alternative, we will be required to compute the T_2 statistics of (3.2.27) rather than T_{2c} as shown in (3.2.29).

(iii) Wald's Quadratic Form Criterion

In order to test $H_0 [\underset{tx1}{\underline{R}}\ (\underline{\phi}) = \underline{0}]$, Wald's (1943) quadratic form statistic is given by

$$W = N\underline{R}'\ (\underset{\sim}{F}'\ \underset{\sim}{Z}^{-1}\ \underset{\sim}{F})^{-1}\ \underline{R} \qquad (3.2.30)$$

where $\underset{\sim}{F} = \left(\left(\frac{\partial R_i(\underline{\phi})}{\partial \phi_j}\right)\right)_{rxt}$ and $\underset{\sim}{Z}$ is as defined in equation (3.2.19). Wald (1943) showed that the statistic W (when computed on the basis of unrestricted MLE of $\underline{\phi}$) is asymptotically distributed as χ^2 with t d.f. .

Under multinormal assumption, paralleling the treatment of (ii), let

$$H_0 : \langle\underset{\sim}{\Sigma}\rangle = \langle\underset{\sim}{\Sigma}(\underline{\theta})\rangle$$

or, equivalently, $H_0 : \underline{R}(\langle\underset{\sim}{\Sigma}\rangle) = \underline{0}\ (\upsilon \times 1)$. (3.2.31)

Since unrestricted MLE of $\underset{\sim}{\Sigma}$ is $\underset{\sim}{S}$, (3.2.30) may be expressed as

$$W = N\underline{R}'(\langle\underset{\sim}{S}\rangle)\,[\underset{\sim}{F}'\underset{\sim}{\Phi}(\underset{\sim}{S})\underset{\sim}{F}]^{-1}\,\underline{R}(\langle\underset{\sim}{S}\rangle)\ . \qquad (3.2.32)$$

The asymptotic distribution of W under H_o(3.2.31) is a χ^2 with υ d.f. .

We shall consider now the statistic (3.2.32) with reference to H_o where $\underset{\sim}{\Sigma}$ is linearly structured [vide ch. 2, equ. (2.1.1)] with equivalent expression given by

$$H_o : \langle\underset{\sim}{\Sigma}\rangle = \underset{\sim}{H}\,\underline{\theta} \qquad (3.2.33)$$

where $\underset{\sim}{H} = (\langle\underset{\sim}{H}_1\rangle\ \langle\underset{\sim}{H}_2\rangle\ \cdots\ \langle\underset{\sim}{H}_k\rangle)_{p(p+1)/2\ x\ k}$. Given $\underset{\sim}{H}_i$-matrices, $\langle\underset{\sim}{\Sigma}\rangle$ may be expressed as a partial sum of k vectors, e.g. $\langle\underset{\sim}{H}_1\rangle^*\ \langle\underset{\sim}{\Sigma}\rangle$, i = 1(1)k where the symbol * represents the Schur product (term by term product). From the components of $\langle H_i\rangle\ ^*\ \langle\underset{\sim}{\Sigma}\rangle$, there may be as many as (f_i-1) linearly independent contrasts formed where f_i is the number of 1's in the sparse vector $\langle\underset{\sim}{H}_i\rangle$, $f_i \geq 2$. For $f_i = 1$, however, no such contrast could be formed. The total number of contrasts is $\sum_{i=1}^{k}(f_i-1)$, i.e. υ in the case of $\sum_{i=1}^{k} f_i = p(p+1)/2$. Denoting the contrast vectors by $\langle\underset{\sim}{M}_\alpha\rangle$, we thus reexpress (3.2.33) by υ contrasts, e.g. $\langle\underset{\sim}{M}_\alpha\rangle'\ \langle\underset{\sim}{\Sigma}\rangle = 0$, $\alpha = 1(1)\upsilon$. In a more compact way, letting the contrast matrix

$$\underset{\sim}{M} = (\langle\underset{\sim}{M}_1\rangle\ \langle\underset{\sim}{M}_2\rangle\ \cdots\ \langle\underset{\sim}{M}_\upsilon\rangle)_{p(p+1)/2\ x\ \upsilon}\ , \qquad (3.2.33)$$

is thus equivalent to

$$H_o : \underset{\sim}{M}'\langle\underset{\sim}{\Sigma}\rangle = \underline{0}\ (\upsilon\ x\ 1)\ . \qquad (3.2.34)$$

Recalling (3.2.32), Wald's statistic would take the form

$$W_c = N\langle \underset{\sim}{S} \rangle' \underset{\sim}{M} [\underset{\sim}{M}' \underset{\sim}{\Phi}(\underset{\sim}{S}) \underset{\sim}{M}]^{-1} \underset{\sim}{M}' \langle \underset{\sim}{S} \rangle \qquad (3.2.35)$$

which, under H_o (3.2.34), is asymptotically distributed as a χ^2 with υ d.f. .

As regards the construction of the contrast matrix $\underset{\sim}{M}$, it may be seen elsewhere [cf. Kshirsagar, 1972, ch. 5] that for two different contrast matrices $\underset{\sim}{M}^{(1)}$ and $\underset{\sim}{M}^{(2)}$ both representing (3.2.34), $\underset{\sim}{M}^{(1)} = \underset{\sim}{M}^{(2)} \underset{\sim}{B}$ where $\underset{\sim}{B}(\upsilon \times \upsilon)$ is a nonsingular matrix. In this event, W_c is invariant with respect to the choice of contrast matrix defining (3.2.34). We prefer the simplest possible way of contructing $\underset{\sim}{M}$-matrix as shown below with the aid of an example of 3x3 Toeplitz matrix.

$$H_o : \underset{\sim}{\Sigma} = \begin{bmatrix} a_o & a_1 & a_2 \\ a_1 & a_o & a_1 \\ a_2 & a_1 & a_o \end{bmatrix}$$

Equivalently, $H_o : \langle \underset{\sim}{\Sigma} \rangle = (\langle \underset{\sim}{H}_1 \rangle \langle \underset{\sim}{H}_2 \rangle \langle \underset{\sim}{H}_3 \rangle)(a_o\ a_1\ a_2)'$ where $\langle \underset{\sim}{H}_1 \rangle = (111000)'$, $\langle \underset{\sim}{H}_2 \rangle = (000101)'$ and $\langle \underset{\sim}{H}_3 \rangle = (000001)'$. We choose

$$\underset{\sim}{M}_{6x3} = (\langle \underset{\sim}{M}_1 \rangle, \langle \underset{\sim}{M}_2 \rangle, \langle \underset{\sim}{M}_3 \rangle) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \qquad (3.2.36)$$

where $\underset{\sim}{M}_1$ and $\underset{\sim}{M}_2$ are based on $\underset{\sim}{H}_1$ while $\underset{\sim}{M}_3$ on $\underset{\sim}{H}_2$.

When the elements of $\underset{\sim}{\Sigma}$ under H_o are unconstrained or specified, $\underset{\sim}{M}$ matrix may be constructed accordingly. In such cases, columns of $\underset{\sim}{M}$ will not be contrast vectors. Thus as, in general, Σ under H_o may have constrained, unconstrained and/or specified elements, $\underset{\sim}{M}$ matrix would consist of contrast or non-contrast vectors all of which are necessarily linearly independent.

Remark 3.4 We may note in this connection that when we adopt Q-procedure to obtain MLE in the case of linearly structured $\underset{\sim}{\Sigma}$-matrix, $\underset{\sim}{Q}$ matrix under likelihood equations may be expressed as $\underset{\sim}{Q} = \sum_{\alpha=1}^{\upsilon} \underset{\sim}{M}_\alpha \, q_\alpha$ where $\underset{\sim}{M}_\alpha$'s are linearly independent symmetric matrices [vide ch. 2, lemma 2.1] and may be eventually useful here to construct the contrast matrix $\underset{\sim}{M}$. Consequently, T_{2c} may have an alternative expression parallel to that of W_c deduced as follows :

Since $\langle \hat{\underset{\sim}{Q}} \rangle = \sum_{\alpha=1}^{\upsilon} \langle \underset{\sim}{M}_\alpha \rangle \, \hat{q}_\alpha = \underset{\sim}{M} \, \hat{\underline{q}}$ where $\underset{\sim}{M} = (\langle \underset{\sim}{M}_1 \rangle \langle \underset{\sim}{M}_2 \rangle \cdots \langle \underset{\sim}{M}_\upsilon \rangle)$ and $\hat{\underline{q}} = (\hat{q}_1 \; \hat{q}_2 \cdots \hat{q}_\upsilon)'$, then

$$T_{2c} = -\frac{N}{2} \operatorname{tr}(\hat{\underset{\sim}{Q}} \, \underset{\sim}{S}) = -N \langle \hat{Q} \rangle_{\Lambda}^{\Phi^{-1}(I_p)} \langle S \rangle = -N \, \hat{\underline{q}}' \underset{\sim}{M}'_{\Lambda}^{\Phi^{-1}(I_p)} \langle \underset{\sim}{S} \rangle \; . \quad (3.2.37)$$

According to Q-procedure, $\hat{\underline{q}} = \underset{\sim}{A}^{-1} \, \underline{b}$ where $\underset{\sim}{A} = \underset{\sim}{A}' = ((\operatorname{tr}(\underset{\sim}{M}_\alpha \hat{\underset{\sim}{\Sigma}} \, \underset{\sim}{M}_\beta \hat{\underset{\sim}{\Sigma}})))_{\upsilon x \upsilon}$ and $\underline{b} = (-\operatorname{tr}(\underset{\sim}{M}_\alpha \underset{\sim}{S}))_{\upsilon x 1}$ [vide ch. 2, equ. (2.2.16)]. By Anderson's notation,

$\text{tr}(\underset{\sim}{M}_\alpha \hat{\underset{\sim}{\Sigma}} \underset{\sim}{M}_\beta \hat{\underset{\sim}{\Sigma}}) = 2\langle \underset{\sim}{M}_\alpha \rangle' \; \underset{\sim}{\Phi}^{-1}(\hat{\underset{\sim}{\Sigma}}^{-1}) \langle \underset{\sim}{M}_\beta \rangle$ and $\text{tr}(\underset{\sim}{M}_\alpha \underset{\sim}{S}) = 2\langle \underset{\sim}{M}_\alpha \rangle' \underset{\lambda}{\Phi^{-1}(I_p)} \langle \underset{\sim}{S} \rangle$

so that $\underset{\sim}{A} = \underset{\sim}{M}' \underset{\sim}{\Phi}^{-1}(\hat{\underset{\sim}{\Sigma}}^{-1}) \underset{\sim}{M}$, $b = - \underset{\sim}{M}' \underset{\lambda}{\Phi^{-1}(I_p)} \langle \underset{\sim}{S} \rangle$ and finally,

$$T_{2c} = N\langle \underset{\sim}{S} \rangle \underset{\sim}{\Phi}^{-1} \underset{\sim}{M} [\underset{\sim}{M}' \underset{\sim}{\Phi}^{-1}(\hat{\underset{\sim}{\Sigma}}^{-1}) \underset{\sim}{M}]^{-1} \underset{\sim}{M}' \underset{\lambda}{\Phi^{-1}(I_p)} \langle \underset{\sim}{S} \rangle \qquad (3.2.38)$$

where $\hat{\underset{\sim}{\Sigma}}$ is the MLE of linearly structured $\underset{\sim}{\Sigma}$ (under H_o).

Three large sample test criteria described above are asymptotically equivalent [Rao, 1974, p. 418; Silvey, 1959]. There is no adequate discussion as yet in the statistical literature on the relative merits of these tests in detecting departures from the null hypothesis.

From the practical point of view, however, the statistic T_1 requires computation of two determinants, e.g. det $\underset{\sim}{S}$ and det $\hat{\underset{\sim}{\Sigma}}$ both of which are much sensitive to computational errors due to rounding. Sometimes mistakes emerge from the " ill-condition" of $\underset{\sim}{S}$ and $\hat{\underset{\sim}{\Sigma}}$ matrices while T_2 is not so much influenced by these factors. When $\underset{\sim}{\Sigma}$ is properly identifiable by the parameter $\underline{\theta}(kx1)$, the inverse of $\underset{\sim}{F}' \underset{\sim}{\Phi}(\underset{\sim}{S}) \underset{\sim}{F}$ of W will exist. The statistic W is very easy to calculate since it does not require MLE under H_o as required in calculating T_1 and T_2 . The procedure of calculating W_c in the case of linearly structured $\underset{\sim}{\Sigma}$ is already adopted by many authors, e.g. Harris (1985), Levy (1976), Choi and Wette (1972), Han (1968), particularly in the field of testing homogeneity of variances and/or covariances.

Remark 3.5 In the case of testing $H_o : \underset{\sim}{\Sigma} = \underset{\sim}{I}_p$, the statistic T_2 and W_c do not call for the MLE and eventually they are found to be equivalent, e.g. both T_2 and W_c boil down to

$$T_2 = W_c = \frac{N}{2} \operatorname{tr}(\underset{\sim}{S} - \underset{\sim}{I}_p)^2 \cdot \qquad (3.2.39)$$

3.2.4 A Numerical Illustration of the Test Procedures

Let us consider the sample covariance matrix ($\underset{\sim}{C}$) originated in a study of proficiency in a two-hand co-ordination task conducted by Bilodeau (1957). Table 3.1 shows the transformed covariance matrix $\underset{\sim}{S} = \underset{\sim}{T}^{-1} \underset{\sim}{C} \underset{\sim}{T}^{'-1}$ where $\underset{\sim}{T}$ is a known lower triangular matrix with all its non-zero elements equal to unity. The variables represented by the rows and columns of Table 3.1 are measures of proficiency (improvement from the preceding trial) in the two-hand co-ordination task at six stages of practice. The data were collected from 152 basic airmen who were asked to move a pin around a clover-shaped runway by the co-ordinated turning of two control handles. The proficiency was originally measured in eight trials, each being 60 seconds long with 30 seconds rest interval. For certain convenience, the data for the first and fifth trials have not been considered. The reduced data have been analyzed previously by Mukherjee (1963), Bock and Bargmann (1966), Jöreskog (1970) and Lee (1980).

Table 3.1

Transformed Covariance Matrix for Two-hand Co-ordination Test Based on 152 Basic Airman (Bilodeau, 1960)

Trial	2	3	4	6	7	8
2	521					
3	-44	143				
4	7	-47	105	Symmetric		
6	26	13	-47	170		
7	13	-8	0	-42	116	
8	5	-1	11	-11	-50	97

Let the population covariance matrix $\underset{\sim}{\Sigma}$ be a 6x6 tri-diagonal Jacobi matrix [vide ch. 2, subsec. 2.2.3(c)]. Under normality assumption, the MLE of the concerning 11 parameters are obtained by the Q-procedure [vide ch. 2, equs. (2.2.34) and (2.2.35)] as follows :

$$\hat{a}_1 = 521.00000,\ \hat{a}_2 = 141.91208,\ \hat{a}_3 = 103.07980,$$

$$\hat{a}_4 = 168.34507,\ \hat{a}_5 = 118.01643,\ \hat{a}_6 = 97.00000,$$

$$\hat{b}_1 = -36.16400,\ \hat{b}_2 = -43.43269,\ \hat{b}_3 = -40.56060,$$

$$\hat{b}_4 = -46.26969 \text{ and } \hat{b}_5 = -51.88496.$$

The corresponding MLE of $\underset{\sim}{Q} = ((q_{ij}))_{6x6}$ matrix is given below.

$$\underset{\sim}{\hat{Q}} = 10^{-4}\begin{bmatrix} 0 & & & & & \\ 0 & .0002 & & & & \\ -3.4629 & 0 & .0004 & & \text{Sym.} & \\ -5.8663 & -5.4497 & 0 & .0001 & & \\ -6.0247 & 1.7510 & -4.4089 & 0 & .0001 & \\ -4.4061 & -1.8605 & -12.8856 & 4.1999 & .0001 & .0002 \end{bmatrix}$$

The converged solution of $\underset{\sim}{\hat{Q}}$ as given above is obtained after 10 iterations and is found correct to 9 places of decimal.

We now calculate $\frac{1}{N}$ times the Hessian matrix at the estimated parameter vector $\underline{\hat{\theta}}$, using the following partial second-order derivatives :

$$\frac{1}{N}\frac{\partial^2 \log L}{\partial a_i \partial a_j} = q_{ij}\sigma^{ij} - (\sigma^{ij})^2/2 \ , \ i,j = 1(1)6 \qquad (3.2.40)$$

$$\frac{1}{N}\frac{\partial^2 \log L}{\partial a_i \partial b_j} = (q_{i,j+1} - \sigma^{i,j+1})\sigma^{ij} + q_{ij}\sigma^{i,j+1} \ , \qquad (3.2.41)$$

$$i = 1(1)6, \ j = 1(1)5$$

$$\frac{1}{N}\frac{\partial^2 \log L}{\partial b_i \partial b_j} = (q_{i,j+1} - \sigma^{i,j+1})\sigma^{i+1,j}$$

$$+ (q_{i+1,j+1} - \sigma^{i+1,j+1})\sigma^{ij}$$

$$+ q_{i+1,j}\,\sigma^{i,j+1} + q_{ij}\sigma^{i+1,j+1}, \ i,j = 1(1)5 \qquad (3.2.42)$$

where the notation σ^{ij} and q_{ij} denote the (i,j) element

of $\underset{\sim}{\Sigma}^{-1}$ and $\underset{\sim}{Q}$ respectively. Table 3.2 shows the Hessian matrix ($J(\underline{\theta})$) obtained at $\underline{\theta} = \underline{\hat{\theta}}$ which satisfies the checking formula provided in section 2.4 (Result 2.4.1). For comparative purpose, we also report in Table 3.3 the Fisher's information matrix ($\underset{\sim}{\mathcal{J}}(\underline{\theta})$) at $\underline{\theta} = \underline{\hat{\theta}}$. For obtaining the estimate of the standard error (SE) of components of $\underline{\hat{\theta}}$ by two different methods, namely E_1 and E_2 as was suggested in section 2.4, we computed the inverses of both $-\underset{\sim}{J}(\underline{\hat{\theta}})$ and $-\underset{\sim}{\mathcal{J}}(\underline{\hat{\theta}})$. The inverse of $-\underset{\sim}{\mathcal{J}}(\underline{\hat{\theta}})$ is shown in Table 3.4. It is seen that this inverse matrix has also an interesting structure which is not so clearly displayed by the inverse of $-\underset{\sim}{J}(\underline{\hat{\theta}})$. The square root of the diagonal entry provides the estimate of standard error (SE) of the corresponding component of $\underline{\hat{\theta}}$ vector. For the sake of comparison, we provide below these SE's .

$$SE(\underline{\hat{\theta}}) = SE\,[\hat{a}_i\ ,\ i = 1(1)6;\ \hat{b}_j\ ,\ j = 1(1)5]$$

$$= [59.7,\ 16.1,\ 11.5,\ 18.9,\ 13.3,\ 11.1;\ 20.6,\ 9.9,\ 9.8,\ 10.2,\ 9.2\ ] \quad \text{(using } \underset{\sim}{J}(\underline{\hat{\theta}})\text{)}$$

$$= [59.8,\ 16.2,\ 11.6,\ 19.0,\ 13.1,\ 11.1;\ 20.6,\ 9.8,\ 9.8,\ 10.1,\ 9.2] \quad \text{(using } \underset{\sim}{\mathcal{J}}(\underline{\hat{\theta}})\text{)}$$

For carrying out large sample tests by the methods as mentioned in this section, we set the null hypothesis H_o: $\underset{\sim}{\Sigma}$ is of Tri-diagonal Jacobi structure against alternative

hypothesis H : $\underset{\sim}{\Sigma}$ is any arbitrary p.d. matrix . We calculate

$$\det \hat{\underset{\sim}{\Sigma}} \text{ (under } H_o) = 7.189418 \times 10^{12}$$

$$\det \underset{\sim}{S} = 6.801967 \times 10^{12}, \quad \mathrm{tr}(\hat{\underset{\sim}{Q}}\underset{\sim}{S}) = -0.1095968.$$

Likelihood ratio statistic (3.2.16) and Rao's efficient score statistic (3.2.29) are calculated respectively as

$$T_1 = 8.4206, \quad T_{2c} = 8.3294$$

To calculate Wald's statistic W_c, we rearrange H_o as

$$H_o' : \underset{\sim}{M}' \langle \underset{\sim}{\Sigma} \rangle = \underline{0} \ (10 \times 1) \qquad (3.2.48)$$

where $\underset{\sim}{M} = (\langle \underset{\sim}{M}_{ij} \rangle , \ i = 1,2,3,4; \ j = i+2, \ \ldots \ , 6)$.

The vector $\langle \underset{\sim}{M}_{ij} \rangle$ is a sparsevector of order (21x1) with 1 at the location where σ_{ij} occurs in the vector $\langle M \rangle$. Recalling (3.2.36), the statistic W_c for testing (3.2.43) simplies to

$$W_c = N \langle \underset{\sim}{S} \rangle^{*\prime} \ \underset{\sim}{\Phi}^{*-1}(\underset{\sim}{S}) \ \langle \underset{\sim}{S} \rangle^{*} \qquad (3.2.44)$$

where $\langle \underset{\sim}{S} \rangle^{*} = (s_{ij} , \ i = 1,2,3,4; \ j = i+2, \ \ldots \ , 6)_{10x1}$,

$\underset{\sim}{\Phi}^{*}(\underset{\sim}{S}) = ((\mathcal{P}_{ij,gh}))_{10x10}$ with $\mathcal{P}_{ij,gh} = s_{ig}s_{jh}+s_{ih}s_{jg}$,

$i,g = 1,2,3,4; \ j = i+2, \ \ldots \ , 6; \ h = g+2, \ \ldots \ , 6.$

From $\underset{\sim}{S}$ matrix (vide Table 3.1), we calculate $\underset{\sim}{\Phi}^{*}(\underset{\sim}{S})$ and its inverse $\underset{\sim}{\Phi}^{*-1}(\underset{\sim}{S})$ to obtain W_c from (3.2.44). The statistic turns out to be as follows :

Hessian matrix* at $\underline{\theta} = \underline{\hat{\theta}}$

	a_1	a_2	a_3	a_4	a_5	a_6	b_1	b_2	b_3	b_4	b_5
a_1	1.9107										
a_2	0.1106	35.4084									
a_3	0.1341	7.9424	81.8202								
a_4	0.0485	1.2338	6.4353	31.3669							
a_5	0.0248	0.0639	2.5018	8.2421	86.5915						
a_6	0.0097	0.1039	1.7513	1.4461	24.7750	99.0458					
b_1	1.1145	4.9155	2.4827	0.7847	0.3610	0.1455	16.8346				
b_2	0.3639	33.5397	50.9843	5.9651	0.9843	0.8794	7.5702	123.5346			
b_3	0.2106	6.6269	45.8928	28.4151	9.2566	4.5216	3.4944	35.5685	114.1908		
b_4	0.0732	0.7569	8.1793	32.1577	53.4303	12.3308	1.1065	4.9779	32.6236	120.7169	
b_5	0.0312	0.2285	4.6123	7.1120	92.6349	99.0729	0.4652	2.1706	13.4045	51.6324	234.7693

*Multiply each element by $-N/10^6$.

Table 3.3

Fisher's Information matrix at $\underline{\theta} = \hat{\underline{\theta}}$

Parameters	a_1	a_2	a_3	a_4	a_5	a_6	b_1	b_2	b_3	b_4	b_5
a_1	1.9207										
a_2	0.1706	35.4084									
a_3	0.0383	7.9425	81.8202								
a_4	0.0030	0.6247	6.4353	31.3669							
a_5	0.0008	0.1642	1.6910	8.2422	86.5916						
a_6	0.0003	0.0469	0.4838	2.3582	24.7750	99.0459					
b_1	1.1449	4.9156	1.1026	0.0867	0.0228	0.0070	16.8346				
b_2	0.1616	33.5397	50.9844	4.0100	1.0537	0.3015	4.6561	123.5346			
b_3	0.0215	4.4549	45.8928	28.4151	7.4666	2.1363	0.6185	28.5971	114.1908		
b_4	0.0031	0.6404	6.5976	32.1577	53.4303	15.2872	0.0889	4.1111	29.1316	120.7169	
b_5	0.0009	0.1756	1.8090	8.8172	92.6350	99.0730	0.0244	1.1273	7.9877	57.1594	234.7693

*Multiply each element by - $N/10^6$.

Table 3.4

Inverse of the negative of Fisher's information matrix at $\underline{\theta} = \hat{\underline{\theta}}$

Parameters											
a_1	3571.54										
a_2	17.21	263.62									
a_3	0	24.82	135.76								
a_4	0	0	21.64	361.43							
a_5	0	0	0	28.16	172.83						
a_6	0	0	0	0	35.42	123.80		Symmetric			
b_1	-247.92	- 57.74	0	0	0	0	424.52				
b_2	0	- 79.66	-52.41	0	0	0	0	96.51			
b_3	0	0	-47.79	-77.18	0	0	0	0	96.03		
b_4	0	0	0	-91.30	-54.95	0	0	0	0	103.14	
b_5	0	0	0	0	-70.82	- 66.22	0	0	0	0	83.91

$$W_c = 8.0690$$

We now observe that each of the computed T_1, T_{2c} and W_c are more or less comparable and less than the critical value $\chi^2_{.05(10)}$ (= 18.3070) required for the 5 percent level of significance with d.f. $\upsilon = 10$. So we are obliged to accept the null hypothesis that the structure of the transformed covariance matrix for the two-hand co-ordination task is consistent with the hypothesized tri-diagonal Jacobi structure. Since this structure is a nonsingular transformation of the Guttman additive quasi-Weiner-simplex covariance structure by a known triangular matrix $\underset{\sim}{T}$ (Mukherjee, 1963), the findings confirm the assertion that the original sample covariance matrix $\underset{\sim}{C} = \underset{\sim}{T}\ \underset{\sim}{S}\ \underset{\sim}{T}'$ has been drawn from a population which has a covariance matrix of the quasi-Weiner simplex form. This conclusion supports the inference independently arrived at by Mukherjee (1963), Bock and Bargmann (1966) as well as by Jöreskog (1970) who analyzed the same data.

3.2.5 A Monte Carlo Experiment on Power of Test Statistics

A 'Monte Carlo' simulation was designed to compare the power of the test statistics mentioned in this section to test the null hypothesis $H_o : [\underset{\sim}{\Sigma} = \underset{\sim}{I}_p]$ against the alternative $H_a : [\underset{\sim}{\Sigma}$ has an intraclass structure]. This was done by drawing samples from a 8-variate (p = 8) normal population with zero mean vector.

The simulation study was conducted using an IBM compatible personal computer with random normal deviates generated by Box and Muller (1958) method. We considered H_a , in particular, with diagonal elements unity and a constant off-diagonal element ρ . We took into consideration the fact that intraclass $\underset{\sim}{\Sigma}$ is diagonally reducible by an orthogonal matrix (here 8x8 Hadamard matrix) whose elements are independent of the parameter ρ . Using that orthogonal matrix, we generated observational vectors from an 8-variate intraclass normal population having null mean vector. For each population correlation coefficient ρ , we calculated 100 sample covariance matrices (replications) for sample sizes N = 20, 40 and 100. From the admissible interval $(-\frac{1}{p-1}, 1)$ of ρ , we chose the value of ρ as -.10, -.05, 0, .05, .1 , .2, .3, .5 and .9 only in order to specify H_a for power calculation. Thus a total of 9x3x100 = 2,700 sample covariance matrices each of size 8x8 were generated. The LR criterion gives rise to the test statistic (cf. Anderson, 1984, p. 435) expressible as

$$T_1 = N\,[\,\mathrm{tr}\,\underset{\sim}{S} - \log(\det \underset{\sim}{S}) - p\,] \qquad (3.2.45)$$

which follow, asymptotically a chi-square distribution with $p(p+1)/2$ d.f. under H_o . Bartlett (1950) made certain modification of (3.2.45) to obtain improved chi-square test of scale-free sphericity, the statistic for which in the present case may be written as

$$T_{1B} = (N-1 - \frac{2p+5}{6}) \frac{T_1}{N} \cdot \qquad (3.2.46)$$

The above mentioned statistic also follows asymptotically a χ^2 distribution with the same $p(p+1)/2$ d.f. under H_o .

From Remark 3.5, the statistic

$$W_o = T_2 = \frac{N}{2} \text{ tr } (\underset{\sim}{S}-\underset{\sim}{I}_p)^2 \qquad (3.2.47)$$

which also follows asymptotically a χ^2 distribution with $p(p+1)/2$ d.f. under H_o . Nagao (1973) has suggested the same test as a modified likelihood ratio criterion for the above H_o . We are also interested in the performance of Steiger's (1980) chi-square statistic given by

$$T_3 = (N-3) \sum_{i=1}^{p} \sum_{j>1} Z_{ij}^2 \qquad (3.2.48)$$

where Z_{ij} is Fisher's $\tanh^{-1}$ - transformation upon the elements r_{ij} of the sample correlation matrix. This test is "worthy of immediate attention primarily because the computational formula requires no matrix inversion, would greatly reduce time spent in testing $\underset{\sim}{\rho} = \underset{\sim}{I}_p$, and is simple enough to encourage its use prior to performing factor analysis' (Wilson and Martin, 1983, p. 12). Under H_o, the statistic T_3 is also asymptotically distributed as a χ^2 variate with $p(p+1)/2$ d.f. .

To assess the power of the four statistics T_1, T_{1B}, W_o and T_3 , we calculated their rejection rate empirically

for each sample covariance matrix and compared at the 5 percent and 1 percent levels of significance (nominal rejection values respectively being 50.76 and 57.80 for d.f. 36, 36.4 and 48.3 for d.f. 28). We expected that the proportion of times that the model misspecification (i.e., a sample from H_a) is detected by each of the four competing tests at 5 percent and 1 percent levels of significance would increase with the sample size and the magnitude of model misspecification (i.e. the value of ρ).

The results of the simulation study are summarized in Table 3.5 and are somewhat comparable to those of Wilson and Martin (1983) who compared the performances of Bartlett's test (T_{1B}) and Steiger's test (T_3 for only positive values of ρ .

We shall call a test unbiased if its rejection rate under the true H_o [here for $\rho = 0$] is more or less close to the nominal level of significance for any sample size (N), and a test is consistent if its rejection rate under H_a [here for admissible $\rho \neq 0$] increases to 100 as N increases. Keeping these in view, we have the following observationsfrom the present simulation study.

(a) T_1 and W_c are slightly biased. Bias in T_1 is relatively more than that in W_c . These biases, however, decrease as N increases. The amount of bias in T_{1B} is considerably

N	Stat	ρ= -.10	ρ=-.05	ρ= 0	ρ= 0.05	ρ= .10	ρ= .20	ρ= .30	ρ= .50	ρ= .90
20	T1	65	32	35	44	41	61	91	100	100
	T1B	28	7	8	11	13	31	70	98	100
	WC	24	6	7	10	17	43	72	93	100
	T3	35	16	14	23	32	64	91	100	100
40	T1	83	28	17	18	43	85	100	100	100
	T1B	63	16	5	13	25	77	99	100	100
	WC	29	18	10	16	30	76	98	100	100
	T3	60	30	13	24	55	93	100	100	100
100	T1	100	34	11	25	64	100	100	100	100
	T1B	100	27	9	19	60	100	100	100	100
	WC	62	23	6	18	67	100	100	100	100
	T3	78	45	17	42	88	100	100	100	100

Significance level α= .01

20	T1	45	14	10	20	29	43	83	100	100
	T1B	6	1	1	5	1	12	49	97	100
	WC	10	2	2	9	11	32	58	90	100
	T3	8	2	1	4	9	12	78	100	100
40	T1	58	15	2	10	24	77	98	100	100
	T1B	58	10	1	4	12	60	94	100	100
	WC	16	9	3	9	16	66	96	100	100
	T3	10	5	2	7	25	82	100	100	100
100	T1	100	20	4	12	51	100	100	100	100
	T1B	100	16	4	6	45	100	100	100	100
	WC	62	16	4	12	56	100	100	100	100
	T3	78	7	1	9	63	100	100	100	100

small for both the levels of significance even though such small amount of bias does not show any trend of systematic reduction with the increase of N. The T_3 statistic also shows a very high Type-I error for small samples when the alpha is set at 5 percent level.

(b) The T_3 statistic exhibits sensitivity to non-sphericity over a wide range of alternative distributions. In most cases, it has better power than the other procedures selected for testing the scale-free sphericity hypothesis. Power wise, the statistic W_c performs relatively better than T_{1B} when the values of ρ under H_a are all positive and equal to or less than 0.9.

(c) It appears that one may achieve a reasonably high power with quite small sample size if T_3 is chosen when Type I error is set at 1 percent. The T_3 statistic clearly provides a superior omnibus indicator of non-sphericity judged over the various alternatives and over all the sample sizes used when Type I error is set to 1 percent. For this test, sensitive assessment (approximately 50 percent power at 5 percent level) of even slight non-sphericity ($\rho = \pm .10$) is possible in samples as small as N = 40. Furthermore, moderate non-sphericity (e.g., $\rho = .30$) can be detected quite correctly with small sample of size 20.

(d) The T_3 statistic is also unbiased at the 1 percent level of significance but showed a higher-than-nominal probability of Type I error, especially when alpha is set at .05. There is also no indication of reduction of bias as N increases at 5 percent level of significance. However, it appears that the performance of W_c in respect of Type I error is better as compared to other test statistics at the 5 percent level.

(e) All the tests are consistent. Most test statistics show a substantial increase in power as sample size increases. With respect to power, T_3 performs better than T_{1B} for $\rho > 0$ but the situation is quite the reverse for admissible negative value of ρ . Wilson and Martin (1983) have got the same conclusion for $\rho > 0$, but they did not consider the cases of negative ρ-values.

3.3 MEASURES OF GOODNESS-OF-FIT IN ACOVS

Having obtained satisfactory estimates of the model parameters, the crucial question that arises in ACOVS is whether or not the hypothesized covariance structure fits adequately the sample covariance matrix ($\underset{\sim}{S}$). For this purpose, if $\hat{\underset{\sim}{\Sigma}}$ is the estimated hypothesized covariance matrix by any suitable method of estimation, the elements of residuals ($\underset{\sim}{S}-\hat{\underset{\sim}{\Sigma}}$) matrix are generally considered. If they are so small

that they can be ascribable to chance errors in the observations, then the fit is called satisfactory. On the other hand, if the elements of $(\underset{\sim}{S} - \hat{\underset{\sim}{\Sigma}})$ matrix are so large, or of such a non-random nature that they cannot be ascribed to chance errors under the structural hypothesis, then it indicates a poor fit of the structure.

The failure to fit a model generally implies that the theoretical model chosen does not reflect adequately the inter-relationships among the observed variables under study. However, whereas a lack of fit or model misspecification is a justification for rejecting the hypothesized structure or at least an indication for modifying the structure, a good fit does not necessarily indicate that the structure is correct (Mukherjee, 1976). A good fit merely establishes that there is no reason to reject the structure on the basis of the data at hand. There is always some possibility of the existence of some more intuitively appealing models which might show a better fit to the given data.

In ACOVS , the classical hypothesis - testing approach suffers from a serious practical limitation because the test of the null hypothesis (H_0) regarding a structure invariably leads to rejection of the H_0 that $\underset{\sim}{\Sigma} = \underset{\sim}{\Sigma}_0$ whenever the sample size is large. Similarly, any structural hypothesis can be readily accepted if the sample size is very small.

In such situations, the measure of goodness-of-fit in ACOVS is supposed to help in determining which structural model best approximates $\underset{\sim}{\Sigma}$ to $\underset{\sim}{\Sigma}_o$ given a sample of specified size. We also ask the question as to what might be a sound theoretical reason for considering a particular structure as more plausible than others. Thus, goodness-of-fit measures show the correspondence between the model and the sample data and thereby help us in determining the extent to which the model with its parameter estimates can reproduce the covariances between the observed variables.

An index measuring goodness-of-fit should be 'sensitive' in the sense that any perceptible departure between the sample covariance matrix $\underset{\sim}{S}$ and the hypothesized population covariance matrix $\underset{\sim}{\Sigma}$ should be reflected in the magnitude of the index. The closer the fit, the higher is the fit index and therefore, less is the model misspecification. With reference to testing of hypothesis, a reasonably sensitive fit index (measured in percentage form) should be very close to 100 whenever the hypothesized structure of $\underset{\sim}{\Sigma}$ resembles the actual structure $\underset{\sim}{\Sigma}_o$, i.e., when $\underset{\sim}{S}$ conforms to $\underset{\sim}{\Sigma}_o$. In other words, the index would be called sensitive if sample covariance matrices drawn from a structured population covariance matrix are identified as having the same structure. Such a sensitive index should be expected to possess also the property of specificity in the sense that whenever the hypothesized structure of $\underset{\sim}{\Sigma}$ is quite

different from $\underset{\sim}{\Sigma}_0$, the index should be very close to zero, thereby showing a very high percentage of the possibility of model misspecification. It is worthwhile to mention here that sensitivity and specificity of a fit index are comparable respectively to one minus the nominal level of significance and the power of a test of significance [cf. Fleiss, 1973, p. 3-13].

While constructing a goodness-of-fit index for use in ACOVS, one should take care of at least the following properties :

(a) Since the overall goodness-of-fit usually is a unidimensional concept, it should be based on a single value for ease of comparison.

(b) The index should be invariant under affine transformation upon the variables.

(c) As far as possible, the index should depend on the amount of departure of the fitted model to sample covariance matrix ($\underset{\sim}{S}$) alone and not on the sample size. The value of the index should not also depend upon the dimension or order of the $\underset{\sim}{S}$-matrix.

(d) The sensitivity and specificity of the index should be quite high and should not be affected by a monotonic transformation of the variables.

(e) The index should have a rational basis of a meaningful zero point and range from 0 to 1. However, for convenience it may be expressed in percentage form.

While, condition (e) is not strictly necessary, it enables us to employ it readily for comparative purposes. However, it is generally possible to perform a suitable transformation on the index to vary within the range of 0 to 1 without changing its order properties.

There are quite a few indices of goodness-of-fit which are currently being used in ACOVS, but not all of them are consistent with the properties mentioned above. A major limitation of most of the currently available goodness-of-fit indices is that one gets the index value approaching unity just by " freezing up more parameters in a model" as has been pointed out by James et al. (1982).

While constructing a fit index, the degree of dissimilarity of $\underset{\sim}{\Sigma}$ to $\underset{\sim}{S}$ is generally given by the " loss" or " discrepancy" function (Browne, 1984, p. 72) which is minimized for obtaining 'consistent' estimates of the parameters involved in $\underset{\sim}{\Sigma}$. Two important functions are given by

$$F(\underset{\sim}{S}, \underset{\sim}{\Sigma}) = \log \det (\underset{\sim}{S}\ \underset{\sim}{\Sigma}^{-1}) + \mathrm{tr}(\underset{\sim}{S}\ \underset{\sim}{\Sigma}^{-1}) - p \ , \qquad (3.3.1)$$

and

$$G(\underset{\sim}{S}, \underset{\sim}{\Sigma}) = \frac{1}{2}\ \mathrm{tr}\left\{(\underset{\sim}{\Sigma} - \underset{\sim}{S})\ \underset{\sim}{V}\right\}^2 \qquad (3.3.2)$$

where $\underset{\sim}{V}$ is a suitable p.d. weight matrix. These would lead

respectively to MLE (under multinormal set-up) and GLSE of the parameters of $\underset{\sim}{\Sigma}$. Under multinormality assumption, the minimum of F function based on MLE of $\underset{\sim}{\Sigma}$, and the minimum G based on GLSE of $\underset{\sim}{\Sigma}$ with $\underset{\sim}{V} = \underset{\sim}{S}^{-1}$ would provide the goodness-of-fit indices [cf. Anderson and Gerbing, 1984; Bentler and Bonett, 1980] in terms of the probability value associated with the chi-square as

$$\text{PRCHI}_Q = P[\ \chi^2_\upsilon \geq (N-1)F(\underset{\sim}{S}, \hat{\underset{\sim}{\Sigma}}_Q)] \qquad (3.3.3)$$

and

$$\text{PRCHI}_R = P[\ \chi^2_\upsilon \geq (N-1)\ G(\underset{\sim}{S}, \hat{\underset{\sim}{\Sigma}}_R)] \qquad (3.3.4)$$

where $\upsilon = p(p+1)/2-k$, and $\hat{\underset{\sim}{\Sigma}}_Q$ and $\hat{\underset{\sim}{\Sigma}}_R$ are the MLE and GLSE of $\underset{\sim}{\Sigma} = \underset{\sim}{\Sigma}(\underline{\theta})$. Although $(N-1)F(\underset{\sim}{S}, \hat{\underset{\sim}{\Sigma}}_Q)$ and $(N-1)G(\underset{\sim}{S}, \hat{\underset{\sim}{\Sigma}}_R)$ would have a common limiting χ^2-distribution which is valid even for a wide class of nonnormal situations, equs. (3.3.3) and (3.3.4) are highly sensitive to sample size (N) [Bentler and Bonett, 1980; Jöreskog, 1978; 1981; Lawley and Maxwell, 1971]. Hence, very often misleading conclusions may be reached for samples of either smaller or larger sizes if these indices are employed. To avoid direct involvement of N, Jöreskog and Sörbom (1981) proposed the goodness-of-fit indices for MLE and OLSE as

$$\text{GFIQ} = 1 - \frac{\text{tr}(\hat{\underset{\sim}{\Sigma}}_Q^{-1}\underset{\sim}{S} - \underset{\sim}{I}_p)^2}{\text{tr}(\hat{\underset{\sim}{\Sigma}}_Q^{-1}\underset{\sim}{S})^2} \qquad (3.3.5)$$

$$GFIU = 1 - \frac{tr(\underset{\sim}{S}-\hat{\underset{\sim}{\Sigma}}_U)^2}{tr\,S^2} \qquad (3.3.6)$$

where $\hat{\underset{\sim}{\Sigma}}_Q$ and $\hat{\underset{\sim}{\Sigma}}_U$ are respectively the MLE and OLSE (ordinary or unweighted LSE of $\underset{\sim}{\Sigma}$. Tanaka and Huba (1985) found out the statistical origin of these fit indices. They also provided the formula for fit index using GLSE ($\hat{\underset{\sim}{\Sigma}}_R$) of $\underset{\sim}{\Sigma}$ as

$$GFIR = 1 - \frac{tr(\hat{\underset{\sim}{\Sigma}}_R S^{-1} - \underset{\sim}{I}_p)^2}{p} \quad . \qquad (3.3.7)$$

Adjusting the degrees of freedom, Jöreskog and Sörbom modified their GFIQ index as

$$AGFIQ = 1 - \frac{p(p+1)}{2v}\,(1 - GFIQ) \quad . \qquad (3.3.8)$$

The indices appearing as (3.3.5) to (3.3.8) are claimed to be independent of N and relatively robust against departure from normality. Although their statistical distributions are still not known, their properties and limitations are discussed in the implementation of LISREL V programme (Jöreskog and Sörbom, 1981). The effects of N upon these indices are primarily studied by Boomsma (1982) and more extensively and elaborately by Anderson and Gerbing (1984) through simulation in the context of confirmatory factor analysis. Mulaik *et al.* (1989) have discussed the inadequacies of AGFI in the context of covariance structure modelling.

Without employing a statistical basis, Bentler and Bonett (1980) suggested a measure, called <u>incremental fit index</u> (ρ), as an alternative to certain type of reliability coefficient proposed by Tucker and Lewis (1973). Their incremental fit index is given by the formula :

$$\rho_{tu} = \frac{U_t - U_u}{U_o} \qquad (3.3.9)$$

where the fit function U_t is either $F(\underset{\sim}{S}, \hat{\underset{\sim}{\Sigma}}_Q)$ or $G(\underset{\sim}{S}, \hat{\underset{\sim}{\Sigma}}_R)$. The suffixes 't' and 'u' refer to two nested models. The suffix 'o' is related to what Bentler and Bonett (1980) called the " null model" . The null model is the simplest, the most <u>restrictive</u> model that may be taken as the standard to which we compare less restrictive models subscripted by 't' and 'u'. La Du and Tanaka (1989) have shown that the normed fit index (NFI) of Bentler and Bonett (1980) " is not a good summary of model fit, particularly when different methods of estimation are being compared" (p. 634). Earlier, Bollen (1986) pointed out that the incremental fit index (ρ) is severely affected by sample size (N). His modified version is given by

$$\rho^*_{tu} = \frac{U_t/\upsilon_t - U_u/\upsilon_u}{U_o/\upsilon_o} \qquad (3.3.10)$$

where the symbol υ refers to the degrees of freedom of the model indicated by subscript following υ . ρ_{tu} may have the possibility to be greater than unity or sometimes negative while ρ^*_{tu} is not so.

The consideration of the residual matrix $(S - \hat{\underset{\sim}{\Sigma}})$ where $\hat{\underset{\sim}{\Sigma}}$ is the estimated $\underset{\sim}{\Sigma}$ by any method of estimation, enables us to suggest a number of different measures of goodness-of-fit. The root-mean-square residual (RMR) is defined by Jöreskog and Sörbom (1981) as

$$RMR = \Big[2 \sum_{\substack{i=1 \\ i \geq j}}^{p} \sum_{j=1}^{p} (s_{ij} - \hat{\sigma}_{ij})^2 / p(p+1)\Big]^{1/2} . \qquad (3.3.11)$$

Another similar measure as used by Werts *et al.* (1981) is defined in terms of the average deviation as

$$AD = \frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{p} \left| s_{ij} - \hat{\sigma}_{ij} \right| \quad . \qquad (3.3.12)$$

Mukherjee and Maiti (1988b) have considered a modified version of AD, called average relative deviation which is defined as

$$ARD = \frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{p} \left| \frac{s_{ij} - \hat{\sigma}_{ij}}{s_{ij}} \right| \qquad (3.3.13)$$

This measure is meaningful provided the s_{ij}'s are non-zero for all (i,j) elements of $\underset{\sim}{S}$ matrix.

Marsh, Balla and McDonald (1988), as well as McDonald (1989) have suggested fit indices in terms of suitable non-centrality parameters. These fit indices appear to be consistent over variations in sample size.

3.3.1 Distributional Properties of Jöreskog and Sörbom Fit Indices

Jöreskog and Sörbom (1981, 1982, 1988) discussed only the properties and limitations of their fit indices GFIQ and AGFIQ [vide equs.(3.3.5) and (3.3.8)] in the implementation of the LISREL program, but their sampling distribution could not be derived " even under idealized assumptions" (Jöreskog and Sörbom, 1982, p. 409). We shall derive their asymptotic distributions and discuss some distributional properties in the context of a wide class ($\mathcal{C}$) of both linearly and nonlinearly structured covariance matrices whose MLE satisfies $tr(\hat{Q}\,\hat{\Sigma}) = 0$ [vide Remark 3.2]. When $\underset{\sim}{\Sigma} \in \mathcal{C}$, GFIQ reduces to the simpler form given by

$$g_o = \frac{p}{p - tr(\hat{Q}\,S)} \,. \qquad (3.3.14)$$

We may note here that $tr(\hat{Q}\,S)$ is involved in the ACOVS version of Rao's efficient score criterion expressed in equation (3.2.29). Recalling this equation, the asymptotic distribution of g_o is given by

$$\text{Prob}\,[g_o \leq g_o] = \text{Prob}\,[\chi^2_\upsilon \geq \chi^2_o] \qquad (3.3.15)$$

where $$\chi^2_o = \frac{Np}{2}\left(\frac{1}{g_o} - 1\right). \qquad (3.3.16)$$

Since $AGFIQ = 1 - \frac{p(p+1)}{2\upsilon}(1 - GFIQ)$, the asymptotic distribution

of AGFIQ , when $\underset{\sim}{\Sigma} \in \mathcal{C}$, is similarly given by

$$\text{Prof}\,[\,a_o < a_o] = \text{Prob}\,[\chi^2_\upsilon \geq \chi^{*2}_o\,] \qquad (3.3.17)$$

where $$\chi^{*2}_o = \frac{Np}{2}\left[\,\frac{p(p+1)}{2\upsilon(1-a_o)} - 1\,\right]. \qquad (3.3.18)$$

Thus, both the GFIQ and AGFIQ are exact functions of the χ^2-variable.

Involvement of N in (3.3.16) or (3.3.18) provides a clear insight into the dependence of the distribution of GFIQ or AGFIQ on sample size. The fact that GFIQ is not independent of sample size was previously noted by a number of authors such as Bearden, Sharma and Teel (1982), Hoelter (1983), Anderson and Gerbing (1984), as well as Marsh, Balla and McDonald (1988) on the basis of different Monte Carlo experiments. Although Jörekog and Sörbom (1982, p. 408) stated that " unlike χ^2, GFIQ is independent of the sample size" , they meant that GFIQ does not depend upon N in its calculation. The sampling distribution of both GFIQ and AGFIQ must, of course, depend on N since the distribution of any function of sample moments (covariances in this case) is dependent on sample size. This distribution was not then known but has now been obtained for covariance structures in class $\mathcal{C}$.

Since the fit indices are usually intended for use in situations where a model serves as an approximation, not

necessarily when it is correct and the null hypothesis holds, we are also interested in the non-null distribution of g_o . When $\underset{\sim}{\Sigma} = \underset{\sim}{\Sigma}(\underline{\theta})$ (the null model) is not true, let $\underset{\sim}{\Sigma}_o$ be the true population covariance matrix and let $\hat{\underset{\sim}{\Sigma}}_o$ be the reproduced covariance matrix obtained if the model were fitted to $\underset{\sim}{\Sigma}_o$ by ML method. Since the model is not assumed to fit, there is no assumption that $\hat{\underset{\sim}{\Sigma}}_o = \underset{\sim}{\Sigma}_o$.

$$\text{Let} \quad \hat{f} = \frac{1}{2} \operatorname{tr} [\hat{\underset{\sim}{Q}} (\hat{\underset{\sim}{\Sigma}} - \underset{\sim}{S})]$$

$$\text{and} \qquad f_o = -\frac{1}{2} \operatorname{tr} [\underset{\sim}{U} (\hat{\underset{\sim}{\Sigma}}_o - \underset{\sim}{\Sigma}_o)] \; . \qquad (3.3.19)$$

$$\text{where} \qquad \underset{\sim}{U} = \hat{\underset{\sim}{\Sigma}}_o^{-1} - \hat{\underset{\sim}{\Sigma}}_o^{-1} \underset{\sim}{\Sigma}_o \hat{\underset{\sim}{\Sigma}}_o^{-1} \; . \qquad (3.3.20)$$

Then, if N is large, the distribution of Nf [i.e. of T_2 , vide equ. (3.2.27)] is approximately noncentral chi-square with υ d.f. and noncentrality parameter Nf_o [cf. Browne, 1984, Cor. 4.1]. In this connection, we refer to the recent results of McDonald (1989) as well as McDonald and Marsh (1989) showing how a number of goodness-of-fit indices may be expressed as functions of their corresponding noncentrality parameters. Under non-null situations, $E(\hat{f}) \approx f_o + \frac{\upsilon}{N}$ and $\operatorname{var}(\hat{f}) \approx \frac{2}{N} (\frac{\upsilon}{N} + 2f_o)$ [cf. Kendall and Stuart, 1967, p.229].

It is worthwhile to examine here the biasedness of the sample values of GFIQ for $\underset{\sim}{\Sigma} \in \mathcal{C}$. We observe that

$$E\ (g_c) \approx \frac{p}{p+E\ [-tr(\hat{Q}\ S)]}$$

$$\approx \frac{p}{p + 2f_o + \frac{2}{N}\ v} \qquad (3.3.21)$$

while the population value of GFIQ is given by

$$\gamma_{oc} = \frac{p}{p + 2f_o} \qquad (3.3.22)$$

so that

$$E(g_o) \leq \gamma_{oc} \qquad (3.3.23)$$

Thus, g_c is a consistent but biased estimator of γ_{oc}.

Obviously, γ_{oc} is low for a poor fit of the model, and it increases toward 1 as goodness-of-fit improves. A bias-corrected estimator of γ_{oc} is given by

$$\widetilde{g_c} = \frac{p}{p-tr(\hat{Q}S) - \frac{2v}{N}}$$

$$= 1 - \frac{tr(\hat{\Sigma}^{-1}S - I_p)^2 - \frac{2v}{N}}{tr(\hat{\Sigma}^{-1}S)^2 - \frac{2v}{N}} \qquad (3.3.24)$$

Because $tr(\hat{\Sigma}^{-1}S)^2 > \frac{tr^2(\hat{\Sigma}^{-1}S)}{p}$, the biased estimator g_c lies in the interval $[\frac{1}{p}, 1]$. But the (approximately) unbiased estimator $\widetilde{g_c}$ can assume values outside this interval when $tr(\hat{Q}\ S)$ is either greater than $-\frac{2v}{N}$, or

less than $-\left[\frac{2\upsilon}{N} + p(p-1)\right]$. However, for practical purposes, $\widetilde{g}_c$ can be used as a measure of goodness-of-fit [see McDonald, 1989].

As in the case of GFIQ, we observe that the approximately unbiased estimator, $\widetilde{a}_c$, obtainable from the biased estimator AGFIQ, may even assume (apart from values greater than unity) negative values when $tr(\hat{\underset{\sim}{Q}}\ \underset{\sim}{S})$ is less than

$-\left[\frac{2\upsilon}{N} + \frac{2\upsilon\, p}{p(p+1)-2\upsilon}\right]$, although the population AGFIQ never assumes negative values. Since the sampled value of AGFIQ like the correction for bias of the squared multiple correlation (SMC), can take on negative values, this is regarded as one of its inadequate features (Mulaik et al.,1989).

The GFIQ or AGFIQ is analogous to the squared multiple correlation coefficient (SMC) [Tanaka and Huba, 1985]. It is interesting that the sample GFIQ (or AGFIQ) is a negatively biased estimator of the population GFIQ (or AGFIQ), whereas the sample SMC is a positively biased estimator of the population SMC (cf. Kendall and Stuart, 1967, p. 341).

When $tr(\hat{\underset{\sim}{Q}}\ \underset{\sim}{S})$ is small, an approximation of g_c can be given by a first-order Taylor series as

$$g_c \approx 1 + \frac{tr(\hat{\underset{\sim}{Q}}\ \underset{\sim}{S})}{p} \qquad (3.3.25)$$

which is greater than $1 + \frac{tr(\hat{\underset{\sim}{Q}}\ \underset{\sim}{S})}{2}$ so that GFI (g_c) gives

a more optimistic impression of the closeness of the model fit than does the residual quadratic form $\frac{1}{2}\ tr(\hat{\underset{\sim}{\Sigma}}^{-1}\underset{\sim}{S}-\underset{\sim}{I}_p)^2$.

The first order approximation (3.3.25) of g_c is comparable to the fit index (g^*) for the GLSE ($\hat{\Sigma}_R$) which is given by (Tanaka and Huba, 1985) :

$$g^* = 1 - \frac{\hat{f}^*}{p/2} \qquad (3.3.26)$$

where $\hat{f}^* = \frac{1}{2}\ tr(\underset{\sim}{I}_p - \underset{\sim}{S}^{-1}\ \hat{\underset{\sim}{\Sigma}}_R)^2$ and $\hat{\Sigma}_R$ is obtained by minimizing the GLS fit function with weight matrix $\underset{\sim}{S}^{-1}$ [vide Ch.2, sec. 2].

3.3.2 Some Matrix Results on $\underset{\sim}{Q}$ and $\underset{\sim}{R}$ Matrices

To derive some simple matrix results on $\underset{\sim}{Q} = \underset{\sim}{\Sigma}^{-1} - \underset{\sim}{\Sigma}^{-1}\underset{\sim}{S}\underset{\sim}{\Sigma}^{-1}$ and $\underset{\sim}{R} = \underset{\sim}{S}^{-1} - \underset{\sim}{S}^{-1}\ \underset{\sim}{\Sigma}\underset{\sim}{S}^{-1}$ (as defined in Ch. 2, sec. 2.1) we shall first state some lemmas whose proofs are either simple or available elsewhere.

Lemma 3.3.1 : Let $\underset{\sim}{A}(p \times p)$ be such that all its eigenvalues are positive and $tr(\underset{\sim}{A}) = p$. Then $0 < \det \underset{\sim}{A} \leq 1$.

Hint : Apply the result stating a.m. $\geq$ g.m. upon the eigenvalues of $\underset{\sim}{A}$.

Lemma 3.3.2 : Let $\underset{\sim}{A}(p \times p)$ be such that all its eigenvalues are positive. Then (i) $tr(\underset{\sim}{A}^2) \in [(tr\underset{\sim}{A})^2/p, (tr\underset{\sim}{A})^2]$ and (ii) $tr\underset{\sim}{A} > p^2/tr(\underset{\sim}{A}^{-1})$.

Hint : (i) Let λ_i (> 0), $i = 1, \cdots, p$ be the eigenvalues of $\underset{\sim}{A}$. Then the inequality $(\sum_{i=1}^{p} \lambda_i)^2/p < \sum_{i=1}^{p} \lambda_i^2 < (\sum_{i=1}^{p} \lambda_i)^2$ may be applied to complete the proof.

(ii) Apply a.m. > h.m. result upon λ_i's . ■

Lemma 3.3.3 : For two p.d. matrices $\underset{\sim}{A}$ and $\underset{\sim}{B}$, $\underset{\sim}{A}\underset{\sim}{B}$ would have positive eigenvalues.

Hint : See [Ikramov (1983, p. 211) in this connection.

Lemma 3.3.4 : (Weyl's inequality)

For any two symmetric matrices $\underset{\sim}{A}(p \times p)$ and $\underset{\sim}{B}(p \times p)$ with eigenvalues $\lambda_1(\underset{\sim}{A}) \geq \lambda_2(\underset{\sim}{A}) \geq \cdots \geq \lambda_p(\underset{\sim}{A})$ and $\lambda_1(\underset{\sim}{B}) \geq \lambda_2(\underset{\sim}{B}) \geq \cdots \geq \lambda_p(\underset{\sim}{B})$, then

$$\lambda_1(\underset{\sim}{A}) + \lambda_p(\underset{\sim}{B}) \leq \lambda_1(\underset{\sim}{A} + \underset{\sim}{B}) \leq \lambda_1(\underset{\sim}{A}) + \lambda_1(\underset{\sim}{B}).$$

In particular, for positive semi-definite $\underset{\sim}{B}$, $\lambda_i(\underset{\sim}{A}+\underset{\sim}{B}) \geq \lambda_i(\underset{\sim}{A})$, $i = 1,2, \cdots, p$.

Hint : See Franklin (1968, p. 157) for the proof.

As regards bounds of the eigenvalues of $\underset{\sim}{Q}$, we may use lemma 3.3.4 by putting $\underset{\sim}{A} = \underset{\sim}{Q}$ and $\underset{\sim}{B} = \underset{\sim}{\Sigma}^{-1}\underset{\sim}{S}\underset{\sim}{\Sigma}^{-1}$ and get the following.

Proposition 3.3.1 :

(i) $\lambda_1(\underset{\sim}{\Sigma}^{-1}) - \lambda_1(\underset{\sim}{\Sigma}^{-1}\underset{\sim}{S}\,\underset{\sim}{\Sigma}^{-1}) \leq \lambda_1(\underset{\sim}{Q}) \leq \lambda_1(\underset{\sim}{\Sigma}^{-1}) - \lambda_p(\underset{\sim}{\Sigma}^{-1}\underset{\sim}{S}\,\underset{\sim}{\Sigma}^{-1})$

(ii) $\underline{\lambda_i(\underset{\sim}{Q})\ \lambda_{p-i+1}\ (\underset{\sim}{\Sigma}) \leq 1}$

$\underline{i = 1, 2, \ldots p.}$ In particular, $\underline{\lambda_{max}(\underset{\sim}{Q}) \leq 1/\lambda_{min}(\underset{\sim}{\Sigma}).}$

Remark 3.6 : In the version of spectral norm [cf. Franklin, 1968, p. 159] defined by

$$\|\underset{\sim}{P}\|_\lambda = \max\left\{\left|\lambda_{max}(\underset{\sim}{P})\right|,\ \left|\lambda_{min}(\underset{\sim}{P})\right|\right\} \qquad (3.3.27)$$

for a symmetric matrix $\underset{\sim}{P}$,

$$\|\underset{\sim}{Q}\|_\lambda \leq \|\underset{\sim}{\Sigma}^{-1}\|_\lambda . \qquad (3.3.28)$$

Thus $\underset{\sim}{Q}$ would have finite spectral norm if $\underset{\sim}{\Sigma}$ is not " ill-conditioned" .

Proposition 3.3.2 : Given the p.d. matrix $\underset{\sim}{S}(p \times p)$, let $\underset{\sim}{\Sigma}$ (p x p) be so chosen, say $\hat{\underset{\sim}{\Sigma}}_Q$, that its corresponding $\underset{\sim}{Q}$, say $\hat{\underset{\sim}{Q}}$, satisfies the equation $\mathrm{tr}(\hat{\underset{\sim}{\Sigma}}_Q\ \hat{\underset{\sim}{Q}}) = 0$. Then

(i) $\mathrm{tr}(\hat{\underset{\sim}{\Sigma}}_Q\ \underset{\sim}{S}^{-1}) \geq p$, (ii) $\det \hat{\underset{\sim}{\Sigma}}_Q \geq \det \underset{\sim}{S}$ and (iii) $\|\hat{\underset{\sim}{\Sigma}}_Q\|_\lambda \geq \|\underset{\sim}{S}\|_\lambda$.

Proof: (i) $\mathrm{tr}(\hat{\underset{\sim}{\Sigma}}_Q\ \hat{\underset{\sim}{Q}}) = 0 \Rightarrow \mathrm{tr}(\hat{\underset{\sim}{\Sigma}}_Q^{-1}\ \underset{\sim}{S}) = p$. As $\hat{\underset{\sim}{\Sigma}}_Q^{-1}$ and $\underset{\sim}{S}$ are p.d. matrices, lemmas 3.3.2 and 3.3.3 may be applied to complete proof.

(ii) Apply lemmas 3.3.1 and 3.3.3.

(iii) It is simple to derive from (ii). ■

Recalling the definition of $\underset{\sim}{R}$, Proposition 3.3.1 may be similarly established for $\underset{\sim}{R}$ but with the substitutions of $\underset{\sim}{\Sigma}$ by $\underset{\sim}{S}$, $\underset{\sim}{S}$ by $\underset{\sim}{\Sigma}$ and $\underset{\sim}{Q}$ by $\underset{\sim}{R}$.

From the very definition of $\underset{\sim}{Q}$ and $\underset{\sim}{R}$, we note that

$$\underset{\sim}{\Sigma}\ \underset{\sim}{Q}\ \underset{\sim}{\Sigma} + \underset{\sim}{S}\ \underset{\sim}{R}\ \underset{\sim}{S} = \underset{\sim}{Q} . \qquad (3.3.29)$$

Since $\underset{\sim}{\Sigma}^{-1}\underset{\sim}{S}$ (say, $\underset{\sim}{T}$) is nonsingular (n.s.) and $\underset{\sim}{Q} = -\ \underset{\sim}{T}\underset{\sim}{R}\underset{\sim}{T}'$ or $\underset{\sim}{R} = -\ \underset{\sim}{T}^{-1}\ \underset{\sim}{Q}\underset{\sim}{T}'^{-1}$, $\underset{\sim}{Q}$ and $-\underset{\sim}{R}$ (or $-\ \underset{\sim}{Q}$ and $\underset{\sim}{R}$) may be said to be congruent matrices. When det $\underset{\sim}{Q} = 0$, det $\underset{\sim}{R} = 0$ and vice versa. In the case of n.s. $\underset{\sim}{Q}$ and $\underset{\sim}{R}$

$$\det \underset{\sim}{Q}/\det \underset{\sim}{R} = (-1)^{p}\ (\det \underset{\sim}{S}/\det \underset{\sim}{\Sigma})^{2} . \qquad (3.3.30)$$

Due to congruence, rank ($\underset{\sim}{Q}$) = rank ($\underset{\sim}{R}$) = rank ($\underset{\sim}{\Sigma} - \underset{\sim}{S}$). As regards the norm of these matrices, it may be noticed that when the matrix norm $\|\underset{\sim}{Q}\|$ of $\underset{\sim}{Q}$ equals 0 , $\|\underset{\sim}{R}\|$ is also implied to be 0 and vice versa.

It is interesting further to observe that the four matrices ($\underset{\sim}{Q}\ \underset{\sim}{\Sigma}$, $\underset{\sim}{Q}\ \underset{\sim}{S}$, $\underset{\sim}{R}\ \underset{\sim}{\Sigma}$, $\underset{\sim}{R}\ \underset{\sim}{S}$) are commutative with respect to multiplication. Given $\underset{\sim}{S}$ and $\underset{\sim}{\Sigma}$, $\underset{\sim}{Q}$ and $\underset{\sim}{R}$ may be treated as a pair of parallel matrices for their expressional similarities as well as for the above-mentioned properties.

Applying any matrix norm $\|\cdot\|$ upon $\underset{\sim}{Q} = -\underset{\sim}{T}\underset{\sim}{R}\underset{\sim}{T}'$ and $\underset{\sim}{R} = -\ \underset{\sim}{T}^{-1}\ \underset{\sim}{Q}\underset{\sim}{T}'^{-1}$ with $\underset{\sim}{T} = \underset{\sim}{\Sigma}^{-1}\underset{\sim}{S}$, we arrive at the following propositions, the proof of some of which are quite easy and

and therefore omitted.

Proposition 3.3.3 : For non-null matrices $\underset{\sim}{Q}$ and $\underset{\sim}{R}$, $\|\underset{\sim}{Q}\| / \|\underset{\sim}{R}\| \in [1/(\|\underset{\sim}{S}^{-1}\| \|\underset{\sim}{\Sigma}\|)^2, (\|\underset{\sim}{\Sigma}^{-1}\| \|\underset{\sim}{S}\|)^2]$. In particular for spectral norm,

$$\|\underset{\sim}{Q}\|_\lambda / \|\underset{\sim}{R}\|_\lambda \in [\lambda^2_{min}(\underset{\sim}{S})/\lambda_{max}(\underset{\sim}{\Sigma}), \lambda^2_{max}(\underset{\sim}{S})/\lambda^2_{min}(\underset{\sim}{\Sigma})] \ .$$

Proposition 3.3.4 : Given p.d. matrix $\underset{\sim}{S}(p \times p)$, let $\underset{\sim}{\Sigma}(p \times p)$ be so chosen, say $\hat{\underset{\sim}{\Sigma}}_R$, that the corresponding $\underset{\sim}{R}$, say $\hat{\underset{\sim}{R}}$, satisfies the equation $tr(\hat{\underset{\sim}{\Sigma}}_R \hat{\underset{\sim}{R}}) = 0$. Then (i) $tr(\hat{\underset{\sim}{\Sigma}}_R \underset{\sim}{S}^{-1}) \in (0,p)$, (ii) $\det \hat{\underset{\sim}{\Sigma}}_R \leq \det \underset{\sim}{S}$ and (ii) $\|\hat{\underset{\sim}{\Sigma}}_R\|_\lambda \leq \|\underset{\sim}{S}\|_\lambda$.

Proof : From $tr(\hat{\underset{\sim}{\Sigma}}_R \hat{\underset{\sim}{R}}) = 0$, $tr(\hat{\underset{\sim}{\Sigma}}_R \underset{\sim}{S}^{-1})^2 = tr(\hat{\underset{\sim}{\Sigma}}_R \underset{\sim}{S}^{-1})$.

Due to lemma 3.3.3, eigenvalues of $\hat{\underset{\sim}{\Sigma}}_R \underset{\sim}{S}^{-1}$, say δ_i , $i = 1$, ... , p are all positive so that their average $\bar{\delta} > 0$ and hence $\sum_{i=1}^{p} (\delta_i - \bar{\delta})^2 \geq 0$ implies that $\bar{\delta} \in (0,1)$ which would complete the proof.

(ii) and (iii) may be proved similar to proving (ii) and (iii) of Proposition 3.3.2.

Remark 3.7 By combining Propositions 3.3.2 and 3.3.4, we get the following inequalities :

$$(i)\ tr(\hat{\underset{\sim}{\Sigma}}_R \underset{\sim}{S}^{-1}) \leq p \leq tr(\hat{\underset{\sim}{\Sigma}}_Q \underset{\sim}{S}^{-1}) \qquad (3.3.31)$$

(ii) $\det \hat{\underset{\sim}{\Sigma}}_R \leq \det \underset{\sim}{S} \leq \det \hat{\underset{\sim}{\Sigma}}_Q$ (3.3.32)

(iii) $\|\hat{\underset{\sim}{\Sigma}}_R\|_\lambda \leq \|\underset{\sim}{S}\|_\lambda \leq \|\hat{\underset{\sim}{\Sigma}}_Q\|_\lambda$. (3.3.33)

3.3.3 The Notion of Structural Closeness

Let us consider the spherical structure $\sigma^2 \underset{\sim}{I}_p$, as the hypothesized structure of $\underset{\sim}{\Sigma}$. The estimates of σ^2 by ML and GLS method are given by

$$\hat{\sigma}^2_Q = (\mathrm{tr}S)/p \ , \ \hat{\sigma}^2_R = \mathrm{tr}(S^{-1})/\mathrm{tr}(S^{-2}) \quad (3.3.34)$$

so that from (3.3.5) and (3.3.7),

$$GFIQ = (\mathrm{tr}\underset{\sim}{S})^2/[p\ \mathrm{tr}(\underset{\sim}{S}^2)], \ GFIR = (\mathrm{tr}\underset{\sim}{S}^{-1})^2/[p\mathrm{tr}(\underset{\sim}{S}^{-2})]. \quad (3.3.35)$$

As the above expressions are functionally different, the amount of " goodness" as measured by MLE (Q-procedure) and by GLSE (R-procedure) will generally differ so that sometimes conflicting decisions about the same hypothesized structure may be arrived at. For example, Tanaka (1987) provided an empirical illustration in which the Bentler-Bonett Normed Fit Index (NFI) applied to the same data and model was affected by choice of estimator. In general, it is thus clear that the consideration of only one particular method of estimation of the structured $\underset{\sim}{\Sigma}$ in deriving a measure of " fit " may not give us the complete picture about the closeness.

So long as we are interested in the closeness of given hypothesized structure of $\underset{\sim}{\Sigma}$ to $\underset{\sim}{S}$, we should take into consideration the structural behaviour of S in stead of caring much about the magnitude of the sample covariance and/or the statistical significance of the unknown parameters so estimated. The detection of the best structure requires, in a sense, the study of the 'quality' not the 'quantity' present in $\underset{\sim}{S}$. Alternatively, we need the study of " closeness" in stead of " goodness" of a structure to $\underset{\sim}{S}$. From such a view-point, given a number of structures simultaneously hypothesized for $\underset{\sim}{\Sigma}$, the best structure is that which is structurally closest to $\underset{\sim}{S}$. In this connection, we shall reproduce a result from Mukherjee and Maiti (1988d, Result 2) in the form of a lemma.

Lemma 3.3.5 : If $\underset{\sim}{S}$ has the same structure as a linearly structured $\underset{\sim}{\Sigma}$, the MLE , OLSE and GLSE with any p.d. weight matrix ($\underset{\sim}{V}$) , for $\underset{\sim}{\Sigma}$, are identically equal to $\underset{\sim}{S}$.

A $\underset{\sim}{\Sigma}$-matrix which is structurally different from $\underset{\sim}{S}$, would have MLE , OLSE and GLSE different from each other and also different from $\underset{\sim}{S}$. From the technical point of view, we shall adjudge the structural closeness of $\underset{\sim}{S}$ to $\underset{\sim}{\Sigma}$ by using two distinctly different methods (in stead of a single one) of estimation and exploit certain common property present in both the estimates. We shall now construct two important

properties of the estimates obtained by Q-procedure and R-procedure assuming a linearly structured $\underset{\sim}{\Sigma}$ as expressed in Ch. 2, equ. (2.1.1).

3.3.4 How to choose the best fitting structure

Two linearly structured $\underset{\sim}{\Sigma}$-matrices will be called structurally equivalent when both of them would have equal number of parameters and the same design matrices. Given a number of competing (structurally different) hypothesized $\underset{\sim}{\Sigma}$-matrices, we shall now deal with the problem of ascertaining how good is each one to $\underset{\sim}{S}$ in terms of structural closeness by utilizing both the Q-procedure and R-procedure.

Quite trivially when the number of unknown parameters (k) is equal to $\binom{p+1}{2}$, the same estimate $\underset{\sim}{S}$ will be obtained by both the Q-and R-procedure as is seen from the respective normal equations [vide Ch. 2, equs. (2.2.3) and (2.6.5)]. Obviously in such a case, $\hat{\underset{\sim}{\Sigma}}_Q = \hat{\underset{\sim}{\Sigma}}_R = \underset{\sim}{S}$ so that $\mathrm{tr}(\hat{\underset{\sim}{\Sigma}}_Q \underset{\sim}{S}^{-1}) = \mathrm{tr}(\hat{\underset{\sim}{\Sigma}}_R \underset{\sim}{S}^{-1}) = p$ and $\det \hat{\underset{\sim}{\Sigma}}_Q = \det \hat{\underset{\sim}{\Sigma}}_R = \det \underset{\sim}{S}$.

Are there any non-trivial situations where such equalities on traces or on determinants (vide Remark 3.7) might hold true ? The answer is in the affirmative as can be seen from the following propositions.

Proposition 3.3.5 : $\mathrm{tr}(\hat{\underset{\sim}{\Sigma}}_Q \underset{\sim}{S}^{-1}) = p \iff S$ has the same structure as that of $\underset{\sim}{\Sigma}$.

Proof : By construction of normal equations (2.2.3) [vide Ch. 2],

$$tr(\underset{\sim}{H}_i \hat{\underset{\sim}{\Sigma}}_Q^{-1} \underset{\sim}{S} \hat{\underset{\sim}{\Sigma}}_Q^{-1}) = tr(\underset{\sim}{H}_i \hat{\underset{\sim}{\Sigma}}_Q^{-1}), \quad i = 1, \ldots, k.$$

Multiplying both sides by $\hat{\theta}_{iQ}$, the MLE of θ_i and summing over i ,

$$tr(\hat{\underset{\sim}{\Sigma}}_Q^{-1} \underset{\sim}{S}) = p = tr(\hat{\underset{\sim}{\Sigma}}_Q \underset{\sim}{S}^{-1}) . \qquad (3.3.36)$$

As the eigenvalues of $\hat{\underset{\sim}{\Sigma}}_Q^{-1} \underset{\sim}{S}$ are the reciprocals of those of $\hat{\underset{\sim}{\Sigma}}_Q \underset{\sim}{S}^{-1}$ (which are all positive), equ. (3.3.36) implies that their a.m. is equal to their h.m. so that all the eigenvalues are equal to unity.

Let $\hat{\underset{\sim}{\Sigma}}_Q^{1/2}$ and $\underset{\sim}{S}^{1/2}$ be the square root matrices of $\hat{\underset{\sim}{\Sigma}}_Q$ and $\underset{\sim}{S}$ respectively. i.e., $\hat{\underset{\sim}{\Sigma}}_Q = \hat{\underset{\sim}{\Sigma}}_Q^{1/2} \hat{\underset{\sim}{\Sigma}}_Q^{1/2}$ and $\underset{\sim}{S} = \underset{\sim}{S}^{1/2} \underset{\sim}{S}^{1/2}$. Then it is easy to show from the characteristic equation of $\hat{\underset{\sim}{\Sigma}}_Q \underset{\sim}{S}^{-1}$ that the eigenvalues of the symmetric matrix $(\hat{\underset{\sim}{\Sigma}}_Q^{1/2} \underset{\sim}{S}^{-1/2})(\hat{\underset{\sim}{\Sigma}}_Q^{1/2} \underset{\sim}{S}^{-1/2})'$ are all equal to unity. Hence, there exists an orthogonal matrix, say $\underset{\sim}{P}$, such that

$$\underset{\sim}{P}(\hat{\underset{\sim}{\Sigma}}_Q^{1/2} \underset{\sim}{S}^{-1/2})(\hat{\underset{\sim}{\Sigma}}_Q^{1/2} \underset{\sim}{S}^{-1/2})' \underset{\sim}{P}' = \underset{\sim}{I}_p$$

implying that $\underset{\sim}{S} = \hat{\underset{\sim}{\Sigma}}_Q = \sum_{i=1}^{k} \underset{\sim}{H}_i \hat{\theta}_{iQ}$. Since $\underset{\sim}{H}_i$'s are linearly independent, $\underset{\sim}{S}$ is structurally similar to $\underset{\sim}{\Sigma}$.

Conversely, when $\underset{\sim}{S}$ has a structure similar to $\underset{\sim}{\Sigma}$, say $\underset{\sim}{S} = \sum_{i=1}^{k} \underset{\sim}{H}_i u_i$, then $tr(\underset{\sim}{Q} \underset{\sim}{H}_i) = 0$ and this implies that

$$tr(\hat{\underset{\sim}{\Sigma}}_Q^{-1} \underset{\sim}{H}_i) = tr(\hat{\underset{\sim}{\Sigma}}_Q^{-1} \underset{\sim}{S} \hat{\underset{\sim}{\Sigma}}_Q^{-1} \underset{\sim}{H}_i)$$

$$\text{i.e,} \quad tr(\hat{\underset{\sim}{\Sigma}}_Q^{-1} \underset{\sim}{H}_i u_i) = tr(\hat{\underset{\sim}{\Sigma}}_Q^{-1} \underset{\sim}{S} \hat{\underset{\sim}{\Sigma}}_Q^{-1} \underset{\sim}{H}_i u_i). \qquad (3.3.37)$$

Summing over i ,

$$tr(\hat{\underset{\sim}{\Sigma}}_Q^{-1}\ \underset{\sim}{S}) = tr(\hat{\underset{\sim}{\Sigma}}_Q^{-1}\ \underset{\sim}{S})^2 \ . \qquad (3.3.38)$$

Similarly, by replacing u_1 by $\hat{\theta}_{iQ}$ in (3.3.37) for all i = 1, ... , k and summing over i ,

$$tr(\hat{\underset{\sim}{\Sigma}}_Q^{-1}\ \underset{\sim}{S}) = p \ . \qquad (3.3.39)$$

Due to (3.3.38) and (3.3.39), the eigenvalues of $\hat{\underset{\sim}{\Sigma}}_Q^{-1}\underset{\sim}{S}$ are all equal to unity. Hence, the eigenvalues of $\hat{\underset{\sim}{\Sigma}}_Q\underset{\sim}{S}^{-1}$, being the reciprocal of those of $\hat{\underset{\sim}{\Sigma}}_Q^{-1}\ \underset{\sim}{S}$, are all equal to unity so that $tr(\hat{\underset{\sim}{\Sigma}}_Q\underset{\sim}{S}^{-1}) = p$. Therefore, in the case of linear structures, $\underset{\sim}{S}$ will have identically the same structure as the hypothesized $\underset{\sim}{\Sigma}$ whenever $tr(\hat{\underset{\sim}{\Sigma}}_Q\underset{\sim}{S}^{-1}) = p = tr(\hat{\underset{\sim}{\Sigma}}_Q^{-1}\underset{\sim}{S})$.

Proposition 3.3.6 : $tr(\hat{\underset{\sim}{\Sigma}}_R\underset{\sim}{S}^{-1}) = p \iff \underset{\sim}{S}$ has the same structure as that of $\underset{\sim}{\Sigma}$.

Proof : By construction of normal equation (2.6.5) [vide Ch.2],

$$tr(\underset{\sim}{H}_1\hat{\underset{\sim}{R}}) = 0 \ , \ i = 1,2, \ldots , k \qquad (3.3.40)$$

$$\text{i.e.,} \quad tr(\hat{\underset{\sim}{\Sigma}}_R\ \underset{\sim}{S}^{-1})^2 = tr(\hat{\underset{\sim}{\Sigma}}_R\underset{\sim}{S}^{-1}) = p \quad \text{(by assumption).} \qquad (3.3.41)$$

From (3.3.41), all the eigenvalues of $\hat{\underset{\sim}{\Sigma}}_R\underset{\sim}{S}^{-1}$ are equal to unity. Following the same steps as used in Proposition 3.3.5, we may now prove that $\hat{\underset{\sim}{\Sigma}}_R\underset{\sim}{S}^{-1} = \underset{\sim}{I}_p$ and also that $\underset{\sim}{S} = \hat{\underset{\sim}{\Sigma}}_R = \sum_{i=1}^{k} \underset{\sim}{H}_i\hat{\theta}_{iR}$ to assert that $\underset{\sim}{S}$ is structurally the same as $\underset{\sim}{\Sigma}$.

Conversely, when S has a structure similar to $\underset{\sim}{\Sigma}$, say $\underset{\sim}{S} = \sum_{i=1}^{k} \underset{\sim}{H}_i u_i$, then $tr(\underset{\sim}{\hat{R}} \underset{\sim}{H}_i) = 0$ will imply that

$$tr(\underset{\sim}{H}_i \underset{\sim}{S}^{-1}) = tr(\underset{\sim}{H}_i \underset{\sim}{S}^{-1} \underset{\sim}{\hat{\Sigma}}_R \underset{\sim}{S}^{-1})$$

i.e., $$tr(\underset{\sim}{H}_i \underset{\sim}{S}^{-1} u_i) = tr(\underset{\sim}{H}_i \underset{\sim}{S}^{-1} \underset{\sim}{\hat{\Sigma}}_R \underset{\sim}{S}^{-1} u_i).$$

Summing over i, we obtain

$$tr(\underset{\sim}{S}^{-1} \underset{\sim}{S}) = tr(\underset{\sim}{S}^{-1} \underset{\sim}{\hat{\Sigma}}_R \underset{\sim}{S}^{-1})$$

or $$tr(\underset{\sim}{\hat{\Sigma}}_R \underset{\sim}{S}^{-1}) = p \ .$$

Remark 3.8 The Propositions 3.3.5 and 3.3.6 would also show that the fit indices GFIQ and GFIR [vide equs. (3.3.5) and (3.3.7)] in the case of linearly structured $\underset{\sim}{\Sigma}$, would attain unity when and only when $\underset{\sim}{S}$ has the same structure as $\underset{\sim}{\Sigma}$.

3.3.5 The Proposed Measures of Structural Closeness

The trace function $tr(\underset{\sim}{\Sigma} \underset{\sim}{S}^{-1})$ for a given hypothesized of $\underset{\sim}{\Sigma}$ may be estimated by Q-procedure and R-procedure respectively as $tr(\underset{\sim}{\hat{\Sigma}}_Q S^{-1})$ and $tr(\underset{\sim}{\hat{\Sigma}}_R S^{-1})$. These trace functions are respectively greater and less than p . We find that $tr(\underset{\sim}{\hat{\Sigma}}_Q S^{-1})$ would be equal $tr(\underset{\sim}{\hat{\Sigma}}_R S^{-1})$ (common value being p) when and only when $\underset{\sim}{S}$ is structurally similar to the hypothesized structure of $\underset{\sim}{\Sigma}$. So, in practice, when the two indicators e.g. $tr(\underset{\sim}{\hat{\Sigma}}_Q S^{-1})$ and $tr(\underset{\sim}{\hat{\Sigma}}_R \underset{\sim}{S}^{-1})$ are closer and closer to p , $\underset{\sim}{S}$ may be recognized to have greater and greater resemblance to the hypothesized structure of $\underset{\sim}{\Sigma}$.

Combining these two indicators, we construct an index of structural closeness (ISC_1) as

$$ISC_1 = tr(\hat{\Sigma}_R S^{-1})/tr\,(\hat{\Sigma}_Q S^{-1}). \qquad (3.3.42)$$

$ISC_1 \in (0, 1)$ and may be expressed in percentages. We calculate ISC_1-values for all the hypothesized structures of Σ under consideration and choose that structure as the "best" for which ISC_1 is the largest, provided such ISC_1 in percentage form would be close to 100 .

Following the same argument as stated above, we may be also able to construct another index of structural closeness (ISC_2) which is based on two different indicators, namely, $det(\hat{\Sigma}_Q S^{-1})$ and $det(\hat{\Sigma}_R S^{-1})$ as shown below :

$$ISC_2 = [det(\hat{\Sigma}_R S^{-1})/det(\hat{\Sigma}_Q S^{-1})]^{1/p} = (det\,\hat{\Sigma}_R/det\,\hat{\Sigma}_Q)^{1/p} . \qquad (3.3.43)$$

From the key inequality (3.3.32), the indicators $det(\hat{\Sigma}_R S^{-1}) \in (0,1)$ and $det(\hat{\Sigma}_Q S^{-1}) \in (1,\infty)$ so that $ISC_2 \in (0,1)$. Parallel to the Propositions 3.3.5 and 3.3.6 for the indicators $det(\hat{\Sigma}_Q S^{-1})$ and $det(\hat{\Sigma}_R S^{-1})$ respectively, we now state a sufficient condition for ISC_2 to be equal to unity. The condition is that S should be structurally the same as Σ .

Algebraically, we may derive shorter interval for ISC_1 or ISC_2 as implied in the following propositions :

Proposition 3.3.7 : For a linearly structured $\underset{\sim}{\Sigma}$, ISC_1 or $ISC_2 \in [ab, \min(c/b, 1)]$ where $a = \lambda_{min}(\hat{\underset{\sim}{\Sigma}}_R)/\lambda_{min}(\hat{\underset{\sim}{\Sigma}}_Q)$, $b = \lambda_{min}(\underset{\sim}{S})/\lambda_{max}(\underset{\sim}{S})$ and $c = \lambda_{max}(\hat{\underset{\sim}{\Sigma}}_R)/\lambda_{min}(\hat{\underset{\sim}{\Sigma}}_Q)$.

Proof : Denoting the square root matrix of $\underset{\sim}{\Sigma}$ by $\underset{\sim}{\Sigma}^{1/2}$, we note that

$$\lambda_{max}(\underset{\sim}{\Sigma}\underset{\sim}{S}^{-1}) = \lambda_{max}(\underset{\sim}{\Sigma}^{1/2}\ \underset{\sim}{S}\ \underset{\sim}{\Sigma}^{1/2})$$

$$\leq \lambda_{max}(\underset{\sim}{S}^{-1})\ \lambda_{max}(\underset{\sim}{\Sigma}^{1/2}\ \underset{\sim}{\Sigma}^{1/2})$$

$$= \lambda_{max}(\underset{\sim}{\Sigma})/\lambda_{min}(\underset{\sim}{S}) \cdot \qquad (3.3.44)$$

and that

$$\lambda_{min}(\underset{\sim}{\Sigma}\underset{\sim}{S}^{-1}) = 1/\lambda_{max}(\underset{\sim}{\Sigma}^{-1}\underset{\sim}{S})$$

$$\geq \lambda_{min}(\underset{\sim}{\Sigma})/\lambda_{max}(\underset{\sim}{S}) \cdot \qquad (3.3.45)$$

Then, since

$$p\ \lambda_{max}(\underset{\sim}{\Sigma}\underset{\sim}{S}^{-1}) < tr(\underset{\sim}{\Sigma}\underset{\sim}{S}^{-1}) < p\ \lambda_{max}(\underset{\sim}{\Sigma}\underset{\sim}{S}^{-1}),$$

using (3.3.44) and (3.3.45) we have

$$p \cdot \lambda_{min}(\underset{\sim}{\Sigma})\ \lambda_{max}(\underset{\sim}{S}) < tr(\underset{\sim}{\Sigma}\underset{\sim}{S}^{-1}) < p \cdot \lambda_{max}(\underset{\sim}{\Sigma}) \cdot \lambda_{min}(\underset{\sim}{S}) \cdot \qquad (3.3.46)$$

Similarly,

$$\lambda^p_{min}(\underset{\sim}{\Sigma}\underset{\sim}{S}^{-1}) < det(\underset{\sim}{\Sigma}\underset{\sim}{S}^{-1}) < \lambda^p_{max}(\underset{\sim}{\Sigma}\underset{\sim}{S}^{-1})$$

so that

$$[\lambda_{min}(\underset{\sim}{\Sigma})/\lambda_{max}(\underset{\sim}{S})]^p < \det(\underset{\sim}{\Sigma}\underset{\sim}{S}^{-1}) < [\lambda_{max}(\underset{\sim}{\Sigma})/\lambda_{min}(\underset{\sim}{S})]^p \cdot \qquad (3.3.47)$$

Using inequalities (3.3.46) and (3.3.47) upon the definitions of ISC_1 and ISC_2 , the proof may be completed. ■

Remark 3.9 : For the most restrictive hypothesized structure, namely the spherical structure $\underset{\sim}{\Sigma} = \sigma^2 \underset{\sim}{I}_p$, it is easy to check that

$$ISC_1 = ISC_2 = \frac{p\ tr(\underset{\sim}{S}^{-1})}{tr(\underset{\sim}{S})\ tr(\underset{\sim}{S}^{-2})} \cdot \qquad (3.3.48)$$

Proposition 3.3.8 : For a linearly structured $\underset{\sim}{\Sigma}$, ISC_1 or ISC_2 does not exceed GFIR .

Proof : By definition, $ISC_1 = \frac{tr(\hat{\underset{\sim}{\Sigma}}_R S^{-1})}{p} \cdot \frac{p}{tr(\hat{\underset{\sim}{\Sigma}}_Q \underset{\sim}{S}^{-1})}$

$< tr(\hat{\underset{\sim}{\Sigma}}_R S^{-1})/p$ [using (3.3.31)]

= GFIR . [using (3.3.40) and (3.3.7)]

Similarly, using (3.3.32),

$ISC_2 < (\det \hat{\underset{\sim}{\Sigma}}_R \underset{\sim}{S}^{-1})^{1/p}$

$< tr(\hat{\underset{\sim}{\Sigma}}_R \underset{\sim}{S}^{-1})/p$ = GFIR. [since a.m. > g.m.] ■

Remark 3.10 : When the hypothesized $\underset{\sim}{\Sigma}$ is structurally the same as $\underset{\sim}{S}$, then $\underset{\sim}{S}$ becomes the estimate of $\underset{\sim}{\Sigma}$ by both Q-and R-procedures. As a consequence, $\hat{\underset{\sim}{Q}}$ and $\hat{\underset{\sim}{R}}$ would be null matrices. We may then find that measures such as $\|\hat{\underset{\sim}{Q}}\|$ and

$\|\hat{R}\|$ would be zero for the "best" fitted structure. Instead of the pair ($\|\hat{Q}\|$, $\|\hat{R}\|$), we may suggest Max($\|\hat{Q}\|$, $\|\hat{R}\|$) or $(\|\hat{Q}\|^2 + \|\hat{R}\|^2)^{1/2}$ as a single measure of structural closeness when these are not equal to zero.

3.3.6 The Effect of Number of Variables on the Lower limits of GFIQ, GFIR and ISC Indices

For the purpose of demonstrating the effect of p (number of variables), let us consider a 3 x 3 spherical Σ-matrix, i.e., $\Sigma = \sigma^2 I_3$. For experimental purpose, we consider various linearly structured sample covariance matrices (S_i) imposing different types of constraints upon its 6 non-duplicated elements. For simplicity, we list below some 8 such structured matrices, namely, S_i, i = 1,2, ... , 8 as having only one or two unknown elements with admissible range for positive definiteness.

$$\text{(i)} \quad S_1 = \begin{bmatrix} 1 & a & 0 \\ a & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} , \quad |a| < 1$$

$$\text{(ii)} \quad S_2 = \begin{bmatrix} 1 & a & 0 \\ a & 1 & a \\ 0 & a & 1 \end{bmatrix} , \quad |a| < 1/\sqrt{2}$$

$$\text{(iii)} \quad S_3 = \begin{bmatrix} 1 & a & a \\ a & 1 & a \\ a & a & 1 \end{bmatrix} , \quad -1/2 < a < 1$$

(iv) $\underset{\sim}{S}_4 = \text{diag.}\ (1, a, b),\ a > 0,\ b > 0$

(v) $$\underset{\sim}{S}_5 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & a & a \\ 1 & a & b \end{bmatrix},\ 1 < a < b < \infty$$

(vi) $$\underset{\sim}{S}_6 = \begin{bmatrix} 1 & a & 0 \\ a & 1 & 0 \\ 0 & 0 & b \end{bmatrix},\ |a| < 1,\ b > 0$$

(vii) $$\underset{\sim}{S}_7 = \begin{bmatrix} 1 & a & 0 \\ a & b & 0 \\ 0 & 0 & 1 \end{bmatrix},\ |a| < \sqrt{b},\ b > 0$$

(viii) $$\underset{\sim}{S}_8 = \begin{bmatrix} 1 & a & b \\ a & 1 & a \\ b & a & 1 \end{bmatrix},\ |a| < \sqrt{(1+b)/2},\ |b| < 1$$

Using the GFIQ, GFIR and ISC formula simplified for $p = 3$ and for $\underset{\sim}{\Sigma} = \sigma^2 \underset{\sim}{I}_3$ [vide (3.3.35) and (3.3.48)], we construct Tables 3.6 and 3.7 showing limiting values of these indices when the values of a (and b) either make the $\underset{\sim}{S}$-matrix to be just positive semi-definite or make the $\underset{\sim}{S}$-matrix to be perfectly spherical.

Table 3.6

Limiting Values of the Selected Indices for Sample Covariance Matrices having Structures with Single Unknown Element

Structure	Given value of 'a'	Indices of Goodness-of-fit GFIQ	GFIR	ISC
S_1	0	1	1	1
	$\pm$ 1	3/5	1/3	0
S_2	0	1	1	1
	$\pm 1/\sqrt{2}$	3/5	1/3	0
S_3	-1/2	2/3	1/3	0
	0	1	1	1
	1	1/3	2/3	0

It is seen from Tables 3.6 and 3.7 that S has a perfect fit to the spherical structure, the values of the fit indices are all unity. But when S assumes a structure other than spherical, e.g. S_1, and the value of the unknown element(s), e.g. a is set equal to $\pm$ 1, only ISC index turns out to be zero, as it should be under any model misspecification. In contrast to this, under model misspecification, both GFIQ and GFIR assumes significant non-zero values. Thus, the limiting values of these two indices fail to reach the desired value of zero necessary for the unequivocal rejection of the sphericity structure.

Table 3.7

Limiting Values of the Selected Indices for Sample Covariance Matrices having Structures with Two Unknown Elements

Structure	Given values of b\a	Measures of Goodness-of-fit								
		GFIQ			GFIR			ISC		
		0	1	∞	0	1	∞	0	1	∞
S_4	0	1/3	2/3	1/3	2/3	1/3	1/3	0	0	0
	1	2/3	1	1/3	1/3	1	2/3	0	1	0
	∞	1/3	1/3	2/3	1/3	2/3	1/3	0	0	0
S_5	0	*	*	*	*	*	*	*	*	*
	1	*	1/3	*	*	1/3	*	*	0	*
	∞	*	1/3	1/3	*	1/3	1/3	*	0	0
	b\a	0	±1	*	0	±1	*	0	±1	∞
S_6	0	2/3	1/3	*	1/3	4/9	*	0	0	*
	1	1	3/5	*	1	4/9	*	1	0	*
	∞	1/3	1/3	*	2/3	4/9	*	0	0	*
S_7	0	2/3	*	*	1/3	*	*	0	*	∞
	1	*	3/5	*	1	1/3	*	1	0	*
	∞	1/3	1/3	*	2/3	2/3	*	0	0	*
	b\a	0	$\pm 1/\sqrt{2}$	1	0	$\pm 1/\sqrt{2}$	1	0	$\pm 1/\sqrt{2}$	1
S_8	-1	3/5	*	*	1/3	*	*	0	*	*
	0	1	3/5	*	1	1/3	*	1	0	*
	1	3/5	3/7	1/3	1/3	1/3	2/3	0	0	0

*Cannot be determined due to indefiniteness.

From Tables 3.6 and 3.7, we also find that only the ISC index ensures the zero lower limit, and non-zero lower limits of GFIQ and GFIR, so obtained, are essentially due to the effect of dimension p . To clarify this, let us construct Table 3. considering the structures $\underset{\sim}{S}_1$ and $\underset{\sim}{S}_3$ since their inverses are very simple to express algebraically for any value of p .

Table 3.8

Limiting Values of the Selected Indices for two Chosen Structures of Sample Covariance Matrices of p Dimension

Structure	Given value of 'a'	Indices of Goodness-of-fit		
		GFIQ	GFIR	ISC
$\underset{\sim}{S}_1$	0	1	1	1
	1	p/(p+2)	1/p	0
$\underset{\sim}{S}_3$	-1/(p-1)	(p-1)/p	1/p	0
	0	1	1	0
	1	1/p	(p-1)/p	0

Table 3.8 indicates that as $p \longrightarrow \infty$, the lower limits of GFIQ and GFIR tend to zero. While ISC attains its lower limit of exact zero without depending on p , the lower limits of GFIQ and GFIR are affected by p. Such influence of p upon the values of the indices is unwanted. From this point of view, the proposed ISC index is seen to have acquired superiority over GFIQ and GFIR indices. However, our comment

at this juncture is limited only to spherically structured $\underset{\sim}{\Sigma}$.

Remark 3.11 : The effect of p upon GFIQ (and GFIR) as mentioned above, is not possible to remove. The GFIQ index may also assume negative values in case $tr(\hat{\underset{\sim}{\Sigma}}_Q^{-1}\underset{\sim}{S}) < p/2$. However, for linearly structured $\underset{\sim}{\Sigma}$, $tr(\hat{\underset{\sim}{\Sigma}}_Q^{-1}\underset{\sim}{S}) = p$ and $GFIQ = p/tr(\hat{\underset{\sim}{\Sigma}}_Q^{-1}\underset{\sim}{S})^2$ which ranges from $1/p$ to 1 since $tr(\hat{\underset{\sim}{\Sigma}}_Q^{-1}\underset{\sim}{S})^2 > tr^2(\hat{\underset{\sim}{\Sigma}}_Q^{-1}S)/p$. In order to make the GFIQ ranging from 0 to 1, we may suggest the following transformation :

$$GFIQ^* = \frac{GFIQ - 1/p}{1 - 1/p} . \qquad (3.3.49)$$

The effect of p may be seen to be somewhat reduced when we use $GFIQ^*$ in stead of GFIQ. In the case of GFIR, no such transformation is possible. ■

3.3.7 Illustrative Example : Analysis of GRE Repeaters Data

For the purpose of illustration, we consider an example based on real life data assumed to be a random sample from a trivariate normal population with null mean vector but with structured $\underset{\sim}{\Sigma}$ matrix. As to the structure of $\underset{\sim}{\Sigma}$, we fit one by one as many as 13 different linear covariance structures, which appear below according to the increase in the number of parameters (k) involved in each.

(i) Spherical : $\underset{\sim}{\Sigma} = a\,\underset{\sim}{I}_3$, $k = 1$

(ii) Intraclass : $\underset{\sim}{\Sigma} = \begin{bmatrix} a & b & b \\ b & a & b \\ b & b & a \end{bmatrix}$, $k = 2$

(iii) Tri-diagonal Moving Average : $\underset{\sim}{\Sigma} = \begin{bmatrix} a & b & 0 \\ b & a & b \\ 0 & b & a \end{bmatrix}$, k = 1

(iv) Quasi-Spherical : $\underset{\sim}{\Sigma}$ = diag. (a_1, a_2, a_3), k = 3

(v) Toeplitz : $\underset{\sim}{\Sigma} = \begin{bmatrix} a & b & c \\ b & a & b \\ c & b & a \end{bmatrix}$, k = 3

(vi) Guttman-Simplex : $\underset{\sim}{\Sigma} = \begin{bmatrix} a & a & a \\ a & b & b \\ a & b & c \end{bmatrix}$, k = 3

(vii) Equi-variance : $\underset{\sim}{\Sigma} = \begin{bmatrix} a & b & c \\ b & a & d \\ c & d & a \end{bmatrix}$, k = 4

(viii) Quasi-Intraclass: $\underset{\sim}{\Sigma} = \begin{bmatrix} a_1 & b & b \\ b & a_2 & b \\ b & b & a_3 \end{bmatrix}$, k = 4

(ix) Quasi-Toeplitz : $\underset{\sim}{\Sigma} = \begin{bmatrix} a_1 & b & c \\ b & a_2 & b \\ c & b & a_3 \end{bmatrix}$, k = 5

(x) Quasi-Simplex(increasing): $\underset{\sim}{\Sigma} = \begin{bmatrix} a_1 & b & b \\ b & a_2 & c \\ b & c & a_3 \end{bmatrix}$, k = 5

(xi) Quasi-Simplex (decreasing) : $\underset{\sim}{\Sigma} = \begin{bmatrix} a_1 & b & c \\ b & a_2 & c \\ c & c & a_3 \end{bmatrix}$, k = 5

(xii) Tri-diagonal Jacobi: $\underset{\sim}{\Sigma} = \begin{bmatrix} a_1 & b_1 & 0 \\ b_1 & a_2 & b_2 \\ 0 & b_2 & a_3 \end{bmatrix}$, k = 5

(xiii) Persymmetric : $\underset{\sim}{\Sigma} = \begin{bmatrix} a & d & b \\ d & b & e \\ b & e & c \end{bmatrix}$, k = 5

Our aim here is to identify which among such competing structural forms will closely approximate the sample $\underset{\sim}{S}$ in such a way that the discrepancy between $\underset{\sim}{S}$ and $\underset{\sim}{\Sigma}$ is minimum in suitable sense. To do that, we confine ourselves only to the columns calculated and shown in Table 3.10 in relation to GFIQ, GFIR, ISC_1 and ISC_2 . We further include the columns of $\|\hat{\underset{\sim}{Q}}\|_F$ and $\|\hat{\underset{\sim}{R}}\|_F$. We expect that these columns would be sufficient to examine the comparable discriminatory power of the proposed indices of closeness as a measure of goodness-of-fit of a given covariance structure.

From an unpublished study of Rock and Werts (not dated), Werts *et al.* (1981) analysed the sample covariance matrices of the scores obtained from candidates who appeared in the Graduate Record Examination (GRE) of Educational

Testing Service, Princeton, New Jersey, U.S.A.. Groups of candidates took the GRE twice, thrice, four times and five times. Werts et al. (1981) attempted to fit the quasi-Markov simplex structure to the whole set of such longitudinal data. We shall reproduce below only the sample covariance matrix of the scores of 5,072 respondents who took the GRE thrice. In Table 3.9, the diagonal, lower diagonal and upper diagonal elements correspond to the variances, covariances and intercorrelations respectively.

Table 3.9

Sample Variances, Covariances and Product-moment Intercorrelations for the GRE Repeaters(N=5,072)

Year of Administration	1	2	3
1	14,388	0.8712	0.8568
2	11,809	12,771	0.8822
3	11,049	10,719	11,559

In stead of a less parsimonious model of quasi-Markov simplex for the 3 x 3 GRE data, we first consider the hypothesis of the additive Guttman simplex structure and quasi-simplex structures (both increasing and decreasing). The main reason is that the correlations tend to decrease as one moves away from the main diagonal of the sample intercorrelation matrix. We shall, however, presume that more

than one competing structure may exist depending upon various casual arguments in generating the structure. Hence, there is a need for a study of structural closeness of various possible structures and the detection of the " best" possible structure out of them.

The sequencing of different structures in the first column of Table 3.10 is done according to the increasing value of the indices as obtained by using GFIQ. There are as many as 8 structures, e.g. from intraclass to decreasing quasi-simplex which are found to have a very high GFIQ, each being close to 100. Although the decreasing quasi-simplex structure has the largest GFIQ-value, the importance of other seven structures can not be summarily dismissed. Similar discriminatory problems become obvious while we inspect GFIR-column. The prescribed indices, e.g. GFIQ and GFIR by two important methods - ML and GLS procedures - are thus found to be not so efficient in detecting the best structure. Their values for screening the other models which are not so relevant can not be however denied.

The columns of ISC_1 and ISC_2 have shown a considerable power in detecting the " best" structure out of 13 competing structures. The best fitted structure in terms of these two indices is obviously the decreasing quasi-simplex which is also well-supported by the smallest magnitudes of $\|\hat{\underset{\sim}{Q}}\|_F$ and $\|\hat{\underset{\sim}{R}}\|_F$.

Table 3.10

Measures of Goodness-of-fit obtained for the GRE Data under Different Model Specifications

Covariance structure hypothesized	Number of parameters (k)	GFIQ	GFIR	ISC_1	ISC_2	$\|\hat{Q}\|_F$	$\|\hat{R}\|_F$
Spherical	1	39.8	68.1	12.7	12.6	1.72	1.35
Quasi-Spherical	3	40.3	69.7	13.1	13.0	1.10	1.19
Tridiagonal (M.A.)	2	52.8	70.1	24.9	24.2	2.02	1.14
Tridiagonal (Jacobi)	5	67.1	71.8	25.3	28.6	0.94	0.54
Guttman-Simplex	3	83.0	74.1	54.4	54.3	4.23	4.27
Intraclass	2	95.6	96.3	92.2	92.2	1.41	1.19
Toeplitz	3	96.7	96.8	93.5	93.5	1.01	0.98
Equivariance	4	97.2	97.2	94.4	94.4	0.43	0.42
Quasi-Intraclass	4	98.1	98.3	96.6	96.5	0.51	0.45
Persymmetric	5	98.3	98.2	96.4	96.4	0.37	0.39
Quasi-Toeplitz	5	98.4	98.4	96.8	96.8	0.44	0.43
Increasing Quasi-Simplex	5	99.1	99.1	98.2	98.2	0.36	0.36
Decreasing Quasi-Simplex	5	99.4	99.8	99.6	99.6	0.16	0.16

As regards the comparative performance of different structures, it is clear from Table 3.10 that the involvement of too many parameters would not necessarily increase the values of " goodness" of fit. The intraclass structure, for example, involving only 2 parameters shows somewhat better performance than some of the structures mentioned in the first 5 rows of Table 3.10. These structures have as many as 5 parameters. Yet, their fit seems to be quite poor. The tri-diagonal Jacobi structure, for example, having as many as 5 parameters has a very bad fitting performance as evidenced by every column of Table 3.10. Thus, even with maximum number of parameters (equal to $p(p+1)/2-1$) for a structured $\underset{\sim}{\Sigma}$, the fit may not be necessarily satisfactory in all cases unless the process generating $\underset{\sim}{\Sigma}$ matrix is also reflected by the structure of $\underset{\sim}{S}$.

The hypothesis of decreasing quasi-simplex covariance structure (Mukherjee, 1970) seems to be quite appealing for the given GRE data. It is well known that with gradual practice, the inter-individual variability diminishes. In the case of GRE, the examinees becomes more and more test-wise with practice. As a result of reduced inter-individual variability and greater familiarity with the nature of the task, the correlations between two adjacent periods also tend to increase. This, in turn, produces a kind of 'diagonal

ridge' pattern in the inter-correlation matrix indicating that with increasing practice, the neighbouring variables tend to correlate more and more.

However, with the passage of time, there is a kind of 'plateau' or stability in the magnitude of correlation between the adjacent periods. Therefore, if the practice session is extended for a much longer period, the inter-correlations among the adjacent points should be more or less equal and the entire correlation matrix for the concluding phase should have a Laurent or Toeplitz structure (Vide Chap. 5, subsec. 5.4.2).

3.3.8 A Monte Carlo Experiment on Measures of Goodness-of-Fit

We considered the same 'Monte Carlo' simulation experiment as described in subsection 3.2.5. In order to compare the relative performances of the proposed indices of structural closeness vis-a-vis GFIQ and GFIR, sample covariance matrices were generated from an 8-variate normal population with zero mean vector and intraclass covariance matrix with common variance unity and common correlation ρ. We chose eight different values of ρ as -.1, -.05, 0, .05, .1, .3, .5 and .9 and generated for each ρ a total of 100 sample covariance matrices, each of sizes N = 20, 40, 100 and 200. Thus, a total of 8 x 4 x 100 = 3,200 sample covariance

matrices of order 8 x 8 were generated independently for the study.

We computed the values of the fit indices — GFIQS, GFIRS, and $ISCS_1$ and $ISCS_2$ first by fitting the spherical structure and next the values of the fit indices GFIQI, GFIRI, and $ISCI_1$ and $ISCI_2$ by fitting the intraclass structure to all such sample covariance matrices actually drawn from the intraclass pupulation. Since $ISCS_1 = ISCS_2$ [vide (3.2.48)], we denote the common value by ISCS. After computation, it was found that $ISCI_1$ and $ISCI_2$ values (in percentage form) were more or less identical at least upto first place of decimal. For all practical purposes, we shall therefore consider $ISCI_1 \approx ISCI_2$ common value being ISCI. Since the sample covariance matrices are actually sampled from intraclass population, the indices, namely, GFIQI, GFIRI and ISCI are expected to measure the goodness-of-fit . Conversely, GFIQS, GFIRS and ISCS are expected to measure the badness-of-fit. Trivially, for $\rho = 0$, it is expected that GFIQI = GFIQS, GFIRI = GFIRS and ISCI = ISCS.

Table 3.11 shows the mean and mean squares values of the above-mentioned indices. For increasing sample size, all the indices measuring goodness-of-fit are found to be increasing and becoming more and more stable for any value of ρ . For example, for N = 200, GFIQI = 90, GFIRI = 96, ISCI = 92 (approx.). But as expected, GFIQS, GFIRS and ISCS are

systematically decreasing as ρ-value departs from 0 more and more. These values are also increasing with increasing sample size. As sample size increases, the values of GFIQI, GFIRI and ISCI are required to indicate acceptable model fit. For example, when N = 20, the critical value for rejection of intraclass model on the basis of GFIQI should be 68 and below. For N = 200, the critical value of GFIQI should be 95 and below. In the case of ISCI, the critical value is 34 or less when N = 20. But when N = 200, it is 90 or below.

All the indices have decreasing variability (measured by mean squares) for increasing sample size. In terms of measures of skewness and kurtosis [from the calculated values of sample β_1 and β_2 , not produced here for paucity of space, but available in Maiti and Mukherjee (1989a)] none of the indices show that their distributions can be approximated by normal distributions. This is true even for sample size as large as 200, except possibly that all of them are negatively skewed.

While comparing the indices among themselves, the mean GFIQ-value is the largest while the mean ISC-value is the smallest for any sample size. Because of the variation of sample size from 20 to 200, while the GFIQI mean values vary from 69 to 96, the ISCI mean values increase from 34

to 92. This clearly implies that ISC-value is more influenced by sample size than GFIQ-value.

In general, the indices are highly intercorrelated (not presented here for paucity of space). All intercorrelations are found to increase for increasing sample size. For example, the median correlation between GFIQI and ISCI for N = 20, 40, 100 and 200, are 0.59, 0.70, 0.82 and 0.93 respectively. [We take the median of 8 ρ-values under consideration for any sample size]. Thus for large sample sizes, the new index ISC will perform as efficiently as GFIQ and GFIR indices.

Of all indices selected for this study, the GFIR is the least sensitive to model misspecification, especially when the sample size is small. For example, it may be seen from Table 3.11 that when the sample size is as small as 20, and we calculate the GFIRS index for fitting the generated covariance matrix of intraclass structure with 0.9 as the off-diagonal values to a wrongly specified sphericity model, the index is found to be 53. Under the correct model specification, the GFIRI index after averaging across 100 samples is found to be 59 only. In the case of GFIQ index, the corresponding values are 15 and 69 respectively, thus showing a substantial increase due to correct choice of the model. The GFIRS also shows low specificity in rejecting

Table 3.11

Mean & Mean Squares of the Selected Indices of Goodness-of-fit obtained for Covariance Matrices Generated from an Intraclass Structures for Different Sample Sizes

Sample size	Rho value	GFIQS Mean	GFIQS Mean Sq.	GFIRS Mean	GFIRS Mean Sq.	ISCS Mean	ISCS Mean Sq.	GFIQI Mean	GFIQI Mean Sq.	GFIRI Mean	GFIRI Mean Sq.	ISCI Mean	ISCI Mean Sq.
N = 20	-.1	66	26.0	49	90.1	23	76.0	69	26.8	60	76.2	34	96.8
	-.05	68	20.7	57	99.4	31	99.0	70	18.6	59	94.2	34	104.6
	0	69	20.8	58	108.9	33	102.8	69	19.7	60	100.6	34	102.8
	.05	68	34.3	56	106.8	31	112.3	69	24.7	58	98.6	33	116.5
	.1	67	37.1	56	98.6	30	105.1	70	26.6	58	103.8	33	116.5
	.3	51	86.3	54	116.4	24	83.5	70	21.6	58	114.6	33	127.2
	.5	35	54.4	53	85.5	17	39.4	69	22.1	58	90.6	33	95.6
	.9	15	1.6	53	68.5	3	3.5	69	22.2	59	84.3	33	87.4
N =40	-.1	78	100.0	60	77.1	39	82.5	82	9.9	79	30.0	62	55.5
	-.05	81	16.0	76	53.9	58	96.5	82	15.1	78	43.2	61	80.5
	0	82	11.5	79	41.1	61	77.0	82	11.9	79	35.6	62	72.2
	.05	81	17.3	78	29.6	60	59.3	83	13.6	79	26.6	63	57.2
	.1	77	24.2	76	37.8	57	74.8	82	12.5	79	33.7	61	68.8
	.3	56	49.8	73	27.9	45	52.1	82	12.9	79	25.2	62	55.2
	.5	36	36.9	71	26.8	32	40.2	82	13.1	78	26.3	61	56.5
	.9	15	0.5	70	27.5	6	3.3	83	12.9	78	35.3	61	73.3
N =100	-.1	86	2.8	67	40.7	51	53.8	92	3.0	92	4.3	84	13.6
	-.05	90	4.2	88	10.0	78	24.7	92	3.5	91	4.3	83	13.5
	0	92	3.2	91	6.4	83	18.2	92	3.2	91	6.3	83	18.3
	.05	90	4.6	90	5.1	81	15.6	92	3.3	91	4.8	83	14.7
	.1	87	13.6	89	5.9	78	20.6	92	3.3	92	4.4	84	14.0
	.3	60	32.9	85	4.1	61	19.6	92	2.6	92	4.2	84	12.5
	.5	36	12.4	83	3.8	43	16.4	92	2.6	91	3.9	83	11.9
	.9	15	0.2	81	4.0	9	1.8	92	3.1	92	5.0	84	14.3
N =200	-.1	90	1.0	68	16.4	54	23.8	96	0.8	96	1.1	92	3.9
	-.05	94	0.8	92	3.2	86	7.9	96	0.7	96	0.8	92	2.8
	0	96	0.8	95	0.9	91	3.2	96	0.8	96	0.8	91	3.0
	.05	94	1.9	95	1.5	90	5.5	96	0.9	96	1.2	92	4.0
	.1	90	5.6	93	1.5	85	6.8	96	0.9	96	0.9	91	3.4
	.3	61	20.0	89	1.0	66	10.3	96	0.7	96	0.9	92	3.0
	.5	36	7.7	86	1.0	46	9.6	96	0.7	96	1.0	91	3.4
	.9	15	0.1	85	0.7	9	0.9	96	0.7	96	0.8	92	2.9

intraclass structures with as high value of common correlation as 0.30 .

With the increase in sample size, the proposed ISC index as compared to others, shows more and more sensitivity to model misspecification. For example, with a sample of size 200, the calculated mean value of this index (here, ISCS) is found to be 9 only when a grossly wrong model (sphericity) is specified. Under the correct model specification (intraclass with $\rho = 0.9$), the corresponding ISCI index rises to 92. In the case of GFIQ, the corresponding mean values are 15 and 96. When the true model is intraclass with $\rho = -0.1$, the calculated mean value of GFIQI for $N = 200$ is found to be 96. It turns out to be 90 for incorrect model fitting. In contrast to this, the corresponding mean value of ISC under incorrect model specification is 54 which increases to 92 when the correct intraclass model with $\rho = -0.1$ is fitted. Thus, ISC index seems to show more and more sensitivity than other indices as the true value of ρ tends to reach the inadmissible lower limit of $-1/(p-1)$.

In order to study the structural sensitivity and structural specificity of the GFIQ, GFIR and ISC indices, we shall consider GFIQS, GFIRS and ISCS values as the badness-of-fit measured by them due to fitting a wrong structure, namely a spherical structure in stead of the correct structure, namely the intraclass structure. It is expected that badness-

Table 3.12

Assessment of Specificity and Sensitivity of GFIQ, GFIR and ISC Indices

Sample size	Rho value	Specificity			Sensitivity		
		GFIQ	GFIR	ISC	GFIQ	GFIR	ISC
N = 20	-.1	10	12	16	6	6	5
	-.05	2	9	3	2	6	2
	0	4	3	2	3	3	2
	.05	4	8	3	2	8	4
	.1	15	4	3	3	2	2
	.3	77	2	2	30	2	2
	.5	100	6	29	100	14	39
	.9	100	11	100	100	20	100
N = 40	-.1	10	8	59	25	65	73
	-.05	3	3	3	2	11	4
	0	1	2	2	3	5	3
	.05	9	2	2	4	2	2
	.1	9	6	5	11	4	3
	.3	100	6	42	100	27	59
	.5	100	10	94	100	43	95
	.9	100	10	100	100	50	100
N = 100	-.1	71	100	100	85	100	100
	-.05	7	27	23	7	10	9
	0	1	2	0	1	1	1
	.05	7	2	10	12	5	19
	.1	49	3	7	17	7	10
	.3	100	29	100	100	88	100
	.5	100	93	100	100	99	100
	.9	100	100	100	100	100	100
N = 200	-.1	100	100	100	100	100	100
	-.05	42	63	63	23	56	50
	0	1	2	2	2	2	2
	.05	15	12	13	10	9	10
	.1	91	56	91	91	56	91
	.3	100	100	100	100	100	100
	.5	100	100	100	100	100	100

of-fit values with reference to any sample would be less than the corresponding goodness-of-fit value for any fit index.

An ideally structurally-specific and structurally-sensitive fit index should show that all the badness-of-fit values obtained from a given $\underset{\sim}{S}$ do not overlap with the goodness-of-fit values. Heuristically, we may assess the specificity of any fit index by the percentage of the $\underset{\sim}{S}$ matrices for which badness-of-fit values are less than the minimum of the goodness-of-fit values obtained for the same set of $\underset{\sim}{S}$ matrices. We may also assess the sensitivity of any fit index by the percentage by which goodness-of-fit values are greater than the maximum of the badness-of-fit values.

Keeping these in view, we report in Table 3.12, all the three indices which have increasing sensitivity and specificity with increasing sample size. For positive ρ-values, both the GFIQ and ISC are certainly more satisfactory than GFIR.

However, the GFIQ shows in general better performance than ISC. For negative ρ-values, ISC performs better than GFIQ but does not do so over GFIR.

If we agree, for simplicity, that more the difference in the mean values of an index between correct and incorrect structures, the better the index, then the dominating

Table 3.13

Difference of mean values of the three indices between correct and incorrect structures as obtained from Table 3.11.

Sample size	value	Difference in the Mean values		
		GFIQI-GFIQS	GFIRI-GFIRS	ISCI-ISCS
N = 20	-.1	3	11	11
	-.05	2	2	3
	0	0	2	1
	.05	1	2	2
	.1	3	2	3
	.3	19	4	7
	.5	34	5	16
	.9	54	6	30
N = 40	-.1	4	19	23
	-.05	1	2	3
	0	0	0	1
	.05	2	1	3
	.1	5	3	4
	.3	26	6	17
	.5	46	7	29
	.9	68	8	55
N = 100	-.1	6	25	33
	-.05	2	3	5
	0	0	0	0
	.05	2	1	2
	.1	5	3	6
	.3	32	7	23
	.5	56	8	40
	.9	77	9	75
N= 200	-.1	6	32	38
	-.05	2	4	6
	0	0	1	0
	.05	2	1	2
	.1	6	6	6
	.3	35	7	26
	.5	60	10	45
	.9	81	11	81

behaviour of the indices as mentioned above, may be similarly observed from Table 3.13.

In the absence of ISC, the GFIQ shows better performance both in terms of sensitivity and specificity when the value ρ is positive. But when ρ assumes a negative value, the GFIR shows better result than GFIQ. Thus, when the value of ρ is unknown, we are not sure as to the choice of the index. The use of a double pronged approach by employing both the indices simultaneously should be the simplest strategy.

3.4 DISCUSSION

Traditionally, the ML minimization function $F(\underset{\sim}{S},\underset{\sim}{\Sigma})$ [vide (3.3.1)] has been used in ACOVS as an overall measure of goodness-of-fit of $\underset{\sim}{\Sigma}$ to $\underset{\sim}{S}$. As an alternative to this, we may propose a measure based on the minimization of the residual quadratic form $N < \underset{\sim}{S} - \underset{\sim}{\Sigma} >' \underset{\sim}{\Phi}^{-1}(\underset{\sim}{\Sigma}) <\underset{\sim}{S} - \underset{\sim}{\Sigma}>$. Under multinormal set up, this measure would be $\frac{N}{2}$ tr$(\hat{\underset{\sim}{Q}}\,\hat{\underset{\sim}{\Sigma}} - \hat{\underset{\sim}{Q}}\,\underset{\sim}{S})$ (where $\hat{\underset{\sim}{\Sigma}}$ is the MLE of $\underset{\sim}{\Sigma}$) and is thus related to a more fundamental approach to the problem of estimation (Berkson, 1980). The two statistics $NF(\underset{\sim}{S}, \hat{\underset{\sim}{\Sigma}})$ and $\frac{N}{2}$ tr$(\hat{\underset{\sim}{Q}}\,\hat{\underset{\sim}{\Sigma}} - \hat{\underset{\sim}{Q}}\,S)$ are both distributed asymptotically as a chi-square variable with υ d.f. and are asymptotically equivalent in the sense that as $N \longrightarrow \infty$, the difference between these two statistics converges in probability to zero. Marsh et al.(1988) in the appendix of their paper obtained the asymptotic distribution of GFIQ based on the distribution of $NF(\underset{\sim}{S}, \hat{\underset{\sim}{\Sigma}})$. We have derived the same independently from the distribution of $\hat{\underset{\sim}{Q}}$.

The equivalence between the LR test statistic and the Wald type of test statistics is proven by Lee (1985) in the context of ACOVS. Very recently, Satorra (1989) provided a unified approach to the asymptotic theory of these test criteria along with Rao's efficient score criterion. Incidentally we found the algebraic identity between the Wald's test statistic and Rao's efficient score criterion for testing a very special covariance hypothesis, namely $\underset{\sim}{\Sigma} = \underset{\sim}{I}_p$.

However, three procedures examined in this chapter will not yield, in general, identical results or even comparable values in small sample cases. So it is possible for these tests to yield conflicting conclusions in empirical situations, as has been noted by various authors (e.g. Berndt and Savin, 1977; Breusch, 1979; Evans and Savin, 1982) in testing linear restrictions on the coefficients of certain linear models useful in econometrics.

Breusch and Pagan (1980) observed in the above context that in the absence of any information on the finite sample proportions of these test, the choice between them is likely to depend on computational covenience. If the unrestricted estimates are easier to compute, the Wald test should be preferred. However, in ACOVS, the efficient score test of Rao which is the same as the Lagrange Multiplier (LM) test of Aitchison and Silvey (1985), yields test statistic of the form $\frac{N}{2}\ \mathrm{tr}[\hat{\underset{\sim}{Q}}\ (\hat{\underset{\sim}{\Sigma}} - \underset{\sim}{S})]$ which is computationally much

convenient particularly when we adopt Q-procedure to obtain $\hat{\underset{\sim}{\Sigma}}$ and $\hat{\underset{\sim}{Q}}$ simultaneously in the case of linearly structured $\underset{\sim}{\Sigma}$. As such LM test is also equally competing for testing hypothesis in ACOVS. If, however, the power of the test is the main guiding principle, there is no obvious reason to prefer one test over the others, as evidenced by our Monte Carlo simulation study. According to Satorra and Saries (1985), the conventional LR test appears to be as powerful as the other two tests.

We have not investigated the possible numerical inequality among three test statistics in the context of ACOVS. Savin (1976), Breusch (1979) as well as Berndt and Savin (1977) showed that a systematic numerical inequality exists between the three test statistics in the multinormal case when we are testing certain linear restrictions on the regression coefficients. The inequality relation between the values of these test statistics is

$$\text{Wald} \geq \text{LR} \geq \text{Lagrange (Rao)}.$$

The main significance of this inequality appears to be that one outcome can be favoured over another by prior choice of testing procedure. In particular, rejection of H_0 that $\underset{\sim}{\Sigma} = \underset{\sim}{\Sigma}_0$ can be favoured by selecting Wald statistics, <u>a priori</u>, while the probability of acceptance of H_0 can be enhanced by choosing the Rao's criterion <u>a priori</u>. Through simulation studies, we

need to verify, therefore, whether for testing various structural hypotheses in ACOVS, the probability of Type I error will be less than 0.05 for the Rao's criterion and greater than 0.05 for the Wald's criterion when a (large-sample) 5 percent significance level is employed and the LR test correctly states the probability of Type I error as 0.05 under H_o .

We also note the necessity of investigating the small sample power properties of these tests and their robustness against violation of multinormality assumptions. Through simulation studies, we also need to study the usefulness of the goodness-of-fit measures in reaching correct decision regarding the covariance structure under different conditions.

The proposed indices of goodness-of-fit, namely, ISC_1 and ISC_2 fulfill almost all the properties as mentioned in the beginning of section 3.2. These indices show the general property of 'sensitivity' and power of discrimination in the sense that these can help in identifying covariance structures which best fit the observed data not only by yielding the lowest "loss-function" but also by showing the closest fit for a given sample. These indices are based on both MLE and GLSE of the covariance parameters in an anticipation that these estimates might be close to each other when the hypothesized structure is tenable.

Although the sampling distributions are unknown, these indices can nevertheless serve the purpose of deciding whether or not it is plausible to assume that a given sample covariance matrix could have arisen from a postulated covariance structure. Through a case study, it has also been demonstrated that the proposed indices are more sensitive than GFIQ and GFIR indices which are currently being used in ACOVS. If we consider an ISC_1 or ISC_2 value as high as 97 or above as a cut-off point for judging the fit as satisfactory for a sample size of 5,000 or more, then the quasi-simplex structures are the only models that fit the GRE data. If on the other hand, we use Jöreskog and Sörbom's (1981) GFIQ or Tanaka and Huba's (1985) GFIR indices, as many as 6 covariance structures will be considered as satisfactorily fit to GRE data. Thus, the proposed indices provide a means of assessing how well the chosen model for each set of data captures most of the information of the sample covariance matrix. These indices thus also provide a non-statistical assessment of the adequacy of structural hypothesis in fitting a set of data.

Although limited by its scope for generalization, the findings of the Monte Carlo study reported here provides the expected sample values of three different goodness-of-fit indices for two different covariance structures each of order 8. These tabled values allow researchers to compare their results

with values that would be obtained from correctly specified intraclass structures with multivariate normally distributed data given a particular sample size. The findings of this study clearly show that all the indices of goodness-of-fit, including GFIQ and GFIR are sensitive to sample size, as has been found in previous simulation studies reported by Bearden, Sharma and Teel (1982), Hoelter (1983), Anderson and Garbing (1984), as well as Marsh, Balla and McDonald (1988). As sample size increases, relatively larger values of GFIQ, GFIR and ISC indices are required for a decisively acceptable model fit.

The simulation experiment also reveals that for moderate and small samples, none of the fit indices under study, including the GFIQ appears to have a chi-square distribution. Their distributions, even for samples as large as 200, can not be approximately normal. However, the variability of these indices approaches zero as $N \longrightarrow \infty$. This finding suggests that the estimates of the indices increase in accuracy with sample size.

As to the sensitivity to model misspecification, the GFIR shows very small changes in its performance in comparison to both GFIQ and ISC indices when greater and greater departures are made in $\underset{\sim}{S}$ from the hypothesized $\underset{\sim}{\Sigma}$ (here sphericity structure).

From the results of the simulation study, both GFIQ and ISC indices seem to be quite sensitive to model mis-specification, but the former index appears to suffer from two drawbacks, namely, (a) it never attains the desired lower limit of zero unless $p = \infty$, and (b) for an 8-variate intraclass structure with the common correlation term ρ assuming admissible negative values, the ISC index shows better efficiency in rejecting an erroneous sphericity hypothesis than the GFIQ index and this sensitivity increases as ρ tends to $-1/(p-1)$, the lower admissible limit. For all positive values of ρ , both the indices show more or less the same amount of efficiency especially for large sample sizes. This observation is also supported by their increasingly high correlation as sample size increases. Thus if the sample size is very large and the order of the $\underset{\sim}{\Sigma}$-matrix is quite small, then for all practical purposes, the use of ISC index is recommended.

However, even after using such a sensitive index, if two or more structural hypotheses with varying degrees of freedom are accepted as plausible for a given sample size, then the researcher should use the efficiency index of Khattab and Hocevar (1982). This index is useful in deciding how efficient is the increase in fit in going from one model with many degrees of freedom to another model with just a few degrees of freedom. Another alternative approach will be to

calculate relative efficiency index of Szatrowski (1982) in order to determine the loss of efficiency in estimation when we estimate by MLE, the linearly structured covariance matrix under two or more nested models with the assumption that the more restrictive pattern is the true state. The model for which the relative efficiency is the highest will be chosen as the best one by this approach.

Should we desire not to use any index which is sentitive to sample size, then McDonald's (1989) index of goodness-of-fit based on noncentrality or the modified Tucker and Lewis (1973) coefficient may be used. A still better way will be to use the parsimonious fit index suggested by James, Mulaik and Brett (1982, p. 155) although there are some problems involved in it (McDonald and Marsh, 1989). Mulaik, James, Van Alstine, Bennett, Lind and Stilwell (1989) have revived the rationale for parsimony in the context of evaluation of goodness-of-fit in ACOVS along with some new measures of parsimonious fit. Bentler and Mooijaart (1989) advocated the use of parsimony as a criterion for choice between two otherwise acceptable models since, in large samples, they found that the more parsimonious of two rival nested models yields an estimator of the common parameters that has smaller sampling variance. We emphasize in this connection not only the parametric parsimony but also scientific parsimony in terms of

practical meaningfulness of the fitted results in relation to existing theories. The best fitted model should not only provide the best estimate of Σ given a sample of a specified size but also must prove useful for the interpretation of data in a parsimonious and meaningful way.

It is possible that for the sake of goodness-of-fit we will have to try new models incorporating in the original linearly structured models certain interactive and non-linear components (Kenny and Judd, 1984). However, if the fitted models are such that these can not be interpreted meaningfully, then we have to be content with the linear models. Even then, such models must be viewed with caution and may be considered plausible only if enough justification can be given on the basis of substantive theory. Careful attention to the assumption of a model is also essential for the use of any covariance structure analysis (ACOVS). The use of covariance structure modeling with non-experimental data requires both a strong theory of the substantive area connected with the model and careful attention to the specification of the model. From this point of view, the model modification procedure, as suggested by Sörbom (1989) merits special attention since he also emphasizes the acceptance of those '' best" fitting models " that are relevant, meaningful and interpretable" (p. 384).

3.5 SUMMARY

The conventional LR test asymptotically distributed as χ^2 under the multinormal assumption has so far played an important role in ACOVS. As alternatives, two new test statistics (which are the ACOVS-versions of Rao's efficient score criterion and Wald's quadratic form criterion) are proposed here. These test statistics have the same asymptotic χ^2-distribution as the LR test and based on ML estimate of $\underset{\sim}{Q} = \underset{\sim}{\Sigma}^{-1} - \underset{\sim}{\Sigma}^{-1}\ \underset{\sim}{S}\ \underset{\sim}{\Sigma}^{-1}$ which is readily available as a by-product when the Q-procedure (as described in Chap. 2) is employed. One of these statistics is also related to the most popular Jöreskog and Sörbom's goodness-of-fit index (GFIQ) for the linearly structured covariance matrix ($\underset{\sim}{\Sigma}$) to $\underset{\sim}{S}$. A Monte Carlo simulation study is reported which deals with the power of these statistics in rejecting a scale-free sphericity hypothesis when samples from intraclass covariance structures are drawn randomly.

The large sample distribution of GFIQ index is derived in the case of a certain wide class of structured covariance matrices. Additionally, two new fit indices each based on MLE and GLSE of the parameters are proposed. A case study has been presented to show how the proposed indices in comparison to GFIQ are more efficient in detecting the " best" structure. Algebraically, it has been shown that the limiting value of GFIQ fails to reach the desired value of zero under model misspecification. A Monte Carlo study also shows that with increasing

sample size, the proposed indices show more and more sensitivity to model misspecification.

CHAPTER 4

ALGEBRAIC AND STATISTICAL ASPECTS OF POSITIVE DEFINITE TOEPLITZ MATRICES

4.1 INTRODUCTION

The Toeplitz matrix occupies an active area of research since the beginning of this century. The main reason is perhaps its occurrence in a large variety of area not only in pure and applied mathematics, but also in physical and social sciences.

A matrix is called a Toeplitz matrix when its entries depend only upon the difference of the indices of rows and columns. This is also called Laurent matrix as its entries may be generated from the Laurent expansion of a suitably defined rational function.

In 1910, O. Toeplitz studied various forms of the Toeplitz matrix in relation to Laurent power series and called them L-forms, without assuming that they are symmetric. A sufficiently well structured theory existed in the Memoirs of Frobenius (1894, 1912). Granander and Szego (1958) dealt with analytical problems with regard to " Toeplitz forms" associated with various analytical functions. As regards operator theory and algebraic theory on the Toeplitz matrix, the books by Heinig and Rost (1984) and also by Iohvidov (1982) cover

recent progress in this area including an exhaustive list of related references.

4.1.1 Statistical Applications of Toeplitz matrices in Various Fields

We shall give here an exposition of some situations in various applied fields where positive definite (p.d.) Toeplitz matrix comes as an associated covariance matrix from the point of statistical consideration of stochastic process.

Let there be a discrete stochastic process producing a sequence of m realization $\{X_i = X(t_i),\ i = 1, \ldots, m\}$ at m time points and $\{Y_i = X(t_i+\alpha),\ i = 1, \ldots, m\}$ be a translated sequence due to a translation of α time units. This stochastic process is called stationary (in strict sense) when it maintains a statistical equilibrium in the sense that the joint density of $Y_1, Y_2, \ldots, Y_m$ is identical to that of $X_1, X_2, \ldots, X_m$ for any $m \geq 1$. In practice, we however tackle upto second order ($m = 2$) density functions and get what is called covariance stationarity or wide-sense stationary process obeying $E(X_i)$ independent of time t_i and $E(X_i, X_j)$ dependent only on time difference $|t_i - t_j|$. For a Gaussian (normal) stochastic process, covariance stationarity results to strict-sense stationarity, and the stationary process is called Gaussian Stationary process.

For equidistant time points, the covariance matrix for such sequence $\{X_i, i = 1,2, \ldots, m\}$ is a p.d. Toeplitz matrix or " stripe matrix" given by

$$\underset{\sim}{\Sigma}_m = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{m-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{m-2} \\ \vdots & & & & \\ \rho_{m-1} & \rho_{m-2} & \rho_{m-3} & \cdots & 1 \end{bmatrix} \qquad (4.1.1)$$

where $\sigma^2 = v(X_i)$ and $\rho_{|i-j|}\,\sigma^2 = Cov(X_i, X_j)$. The term $\rho_{|i-j|}$ is called autocorrelation reflecting in a sense correlation on the same variable at two different time points t_i and t_j. Such autocorrelation functions are generally observed in growth studies and in longitudinal studies of various psychological processes such as learning, forgetting and development of abilities.

In the case of repeated measurements, the error of measurements (i.e., stochastic disturbance term) at one observation point will be related to those at nearby observation points and not remain independent as is assumed in the classical psychological test theory. Such autocorrelations are thus " fairly typical of behavioural data". (Gottman and Glass, 1978, p. 202).

In psychological research, we encounter data consisting of random replications of short time series where a systematic error arises at each step. Due to practical problem of obtaining a large number of replications on the same respondents, such time series data in the psychological measurements are usually quite short in comparison with those long, single series found, for example, in econometric or biometric studies. In such psychological studies, we try to fit some stochastic model to the data on the assumption that the residual covariance matrix will exhibit a particular kind of pattern such as a Toeplitz structure.

In most of the applied fields, one deals with the direct application of the Gaussian stationary process, e.g. econometrics, signal processing, speech processing, electrical engineering, statistical mechanics etc.

To assure a steady-state technological process, or to analyse data in presence of white noise (random error), or to search for the hidden periodicities in the field such as geophysics, mateorology etc., stationarity in the underlying processes is studied via autocorrelations. Thus, as a consequence, p.d. Toeplitz matrix gets its own importance in these areas. We shall now give an overview of studies in these areas.

In the case of non-stationary stochastic process (such as econometric time series), devices like auto-regressive (AR),

Moving-average (MA) or mixed auto-regressive moving-average (ARMA) are used for modeling so as to obtain a residual series reasonably stationary which can give rise to a Toeplitz covariance matrix whose elements are functions of the un-known constants of the device chosen and used. For example, in the case of auto-regressive device of order, e.g.,

AR(r) : $X_i = \sum_{j=1}^{r} \varphi_j X_{i-j} + \epsilon_i$, ϵ_i is the random error component. The optimum estimate of the unknown constants, $\varphi_1, \cdots, \varphi_r$ requires solving Yule-Walker equation given by

$$\underset{\sim}{\Sigma}_r \underline{\varphi} = \underline{a} \qquad (4.1.2)$$

where $\underset{\sim}{\Sigma}_r$ is a symmetric Toeplitz matrix with parameters $a_0, a_1, \cdots, a_{r-1}$; $\underline{a} = (a_1, a_2, \cdots, a_r)'$ and $\underline{\varphi} = (\varphi_1, \varphi_2, \cdots, \varphi_r)$. To solve (4.1.2), Levinson-Durbin algorithm is very much in use once the solution corresponding to AR(r-1) is known. Denoting the j-th φ-coefficient of AR(r) by φ_{rj} ,

$$\varphi_{rj} = \varphi_{r-1,j} - \varphi_{rr} \varphi_{r-1,r-j} , \quad j = 1,2, \cdots, (r-1)$$

$$\varphi_{rr} = \frac{a_r - \sum_{j=1}^{r-1} \varphi_{r-1,j} a_{r-j}}{a_0 - \sum_{j=1}^{r-1} \varphi_{r-1,j} a_j} . \qquad (4.1.3)$$

Except for sign, φ_{rr} (called the reflection coefficient)

is functionally the same as the partial auto-correlation function of lag (r-1) being defined as

$$\pi\,(r-1) = \frac{\text{cofactor of } (1,r) \text{ element of } \underset{\sim}{\Sigma}_r}{\text{cofactor of } (r,r) \text{ element of } \underset{\sim}{\Sigma}_r} \cdot \qquad (4.1.4)$$

Such statistical measure $\pi(\cdot)$ has a property that if the process is actually AR(m),

$$\begin{aligned} \pi(j) &\neq 0 \quad \text{for } j \leq m \\ &= 0 \quad \text{for } j > m \cdot \end{aligned} \qquad (4.1.5)$$

To ensure the order r of AR(r), $\mathscr{P}_{r+1,r+1}$ should be very close to zero. Thus Levinson-Durbin algorithm is found to be based essentially upon the very structure of Toeplitz matrix $\underset{\sim}{\Sigma}_r$.

In the area of digital signal processing, useful mechanical devices are frequently of such nature that their characteristics are not affected by time and temparature to generate or alter analog signals. The determination of the digital processor (called filter) giving the best signal estimate is a problem relevant to engineers. A more general approach is statistically based and involves developing the processor from the classical view point of extracting signal sequence from the noise-corrupted sequence. Conceivably, signal is modelled as a linear combination of its

past value and present and past values of noise (called input). In the frequency domain, this is equivalent to ARMA modeling (Makhoul, 1975).

Similarly, in the area of digital speech processing, while producing speech, speech waveform is quantized in a discrete amplitude representation or sampled in time. A "long-time" speech waveform may be treated as locally stationary [Giordano and Hsu, 1985, Ch. 11]. The present speech sample is predicted in terms of a linear combination of the previous sample or is equivalently modelled by an AR-model.

After the enormous development of computer technology, various other types of processing are also being developed, e.g. digital image processing, graphics and pattern recognition etc., all of which are treated specifically by appropriate stationary stochastic process through modeling by AR, MA or ARMA models [Pratt (1978), Rosenfled and Kak (1982)].

In the field of communication engineering, a common example involving Toeplitz matrx is as follows. Let us consider a point source located at the geocentre of a regular polygon with p vertices possessing identical signal receivers. Assume that the point source is capable of transmitting a signal of same strength in all directions and that covariances

depend only on the number of vertices separating the two receivers. Then the resulting covariance matrix is a p.d. Toeplitz matrix [cf. Olkin and Press, 1969].

By knowing the elements of the symmetric Toeplitz matrix ($\underset{\sim}{\Sigma}_m$), it is possible also to find out the central mass sequence ($\gamma_1, \gamma_2, \cdots, \gamma_m$) for $\underset{\sim}{\Sigma}_m$ defined as

$$\gamma_k = \text{Sup} \left\{ \gamma : (\underset{\sim}{\Sigma}_k - \gamma \underset{\sim}{J}_k) \text{ is p.d.} \right\}$$

$k = 1,2, \cdots, m$ in which $\underset{\sim}{J}_k$ is a $k \times k$ matrix with all elements unity. Caflisch (1981) has shown how this knowledge helps in solving direct and inverse problems of transmission lines consisting of piecewise constant components. To compute impedence ratio between two sections, the reflection coefficient [vide φ_{rr} in equation (4.4.3)] is made useful in Transmission line theory and in speech analysis (Dewilde, Vieria and Kailath, 1978).

In one-dimensional inverse problem of reflection seismology (Bube and Burridge, 1983), the problem of discretization of the Gopinath-Sondhi integral equation requires solving the system given by

$$\underset{\sim}{\Sigma}_\upsilon \underset{\sim}{\mathcal{L}}_\upsilon = \underset{\sim}{1}_\upsilon \ , \ 1 \leq \upsilon \leq m$$

where $\underset{\sim}{\Sigma}_\upsilon$ is the principal submatrix of Toeplitz matrix

$\underset{\sim}{\Sigma}_m$, $\underline{\varphi}_\upsilon$ is the υ-component vector of unknown and $\underline{1}_\upsilon = (1,1, \cdots , 1)'_{\upsilon x 1}$.

Multichannel and multidimensional processes very often have covariance matrices of block Toeplitz form. Akaike (1973), Kailath and Kalitracht (1986) have successfully dealt with the inversion problem of such forms. In multivariate spectral analysis in general and in power density spectrum analysis in particular, the essential ingradients are the auto-correlations. The eigenvalues of relevant p.d. Toeplitz matrix will trace out the so-called spectral curve of the process generating the data (Wise, 1955).

4.1.2 Plan of the Chapter

Various algebraic and statistical aspects of the p.d. Toeplitz matrix are discussed in this chapter. In section 4.2, we assemble important findings on some special matrices. This is followed by interesting and useful algebraic matrix results on the p.d. Toeplitz matrix. Although most of these results are previously established by various authors, these are available from widely scattered sources. We ,therefore, present these results in an organized way. Some propositions (not claimed to be original) are proved here in a simple manner, breeding an expectation that they might be useful in the theory of statistical analysis involving p.d. Toeplitz covariance structures.

being In the event of the patterned covariance matrix the Toeplitz covariance matrix, the multivariate statistical analysis very often becomes considerably simplified and shows interesting features. Conversely, these features may be otherwise helpful to anticipate the pattern of the covariance matrix or the model generating it. In section 4.3, we present various interesting results connected with the different measures of multivariate relationships such as multiple, partial, canonical correlations etc., and principal component analysis to highlight the computational simplicity arising from the Toeplitz pattern.

Although there are some general large sample test procedures which may be applied to test the tenability of the Toeplitz covariance structure, two-stage test procedures are proposed in section 4.4. These tests are conservative and based on centrosymmetric property of the Toeplitz matrix. The chapter is concluded by some remarks on various mathematical properties of the p.d. Toeplitz matrix and their connections with the estimation and testing of Toeplitz covariance structure.

4.2 SOME ALGEBRAIC ASPECTS

4.2.1 Certain Special Matrices and Associated Results

(i) The Flip Matrix : The flip matrix $\underset{\sim}{K}_m$ is an m x m

matrix with (i,j) element = 1 for j = m-i+1, and = 0 elsewhere. For example,

$$\underset{\sim}{K}_3 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

It is easy to check that

(a) $\underset{\sim}{K}_m = \underset{\sim}{K}_m' = \underset{\sim}{K}_m^{-1}$, $\underset{\sim}{K}_m^2 = \underset{\sim}{I}_m$ = an identity matrix of order m .

(b) If $\underset{\sim}{B}_{mxn} = \underset{\sim}{K}_m \underset{\sim}{A}_{mxn}$ then i-th row of $\underset{\sim}{B}$ = (m-i+1)th row of $\underset{\sim}{A}$. Similarly, if $\underset{\sim}{C}_{mxn} = \underset{\sim}{A}_{mxn} \underset{\sim}{K}_n$, the j-th column of $\underset{\sim}{C}$ = (m-j+1)th column of $\underset{\sim}{A}$. And if $\underset{\sim}{D}_{mxn} = \underset{\sim}{K}_m \underset{\sim}{A}_{mxn} \underset{\sim}{K}_n$, then (i,j) element of $\underset{\sim}{D}$ = (m-i+1 , m-j+1) element of $\underset{\sim}{A}$.

Thus, the flip matrix reverses the order of the rows and columns of any matrix on pre-and post-multiplication.

Owing to such property, the flip matrix has various other names, e.g., exchange matrix, contra identity matrix, counter identity matrix, reflection matrix etc.

(ii) Symmetric and Skew Symmetric Vectors

An m-dimensional vector $\underline{X} = (x_1, \cdots, x_m)$ will be called symmetric or skewsymmetric according as its components obeying the rule $x_j = x_{m+j-1}$ or $x_j = -x_{m+j-1}$, j = 1(1)m.

Introducing flip matrix $\underset{\sim}{K}_m$, $\underline{X}$ is symmetric or skew

symmetric according as $\underset{\sim}{K}_m \underline{X} = \underline{X}$ or $\underset{\sim}{K}_m \underline{X} = -\underline{X}$. In partitioned form, a symmetric vector is characterized by

$$\underset{p \times 1}{\underline{X}} = \left(\underset{m \times 1}{\underline{X}_1} , \underset{\sim}{K}_m \underline{X}_1 \right)' \quad \text{for } p = 2m$$

$$= \left(\underset{m \times 1}{\underline{X}_1} , x_{m+1} , \underset{\sim}{K}_m \underline{X}_1 \right)' \quad \text{for } p = 2m+1.$$

Also, a skew symmetric vector is characterized by

$$\underset{p \times 1}{\underline{X}} = \left(\underset{m \times 1}{\underline{X}_1} , -\underset{\sim}{K}_m X_1 \right)' \quad \text{for } p = 2m$$

$$= \left(\underset{m \times 1}{\underline{X}_1} , 0, -\underset{\sim}{K}_m \underline{X}_1 \right) \quad \text{for } p = 2m+1.$$

The relevance of symmetric vectors may be observed in the study of centrosymmetric matrices discussed in the next subsection.

(iii) Centrosymmetric Matrix

A square matrix $\underset{\sim}{A}(m \times m)$ is called centrosymmetric if $\underset{\sim}{K}_m \underset{\sim}{A} \underset{\sim}{K}_m = \underset{\sim}{A}$. This means that every element a_{ij} would obey the rule $a_{ij} = a_{m-i+1,\, m-j+1}$.

Many mathematicians (e.g. Haung and Cline, 1972) treat persymmetry and centrosymmetry as identical. We will maintain here a distinction following Aitken (1967). The rule for a persymmetric matrix is that $a_{ij} = a_{i+j-1}$ so that the elements on the secondary diagonal are same. Contrarily,

a centrosymmetric matrix is symmetric to its secondary diagonal. A symmetric centrosymmetric matrix (SC) is symmetric to both the principal diagonal and secondary diagonal. This type of matrix (SC) has some interesting properties as follows.

(a) $\underset{\sim}{K}_m \underset{\sim}{A}$ or $\underset{\sim}{A}\,\underset{\sim}{K}_m$ is a sym. centrosymmetric matrix (SC) whenever $\underset{\sim}{A}$ is SC .

(Its proof is elementary and is omitted.)

(b) $\underset{\sim}{A}^{-1}$ is SC whenever $\underset{\sim}{A}$ is a nonsingular SC.

Proof : Let $\underset{\sim}{M}_{ij}$ be the submatrix obtained by deleting i-th row and j-th column of $\underset{\sim}{A}$, i,j = 1(1) m . It is easy to check that $\underset{\sim}{M}_{ij} = \underset{\sim}{K}_{m-1}\, \underset{\sim}{M}_{m-i+1,m-j+1}\, \underset{\sim}{K}_{m-1}$ ensuring that the cofactors corresponding to (i,j) and (m-i+1,m-j+1) elements of $\underset{\sim}{A}$, e.g. $(-1)^{i+j} \det(\underset{\sim}{M}_{ij})$ and $(-1)^{(m-i+1)+(m-j+1)} \det(\underset{\sim}{M}_{m-i+1,m-j+1})$, are equal. Asaconsequence, adjugate matrix and hence the inverse of $\underset{\sim}{A}$ are SC.

(c) Deletion of any r-th row and corresponding column and (m-r+1)th row and corresponding column of an SC matrix preserves the SC property.

Proof : Let b_{ij} be the typical element of the new matrix $\underset{\sim}{B}(\overline{m-2} \times \overline{m-2})$ as proposed from an SC matrix $\underset{\sim}{A}(m \times m)$. Let us consider the three index sets, namely I_1 : $(1 \leq i \leq r-1)$, I_2 : $(r \leq i \leq m-r-1)$, I_3 : $(m-r \leq i \leq m-2)$. It is easy to

establish that (a) for any $i \in I_k$, $\overline{m-2}-i+1 \in I_{4-k}$, $k = 1,2,3$, and (b) for any $i \in I_k$ and for any $j \in I_\ell$, $b_{ij} = a_{i+k-1, j+\ell-1}$, $k, \ell = 1,2,3$.

Let us now consider any certain $i \in I_k$ and $j \in I_\ell$. Then due to (a), $\overline{m-2}-i+1 \in I_{4-k}$ and $\overline{m-2}-j+1 \in I_{4-\ell}$ so that due to (b),

$$b_{\overline{m-2}-i+1, \overline{m-2}-j+1} = a_{\overline{m-2}-i+1+\overline{4-k}-1,\ m-2-j+1+\overline{4-\ell}-1}$$

$$= a_{m-i-k+2,\ m-j-\ell+2}$$

$$= a_{i+k-1,\ j+\ell-1} \qquad \text{(due to SC } \underset{\sim}{A}\text{)}$$

$$= b_{ij}, \qquad \text{(due to (b))}$$

ensuring $\underset{\sim}{B}$ as an SC matrix. ∎

Remark 4.1 Intuitively, it is the (m-i+1)th row (or column) which is related to i-th row (or column) for an SC matrix. So omission of i-th and (m-i+1)-th row and the corresponding columns will not affect the centrosymmetric property of the remaining (m-2) rows and (m-2) columns. ∎

The characterization of SC matrices may be given in partitioned form as follows.

$$\underset{p \times p}{\underset{\sim}{A}} = \begin{bmatrix} \underset{m \times m}{\underset{\sim}{A}_{11}} & \underset{m \times m}{\underset{\sim}{A}_{12}} \\ \underset{\sim}{K}_m \underset{\sim}{A}_{12} \underset{\sim}{K}_m & \underset{\sim}{K}_m \underset{\sim}{A}_{11} \underset{\sim}{K}_m \end{bmatrix} \quad \text{for } p = 2m \qquad (4.2.1)$$

$$= \begin{bmatrix} \underset{m \times m}{\underset{\sim}{A}_{11}} & \underset{m \times 1}{\underline{a}} & \underset{m \times m}{\underset{\sim}{A}_{12}} \\ \underline{a}' & a_{m+1,m+1} & \underline{a}' \underset{\sim}{K}_m \\ \underset{\sim}{K}_m \underset{\sim}{A}_{12} \underset{\sim}{K}_m & \underset{\sim}{K}_m \underline{a} & \underset{\sim}{K}_m \underset{\sim}{A}_{11} \underset{\sim}{K}_m \end{bmatrix} \quad \text{for } p = 2m+1. \qquad (4.2.2)$$

where $\underset{\sim}{A}_{11} = \underset{\sim}{A}'_{11}$ and $\underset{\sim}{A}_{12} = \underset{\sim}{K}_m \underset{\sim}{A}_{12} \underset{\sim}{K}_m$. We can check SC property of any symmetric matrix $\underset{\sim}{A}$ from its partitioned form given by

$$\underset{p \times p}{\underset{\sim}{A}} = \begin{bmatrix} \underset{m \times m}{\underset{\sim}{A}_{11}} & \underset{\sim}{A}_{12} \\ \underset{\sim}{A}'_{12} & \underset{m \times m}{\underset{\sim}{A}_{22}} \end{bmatrix} \quad \text{for } p = 2m. \qquad (4.2.3)$$

$$= \begin{bmatrix} \underset{\sim}{A}_{11} & \underline{a} & \underset{\sim}{A}_{12} \\ \underline{a}' & a_{m+1,m+1} & \underline{b} \\ \underset{\sim}{A}'_{12} & \underline{b}' & \underset{\sim}{A}_{22} \end{bmatrix} \quad \text{for } p = 2m+1. \qquad (4.2.4)$$

To do this, we may use an orthogonal matrix $\underset{\sim}{P}_p$ given by

$$\underset{\sim}{P}_p = \frac{1}{\sqrt{2}} \begin{bmatrix} \underset{\sim}{I}_m & \underset{\sim}{K}_m \\ \underset{\sim}{I}_m & -\underset{\sim}{K}_m \end{bmatrix} \quad \text{for } p = 2m \qquad (4.2.5)$$

$$= \frac{1}{\sqrt{2}} \begin{bmatrix} \underset{\sim}{I}_m & 0 & \underset{\sim}{K}_m \\ 0 & \sqrt{2} & 0 \\ \underset{\sim}{I}_m & 0 & -\underset{\sim}{K}_m \end{bmatrix} \quad \text{for } p = 2m+1 \qquad (4.2.6)$$

to calculate $\underset{\sim}{P}_p \underset{\sim}{A} \underset{\sim}{P}_p'$. It is easy to verify that the off-diagonal blocks of $\underset{\sim}{P}_p \underset{\sim}{A} \underset{\sim}{P}_p'$ would vanish when $\underset{\sim}{A}_{22} = \underset{\sim}{K}_m \underset{\sim}{A}_{11} \underset{\sim}{K}_m$, $\underset{\sim}{A}_{12}' = \underset{\sim}{K}_m \underset{\sim}{A}_{12} \underset{\sim}{K}_m$ and $\underline{b} = \underset{\sim}{K}_{m-1}\, \underline{a}$, thus ensuring SC property of $\underset{\sim}{A}$. In effect, an SC matrix $\underset{\sim}{A}$ is orthogonally similar to Diag. $(\underset{\sim}{A}_{11} + \underset{\sim}{A}_{12}\underset{\sim}{K}_m\ ,\ \underset{\sim}{A}_{11} - \underset{\sim}{A}_{12}\underset{\sim}{K}_m)$ for p = 2m, or to

$$\begin{bmatrix} \underset{\sim}{A}_{11} + \underset{\sim}{A}_{12}\underset{\sim}{K}_m & \sqrt{2}\ \underline{a} & 0 \\ \sqrt{2}\ \underline{a} & a_{m+1,m+1} & 0 \\ 0 & 0 & \underset{\sim}{A}_{11} - \underset{\sim}{A}_{12}\underset{\sim}{K}_m \end{bmatrix} \quad \text{for } p = 2m+1.$$

It is now obvious that eigenvalues of SC $\underset{\sim}{A}(pxp)$ are the eigenvalues of $\underset{\sim}{A}_{11} + \underset{\sim}{A}_{12}\underset{\sim}{K}_m$ and $\underset{\sim}{A}_{11} - \underset{\sim}{A}_{12}\underset{\sim}{K}_m$ for even $p = 2m$, or of $\begin{bmatrix} \underset{\sim}{A}_{11} + \underset{\sim}{A}_{12}\underset{\sim}{K}_m & \sqrt{2}\ \underline{a} \\ \sqrt{2}\ \underline{a} & a_{m+1,m+1} \end{bmatrix}$ and $\underset{\sim}{A}_{11} - \underset{\sim}{A}_{12}\underset{\sim}{K}_m$ for odd $p = 2m+1$.

It may be shown that a sufficient (but not necessary) condition for a pxp matrix $\underset{\sim}{A}$ to be centrosymmetric is that it has p linearly independent eigenvectors in the set ($\mathcal{S}$) of vectors which are either symmetric or skewsymmetric. Furthermore, a necessary condition (not sufficient) for $\underset{\sim}{A}$ to be centrosymmetric is that every eigenspace has a basis consisting of vectors in $\mathcal{S}$ [Andrew, 1973]. Thus, in particular, if the eigenvalues of $\underset{\sim}{A}$ are all distinct, then $\underset{\sim}{A}$ is centrosymmetric if and only if all its eigenvectors are in $\mathcal{S}$

[Collar (1962), Cord and Sylvester (1962)]. In fact, the possibility of getting neither symmetric nor skewsymmetric eigenvectors is obvious when both the diagonal blocks (for example, $\underset{\sim}{A}_{11} \pm \underset{\sim}{A}_{12}\underset{\sim}{K}_m$ for even $p = 2m$) have one eigenvalue in common [Cantoni and Butler (1976)].

(iv) Symmetric Toeplitz Matrix

A matrix $\underset{\sim}{\Sigma}_m(m \times m)$ is called real symmetric Toeplitz matrix if its elements σ_{ij} obey the rule $\sigma_{ij} = a_{|i-j|}$ for all $i,j = 1(1)m$. The matrix $\underset{\sim}{\Sigma}_m$ is a function of m parameters, i.e. $\underset{\sim}{\Sigma}_m = \underset{\sim}{\Sigma}_m(a_0, a_1, \cdots, a_{m-1})$ being expressed as follows:

$$\underset{\sim}{\Sigma}_m = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{m-1} \\ a_1 & a_0 & a_1 & \cdots & a_{m-2} \\ a_2 & a_1 & a_0 & \cdots & a_{m-3} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m-1} & a_{m-2} & a_{m-3} & \cdots & a_0 \end{bmatrix} . \qquad (4.2.7)$$

Glass, Wilson and Gottman (1975) refer to such a matrix as a " stripe" matrix. In its most general (nonsymmetric) form, Toeplitz matrix has the structure $\underset{\sim}{T}_m = ((t_{ij}))_{m \times m}$ with $t_{ij} = a_{i-j}$.

A nonsingular matrix T_m may always be factored as

$$\underset{\sim}{T}_m = \underset{\sim}{A}\ \underset{\sim}{D}\ \underset{\sim}{B} \qquad (4.2.8)$$

where $\underset{\sim}{D}$ is a suitable diagonal matrix, and $\underset{\sim}{A}$ and $\underset{\sim}{B}$ are two Vandermonde matrices, the rows of $\underset{\sim}{A}$ being $(1,1, \cdots , 1)$, $(\gamma_1^{-1} , \gamma_2^{-1} , \cdots , \gamma_m^{-1})$, $(\gamma_1^{-2} , \gamma_2^{-2} , \cdots , \gamma_m^{-2})$ and so on. The columns of $\underset{\sim}{B}$ are $(1,1, \cdots, 1)'$, $(\gamma_1, \gamma_2 \cdots , \gamma_m)'$, $(\gamma_1^2 , \gamma_2^2 , \cdots , \gamma_m^2)'$ and so on where $\gamma_1, \gamma_2 \cdots , \gamma_m$ are suitable nonzero numbers.

Matrix $\underset{\sim}{\Sigma}_m$ allows some useful representations as follows.

$$\underset{\sim}{\Sigma}_m = a_o \sum_{k=o}^{m-1} \rho_k \underset{\sim}{H}_k = a_o \underset{\sim}{\rho}_m \qquad (4.2.9)$$

where $a_o \neq 0$, $\rho_k = a_k / a_o$, and each of the design matrices $\underset{\sim}{H}_k$ has components $h_{ij}^{(k)} = 1$ whenever $|i-j| = k$, and $= 0$ elsewhere. $\underset{\sim}{\rho}_m = \sum_{k=o}^{m-1} \rho_k \underset{\sim}{H}_k$ is called m x m Toeplitz correlation matrix, very often expressible as the difference between the products of two sets of m x m triangular Toeplitz matrices as

$$\underset{\sim}{\rho}_m = \underset{\sim}{G}\underset{\sim}{G}' - (\underset{\sim}{G} - \underset{\sim}{I}_m)(\underset{\sim}{G} - \underset{\sim}{I}_m)' \qquad (4.2.10)$$

where $\underset{\sim}{G}$ is a lower triangular matrix with unity in the principal diagonal and the other elements exactly the same as in the lower triangular part of the matrix $\underset{\sim}{\rho}_m$.

The "stripe" matrix $\underset{\sim}{\Sigma}_m$ allows another representation as a finite "power series expansion" [Whittle, 1951, p. 33] given by

$$\underset{\sim}{\Sigma}_m = a_o \underset{\sim}{I}_m + a_1(\underset{\sim}{U}_m + \underset{\sim}{U}_m') + a_2(\underset{\sim}{U}_m^2 + \underset{\sim}{U}_m'^2) + \cdots$$

$$\cdots + a_{m-1}(\underset{\sim}{U}_m^{m-1} + \underset{\sim}{U}_m'^{m-1}), \qquad (4.2.11)$$

where $\underset{\sim}{U}_m$ is the shift matrix as defined latter in (vii). We note here that $\underset{\sim}{H}_m = \underset{\sim}{U}_m^k + \underset{\sim}{U}_m'^k$, k = 1(1)(m-1) and conventionally, $\underset{\sim}{H}_o = \underset{\sim}{I}_m$.

(v) Hankel matrix

Any matrix $\underset{\sim}{F} = ((f_{ij}))_{mxm}$ satisfying $f_{ij} = f_{i+j}$, i, j = 0(1)(m-1) for some arbitrary numbers f_o, f_1, ..., f_{2m-2} is called a Hankel matrix. (Gantmacher, 1959).

A Hankel matrix is always persymmetric, but this is not true of a symmetric Toeplitz matrix. But, if $\underset{\sim}{F}$(mxm) is Hankel, $\underset{\sim}{K}_m\underset{\sim}{F}$ and $\underset{\sim}{F}\underset{\sim}{K}_m$ are nonsymmetric Toeplitz, and vice versa, if $\underset{\sim}{T}$ is a nonsymmetric Toeplitz, $\underset{\sim}{K}_m\underset{\sim}{T}$ and $\underset{\sim}{T}\underset{\sim}{K}_m$ are Hankel. Algebraic and operator theory of Hankel matrices are available in the books of Iohvidov (1982) as well as Heinig and Rost (1984).

(vi) Schur Complement

When $\underset{\sim}{A}$(or $\underset{\sim}{D}$)is nonsingular, the matrix $\underset{\sim}{D}-\underset{\sim}{C}\underset{\sim}{A}^{-1}\underset{\sim}{B}$ (or $\underset{\sim}{A}-\underset{\sim}{H}\underset{\sim}{D}^{-1}\underset{\sim}{C}$), appearing in the calculation of the determinant of any supermatrix $\underset{\sim}{F}$ by a partition method such as

$$\det(\underset{\sim}{F}) = \det\begin{bmatrix} \underset{\sim}{A} & \underset{\sim}{B} \\ \underset{\sim}{C} & \underset{\sim}{D} \end{bmatrix} = \det(\underset{\sim}{D})\ \det\ (\underset{\sim}{A}-\underset{\sim}{B}\underset{\sim}{D}^{-1}\underset{\sim}{C})$$

$$= \det(\underset{\sim}{A})\ \det(\underset{\sim}{D}-\underset{\sim}{C}\underset{\sim}{A}^{-1}\underset{\sim}{B}) \qquad (4.2.12)$$

is called the Schur complement of $\underset{\sim}{A}$(or $\underset{\sim}{D}$) in $\underset{\sim}{F}$. We refer to Haynsworth (1968) for the definition and to Ouellet (1981) for a general discussion of Schur complements in Statistics.

Huang and Cline (1972) proved the theorem that a nonsingular $\underset{\sim}{E}$ is Toeplitz iff

(i) $\underset{\sim}{E}^{-1}$ is centrosymmetric and

(ii) the Schur complement of $\underset{\sim}{A}$(or $\underset{\sim}{D}$) in $\underset{\sim}{E}$ is also centrosymmetric.

Following Haynsworth (1968), it can also be shown that the inertia of a partitioned Toeplitz matrix $\underset{\sim}{E}$ as shown above can be expressed as

$$In(\underset{\sim}{E}) = In(\underset{\sim}{A}) + In(\underset{\sim}{D}-\underset{\sim}{C}\underset{\sim}{A}^{-1}\underset{\sim}{B})$$

$$= In(\underset{\sim}{D}) + In(\underset{\sim}{A}-\underset{\sim}{B}\underset{\sim}{D}^{-1}\underset{\sim}{C}) \qquad (4.2.13)$$

where the symbol $In(\underset{\sim}{X})$ refers to the inertia of a square matrix $\underset{\sim}{X}$, which is a triplet consisting of number of positive eigenvalues, number of negative eigenvalues and number of zero eigenvalues of the $\underset{\sim}{X}$ matrix.

(vii) Shift Matrix

A matrix $\underset{\sim}{U}_m = ((u_{ij}))_{mxm}$ with $u_{ij} = 1$ for $i = j+1$, $j = 1,2, \cdots , (m-1)$, and $= 0$ elsewhere, is called a shift matrix. For example,

$$\underset{\sim}{U}_4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} .$$

More specifically, $\underset{\sim}{U}_m$ is a matrix of forward shift operator in the sense that for $\underline{X} = (x_1, x_2, \cdots , x_m)'$,

$$\underset{\sim}{U}_m \underline{X} = (0, x_1, x_2, \cdots , x_{m-1})' . \qquad (4.2.14)$$

$\underset{\sim}{U}_m'$ is, on the other hand, called a matrix of backward shift operator, as

$$U_m' \underline{X} = (x_2, x_3, \cdots , x_m, 0)' . \qquad (4.2.15)$$

For such $\underset{\sim}{U}_m$-matrices, we have got the following useful results:

(a) $\underset{\sim}{U}_m^{-1} = \underset{\sim}{U}_m' = \underset{\sim}{K}_m \underset{\sim}{U}_m \underset{\sim}{K}_m$.

(b) t-th power of $\underset{\sim}{U}_m = \underset{\sim}{U}_m^t = \begin{bmatrix} \underset{\sim}{0} & \underset{\sim}{0} \\ t-1 \underset{\sim}{x} t-1 & \\ \underset{\sim}{U}_{m-t+1} & \underset{\sim}{0} \end{bmatrix}$ for $t < m$

$= \underset{\sim}{0}$ for $t \geq m$.

Hence, any $\underset{\sim}{U}_m^t$ is a nilpotent matrix with index m-t, t < m.

(c) $\underset{\sim}{U}_m^r \; \underset{\sim}{U}_m^{'s} = \text{diag.}\ (\underset{\sim}{0}_r\ ,\ \underset{\sim}{U}_{m-r}^{'s-r})$ for r < s

$= \text{diag.}\ (\underset{\sim}{0}_r\ ,\ \underset{\sim}{I}_{m-r})$ for r = s

$= \text{diag.}\ (\underset{\sim}{0}_s\ ,\ \underset{\sim}{U}_{m-s}^{r-s})$ for r > s .

Similarly,

$\underset{\sim}{U}_m^{'r} \; \underset{\sim}{U}_m^s = \text{diag.}\ (\underset{\sim}{U}_{m-r}^{s-r}\ ,\ \underset{\sim}{0}_r)$ for r < s

$= \text{diag.}\ (\underset{\sim}{I}_{m-r}\ ,\ \underset{\sim}{0}_r)$ for r = s

$= \text{diag.}\ (\underset{\sim}{U}_{m-s}^{'r-s}\ ,\ \underset{\sim}{0}_s)$ for r > s .

(d) Singular values of $\underset{\sim}{U}_m^t$ are the positive square root of the eigenvalues of $\underset{\sim}{U}_m^t \; \underset{\sim}{U}_m^{'t}$, which are only 1 and 0 appearing (m-t) times and t times respectively. This ensures that for all $\underset{\sim}{U}_m^t$, t = 1(1)(m-1), spectral radii (spectral norms) are the same, namely, unity. But contrarily, all the eigenvalues of $\underset{\sim}{U}_m^t$ are nothing but zero.

(e) Following Bellman (1974, p. 219), $\det(\underset{\sim}{U}_m^t + \underset{\sim}{A})$ remains the same as $\det(\underset{\sim}{A})$ for any mxm matrix $\underset{\sim}{A}$.

(f) $\text{Tr}(\underset{\sim}{U}_m^t) = 0$, $\text{Tr}(\underset{\sim}{U}_m^t \; \underset{\sim}{U}_m^{'s}) = 0$ for all t, s.

The proofs of (a) - (f) are straight forward and are omitted.

We shall now consider an effective use of the shift matrix.

(g) For any matrix $\underset{\sim}{S}(m \times m)$, Kailath, Kung and Morf (1979) defined the ranks of $\rfloor\underset{\sim}{S}$ and $\lceil\underset{\sim}{S}$ as the pair of <u>displacement ranks</u> where

$$\rfloor\underset{\sim}{S} = \underset{\sim}{S} - \underset{\sim}{U}_m \underset{\sim}{S} \underset{\sim}{U}_m' \quad , \quad \lceil\underset{\sim}{S} = \underset{\sim}{S} - \underset{\sim}{U}_m' \underset{\sim}{S} \underset{\sim}{U}_m \quad . \tag{4.2.16}$$

The difference of the aforesaid displacement ranks does not exceed 2. For a nonsingular $\underset{\sim}{S}$, rank$(\rfloor\underset{\sim}{S})$ = rank$(\lceil\underset{\sim}{S}^{-1})$ and rank $(\lceil\underset{\sim}{S})$ = rank $(\rfloor\underset{\sim}{S}^{-1})$. Interestingly, for a symmetric Toeplitz matrix $(\underset{\sim}{\Sigma}_m)$, the two displacement ranks are both equal to 2(unless $\underset{\sim}{\Sigma}_m$ is diagonal or null), as we note here that

$$\rfloor\underset{\sim}{\Sigma}_m = \begin{bmatrix} a_0 & a_1 \cdots a_{m-1} \\ a_1 & \\ \vdots & 0 \\ a_{m-1} & \end{bmatrix} , \lceil\underset{\sim}{\Sigma}_m = \begin{bmatrix} & & a_{m-1} \\ & & \vdots \\ & 0 & \vdots \\ & & a_1 \\ a_{m-1} \cdots & a_1 & a_0 \end{bmatrix} . \tag{4.2.17}$$

Keeping the idea of displacement rank in view, we are now employing, as its generalization, a pair of k-displacement ranks of any $\underset{\sim}{S}(m \times m)$ as

$$\rfloor_k \underset{\sim}{S} = \underset{\sim}{S} - \underset{\sim}{U}_m^k \underset{\sim}{S} \underset{\sim}{U}_m'^k \quad , \quad \lceil_k \underset{\sim}{S} = \underset{\sim}{S} - \underset{\sim}{U}_m'^k \underset{\sim}{S} \underset{\sim}{U}_m^k \quad .$$

The corresponding pair of ranks may be called the k-displacement ranks of $\underset{\sim}{S}$. Since $\underset{\sim}{U}_m^k \underset{\sim}{\Sigma}_m \underset{\sim}{U}_m'^k$ = diag.$(\underset{\sim}{O}_k, \underset{\sim}{\Sigma}_{m-k})$

and $\underset{\sim}{U}_m^{'k}\ \underset{\sim}{\Sigma}_m\ \underset{\sim}{U}_m^{k} = \text{diag.}(\underset{\sim}{\Sigma}_{m-k},\ \underset{\sim}{O}_k)$, the k-displacement ranks of $\underset{\sim}{\Sigma}_m$ are equal to (k+1) for k = 1,2, ... , (m-1).

An extensive study, in this connection, is available in Heinig and Rost (1984, Ch. II(2)) where they defined a single rank, called Toeplitz-rank based on $(\underset{\sim\sim}{S}\underset{\sim}{U}_m - \underset{\sim}{U}_m\underset{\sim}{S})$, as a simplification of the notion of pair of displacement ranks of $\underset{\sim}{S}$.

(viii) Some Simple Results on the Design Matrices $\underset{\sim}{H}_k$.

In order to obtain some results based on different orders of $\underset{\sim}{H}_k$-matrices, let us henceforth denote an m x m $\underset{\sim}{H}_k$-matrix by $\underset{\sim}{H}_k(m)$. The following six results are given below in this connection :

(a) $\underset{\sim}{H}_k(m) = \underset{\sim}{I}_m^k + \underset{\sim}{U}_m^{'k}$, k = 1(1)(m-1)

$\underset{\sim}{H}_o(m) = \underset{\sim}{I}_m$(conventionally) and $\underset{\sim}{H}_k(m) = \underset{\sim}{O}$ for $k \geq m$.

(b) $\underset{\sim}{H}_r^2(m) = \underset{\sim}{H}_{2r}(m) + \underset{\sim}{I}_m + \text{diag.}(\underset{\sim}{O}_r, \underset{\sim}{I}_{m-2r}, \underset{\sim}{O}_r)$, for $m > 2r$

$= \underset{\sim}{I}_m$, for $m = 2r$

$= \underset{\sim}{H}_{2r}(m) + \underset{\sim}{I}_m - \text{diag.}(\underset{\sim}{O}_{m-r}, \underset{\sim}{I}_{2r-m}, \underset{\sim}{O}_{m-r})$, for $m < 2r$.

(c) $\underset{\sim}{H}_r(m)\underset{\sim}{H}_s(m) = \underset{\sim}{H}_{r+s}(m) + \text{diag.}(\underset{\sim}{U}_{m-r}^{s-r}, \underset{\sim}{O}_r) + \text{diag.}(\underset{\sim}{O}_r, \underset{\sim}{U}_{m-r}^{'s-r})$, for $r < s$

$= \underset{\sim}{H}_{r+s}(m) + \text{diag.}(\underset{\sim}{U}_{m-r}^{'s-r}, \underset{\sim}{O}_r) + \text{diag.}(\underset{\sim}{O}_r, \underset{\sim}{U}_{m-r}^{s-r})$, for $r > s$.

(d) $\underset{\sim}{H}_r(m)\underset{\sim}{H}_s(m) + \underset{\sim}{H}_s(m)\underset{\sim}{H}_r(m)$

$= 2\underset{\sim}{H}_{r+s}(m) + \text{diag.}\ (\underset{\sim}{H}_{s-r}(m-r),\ \underset{\sim}{0}_r)$

$+ \text{diag.}(\underset{\sim}{0}_r, \underset{\sim}{H}_{s-r}(m-r))$, for $r < s$.

From this

$$\underset{\sim}{H}_r(m)\underset{\sim}{H}_s(m) + \underset{\sim}{H}_s(m)\underset{\sim}{H}_r(m) = 2\underset{\sim}{H}_{r+s}(m) + \underset{\sim}{H}_{s-r}(m)$$

$$+ \text{diag.}(\underset{\sim}{0}_r, \underset{\sim}{H}_{s-r}(m-2r), \underset{\sim}{0}_r)$$

for $s-r<m-2r$

$$= 2\underset{\sim}{H}_{r+s}(m) + \underset{\sim}{H}_{s-r}(m)$$

$$- \text{diag.}(\underset{\sim}{0}_{m-s}, \underset{\sim}{H}_{s-r}(2s-m), \underset{\sim}{0}_{m-s})$$

for $s-r \geq m-2r$.

The results (a) - (d) may be proved easily using the properties of $\underset{\sim}{U}_m$-matrix.

(e) Eigenvalues of $\underset{\sim}{H}_t(m)$ are less than or equal to 2 for all $t < m$.

Proof : Apply a theorem by KyFan and Hoffman (1955) stating that for a square matrix $\underset{\sim}{A}$, any eigenvalue of $(\underset{\sim}{A}+\underset{\sim}{A}')$ is less than or equal to twice the corresponding singular value of $\underset{\sim}{A}$.

(f) Maximum eigenvalue of $\underset{\sim}{H}_t(m)$ is $2 \cos \{\pi/(1+[(m+t-1)/t])\}$ which is less than or equal to $2 \cos \{t\pi/(m+2t-1)\}$ where notation [p] stands for the greatest integer contained in p .

Further, minimum eigenvalue of $\underset{\sim}{H}_t(m)$ is the negative of the maximum eigenvalue.

Proof : We refer to Siddiqui (1958) for a proof.

4.2.2 Some Propositions on Symmetric Toeplitz Matrix

Unless otherwise specified, we shall consider in this section Toeplitz matrices which are real and symmetric.

Proposition 4.2.1 : Any principal submatrix of order r x r of a Toeplitz matrix of order m x m, say $\underset{\sim}{\Sigma}_m (r < m)$, is a Toeplitz matrix.

Proof : This is so because omission of successive rows starting from the 1st row (or from the last row) and the corresponding columns from a Toeplitz matrix does not destroy the Toeplitz property. ■

Using the shift-matrix $\underset{\sim}{U}_m$, we may give still another representation of the symmetric Toeplitz matrix as a sequel to Proposition 4.2.1. Let $\underset{\sim}{U}_m^*$ be the matrix $\underset{\sim}{U}_m$ after its last k columns have been deleted. It may be checked now that

$$\underset{\sim}{\Sigma}_{m-k} = \underset{\sim}{U}_m^{*\prime} \underset{\sim}{\Sigma}_m \underset{\sim}{U}_m^* \qquad (4.2.18)$$

where $\underset{\sim}{\Sigma}_{m-k}$ is the resultant symmetric Toeplitz matrix of dimension (m-k), k = 1, 2, ... , m-1.

Proposition 4.2.2 : A symmetric Toeplitz matrix $\underset{\sim}{\Sigma}_m$ is always symmetric centrosymmetric (SC), but the converse is not necessarily true.

Proof : The (i,j) element of $\underset{\sim}{\Sigma}_m$ is given by

$$\sigma_{ij} = a^{|i-j|} = a^{|m-i+1-m+j-1|} = \sigma_{m-i+1,m-j+1}$$

$\forall$ i,j = 1(1)m . This obeys the rule of centrosymmetry. Clearly, the converse is not true. ∎

Due to Proposition 4.2.2, it is obvious that a Toeplitz matrix must possess all the properties of SC matrix.

We may now characterize a Toeplitz matrix by a pair of centrosymmetric matrices in the following proposition.

Proposition 4.2.3 : Let $\underset{\sim}{A}_{m-1}(\overline{m-1} \times \overline{m-1})$ be the leading principal submatrix obtained by deleting the last (or first) row and last (or first) column of a matrix $\underset{\sim}{A}_m(m \times m)$. Then $\underset{\sim}{A}_m$ is a Toeplitz matrix if and only if both $\underset{\sim}{A}_{m-1}$ and $\underset{\sim}{A}_m$ are centrosymmetric.

Proof : The proof for sufficiency of the condition is obvious via Proposition 4.2.2. To prove the necessity part of the condition, we proceed as follows :

As $\underset{\sim}{A}_{m-1}$ and $\underset{\sim}{A}_m$ are centrosymmetric,

$$a_{ij} = a_{m-i,m-j} \ , \ i,j = 1(1)(m-1) \ , \qquad (4.2.19)$$

and

$$a_{ij} = a_{m-i+1,\ m-j+1} \ , \ i,j = 1(1)m \ . \qquad (4.2.20)$$

Putting $j = i+k$, we obtain

$$a_{i,i+k} = a_{m-i+1,\ m-i+1-k} = a_{m-i,m-i-k} \qquad (4.2.21)$$

and

$$a_{m-k,m} = a_{m+1,1} \qquad (4.2.22)$$

for all $i = 1(1)\ (m-1-k)$, $k = 0(1)(m-1)$. From these we have, in addition to the symmetry of $\underset{\sim}{A}_m$,

$$a_{1,k+1} = a_{2,k+2} = \cdots = a_{m-k,m} \ , \ k = 0(1)(m-1). \qquad (4.2.23)$$

This ensures that $\underset{\sim}{A}_m$ is a m x m array of m distinct elements as in (4.2.7). ■

A characterization of a Toeplitz matrix similar to Proposition 4.2.3 has been given by Huang and Cline (1972) in an alternative manner. They also obtained a simple criterion for recognizing an inverse of a matrix to be a Toeplitz matrix.

We shall now give the general representation of the Toeplitz matrix (4.2.7) in the light of centrosymmetric matrix in partitioned form (vide equations (4.2.1) and (4.2.2)). For even order,

$$\underset{\sim}{\Sigma}_{2m} = \underset{\sim}{\Sigma}_{2m}\,(a_o,\ a_1,\ \cdots\ ,\ a_{m-1};\ a_m,\ \cdots\ ,\ a_{2m-1})$$

$$= \begin{bmatrix} \underset{\sim}{\Sigma}_m & \underset{\sim}{\Delta}_m \\ \underset{\sim}{\Delta}_m' & \underset{\sim}{\Sigma}_m \end{bmatrix}, \qquad (4.2.24)$$

where $\underset{\sim}{\Sigma}_m = \underset{\sim}{\Sigma}_m(a_o,\ a_1,\ \cdots\ ,\ a_{m-1})$ and

$$\underset{\sim}{\Delta}_m = \begin{bmatrix} a_m & a_{m+1} & \cdots & a_{2m-1} \\ a_{m-1} & a_m & \cdots & a_{2m-2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_1 & a_2 & & a_m \end{bmatrix} = \underset{\sim}{K}_m\,\underset{\sim}{\Delta}_m'\,\underset{\sim}{K}_m\,. \qquad (4.2.25)$$

For odd order,

$$\underset{\sim}{\Sigma}_{2m+1} = \underset{\sim}{\Sigma}_{2m+1}\,(a_o,\ a_1,\ \cdots\ ,\ a_{m-1};\ a_m;\ a_{m+1};\ \cdots,\ a_{2m})$$

$$= \begin{bmatrix} \underset{\sim}{\Sigma}_m & \underset{\sim}{K}_m\underline{\beta}_m & \underset{\sim}{\Delta}_m^* \\ \underline{\beta}_m'\underset{\sim}{K}_m & a_o & \underline{\beta}_m' \\ \underset{\sim}{\Delta}_m^{*\prime} & \underline{\beta}_m & \underset{\sim}{\Sigma}_m \end{bmatrix} \qquad (4.2.26)$$

where $\underline{\beta}_m = (a_1,\ a_2,\ \cdots\ ,\ a_{m-1}\ ,\ a_m)'$ and

$$\underset{\sim}{\Delta}_m^* = \begin{bmatrix} a_{m+1} & a_{m+2} & \cdots & a_{2m} \\ a_m & a_{m+1} & \cdots & a_{2m-1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_2 & a_3 & & a_{m+1} \end{bmatrix} = \underset{\sim}{K}_m\underset{\sim}{\Delta}_m^{*\prime}\underset{\sim}{K}_m\,. \qquad (4.2.27)$$

4.2.3 Determinant of a Symmetric Toeplitz Matrix

The determinant of a positive definite (p.d.) Toeplitz matrix has been approximated by Whittle (1951, p. 39). But this does not work in the case of Toeplitz correlation matrix $\underset{\sim}{\rho}_m$ (vide (4.2.9)). Daniel (1956) has shown that $\det(\underset{\sim}{\rho}_m)$ can be expressed exactly in terms of partial correlations. If we denote ρ_{j*} to mean the partial correlation coefficient between X_t and X_{t+j} conditional on fixed X_{t+1}, ... , X_{t+j-1} (which may be called the j-th leading partial serial correlation coefficient), then it can be shown that

$$\det(\underset{\sim}{\rho}_m) = (1-\rho_1^2)^{m-1} (1 - \rho_{2*}^2)^{m-2} \ldots (1-\rho_{m-1*}^2). \qquad (4.2.28)$$

If $\det(\underset{\sim}{\rho}_{m-1})$ and $\underset{\sim}{\rho}_{m-1}^{-1}$ are already known, then by direct use of the Schur complement,

$$\det(\underset{\sim}{\rho}_m) = (1-\underline{\rho}^{\ell\prime}\ \underset{\sim}{\rho}_{m-1}^{-1}\ \underline{\rho}^{\ell})\det(\underset{\sim}{\rho}_{m-1}) \qquad (4.2.29)$$

where $\underline{\rho}^{\ell} = (\rho_1 , \rho_2, \ldots , \rho_{m-1})'$.

In general, for calculating $\det(\underset{\sim}{\Sigma}_m)$ (when $\underset{\sim}{\Sigma}_m$ is not necessarily p.d.), we may apply the following three different approaches, each of which can be easily extended to block symmetric Toeplitz matrices.

(i) Reduction by Splitting into Halves

Using a suitable orthogonal matrix (vide (4.2.5) and (4.2.6)), $\underset{\sim}{\Sigma}_m$ be block diagonalized so that its determinant

is split into a product of two determinants. For even order,

$$\det(\underset{\sim}{\Sigma}_{2m}) = \det(\underset{\sim}{\Sigma}_m + \underset{\sim}{\Delta}_m\underset{\sim}{K}_m)\ \det(\underset{\sim}{\Sigma}_m - \underset{\sim}{\Delta}_m\underset{\sim}{K}_m)\ . \qquad (4.2.30)$$

For odd order,

$$\det(\underset{\sim}{\Sigma}_{2m+1}) = \det\begin{bmatrix} \underset{\sim}{\Sigma}_m + \underset{\sim}{\Delta}^*\underset{\sim}{K}_m & \sqrt{2}\underset{\sim}{K}_m\underset{\sim}{\beta}_m \\ \sqrt{2}\ \underset{\sim}{\beta}_m'\ \underset{\sim}{K}_m & a_o \end{bmatrix} \det(\underset{\sim}{\Sigma}_m - \underset{\sim}{\Delta}^*\underset{\sim}{K}_m)\ . \qquad (4.2.31)$$

Caflisch (1981) also proved essentially the same result but in a different manner.

(ii) Simple Reduction by Partitioning

For any order m, by direct use of the Schur complement,

$$\det(\underset{\sim}{\Sigma}_m) = \det\begin{bmatrix} \underset{\sim}{\Sigma}_{m-1} & \underset{\sim}{K}_{m-1}\underset{\sim}{\beta}_{m-1} \\ \underset{\sim}{\beta}_{m-1}'\underset{\sim}{K}_{m-1} & a_o \end{bmatrix}$$

$$= \det(\underset{\sim}{\Sigma}_{m-1})\cdot\ (a_o - \underset{\sim}{\beta}_{m-1}'\ \underset{\sim}{\Sigma}_{m-1}^{-1}\ \underset{\sim}{\beta}_{m-1}) \qquad (4.2.32)$$

(provided $\underset{\sim}{\Sigma}_{m-1}^{-1}$ exists).

$$= a_o\ \det[\underset{\sim}{\Sigma}_{m-1} - (\underset{\sim}{\beta}_{m-1}\ \underset{\sim}{\beta}_{m-1}')/a_o]\ . \qquad (4.2.33)$$

(provided $a_o \neq 0$)

(iii) Reduction as a Function of $\det(\underset{\sim}{\Sigma}_{m-1})$ and $\det(\underset{\sim}{\Sigma}_{m-2})$

Using the Sylvester identity for border determinants (cf. Iohvidov, 1982, p. 7), we may express $\det(\underset{\sim}{\Sigma}_m)$ as

$$\det(\underset{\sim}{\Sigma}_m)\det(\underset{\sim}{\Sigma}_{m-2}) = [\det(\underset{\sim}{\Sigma}_{m-1})]^2 - [\det(\underset{\sim}{\Phi}_{m-1})]^2 \qquad (4.2.34)$$

where

$$\underset{\sim}{\Phi}_{m-1} = \left[\begin{array}{c|ccccc} a_1 & a_0 & a_1 & \cdots & a_{m-3} \\ a_2 & a_1 & a_0 & \cdots & a_{m-4} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m-2} & a_{m-3} & a_{m-4} & \cdots & a_0 \\ \hline a_{m-1} & a_{m-2} & a_{m-3} & \cdots & a_1 \end{array}\right],$$

so that

$$\det(\underset{\sim}{\Phi}_{m-1}) = (-1)^{m-2}\det(\underset{\sim}{\Sigma}_{m-2})(a_{m-1} - \underline{\beta}'_{m-2}\underset{\sim}{\Sigma}^{-1}_{m-2}\underset{\sim}{K}_{m-2}\underline{\beta}_{m-2}) \qquad (4.2.35)$$

provided $\underset{\sim}{\Sigma}^{-1}_{m-2}$ exists.

Computationally, the method (i) seems to be the most efficient, as it requires considerably less time and labour than the other methods discussed above. Once $\det(\underset{\sim}{\Sigma}_{m-1})$, $\underset{\sim}{\Sigma}^{-1}_{m-1}$, $\det(\underset{\sim}{\Sigma}_{m-2})$ etc. are known, the methods (ii) and (iii) may be used accordingly. The procedure (iii) has an additional utilization in the "nonsingular extension" of $\underset{\sim}{\Sigma}_{m-1}$ to $\underset{\sim}{\Sigma}_m$. To make this possible, we require only one more element, e.g. a_{m-1}, such that $\underset{\sim}{\Sigma}_m$ will be nonsingular. This will happen when

$$a_{m-1} \neq [\det(\underset{\sim}{\Sigma}_{m-2})]^{-1}[\pm \det(\underset{\sim}{\Sigma}_{m-1}) + \det(\underset{\sim}{\Phi}_{m-1}(0))] \qquad (4.2.36)$$

where $\underset{\sim}{\Phi}_{m-1}(0)$ is the matrix in (4.2.35) with a_{m-1} replaced

by zero. The expression (4.2.36) is simplified consequently to

$$a_{m-1} \neq \pm \frac{\det(\underset{\sim}{\Sigma}_{m-1})}{\det(\underset{\sim}{\Sigma}_{m-2})} + (-1)^{m-1} \underline{\beta}'_{m-2} \underset{\sim}{\Sigma}^{-1}_{m-2} \underset{\sim}{K}_{m-2} \underline{\beta}_{m-2} \cdot \qquad (4.2.37)$$

The notion of nonsingular extension (Iohvidov, 1982) leads to the following obvious proposition.

Proposition 4.2.4 It is possible to write out a sequence of nonsingular $\underset{\sim}{\Sigma}_i$'s, $i = 1,2,3, \ldots$ provided a sequence of elements a_{i-1}, $i = 1,2,3, \ldots$ exists satisfying the inequalities

$$a_0 \neq 0 \,, \quad a_1 \neq \pm a_0 \,,$$

$$a_i \neq \pm \frac{\det(\underset{\sim}{\Sigma}_i)}{\det(\underset{\sim}{\Sigma}_{i-1})} + (-1)^i \underline{\beta}'_{i-1} \underset{\sim}{\Sigma}^{-1}_{i-1} \underset{\sim}{K}_{i-1} \underline{\beta}_{i-1} \,, \qquad (4.2.38)$$

$i = 2,3,4, \ldots$.

(iv) Cholesky Decomposition

The symmetric Toeplitz matrix (4.2.7) can always be written in the form of the product of a lower and an upper triangular matrix through Cholesky decomposition.

$$\underset{\sim}{\Sigma}_m = \underset{\sim}{L}^* \underset{\sim}{L}^{*\prime} = \underset{\sim}{L} \underset{\sim}{D} \underset{\sim}{L}' \qquad (4.2.39)$$

where $\underset{\sim}{D}$ is a diagonal matrix and the principal diagonal of $\underset{\sim}{L}$ consists of elements equal to unity. Then $\det(\underset{\sim}{\Sigma}_m)$ is the

product of all the diagonal elements of $\underset{\sim}{D}$ [cf. Nehorai and Morf, 1985]. The Levinson-Durbin recursive algorithm provides a way of computing the successive rows of the Cholesky factor $\underset{\sim}{L}$ and the diagonal elements of $\underset{\sim}{D}$.

We also note here that for any $\underset{\sim}{\Sigma}_m = a_o \underset{\sim}{\rho}_m$ as defined in (4.2.9), we can always find, following Grunbaum (1981a), a unique tridiagonal matrix B of the same order having a simple spectrum which commutes with $\underset{\sim}{\Sigma}_m$. The determinant and the eigenvalues as well as eigenvectors of the tridiagonal matrix $\underset{\sim}{B}$ are each equal to the corresponding quantities of $\underset{\sim}{\Sigma}_m$. Hence, once the problem of finding this matrix B is solved, the evaluation of the determinant is done routinely and the solution of the eigenproblem corresponding to $\underset{\sim}{\Sigma}_m$ is easily achieved.

4.2.4 Inversion of a Nonsingular Toeplitz Matrix

The inversion of Toeplitz matrices has been approached in the literature [Trench (1964, 1967); Justice (1972); Calderon, Spitzer and Widom (1959); Widom (1960), Gohberg and Feldman (1974); Heinig and Rost (1984); Böttcher and Silbermann (1983)] as a problem of finding an explicit form of the inverse and also for evaluating the computational merit of the algorithm used in the inversion. Levinson (1947) showed that a nonsingular m x m Toeplitz matrix with nonzero leading minors can be inverted with the order of m^2 multiplication operations

as compared to the order of m^3 multiplication generally required for a non-Toeplitz matrix. Gohberg and Semencul (1972) have reviewed and established various methods of finding the inverse of a general Toeplitz matrix. Ben-Artzi and Sholom (1986) proved that three columns of an inverse of a Toeplitz matrix when properly chosen, are always enough to reconstruct it. Kailath , Vieira and Morf (1978) have also reviewed this field. Mentz (1976) developed a procedure to find the components of the inverse matrix in closed form when the m x m matrix has only (2p+1) nonvanishing (central)diagonals ($1 \leq p < m$). The procedure consists in posing difference equations for the components of the inverse and solving them explicitly.

Exploiting the very special structure as possessed by the Toeplitz matrix, many researchers (e.g. Kumar (1984), Zohar (1974), Morf (1980)) devise computationally fast algorithms for solving Toeplitz system of linear equations without calculating Toeplitz-inverse as such.

Trench (1964) gave a recursive formula for a non-singular scalar entried correlation matrix of Toeplitz form (4.2.9) which has since been rederived, elaborated, and extended by several workers, particularly Akaike (1973) and Zohar (1969, 1974). The formula for computing the inverse of the correlation matrix $\underset{\sim}{\rho}_m$ is

$$[\rho_m^{-1}]_{i+1,j+1} = [\rho_m^{-1}]_{i,j} + \frac{1}{K_o}\{b_{i,m}c_{j,m} - c_{m-i+1,m}b_{m-j+1,m}\}$$

where the coefficients K_o , $\{b_{i,m}\}$, $\{c_{i,m}\}$ are found from certain simple equations, such as the Yule-Walker equations (Pagano (1972)), which can be efficiently solved by the Levinson-Durbin algorithm (Levinson (1947); Durbin (1966)). Wise (1955) has given a method of finding $\underset{\sim}{\rho}_m^{-1}$ of (4.2.9) based on the spectral density function. Exploiting the symmetry, Siddiqui (1958) proposed an alternative procedure of computing the inverse of a Toeplitz covariance matrix with the sequence of elements $\{a_i , i = 0,1, \ldots , (m-1)\}$ as the auto-correlation sequence of a stationary autoregressive process.

We shall mention here two procedures (not to be treated as original) which might be helpful due to their being less complicated. The first one is based on expressing the inverse as the difference between two matrices. The second one involves a procedure of inversion by splitting the matrix into two halves. Each procedure is described below.

(i) <u>Inverse as the difference of two matrices</u>

Analogous to (4.2.10), the inverse matrix may be expressed as

$$\underset{\sim}{\Sigma}_m^{-1} = \alpha_o^{-1} \left(\underset{\sim}{B}\ \underset{\sim}{B}' - \underset{\sim}{C}\underset{\sim}{C}' \right) \qquad (4.2.40)$$

where $\underset{\sim}{B}_{mxm} =$

$$\begin{bmatrix} \alpha_o & & & \\ \alpha_1 & \alpha_o & & \\ \alpha_2 & \alpha_1 & \alpha_o & \\ \vdots & & & \\ \alpha_{m-1} & \alpha_{m-2} & \alpha_{m-3} \cdots & \alpha_o \end{bmatrix},$$

$$\underset{\sim}{C}_{mxm} = \begin{bmatrix} 0 & & & \\ \alpha_{m-1} & 0 & & \\ \alpha_{m-2} & \alpha_{m-1} & 0 & \\ \vdots & & & \\ \alpha_1 & \alpha_2 & \alpha_3 \cdots & \alpha_{m-1} \quad 0 \end{bmatrix},$$

$$\alpha_o = \det(\underset{\sim}{\Sigma}_{m-1})/\det(\underset{\sim}{\Sigma}_m) ,$$

$$\underset{\sim}{\delta}_{m-1} = (\alpha_1, \alpha_2, \cdots, \alpha_{m-1})' = -\alpha_o \underset{\sim}{\Sigma}_{m-1}^{-1} \underset{\sim}{\beta}_{m-1}$$

and $\underset{\sim}{\beta}_{m-1} = (a_1, a_2, \cdots, a_{m-2}, a_{m-1})'$,

provided $\det(\underset{\sim}{\Sigma}_m) \neq 0$ and $\det(\underset{\sim}{\Sigma}_{m-1}) \neq 0$. Essentially the same result has been reported by Kailath , Buckstein and Morgan (1986). Kailath, Vieira and Morf (1978) credited

this formula to Gohberg and Semencul (1972) and used to compute bilinear forms such as $\underset{\sim}{y}' \underset{\sim}{R}^{-1} \underset{\sim}{x}$ where $\underline{y}$ and $\underline{x}$ are given vectors. Some other variant expressions are available in the book of Iohvidov (1982, Sec. 18).

Proposition 4.2.5 The inverse of a nonsingular Toeplitz matrix ($\underset{\sim}{\Sigma}_m$) is a symmetric centrosymmetric matrix (SC).

Proof : Owing to centrosymmetric property of $\underset{\sim}{\Sigma}_m$ (vide Proposition 4.2.2), property (b) of symmetric centrosymmetric matrix (vide subsection 4.2.1(iii)) will complete the proof. ■

In particular, $\underset{\sim}{\Sigma}_m$ may have a Toeplitz inverse iff $\underline{\beta}_{m-1} = \underset{\sim}{K}_{m-1} \underline{\beta}_{m-1}$ (Graville, 1983) implying a circulant Toeplitz matrix.

(ii) Inversion by splitting into halves

For even order $\underset{\sim}{\Sigma}_{2m}$ (vide (4.2.24)) and orthogonal matrix $\underset{\sim}{P}_{2m}$ (vide (4.2.5)),

$$\underset{\sim}{P}_{2m} \underset{\sim}{\Sigma}_{2m} \underset{\sim}{P}_{2m}' = \text{diag.}(\underset{\sim}{\Sigma}_m + \underset{\sim}{\Delta}_m \underset{\sim}{K}_m \ , \ \underset{\sim}{\Sigma}_m - \underset{\sim}{\Delta}_m \underset{\sim}{K}_m) \qquad (4.2.41)$$

implying

$$\underset{\sim}{\Sigma}_{2m}^{-1} = \underset{\sim}{P}_{2m}' \cdot \text{diag.}[(\underset{\sim}{\Sigma}_m + \underset{\sim}{\Delta}_m \underset{\sim}{K}_m)^{-1}, (\underset{\sim}{\Sigma}_m - \underset{\sim}{\Delta}_m \underset{\sim}{K}_m)^{-1}] \cdot \underset{\sim}{P}_{2m} \qquad (4.2.42)$$

Writing $\underset{\sim}{\Sigma}_{2m}^{-1}$ (which is of SC nature) in partitioned form

$$\underset{\sim}{\Sigma}_{2m}^{-1} = \begin{bmatrix} \underset{\sim}{B}_{11} & \underset{\sim}{B}_{12} \\ \underset{\sim}{K}_m \underset{\sim}{B}_{12} \underset{\sim}{K}_m & \underset{\sim}{K}_m \underset{\sim}{B}_{11} \underset{\sim}{K}_m \end{bmatrix} \qquad (4.2.43)$$

where $\underset{\sim}{B}_{11}$ and $\underset{\sim}{B}_{12}$ may be obtained directly from (4.2.42) as

$$\underset{\sim}{B}_{11} = \frac{1}{2}\left[(\underset{\sim}{\Sigma}_m + \underset{\sim}{\Delta}_m\underset{\sim}{K}_m)^{-1} + (\underset{\sim}{\Sigma}_m - \underset{\sim}{\Delta}_m\underset{\sim}{K}_m)^{-1}\right] \qquad (4.2.44)$$

$$\underset{\sim}{B}_{12} = \left[(\underset{\sim}{\Sigma}_m + \underset{\sim}{\Delta}_m\underset{\sim}{K}_m)^{-1} - (\underset{\sim}{\Sigma}_m - \underset{\sim}{\Delta}_m\underset{\sim}{K}_m)^{-1}\right]\underset{\sim}{K}_m \cdot \qquad (4.2.45)$$

Similarly, for odd order $\underset{\sim}{\Sigma}_{2m+1}$ (vide (4.2.26)) we use $\underset{\sim}{P}_{2m+1}$ (vide (4.2.6)) and get

$$\underset{\sim}{\Sigma}_{2m+1}^{-1} = \begin{bmatrix} \underset{\sim}{B}_{11} & \underline{\epsilon} & \underset{\sim}{B}_{12} \\ \underline{\epsilon}' & b_{m+1,m+1} & \underline{\epsilon}'\underset{\sim}{K}_m \\ \underset{\sim}{K}_m\underset{\sim}{B}_{12}\underset{\sim}{K}_m & \underset{\sim}{K}_m\underline{\epsilon} & \underset{\sim}{K}_m\underset{\sim}{B}_{11}\underset{\sim}{K}_m \end{bmatrix} \qquad (4.2.46)$$

in which $\underset{\sim}{B}_{11}$, $\underset{\sim}{B}_{12}$, $\underline{\epsilon}$ and $b_{m+1,\,m+1}$ are to be found from

$$\begin{bmatrix} \underset{\sim}{B}_{11} + \underset{\sim}{B}_{12}\underset{\sim}{K}_m & \sqrt{2}\,\underline{\epsilon} \\ \sqrt{2}\,\underline{\epsilon}' & b_{m+1,m+1} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{\Sigma}_m + \underset{\sim}{\Delta}_m^*\underset{\sim}{K}_m & \sqrt{2}\,\underset{\sim}{K}_m\underline{\beta}_m \\ \sqrt{2}\,\underline{\beta}_m'\underset{\sim}{K}_m & a_o \end{bmatrix}^{-1} \qquad (4.2.47)$$

and
$$\underset{\sim}{B}_{11} - \underset{\sim}{B}_{12}\underset{\sim}{K}_m = (\underset{\sim}{\Sigma}_m - \underset{\sim}{\Delta}_m^*\,\underset{\sim}{K}_m)^{-1} \cdot \qquad (4.2.48)$$

4.2.5 Bound on the Determinant of a P.D. Toeplitz Matrix

We shall give here, in the form of lemmas, various results related to bounds on $\det(\underset{\sim}{\Sigma}_m)$ when $\underset{\sim}{\Sigma}_m$ is a positive definite (p.d.) symmetric Toeplitz matrix of order $m \times m$.

Lemma 4.2.1 The sequence $\left\{\det(\underset{\sim}{\Sigma}_m)/\det(\underset{\sim}{\Sigma}_{m-1}),\ i = 2,3, \ldots\right\}$ is a decreasing sequence of positive numbers less than $a_o(>0)$.

Proof : From the Sylvester identity (4.2.34) we have

$$\det(\underset{\sim}{\Sigma}_m)\ \det(\underset{\sim}{\Sigma}_{m-2}) \leq [\det(\underset{\sim}{\Sigma}_{m-1})]^2 \ . \qquad (4.2.49)$$

As $\underset{\sim}{\Sigma}_m$ is p.d., $\underset{\sim}{\Sigma}_{m-1}$, $\underset{\sim}{\Sigma}_{m-2}$ are also p.d., so that

$$0 < \frac{\det(\underset{\sim}{\Sigma}_m)}{\det(\underset{\sim}{\Sigma}_{m-1})} \leq \frac{\det(\underset{\sim}{\Sigma}_{m-1})}{\det(\underset{\sim}{\Sigma}_{m-2})} \ . \qquad \blacksquare$$

Very recently, Cover and Thomas (1988) have proved the following lemma:

Lemma 4.2.2 $\left\{[\det(\underset{\sim}{\Sigma}_i)]^{1/i}\right\}$ is a decreasing sequence in i .

As regards the uppermost bound of $\underset{\sim}{\Sigma}_m$, we establish the following inequality:

Lemma 4.2.3 $\det(\underset{\sim}{\Sigma}_m) < (1-a_1^2/a_o^2)^{m-1}\ a_o^m$. (4.2.50)

Proof : From Lemma 4.2.1,

$$\frac{\det(\underset{\sim}{\Sigma}_m)}{\det(\underset{\sim}{\Sigma}_1)} = \prod_{i=2}^{m} \frac{\det(\underset{\sim}{\Sigma}_i)}{\det(\underset{\sim}{\Sigma}_{i-1})} \leq \left(\frac{\det(\underset{\sim}{\Sigma}_2)}{\det(\underset{\sim}{\Sigma}_1)}\right)^{m-1} = \left(\frac{a_o - a_1^2}{a_o}\right)^{m-1} . \qquad \blacksquare$$

Lemma 4.2.4 $\det(\underset{\sim}{\Sigma}_{2m}) \leq \prod_{j=o}^{m-1} (a_o^2 - a_{2m-2j-1}^2)$ (4.2.51)

and $\det(\underset{\sim}{\Sigma}_{2m+1}) \leq a_o \prod_{j-o}^{m-1} (a_o^2 - a_{2m-2j}^2)$. (4.2.52)

Proof : To prove these results, it is sufficient to note

the well-known Hadamard inequality [cf. Bellman (1974, p. 129)] :

If $\underset{\sim}{W}(mxm)$ is a p.d. matrix, $\det(\underset{\sim}{W}) \leq \prod_{i=1}^{m} W_{ii}$. The equality holds when $\underset{\sim}{W}$ is a diagonal matrix.

This inequality may be applied to p.d. matrices $\underset{\sim}{\Sigma}_m \pm \underset{\sim}{\Delta}_m \underset{\sim}{K}_m$ [vide eqn. (4.2.30)] to establish (4.2.51). Similarly, (4.2.52) may be established from (4.2.31).

Lemma 4.2.5 $$\sup_{p.d.\Sigma_m} \det(\underset{\sim}{\Sigma}_m) = a_o^m \left(1 - \frac{a_{m-1}^2}{a_o^2}\right). \quad (4.2.53)$$

Proof : Applying the condition of equality (from the above mentioned Hadamard inequality) to $\underset{\sim}{\Sigma}_m \pm \underset{\sim}{\Delta}_m \underset{\sim}{K}_m$, we get the off-diagonal elements zero, so that

$$a_1 \pm a_{2m-2} = a_2 \pm a_{2m-3} = \cdots = a_{m-1} \pm a_m = 0 .$$

Therefore, $a_1 = a_2 = \cdots = a_{2m-2} = 0$. Consequently, the uppermost bound on $\det(\underset{\sim}{\Sigma}_{2m})$ in (4.2.51) is

$$(a_o^2 - a_{2m-1}^2)\, a_o^{2(m-1)} . \quad (4.2.54)$$

Similarly, for odd order $\underset{\sim}{\Sigma}_{2m+1}$, the upper most bound on the determinant is $(a_o^2 - a_{2m}^2)\, a_o^{2m-1}$. ∎

Lemma 4.2.6 $\det(\underset{\sim}{\Sigma}_m) < \det(\underset{\sim}{\bar{\Sigma}}_m)$ (4.2.55)

where $\underset{\sim}{\bar{\Sigma}}_m$ is an mxm matrix with a_o as diagonal elements and $[(m-1)a_1 + (m-2)a_2 + \cdots + a_{m-1}]/[m(m-1)/2]$ (= b, say)

as off-diagonal elements, so that

$$\det(\overline{\Sigma}_m) = (a_o-b)^m[a_o+(m-1)b]. \qquad (4.2.56)$$

Proof : We refer to Aitkin et al. (1968) for a proof. ■

Lemma 4.2.7 If $\underset{\sim}{\rho}_m$ is the m x m symmetric correlation matrix of Toeplitz form as defined in (4.2.10) and $\det(\underset{\sim}{\rho}^*_m)$ is the determinant of the cofactor of the element in the first column and m-th row of $\underset{\sim}{\rho}_m$, then the sequence $[\det(\rho^*_i)]^2/[\det(\rho_{i-1})]^2$, i = 2,3, ... is an increasing sequence, and the index

$$\pi_{i-1} = \frac{\det(\rho^*_i)}{\det(\rho_{i-1})} \qquad (4.2.57)$$

lies between -1 and +1 .

Proof : This is a clear application of the Sylvester identity to $\underset{\sim}{\rho}_m$ as shown in (4.2.34) and lemma 4.2.1. ■

4.2.6 Bounds on Eigenvalues

The problem of obtaining an explicit solution of the eigenvalues of general Toeplitz matrices has been considered by many researchers. Whittle (1951) obtained an approximate expressions of eigenvalues of $\underset{\sim}{\Sigma}_m$. Grunbaum (1981a, 1981b), Trench (1985, 1986, 1989), Day (1975), Dickinson (1980, 1984) have extensive contribution to the problem of determining eigenvalues of Toeplitz and Band Toeplitz matrices generated

particularly by a formal Laurent series of a rational functions. Bini and Pan (1988) developed efficient algorithm of eigenvalues of Block Banded Toeplitz matrix.

Asymptotic distribution of eigenvalues of the m × m Toeplitz matrices as m $\longrightarrow \infty$ has been studied by many authors (e.g. Widom (1958), Dickinson (1980)) in the case where the Toeplitz elements are the Fourier coefficients of some suitable integrable function. Several authors [e.g. Makhoul (1981), Wilks and Hayes (1987), Cybenko (1984)] have dealt with the spectral structure of Hermitian Toeplitz matrices. Numerical methods of finding the smallest eigenvalues of a p.d. Hermitian Toeplitz matrix have been suggested by Cybenko and Van Loan (1984). However, the development of efficient methods designed specially to compute the eigenvalues and eigenvectors of Toeplitz matrices still requires more investigation.

We shall present here some simple bounds on the eigenvalues ($\lambda_i > 0$, i = 1, ... , m) of a p.d. $\underset{\sim}{\Sigma}_m$. Since these results can be derived without much difficulty, we only refer to original sources. We also refer to famous Parron-Frobenius theorem (cf. Gantmacher, 1959, p. 65), which guarantees that any nonnegative matrix has an eigenvalue of maximum magnitude that is real and positive. We also keep in view Gerschgorin's theorem (cf. Pullman, 1976, p. 213) that for $\underset{\sim}{A}(m \times m) = ((a_{jk}))$ every eigenvalue lies in the union of the m closed intervals

$$\left[a_{jj} - \sum_{j \neq k} |a_{jk}| , \quad a_{jj} + \sum_{j \neq k} |a_{jk}| \right], \qquad (4.2.58)$$

$j = 1, \ldots, m.$

Hoffman (1965) also proved a theorem that every eigenvalue of $\underset{\sim}{A}$ lies in the union of m closed intervals

$$\left[\sum_{k} a_{jk} - m(\max_{j \neq k} a_{jk})_{+} , \sum_{k} a_{jk} - m(\min_{j \neq k} a_{jk})_{-} \right], \qquad (4.2.59)$$

$j = 1, \ldots, m$, where $a_{+} = |a|$ if $a > 0$ and $= 0$ otherwise, similarly, $a_{-} = -|a|$ if $a < 0$ and $= 0$ otherwise. These results (e.g. (4.2.58) and (4.2.59) can be simplified for the Toeplitz matrix $\underset{\sim}{\Sigma}_m$. In addition, we present below some results in the form of propositions.

<u>Proposition 4.2.6</u> The inequality $0 < \lambda_{min} < a_o + \frac{2}{m} \sum_{i=1}^{m-1} (m-i) a_i \leq \lambda_{max}$ holds true.

<u>Proof</u> : We refer to Pullman (1976, p. 232) or Graybill (1969, p. 312) for the proof. ∎

Using a result of Bellman (1974, p. 41), we have simplified inequalities

$$0 < \lambda_{min} \leq a_o \leq \lambda_{max} \leq m a_o \qquad (4.2.60)$$

and

$$m\, b_m \leq \lambda_{max} \leq m\, a_o \qquad (4.2.61)$$

where $b_m = \underset{0 \le i \le m-1}{\text{Min}} a_i$. These are particularly helpful to quantify λ_{max} or λ_{min} at a glance from the elements of $\underset{\sim}{\Sigma}_m$. Further refinements on the bounds are obtained in the following proposition:

Proposition 4.2.7 $\quad \text{Max}(L_1 , L_2) \le \lambda_{max} \le U$

where $L_1 = a_o + \frac{2}{m} \sum_{i=1}^{m-1} (m-i)a_i$, $L_2 = \underset{1 \le i \le [(m+1)/2]}{\max} \left(\sum_{j=1}^{m} a^2_{|i-j|} \right)^{1/2}$

and $U = \underset{1 \le i \le [(m+1)/2]}{\text{Max}} \left(\sum_{j=1}^{m} \left| a_{|i-j|} \right| \right)$.

Proof : Using (4.2.57)

$$\lambda_{max} \le \underset{1 \le i \le m}{\text{Max}} \sum_{j=1}^{m} \left| a_{|i-j|} \right| \qquad (4.2.62)$$

$$\text{Since} \quad \sum_{j=1}^{m} \left| a_{|i-j|} \right| = \sum_{j=1}^{m} \left| a_{|(m-i+1)-(m-j+1)|} \right|$$

$$= \sum_{j=1}^{m} \left| a_{|(m-i+1)-j|} \right| , \qquad (4.2.63)$$

maximization in the r.h.s. of (4.2.62) is equivalent to maximization over $1 \le i \le [\ \frac{m+1}{2}\ ]$, so that $\lambda_{max} \le U$.

To establish that $\lambda_{max} \ge L_1$ and $\lambda_{max} \ge L_2$ we should take help of two results from Graybill (1969, p. 312) stated as follows.

(a) For any symmetric matrix $\underset{\sim}{A}(m \times m)$,

$$\lambda_{max} \geq \frac{1}{m} \sum_{i}^{m} \sum_{j}^{m} a_{ij} \qquad (4.2.64)$$

(b) For any p.d. or p.s.d. matrix $\underset{\sim}{A}(m \times m)$

$$\lambda_{max}^2 \geq \sum_{j=1}^{m} a_{ij}^2 \qquad (4.2.65)$$

for all $i = 1,2, \ldots , m$.

Remark 4.2 (i) L_1 may be found reasonably smaller than L_2 when some of the off-diagonal elements of $\underset{\sim}{\Sigma}_m$ are negative. On the other hand, L_2 may be found smaller than L_1 when all the elements of $\underset{\sim}{\Sigma}_m$ are positive. In this case, the interval (L_1, U) is very close to contain λ_{max}.

(ii) The limits $\text{Max}(L_1, L_2)$ and U are not claimed respectively as the greatest lower bound and least upper bound of λ_{max}, but are helpful in assessing the magnitude of λ_{max}. The following proposition will be similarly helpful for assessing λ_{min}.

Proposition 4.2.8 The inequality of $0 < \lambda_{min} \leq a_o - \overset{*}{\text{Max}} |a_k|$ holds true whenever $\overset{*}{\text{Max}}$ denotes the maximization taken over $k = 1, 3, 5, \ldots , (2m-1)$ for even order $\underset{\sim}{\Sigma}_{2m}$ and $k = 2,4,6, \ldots , 2m$ for odd order $\underset{\sim}{\Sigma}_{2m+1}$.

Proof : As eigenvalues are similarity invariant, let us

consider the eigenvalues of $\underset{\sim}{P}_{2m}\underset{\sim}{\Sigma}_{2m}\underset{\sim}{P}'_{2m}$ [vide eqn. (4.2.41)] (or $\underset{\sim}{P}_{2m+1}\underset{\sim}{\Sigma}_{2m+1}\underset{\sim}{P}'_{2m+1}$) and apply the theorem that minimum eigenvalue is less than the minimum diagonal element, to complete the proof. ■

<u>Proposition 4.2.9</u> (a) $\bar{\lambda} = \frac{1}{m}\sum_{i=1}^{m}\lambda_i = a_o$, $\prod_{i=1}^{m}\lambda_i \leq a_o^m$.

(b) Defining spread of the matrix (cf. Bellman, 1974, p.223)

$\underset{\sim}{\Sigma}_m$ as $R_\lambda = \lambda_{max} - \lambda_{min}$, $\frac{\overset{*}{\text{Max}}|a_k|}{a_o} < \frac{R_\lambda}{\bar{\lambda}} < m$.

(c) $a_o \in (\lambda_{min} , \lambda_{max})$ and $a_j \in (-jR_\lambda , jR_\lambda)$, $j = 1,2, \ldots,(m-1)$.

<u>Proof</u> : The proof of (a) is elementary. That of (b) may be done by using (4.2.66) and (4.2.60). By using a result of Graybill (1969, p. 314), (c) may be proved. ■

Following are some additional inequalities related to the monotone behaviour of the eigenvalues of $\underset{\sim}{\Sigma}_m$ of different orders. These are mainly based on the Sturmian separation theorem (cf. Bellman, 1974, p. 117) stated as follows :

Let $\lambda_k(\underset{\sim}{A}_r)$, $k = 1, \ldots , r$; $r = 1,2, \ldots$ be the eigenvalues of symmetric matrices $\underset{\sim}{A}_r(r \times r)$, and $\lambda_1(\underset{\sim}{A}_r) \geq \lambda_2(\underset{\sim}{A}_r) \geq \ldots \geq \lambda_r(\underset{\sim}{A}_r)$. Then $\lambda_{k+1}(\underset{\sim}{A}_{r+1}) \leq \lambda_k(\underset{\sim}{A}_r) \leq \lambda_k(\underset{\sim}{A}_{r+1})$ for any $k \leq r$.

Above mentioned inequality naturally holds for $\underset{\sim}{\Sigma}_m$. Furthermore, we get other inequalities as an immediate consequence.

$$a_o = \lambda_{max}(\underset{\sim}{\Sigma}_1) \leq \lambda_{max}(\underset{\sim}{\Sigma}_2) \leq \cdots \leq \lambda_{max}(\underset{\sim}{\Sigma}_m) \leq ma_o \ , \qquad (4.2.66)$$

$$a_o = \lambda_{min}(\underset{\sim}{\Sigma}_1) \geq \lambda_{min}(\underset{\sim}{\Sigma}_2) \geq \cdots \geq \lambda_{min}(\underset{\sim}{\Sigma}_m) \geq 0 \ , \qquad (4.2.67)$$

$$0 = R_\lambda(\underset{\sim}{\Sigma}_1) \leq R_\lambda(\underset{\sim}{\Sigma}_2) \leq \cdots \leq R_\lambda(\underset{\sim}{\Sigma}_m) \leq ma_o \ , \qquad (4.2.68)$$

where $R_\lambda(\underset{\sim}{\Sigma}_i) = \lambda_{max}(\underset{\sim}{\Sigma}_i) - \lambda_{min}(\underset{\sim}{\Sigma}_i)$. Using Proposition 4.2.6 and (4.2.67), we establish further that

$$0 < \lambda_{max}(\underset{\sim}{\Sigma}_m) - \lambda_{max}(\underset{\sim}{\Sigma}_{m-1}) \leq (m-1) R_a(\underset{\sim}{\Sigma}_{m-1})$$

where $R_a(\underset{\sim}{\Sigma}_{m-1}) = Max(a_o, a_1, \cdots, a_{m-2}) - Min(a_o, a_1, \cdots, a_{m-2})$. ■

4.3 STATISTICAL MEASURES OF ASSOCIATION AMONG TOEPLITZ VARIABLES

Let a random vector $\underline{X}(p \times 1)$ follow $N_p(\underline{\mu}, \underset{\sim}{\Sigma})$ where $\underline{\mu}$ is an arbitrary vector and $\underset{\sim}{\Sigma}$ is a p.d. Toeplitz matrix having either of the equivalent expressions (4.2.7) or (4.2.9). It may be sometimes found that all the diagonal elements of $\underset{\sim}{\Sigma}$ are not equal but its off-diagonal elements are such that corresponding correlation matrix has a Toeplitz pattern. To distinguish, we shall call the latter situation a Toeplitz correlation structure of $\underset{\sim}{\Sigma}$ while the former is termed, as usual, a Toeplitz covariance structure of $\underset{\sim}{\Sigma}$. Correspondingly, the variables are said to be Toeplitz variables if these generate a Toeplitz covariance matrix. In the

event of generating a Toeplitz correlation matrix only, the variables are called scale-free or standardized Toeplitz variables.

It may be noted that since the correlation between any pair of variables is invariant under change of location and scale, the statistical measures based on correlation coefficients will be identical for both the Toeplitz covariance and Toeplitz correlation structures. In the discussions to follow, we shall make use of Toeplitz covariance structure of $\underset{\sim}{\Sigma}$ unless otherwise mentioned. [For notational simplification, here we avoid the notation $\underset{\sim}{\Sigma}_p$ of subsection 4.2.]

4.3.1 Equality on Conditional Dispersion Matrices

Let the random vector $\underline{X}$ be partitioned into two sub-vectors in two ways as follows.

$$\underline{X} = \begin{bmatrix} \underset{r\times 1}{\underline{X}_1} \\ \\ \underset{(p-r)\times 1}{\underline{X}_2} \end{bmatrix} = \begin{bmatrix} \underset{(p-r)\times 1}{\underline{X}_1^*} \\ \\ \underset{r\times 1}{\underline{X}_2^*} \end{bmatrix} \qquad (4.3.1)$$

Under the multinormal set-up, we may easily derive the marginal and conditional distributions of the subvectors. As the association (relationship) of the random vectors usually requires the study of the conditional distributions, we find it less important to consider the marginal distributions and pay attention only to conditional distributions.

In this connection, we hereunder formulate five propositions.

Proposition 4.3.1 (i) The conditional dispersion matrix of $\underset{\sim}{X}_1$ given $\underset{\sim}{X}_2 = \underset{\sim}{x}_2$ is the same as that of $\underset{\sim}{K}_r \underset{\sim}{X}_2^*$ given $\underset{\sim}{X}_1^* = \underset{\sim}{x}_1^*$.

(ii) The conditional dispersion matrix of $\underset{\sim}{X}_2$ given $\underset{\sim}{X}_1 = \underset{\sim}{x}_1$ is the same as that of $\underset{\sim}{K}_{p-r} \underset{\sim}{X}_1^*$ given $\underset{\sim}{X}_2^* = \underset{\sim}{x}_2^*$.

Proof : Due to multinormality, conditional dispersion matrix of $\underset{\sim}{X}_1$ given $\underset{\sim}{X}_2 = \underset{\sim}{x}_2$ is independent of $\underset{\sim}{x}_2$ and is given by

$$\underset{\sim}{\Sigma}_{11.2} = \underset{\sim}{\Sigma}_{11} - \underset{\sim}{\Sigma}_{12}\underset{\sim}{\Sigma}_{22}^{-1}\underset{\sim}{\Sigma}_{21} \qquad (4.3.2)$$

and similarly, that of $\underset{\sim}{K}_r \underset{\sim}{X}_2^*$ given $\underset{\sim}{X}_1 = \underset{\sim}{x}_1^*$ is given by

$$\underset{\sim}{\Sigma}^*_{22.1} = \underset{\sim}{K}_r(\underset{\sim}{\Sigma}^*_{22} - \underset{\sim}{\Sigma}^*_{21}\underset{\sim}{\Sigma}_{11}^{-1}\underset{\sim}{\Sigma}^*_{12})\underset{\sim}{K}_r \qquad (4.3.3)$$

with $\underset{\sim}{\Sigma}$ partitioned as

$$\underset{r \times r}{\underset{\sim}{\Sigma}} = \begin{bmatrix} \underset{rxr}{\underset{\sim}{\Sigma}_{11}} & \underset{\sim}{\Sigma}_{12} \\ \underset{\sim}{\Sigma}_{21} & \underset{\sim}{\Sigma}_{22} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{\Sigma}^*_{11} & \underset{\sim}{\Sigma}^*_{12} \\ \underset{\sim}{\Sigma}^*_{21} & \underset{rxr}{\underset{\sim}{\Sigma}^*_{22}} \end{bmatrix}. \qquad (4.3.4)$$

Due to Toeplitz property of $\underset{\sim}{\Sigma}$, $\underset{\sim}{\Sigma}^*_{11} = \underset{\sim}{\Sigma}_{22}$, $\underset{\sim}{\Sigma}^*_{22} = \underset{\sim}{\Sigma}_{11}$, $\underset{\sim}{\Sigma}^*_{12} = \underset{\sim}{\Sigma}^{*\prime}_{21} = \underset{\sim}{K}_{p-r}\underset{\sim}{\Sigma}_{21}\underset{\sim}{K}_r$. Since $\underset{\sim}{K}_p^2 = \underset{\sim}{I}_p$, we deduce from (4.3.3) that

$$\underset{\sim}{\Sigma}_{22.1} = \underset{\sim}{K}_r(\underset{\sim}{\Sigma}_{11} - \underset{\sim}{K}_r\underset{\sim}{\Sigma}_{12}\underset{\sim}{K}_{p-r}\,\underset{\sim}{\Sigma}_{22}^{-1}\,\underset{\sim}{K}_{p-r}\underset{\sim}{\Sigma}_{21}\underset{\sim}{K}_r)\underset{\sim}{K}_r$$

$$= \underset{\sim}{K}_r(\underset{\sim}{\Sigma}_{11} - \underset{\sim}{K}_r\underset{\sim}{\Sigma}_{12}\underset{\sim}{\Sigma}_{22}^{-1}\,\underset{\sim}{\Sigma}_{21}\underset{\sim}{K}_r)\underset{\sim}{K}_r$$

$$= \underset{\sim}{\Sigma}_{11} - \underset{\sim}{\Sigma}_{12}\underset{\sim}{\Sigma}_{22}^{-1}\underset{\sim}{\Sigma}_{21} = \underset{\sim}{\Sigma}_{11.2} \qquad (4.3.5)$$

completing the proof of (i). Similarly, (ii) may be proved. ∎

The Proposition 4.3.1 would imply some equalities in the measures of association among the two sets of Toeplitz variables, e.g. Multiple correlation ($\rho_{(j)}$) is a measure of association of one variable (X_j) over the other (p-1) variables and is defined by

$$\rho^2_{(j)} = 1 - \frac{\text{Conditional variance of } X_j \text{ given the values of other (p-1) variables}}{\text{(Marginal) variance of } X_j}. \qquad (4.3.6)$$

Since $\underset{\sim}{\Sigma}_{11} = \underset{\sim}{\Sigma}^*_{22}$ (as available from the Toeplitz matrix $\underset{\sim}{\Sigma}$), the marginal dispersion matrix of $\underline{X}_1$ equals that of $\underline{X}^*_2$. For r = 1, this equality and the Proposition 4.3.1 will lead to the following equality:

<u>Proposition 4.3.2</u> $\rho^2_{(j)} = \rho^2_{(p-j+1)}$, $j = 1(1)[p/2]$. (4.3.7)

<u>Remark 4.3</u> The p-component vector formed by $\rho^2_{(j)}$, $j = 1(1)p$ would be a symmetric vector (vide subsection 4.2.1(ii)).

Remark 4.4 It is known that j-th diagonal element of the inverse of the correlation matrix is called the variance inflation factor (VIF) for the j-th variable X_j [Chatterjee and Price, 1977]. In terms of multiple correlation $\rho_{(j)}$, $VIF_{(j)} = 1/(1 - \rho^2_{(j)})$. For the Toeplitz situation, due to Proposition 4.3.2,

$$VIF_{(j)} = VIF_{(p-j+1)}\ ,\ j = 1(1)[p/2]. \qquad (4.3.8)$$

The values of $VIF_{(j)}$ are used to detect the collinearity in the intercorrelation matrix.

As regards partial correlation matrices (typically, $\rho_{\underline{X}_1 \cdot \underline{X}_2}$ of $\underline{X}_1$ given $\underline{X}_2 = \underline{x}_2$) we get the same sort of equality via Proposition 4.3.1 as follows :

Proposition 4.3.3 (i) $$\underset{\sim}{\rho}_{\underline{X}_1 \cdot \underline{X}_2} = \underset{\sim}{\rho}_{\underline{X}_2^{**} \cdot \underline{X}_1^{*}} \qquad (4.3.9)$$

(ii) $$\underset{\sim}{\rho}_{\underline{X}_2 \cdot \underline{X}_1} = \underset{\sim}{\rho}_{\underline{X}_1^{**} \cdot \underline{X}_2^{*}} \qquad (4.3.10)$$

where $\underline{X}_2^{**} = \underset{\sim}{K}_r \underline{X}_2^{*}$, $\underline{X}_1^{**} = \underset{\sim}{K}_{p-r}\, \underline{X}_1^{*}$, and the order of $\underset{\sim}{\rho}$ in (i) and (ii) are r x r and $\overline{p-r}$ x $\overline{p-r}$ respectively.

In particular, when r = 2, the partial correlation coefficients between X_i and X_j ($i \neq j$) when other (p-2) variables are held fixed, is given by

$$\rho_{(ij)} = -\sigma^{ij}/(\sigma^{ii}\sigma^{jj})^{1/2} \qquad (4.3.11)$$

where σ^{ij} denotes the (i,j) element of $\underset{\sim}{\Sigma}^{-1}$, $i \neq j = 1(1)p$. Symbolically, $\rho_{(ii)} = 1$. Now, since Toeplitz $\underset{\sim}{\Sigma}$ has a centro-symmetric inverse (vide Proposition 4.2.5), $\sigma^{ij} = \sigma^{p-i+1,p-j+1}$ so that

$$\rho_{(ij)} = \rho_{(p-i+1,p-j+1)} \qquad (4.3.12)$$

$\forall$ i,j = 1(1)p. Thus, the matrix of partial correlations formed by these $\rho_{(ij)}$-elements is also centrosymmetric.

Let us consider the canonical correlation between a pair of random vectors, and the corresponding pair of canonical variables. For the random vectors $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$, the squared canonical correlations between $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ are the eigenvalues of $\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. Specifically, for the j-th canonical correlation (say, φ_j), we should have a standardised canonical pair $(\underset{\sim}{\alpha}_j' \underset{\sim}{X}_1 , \underset{\sim}{\beta}_j' \underset{\sim}{X}_2)$ where canonical coefficient vectors $\underset{\sim}{\alpha}_j$ and $\underset{\sim}{\beta}_j$ should be obtained from

$$(\underset{\sim}{\Sigma}_{12}\underset{\sim}{\Sigma}_{22}^{-1}\underset{\sim}{\Sigma}_{21} - \varphi_j^2\underset{\sim}{\Sigma}_{11})\underset{\sim}{\alpha}_j = 0 \quad \text{with} \quad \underset{\sim}{\alpha}_j'\underset{\sim}{\Sigma}_{11}\underset{\sim}{\alpha}_j = 1 \ , \qquad (4.3.13)$$

and

$$(\underset{\sim}{\Sigma}_{21}\underset{\sim}{\Sigma}_{11}^{-1}\underset{\sim}{\Sigma}_{12} - \varphi_j^2\underset{\sim}{\Sigma}_{22})\underset{\sim}{\beta}_j = 0 \quad \text{with} \quad \underset{\sim}{\beta}_j'\underset{\sim}{\Sigma}_{22}\underset{\sim}{\beta}_j = 1 \qquad (4.3.14)$$

$j = 1,2, \ldots , q$ where $q = \min.(r, p-r)$.

Let us similarly define φ_j^* , $\underset{\sim}{a}_j^{*\prime} \underset{\sim}{X}_1^*$, $\underset{\sim}{\beta}_j^{*\prime} \underset{\sim}{X}_2^*$ for the random vectors $\underset{\sim}{X}_1^*$ and $\underset{\sim}{X}_2^*$. Then using the relations $\underset{\sim}{\Sigma}_{11}^* = \underset{\sim}{\Sigma}_{22}$, $\underset{\sim}{\Sigma}_{22}^* = \underset{\sim}{\Sigma}_{11}$, $\underset{\sim}{\Sigma}_{12}^* = \underset{\sim}{\Sigma}_{21}^{*\prime} = \underset{\sim}{K}_{p-r} \underset{\sim}{\Sigma}_{21} \underset{\sim}{K}_r$, as used in proving (4.3.5), we may easily obtain the following proposition :

Proposition 4.3.4 : (a) The canonical correlations between $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ are the same as those between $\underset{\sim}{X}_1^*$ and $\underset{\sim}{X}_2^*$, e.g. $\varphi_j = \varphi_j^*$, $j = 1, 2, \cdots, q$.

(b) The canonical coefficient vectors $\underset{\sim}{a}_j$, $\underset{\sim}{\beta}_j$, $\underset{\sim}{a}_j^*$, $\underset{\sim}{\beta}_j^*$ would obey the relationship

$$\underset{\sim}{a}_j^* = \underset{\sim}{K}_{p-r} \underset{\sim}{\beta}_j, \quad \underset{\sim}{\beta}_j^* = \underset{\sim}{K}_r \underset{\sim}{a}_j$$

$j = 1, 2, \cdots, q$. ∎

Remark 4.5 Defining canonical coefficient matrices $\underset{\sim}{L}_1$ (q x r) and $\underset{\sim}{L}_2$(q x $\overline{p-r}$) as

$$\underset{\sim}{L}_1 = (\underset{\sim}{a}_1 \ \underset{\sim}{a}_2 \cdots \underset{\sim}{a}_q)' ; \ \underset{\sim}{L}_2 = (\underset{\sim}{\beta}_1 , \underset{\sim}{\beta}_2 \cdots \underset{\sim}{\beta}_q)'$$

for q pairs of canonical variables, and similarly defining $\underset{\sim}{L}_1^*$ and $\underset{\sim}{L}_2^*$, we may easily see that

$$\underset{\sim}{L}_1^* = \underset{\sim}{L}_2 \underset{\sim}{K}_{p-r} , \ \underset{\sim}{L}_2^* = \underset{\sim}{L}_1 \underset{\sim}{K}_r \ .$$

Remark 4.6 The correlation vectors of canonical variables with original variable (technically called the vector of canonical variable loadings) for the two situations ($\underset{\sim}{X}_1$,$\underset{\sim}{X}_2$)

and $(\underline{X}_2^{*}\ ,\ \underline{X}_2^{*})$ would also maintain the same relationship as proposed in Proposition 4.3.4(b).

Remark 4.7 Various measures of associations (between two vector variates) which directly involve canonical correlations would have the same sort of equality as stated in Proposition 4.3.4(a). Some such measures are vector alienation coefficient (VAC), vector correlation coefficient (VCC), generalized coefficient of determination (GCD) [cf. Takeuchi, et al. (1981), Chap. 6].

Proposition 4.3.5 (a) Let $\underline{X}_1^{(i)} = (x_i, x_{i+h}, \cdots, x_{i+(k-1)h})'$ and $\underline{X}_2^{(j)} = (x_j, x_{j+h}, x_{j+2h}, \cdots, x_{j+(k-1)h})'$ be two k-component subvectors with the Toeplitz property (e.g. $\mathrm{cov}(x_t, x_u) = a^{|t-u|}$ for any t and u) among the components. If we denote by $\underline{\alpha}_v$ and $\underline{\beta}_v$ as the k-component coefficient vectors of the pair of canonical variables corresponding to the v-th canonical correlation (φ_v) between $\underline{X}_1^{(i)}$ and $\underline{X}_2^{(j)}$, then

$$\underline{\beta}_v = K_k\, \underline{\alpha}_v \quad , \tag{4.3.15}$$

$v = 1(1)k$.

(b) The above-mentioned canonical correlations (φ_v) and the coefficient vectors ($\underline{\alpha}_v\ ,\ \underline{\beta}_v$) for each $v = 1(1)k$ remains unchanged so long as the difference of i and j remains the same.

Proof : (a) The relevant submatrices $\underset{\sim}{\Sigma}_{11}$, $\underset{\sim}{\Sigma}_{12}$ and $\underset{\sim}{\Sigma}_{22}$ may be easily formed and found that

$$\underset{\sim}{\Sigma}_{11} = \underset{\sim}{\Sigma}_{22} = \underset{\sim}{K}_k\underset{\sim}{\Sigma}_{11}\underset{\sim}{K}_k = \underset{\sim}{K}_k\underset{\sim}{\Sigma}_{22}\underset{\sim}{K}_k \quad \text{and}$$

$\underset{\sim}{\Sigma}_{21} = \underset{\sim}{\Sigma}'_{12} = \underset{\sim}{K}_k\underset{\sim}{\Sigma}_{12}\underset{\sim}{K}_k$. Using these upon (4.3.13) and (4.3.14), (4.3.15) may be easily proved.

(b) Since $\underset{\sim}{\Sigma}_{12}$-matrix remains invariant as long as the difference between i and j remains the same, the statement is very simple to establish. ∎

4.3.2 Partitioning of Observation Vector into Three Sets and Equalities on Conditional Dispersion Matrices

Let $\underline{X}$ be partitioned into 3 sets of variables with similarly partitioned Σ as follows :

$$\underset{p\times 1}{\underline{X}} = \begin{bmatrix} \underline{X}_1 \\ r\times 1 \\ \underline{X}_2 \\ (p-2r)\times 1 \\ \underline{X}_3 \\ r\times 1 \end{bmatrix}, \quad \underset{\sim}{\Sigma} = \begin{bmatrix} \underset{\sim}{\Sigma}_{11} & \underset{\sim}{\Sigma}_{12} & \underset{\sim}{\Sigma}_{13} \\ \underset{\sim}{\Sigma}_{21} & \underset{\sim}{\Sigma}_{22} & \underset{\sim}{\Sigma}_{23} \\ \underset{\sim}{\Sigma}_{31} & \underset{\sim}{\Sigma}_{32} & \underset{\sim}{\Sigma}_{33} \end{bmatrix} . \quad (4.3.16)$$

In general, the conditional dispersion matrix of $\begin{pmatrix} \underline{X}_1 \\ \underline{X}_3 \end{pmatrix}$ given $\underline{X}_2 = \underline{x}_2$ is given by

$$\underset{\sim}{\Sigma}_{\underline{X}_1\underline{X}_3\cdot\underline{X}_2} = \begin{bmatrix} \underset{\sim}{\Sigma}_{11.2} & \underset{\sim}{\Sigma}_{13.2} \\ \Sigma_{31.2} & \Sigma_{33.2} \end{bmatrix} \quad (4.3.17)$$

where $\underset{\sim}{\Sigma}_{ij.2} = \underset{\sim}{\Sigma}_{ij} - \underset{\sim}{\Sigma}_{12}\,\underset{\sim}{\Sigma}_{22}^{-1}\,\underset{\sim}{\Sigma}_{2j}$, i, j = 1,3 .

For Toeplitz $\underset{\sim}{\Sigma}$, the $\underset{\sim}{\Sigma}_{ii}$-matrices are Toeplitz, $i = 1,2,3$, and $\underset{\sim}{\Sigma}_{11} = \underset{\sim}{\Sigma}_{33}$, $\underset{\sim}{\Sigma}_{13} = \underset{\sim}{K}_r \underset{\sim}{\Sigma}_{31} \underset{\sim}{K}_r$, and $\underset{\sim}{\Sigma}_{23} = \underset{\sim}{K}_{p-2r} \underset{\sim}{\Sigma}_{21} \underset{\sim}{K}_r$. Using all of these, we may easily establish the following result :

<u>Proposition 4.3.6</u> : <u>Given $\underline{X}_2 = \underline{x}_2$, the conditional dispersion matrices of $\underline{X}_1$ and $\underset{\sim}{K}_r \underline{X}_3$ are equal (i.e. $\underset{\sim}{\Sigma}_{11.2} = \underset{\sim}{K}_r \underset{\sim}{\Sigma}_{33.2} \underset{\sim}{K}_r$)</u> .

As an immediate consequence, we have the equality on the two relevant partial correlation matrices formed from $\underline{X}_1$ and $\underset{\sim}{K}_r \underline{X}_3$ ($= \underline{X}_3^*$, say), when $\underline{X}_2$ is partialled out.

<u>Proposition 4.3.7</u> $\underset{r \times r}{\underset{\sim}{\rho}_{\underline{X}_1 \cdot \underline{X}_2}} = \underset{r \times r}{\underset{\sim}{\rho}_{\underline{X}_3^* \cdot \underline{X}_2}}$.

<u>Remark 4.8</u> We may note here that Proposition 43.7 should not be confused with Proposition 4.3.3, as only the common $(p-2r)$ variables [e.g. x_i, $i = (r+1)(1)(p-r+1)$] are partialled out in the latter case.

In particular, for a single conditioning variable, we may obtain interesting expressions of the partial correlations between pair of suitably spaced variables [cf. Anderson, 1984, p. 43]. Three such expressions are as follows.

$$\rho_{ij.t} = \rho_{p-i+1,p-j+1.\ p-t+1}, \quad i \neq j \neq t = 1,2,\ldots, p. \tag{4.3.19}$$

$$\rho_{i-k,\ i+k.i} = (\rho_{2k} - \rho_k^2)/(1-\rho_k^2), \qquad (4.3.19)$$

$i = k+1, k+2, \ldots, p-k$ for some given $k \in \{1,2, \ldots, [(p-1)/2]\}$. Alternatively, for any given pair (i,j) $1 < i, j \leq p$,

$$\rho_{k+i,k+j.k} = \rho_{k-i,k-j.k} = \frac{\rho_{|i-j|} - \rho_i\rho_j}{\sqrt{1-\rho_i^2}\sqrt{1-\rho_j^2}} \qquad (4.3.20)$$

$\forall$ k such that (k+i) and (k+j) should not exceed p .

Similar to Proposition 4.3.4, we may easily obtain that eigenvalues of $\underset{\sim}{\Sigma}_{12}\underset{\sim}{\Sigma}_{22}^{-1}\underset{\sim}{\Sigma}_{21}\underset{\sim}{\Sigma}_{11}^{-1}$ are the same as those of $\underset{\sim}{\Sigma}_{32}\underset{\sim}{\Sigma}_{22}^{-1}\underset{\sim}{\Sigma}_{23}\underset{\sim}{\Sigma}_{33}^{-1}$ for a Toeplitz $\underset{\sim}{\Sigma}$, so that we arrive at the following equality on canonical correlations.

Proposition 4.3.8 (a) The canonical correlations between $\underline{X}_1$ and $\underline{X}_2$ equal those between $\underline{X}_3$ and $\underline{X}_2$. [i.e. symbolically, $\rho_{\underline{X}_1,\ \underline{X}_2} = \rho_{\underline{X}_3,\underline{X}_2}$].

(b) The coefficient vectors of the canonical variables of the former pair of subsets $(\underline{X}_1,\ \underline{X}_2)$ are those (in the reverse order) of the latter pair of subsets. ■

We may now introduce the concept of partial canonical correlation (Rao, 1969). Let the (linear) effect of $\underline{X}_2$ be partialled out from $\underline{X}_1$ and $\underline{X}_3$. Then the conditional dispersion matrix of $\begin{pmatrix}\underline{X}_1\\ \underline{X}_3\end{pmatrix}$ given $\underline{X}_2$ is $\underset{\sim}{\Sigma}_{\underline{X}_1,\underline{X}_3.\underline{X}_2}$ as expressed in(4.3.17). Partial canonical correlations ($\rho_{\underline{X}_1,\underline{X}_3.\underline{X}_2}$) are the square root

of the eigenvalues of $\underset{\sim}{\Sigma}_{13.2}\,\underset{\sim}{\Sigma}_{33.2}^{-1}\,\underset{\sim}{\Sigma}_{31.2}\underset{\sim}{\Sigma}_{11.2}^{-1}$. We may establish some equality on these partial canonical correlations when the vector $\underline{X}$ is partitioned into 4 sets with $p = 4r$ (for simplicity).

4.3.3 Partitioning of Observation Vector into Four Sets and Equality on Partial Canonical Correlations

Let $\underline{X}$ be partitioned into 4 sets, $\underline{X}_i$ being the i-th set of order $(r \times 1)$, $i = 1, 2, 3, 4$. $p = 4r$. The Toeplitz covariance matrix $\underset{\sim}{\Sigma}(4r \times 4r)$ in its partitioned form appears as

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \underset{\sim}{\Sigma}_0 & \underset{\sim}{\Sigma}_1 & \underset{\sim}{\Sigma}_2 & \underset{\sim}{\Sigma}_3 \\ \underset{rxr}{} & & & \\ \underset{\sim}{\Sigma}_1' & \underset{\sim}{\Sigma}_0 & \underset{\sim}{\Sigma}_1 & \underset{\sim}{\Sigma}_2 \\ \underset{\sim}{\Sigma}_2' & \underset{\sim}{\Sigma}_1' & \underset{\sim}{\Sigma}_0 & \underset{\sim}{\Sigma}_1 \\ \underset{\sim}{\Sigma}_3' & \underset{\sim}{\Sigma}_2' & \underset{\sim}{\Sigma}_1' & \underset{\sim}{\Sigma}_0 \end{bmatrix} \qquad (4.3.21)$$

We propose below some equalities in the partial canonical correlations.

Proposition 4.3.9 (i) $\rho_{\underline{X}_1, \underline{X}_2 \cdot \underline{X}_3} = \rho_{\underline{X}_2, \underline{X}_3 \cdot \underline{X}_4} = \rho_{\underline{X}_3, \underline{X}_4 \cdot \underline{X}_2}$

(ii) $\rho_{\underline{X}_1, \underline{X}_3 \cdot \underline{X}_2} = \rho_{\underline{X}_2, \underline{X}_4 \cdot \underline{X}_3}$

Proof : $\rho^2_{\underline{X}_1, \underline{X}_2 \cdot \underline{X}_3}$ are the eigenvalues of $\underset{\sim}{\Sigma}_{12.3}\underset{\sim}{\Sigma}_{22.3}^{-1}\underset{\sim}{\Sigma}_{21.3}\underset{\sim}{\Sigma}_{11.3}^{-1}$.

We note from (4.3.21), that

$$\underset{\sim}{\Sigma}_{12.3} = \underset{\sim}{\Sigma}_{12} - \underset{\sim}{\Sigma}_{13}\underset{\sim}{\Sigma}_{33}^{-1}\underset{\sim}{\Sigma}_{32} = \underset{\sim}{\Sigma}_{1} - \underset{\sim}{\Sigma}_{2}\underset{\sim}{\Sigma}_{o}^{-1}\underset{\sim}{\Sigma}_{1}' = \underset{\sim}{\Sigma}_{21.3}' ,$$

$$\underset{\sim}{\Sigma}_{22.3} = \underset{\sim}{\Sigma}_{22} - \underset{\sim}{\Sigma}_{23}\underset{\sim}{\Sigma}_{33}^{-1}\underset{\sim}{\Sigma}_{32} = \underset{\sim}{\Sigma}_{o} - \underset{\sim}{\Sigma}_{1}\underset{\sim}{\Sigma}_{o}^{-1}\underset{\sim}{\Sigma}_{1}' ,$$

$$\underset{\sim}{\Sigma}_{11.3} = \underset{\sim}{\Sigma}_{11} - \underset{\sim}{\Sigma}_{13}\underset{\sim}{\Sigma}_{33}^{-1}\underset{\sim}{\Sigma}_{31} = \underset{\sim}{\Sigma}_{o} - \underset{\sim}{\Sigma}_{2}\underset{\sim}{\Sigma}_{o}^{-1}\underset{\sim}{\Sigma}_{2}' .$$

Similarly, for $\rho^2_{\underline{X}_2 , \underline{X}_3 \cdot \underline{X}_4}$,

$$\underset{\sim}{\Sigma}_{23.4} = \underset{\sim}{\Sigma}_{1} - \underset{\sim}{\Sigma}_{2}\underset{\sim}{\Sigma}_{o}^{-1}\underset{\sim}{\Sigma}_{1}' = \underset{\sim}{\Sigma}_{32.4}' = \underset{\sim}{\Sigma}_{12.3} = \underset{\sim}{\Sigma}_{21.3}' ,$$

$$\underset{\sim}{\Sigma}_{22.4} = \underset{\sim}{\Sigma}_{11.3} , \underset{\sim}{\Sigma}_{33.4} = \underset{\sim}{\Sigma}_{22.3} .$$

Due to centrosymmetric property of $\underset{\sim}{\Sigma}_{o}$, $\underset{\sim}{\Sigma}_{12.3} = \underset{\sim}{K}_{r}\underset{\sim}{\Sigma}_{23.4}\underset{\sim}{K}_{r}$, $\underset{\sim}{\Sigma}_{11.3} = \underset{\sim}{K}_{r}\underset{\sim}{\Sigma}_{22.4}\underset{\sim}{K}_{r}$ and $\underset{\sim}{\Sigma}_{22.3} = \underset{\sim}{K}_{r}\underset{\sim}{\Sigma}_{33.4}\underset{\sim}{K}_{r}$. These results lead to the fact that eigenvalues of $\underset{\sim}{\Sigma}_{12.3}\underset{\sim}{\Sigma}_{22.3}^{-1}\underset{\sim}{\Sigma}_{21.3}\underset{\sim}{\Sigma}_{11.3}^{-1}$ are same as those of $\underset{\sim}{K}_{r}(\underset{\sim}{\Sigma}_{12.3}\underset{\sim}{\Sigma}_{22.3}^{-1}\underset{\sim}{\Sigma}_{21.3}\underset{\sim}{\Sigma}_{11.3}^{-1})\underset{\sim}{K}_{r}$, which are again the same as those of $\underset{\sim}{\Sigma}_{23.4}\underset{\sim}{\Sigma}_{33.4}^{-1}\underset{\sim}{\Sigma}_{32.4}\underset{\sim}{\Sigma}_{22.4}^{-1}$. This indicates that $\rho_{\underline{X}_1 , \underline{X}_2 \cdot \underline{X}_3} = \rho_{\underline{X}_2 , \underline{X}_3 \cdot \underline{X}_4}$ completing the proof of (i). Similarly, the proof of (ii) can be advanced. ∎

The concept of bi-partial canonical correlation [Timm and Carlson, 1976] may now be introduced as it requires essentially the four subvectors, e.g. $\underline{X}_i$, $i = 1, 2, 3, 4$.

Bipartial canonical correlations $\rho_{(\underline{X}_1 \cdot \underline{X}_2), (\underline{X}_3 \cdot \underline{X}_4)}$ between $\underline{X}_1$ and $\underline{X}_3$ when the (linear) effects of $\underline{X}_2$ from $\underline{X}_1$ and those of $\underline{X}_4$ from $\underline{X}_3$ are partialled out, is defined as the square root of the eigenvalues of

$\Sigma^{*}_{13}\Sigma^{*-1}_{33}\Sigma^{*}_{31}\Sigma^{*-1}_{11}$ where $\Sigma^{*}_{11} = \Sigma_{11}-\Sigma_{12}\Sigma^{-1}_{22}\Sigma_{21}$,

$\Sigma^{*}_{13} = \Sigma^{*'}_{31} = \Sigma_{13} - \Sigma_{12}\Sigma^{-1}_{22}\Sigma_{23} - \Sigma_{14}\Sigma^{-1}_{44}\Sigma_{43}+\Sigma_{12}\Sigma^{-1}_{22}\Sigma_{24}\Sigma^{-1}_{44}\Sigma_{43}$,

$\Sigma^{*}_{33} = \Sigma_{33} - \Sigma_{34}\Sigma^{-1}_{44}\Sigma_{43}$.

If the variates in $\underline{X}_2$ are uncorrelated with the variates in $\underline{X}_1$ (i.e. when $\Sigma_{12} = 0$), we get a particular case of bipartial canonical correlation, which is termed as part canonical correlation (Timm and Carlson, 1976] between $\underline{X}_1$ and $\underline{X}_3$ when the effect of $\underline{X}_4$ from $\underline{X}_3$ are only eliminated. Notationally, part canonical correlation $\rho_{\underline{X}_1,(\underline{X}_3\cdot\underline{X}_4)}$ are the positive square roots of the eigenvalues of $\Sigma_{13.4}\Sigma^{-1}_{33.4}\Sigma_{31.4}\Sigma^{-1}_{11}$ where the notations $\Sigma_{13.4}$, $\Sigma_{31.4}$ and $\Sigma_{33.4}$ are defined following (4.3.17).

The following proposition may be easily proved paralleling Proposition 4.3.9 for p = 4r.

<u>Proposition 4.3.10</u> (i) $\rho_{(\underline{X}_1\cdot\underline{X}_2),\,(\underline{X}_3\cdot\underline{X}_4)} = \rho_{(\underline{X}_2\cdot\underline{X}_1),(\underline{X}_4\cdot\underline{X}_3)}$

(ii) $\rho_{\underline{X}_1,(\underline{X}_3\cdot\underline{X}_4)} = \rho_{\underline{X}_2,(\underline{X}_4\cdot\underline{X}_3)}$. ■

4.3.4 Partitioning of Observation Vector into more than Four Sets and Block Toeplitz Matrices

Let p = br where b and r are integers, $r \geq 1$ but $b > 4$. Then vector $\underline{X}$ may be partitioned into b subvectors

of r components each. The Toeplitz $\underset{\sim}{\Sigma}$(br x br) may then be partitioned accordingly into a b x b array of blocks of size r x r each. Typically, any (i,j) block would be equal to $\underset{\sim}{\Sigma}_{|j-i|}$ if $i \leq j$ and equal to $\underset{\sim}{\Sigma}'_{|j-i|}$ if $i > j$ and $\underset{\sim}{K}_r \underset{\sim}{\Sigma}_{|j-i|} \underset{\sim}{K}_r = \underset{\sim}{\Sigma}'_{|j-i|}$ $\forall$ i,j = 1(1)b . Such partitioned - form Toeplitz matrix is termed as Block Toeplitz matrix.

All sorts of measures of association as mentioned above may be extended accordingly. We only add one set of equalities on bipartial canonical, apart from its extension based on Proposition 4.3.10(i).

Proposition 4.3.11 $\underline{\rho_{(\underline{X}_1 \cdot \underline{X}_2),(\underline{X}_3 \cdot \underline{X}_4)}} = \underline{\rho_{(\underline{X}_2 \cdot \underline{X}_3),(\underline{X}_4 \cdot \underline{X}_5)}} = \underline{\rho_{(\underline{X}_3 \cdot \underline{X}_4),\ (\underline{X}_5 \cdot \underline{X}_6)}} = \cdots\cdots$. ∎

4.3.5 Principal Component Analysis

Principal component analysis (PCA) is often useful in reducing the dimension of the multivariate data so that most of the relevant information is contained in the data set of smaller dimension [Hotelling, 1933]. Specific treatments are available in any book on multivariate analysis [cf. Anderson (1984), Morrison (1976), Muirhead (1982), Seber (1984), Takeuchi et al. (1981)].

We shall consider here PCA of random variable $\underline{X}$(p x 1) having a Toeplitz covariance matrix $\underset{\sim}{\Sigma}_p$. To do this, we may recall various results from section 4.2. Consider

the partitioned form of $\underset{\sim}{\Sigma}_p$ and orthogonal matrix $\underset{\sim}{P}_p$[vide (4.2.24), (4.2.26), (4.2.5) and (4.2.6)].

For even p(= 2m), principal components are given by

$$\underset{2m \times 1}{\underline{Y}} = \underset{\sim}{P}^*_{2m}\,\underline{X} \qquad (4.3.22)$$

where $\underset{\sim}{P}^*_{2m}$ = diag. $(\underset{\sim}{\Gamma}_1', \underset{\sim}{\Gamma}_2')\,\underset{\sim}{P}_{2m}$ with orthogonal matrices $\underset{\sim}{\Gamma}_1$ and $\underset{\sim}{\Gamma}_2$ such that

$$\underset{\sim}{\Gamma}_1'(\underset{\sim}{\Sigma}_m + \underset{\sim}{\Delta}_m\underset{\sim}{K}_m)\,\underset{\sim}{\Gamma}_1 = \text{diag.}(\lambda_1, \lambda_2, \cdots, \lambda_m) \qquad (4.3.23)$$

and

$$\underset{\sim}{\Gamma}_2'(\underset{\sim}{\Sigma}_m - \underset{\sim}{\Delta}_m\underset{\sim}{K}_m)\,\underset{\sim}{\Gamma}_2 = \text{diag.}(\lambda_{m+1}, \lambda_{m+2}, \cdots, \lambda_{2m}). \qquad (4.3.24)$$

$\lambda_1, \lambda_2, \cdots, \lambda_{2m}$ are the eigenvalues of $\underset{\sim}{\Sigma}_{2m}$. We may note that

$$\underset{\sim}{P}^*_{2m} = \frac{1}{\sqrt{2}} \begin{bmatrix} \underset{\sim}{\Gamma}_1' & \underset{\sim}{\Gamma}_1'\underset{\sim}{K}_m \\ \underset{\sim}{\Gamma}_2' & -\underset{\sim}{\Gamma}_2'\underset{\sim}{K}_m \end{bmatrix} \qquad (4.3.25)$$

so that the vectors of component loadings related to each principal component are, in general, either symmetric or skew symmetric. [Only exception may occur when some λ_i in both (4.3.23) and (4.3.24) is found in common. In that case related vectors are neither symmetric nor skew symmetric].

Furthermore, 2m linear combinations (functions) of 2m components of $\underline{X}$ are uncorrelated. Out of $(2m)^2$ component

loadings [i.e. the elements of $\underset{\sim}{P}^{*}_{2m}$ in (4.2.25)], only $2m^2$ component loadings (namely the elements of $\underset{\sim}{\Gamma}_1$ and $\underset{\sim}{\Gamma}_2$) are necessary to express $\underset{\sim}{P}^{*}_{2m}$.

In the same manner, for odd p (= 2m+1), the principal components are given by

$$\underset{(2m+1)\ x\ 1}{\underline{Y}} = \underset{\sim}{P}^{*}_{2m+1}\ \underline{X} \qquad (4.3.26)$$

with

$$\underset{\sim}{P}^{*}_{2m+1} = \frac{1}{\sqrt{2}} \begin{bmatrix} \underset{\sim}{\Gamma}^{*\prime}_1 & \sqrt{2}\ \underline{e} & \underset{\sim}{\Gamma}^{*\prime}_1\ \underset{\sim}{K}_m \\ \underline{\gamma}'_{m+1} & \sqrt{2}\ e_o & \underline{\gamma}_{m+1}\underset{\sim}{K}_m \\ \underset{\sim}{\Gamma}'_2 & \underline{0} & -\underset{\sim}{\Gamma}'_2\ \underset{\sim}{K}_m \end{bmatrix}, \qquad (4.3.27)$$

and orthogonal matrices

$$\underset{\sim}{\Gamma}_1 = \underset{m+1\ x\ m+1}{} \begin{bmatrix} \underset{\sim}{\Gamma}^{*}_1 & \underline{\gamma}_{m+1} \\ \underline{e}' & e_o \end{bmatrix}$$

and $\underset{\sim}{\Gamma}_2$ are so chosen that

$$\underset{\sim}{\Gamma}'_1 \begin{bmatrix} \underset{\sim}{\Sigma}_m + \underset{\sim}{\Delta}_m\underset{\sim}{K}_m & \sqrt{2}\ \underset{\sim}{K}_m\ \underline{\xi}_m \\ \sqrt{2}\ \underline{\xi}'_m\ \underset{\sim}{K}_m & a_o \end{bmatrix} \underset{\sim}{\Gamma}_1 = \mathrm{diag}(\lambda_1, \cdots, \lambda_{m+1}) \qquad (4.3.28)$$

and
$$\Gamma_2' (\Sigma_m - \Delta_m K_m) \Gamma_2 = \text{diag}(\lambda_{m+2}, \cdots, \lambda_{2m+1}). \quad (4.3.29)$$

Expression (4.3.23) of P^*_{2m+1} ensures that out of $(2m+1)^2$ component loadings, we need to know only $m^2+(m+1)^2$ component loadings which are the elements of Γ_1 and Γ_2.

(i) An Assessment of Block-superiority

As $P_{2m}\Sigma_{2m}P_{2m}'$ is a block diagonal matrix, the sub-vectors $Z_1(mx1)$ and $Z_2(mx1)$ of the partitioned $Z(2mx1) = P_{2m} X = (Z_1' Z_2')'$, are uncorrelated. The block-superiority of Z_1 and Z_2 will be measured by the largeness of their explained proportions of the total variance, namely $tr(\Sigma_{2m})$. If α and β be such proportions, we note that

$$\alpha,\beta = \frac{tr(\Sigma_m \pm \Delta_m K_m)}{tr(\Sigma_{2m})} = \frac{1}{2} \pm \frac{1}{2m} (\rho_1 + \rho_3 + \cdots + \rho_{2m-1}) \quad (4.3.30)$$

with the notation $\rho_i = a_i/a_o$.

Similarly, for $p = 2m+1$,

$$\alpha,\beta = \frac{m+1}{2m+1} \pm \frac{1}{2m+1} (\rho_2 + \rho_4 + \cdots + \rho_{2m}). \quad (4.3.31)$$

Thus, the values of α and β depend upon the same subset of ρ_i-values which may be positive or negative proper fractions as Σ_{2m} or Σ_{2m+1} is positive definite.

When α is so large that β may be neglected, we shall select Z_1 as the principal (sub) vector with regard

to a consideration that proportion (α) explained by $\underset{\sim}{Z}_1$ is considerably much more than that (β) by $\underset{\sim}{Z}_2$, indicating certain block-superiority of $\underset{\sim}{Z}_1$ to $\underset{\sim}{Z}_2$.

Since $\underset{\sim}{P}_{2m}$ or $\underset{\sim}{P}_{2m+1}$ has very simple expression independent of a_i-values, construction of $\underset{\sim}{Z}_1$ and $\underset{\sim}{Z}_2$ is no longer difficult.

(ii) Assessment by the First and Last Principal Components

Mathematically, the assessment of " the proportions of variance accounted for" can be done by computing the 1st and last principal components (i.e., the pair of components in $\underset{\sim}{Y}$ (vide 4.3.22))in relation to the largest and smallest eigenvalues of $\underset{\sim}{\Sigma}_p$.

Recall in this connection Propositions 4.2.6 and 4.2.7 to calculate primarily the limits of the largest and the smallest proportions (π_1 and π_2, say) of actually computing them by principal component analysis (PCA). A formal PCA would be effective if at least the lower limits of π_1 is not much less as well as if the upper limit of π_p is very small. If the calculated upper bound of π_1 is found to be very low (say, less than 50%), the reduction of dimension by means of PCA will be no longer meaningful. On the other hand, while observing the lower limit of π_1 as reasonably high, we may readily go forward to determine the 1st principal component. The reduction to k(< p) principal

components is then done by checking successively the proportions explained. We ignore other (p-k) principal components that have smaller variances without significantly effecting the total variance.

In order to judge how good the first k(< p) principal components are in approximating the original p variables, we find multiple correlations between x_i and $y_1, \cdots, y_k$, and the residual variances $var(x_i/y_1, \cdots, y_k)$ which are given by

$$\rho^2_{x_i(y_1,\cdots,y_k)} = \sum_{j=1}^{k} \lambda_j p^{*2}_{ij}/\sigma_{ii} \qquad (4.3.32)$$

$$\sigma^2_{x_i(y_1,\cdots,y_k)} = \sigma_{ii} - \sum_{j=1}^{k} \lambda_j p^{*2}_{ij} \qquad (4.3.33)$$

where p^*_{ij}'s are the 1st k elements in the i-th row of the orthogonal matrix $\underset{\sim}{P}^*_{2m}$(or $\underset{\sim}{P}^*_{2m+1}$) and σ_{ii} is the variance of x_i which is equal to as $\forall$ i for Toeplitz covariance matrix $\underset{\sim}{\Sigma}_{2m}$ or $\underset{\sim}{\Sigma}_{2m+1}$.

(iii) Detection of Toeplitz Structure via a " Formal" Principal Component Analysis

Given a sample covariance matrix $\underset{\sim}{S}$(pxp), we can judge the closeness of its structure to centrosymmetric by the closeness of $\underset{\sim}{S}$ and $\underset{\sim p}{K}\underset{\sim}{S}\underset{\sim p}{K}$. To do such an analysis, one may check whether the largest eigenvalue of $\underset{\sim}{S}-\underset{\sim p}{K}\underset{\sim}{S}\underset{\sim p}{K}$ is very close to zero.

For the detection of Toeplitz structure, we may also consider the displacement ranks of $\underset{\sim}{S}$ [vide (vii(g)) of subsec. 4.2.1]. For a Toeplitz structure, such ranks are exactly equal to 2. $\underset{\sim}{S}(pxp)$ would be " close" to Toeplitz if $\underset{\sim}{S}-\underset{\sim}{U}_p \underset{\sim}{S} \underset{\sim}{U}_p'$ and $\underset{\sim}{S}-\underset{\sim}{U}_p' \underset{\sim}{S} \underset{\sim}{U}_p$ have ranks (called displacement ranks of $\underset{\sim}{S}$) approximately equal to 2. The determination of approximate rank may be done in many ways. Since both the matrices $\underset{\sim}{S}-\underset{\sim}{U}_p \underset{\sim}{S}\underset{\sim}{U}_p'$ and $\underset{\sim}{S}-\underset{\sim}{U}_p' \underset{\sim}{S}\underset{\sim}{U}_p$ are symmetric, one important procedure is to calculate their eigenvalues. Due to possible indefiniteness of $\underset{\sim}{S}-\underset{\sim}{U}_p \underset{\sim}{S}\underset{\sim}{U}_p'$ and $\underset{\sim}{S}-\underset{\sim}{U}_p' \underset{\sim}{S} \underset{\sim}{U}_p$, their eigenvalues may assume both the positive and negative values. Since the rank of a symmetric matrix is equal to the number of its non-zero eigenvalues, from the computational point of view, the ranks of $\underset{\sim}{S}- \underset{\sim}{U}_p' \underset{\sim}{S} \underset{\sim}{U}_p$ and $\underset{\sim}{S}- \underset{\sim}{U}_p \underset{\sim}{S} \underset{\sim}{U}_p'$ may be adjudged by the number of non-zero eigenvalues which are, in magnitude, very much different from zero. For the purpose of a comparison, we calculate proportions of the absolute eigenvalues to their total which are nothing but the proportions relative to the total singular values. We expect that when $\underset{\sim}{S}$ is very much close to Toeplitz matrix, two largest singular values of $\underset{\sim}{S} - \underset{\sim}{U}_p' \underset{\sim}{S} \underset{\sim}{U}_p$ (or of $\underset{\sim}{S} - \underset{\sim}{U}_p \underset{\sim}{S} \underset{\sim}{U}_p'$) are fairly large so that the total proportions accounted by them would be very much close to unity. Thus, this observation may be treated as a motivation for a "formal" principal component analysis based on symmetric but (possibly) indefinite matrix

$\underset{\sim}{S} - \underset{\sim}{U}_p' \underset{\sim}{S} \underset{\sim}{U}_p$ (or $\underset{\sim}{S} - \underset{\sim}{U}_p \underset{\sim}{S} \underset{\sim}{U}_p'$). It may be noted that for a Toeplitz matrix $\underset{\sim}{\Sigma}$ (as defined in (4.2.7)), the eigenvalues of $\underset{\sim}{\Sigma} - \underset{\sim}{U}_p \underset{\sim}{\Sigma} \underset{\sim}{U}_p'$ are given by

$$\delta_1 \ , \delta_p = \frac{a_o}{2} [\, 1 \pm (1+4b)^{1/2} \,] \ , \delta_2 = \cdots = \delta_{p-1} = 0$$

with $b = (a_1^2 + a_2^2 + \cdots + a_{p-1}^2)/a_o^2$. Clearly, $\delta_1 > 0$ and $\delta_p < 0$. The corresponding eigenvectors $\underline{Z}_1 , \underline{Z}_2 , \cdots , \underline{Z}_p$ are such that

$$\underline{Z}_1 \propto (\delta_1 ; a_1, a_2, \cdots , a_{p-1})'$$

$$\underline{Z}_p \propto (\delta_p : a_1, a_2, \cdots , a_{p-1})'$$

$$\underline{Z}_j = (0 ; z_{1j}, z_{2j}, \cdots , z_{p-1j})' ,$$

$$j = 2,3, \cdots , p-1$$

where components of $\underline{Z}_j$ are so chosen that (p x p) matrix $(\underline{Z}_1, \underline{Z}_2 , \cdots , \underline{Z}_p)$ is an orthogonal matrix. Similarly, using the property of $\underset{\sim}{U}_p$ that $\underset{\sim}{U}_p' = \underset{\sim}{K}_p \underset{\sim}{U}_p \underset{\sim}{K}_p$ [vide 4.2.1 (vii(a))], the eigenvectors of $\underset{\sim}{\Sigma} - \underset{\sim}{U}_p \underset{\sim}{\Sigma} \underset{\sim}{U}_p$ are $\underset{\sim}{K}_p \underline{Z}_1$, $K_p \underline{Z}_2 , \cdots , K_p \underline{Z}_p$, eigenvalues being the same as δ_1 , δ_2, $\cdots , \delta_p$ respectively.

While one may run a " formal" principal component analysis upon $(\underset{\sim}{\Sigma} - \underset{\sim}{U}_p \underset{\sim}{\Sigma} \underset{\sim}{U}_p')$ or $(\underset{\sim}{\Sigma} - \underset{\sim}{U}_p' \underset{\sim}{\Sigma} \underset{\sim}{U}_p)$; the expressions of (δ_1 , δ_p) and $(\underline{Z}_1 , \underline{Z}_p)$ may be useful to get algebraically the expressions of the Toeplitz parameters $a_o , a_1 , \cdots , a_{p-1}$.

Remark 4.9 Since intraclass structure is a special case of Toeplitz structure when $a_1 = a_2 = \cdots = a_{p-1}$; the last (p-1) components of $\underline{Z}_1$ and $\underline{Z}_p$ would be then equal. This additional phenomenon will be helpful to detect the intraclass structure from the collection of Toeplitz matrices. In practice, when (i) $\delta_2, \cdots, \delta_{p-1}$ are very much close to zero, (ii) the first component of each of $\underline{Z}_2, \cdots, \underline{Z}_{p-1}$ are close to zero, and (iii) the last (p-1) components of the difference $\underline{Z}_1 - \underline{Z}_{p-1}$ are approximately zero, we shall expect the Toeplitz matrix to be the best possible fitting structure. Further if (iv) last (p-1) components of each of $\underline{Z}_1$ and $\underline{Z}_p$ are more or less the same, we may then assert that the intraclass structure would be the most plausible hypothesis for model fitting purpose.

4.3.6 Empirical Verification of Some Propositions by GRE Data

Let us consider the sample (5 x 5) covariance matrix ($\underset{\sim}{S}$) based on GRE data (Werts et al., 1981) and MLE ($\hat{\underset{\sim}{\Sigma}}$) of the Toeplitz covariance matrix as given below. The source and interpretation of $\underset{\sim}{S}$-matrix will be discussed in Chap. 5.

$$\underset{\sim}{S} = \begin{bmatrix} 11008 & & & & \\ 9394 & 11425 & & & \\ 8960 & 9587 & 10721 & \text{Sym.} & \\ 8963 & 9569 & 9388 & 10831 & \\ 8051 & 8547 & 8846 & 9014 & 9726 \end{bmatrix} . \qquad (4.3.34)$$

The ML estimates of the p number of unknown parameters of $\underset{\sim}{\Sigma}$ were obtained following the Q-procedure discussed in Chap.2.

$$\hat{\underset{\sim}{\Sigma}} = \begin{bmatrix} 10659.31 & & & & \\ 9167.53 & 10659.31 & & & \\ 9032.56 & 9167.53 & 10659.31 & \text{Sym.} & \\ 8655.31 & 9032.56 & 9167.53 & 10659.31 & \\ 8347.98 & 8655.31 & 9032.56 & 9167.53 & 10659.31 \end{bmatrix} \quad (4.3.35)$$

Based on both $\underset{\sim}{S}$ and $\hat{\underset{\sim}{\Sigma}}$ matrices, we calculate the square multiple correlations $\rho^2_{(j)}$, $j = 1,2,3,4,5$ and partial correlation matrices $((\rho_{(i,j)}))_{5x5}$ to verify the Propositions 4.3.2 and 4.3.3 for $r = 2$ [vide eqn. (4.3.12)].

Table 4.1

Squared Multiple Correlations based on sample covariance matrix ($\underset{\sim}{S}$) and estimated Toeplitz Covariance matrix ($\hat{\underset{\sim}{\Sigma}}$).

Variable (j)	Squared Multiple correlation $(\rho^2_{(j)})$	
	$\underset{\sim}{S}$	$\hat{\underset{\sim}{\Sigma}}$
X_1	.7543	.7886
X_2	.8189	.8229
X_3	.8447	.8360
X_4	.8496	.8229
X_5	.8142	.7886

In Table 4.1, the 3rd column related to $\hat{\underset{\sim}{\Sigma}}$ satisfies the Proposition 4.3.2 whereas its 2nd column related to $\underset{\sim}{S}$ fails to satisfy even upto two places of decimals.

The matrix of partial correlations obtained from the sample covariance matrix $\underset{\sim}{S}$ is given below :

$$\begin{bmatrix} 1.000 & & & & \\ .330 & 1.000 & & & \\ .198 & .323 & 1.000 & \text{Sym.} & \\ .181 & .282 & .191 & 1.000 & \\ .041 & .017 & .361 & .441 & 1.000 \end{bmatrix} \qquad (4.3.36)$$

From the MLE($\hat{\underset{\sim}{\Sigma}}$) of Toeplitz covariance matrix, corresponding matrix of partial correlations for the population appears as

$$\begin{bmatrix} 1.000 & & & & \\ .405 & 1.000 & & & \\ .290 & .243 & 1.000 & \text{Sym.} & \\ .103 & .248 & .243 & 1.000 & \\ .048 & .103 & .290 & .405 & 1.000 \end{bmatrix} \qquad (4.3.37)$$

For Toeplitz structure, partial correlation matrix [e.g. (4.3.37)] is a centrosymmetric matrix as proposed in (4.3.12). Other sorts of partial correlations (for $r > 2$) may be computed similarly to verify the Proposition 4.3.3, in general. We omit them here.

So far as the sample partial correlation matrix [e.g. (4.3.36)] is concerned, the entries do not show the centrosymmetric behaviour at least approximately. Hence, the pattern of $\underset{\sim}{S}$ to be of Toeplitz form is not supported by this analysis.

We shall now switch over to canonical correlation analysis to verify the Propositions 4.3.4, 4.3.5 and 4.3.8. Table 4.2 shows the canonical correlations computed both on the basis of $\underset{\sim}{S}$ and $\hat{\underset{\sim}{\Sigma}}$ separately for as many as four pairs of comparable subvector pairs $(\underline{X}_1 , \underline{X}_2)$.

Tables 4.3 and 4.4 show the coefficient vectors of the corresponding canonical variables separately for $\underset{\sim}{S}$ and $\hat{\underset{\sim}{\Sigma}}$. It may be noted that the columns of Tables 4.2 and 4.4 satisfy Proposition 4.3.4, by the canonical correlation values reported in the first two rows, and Proposition 4.3.5 by rows 3,4 and 5,6. Rows 7 and 8 satisfy Proposition 4.3.8 although it may be treated as a particular case (obtainable obviously from Principal Component Analysis) as the first subvectors reduce to a single variable only since p = 5 here.

Table 4.2

Canonical Correlation between Two Sets of Variables Based on $\underset{\sim}{S}$ and the Estimated Toeplitz Covariance Matrix $(\hat{\underset{\sim}{\Sigma}})$.

Components of 1st set	Components of 2nd set	Canonical correlations $\underset{\sim}{S}$		Canonical correlations $\hat{\underset{\sim}{\Sigma}}$	
X_1, X_2, X_3	X_4, X_5	.918	.188	.910	.086
X_1, X_2	X_3, X_4, X_5	.912	.022	.910	.086
X_1, X_2	X_3, X_4	.912	.006	.908	.082
X_2, X_3	X_4, X_5	.914	.178	.908	.082
X_1, X_3	X_2, X_4	.924	.070	.922	.163
X_2, X_4	X_3, X_5	.922	.231	.922	.163
X_3	X_4, X_5	.896	-	.887	-
X_3	X_1, X_2	.885	-	.887	-

Table 4.3

Standardised Coefficient Vectors for Canonical Variables Based on $\underset{\sim}{S}$

Components of 1st set	Standardised coefficient vectors of the canonical variables	Components of 2nd set	Standardised coefficient vectors of the canonical variables
x_1, x_2, x_3	(.190,.302,.557),(-.633,-1.531, 2.067)	x_4, x_5	(.630,.400),(-1.994,2.053)
x_1, x_2	(.380,.660),(1.791, -1.708)	x_3, x_4, x_5	(.5 19,.468,.049),(-1.508,-.578, 2.219)
x_1, x_2	(.379,.661),(1.791,-1.707)	x_3, x_4	(.540,.494),(-1.964,1.976)
x_2, x_3	(.395,.538),(1.962,-1.897)	x_4, x_5	(.617,.414),(1.998,-2.050)
x_1, x_3	(.397,.647),(1.723,-1.646)	x_2, x_4	(.532,.505),(1.888, -1.895)
x_2, x_4	(.389,.646),(-1.922,1.852)	x_3, x_5	(.586,.448),(-1.914,1.951)
x_3	1.0	x_4, x_5	(.538, .493)
x_3	1.0	x_1, x_2	(.376, .664)

Table 4.4

Standardised Coefficient Vectors for Canonical Variables Based on the Estimated Toeplitz Covariance Matrix

Components of 1st set	Standardised coefficient vectors of the canonical variables	Components of 2nd set	Standardised coefficient vectors of the canonical variables
x_1, x_2, x_3	(.136,.356,.552),(-.625,-1.523, 2.068)	x_4, x_5	(.613,.422),(-1.879,1.930)
x_1, x_2	(.422,.613),(1.930,-1.879)	x_3, x_4, x_5	(.552,.356,.136),(2.068,-1.523, -.625).
x_1, x_2	(.951,.977),(.309,-.215)	x_3, x_4	(.977,.951),(.215,-.309)
x_2, x_3	(.425,.610),(1.930,-1.880)	x_4, x_5	(.610,.425),(1.880,-1.930)
x_1, x_3	(.442,.597),(1.842,-1.798)	x_2, x_4	(.597,.442),(1.798,-1.842)
x_2, x_4	(.442,.597),(1.842,-1.798)	x_3, x_5	(.597,.442),(1.798,-1.842)
x_3	1.0	x_4, x_5	(.572,.464)
x_3	1.0	x_1, x_2	(.464,.572)

The results of Principal Component Analysis (PCA) based on $\underset{\sim}{S}$ and $\hat{\underset{\sim}{\Sigma}}$ are presented in Tables 4.5, 4.6 and 4.7. We observe that more than 87 % of the total variance is explained by the 1st principal component of both $\underset{\sim}{S}$ and $\hat{\underset{\sim}{\Sigma}}$. The percentages related to other components are quite low and systematically smaller for both $\underset{\sim}{S}$ and $\hat{\underset{\sim}{\Sigma}}$. As regards the vectors of component loadings, those for $\hat{\underset{\sim}{\Sigma}}$ are either symmetric or skew-symmetric [vide equ. (4.3.26)]. The two vectors of component loadings related to the first principal components of $\underset{\sim}{S}$ and $\hat{\underset{\sim}{\Sigma}}$ are quite close to each other, although those corresponding to other principal components differ considerably.

For the possible verification of Toeplitz structure on the basis of $\underset{\sim}{S}$, we report below Table 4.8 on the eigenvalues and proportion of singular values related to each of $\underset{\sim}{S}-\underset{\sim}{U}_p\underset{\sim}{S}\underset{\sim}{U}_p'$, $\underset{\sim}{S}-\underset{\sim}{U}_p'\underset{\sim}{S}\underset{\sim}{U}_p$, $\hat{\underset{\sim}{\Sigma}} - \underset{\sim}{U}_p\hat{\underset{\sim}{\Sigma}}\underset{\sim}{U}_p'$ and $\hat{\underset{\sim}{\Sigma}}-\underset{\sim}{U}_p'\hat{\underset{\sim}{\Sigma}}\underset{\sim}{U}_p$ with $p = 5$. Only the two largest singular values of either $\underset{\sim}{S}-\underset{\sim}{U}_p\underset{\sim}{S}\underset{\sim}{U}_p'$ or $\underset{\sim}{S}-\underset{\sim}{U}_p'\underset{\sim}{S}\underset{\sim}{U}_p$ not only account for more than 97 % of the total singular values, but also very much close to proportions of corresponding singular values of $\hat{\underset{\sim}{\Sigma}} - \underset{\sim}{U}_p\hat{\underset{\sim}{\Sigma}}\underset{\sim}{U}_p'$ (or of $\hat{\underset{\sim}{\Sigma}} - \underset{\sim}{U}_p'\hat{\underset{\sim}{\Sigma}}\underset{\sim}{U}_p$). Additionally, the proportions corresponding to other singular values are very much nearer to zero. Therefore, the ranks of $\underset{\sim}{S}-\underset{\sim}{U}_p\underset{\sim}{S}\underset{\sim}{U}_p'$ and $\underset{\sim}{S}-\underset{\sim}{U}_p'\underset{\sim}{S}\underset{\sim}{U}_p$ may be heuristically taken to be 2 so that the fitting of Toeplitz structure to $\underset{\sim}{S}$ is quite reasonable from the mathematical point of view [vide subsec. 4.3.5(iii)].

Table 4.5

Results of PCA of $\underset{\sim}{S}$ and $\underset{\sim}{\hat{\Sigma}}$: Percentage of Variation Explained by Principal Components

Principal Components	Explained Variance		Percentage of Total Variance explained	
	$\underset{\sim}{S}$	$\underset{\sim}{\hat{\Sigma}}$	$\underset{\sim}{S}$	$\underset{\sim}{\hat{\Sigma}}$
I	46920.16	46435.06	87.36	87.12
II	2562.95	2585.11	4.77	4.85
III	1756.37	1532.89	3.27	2.88
IV	1378.27	1390.51	2.57	2.61
V	1093.25	1352.98	2.03	2.54

Table 4.6

Results of PCA of $\underset{\sim}{S}$: Component Loadings

Variables	Principal Components of $\underset{\sim}{S}$				
	I	II	III	IV	V
X_1	.442	-.722	.529	-.014	.053
X_2	.464	-.254	-.763	.160	.337
X_3	.453	.172	-.123	-.763	-.410
X_4	.456	.275	.067	.626	-.566
X_5	.420	.556	.344	-.018	.629

Table 4.7

Results of PCA of $\hat{\Sigma}$: Component Loadings

Variables	Principal Components of $\hat{\Sigma}$				
	I	II	III	IV	V
X_1	.441	-.624	.430	.347	-.333
X_2	.450	-.333	-.538	.094	.624
X_3	.453	0	.229	-.861	0
X_4	.450	.333	-.538	.094	-.624
X_5	.441	.624	.430	.347	-.333

Table 4.8

Selected Findings Connected with the Formal PCA of the Resulting Matrices Obtained for Displacement Rank Analyses Based on $\underset{\sim}{S}$ and $\underset{\sim}{\hat{\Sigma}}$

Eigenvalues			Proportion of singular value to the total		
$\underset{\sim}{S}-\underset{\sim}{U}_p\underset{\sim}{S}\underset{\sim}{U}_p'$	$\underset{\sim}{\hat{\Sigma}}-\underset{\sim}{U}_p\underset{\sim}{\hat{\Sigma}}\underset{\sim}{U}_p'$ or $\underset{\sim}{\hat{\Sigma}}-\underset{\sim}{U}_p'\underset{\sim}{\hat{\Sigma}}\underset{\sim}{U}_p$	$\underset{\sim}{S}-\underset{\sim}{U}_p'\underset{\sim}{S}\underset{\sim}{U}_p$	$\underset{\sim}{S}-\underset{\sim}{U}_p\underset{\sim}{S}\underset{\sim}{U}_p'$	$\underset{\sim}{\hat{\Sigma}}-\underset{\sim}{U}_p\underset{\sim}{\hat{\Sigma}}\underset{\sim}{U}_p'$ or $\underset{\sim}{\hat{\Sigma}}-\underset{\sim}{U}_p'\underset{\sim}{\hat{\Sigma}}\underset{\sim}{U}_p$	$\underset{\sim}{S}-\underset{\sim}{U}_p'\underset{\sim}{S}\underset{\sim}{U}_p$
23855.4974	23731.8052	23123.4339	.6271	.6445	.6335
27.3263	0	507.3104	.0007	0	.0139
-94.8342	0	123.8496	.0025	0	.0034
-509.9972	0	-144.0031	.0134	0	.0039
-13551.9923	-13072.4952	-12602.5909	.3563	.3555	.3453

4.4 LARGE SAMPLE TESTS FOR TESTING TOEPLITZ COVARIANCE STRUCTURE

We shall consider in this section some large sample test procedures which are very specific to test the null hypothesis (under multinormality assumption) given by

$$H_o : \{\text{Population covariance matrix } (\underset{\sim}{\Sigma}) \text{ is a p.d. Toeplitz matrix}\} \qquad (4.4.1)$$

against the alternative

$$H : \{\underset{\sim}{\Sigma} \text{ is simply a p.d. matrix}\} . \qquad (4.4.2)$$

Three single-stage test criteria which are, in general, applicable to ACOVS (as discussed in Ch. 3, subsec. 3.1.3) may be applied easily to test H_o against H. We shall now propose two-stage test procedures employing the idea of centrosymmetric matrices (vide subsec. 4.2.1) for specifically testing the Toeplitz covariance structure.

4.4.1 Two-stage Tests

(i) Successive tests of Centrosymmetric hypotheses

Keeping in view the Proposition 4.2.3 stating a characterization of symmetric Toeplitz (ST) matrix based on a pair of symmetric centrosymmetric (SC) matrices, the H_o (6.1.1) may be treated as $H_o \equiv H_{o1} \cap H_{o2}$ where

$$H_{o1} : \{ \underset{\sim}{\Sigma}_{p-1} \text{ is an SC matrix} \} \qquad (4.4.3)$$

and

$$H_{o2} : \{ \underset{\sim}{\Sigma} \text{ is an SC matrix} \} . \qquad (4.4.4)$$

The notation $\underset{\sim}{\Sigma}_{p-1}$ in (4.4.3) denotes the population covariance matrix of the first (p-1) components (say, subvector $\underline{X}^*$) of $\underline{X}(px1)$. Both H_{o1} and H_{o2} are to be tested individually against unrestricted alternatives. Acceptance of both H_{o1} and H_{o2} implies acceptance of H_o . On the contrary, H_o will be rejected if either or both H_{o1} and H_{o2} are rejected. These two subtests may be performed by the conventional LR procedure after obtaining the explicit MLE of an SC matrix at each stage [vide Ch. 2, subsec. 2.2.3(d)] as follows :

$$\lambda_1 = \det \underset{\sim}{S}_{p-1} / \det[(\underset{\sim}{S}_{p-1} + \underset{\sim}{K}_{p-1} \underset{\sim}{S} \underset{\sim}{K}_{p-1})/2] \qquad (4.4.5)$$

$$\lambda_2 = \det \underset{\sim}{S} / \det[(\underset{\sim}{S} + \underset{\sim}{K}_p \underset{\sim}{S} \underset{\sim}{K}_p/2] \qquad (4.4.6)$$

with $\underset{\sim}{S}_{p-1}$ as the principal submatrix obtained from S(pxp) by deleting its last row and last column. The notation $\underset{\sim}{K}_m$ stands for a flip matrix of order m (vide subsec. 4.2.1(1)). Under H_{oi} , the statistic

$$T_i = - N \log_e \lambda_i \qquad (4.4.7)$$

is distributed asymptotically as a χ^2 with v_i degrees of freedom (d.f.) with $v_i = p_i(p_i+1)/2 - k_i$. k_i = number of estimated parameters under $H_{oi} = [(p_i+1)/2] [(p_i+2)/2]$. $p_i = p-1$ for i = 1, and $p_i = p$ for i = 2.

A level -α_i test rule is to reject H_{oi} if the computed value of T_i exceeds $\chi^2_{\alpha_i}(\upsilon_i)$, the critical value of χ^2 with υ_i d.f. for α_i level of significance, i = 1,2.

As regards control of test size, any two-stage procedure is more conservative than the single-stage level-α test when we choose $\alpha_1 + \alpha_2 = \alpha$ and the level of significance for testing H_{oi} does not exceed α_i , i = 1,2. Its proof is based on the Bonferroni's inequality.

If further, the dependence between two events, E_1 and E_2 , is assumed to be very small, $P(E_2/E_1) \approx P(E_2)$, where E_i denotes the acceptance of H_{oi} given that H_{oi} is true, i = 1,2, then approximate size (α^*) of the two-stage test is given by

$$1 - \alpha^* \approx (1-\alpha_1)(1-\alpha_2) \cdot \qquad (4.4.8)$$

Thus, the two-stage test may provide a reasonably high degree of protection against type I error for the entire null hypothesis.

(ii) <u>Nested hypothesis testing</u>

It may be noted that a class $\mathcal{Y}_p(SC)$ of symmetric centrosymmetric (SC) matrices forms an Abelian group under addition while the nonsingular members of $\mathcal{Y}_p(SC)$ form a non-Abelian group under multiplication. Symmetric Toeplitz (ST matrices form a subclass $\mathcal{Y}_p(ST) \subset \mathcal{Y}_p(SC)$.

Keeping in view the null hypothesis $H_o(4.4.1)$, we may construct a nested frame as follows. At the first stage, we shall test

$$H^*_{o1} : \{\underset{\sim}{\Sigma} \in \mathcal{V}_p(SC)\} . \qquad (4.49)$$

If H^*_{o1} is tenable, we shall next test the conditional hypothesis

$$H^*_{o2/1} : \{\underset{\sim}{\Sigma} \in \mathcal{V}_p(ST)/\underset{\sim}{\Sigma} \in \mathcal{V}_p(SC)\} . \qquad (4.4.10)$$

We are allowed to regard the H_o as not being unreasonable if both H^*_{o1} and $H^*_{o2/1}$ are <u>not</u> rejected. The alternative hypothesis corresponding to H^*_{o1} is the <u>unrestricted</u> $\underset{\sim}{\Sigma}$-matrix without any particular pattern while that corresponding to $H^*_{o2/1}$ is naturally the centrosymmetric $\underset{\sim}{\Sigma}$-matrix. Using LR principle, we define

$$\lambda^*_1 = \det \underset{\sim}{S} / \det[(\underset{\sim}{S} + \underset{\sim}{K}_p \underset{\sim}{S} \underset{\sim}{K}_p)/2] \qquad (4.4.11)$$

$$\lambda^*_{2/1} = \det[(\underset{\sim}{S} + \underset{\sim}{K}_p \underset{\sim}{S} \underset{\sim}{K}_p)/2] / \det \underset{\sim}{\hat{\Sigma}} \qquad (4.4.12)$$

where $\underset{\sim}{\hat{\Sigma}}$ is the MLE of the Toeplitz $\underset{\sim}{\Sigma}$-matrix. Under H_{o1} ,

$$T^*_1 = - N \log_e \lambda^*_1 \qquad (4.4.13)$$

follows asymptotically a χ^2 with $\upsilon^*_1 = p(p+1)/2 - [(p+1)/2][p+2)/2]$.

Similarly, under $H^*_{o2/1}$,

$$T^*_{2/1} = - N \log_e \lambda^*_{2/1} \qquad (4.4.13)$$

follows asymptotically a χ^2 with $\upsilon^*_{2/1} = [(p+1)/2][(p+2)/2]-p$. Test rules for these nested hypotheses may be framed accordingly.

Nested hypothesis testing procedure is generally undertaken to control the probability of Type I error so that it does not exceed the nominal significance level. Defining

$$A = \left\{\text{rejection of } H^*_{01}\right\}, \ \bar{A} = \left\{\text{Acceptance of } H^*_{01}\right\},$$

$$B = \left\{\text{rejection of } H^*_{02/1}\right\}, \ C = \left\{\text{rejection of } H_o\right\},$$

actual nested probability of type I error is expressible as

$$P(C/H_o) = P(A/H_o) + P(\bar{A}/H_o)P(B/\bar{A}\ H_o). \qquad (4.4.14)$$

Let α_1 be the level of significance of testing H^*_{01}. For a given α_1 which is lower than α, one should expect that

$$\alpha_1 + (1-\alpha_1) \sup_{H_o} P(B/\bar{A}\ H_o) \leq \alpha$$

$$\text{i.e., } \sup_{H_o} P(B/\bar{A}\ H_o) \leq \frac{\alpha-\alpha_1}{1-\alpha_1}. \qquad (4.4.15)$$

The suitability of such test procedures depends on how closely $\sup P(B/H_o)$ approximates $\sup P(B/\bar{A}\ H_o)$ [Schott, 1987, sec. 4].

Remark 4.10 Procedure (i) has an arbitrariness in the sense that this may lead sometimes to different conclusions when the testing is done by deleting the first component variable of $\underline{X}(p \times 1)$ in stead of the last one. Furthermore, this procedure is not capable of providing the MLE of the Toeplitz parameters once H_o is accepted.

4.4.2 Optimum Properties of the Subtests

Although in the two-stage tests (i) and (ii) described in the previous section, we do not obtain the exact sizes of the tests, the subtests related to H_{o1}, H_{o2} and H^{*}_{o1} possess some optimum properties. Such properties are common in the case of LR statistic applicable to testing the centrosymmetric covariance matrix. Without any loss, we shall discuss these properties considering only H^{*}_{o1} as follows :

Since LR test is always invariant [cf. Lehmann, 1959, p. 252; Eaton, 1983, p. 263], we apply orthogonal transformation $\underset{\sim}{Z} = \underset{\sim}{P}_{p}\ \underset{\sim}{X}$ with $\underset{\sim}{P}_{p}$ as defined in equation (4.2.5) and (4.2.6). It is then possible to partition $\underset{\sim}{Z}$ as

$$\underset{1xp}{\underline{Z}'} = (\ \underset{1x[\frac{p+1}{2}]}{\underline{Z}_1'}\qquad \underset{1x[\frac{p}{2}]}{\underline{Z}_2'}\ ).$$

We then equivalently test H^{**}_{o1}:{the stochastic independence of two subvectors $\underline{Z}_1$ and $\underline{Z}_2$}, since a centrosymmetric covariance matrix of $\underline{X}$ would be block diogonalized by the above mentioned orthogonal transformation. Following Narain (1950), such a test is strictly unbiased, i.e., the probability of rejecting H_{o1} is greater than the nominal significance level if H^{**}_{o1} is not true. Furthermore, it is admissible adopting a (0, 1) loss function [Keifer and Schwartz, 1965]. As this test is invariant under the group of affine transformation

within each component subvectors, its result may be used for testing the nullity of all canonical correlations between component subvectors $\underline{Z}_1$ and $\underline{Z}_2$. The monotonicity of the power function based on λ_1^* is confirmed along the concept of Anderson and Das Gupta (1964). Moreover, by using available non-null distribution of such a λ-statistic, approximate power function may be calculated for a very specific set of local alternatives each formulated in terms of the largest canonical correlation [cf. Muirhead, 1982, Ch. 11].

4.4.3 An Illustrative Example

For the purpose of illustration, we shall consider the sample covariance matrix ($\underset{\sim}{S}$) of the same GRE data as used in subsec. 4.3. To test H_o (4.4.1) under normality assumption, we calculate

$$\det \underset{\sim}{S} = 3.1825 \times 10^{17} \quad , \det \hat{\underset{\sim}{\Sigma}} = 3.4617 \times 10^{17}$$

$$\mathrm{tr}(\hat{\underset{\sim}{Q}}\,\underset{\sim}{S}) = -\,0.1687 \text{ where } \hat{\underset{\sim}{Q}} = \hat{\underset{\sim}{\Sigma}}^{-1} - \hat{\underset{\sim}{\Sigma}}^{-1}\,\underset{\sim}{S}\,\hat{\underset{\sim}{\Sigma}}^{-1} \ . \quad \text{The}$$

matrices $\underset{\sim}{S}$ and $\hat{\underset{\sim}{\Sigma}}$ are already referred to in equations (4.3.34) and (4.3.35). We now apply the single-stage likelihood ratio criterion as well as Rao's criterion [vide Ch. 3, eqns.(3.1.16) and (3.1.29)] and calculate

$$-N \log_e (\det \underset{\sim}{S}/\det \hat{\underset{\sim}{\Sigma}}) = 18.25, -(N/2)\mathrm{tr}(\hat{\underset{\sim}{Q}}\,\underset{\sim}{S}) = 18.30$$

both of which are less than $\chi^2_{.05}(10)$ (= 18.31). The result suggests the tenability of the Toeplitz structure H_o(4.4.1).

But as these computed values are very close to the critical value $\chi^2_{.05}(10)$, we feel some hesitation to accept H_o from the point of view of goodness-of-fit of the Toeplitz structure. The frequently used measure of goodness-of-fit, namely probability above the likelihood ratio chi-square value (PRCHI) [vide Ch. 3, subsec. 3.2] is found to be $P[\chi^2(10) > 18.25]$ which is as small as 5.5 percent only indicating a poor fit of the Toeplitz structure to $\underset{\sim}{S}$.

Now, in order to apply the proposed two-stage tests, we calculate $\det[(\underset{\sim}{S} + \underset{\sim}{K}_5 \underset{\sim}{S}\underset{\sim}{K}_5)/2] = 3.4114 \times 10^{17}$, $\det \underset{\sim}{S}_4 = 1.7618 \times 10^{14}$ and $\det[(\underset{\sim}{S}_4 + \underset{\sim}{K}_4 \underset{\sim}{S}_4 \underset{\sim}{K}_4)/2] = 1.8141 \times 10^{14}$ where $\underset{\sim}{S}_4$ is the $\underset{\sim}{S}$-matrix with last row and last column deleted. Using (4.4.7), we next calculate $T_1 = 6.34$ and $T_2 = 15.00$ which are respectively less than $\chi^2_{.05}(4)$ (= 9.49) and greater than $\chi^2_{.05}(6)$ (= 12.59). Thus, although H_{o1} is tanable, H_{o2} is not so. As a result, we reject H_o . Large value of T_2 compared to T_1 is perhaps due to the introduction of last row and last column of $\underset{\sim}{S}$ in computing T_2 . The two-stage test therefore has a feature of diagnostic identification for possible rejection of the H_o .

If we delete the first row and first column of $\underset{\sim}{S}$-matrix in stead of its last row and last column, the statistic T_1 would be changed to 5.69 which is less than $\chi^2_{.05}(4)$ and hence supports the tenability of H_{o1} . However,

the hypothesis of equality of the elements σ_{41} and σ_{52} is very much questionable. But as the calculated value (15.00) of T_2 exceeds $\chi^2_{.05}(6)$ as seen earlier, H_{o2} is no longer tenable. Hence, the overall H_o remains untenable on the basis of the proposed two-stage test.

As regards the testing of nested hypothesis by (ii), we observe that first subhypothesis H^*_{o1} is the same as H_{o2} [of procedure (i)] which has already been tested and found not tenable. Hence, we are no longer interested to test the second hypothesis $H^*_{o2/1}$. Acceptance of Toeplitz null hypothesis cannot be, therefore, supported by both the two-stage procedures (i) and (ii).

It may be worthwhile to mention here that for the GRE data, it is more appropriate to think of the plausibility of Toeplitz correlation structure than other structures [Mukherjee and Maiti, 1988b]. The results in this connection are discussed, in brief, in subsection 5.4.2 of Chapter 5.

4.5 DISCUSSION

Statistical tests of significance may be sometimes affected by serial correlations when data are collected in time. It is usually necessary to assume a time series model to represent and explain serial correlations. One of the finite parameter models for stationary time series process evolves Toeplitz covariance matrix in which the covariance

elements depend only upon the difference of two time points. A p.d. Toeplitz matrix has interesting applications not only in the area of spectral analysis of time series, but also in the analysis of repeated measurements and in nonparametric theory. Various fields of its applications were discussed, in brief, in the introductory section.

Mathematical properties and forms of the Toeplitz matrix lead to particularly simple expressions for eigenvalues and eigenvectors, many of which are conventionally applied in structural engineering. Although a symmetric Toeplitz matrix is not diagonally reducible by a nonsingular matrix having elements independent of the Toeplitz matrix, some of the results reported here are expected to help in an adhoc checking of the hypothesis as to whether an empirical time series can be fitted by an autoregressive process of a given order and also in the testing of autocovariance structure.

Since Toeplitz matrix and Hankel matrix have a suitable relationship (vide subsec. 4.2.1(v)), all results about quasidirect decomposition of Hankel matrices apply to results about quasidirect decomposition of Toeplitz matrices and conversely (Fiedler, 1984). This result, together with the fact that a p.d. Toeplitz matrix of order m can be uniquely factored with $O(m^2)$ element operations [Bareiss (1969), Rissanen (1973)] may have important implications for

any new computational algorithm relating to Toeplitz matrices. Zellini (1979) has furthermore shown that the matrix product $\underset{\sim}{A}\,\underline{b}$ over a real field when $\underset{\sim}{A}$ is an m x m symmetric Toeplitz matrix requires at least (2m-1) multiplications, since the matrix $\underset{\sim}{A}$ has tensor rank 2(m-1) in the real field. This has important implications for the computation of a finite set of bilinear forms formulated as a matrix product.

The Levinson - Durbin algorithm for inverting a Toeplitz matrix has been generalized to accommodate arbitrary nonsingular matrices via the introduction of a Toeplitz distance concept [e.g.Friedlander , Morf, Kailath and Ljung (1979); Kailath, Kung and Morf (1979)]. The Toeplitz distance d of a matrix reflects the complexity of the algebraic structure of the matrix considered in relation to the Toeplitz structure. Computational efficiency for inverting a given matrix decreases linearly with increase in its Toeplitz distance d . More precisely, with the help of the generalized Levinson algorithm the inversion can be performed at the cost of $(d+2)m^2$ operations.

Toeplitz structure has been generalized to the block form (Akaike, 1973) . Kailath and Kolitracht (1986) have obtained conditions for a nonsingular matrix to have a block Toeplitz inverse. Greville and Trench (1979) found conditions for a nonsingular band matrix to have an inverse with the Toeplitz structure.

Symmetric Toeplitz matrices of order m and their block extensions are members of an m-dimensional quadratic subspace $\mathcal{B}$ of real symmetric matrices as defined by Seely (1971). The subspace $\mathcal{B}$ is quadratic if and only if $\underset{\sim}{A}^2 \in \mathcal{B}$ i.e., $\mathcal{B}$ is closed under the multiplication $\underset{\sim}{A} \circ \underset{\sim}{C} = \frac{1}{2}(\underset{\sim}{A}\underset{\sim}{C}+\underset{\sim}{C}\underset{\sim}{A})$. Jensen (1975, 1977) observes that the latter property makes $\mathcal{B}$ an m-dimensional special Jordan algebra. By Seely's Lemma 2(a), inverses of these matrices are members of a quadratic subspace. Thus, all resulting matrices form part of the special Jordan algebra (Jacobson, 1968). Hence, the useful properties of Jordan algebra can be fruitfully applied in solving various mathematical and statistical problems connected with nonsingular Toeplitz matrix, such as variance components analysis using MINQUE (cf. Rao, 1974, Ch. 4).

By exploiting the special feature of the Toeplitz covariance matrix (or Toeplitz correlation matrix), computational short-cuts have been obtained for various measures of association. A formal application of the Principal Component Analysis (PCA) has also been made to ascertain whether or not the rank of a matrix of order p of the form $\underset{\sim}{S}-\underset{\sim}{U}_p\underset{\sim}{S}\underset{\sim}{U}_p'$ or $\underset{\sim}{S}-\underset{\sim}{U}_p\underset{\sim}{S}\underset{\sim}{U}_p'$ is close to 2. If this result holds approximately, then we can formally proceed to use a suitable test of significance of a "striped" (Toeplitz) covariance matrix as proposed in subsection 4.4. In the event of sample size being either too large or too small, this heuristic procedure in

addition seems to be quite useful when the conventional LR test fails to detect the hidden structure because of its dependence on sample size. In the case of sample correlation matrix $\underset{\sim}{R}$, we may use similar formal PCA upon $\underset{\sim}{R}-\underset{\sim}{U}_p'\underset{\sim}{R}\underset{\sim}{U}_p$ or $\underset{\sim}{R}-\underset{\sim}{U}_p\underset{\sim}{R}\underset{\sim}{U}_p'$ to see if the rank is close to 2 in order to ascertain whether or not the Toeplitz correlation structure is embedded in the covariance matrix.

For testing the tenability of the Toeplitz covariance hypothesis, two-stage sample tests have been proposed. Although these are conservative tests, the optimum properties of the subtests are discussed. The proposed two-stage tests are different from the two-stage (double-sample) procedure suggested by Hewett and Spurrier (1983), but similar in spirit to those proposed by Anderson (1962), Hogg (1962) and Mustafi (1967). These authors were interested in testing iteratively a sequence of hypotheses H_r , $r = 1, \ldots$, m against a sequence of alternatives $\bar{H}_r$, $r = 1, 2, \ldots$, m. Although these authors treated the problem from the standpoint of multiple decision theory, we have not shown here explicitly how the assumption of a UMP similar test which is essential for testing such a sequence of null hypotheses is valid in the present case.

Despite certain arbitrary features of the proposed tests, the technique involved can be easily applied to more general situations for analysis of a wide class of covariance structures. For example, we can make use of the nesting approach for testing the equality of the elements in the

principal diagonal and the two adjoining subdiagonals of the covariance matrix arising from the first-order autoregressive AR(1) model. For this purpose, we may test at the first stage that the matrix has a tri-diagonal Jacobi structure against arbitrary p.d. alternative. Using the LR test proposed by Mukherjee (1966, 1970), we can verify if the hypothesis is accepted. If it is accepted then we take up at the second stage the testing concerned with further restrictions imposed on the equality of the elements within principal diagonal and each subdiagonal. For diagnostic checks, the two tests may also be applied in the reverse order, should we find that the two-parameter AR(1) covariance structure (cf. Judge et al., 1985, p. 276) is rejected at the very first stage. Such diagnostic checking is quite essential in time series modelling (McAleer et al., 1988).

4.6 SUMMARY

An overview on the algebraic properties of p.d. Toeplitz matrix is given. Matrices of this kind often arise, from statistical point of view, in econometrics, psychometrics, structural engineering, multichannel filtering, reflection seismology etc. and it is desirable to have techniques which exploit their special structure. Possible applications of the results related to their inverse, determinant and eigenvalue problem are suggested. Specific features of various statistical measures such as multiple, partial, bipartial, canonical

correlation and so on due to involvement of Toeplitz covariance (correlation) structure are studied. With the aid of Principal Component Analysis, an adhoc procedure for detecting the Toeplitz structure of the covariance (correlation) matrix before running a formal test of significance is suggested. Specific large sample two-stage tests are suggested for formal test of significance of the Toeplitz covariance structure. The test is illustrated with life data. Some of the advantages of the test are discussed in the concluding section.

CHAPTER 5

AN ITERATIVE PROCEDURE FOR OBTAINING MLE OF PARAMETERS OF TOEPLITZ CORRELATION STRUCTURE

5.1 INTRODUCTION

Let us consider $N_p(\underset{\sim}{X} ; \underset{\sim}{\mu} , \underset{\sim}{\Sigma})$ where the mean vector $\underset{\sim}{\mu}$ is perfectly arbitrary but the covariance matrix $\underset{\sim}{\Sigma}$ is expressible as $\underset{\sim}{D}_{\sigma}\underset{\sim}{\rho}\underset{\sim}{D}_{\sigma}$ with $\underset{\sim}{D}_{\sigma} = \text{diag}\,(\sigma_1, \sigma_2, \cdots, \sigma_p)$ and $\underset{\sim}{\rho}$, the correlation matrix assumed to have Toeplitz structure as follows :

$$\underset{p\times p}{\rho} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{p-3} \\ \vdots & & & & \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{bmatrix}$$

In short, we can write the "stripe" matrix $\underset{\sim}{\rho}$ as

$$\underset{\sim}{\rho} = \underset{\sim}{I}_p + \sum_{t=1}^{p-1} \rho_t \underset{\sim}{H}_t \qquad (5.1.1)$$

where the design matrix $\underset{\sim}{H}_t$ is a sparse matrix with 1's on the t-th upper and lower subdiagonals parallel to principal diagonal, t = 1(1)(p-1). We assume that p scale parameters $\sigma_1, \sigma_2, \cdots, \sigma_p$ and (p-1) correlation parameters $\rho_1, \rho_2, \cdots, \rho_{p-1}$ are all functionally independent.

We are interested, in this chapter, to employ an iterative scheme which is applicable to the above-mentioned Toeplitz correlation structure. Although the iterative scheme originates essentially from the Q^*-procedure (discussed in Ch. 2, sec. 2.5), by introducing the operation of a new type of sum (we call it as "systematic sum"), we reduce here considerable amount of computational labour in the ML estimation of (2p-1) parameters of the Toeplitz correlation structure. Sections 5.2 and 5.3 are used to develop the iterative scheme. Analysis of GRE data (Werts et al., 1981) is provided in section 5.4 to illustrate the proposed iterative scheme. Lastly, in section 5.5, we make some concluding remarks.

5.2 THE EFFECT OF NORMAL EQUATIONS UPON Q^*-MATRIX

The general Q^*-procedure (vide chap. 2, sec. 2.5) is based on the "matrix equation"

$$\underset{\sim}{\rho} = \underset{\sim}{S}^* + \underset{\sim}{\rho}\ \underset{\sim}{Q}^*\ \underset{\sim}{\rho} \qquad (5.2.1)$$

where $\underset{\sim}{S}^* = \underset{\sim}{D}_\sigma^{-1}\ \underset{\sim}{S}\underset{\sim}{D}_\sigma^{-1}$ and $\underset{\sim}{Q}^* = \underset{\sim}{D}_\sigma\ \underset{\sim}{Q}\ \underset{\sim}{D}_\sigma$.

We write below the normal equations in the case of Toeplitz correlation structure as

$$\text{diag.}\ (\underset{\sim}{Q}^*\underset{\sim}{\rho}) = \underline{0}(p \times 1) \qquad (5.2.2)$$

$$\text{tr}\ (\underset{\sim}{Q}^*\underset{\sim}{H}_t) = 0\ ,\ t = 1(1)(p-1). \qquad (5.2.3)$$

In order to develop the specific iterative scheme, we shall first show the effect of these normal equations on the structure and magnitude of the elements of $\underset{\sim}{Q}^*$.

Since (2p-1) normal equations, e.g. (5.2.2) and (5.2.3) restrict the elements of $\underset{\sim}{Q}^*$-matrix linearly, the matrix $\underset{\sim}{Q}^*$ will thus undergo changes in its elements so as to get $\underset{\sim}{Q}^*$ expressed as

$$\underset{\sim}{Q}^* = \sum_{i=1}^{p-2} \sum_{j>i}^{p-1} \underset{\sim}{M}_{ij}\, q^*_{ij} \qquad (5.2.4)$$

where q^*_{ij}'s are the distinct unknown elements (called auxiliary parameters) of $\underset{\sim}{Q}^*$ and $\underset{\sim}{M}_{ij}$'s are suitable symmetric matrices. [To keep notational simplicity, $\underset{\sim}{M}_{ij}$ is used here in place of $\underset{\sim}{M}^*_{ij}$ which is employed in Q^*-procedure in chap.2, sec. 2.5].

Let us illustrate the case of p equal to 5. Normal equations in terms of elements of $\underset{\sim}{Q}^*$ are given as follows.

$$\begin{aligned}
q^*_{11} + \rho_1 q^*_{12} + \rho_2 q^*_{13} + \rho_3 q^*_{14} + \rho_4 q^*_{15} &= 0 \\
\rho_1 q^*_{12} + q^*_{22} + \rho_1 q^*_{23} + \rho_2 q^*_{24} + \rho_3 q^*_{25} &= 0 \\
\rho_2 q^*_{13} + \rho_1 q^*_{23} + q^*_{33} + \rho_1 q^*_{34} + \rho_2 q^*_{35} &= 0 \\
\rho_3 q^*_{14} + \rho_2 q^*_{24} + \rho_1 q^*_{34} + q^*_{44} + \rho_1 q^*_{45} &= 0 \\
\rho_4 q^*_{15} + \rho_3 q^*_{25} + \rho_2 q^*_{35} + \rho_1 q^*_{45} + q^*_{55} &= 0 \qquad (5.2.5a)
\end{aligned}$$

and

$$q^*_{12} + q^*_{23} + q^*_{34} + q^*_{45} = q^*_{13} + q^*_{24} + q^*_{35} = q^*_{14} +$$

$$q^*_{25} = q^*_{15} = 0 \,. \qquad (5.2.5b)$$

These two sets of equations would help us to express $\underset{\sim}{Q}^*$ in the following form :

$$\underset{\sim}{Q}^* = \underset{\sim}{M}_{12}q^*_{12} + \underset{\sim}{M}_{13}q^*_{13} + \underset{\sim}{M}_{14}q^*_{14} + \underset{\sim}{M}_{23}q^*_{23} + \underset{\sim}{M}_{24}q^*_{24} + \underset{\sim}{M}_{34}q^*_{34} \qquad (5.2.6)$$

where

$$\underset{\sim}{M}_{12} = \begin{bmatrix} -\rho_1 & & & & \\ 1 & -\rho_1 & & & \\ 0 & 0 & 0 & \text{Sym} & \\ 0 & 0 & 0 & \rho_1 & \\ 0 & 0 & 0 & -1 & \rho_1 \end{bmatrix}, \quad \underset{\sim}{M}_{13} = \begin{bmatrix} -\rho_2 & & & & \\ 0 & 0 & & & \\ 1 & 0 & 0 & \text{Sym} & \\ 0 & 0 & 0 & 0 & \\ 0 & 0 & -1 & 0 & \rho_2 \end{bmatrix},$$

$$\underset{\sim}{M}_{14} = \begin{bmatrix} -\rho_3 & & & & \\ 0 & \rho_3 & & & \\ 0 & 0 & 0 & \text{Sys} & \\ 1 & 0 & 0 & -\rho_3 & \\ 0 & -1 & 0 & 0 & \rho_3 \end{bmatrix}, \quad \underset{\sim}{M}_{23} = \begin{bmatrix} 0 & & & & \\ 0 & -\rho_1 & & & \\ 0 & 1 & -\rho_1 & \text{Sym} & \\ 0 & 0 & 0 & \rho_1 & \\ 0 & 0 & 0 & -1 & \rho_1 \end{bmatrix},$$

$$\underset{\sim}{M}_{24} = \begin{bmatrix} 0 & & & & \\ 0 & -\rho_2 & & & \\ 0 & 0 & \rho_2 & \text{Sym} & \\ 0 & 1 & 0 & -\rho_2 & \\ 0 & 0 & -1 & 0 & \rho_2 \end{bmatrix}, \quad \underset{\sim}{M}_{34} = \begin{bmatrix} 0 & & & & \\ 0 & 0 & & & \\ 0 & 0 & -\rho_1 & \text{Sys} & \\ 0 & 0 & 1 & 0 & \\ 0 & 0 & 0 & -1 & \rho_1 \end{bmatrix}.$$

In general, the $\underset{\sim}{M}_{ij}$'s (i<j) of (5.2.4) are p x p symmetric matrices with their typical (u,v) elements of the form :

$$\begin{aligned} m_{ij}^{uv} &= -\rho_{j-i} \quad \text{for } (u,v) = (i,i) \text{ or } (j,j) \\ &= \rho_{j-i} \quad \text{for } (u,v) = (p-j+i, p-j+i) \text{ or } (p,p) \\ &= 1 \quad \text{for } (u,v) = (i,j) \text{ or } (j,i) \\ &= -1 \quad \text{for } (u,v) = (p-j+i,p) \text{ or } (p,p-j+i). \end{aligned} \tag{5.2.7}$$

Adopting a convention that $\rho_o = 1$ and $\rho_{-j} = \rho_j$, any t-th column of $\underset{\sim}{\rho}$-matrix and any typical (u,v) element of $\underset{\sim}{\rho}\underset{\sim}{M}_{ij}\underset{\sim}{\rho}$ matrix are respectively as follows :

$$\underline{\rho}^{(t)} = (\rho_{-(t-1)}\ \rho_{-(t-2)} \cdots \rho_{-1}\ \rho_o\ \rho_1 \cdots \rho_{p-t})', \tag{5.2.8}$$

$t = 1, 2, \cdots, p$.

$$\begin{aligned} \bar{m}_{ij}^{uv} &= tr[\underset{\sim}{M}_{ij}\ \underline{\rho}^{(v)'}\ \underline{\rho}^{(u)}] \\ &= \rho_{j-i}[- \rho_{i-u}\ \rho_{i-v} - \rho_{j-u}\rho_{j-v} + \rho_{p-u}\rho_{p-v} + \rho_{p'-u}\rho_{p'-v}] \\ &\quad + [\rho_{i-u}\ \rho_{j-v} + \rho_{j-u}\ \rho_{i-v}] - [\rho_{p-u}\rho_{p'-v} + \rho_{p'-u}\rho_{p-v}] \end{aligned} \tag{5.2.9}$$

with $p' = p-j+i$; $i = 1,2, \cdots, (p-2)$; $j = i+1, i+2, \ldots, (p-1)$; $u, v = 1,2, \cdots, p$.

5.2.1 Simplification of the " Matrix Equation"

Substituting the expression (5.2.4) of $\underset{\sim}{Q}^*$ in " matrix equation" (5.2.1), we get

$$\underset{\sim}{\rho} = \underset{\sim}{S}^{*} + \sum_{i=1}^{p-2} \sum_{j>1}^{p-1} (\underset{\sim}{\rho} \underset{\sim}{M}_{ij} \underset{\sim}{\rho}) q^{*}_{ij} \cdot \qquad (5.2.10)$$

We define

$$d_{u,u+h} = \sum_{i=1}^{p-2} \sum_{j>i}^{p-1} \underset{\sim}{m}_{ij}^{u,u+h} \; q^{*}_{ij} \qquad (5.2.11)$$

where $d_{u,u+h} = \rho_h - s^{*}_{u,u+h}$, $u = 1(1)(p-1)$, $h = 0(1)(p-u)$. For $h = 0$, from (5.2.11) we get a set of p equations, called 'diagonal set'. For $h = 1(1)(p-u)$, $u = 1(1)(p-1)$, we similarly, get a set of $\binom{p}{2}$ equations, called 'off-diagonal set'.

From the afore-said p and $\binom{p}{2}$ equations, we need to solve $\binom{p-1}{2}$ number of q^{*}_{ij}'s , p number of σ_1's and (p-1) number of ρ_j's simultaneously by an iterative procedure. But it is considerably difficult to tackle these equations. The difficulty may be reduced to some extent by applying certain special type of sum (let us call them "systematic sum") upon the coefficient matrices treating (5.2.11) as a system of non-homogeneous linear equations in q^{*}_{ij}'s .

5.2.2 Effect of "Systematic Sum"

We define the "systematic sum" given by

$$\sum_{u=\ell+r}^{u=p-h-r} d_{u,u+h} \;, \quad \ell = 1,2, \cdots, (p-2), \; r = 0,1,2, \cdots,$$

$[(p-\ell-1)/2]$ and $h = 0,1,2, \cdots, (p-\ell-2r)$. The notation $[m]$ means the greatest integer contained in m. Such type of sums, consequently upon $\underset{\sim}{m}_{ij}^{u,u+h}$ of (5.2.11) would have following special properties :

(a) $\sum_{u=\ell+r}^{u=p-h-r} \bar{m}_{\ell,\ell+t}^{u,u+h} = 0$ for $t = 1,2,3, \ldots, (p-\ell-1)$.

(b) $\sum_{u=\ell+r}^{u=p-h-r} \bar{m}_{\ell+s,\ell+t}^{u,u+h} = \sum_{u=\ell+r}^{u=p-h-r} \bar{m}_{p-t,p-s}^{u,u+h}$

for $s = 1,2, \ldots, [(p-\ell-2)/2]$ and $t = (s+1), (s+2), \ldots,$ $(p-\ell-s-1)$.

(c) In general, regardless of s and t,

$$\sum_{u=\ell+r}^{u=p-h-r} \bar{m}_{\ell+s,\ell+t}^{u,u+h} = \sum_{u=(\ell-1)+(r+1)}^{u=(p-h)-(r+1)} \bar{m}_{\ell+s,\ell+t}^{u,u+h} + \sum_{u=(\ell-1)+r}^{u=(p-h)-r} \bar{m}_{\ell+s,\ell+t}^{u,u+h} - \sum_{u=(\ell-2)+(r+1)}^{u=(p-h)-(r+1)} \bar{m}_{\ell+s,\ell+t}^{u,u+h}$$

for $0 \leq r \leq [(p-\ell-1-h)/2]$,

$$= \sum_{u=(\ell-1)+r}^{u=(p-h)-r} \bar{m}_{(\ell+s,\ell+t}^{u,u+h} - \sum_{u=(\ell-2)+(r+1)}^{u=(p-h)-(r+1)} \bar{m}_{\ell+s,\ell+t}^{u,u+h}$$

for $[(p-\ell-1-h)/2] < r \leq [(p-\ell-h)/2]$. (5.2.12)

As an outline of the proof of these properties, using (5.2.9) we express

$$\bar{m}^{u,u+h}_{\ell,\ell+t} = \rho_t[-\rho_{\ell-u}\,\rho_{\ell-u-h} - \rho_{\ell+t-u}\,\rho_{\ell+t-u-h}$$

$$+ \rho_{p-u}\rho_{p-u-h} + \rho_{p-t-u}\,\rho_{p-t-u-h}]$$

$$+[\rho_{\ell-u}\rho_{\ell+t-u-h} + \rho_{\ell+t-u}\rho_{\ell-u-h}]$$

$$-[\rho_{p-u}\rho_{p-t-u-h} + \rho_{p-t-u}\rho_{p-u-h}]. \qquad (5.2.13)$$

While summing over from $u = \ell+r$ to $u = p-h-r$, we observe that of the eight partial sums in the r.h.s., once we put $u' = (p-h-r)+(\ell+r)-u$ and utilise the notational convenience $\rho_{-i} = \rho_i$, 1st, 2nd, 5th and 6th partial sum cancel respectively the 3rd, 4th, 8th and 7th one. Hence (a) is proved.

To prove (b), without going into the details of expressions, we would only like to mention here that among eight partial sums of l.h.s. and those of r.h.s. (as obtainable using the formula (5.2.9), 3rd, 4th, 7th and 8th ones of l.h.s. coincide with those of r.h.s.. The remaining four ones of l.h.s. would be found equal to the remaining four ones of r.h.s. once we put u' as defined above.

To prove (c) is easy as it depends only upon the ranges of the summations, not on the summands.

Admissible values of s,t,h and/or r as satisfying (5.2.11) and (5.2.12) may be established with a little effort. ■

Consider now (p-2) blocks of equations on both the " diagonal set" and " off-diagonal set" based on $\ell = 1,2,\ldots,(p-2)$. On the basis of equations in (5.2.12(c)) it may be established that blocks of equations based on $\ell = 1$ and $\ell = 2$ are only linearly independent while those based on $\ell = 3,4, \ldots, (p-2)$ are linearly dependent. When h = 0, for $\ell = 1$ and $\ell = 2$, we get $[(p-1)/2] + 1 + [(p-2)/2] + 1 = p$ equations from the same p number of equations of " diagonal set". Similarly, when h > 0, for $\ell = 1$ and $\ell = 2$, we get $[p/2]\{[(p+1)/2] + [(p-1)/2]\} = p(p-1)/2$ equations of " off-diagonal set" .

From property 5.2.12(b), we are able to identify some identical columns in coefficient matrices involved in the blocks of equations. Table 5.1 gives the detailed information regarding the two blocks for $\ell = 1$ and $\ell = 2$. In matrix notation, the " diagonal set" and " off-diagonal"set of equations expressed in equations as obtainable from (5.2.11) may now be expressed as

$$\underset{p \times 1}{\underline{d}(1)} = \underset{p \times \upsilon}{\underset{\sim}{A}_1} \quad \underset{\upsilon \times 1}{\underline{q}^*} , \qquad (5.2.14)$$

$$\underset{(\frac{p}{2}) \times 1}{\underline{d}_{(2)}} = \underset{(\frac{p}{2}) \times \upsilon}{\underset{\sim}{A}_2} \quad \underset{\upsilon \times 1}{\underline{q}^*} . \qquad (5.2.15)$$

In the above-mentioned 'systematic sums', we are, in fact, premultiplying (5.2.14) and (5.2.15) by particular

Table 5.1

Details Regarding the Columns when $\ell = 1$ and $\ell = 2$

Block No.	No. of Columns	No. of Non-zero columns	No. of identical Columns	No. of independent Columns	No. of rows in diagonal set	No. of rows in off-diagonal set
$\ell = 1$	$\binom{p-1}{2}$ $(= \nu)$	$\binom{p-1}{2}-(p-2)$	$([\frac{p-1}{2}]-1)([\frac{p}{2}]-1)$	$[\frac{p-2}{2}][\frac{p-1}{2}]$ $(= \nu_1)$	$[\frac{p+1}{2}]$	$[\frac{p}{2}][\frac{p+1}{2}]$ $(= \nu_2)$
$\ell = 2$	$\binom{p-1}{2}$	$\binom{p-1}{2}-(p-3)$	$([\frac{p-2}{2}]-1)([\frac{p-1}{2}]-1)$	$[\frac{p-1}{2}][\frac{p}{2}]$ $(= \nu_3)$	$[\frac{p}{2}]$	$[\frac{p-1}{2}][\frac{p}{2}]$ $(= \nu_3)$
Total				$\binom{p-1}{2}(= \nu)$	p	$\binom{p}{2}$

type of nonsingular matrices $\underset{\sim}{P}_1(p \times p)$ and $\underset{\sim}{P}_2(\binom{p}{2} \times \binom{p}{2})$[cf. illustrative example in subsec. 5.4.1]. We now define

$$\underset{p \times 1}{\underline{h}} = \underset{\sim}{P}_1\underline{d}(1) = (\underset{\sim}{P}_1\underset{\sim}{A}_1)\underline{g}^* \qquad (5.2.16)$$

and

$$\underset{\binom{p}{2} \times 1}{\underline{g}} = \underset{\sim}{P}_2\underline{d}(2) = (\underset{\sim}{P}_2\underset{\sim}{A}_2)\underline{g}^* \quad . \qquad (5.2.17)$$

Depending upon block 1 and block 2 for $\ell = 1$ and $\ell = 2$, equations (5.2.16) and (5.2.17) may be split as

$$\underset{[(p+1)/2]x1}{\underline{h}_1} = \underset{[(p+1)/2]x\upsilon_1}{\underset{\sim}{B}_1} \; \underset{\upsilon_1 x1}{\underline{g}^*_{(1)}} \qquad (5.2.18)$$

$$\underset{[p/2]x1}{\underline{h}_2} = \underset{[p/2]x\upsilon_3}{\underset{\sim}{B}_2} \; \underset{\upsilon_3 x1}{\underline{g}^*_{(2)}} \qquad (5.2.19)$$

$$\underset{\upsilon_2 x1}{\underline{g}_1} = \underset{\upsilon_2 x\upsilon_1}{\underset{\sim}{B}_3} \; \underset{\upsilon_1 x1}{\underline{g}^*_{(1)}} \qquad (5.2.20)$$

$$\underset{\upsilon_3 x1}{\underline{g}_2} = \underset{\upsilon_3 x\upsilon_3}{\underset{\sim}{B}_4} \; \underset{\upsilon_3 x1}{\underline{g}^*_{(2)}} \qquad (5.2.21)$$

where $\underline{g}^*_{(1)}$ and $\underline{g}^*_{(2)}$ are shrunken column $\underline{g}^*$ after appropriate coalition of its components to avoid identical columns and $\underset{\sim}{B}_1$, $\underset{\sim}{B}_2$, $\underset{\sim}{B}_3$ and $\underset{\sim}{B}_4$ are appropriate coefficient matrices obtained from the r.h.s. of (5.2.16) and (5.2.17).

From Table 5.1, we find that $v_2 = v_1 + (p-1)$. We may further partition $\underset{\sim}{B}_3$ as

$$\underset{v_2 x v_1}{\underset{\sim}{B}_3} = \begin{bmatrix} \underset{\sim}{B}_{31} \\ (p-1)xv_1 \\ \underset{\sim}{B}_{32} \\ v_1 x v_1 \end{bmatrix}$$

so that $g_{(1)}$ would be similarly partitioned as

$$g_{1(1)} = \underset{\sim}{B}_{31} \underline{q}^*_{(1)} \qquad (5.2.22)$$

$$g_{1(2)} = \underset{\sim}{B}_{32} \underline{q}^*_{(1)} \qquad (5.2.23)$$

with the implication that

$$\underline{q}^*_{(1)} = \underset{\sim}{B}_{32}^{-1} \; g_{1(2)} \qquad (5.2.24)$$

and

$$g_{1(1)} = \underset{\sim}{B}_{31} \underset{\sim}{B}_{32}^{-1} \; g_{1(2)} \; . \qquad (5.2.25)$$

From (5.2.18), $\underline{h}_1 = \underset{\sim}{B}_1 \underset{\sim}{B}_{32}^{-1} g_{1(2)}$, (5.2.26)

From (5.3.21), $\underline{q}^*_{(2)} = \underset{\sim}{B}_4^{-1} g_2$. (5.2.27)

From (5.3.19), $\underline{h}_2 = \underset{\sim}{B}_2 \underset{\sim}{B}_4^{-1} g_2$. (5.2.28)

5.2.3 The Proposed Iterative Scheme

An iterative procedure may now be applied upon five equations (5.2.24) through (5.2.28). Once we get the improved estimates $\hat{g}_{1(1)}$, $\hat{\underline{h}}_1$ and $\hat{\underline{h}}_2$, we calculate the estimates

$\hat{\sigma}_i$'s and $\hat{\rho}_j$'s as follows. We write,

$$\underset{\sim}{\hat{h}}_1 = (a_1, a_2, \cdots, a_{[(p+1)/2]})', \underset{\sim}{\hat{h}}_2 = (b_1, b_2, \cdots, b_{[p/2]})'.$$

We may calculate $\hat{d}_{ii}$, i = 1(1)p as follows :

$$\hat{d}_{ii} = a_i - b_i \text{ , for } i = 1,2, \cdots, [\tfrac{p+1}{2}] \qquad (5.2.29)$$

$$= b_{p+1-i} - a_{p+2-i} \text{ , for } i = [\tfrac{p+1}{2}] + 1, \cdots, p \qquad (5.2.30)$$

adopting a convention that $a_{[p/2]+1} = b_{[\frac{p+1}{2}]} = 0$.

Knowing $\hat{d}_{ii}$ values, we calculate

$$\hat{\sigma}_i^2 = s_{ii}/(1-\hat{d}_{ii}),\ i = 1(1)p . \qquad (5.2.31)$$

Note that from (5.2.16),

$$\underset{\sim}{\hat{d}}_{(1)} = \underset{\sim}{P}_1^{-1}\, \underset{\sim}{\hat{h}} = \underset{\sim}{P}_1^{-1} \begin{bmatrix} \underset{\sim}{\hat{h}}_1 \\ \underset{\sim}{\hat{h}}_2 \end{bmatrix} .$$

For obtaining the ML estimates of ρ_j , we write,

$$\underset{\sim}{\hat{g}}_{1(1)} = (c_1\ c_2 \cdots c_{p-1})' = \underset{\sim}{R}_3\, \underset{\sim}{g}^*_{(1)}$$

and

$$\hat{\rho}_j = (c_j + \sum_{i=1}^{p-j} s^{**}_{i,i+j})/(p-j), \qquad (5.2.32)$$

$j = 1(1)(p-1)$, where $s^{**}_{i,i+j} = s_{i,i+j}/(\hat{\sigma}_i \hat{\sigma}_{i+j})$.

For the initial values of $\hat{\sigma}_j$, we take $\sqrt{s_{jj}}$, $j = 1(1)p$.

Similarly, the least squares estimate (LSE) of ρ_j defined as

$$\hat{\rho}_j^{(\ell)} = \sum_{t=1}^{p-j} r_{t,t+j}/(p-j) \qquad (5.2.32)$$

$j = 1(1)(p-1)$, may be employed as the initial value for the iterative solution of $\hat{\rho}_j$. Let $\sigma_1^{(o)}, \sigma_2^{(o)}, \ldots, \sigma_p^{(o)}$ be the initial set of values of the population standard deviation. From these, we first obtain $\rho_j^{(o)}$; $j = 1(1)(p-1)$, which may not be necessarily the same as the LSE of ρ_j from which we calculate $\underset{\sim}{d}_1^{(o)}$ vector which in turn will give a new estimate of σ_j following (5.2.31) for all $j = 1(1)p$. The iteration may be stopped at the i-th stage if $\sigma_1^{(i)}, \sigma_2^{(i)}, \ldots, \sigma_p^{(i)}, \rho_1^{(i)}, \rho_2^{(i)}, \ldots, \rho_{p-1}^{(i)}$ do not differ much from $\sigma_1^{(i+1)}, \ldots, \sigma_p^{(i+1)}, \rho_1^{(i+1)}, \ldots, \rho_{p-1}^{(i+1)}$ respectively.

5.3 PROPOSITIONS ON THE STRUCTURE OF THE ASYMPTOTIC DISPERSION MATRIX OF ML ESTIMATES

Detailed derivation of the asymptotic dispersion matrix for the MLE of the parameters of the population covariance matrix with linearly structured correlation matrix has already been discussed in Chap. 2, subsec. 2.5.2.

For the Toeplitz correlation structure, we shall only mention specific feature of the asymptotic dispersion matrix which is the inverse of the information matrix $\underset{\sim}{\mathcal{I}}(\underline{\theta})$ in the partitioned form as

$$\underset{\sim}{\mathcal{I}}(\underline{\theta}) = N \begin{bmatrix} \underset{\sim}{D}_\sigma^{-1} & \\ & \underset{\sim}{I}_{p-1} \end{bmatrix} \begin{bmatrix} \underset{\sim}{A}_{11} & \underset{\sim}{A}_{12} \\ \underset{\sim}{A}_{21} & \underset{\sim}{A}_{22} \end{bmatrix} \begin{bmatrix} \underset{\sim}{D}_\sigma^{-1} & \\ & \underset{\sim}{I}_{p-1} \end{bmatrix} . \qquad (5.3.1)$$

In eqn. (5.3.1), we define

$$\underset{\sim}{\theta} = (\sigma_1, \sigma_2, \cdots, \sigma_p; \rho_1, \rho_2, \cdots, \rho_{p-1})',$$

$$\underset{\sim}{A}_{11} = \underset{\sim}{I}_p + (\underset{\sim}{\rho} * \underset{\sim}{\rho}^{-1}),$$

$$\underset{p-1xp}{\underset{\sim}{A}_{21}} = \underset{\sim}{A}_{12}' = [\text{diag.}(\underset{\sim}{H}_t \underset{\sim}{\rho}^{-1})]_{t=1(1)(p-1)},$$

and

$$\underset{p-1xp-1}{\underset{\sim}{A}_{22}} = \frac{1}{2}[\text{tr}(\underset{\sim}{\rho}^{-1}\underset{\sim}{H}_t\underset{\sim}{\rho}^{-1}\underset{\sim}{H}_u)]_{t,u=1(1)(p-1)}$$

The notation $\underset{\sim}{A} * \underset{\sim}{B}$ means the Schur product of the matrices $\underset{\sim}{A}$ and $\underset{\sim}{B}$ of same order. Expressions of $\underset{\sim}{\rho}$, $\underset{\sim}{H}_t$'s are as defined in section 5.1.

Proposition 5.2.1 : The matrix $\underset{\sim}{\rho}^{-1}\underset{\sim}{H}_j\underset{\sim}{\rho}^{-1}$ would have the centrosymmetric structure as is the structure of the inverse of $\underset{\sim}{\rho}$-matrix (5.1.1).

Proposition 5.2.2 : Let the inverse of the following non-singular matrix $\underset{\sim}{A}$ be partitioned as :

$$\underset{\sim}{A}^{-1} = \begin{bmatrix} \underset{\sim}{A}_{11} & \underset{\sim}{A}_{12} \\ \underset{\sim}{A}_{21} & \underset{\sim}{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \underset{\sim}{A}^{11} & \underset{\sim}{A}^{12} \\ \underset{\sim}{A}^{21} & \underset{\sim}{A}^{22} \end{bmatrix}$$

when the submatrices $\underset{\sim}{A}_{11}$, $\underset{\sim}{A}_{12}$, $\underset{\sim}{A}_{21}$, $\underset{\sim}{A}_{22}$ are as defined in (5.3.1). Both $\underset{\sim}{A}_{11}$ and $\underset{\sim}{A}^{11}$ have the same centrosymmetric structure as $\underset{\sim}{\rho}^{-1}$. The submatrices $\underset{\sim}{A}_{12}$ and $\underset{\sim}{A}^{12}$ as well as $\underset{\sim}{A}_{21}$ and $\underset{\sim}{A}^{21}$ have the same structure.

5.4 A NUMERICAL ILLUSTRATION BY GRE DATA

For the purpose of illustration, we present here a 5 x 5 sample covariance matrix which appears to exhibit the basic features of a Toeplitz correlation structure. The data for this illustrative analysis were obtained from an unpublished study by Rock and Werts (undated). In this study, persons who repeated the Graduate Record Examination (GRE) of Educational Testing Service, Princeton, NJ (ETS) were segregated into four groups, i.e., those who had taken the test twice, thrice, four times and five times. Werts, Alike, Rock and Grandy (1981) attempted to fit the Markov quasi-simplex model (Jöreskog, 1970b) to these longitudinal data since the intercorrelation matrices showed a decreasing trend in the intercorrelations as one moves away from the main diagonal.

We shall consider only the data based on 217 respondents who took the GRE five times. Sample covariance matrix is reproduced from Werts et al. (1981) in the diagonal and lower diagonal of the Table 5.2, whereas its upper diagonals contain the inter-correlations. Table 5.2 clearly indicates that the Toeplitz correlation structure can be a plausible competing hypothesis so far as this set of GRE data is concerned. The correlations in Table 5.2 can be regarded as " autocorrelation" since a 'parallel' form GRE

Table 5.2

Observed Covariance and Correlation Matrix for GRE Data (from Werts et al.,1981) based on 217 Examinees of EMS

Year of Administration	Successive year of GRE Administration				
	One	Two	Three	Four	Five
One	11,008	0.8377	0.8248	0.8208	0.7781
Two	9,394	11,425	0.8662	0.8602	0.8108
Three	8,960	9,587	10,721	0.8712	0.8663
Four	8,963	9,569	9,388	10,831	0.8782
Five	8,051	8,547	8,846	9,014	9,726

was administered every time. The correlations appearing within any subdiagonal running parallel to the principal diagonal also seem to be equal. For example, the correlations appearing on the subdiagonals adjacent to the principal diagonal are approximately equal to 0.8663. Each of the third subdiagonal elements are similarly comparable.

In the case of sample covariance matrix shown in Table 5.2, the elements on the principal diagonal are approximately equal. Thus, both the Toeplitz correlational hypothesis and the Toeplitz covariance hypothesis may be tested. For the sake of comparison, in addition to these structures, we also take into consideration Quasi-Weiner

simplex (cf. Mukherjee, 1966) and intraclass covariance structure.

The analysis presented in this section is intended to illustrate the computational details of ML estimation of $(2p-1)$ parameters of the Toeplitz correlation matrix $(p = 5)$. We next apply the large sample test procedure and simple measure of goodness-of-fit to observe the plausibility of this structure compared to other three covariance structures as mentioned above, namely, intraclass, Quasi-Weiner simplex and Toeplitz covariance structures.

5.4.1 Computational Details for ML Estimation

In order to demonstrate the steps involved in obtaining the ML estimates of $(2p-1)$ parameters of the $\underset{\sim}{\Sigma}$-matrix showing Toeplitz correlation structure (5.1.1), we recall 5 x 5 case (as illustrated in sec. 5.2) which can be used for the analysis of GRE data.

Here we have $p = 5$, $v = 6$, $v_1 = 2$, $v_2 = 6$, $v_3 = 4$ and the vectors are defined as

$$\underset{5x1}{\underline{d}_1} = (d_{11}, d_{22}, d_{33}, d_{44}, d_{55})'$$

$$\underset{10x1}{\underline{d}_2} = (d_{12} \quad d_{23} \quad d_{34} \quad d_{45} \quad d_{13} \quad d_{24} \quad d_{35} \quad d_{14} \quad d_{25} \quad d_{15})'$$

and

$$\underset{6x1}{\underline{q}^*} = (q_{12}^* \quad q_{13}^* \quad q_{14}^* \quad q_{23}^* \quad q_{24}^* \quad q_{34}^*)'.$$

The nonsingular matrices $\underset{\sim}{P}_1$ and $\underset{\sim}{P}_2$ and their inverses are defined as

$$\underset{\sim}{P}_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}, \quad \underset{\sim}{P}_1^{-1} = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ 0 & -1 & 0 & 1 & 0 \end{bmatrix};$$

$$\underset{10x10}{\underset{\sim}{P}_2} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and its inverse given by

$$\underset{\sim}{P}_2^{-1} = \underset{10x10}{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}}$$

Similarly, following (5.2.11), the coefficient matrices $\underset{\sim}{A}_1$ and $\underset{\sim}{A}_2$ of (5.2.14) and (5.2.15) may be written as

$$\underset{5x6}{\underset{\sim}{A}_1} = \begin{bmatrix} \bar{m}_{12}^{11} & \bar{m}_{13}^{11} & \bar{m}_{14}^{11} & \bar{m}_{23}^{11} & \bar{m}_{24}^{11} & \bar{m}_{34}^{11} \\ \bar{m}_{12}^{22} & \bar{m}_{13}^{22} & \bar{m}_{14}^{22} & \bar{m}_{23}^{22} & \bar{m}_{24}^{22} & \bar{m}_{34}^{22} \\ \bar{m}_{12}^{33} & \bar{m}_{13}^{33} & \bar{m}_{14}^{33} & \bar{m}_{23}^{33} & \bar{m}_{24}^{33} & \bar{m}_{34}^{33} \\ \bar{m}_{12}^{44} & \bar{m}_{13}^{44} & \bar{m}_{14}^{44} & \bar{m}_{23}^{44} & \bar{m}_{24}^{44} & \bar{m}_{34}^{44} \\ \bar{m}_{12}^{55} & \bar{m}_{13}^{55} & \bar{m}_{14}^{55} & \bar{m}_{23}^{55} & \bar{m}_{24}^{55} & \bar{m}_{34}^{55} \end{bmatrix}$$

and

$$\underset{\sim}{A}_{2\ 10x6} = \begin{bmatrix} \bar{m}^{12}_{12} & \bar{m}^{12}_{13} & \bar{m}^{12}_{14} & \bar{m}^{12}_{23} & \bar{m}^{12}_{24} & \bar{m}^{12}_{34} \\ \bar{m}^{23}_{12} & \bar{m}^{23}_{13} & \bar{m}^{23}_{14} & \bar{m}^{23}_{23} & \bar{m}^{23}_{24} & \bar{m}^{23}_{34} \\ \bar{m}^{34}_{12} & \bar{m}^{34}_{13} & \bar{m}^{34}_{14} & \bar{m}^{34}_{23} & \bar{m}^{34}_{24} & \bar{m}^{34}_{34} \\ \bar{m}^{45}_{12} & \bar{m}^{45}_{13} & \bar{m}^{45}_{14} & \bar{m}^{45}_{23} & \bar{m}^{45}_{24} & \bar{m}^{45}_{34} \\ \bar{m}^{13}_{12} & \bar{m}^{13}_{13} & \bar{m}^{13}_{14} & \bar{m}^{13}_{23} & \bar{m}^{13}_{24} & \bar{m}^{13}_{34} \\ \bar{m}^{24}_{12} & \bar{m}^{24}_{13} & \bar{m}^{24}_{14} & \bar{m}^{24}_{23} & \bar{m}^{24}_{24} & \bar{m}^{24}_{34} \\ \bar{m}^{35}_{12} & \bar{m}^{35}_{13} & \bar{m}^{35}_{14} & \bar{m}^{35}_{23} & \bar{m}^{35}_{24} & \bar{m}^{35}_{34} \\ \bar{m}^{14}_{12} & \bar{m}^{14}_{13} & \bar{m}^{14}_{14} & \bar{m}^{14}_{23} & \bar{m}^{14}_{24} & \bar{m}^{14}_{34} \\ \bar{m}^{25}_{12} & \bar{m}^{25}_{13} & \bar{m}^{25}_{14} & \bar{m}^{25}_{23} & \bar{m}^{25}_{24} & \bar{m}^{25}_{34} \\ \bar{m}^{15}_{12} & \bar{m}^{15}_{13} & \bar{m}^{15}_{14} & \bar{m}^{15}_{23} & \bar{m}^{15}_{24} & \bar{m}^{15}_{34} \end{bmatrix}$$

We next obtain

$$\underset{\sim}{P}_1\ \underset{\sim}{A}_{1\ 5\ x\ 6} = \left[\begin{array}{cccccc} 0 & 0 & 0 & \sum_{1}^{5} \bar{m}^{u,u}_{23} & \sum_{1}^{5} \bar{m}^{u,u}_{24} & \sum_{1}^{5} \bar{m}^{u,u}_{34} \\ 0 & 0 & 0 & \sum_{2}^{4} \bar{m}^{u,u}_{23} & \sum_{2}^{4} \bar{m}^{u,u}_{24} & \sum_{2}^{4} \bar{m}^{u,u}_{34} \\ 0 & 0 & 0 & \bar{m}^{3,3}_{23} & \bar{m}^{3,3}_{24} & \bar{m}^{3,3}_{34} \\ \hline \sum_{2}^{5} \bar{m}^{u,u}_{12} & \sum_{2}^{5} \bar{m}^{u,u}_{13} & \sum_{2}^{5} \bar{m}^{u,u}_{14} & 0 & 0 & \sum_{2}^{5} \bar{m}^{u,u}_{34} \\ \sum_{3}^{4} \bar{m}^{u,u}_{12} & \sum_{3}^{4} \bar{m}^{u,u}_{13} & \sum_{3}^{4} \bar{m}^{u,u}_{14} & 0 & 0 & \sum_{3}^{4} \bar{m}^{u,u}_{34} \end{array}\right] \begin{array}{l} \\ \\ \\ \\ \\ \end{array} \begin{array}{l} \text{Block 1} \\ \\ \\ \text{Block 2} \\ \\ \end{array}$$

and $\underset{\sim}{P}_2\ \underset{\sim}{A}_2$ given by

$$
\underset{\sim}{P}_2\underset{\sim}{A}_2 = \underset{10\times 6}{\left[\begin{array}{cccccc}
0 & 0 & 0 & \sum_{1}^{4}\bar{m}^{u,u+1}_{23} & \sum_{1}^{4}\bar{m}^{u,u+1}_{24} & \sum_{1}^{4}\bar{m}^{u,u+1}_{34} \\
0 & 0 & 0 & \sum_{1}^{3}\bar{m}^{u,u+2}_{23} & \sum_{1}^{3}\bar{m}^{u,u+2}_{24} & \sum_{1}^{3}\bar{m}^{u,u+2}_{34} \\
0 & 0 & 0 & \sum_{1}^{2}\bar{m}^{u,u+3}_{23} & \sum_{1}^{2}\bar{m}^{u,u+3}_{24} & \sum_{1}^{2}\bar{m}^{u,u+3}_{34} \\
0 & 0 & 0 & \bar{m}^{15}_{23} & \bar{m}^{15}_{24} & \bar{m}^{15}_{34} \\
0 & 0 & 0 & \sum_{2}^{3}\bar{m}^{u,u+1}_{23} & \sum_{2}^{3}\bar{m}^{u,u+1}_{24} & \sum_{2}^{3}\bar{m}^{u,u+1}_{34} \\
0 & 0 & 0 & \bar{m}^{24}_{23} & \bar{m}^{24}_{24} & \bar{m}^{24}_{34} \\
\hline
\sum_{2}^{4}\bar{m}^{u,u+1}_{12} & \sum_{2}^{4}\bar{m}^{u,u+1}_{13} & \sum_{2}^{4}\bar{m}^{u,u+1}_{14} & 0 & 0 & \sum_{2}^{4}\bar{m}^{u,u+1}_{34} \\
\sum_{2}^{3}\bar{m}^{u,u+2}_{12} & \sum_{2}^{3}\bar{m}^{u,u+2}_{13} & \sum_{2}^{3}\bar{m}^{u,u+2}_{14} & 0 & 0 & \sum_{2}^{3}\bar{m}^{u,u+2}_{34} \\
\bar{m}^{25}_{12} & \bar{m}^{25}_{13} & \bar{m}^{25}_{14} & 0 & 0 & \bar{m}^{25}_{34} \\
\bar{m}^{34}_{12} & \bar{m}^{34}_{13} & \bar{m}^{34}_{14} & 0 & 0 & \bar{m}^{34}_{34}
\end{array}\right]}
\begin{array}{l} \\ \\ \\ \\ \\ \\ \text{Block 1} \\ \\ \\ \\ \text{Block 2} \end{array}
$$

Recall the property 5.2.12(b) of 'systematic sums'. Block 1 and Block 2 are the partitioned matrices in $\underset{\sim}{P}_1\underset{\sim}{A}_1$ or in $\underset{\sim}{P}_2\underset{\sim}{A}_2$. It may be verified that column 4 and column 6 are identical, their elements being the coefficients of q^*_{23} and q^*_{34} .

In order to apply (5.2.24) - (5.2.28), we need the following :

$$\underset{\sim}{q}^{*}_{(1)} = (q^{*}_{23} + q^{*}_{34}\ ,\ q^{*}_{24})'\ ,\ \underset{\sim}{q}^{*}_{(2)} = (q^{*}_{12},\ q^{*}_{13},\ q^{*}_{14},\ q^{*}_{34})',$$

$$\underset{\sim}{B}_{1\ 3\times 2} = \begin{bmatrix} \sum_{1}^{5} \bar{m}^{u,u}_{23} & \sum_{1}^{5} \bar{m}^{u,u}_{24} \\ \sum_{2}^{4} \bar{m}^{u,u}_{23} & \sum_{2}^{4} \bar{m}^{u,u}_{24} \\ \bar{m}^{33}_{23} & \bar{m}^{33}_{24} \end{bmatrix},$$

$$\underset{\sim}{B}_{2\ 2\times 4} = \begin{bmatrix} \sum_{2}^{5} \bar{m}^{u,u}_{12} & \sum_{2}^{5} \bar{m}^{u,u}_{13} & \sum_{2}^{5} \bar{m}^{u,u}_{14} & \sum_{2}^{5} \bar{m}^{u,u}_{34} \\ \sum_{3}^{4} \bar{m}^{u,u}_{12} & \sum_{3}^{4} \bar{m}^{u,u}_{13} & \sum_{3}^{4} \bar{m}^{u,u}_{14} & \sum_{3}^{4} \bar{m}^{u,u}_{34} \end{bmatrix}$$

$$\underset{\sim}{B}_{31\ 4\times 2} = \begin{bmatrix} \sum_{1}^{4} \bar{m}^{u,u+1}_{23} & \sum_{1}^{4} \bar{m}^{u,u+1}_{24} \\ \sum_{1}^{3} \bar{m}^{u,u+2}_{23} & \sum_{1}^{3} \bar{m}^{u,u+2}_{24} \\ \sum_{1}^{2} \bar{m}^{u,u+3}_{23} & \sum_{1}^{2} \bar{m}^{u,u+3}_{24} \\ \bar{m}^{15}_{23} & \bar{m}^{15}_{24} \end{bmatrix}, \quad \underset{\sim}{B}_{32\ 2\times 2} = \begin{bmatrix} \sum_{2}^{3} \bar{m}^{u,u+1} & \sum_{2}^{3} \bar{m}^{u,u+1}_{24} \\ \bar{m}^{24}_{23} & \bar{m}^{24}_{24} \end{bmatrix},$$

$$\underset{\sim}{B}_{4\ 4\times 4} = \begin{bmatrix} \sum_{2}^{4} \bar{m}^{u,u+1}_{12} & \sum_{2}^{4} \bar{m}^{u,u+1}_{13} & \sum_{2}^{4} \bar{m}^{u,u+1}_{14} & \sum_{2}^{4} \bar{m}^{u,u+1}_{34} \\ \sum_{2}^{3} \bar{m}^{u,u+2}_{12} & \sum_{2}^{3} \bar{m}^{u,u+2}_{13} & \sum_{2}^{3} \bar{m}^{u,u+2}_{14} & \sum_{2}^{3} \bar{m}^{u,u+2}_{34} \\ \bar{m}^{25}_{12} & \bar{m}^{25}_{13} & \bar{m}^{25}_{14} & \bar{m}^{25}_{34} \\ \bar{m}^{34}_{12} & \bar{m}^{34}_{13} & \bar{m}^{34}_{14} & \bar{m}^{34}_{34} \end{bmatrix}.$$

Using the proposed iterative procedure, the vectors $\underline{\hat{h}}_1 = (a_1, a_2, a_3)'$, $\underline{\hat{h}}_2 = (b_1, b_2)'$ and $\hat{g}_{1(1)} = (c_1, c_2, c_3, c_4)'$ are calculated by the equations (5.2.24) through (5.2.28). Lastly at each iteration, the estimates of σ_i's and ρ_j's are calculated by (5.2.31) and (5.2.32). This procedure continues until there is a satisfactory convergence in the values of each estimate.

For analysing the GRE data, we use the initial estimates as $\sigma_{i(o)} = \sqrt{s_{ii}}$, i = 1(1)5 and $\rho_{1(o)} = (r_{12} + r_{23} + r_{34} + r_{45})/4$, $\rho_{2(o)} = (r_{13} + r_{24} + r_{35})/3$, $\rho_{3(o)} = (r_{14} + r_{25})/2$, $\rho_{4(o)} = r_{15}$. It was observed in course of computations that ρ_j's converged most rapidly while σ_i's converged at a slower rate and the 'auxiliary parameters' q^*_{ij}'s at the slowest rate. Such a trend has been found with other data sets also. Table 5.3 and Table 5.4 show the successive iterated values upto 10th iteration only. More accurate estimates obtained at 16th iteration are given below:

$$\hat{\sigma}_1 = 106.5915, \quad \hat{\sigma}_2 = 107.5342, \quad \hat{\sigma}_3 = 103.3815$$

$$\hat{\sigma}_4 = 102.8762, \quad \hat{\sigma}_5 = 97.3286, \quad \hat{\rho}_1 = 0.862494$$

$$\hat{\rho}_2 = 0.849285, \quad \hat{\rho}_3 = 0.814078, \quad \hat{\rho}_4 = 0.784101$$

$$\hat{q}^*_{23} + \hat{q}^*_{34} = -0.286121, \quad \hat{q}^*_{24} = -0.199830, \quad \hat{q}^*_{12} = 0.638206 ,$$

$$\hat{q}^*_{13} = 0.646694, \quad \hat{q}^*_{14} = -0.398679, \quad \hat{q}^*_{34} = 0.212643.$$

TABLE 5.3

ML Estimates of the Parameters of Five-variate Toeplitz(Correlation) Matrix

Iteration No.	Parameters of the Population Covariance Matrix								
	σ_1	σ_2	σ_3	σ_4	σ_5	ρ_1	ρ_2	ρ_3	ρ_4
0	104.919	106.888	103.542	104.072	98.620	.86334	.85042	.81583	.77809
1	105.632	107.096	103.477	103.213	98.058	.86283	.84982	.81469	.78208
2	106.081	107.268	103.430	103.052	97.725	.86263	.84954	.81433	.78334
3	106.324	107.383	103.404	102.967	97.540	.86254	.84940	.81419	.78381
4	106.453	107.452	103.391	102.922	97.440	.86251	.84933	.81413	.78400
5	106.521	107.491	103.386	102.899	97.386	.86250	.84931	.81409	.78407
6	106.556	107.511	103.383	102.887	97.358	.86250	.84929	.81408	.78410
7	106.573	107.522	103.382	102.882	97.344	.86249	.84929	.81408	.78410
8	106.583	107.528	103.382	102.879	97.336	.86249	.84928	.81408	.78410
9	106.587	107.531	103.381	102.878	97.332	.86249	.84928	.81408	.78410
10	106.589	107.533	103.381	102.877	97.330	.86249	.84928	.81408	.78410

Table 5.4

Iterated Values of Some Elements of $\underset{\sim}{Q}^*$-matrix
required for solving normal equations for
Five-variate Toeplitz(Correlation)Matrix

Iteration	Estimates of the Elements of Q^*-Matrix					
	$q^*_{23} + q^*_{24}$	q^*_{24}	q^*_{12}	q^*_{13}	q^*_{14}	q^*_{34}
0	-	-	-	-	-	-
1	- .15053	-.13284	.30217	.50735	- .39823	.28614
2	- .22454	-.17552	.45990	.57566	- .39024	.25016
3	- .25910	-.19062	.54613	.61136	- .39200	.23255
4	- .27460	-.19629	.59136	.62936	- .39470	.22296
5	- .28136	-.19848	.61464	.63825	- .39655	.21787
6	- .28423	-.19933	.62645	.64261	- .39676	.21525
7	- .28541	-.19965	.63239	.64473	- .39814	.21393
8	- .28587	-.19978	.63535	.64575	- .39841	.21327
9	- .28605	-.19982	.63682	.64625	- .39855	.21295
10	- .28611	-.19983	.63754	.64648	- .39862	.21279

The estimates of σ_i's are correct up to three decimal places while others are correct up to five decimal places. For verification purposes, we calculate the following directly, remembering $\hat{\underset{\sim}{\Sigma}} = \hat{\underset{\sim}{D}}_\sigma \hat{\underset{\sim}{\rho}} \hat{\underset{\sim}{D}}_\sigma$ and $\hat{\underset{\sim}{Q}} = \hat{\underset{\sim}{\Sigma}}^{-1} - \hat{\underset{\sim}{\Sigma}}^{-1} \underset{\sim}{S} \hat{\underset{\sim}{\Sigma}}^{-1}$.

$$\text{diag}(\hat{\underset{\sim}{\Sigma}}^{-1} \underset{\sim}{S}) = (1.000001,\ 1.000002,\ 1.000000,\ 1.000000, 0.999999).$$

$$\hat{\underset{\sim}{Q}}^{*} = \begin{bmatrix} -.775133 & & & & \\ .638216 & -.275127 & & & \\ .646690 & -.498755 & .077067 & \text{Sym.} & \\ -.398670 & -.199842 & .212643 & .614546 & \\ -.000004 & .398687 & -.446870 & -.352092 & .358637 \end{bmatrix}.$$

Recalling (5.2.2), (5.2.5a) and (5.2.5b), normal equations are also found satisfied by the estimates as follows :

$$\text{diag}(\hat{\underset{\sim}{Q}}^{*}\hat{\underset{\sim}{\rho}}) = (-.000003,\ -.000003,\ .000001,\ .000001,\ -.000001),$$

$\hat{q}^{*}_{11} + \hat{q}^{*}_{22} + \hat{q}^{*}_{33} + \hat{q}^{*}_{44} + \hat{q}^{*}_{55} = -.000010$, $\hat{q}^{*}_{12} + \hat{q}^{*}_{23} + \hat{q}^{*}_{34} + \hat{q}^{*}_{45} = .000012$, $\hat{q}^{*}_{13} + \hat{q}^{*}_{24} + \hat{q}^{*}_{35} = -.000022$, $\hat{q}^{*}_{14} + \hat{q}^{*}_{25} = .000018$ and $\hat{q}_{15} = -.000004$.

In order to get the estimate of the standard error (SE), we replace the parameter in the information matrix (5.3.1) and get its inverse. The square roots of the diagonal elements are the estimate of the SEs of the MLE. Table 5.5 shows the entries of 9x9 symmetric matrix

Table 5.5

Entries of (9x9) matrix $\begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix}$

5.79230								
- 1.85076	6.74486							
- 1.34391	- 1.25908	7.20598						
- .44155	- 1.19347	- 1.25908	6.74486					
- .15609	- .44155	- 1.34391	- 1.85076	5.79230				
- 2.14583	- 3.60564	- 2.91963	- 3.60564	- 2.14583	116.03560			
- 1.58240	- 1.40527	- 3.16480	- 1.40527	- 1.58240	- 41.89980	117.70243		
- .54239	- .54239	0	- .54239	- .54239	- 16.98605	- 28.03344	65.41927	
- .19907	0	0	0	- .19907	4.16341	- 9.63792	-20.35097	23.05581

Sym.

Table 5.6

Entries of the (9x9) Dispersion matrix of $\hat{\underset{\sim}{\theta}}$ at the Convergence Point Expressed as

$$\frac{1}{217}\begin{bmatrix} \hat{\underset{\sim}{D}}_\sigma & \\ & \underset{\sim}{I} \end{bmatrix} \begin{bmatrix} \hat{\underset{\sim}{A}}_{11} & \hat{\underset{\sim}{A}}_{12} \\ \hat{\underset{\sim}{A}}_{21} & \hat{\underset{\sim}{A}}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\underset{\sim}{D}}_\sigma & \\ & \underset{\sim}{I} \end{bmatrix}$$

25.72								
19.43	26.35							
18.31	19.06	24.44						
17.01	18.50	18.24	24.11			Sym.		
15.09	16.24	16.72	16.97	21.44				
.0483	.0513	.0500	.0491	.0441	.00018			
.0527	.0552	.0547	.0528	.0482	.00017	.00022		
.0642	.0671	.0655	.0642	.0586	.00021	.00024	.00038	
.0736	.0763	.0748	.0730	.0672	.00023	.00029	.00041	.00065

$$\begin{bmatrix} \hat{\underset{\sim}{A}}_{11} & \hat{\underset{\sim}{A}}_{12} \\ \hat{\underset{\sim}{A}}_{21} & \hat{\underset{\sim}{A}}_{22} \end{bmatrix}$$

of (5.3.1). Finally, Table 5.6 shows the entries of the estimated dispersion matrix of the MLE s. From this we get

$$SE[\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3, \hat{\sigma}_4, \hat{\sigma}_5; \hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \hat{\rho}_4]$$

$$= (5.07, 5.13, 4.94, 4.91, 4.63; .0134, .0148, .0195, .0255).$$

It is seen that the SEs are generally quite low. As expected, they are especially very small for the correlation parameters.

5.4.2 Plausibility of Toeplitz Correlation Hypothesis for GRE Data

For testing the plausibility of different rival structural hypotheses for the covariance matrix of GRE data, we consider here as many as four structures, namely, (i) intra-class covariance structure, (ii) quasi-Weiner-simplex structure, (iii) Toeplitz covariance structure and (iv) Toeplitz correlation structure. In order to verify the tenability of these hypothesized structures, we may make use of the conventional LR test and Rao's efficient score test [vide ch. 3, subsec. 3.1.3]. Since these tests are largely influenced by sample size (N) , we may also consider a measure, called average relative difference (d) [vide ch. 3, equ. (3.2.13)] given by

$$d = \frac{1}{p^2} \sum_i^p \sum_j^p \left| \frac{s_{ij} - \hat{\sigma}_{ij}}{s_{ij}} \right| . \qquad (5.4.3)$$

This index essentially reflects the relative departure of the estimates ($\hat{\sigma}_{ij}$) from the corresponding sample covariance values (σ_{ij}) and is a measure of goodness-of-fit. [There are other indices of goodness-of-fit described, in brief, in Ch. 3 which may also be used.]

The MLEs of the four structures as mentioned above are obtained by the Q-procedure, except for the parameters of Toeplitz correlation structure which were already obtained in earlier subsection. The MLE of the parameters of Toeplitz covariance matrix has been presented in ch. 4, subsec. 4.3.6. The MLE of the covariance matrices under the first two hypotheses are respectively given by

$$\hat{\underset{\sim}{\Sigma}}_{\text{Intraclass}} = \begin{bmatrix} 10742.2 & & & & \\ 9031.9 & 10742.2 & & & \\ 9031.9 & 9031.9 & 10742.2 & \text{Sym.} & \\ 9031.9 & 9031.9 & 9031.9 & 10742.2 & \\ 9031.9 & 9031.9 & 9031.9 & 9031.9 & 10742.2 \end{bmatrix}.$$

$$\hat{\underset{\sim}{\Sigma}}_{\substack{\text{Quasi-Weiner} \\ \text{-Simplex}}} = \begin{bmatrix} 11008.0 & & & & \\ 8972.7 & 10738.2 & & & \\ 8972.7 & 9176.7 & 10657.7 & \text{Sym.} & \\ 8972.7 & 9176.7 & 9488.6 & 10937.5 & \\ 8972.7 & 9176.7 & 9488.6 & 9839.8 & 11271.2 \end{bmatrix}.$$

Based on these MLE s, LR χ^2 and Rao's efficient score χ^2 and goodness-of-fit measure (d) are calculated for each structural hypothesis and the results are shown in Table 5.7. The results clearly indicate that the Toeplitz correlation structure shows a better fit than the other three structures. A similar conclusion is reached if we use other measures of goodness of fit of the structures such as the one suggested by Bentler and Bonett (1980), Jöreskog (1978) as well as

Table 5.7

Results of Tests and Goodness-of-fit of Various Structural Hypotheses

Structure	Obs. χ^2 (LR method)	Obs. χ^2 (Rao's method)	d.f. (υ)	$\chi^2_{.05}(\upsilon)$	Average Relative Difference (d%)
Intraclass	50.18	51.41	13	22.36	4.06
Quasi-Weiner -Simplex	17.84	18.47	6	12.59	4.91
Toeplitz Covariance	18.25	18.30	10	18.31	3.08
Toeplitz Correlation	10.84	11.00	6	12.59	2.24

Jöreskog and Sörbom (1982). We note in this connection that the residuals in principal diagonal are generally the lowest when the GRE data are used to fit the Toeplitz correlation structure.

In the case of quasi-Weiner-simplex covariance structure fitted to the GRE data, both the observed χ^2-values indicate the desirability of rejecting the null hypothesis. The assertions that the GRE data are drawn from a population having either a Toeplitz covariance or a Toeplitz correlation structure seem to be the only two valid hypotheses which are accepted at the 5% level of significance. Although the Toeplitz covariance hypothesis is more parsimonious because it imposes fewer constraints, the d.f. attached to χ^2 being $p(p-1)/2$, for all criteria, the fit, however, is obviously better in the case of Toeplitz correlation structure. Both the tests based on LR and Rao's efficient score criteria reflect the relative superiority of fit of the Toeplitz correlation hypothesis. The analysis also indicates a slight trend of decrease in the scaling unit with the passage of time. Since there can not have been any change in the range of talent within the same group of examinees, the standard deviation should have been identical if the scale of measurement used is identical. Therefore the change can be ascribed to a slight shift in the scale of measurement. The near constancy of the adjacent year correlations (about 0.86), however indicate that the reliability of GRE is fairly high as well as stable.

It is worthwhile to mention in this connection that Werts et al. (1981) attempted to fit a quasi-Markov simplex

model (Jöreskog, 1970b) by pooling the GRE data of all the four groups of respondents who had taken tests twice, thrice, four times and five times, sample sizes of these four groups being 53813, 5072, 919 and 217 respectively. From the 5x5 residual matrix which they obtained (vide Werts et al. 1981, p. 302), we calculated the average relative difference (d%) and found it as large as 4.06, which is very close to the measure obtained for intraclass structure. The obtained value far exceeds the one based on the Toeplitz correlation hypothesis (vide last column of Table 5.7).

The analysis of the GRE data clearly indicates the possibility that when a particular exchangeable variable is replicated, the errors of which are varying from one replication to another, the autocorrelations may show a kind of Toeplitz structure provided the system is stable in that the errors (e_t) or the disturbances do not grow as $t \longrightarrow \infty$. Neither are the variances of such errors (disturbances) asymptotically zero as $t \longrightarrow \infty$. In such cases, the intercorrelations among the errors (residuals) will not depend on time but on time difference. If the autocorrelations between successive observations decay exponentially with time difference, then in such a case, the Toeplitz structure will emerge. If time difference is not a factor, then an intraclass structure may emerge from the intercorrelation matrix. It appears that, although in the beginning of practice

the successive trials show a decreasing variance quasi-simplex structure, with extensive practice, the correlation matrix between the trials takes the form of a Toeplitz pattern.

5.5 DISCUSSION

We essentially considered in this chapter the problem of estimation of the parameters of the spectral density matrix of the Toeplitz form. The estimation method proposed here is going to help in the analysis and testing of multiple time series and growth curves which take advantage of the Toeplitz correlation structure and hopefully provide better insights than analysis of the multivariate series by common procedures. For example, quite frequently it has been observed that the mean vector and covariance matrices of a multiple time series may not change after some planned experiment, but certain other facets of the data, such as its time dependence, does change. If we, however, rely only on MANOVA, we might falsely conclude that no change had taken place.

It is well known that the significance tests are generally more powerful if the covariance structure, which arises naturally in longitudinal data, is taken into consideration. Lillestol (1976) has considered in depth the statistical problems related to multiple time-series models with selected structured spectral density matrices such as the intraclass covariance structure and the circulant as well

as block circulant structures. The model identification, estimation, extraction and testing problems as well as prediction (Lillestol, 1982 a,b) have been studied from the frequency point of view (spectral analysis) for general classes of models with structured spectral density matrices which include the intraclass structure. Some of the results obtained by Lillestol (1976) have important bearing upon general theories of inference in multiple time-series models with linear patterns and structured spectral density matrices.

In this context, if we take the approaches for the estimation and testing time-series analysis based on finite Fourier transformation, we are led to consider inference problems in complex-normal models with a covariance matrix having a corresponding pattern such as the Toeplitz, intra-class or circulant. A theory can then be developed paralleling existing theories for the real-normal case as applied to analysis of covariance structures. As for the real case, more specific results are obtained if we study each pattern of interest separately (Lillestol, 1982a; Mukherjee, 1966; Olkin, 1973: Olkin and Press, 1969) and take advantage of its specific structures, as has been done in the present chapter.

It is to be noted that the parameters of the auto-correlation matrix are usually estimated not from the observations but from the sample covariance matrix. A number of

papers dealing with the estimation of parameters of the vector autoregressive moving average (VARMA) models have appeared in the past, e.g. Hannan (1969), Kashyap (1970), Osborn (1977) and Whittle (1963). In most of the papers, the parameters of interest are the matrix of coefficients of of the vectors of observable or unobservable random variables and the common covariance matrix of the unobservable random variables. Following Anderson (1975) and Parzen (1971) we take as our parameters the standardised autocovariance matrix (of the observable random variables) which has a Toeplitz correlation structure. In stead of the sample covariance matrix, the computations are based on the sample standard deviation and sample correlations and not on the raw data.

Although it is possible to use the LISREL procedure, we have not used this procedure of Jöreskog and Sörbom (1982) for obtaining ML estimation of the parameters of the Toeplitz correlation structure. This is because the LISREL is not recommended to analyse correlation matrices. According to Boomsma (1985), " it may have serious effects on the estimated covariance of the parameter estimates" (p. 229). Moreover, the possibility of nonconvergence and improper solution has also been reported with this general approach.

When there is no theoretical or empirical basis for specifying the structure of Σ , one can probably assume

an arbitrary p.d. $\underset{\sim}{\Sigma}$ such as in the case of general multivariate models, but unless $p \leq N$, the approach breaks down. According to Ware (1985), the moving average error structure offers an alternative to the autoregressive structure when the analyst is reluctant in specifying the exact form of covariance pattern between successive observations. If, however, the results of goodness-of-fit, assessed empirically by comparing the sample covariance matrix in groups of respondents with a common observation pattern, indicate that the fit is poor, then the estimates of the location parameters of the model equation will be inefficient and the estimators of their standard errors will be biased. In such a case, the adequacy of the Toeplitz covariance or correlation structure hypotheses can be tested. If the fit is adequate, then Aitken's generalized least squares estimators of $\underline{\beta}$ vector of regression coefficients (cf. Press, 1972, p.219) may be obtained for a given set of empirical time-series data by substituting the estimated value of $\underset{\sim}{\Sigma}$ of the Toeplitz form. Thus, a two-stage approach, involving both estimation and prediction is emphasized for the analysis of longitudinal data [Hildreth and Houck (1968)]. To obtain GLSE of the regression coefficient for Zellner's (1962) 'seemingly unrelated regression' equations, the estimation procedure (as suggested here) may be of considerable value under the assumption of a <u>constant</u> Toeplitz $\underset{\sim}{\Sigma}$-matrix from individual to individual.

5.6 SUMMARY

A simple iterative algorithm is proposed for obtaining the MLE of the parameters of Toeplitz correlation structure. A numerical example based on real life data is provided to illustrate the computations involved in the estimation and to demonstrate the feasibility of the proposed algorithm. This chapter also provides expression in explicit form for the computation of the asymptotic dispersion matrix of the ML estimates. For comparative purposes, the Toeplitz correlation hypothesis as well as other rival structural hypotheses have been tested and the results are discussed.

CHAPTER 6

SOME SUGGESTIONS FOR FUTURE RESEARCH

6.1 INTRODUCTION

We have already raised various issues and discussed different problems associated with ACOVS in the discussion-section of every chapter. In this concluding chapter, we discuss a number of useful topics for future research. These topics have received only a modest amount of attention in the ACOVS literature.

6.2 NEED FOR EXPANDING Q-PROCEDURE

Rogers and Young (1977) presented necessary and sufficient conditions for the explicit ML solutions of patterned covariance matrices. Szatrowski (1980) has also obtained necessary and sufficient conditions for the existence of explicit solutions of the patterned mean and patterned covariance matrices. He has also presented the necessary and sufficient conditions for obtaining the explicit solutions of these parameters based on one iteration of Anderson's (1973) algorithm for any positive definite starting points. Such explicit solutions greatly simplify in general the computation of the dispersion matrix of the ML estimates (Szatrowski, 1979). We need to identify a broad collection of such linearly structured covariance matrices which would have explicit ML solutions of their parameters through the application of Q-procedure. A few examples of this class of structures are

already illustrated in Mukherjee and Maiti (1984).

In case the parameters are restricted by constraints apart from the condition of positive definiteness, the application of Q-procedure needs to be suitably modified. Among many existing methods, Lee's (1980) penalty function method of constrained minimization may be helpful. Improper ML solutions leading to offending estimates of the ultimate parameters such as those involved in the Heywood case (e.g. when the estimated correlation is greater than one) or an inadmissible value of negative variance component (Harman, 1971, p. 117) may be handled with the approaches recommended by Dillon, Kumar and Mulani (1987).

We have claimed the superiority of the Q-procedure as compared to Anderson's (1973) iterative scheme for obtaining converged solution under a set of bad starting values on the basis of our repeated practical experience. As shown in Chapter 2, the Q-procedure provides standard errors of the estimates and chi-square goodness-of-fit tests. The Q-procedure also provides quickly and cheaply the consistent estimation of a large ACOVS model. However, for want of standard mathematical yardstick, we still have not been able to advance a rigorous algebraic proof of its superiority over other computational algorithms. We have also limited knowledge about the utility of Q-procedure in the case of block version of linear covariance

structures. Extensive software packages on Q - and Q^*-procedures are yet to be developed. This important project will be soon undertaken.

6.3 SOME PROBLEMS CONNECTED WITH ESTIMATION

Identifiability (cf. Bentler, 1980, p. 442-443) and estimability (cf. Browne, 1982, sec. 1.3) of the actual ultimate parameters of any covariance structures can not be readily proved particularly in the case of non-linear structures. Suitable reparametrization might be helpful to achieve linearization. Unfortunately, reparametrization may be sometimes questionable (non-unique) either due to singular nature of linear or non-linear transformations. We need proper guidelines in this respect.

The presence of non-additive terms, non-linearity, and heteroscedasticity in the disturbance terms and/or among the latent variables of the structural equation model (SEM) poses many interesting issues. It suggests the need for an analytical study of the properties of the various estimators which might be employed in ACOVS. This may help in the construction of new estimators which are efficient in finite samples. In addition to iterative estimators, shrinkage and ridge-type estimators (Vinod and Ullah, 1981) have yet to be fully investigated in this context. The theory of minimum

norm quadratic estimation (MINQUE) developed by Rao (1970; 1974, Ch. 4) can be fruitfully employed to obtain estimators with certain optimum properties. Procedures for testing hypotheses about the parameters and constructing confidence regions are yet to be developed fully in the context of ACOVS. Jöreskog and Sörbom (1988) as well as Steiger (1989) have made some beginning in this area and this must be followed-up.

In case the sample covariance matrix ($\underset{\sim}{S}$) is singular (or ill-conditioned), we do not generally rely upon the original set of variables in ACOVS. Reduction of variables is essential after proper detection of multicollinearity among the variables. The GLS method of estimation is not recommended in this connection as it involves $\underset{\sim}{S}^{-1}$ (vide R-procedure in Ch. 2). The consequences of ill-conditioned $\underset{\sim}{S}$ upon the results of Q-procedure are still not known and this is another area where analytical studies are necessary.

We also need to investigate whether estimators that are computationally easy, such as, OLS, two-stage OLS estimators or Limited Information ML estimators, are as asymptotically efficient as estimators that require greater computational efforts such as the Full Information ML estimators obtained through Q-procedure. A general approach for obtaining the score vector, the information matrix and the Cramer-Rao lower bound for the asymptotic covariance matrix of consistent estimators

of the parameters of different classes of covariance structure models needs to be worked out.

6.4 PROBLEMS CONNECTED WITH ITERATIVE SOLUTION AND CONVERGENCE

We have seen that the likelihood equations (2.2.3) emerging from the Q-procedure are non-linear in the elements of $\underset{\sim}{\Sigma}(\underline{\theta})$. So it is <u>generally</u> not possible to obtain analytical closed-form expression for the corresponding ML estimator of $\underset{\sim}{\Sigma}(\underline{\theta})$. Numerical procedures are therefore required in most cases to obtain ML estimates of the unknown parameter vector $\underline{\theta}$. However, certain computational short-cuts are possible to reduce the number of iterations required for a satisfactory convergence as has been shown for the Toeplitz correlation structure. Therefore, a general procedure needs to be worked out which will not only guarantee relatively better computational efficiency of the algorithm but also ensure that such iterative procedures do converge to yield a solution to the estimating equations. The necessary conditions for convergence as given by Oberhofer and Kmenta (1974) in connection with a general procedure for obtaining ML estimates in generalized regression models should be examined in the context of ACOVS.

The use of non-linear programming for the optimization of likelihood functions obtained for constrained covariance structures where certain restrictions on the order relations

among the parameters of the model must be satisfied as in Lee's (1980) penalty function minimization, should be also explored. In addition to Fletcher-Powell method which Jöreskog (1967) very efficiently used first for the ML exploratory factor analysis and then for ACOVS, attempts should be made to exploit other general purpose function optimization methods for this purpose. The Powell and McDonald (1972) algorithm proposed for the solution of the generalized non-linear least squares problems may be also adapted for ACOVS.

Extensive application of the Q-procedure is necessary upon the life data which are already published and analyzed by various authors. The Q-procedure may be compared with the other existing computing procedures with respect to number of iterations both under good and bad initial solutions as well as its influence upon the condition number of the sample covariance matrix.

We have an intention to conduct an extensive study on the comparative performance of the widely circulated Jöreskog-Sörbom (1988) LISREL programme vis-a-vis our Q-procedure. The two procedures may be compared with respect to the speed of convergence, efficiency in minimizing round-off errors and ease in programming. We also need to compare the role of iterations in improving the efficiency of the estimators obtained by the two procedures.

6.5 PROBLEMS CONNECTED WITH TESTS OF SIGNIFICANCE

In ACOVS, Rao's and Wald's test criteria are asymptotically equivalent (Lee, 1985). Very recently, Satorra (1989) provided a unified approach to asymptotic theory of three test criteria, e.g. LR, Rao's and Wald's. This approach is developed within a general framework that distinguishes whether or not the fit function is asymptotically optimal. We need to compare the power of the test criteria for the same sort of specified alternatives or a sequence of local alternatives as was considered by Satorra and Saris (1985) to obtain the power of the conventional LR test.

Excepting the property of invariance with respect to affine transformation among the variables, very little is known about the optimum properties of the criteria when applied to ACOVS. We have little evidence of the behaviour (e.g. robustness etc.) of these test criteria in non-normal situations. We need investigation in these areas with an emphasis on searching some structure - based non-parametric test criterion which we can apply when the population can not be specified in advance.

We also need to investigate the appropriateness of the application of classical test procedures to form test statistics for the non-linear hypotheses that are of interest in econometric and psychometric models. Godfrey and Breusch

(1981) have worked out the Lagrange Multiplier (LM) test procedure for testing the hypotheses of autoregresive and moving average disturbances in dynamic economic models. Engel (1982) provided the LM test procedure for testing of the block-recursive systems. The use of LM and other such classical tests as the likelihood ratio (LR) procedure may be investigated for testing non-linear covariance structures.

In connection with the covariance structures emerging from time-series models, such as the Toeplitz structure discussed extensively in Chapter 4, there is an urgent need to develop suitable significance tests in order to discriminate among various specifications for the disturbance process of such models. The studies reported by Anderson and de Gooijer (1980), Ansley and Newbold (1980), King (1983), Mentz (1977) as well as Nicholls and Hall (1979) may serve as a useful guide in this context.

Researchers need to consider in the above context the general problem of testing one covariance structure model against another (Mukherjee, 1976, p. 178) and the importance of developing a practical guideline for ensuring an optimal likelihood of success in a specification search, i.e. a sequential process of modifying a covariance structure model so as to improve its fit and/or parsimony which we discussed in Chapter 3 following MacCallum (1986).

The assumption of normality, linearity and additivity of the terms in any covariance structure specification has been taken for granted in this thesis. This aspect of the model's specification may be questionable in many practical situations since its violation may have an important bearing on our inference procedure. For example, if the true relationships between two latent variables of a SEM is non-linear, the present procedures of ACOVS including LISREL will provide misleading results.

Hence, we see an urgent need to address such matters in our future work as testing for normality and linearity, transformation to achieve normality and linearity, inference under alternative competing structures or rival hypotheses and development of procedures which are robust to distributional misspecifications in the context of ACOVS models. A general framework for handling some of these problems might be simply based on the assumption of the existence of the fourth or higher order moments (see Mardia, Kent and Bibby, 1979). The recent work of Browne (1984) as well as Bentler and Dijkstra (1985) may serve as a direction in this area of research.

Browne (1984) developed an estimator and goodness-of-fit test that holds when the variables have arbitrary distributions. This asymptotically distribution-free (ADF) procedure is based on a minimum chi-square (Berkson, 1980) or GLS rationale

applied to the distribution of sample covariances. More research efforts need to be directed in this important area of ACOVS. Applied researchers would greatly benefit from information concerning the relative merits and powers of distribution-free methods in ACOVS vis-a-vis the LR and other tests of significance based on multinormality assumption. The recent empirical studies examining the ADF estimator's behaviour, as those reported by Huba and Harlow (1986, 1987) as well as those of Tanaka and Huba (1987) may be helpful in addressing the robustness issue.

6.6 SOME NEEDED RESEARCH IN GOODNESS-OF-FIT MEASURES

Given a number of measures of assessing the fit of any given structure, it is unclear which of the alternatives might provide the best index of fit with regard to variation of sample size, method of estimation and model modification. La Du and Tanaka (1989) have some specific observations in this regard. We need rigorous and extensive investigation to tackle this problem. Since the indices of structural closeness (ISC_1 and ISC_2) as suggested in Chapter 3 of this thesis involve the estimated parameters both under the ML method and GLS method, these indices might have good large sample properties (behaviours) which are yet to be studied.

We have put forward a concept of 'sensitivity' of the fit indices in Chapter 3 (subsec. 3.3.8) without having

in mind the work on sensitivity analysis as reported recently by Tanaka and Odaka (1989). They considered the sensitivity analysis in terms of an algebraic function as measure of influence of a small change in the data set on the results of maximum likelihood factor analysis. The application of this algebraic function in ACOVS, in general, needs to be investigated. The loss of efficiency of the MLE, GLSE and other estimators vis-a-vis the OLS estimator of $\underset{\sim}{\Sigma}(\underline{\theta})$ due to various misspecification of the model needs to be worked out.

6.7 NON-LINEAR COVARIANCE STRUCTURES

In this dissertation, we have considered ACOVS models which are linear in both the variables and parameters, and in which the errors of measurement are additive. In practive, such a model specification may be unduly restrictive. For example, if the system of equations which we wish to estimate is derived from some underlying microeconomic optimization problem, then nonlinearity may be a feature of the model. In order to handle dynamic models or non-linear causal theories, suitable modifications are required in the ACOVS methodology. Some simple forms of non-linear relationships can often be treated in terms of algebraic transformations and renaming of the variables. However, models which are inherently non-linear such as those discussed by Gallant (1975) must be taken into consideration for this purpose.

Clearly, the proposed Q-procedure and the propositions provided in Ch. 2 and Ch. 3 would not, in general, hold in non-linear cases. Suitable modifications in this context are thus required and expected from future researchers. Development of a general framework for the analysis of covariance structures emerging from non-linear structural equations which Mukherjee (1976, p. 147) suggested as one of the emerging problems has received only scant attention to date. As such, the problem warrants more systematic research efforts especially from econometricians who work with systems of demand equations derived from the constrained maximization of utility functions where non-linearity is a reality.

6.8 SIMULATION STUDIES

Monte Carlo experiment is immensely useful in studying the sampling distribution of any statistic whatever may be the characteristics of the parent population. We need an extensive Monte Carlo experiment to investigate power behaviour, robustness of the test statistics and the bahaviour of fit indices as suggested in Chapter 3.

Monte Carlo studies are quite useful also for improving upon the asymptotic results as well as for studying the performance of estimators in finite samples. Kmenta and Gilbert (1968) carried out an interesting Monte Carlo study

of the OLSE and some other estimators of the SURE (seemingly unrelated regression) model. The analysis reported by Kmenta and Gilbert (1968) provided a number of illuminating results on the unbiasedness of the estimators. Such experiments should be conducted also in the field of ACOVS.

However, we must keep in mind that the reliability, validity and generalizability of the results of any Monte Carlo study will depend upon the design of the study and the analysis adopted. Moreover, the conclusions of such studies are based upon an empirical estimate of the sampling distribution rather than from the true sampling distribution. Thus, the Monte Carlo technique itself is subject to a sampling error, the seriousness of which can not be always assessed or easily controlled. As a consequence, any Monte Carlo experiment of the bias, mean square error etc. may sometimes lead to erroneous conclusious. These facts must be kept in mind when using Monte Carlo studies.

6.9 NEEDED RESEARCH IN SPECIFIC STRUCTURES

When the structure of the covariance matrix is clearly specified, the statistical features based on the measures of association, principal component analysis, canonical analysis etc. may be easily studied. The results may be found to be very interesting and may, in turn, help us to identify the

covariance structure when it is unknown. Following the spirit of Mukherjee (1982) in the uniform covariance structure case, we have obtained some ad hoc indicators (or checking procedure) regarding the Toeplitz covariance structure. There are various other structures e.g. circumplex, centro-symmetric etc. which may be studied in the same manner. Devising techniques to identify appropriate covariance structures for the disturbance terms in time-series data is another equally potential area for future research.

In multiple time series models, Lillestol (1976) has studied spectral density matrices having structures such as spherical, intraclass, circular symmetric etc. in a complex multinormal set-up. Spectral density matrices having Toeplitz or other such pattern may be studied along the lines of Lellestol (1976).

6.10 EXTENSION OF ACOVS

The whole discussion in this thesis is limited to problems of a sample drawn from a single population having non-structured mean vector. ACOVS may be extended to the case of samples across different populations with or without structured mean vector. These are some problems left to future researchers. The need for extending the procedures of ACOVS for studying moment structures (Bentler, 1983) is also emphasized.

By treating latent variables as incomplete data, as has been done by Kiiveri (1987), following Dempster (1972) we can explore the possibility of improving the EM algorithm of Dempster, Laird and Rubin (1977) further for different types of more complicated causal models.

6.11 SUMMARY

Some suggestions for future research have been outlined in this chapter. The need for developing asymptotic theory in connection with the estimation and testing under a distribution-free framework is emphasized. Applied researchers using structural equation models are likely to gain enormously if simple methods of specification search in ACOVS are developed. The development of new measures of goodness-of-fit and/or parsimony is another potential area of research in ACOVS. Many such research problems are discussed in this chapter.

REFERENCES

Aigner, D.J. and Goldberger, A.S. (1977) (Eds.). Latent Variables in Socio-Economic Models. Amsterdam : North-Halland.

Aitchison, J., and Silvey, D. (1958). Maximum likelihood estimation of parameters subject to restraints. Annals of Mathematical Statistics, 29, 813-828.

Aitken, A.C. (1967). Determinants and matrices, Edinburgh : Oliver and Boyed.

Aitkin, M.A., Nelson, W.C. and Reinfurt, K.H. (1968). Tests for Correlation Matrices, Biometrika, 55, 327-334.

Aitkinson, K.E. (1978). An Introduction to Numerical Analysis. New York : Wiley.

Akaike, H. (1973). Block Toeplitz Matrix Inversion, SIAM Jour.of Applied Mathematics, 24, 234-241.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19, 716-723.

Anderson, J.C. and Gerbing, D.W. (1984). The effect of sampling error on convergence, improper solutions and goodness of fit indices for maximum likelihood confirmatory factor analysis. Psychometrika, 49, 155-173.

Anderson, J.C. and Gerbing, D.W. (1988). Structural equation modeling in practice : A review and recommended two-step approach. Psychological Bulletin, 103, 411-423.

Anderson, O.D. and de Gooijer, J.G. (1980). Distinguishing between IMA (1,1) and ARMA (1,1) models : a large scale simulation study of two particular Box-Jenkins time processes. In O.D. Anderson (Ed.), Time series, p. 15-40, Amsterdam : North-Holland.

Anderson, T.W. (1960). Some stochastic process models for intelligence test scores. In K.J. Arrow et al.(Eds.), Mathematical Methods in the Social Sciences, Stanford University Press, Standord, California.

Anderson, T.W. (1962). The choice of the degree of a polynomial regression as a multiple decision problem, Annals of Mathematical Statistics, 33, 255-265.

Anderson, T.W. (1969). Statistical inference for covariance with linear structure. In P.R. Krishnaiah (Ed.), Multivariate Analysis II, New York : Academic Press, p.55-66.

Anderson, T.W. (1970). Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices. In R.C. Bose, I.M. Chakrabarti, P.C. Mahalanobis, C.R. Rao and K.J.C. Smith (Eds.) Essays in Probability and Statistics, Chapel Hill, N.C. : Univ. of North Carolina Press, p. 1-24.

Anderson, T.W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. Annals of Statistics, 1, 135-141.

Anderson, T.W. (1975). Maximum likelihood estimation of parameters of autoregressive processes with moving average residuals and other covariance matrices with linear structure. Annals of Statistics, 3, 1283-1304.

Anderson, T.W. (1984). An Introduction to Multivariate Statistical Analysis. 2nd Ed., New York : Wiley.

Anderson, T.W. and Das Gupta, S. (1964). Monotonicity of the power function of some tests of independence between two sets of variates. Annals of Mathematical Statistics, 35, 206-207.

Andrew, A.L. (1973). Eigenvectors of certain Matrices. Linear Algebra and Its Application, 7, 151-162.

Ansley, C.F. and Newbold, P. (1980). Finite sample properties of estimators for autoregressive moving average models. Journal of Econometrics, 13, 159-183.

Armstrong, J.S. and Soelberg, P. (1968). On the interpretation of factor analysis. Psychological Bulletin, 70, 361-364.

Bagozzi, R.P. (1980). Causal Models in Marketing. New York, Wiley.

Bagozzi, R.P., Fornell, C. and Larcker, D.F. (1981). Canonical Correlation analysis as a special case of a structural relation model. Multivariate Behavioural Research, 16, 437-454.

Bargmann, R.E. (1957). A study of Independence and Dependence in Multivariate Normal Analysis. Institute of Statistics Mimeograph Series No. 186, University of North Carolina, Chapel Hill, N.C.

Barnett, V.D. (1966). Evaluation of the maximum - likelihood estimators where the likelihood equation has multiple roots, Biometrika, 53, 151-165.

Bartlett, M.S. (1950). Test of significance in factor analysis. British Journal of Psychology, 3, 77-85.

Bearden, W.O., Sharma, S., and Teel, J.E. (1982). Sample size effects on chi-square and other statistics used in evaluating causal models, Journal of Marketing Research, 19, 425-430.

Bellman, R. (1974). Introduction to Matrix Analysis, 2nd Ed., New Delhi : Tata McGraw Hill.

Ben-Artzi, A. and Shalom, T. (1986). On inversion of Toeplitz and close to Toeplitz matrices. Linear Algebra and its Applications, 75, 173-192.

Bentler, P.M. (1973). Assessment of developmental factor change at the individual and group level. In Life-Span Developmental Psychology : Methodological Issues edited by J.R. Nesselroade and H.W. Ruse, New York; Academic Press, p. 145-174.

Bentler, P.M. (1976). Multivariate statistical models applied to factor analysis. Multivariate Behavioural Research, 11, 3-25.

Bentler, P.M. (1980). Multivariate analysis with latent variables : Causal modeling. Annual Review of Psychology, 31, 419-456.

Bentler, P.M. (1983). Simultaneous equation systems as moment structure models. Journal of Econometrics, 22, 13-42.

Bentler, P.M. (1986). Structural modeling and Psychometrika : An historic perspective on growth and achievements. Psychometrika, 5, 35-51.

Bentler, P.M. (1989). EQS Structural Equations Program Manual. Los Angeles, CA : BMDP statistical software.

Bentler, P.M. and Bonett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-506.

Bentler, P.M. and Dijkstra, T. (1985). Efficient estimation via linearization in structural models. In P.R. Krishnaiah (Ed.), Multivariate Analysis VI, Amsterdam : North-Holland, p. 9-42.

Bentler, P.M. and Mooijaart, A. (1989). Choice of structural model via Parsimony : A rationale based on Precision. Psychological Bulletin, 106, 315-317.

Bentler, P.M. and Speckart, G. (1981). Attitudes " cause" behaviors : A structural equation analysis. Journal of Personality and Social Psychology, 40, 226-238.

Bentler, P.M. and Weeks, D.G. (1979). Interrelations among models for the analysis of moment structures. Multivariate Behavioural Research, 14, 169-186.

Bentler, P.M., and Weeks, D.G. (1980). Linear structural equations with latent variables. Psychometrika, 45, 289-308.

Bereis, E.H. (1969). Numerical solution of linear equation with Toeplitz and vector Toeplitz matrices. Numerical Mathematics, 13, 404-424.

Berkson, J. (1980). Minimum Chi-square, not maximum likelihood. Annals of Statistics, 8, 457-487 (with discussion).

Berndt, E.R., and Savin, N.E. (1977). Conflict among criteria for testing likelihood in the multivariate regression model. Econometrica, 45, 1263 - 1278.

Bhargava, A.K. (1977). Maximum likelihood estimation in a multivariate 'errors in variables' regression model with unknown error covariance matrix. Communications in Statistics, A6 , 587-601.

Bielby, W.T. and Hauser, R.M. (1977). Structural equation models. Annual Review of Sociology, 3, 137-161.

Bilodeau, E.A. (1960). Unpublished data obtained in connection with the study on "Prediction of complex task proficiency by means of component responses". Perceptual and Motor Skills, 1961, Vol.12, p. 299-306, Tulane University, New Orleans, Louisiana.

Bini, D. and Pan, V. (1988). Efficient algorithms for the evaluation of the eigenvalues of (block) banded Toeplitz matrices. Mathematics of Computation, 50, 431-448.

Blackith, R.E. and Reyment, R.A. (1971). Multivariate Morphometrics. London : Academic.

Blalock, H.M. (1971) (Ed.). Causal Models in the Social Sciences. Chicago : Aldine-Atherton.

Blalock, H.M., Jr. (1985). Causal Models in Panel and Experimental Disigns. New York : Aldine.

Bock, R.D. (1960). Component of variance analysis as a structural and discriminal analysis of psychological tests. British Journal of Statistical Psychology, 13, 151-163.

Bock, R.D. (1963). Multivariate analysis of repeated measurements. In C. Harris (Ed.). Problems of Measuring Change, Madison : Univ. of Wisconsin Press, p. 85-103.

Bock, R.D. (1975). Multivariate Statistical Methods in Behavioural Research. New York : McGraw Hill.

Bock, R.D. and Bargmann, R.E. (1966). Analysis of covariance structures. Psychometrika, 31, 507-534.

Bock, R.D., Dickens, C. and van Pelt, J. (1969). Methodological implications of content-acquiescence correlation in the MMPI. Psychological Bulletin, 71, 127-139.

Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. In K.G. Joreskog and H. Wold (Eds.), Systems Under Indirect Observation : Causality, Structure, Prediction (Part I), 149-173; Amsterdam : North-Holland.

Boomsma, A. (1985). Non-convergence, improper solutions and starting values in LISREL maximum likelihood estimation. Psychometrika, 50, 229-242.

Bottcher, A. and Silbermann, B. (1983). Invertibility and Asymptotics of Toeplitz Matrices. Mathematical Research Band 17, Berlin : Akademic-Verlag.

Box, G.E.P. and Muller, M. (1958). A note on the generation of random normal deviates. Annals of Mathematical Statistics, 29, 610.

Bramble, W.J. and Wiley, D.E. (1974). Estimating content-acquiescence correlation by covariance structure analysis. Multivariate Behavioural Research, 9, 179-190.

Breusch, T.S. (1979). Conflict among criteria for testing hypothesis : Extension and comment. Econometrica, 47, 203-208.

Breusch, T.S. and Pagan, A.R. (1980). The Lagrange multiplier test and its application to model specification in econometrics. Review of Economic Studies, 47, 239-254.

Brown, K.M. (1966). A quadratically convergent method for solving simulteneous nonlinear equations. Ph.D. Thesis, Dept. of Computer Sciences, Purdue Univ., Lafayette, Indiana.

Browne, M.W. (1974). Generalized least squares estimators in the analysis of covariance structures. South African Statistical Journal, 8, 1-24.

Browne, M.W. (1977). The analysis of patterned correlation matrices by generalized least squares. British Journal of Mathematical and Statistical Psychology, 30, 113-124.

Browne, M.W. (1982). Covariance structures. In D.M. Hawkins (Ed.), Topics in Applied Multivariate Analysis, Cambridge : Cambridge University Press, p. 72-141.

Browne, M.W. (1984). Asymptotic distribution free method, for analysis of covariance structures. British Journal of Mathematical and Statistical Psychology, 37, 62-83.

Browne, M.W. and Cudeck, R. (1983). BENWEE (Version 2) : The Computer Programme for Path Analysis with Latent Variables. Pretoria, South Africa : Human Sciences Research Council Technical Report.

Browne, M.W. and Cudeck, R. (1989). Single sample cross-validation indices for covariance structures. Multivariate Behavioral Research, 24,

Browne, M.W. and de Toit, S.H.C. (1987). Automated fitting of nonstandard models. Report No. WS-39. Pretoria, South Africa : Human Science Research Council.

Browne, M.W. and Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. British Journal of Mathematical and Statistical Psychology, 41, 193-208.

Bube, K.P. and Burridge, R. (1983). The one-dimensional inverse problem of reflection seismology. SIAM Review, 25, 497-559.

Burg, J.P., Luenberger, D.G. and Wenger, D.L. (1982). Estimation of structured covariance matrices. Proceedings of IEEE , 70, 963-974.

Burt, C. (1940). The Factors of the Mind. London : University of London Press.

Burt, C. (1947). A comparison of factor analysis and analysis of variance. British Journal of Psychology : Statistics Section, 1, 3-26.

Caflisch, R.E. (1981). An inverse problem for Toeplitz matrices and the synthesis of discrete transmission lines. Linear Algebra and Applications, 38, 207-225.

Calderion, A., Spitzer, E. and Widom, H. (1959). Inversion of Toeplitz matrices. Illinois Journal of Mathematics, 3, 490-498.

Cantoni, A., and Butler, P. (1976). Properties of the eigenvectors of presymmetric matrices with applications to communication theory. IEEE Transactions in Communication, Vol. COM-24, 804-809.

Chatterjee, S. and Price, B. (1977). Regression Analysis By Example. New York, Wiley.

Choi, S.C. and Wette, R. (1972). A test for the homogeneity of variances among correlated variables, Biometrics, 28, 589-592.

Coovert, M.D., Penner, L.A. and MacCallum, R. (1990). Covariance structure modeling in personality and social psychological research : A practical guide. In C. Hendrick and M.S. Clark (Eds.). Review of Personality and Social Psychology (Vol. 11), Newbury Park, CA: Sage.

Corballis, M.C. and Traub, R.E. (1970). Longitudinal factor analysis. Psychometrika, 35, 79-93.

Cover, T.M. and Thomas, J.A. (1988). Determinant inequalities via information theory. SIAM Journal of Matrix Analysis and Application, 9, 384-392.

Cudeck, R. (1986). A note on structural models for the circumplex. Psychometrika, 51, 143-147.

Cudeck, R. (1988). Multiplicative models and MTMM matrices. Journal of Educational Statistics, 13, 131-147.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. Psychological Bulletin, 105, 317-327.

Cudeck, R. and Browne, M.W. (1983). Cross-validation of covariance structures. Multivariate Behavioural Research, 18, 147-167.

Cybenko, G. (1984). On the eigenstructure of Toeplitz matrices. IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. Assp, 32, 918-921.

Cybenko, G. and Van Loan, C.F. (1984). Computing the minimum eigenvalue of a symmetric positive definite Toeplitz matrix. TR 82-527, April, 1984, Dept. of Computer Science, Cornell University, Ithaca, New York.

Day, K.M. (1975). Toeplitz matrices generated by Laurent series expansion of an arbitrary rational function. Transaction of American Mathematical Society, 206, 224-245.

Dempster, A.P. (1966). Estimation in multivariate analysis. In P.R. Krishnaiah (Ed.). Multivariate Analysis. Academic Press, New York, p. 315-334.

Dempster, A.P. (1972). Covariance selection. Biometrics, 28, 157-175.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of Royal Statistical Society, Ser. B, 19, 1-18.

Dewilde, P., Vieira, A.C., and Kailath, T. (1978). On a generalized Szego-Lovinson algorithm for optimal linear predictors based on a network synthesis approach. IEEE Transactions on Circuits System, CAS-25, 663-675.

Dickinson, B.W. (1980). Solution of linear equations with rational Toeplitz matrices. Mathematics for Computation, 34, p. 227-233.

Dickinson, B.W. (1984). An inverse problem for Toeplitz matrices. Linear Algebra and Its Applications, 59, 79-83.

Dillon, W.R., Kumar, A., and Mulani, N. (1987). Offending estimates in covariance structure analysis : Comments on the causes of and solutions to Heywood Cases. Psychological Bulletin, 101, 126-135.

Duncan, O.D. (1975). Introduction to Structural Equation Models. New York : Academic.

Durbin, J. (1960). The fitting of time series models. Review of International Institute of Statistics, 28, 233-243.

Eaton, M.L. (1983). Multivariate Statistics : A vector space approach. New York : Wiley.

Eaves, L.J., Martin, N.G. and Eysenck, S.B.G. (1977). An application of the analysis of covariance structure to the psychogenetical study of impulsiveness. British Journal of Mathematical and Statistical Psychology, 30, 185-197.

Engel, R.F. (1982). A general approach to Lagrange multiplier model diagnostics. Journal of Econometrics, 20, 83-104.

Evans, G.B.A., and Savin, N.E. (1982). Conflict among the criteria revisited: The W, LR and LM tests. Econometrica, 50, 737-748.

Everitt, B.S. (1984). An Introduction to Latent Variable Models. New York : Chapman and Hall.

Fiedler, M. (1984). Quasidirect decomposition of Hankel and Toeplitz matrices. Linear Algebra and its applications, 61, 155-174.

Fisher, F.M. (1966). The Identification Problem in Economics, New York : McGraw-Hill.

Fiske, S.T., Kenny, D.A. and Taylor, S.E. (1982). Structural models for the mediation of salience effects on attribution. Journal of Experimental Social Psychology, 18, 105-127.

Fliess, J. L. (1973). Statistical Methods for Rates and Proportions. London : Wiley.

Fletcher, R. and Powell, M.J.D. (1963). A rapidly convergent descent method for minimization. Computer Journal, 6, 163-168.

Foa, U.G. (1965). New Developments in facet design and analysis. Psychological Review, 72, 262-274.

Folmer, H. (1980). Measurement of the effects of regional policy instruments. Environment and Planning, A, 12, 1191-1202.

Folmer, H. (1981). Measurement of the effects of regional policy instruments by means of linear structural equation models and panel data. Environment and Planning, A, 13, 1435-1448.

Folmer, H. (1983). Measurement of Effects of Regional Economic Policy : Some Methodological Aspects. Ph.D. Thesis. Department of Geography, University of Groningen.

Folmer, H. and Nijkamp, D. (1983). Linear structural equation models with spatio temporal autocorrelation. In G. Behrenberg and M. M. Fischer (Eds.) Theorie and Quantative Methodik in der Geographie. Bremer Beitrage rue Geographie and Raum-planning, 4, University of Bremen, Bremen.

Franklin, J.A. (1968). Matrix Theory, Englewood Cliffs, Prentice-Hall, Inc.

Fraser, C. (1982). COSAN : A Fortran program for Covariance Structure Analysis of the nth order. Armidale, New South Wales, Australia : University of New England, Centre for Behavioural Studies.

Friedlander, B., Morf, M., Kailath, T. and Ljung, L. (1979). New inversion formulas for matrices in terms of their distance from Toeplitz matrices. Linear Algebra and its Applications, 27, 31-60.

Frobenius, G. (1884). Über das Tragheitsgesetz der quardratischen Formem. Sitzungsber. Konigl, Preuss. Akad. Wiss., p. 241-256. [cited in Iohvidov, I.S. (1982)]

Frobenius, G. (1912). Ableitung eines Satzes von Carothodory aus einer Formal von Kronecker. Sitzungsber, Konigl, Preuss. Akad. Wiss., p. 16-31 [cited in Iohvidov, I.S. (1982)].

Gallant, A.R. (1975). Seemingly unrelated non-linear regressions. Journal of Econometrics, 3, 35-50.

Gantmacher, F.R. (1959). Applications of the Theory of Matrices. Interscience, New York.

Geisser, S. (1963). Multivariate analysis of variance for a special covariance case. Journal of American Statistical Association, 58, 660-669.

Geisser, S. (1964). Estimation in the Uniform Covariance Case. Journal of the Royal Statistical Society, Ser. B., 26, 477-483.

Geraci, V.J. (1977). Estimation of simulteneous equation models with measurement errors. Econometrica, 45, 1243-1255.

Ghosh, S. (1987). Structural analysis of the variance-covariance matrices of observations under multidimensional block and 2^m factorial designs. Journal of the Indian Statistical Association, 25, 29-36.

Giordano, A.A. and Hsu, F.M. (1985). Least squares estimation with applications to digital signal processing. New York, John Wiley.

Glass, G.V., Wilson, V.L. and Gottman, J.M. (1975). Design and Analysis of Time-series Experiments. Boulder, Co : Colorado Associated University Press.

Godfrey, L.G. and Breusch, T.D. (1981). A review of recent work on testing for auto correlation in dynamic economic models. In D. Currie, R. Noby and D. Pell. (Eds.), Macro-economic Analysis, Croom Helm : Backemham, p. 63-105.

Goldberger, A.S. (1973). Efficient estimation in over identified models : An interpretive analysis. In A.S. Goldberger and O.D. Duncan (Eds.), Structural equation models in the social sciences. New York : Seminar Press.

Goldberger, A.S. (1974). Unobservable variables in econometrics. In P. Zarambka (Ed.), Frontiers of Econometrics. New York : Academic Press, 193-213.

Goldberger, A.S. and Duncan, O.D. (1973) (Eds.). Structural Equation Models in the Social Sciences. New York : Seminar Press.

Gohberg, I.C. and Feldman, I.A. (1974). Convolution Equations and Projection Methods of their solutions. Transl. Math. Monographs, Vol. 41 , American Mathematical Society, Providence.

Gohberg, I.C. and Semencul, A.A. (1972). On the inversion of finite Toeplitz matrices and their continuous analogs (in Russian), Mat. Issled. 7(2), 201-223.

Gottman, J.M. and Glass, G.V. (1978). Analysis of interrupted time series experiments. In T.R. Krotochwill (Ed.), Single subject research strategies for evaluation of change. New York : Academic Press.

Graham, A. (1981). Kronecker Products and Matrix Calculus with Applications. West Sassex, England : Ellis Horwood Ltd.

Graybill, F.A. (1961). An Introduction to Linear Statistical Models. New York, McGraw-Hill.

Graybill, F.A. (1969). Introduction to Matrices with Applications in Statistics. Bokmont, Calif. : Wadsworth.

Greenberg, B.G. and Sarham, A.E. (1959). Matrix inversions, its interest and application in analysis of data. Journal of American Statistical Association, 54, 755-766.

Greenberg, B.G. and Sarhan, A.E. (1960). Generalization of some results for inversion of partitioned matrices. In I. Olkin et al. (Eds.). Contributions to Probability and Statistics: Essays in Honour of Harold Hotelling. Stanford University Press, Stanford, Calif., p. 216-223.

Grenander, U. and Szego, G. (1958). Toeplitz Forms and Their Applications, University of California Press, Berkeley, Calif.

Greville, T.N.E. (1983). Toeplitz matrices with Toeplitz inverses revisited. Linear Algebra and Its Applications, 55, 87-92.

Greville, T.N.E., and Trench, W.F. (1979). Band matrices with Toeplitz inverses. Linear Algebra and Its Applications, 27, 199-209.

Grunbaum, F.A. (1981a). Toeplitz matrices commutating into tridiagonal matrices. Linear Algebra and Its Applications, 40, 25-36.

Grunbaum, F.A. (1981b). Eigen vectors of a Toeplitz matrix : Discrete version of prolate spheroidal wave functions. SIAM Jour. Algebraic Discrete Methods, 2, 136-141.

Guilford, J.P. (1956). The structure of intellect. Psychological Bulletin, 53, 267-293.

Gulliksen, H. (1950). Theory of Mental Tests. New York: Wiley.

Guttman, L. (1954). A new approach to factor analysis: The radex. In P. Lazarsfeld (Ed.), Mathematical thinking in the social sciences. Glencoe, IL : Free Press, p. 258-348.

Guttman, L. (1957). Empirical verification of the radex structure of mental abilities and personality traits. Educational and Psychological Measurements, 17, 391-407.

Han, C.P. (1968). Testing the homogeneity of a set of correlated variances. Biometrika, 55, 317-26.

Hannon, R.J. (1969). The estimation of mixed moving average autoregressive systems. Biometrika, 56, 579-593.

Harman, H.H. (1971). Modern Factor Analysis. (Second Edn.) Chicago : University of Chicago Press.

Harris, P. (1985). Testing for variance homogeneity of correlated variables. Biometrika, 72, 103-107.

Hauser, R.M. and Goldberger, A.S. (1971). The treatment of unobserable variables in path analysis. In H.L. Costner (Ed.), Sociological Methodology, 1971, San Francisco, Calif. : Jossey-Bass Inc., p. 81-117.

Hausman, J.A. (1977). Errors in variables in simulteneous equation models. Journal of Econometrics, 5, 389-401.

Haynsworth, E.V. (1968). Determination of the inertia of a partitioned Hermitian matrix. Linear Algebra and Their Applications, 1, 73-81.

Heinig , G. and Rost, K. (1984). Algebraic Methods for Toeplitz-like Matrices and Operators, in Mathematical Research, Band 19, Berlin : Akademie-Verlag.

Hewett, J.E. and Spurrier, J. B. (1983). A survey of two-stage tests of hypotheses : Theory and application. Communications in Statistics : Theory and Methods, 12, 2307-2425.

Hildreth, C. and Houck, J.P. (1968). Some estimations for a linear model with random coefficients. Journal of American Statistical Association, 63, 584-595.

Hoelter, J.W. (1983). The analysis of covariance structures : Goodness of fit indices. Sociological Methods and Research, 11, 325-344.

Hoffman, A.J. (1965). On the nonsingularity of real matrices. Mathematics of computation, 19, 56-61.

Hogg, R.V. (1962). Iterated tests of the equality of several distributions. Journal of the American Statistical Association, 57, 579-585.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24, 417-441, 498-520.

Howe, W.G. (1955). Some contributions to Factor Analysis. Unpublished Ph.D. thesis, Univ. of North Carolina, Chapel Hill. Also appeared as Report No. ORNL - 1919, Oak Ridge, Tenn : Oak Ridge National Laboratory, 1955.

Hsiao, C. (1976). Identification and estimation of simulteneous equation models with measurement error. International Economic Review, 17, 319-339.

Huang, N.M. and Cline, R.E. (1972). Inversion of persymmetric matrices having Toeplitz inverses. Journal of Association of Computational Mechanics, 19, 437-444.

Huba, G.J. and Harlow, L.L. (1986). Robust estimation for causal models : A comparison of methods in some developmental datasets. In P.B. Baltes, D.L. Featherman, and R.M. Lerner (Eds.), Life span development and behaviour, (vol. 6, p. 69-111). Hillsdale, NJ : Erlbaum.

Huba, G.J. and Harlow, L.L. (1987). Robust structural equation models : Implications for developmental psychology. Child Development, 58, 147-166.

Ikramov, H.D. (1983). Linear Algebra (English Translation), Moscow : Mir Publishers.

Intriligator, M.D. (1978). Econometric Models, Techniques and Applications. Englewood cliffs, N.J. : Prentice Hall.

Iohvidov, I.S. (1982). Hankel and Toeplitz Matrices and Forms (Algebraic Theory) (translated by G. Philips and A. Thijse), Birkhauser, Boston.

Jackson, J.E. and Morris, R.H. (1957). An application of multivariate quality control to photographic processing. Journal of the American Statistical Association, 52, 186-199.

Jacobson, N. (1968). Structure and Representation of Jordan Algebra. American Mathematical Society, Providence.

James, L.R., Mulaik, S.A. and Brett, J.M. (1982). Causal Analysis : Assumptions, Models and Data. Beverly Hills : Sage.

Jensen, S.T. (1975). Covariance hypotheses which are linear in both the covariance and the inverse covariance. Preprint January 1975, No. 1, Institute of Mathematical Statistics, University of Copenhagen, Denmark.

Jensen, S.T. (1977). An algebraic theory for normal statistical models with unknown covariance. In J.R. Barra et al. (Eds). Recent Developments in Statistics, Amsderdam, North-Halland, 285-288.

Joliceur, P. and Mosimann, J.E. (1960). Size and shape variation in the painted turtle, a principal component analysis. Growth, 24, 339-354.

Jones, M.B. (1960). Molar Correlation Analysis. U.S. Naval School of Aviation Medicine (Monograph 4), Pensacola, Florida.

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, 230-25.

Jöreskog, K.G. (1970b). Estimation and testing of simplex models. British Journal of Mathematical and Statistical Psychology, 23, 121-145.

Jöreskog, K.G. (1971). Simulteneous factor analysis in several populations. Psychometrika, 36, 409-426.

Jöreskog, K.G. (1973a). A general method for estimating a linear structural equation system. In A.S. Goldberger and O.D. Duncan (Eds.), Structural Equation Models in the Social Sciences. New York : Seminar Press, p. 1-18.

Jöreskog, K.G. (1973b). Analysis of covariance structures. In P.R. Krishnaiah (Ed.), Multivariate Analysis - III. New York : Academic Press, p. 263-285.

Jöreskog, K.G. (1974). Analyzing psychological data by structural analysis of covariance matrices. In D.H. Krantz, R.C. Atkinson, R.D. Luce and P. Suppes (Eds.). Contemporary Developments in Mathematical Psychology : Measurement, Psychophysics and Neural Information Processing. (vol. 2). San Fransisco : Freeman, p. 1-56.

Jöreskog, K.G. (1977). Structural equation models in the Social Sciences : Specification, estimation and testing. In P.R. Krishnaiah, (Ed.) Applications of Statistics, Amsterdam : North-Holland, p. 265-287.

Jöreskog, K.G. (1978). Structural analysis of covariance and correlation matrices. Psychometrika, 43, 443-477.

Jöreskog, K.G. (1981). Analysis of covariance structures. Scandinavian Journal of Statistics, 8, 65-92.

Jöreskog, K.G. (1982). The LISREL approach to causal model-building in the social sciences. In K.G. Jöreskog and H. Wold (Eds.) Systems Under Indirect Observation, Causality-Structure-Prediction, Part I. Amsterdam: North-Holland , p. 81-99.

Jöreskog, K.G. and Goldberger, A.S. (1972). Factor analysis by generalized least squares. Psychometrika, 37, 243-260.

Jöreskog, K.G. and Goldberger, A.S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. Journal of the American Statistical Association, 70, 631-639.

Jöreskog, K.G., Klovan, J.E. and Reyment, R.A. (1976). Geological Factor Analysis. Amsterdam, Holland : Elsevier.

Jöreskog, K.G. and Sörbom, D. (1976). Statistical models and methods for test-retest situations. In D.N.M. de Gruijter and L.J. van der Kamp (Eds.), Advances in Psychological and Educational Measurement, New York, Wiley, p.135-157.

Jöreskog, K.G. and Sörbom, D. (1977). Statistical models and methods for analysis of longitudinal data. In D.J. Aigner and A.S. Goldberger (Eds.), Latent variables in socio-economic models. Amsterdam : North-Holland, p. 285-325.

Jöreskog, K.G. and Sörbom, D. (1981). LISREL V : Analysis of Linear Structural Relationships by Maximum Likelihood and Least Squares Method. Research Report 81-8. Department of Statistics, Univ. of Uppsala, Sweden.

Jöreskog, K.G. and Sörbom, D. (1982). Recent Developments in structural equation modeling. Journal of Marketing Research, 19, 404-416.

Jöreskog, K.G. and Sörbom, D. (1985). LISREL VI : User's Guide. Mooresville, Indiana : Scientific Software Inc.

Jöreskog, K.G. and Sörbom, D. (1988). LISREL VII : A guide to the Program and Applications : Chicago : SPSS.

Jöreskog, K.G., van Thillo, M. and Gruvaeus, G.T. (1971). ACOVSM. A general computer program for analysis of covariance structures including generalized MANOVA. Princeton, N.J. : Educational Testing Service.

Judd, C.M., Jessor, R., and Donovan, J.E. (1986). Structure equation models and personality research. Journal of Personality, 54, 149-198.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lutkepohe, H., and Lee, T. (1985). The Theory and Practice of Econometrics, 2nd Edn., New York : Wiley.

Justice, J.H. (1972). An algorithm for inverting positive definite Toeplitz matrices. SIAM Jour. Applied Mathematics, 23, p. 289-291.

Kailath, T., Buckstein, A., and Morgan, D. (1986). Fast matrix factorization via discrete transmission lines. Linear Algebra and Its Applications, 75, 1-25.

Kailath, T. and I. Kolitracht (1986). Matrices with blocks Toeplitz inverses. Linear Algebra and its Application, 75, 145-153.

Kailath, T., Kung, S.Y. and Morf, M. (1979). Displacement ranks of matrices and linear equations. Journal of Mathematical Analysis and Applications, 68, 395-407.

Kailath, T., Vieira, A. and Morf, M. (1978). Inverses of Toeplitz operators, innovations and orthogonal polynomials. SIAM Rev., 20, 106-119.

Kale, B.K. (1982). On the solution of likelihood equations by iteration processes : The multiparametric case. Biometrika, 49, 479-486.

Kashyap, R.L. (1970). Maximum likelihood identification of stochastic linear systems. IEEE Transactions on Automatic Control, AC-15, 24-35.

Keesling, J.W. (1972). Maximum likelihood approaches to causal flow analysis. Unpublished Ph.D. thesis, University of Chicago, Chicago, Illinios.

Keifer, J. and Schwartz, R. (1965). Admissible Bayes character of T^2, R^2 and other fully invariant tests for classical multivariate normal problems. Annals of Mathematical Statistics, 40, 1100-1105.

Kendall, M.G. and Stuart, A. (1966). The Advanced Theory of Statistics, Vol. 3, New York : Hafner.

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

Kenny, D.A. (1979). Correlation and Causality. New York : Wiley.

Kenny, D.A. and Judd, C.M. (1984). Estimating the nonlinear and interactive effects of latent variables. Psychological Bulletin, 96, 201-210.

Khattab, A. and Hocevar, D. (1982). Significance listing in Confirmatory factor analytic models. Paper presented at the Annual Meeting of the American Educational Research Association, New York City, New York, March 22, 1982.

Kiiveri, H.T. (1987). An incomplete data approach to the analysis of covariance structures. Psychometrika, 52, 539-554.

King, M.L. (1983). Testing for autoregressive against moving average errors in the linear regression model. Journal of Econometrics, 21, 35-51.

Kmenta, J. and Gilbert, R.F. (1968). Small sample properties of alternative estimators of seemingly unrelated regressions. Journal of the American Statistical Association, 63, 1180-1200.

Krishnaiah, P.R. and Lee, J.C. (1976). On covariance structures. Sankhya, Ser. B., 38, 357-371.

Kshirsagar, A.M. (1972). Multivariate Analysis. New York : Marcel Decker.

Kumar, R. (1985). A fast algorithm for solving a Toeplitz system of equations. IEEE Transactions on Acoustics, Speech and Signal Processing. ASSP-33, 254-267.

Ky Fan and Hoffman, A.J. (1955). Some metric inequalities in the space of Matrices. Proceedings of American Mathematical Society, 6, 111-116.

La Du, T.J. and Tanaka, J.S. (1989). Influence of sample size, estimation method and model modification on goodness-of-fit assessments in structural equation models. Journal of Applied Psychology, 74, 625-635.

Lawley, D.N. and Maxwell, A.E. (1971). Factor Analysis as a Statistical Method, 2nd ed., Butterworth, London.

Lee, S.Y. (1977). Some algorithms for covariance structure analysis. Unpublished Ph.D. thesis, Dept. of Statistics, University of California, Los Angeles.

Lee, S.Y. (1980). Estimation of covariance structure models with parameters subject to functional restraints. Psychometrika, 45, 309-324.

Lee, S.Y. (1985). On testing functional constraints in structural equation models. Biometrika, 72, 125-131.

Lee, S.Y. and Bentler, P.M. (1980). Some asymptotic properties of constrained generalized least squares estimation in covariance structure models, South African Statistical Journal, 14, 121-136.

Lee, S. Y., and Jennrich, R.T. (1979). A study of algorithms for covariance structure analysis with specific comparisons using factor analysis. Psychometrika, 43, 99-113.

Lehmann, E.L. (1959). Testing Statistical Hypotheses. New York, Wiley.

Levinson, N. (1947). The Wiener RMS (root mean square) error criterion in filter design and prediction. Journal of Mathematical Physics, 25, 261-278.

Levy, K.J. (1976). A procedure for testing the equality of p correlated variances. British Journal of Mathematical and Statistical Psychology, 29, 89-93.

Li, C.C. (1975). Path Analysis - A Primer. Pacific Grove, Calif. : Boxwood.

Lillestol, J. (1976). Some Statistical Problems Related to Multiple Time Series Models with Structured Spectral Density Matrices. Bergen, Norway : The Norwegian School of Economics and Business Administration.

Lillestol, J. (1982a). Multivariate time series models with intraclass structure. Part-I : Identification. In O.D. Anderson (Ed.), Time Series Analysis : Theory and Practice I , Amsterdam : North-Holland. p. 335-347.

Lillestol, J. (1982b). Multivariate time series models with intraclass structure. Part II : Extraction and Prediction. In O.D. Anderson (Ed.); Time Series Analysis : Theory and Practice II, Amsterdam : North-Holland. p. 127-142.

Long, J.S. (1983a). Confirmatory Factor Analysis. Sage University Paper series on Quantitative Applications in the Social Sciences, No. 33. Berverly Hills and London : Sage.

Long, J.S. (1983b). Covariance Structure Models : An Introduction to LISREL . Sage University Paper series on Quantitative Applications in the Social Sciences, No. 34, Beverly Hills and London : Sage.

Lord, F.M. and Novick, M.R. (1968). Statistical Theories of Mental Test Scores. Reading, Mass. : Addison-Wesley.

MacCallum , R.C. (1986). Specification searchers in covariance structure models. Psychological Bulletin, 100, 107-120.

Maiti, S.S. and Mukherjee, B.N. (1989a). Two new tests of significance in the analysis of covariance structures. Technical Report No. ASC/89/12, Applied Statistics, Survey and Computing Division, Indian Statistical Institute, Calcutta.

Maiti, S.S. and Mukherjee, B.N. (1989b). Some test procedures for the Toeplitz covariance structure. Technical Report No.ASC/89/13, Applied Statistics, Survey and Computing Division, Indian Statistical Institute, Calcutta.

Maiti, S.S. and Mukherjee, B.N. (1989c). Two new goodness-of-fit indices for covariance matrices with linear structures. Technical Report No. ASC/89/14, Applied Statistics, Survey and Computing Division, Indian Statistical Institute, Calcutta.

Maiti, S.S. and Mukherjee, B.N. (1990). A note on distributional properties of the Jöreskog-Sörbom fit indices. Psychometrika, In Press.

Makhoul, J. (1975). Linear protection : A tutorial review. Proceedings of IEEE, 63, 561-580.

Makhoul, J. (1981). On the eigenvalue and symmetric Toeplitz matrix. IEEE Transec. Acoust. Spech and Signal processing, Vol. ASSP-29, 868-872.

Mallios, W.S. (1970). The analysis of structural effects in experimental design. Journal of the American Statistical Association, 65, 806-827.

Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979). Multivariate Analysis. London : Academic.

Marsh, H.W., Balla, J.R. and McDonald, R.P. (1988). Goodness-of-fit indices in confirmatory factor analysis : The effect of sample size. Psychological Bulletin, 103, 391-410.

Martin, N.G. and Eaves, L.J. (1977). The genetical analysis of covariance structures. Heredity, 38, 77-95.

Matsueda, R.L. and Bielby, W. (1986). Statistical power in covariance structure models. In N.B. Tuma (Ed.), Sociological Methodology, 1986. Washington, D.C., American Sociological Association, p. 120-158.

McAleer, M., McKenzie, C.R. and Hall, A.D. (1988). Testing separate time series model. Journal of Time Series Analysis, 9, 169-189.

McDonald, R.P. (1969). A generalized common factor analysis based on residual covariance matrices of prescribed structure. British Journal of Mathematical and Statistical Psychology, 22, 149-163.

McDonald, R.P. (1974). Testing pattern hypothesis for covariance matrices. Psychometrika, 39, 189-201.

McDonald, R.P. (1978). A simple comprehensive model for the analysis of covariance structures. British Journal of Mathematical and Statistical Psychology, 31, 59-72.

McDonald, R.P. (1980). A simple comprehensive model for the analysis of covariance structures : Some remarks on applications. British Journal of Mathematical and Statistical Psychology, 33, 161-183.

McDonald, R.P. (1989). An index of goodness-of-fit based on noncentrality. Journal of Classification, 6, 97-103.

McDonald, R.P. and Krane, W.R. (1977). A note on local identifiability and degree of freedom in the asymptotic likelihood ratio test. British Journal of Mathematical and Statistical Psychology, 30, 198-203.

McDonald, R.P. and Krane, W.R. (1979). A note on local identifiability and degrees of freedom in the asymptotic likelihood ratio test. British Journal of Mathematical and Statistical Psychology, 30, 198-203.

McDonald, R.P. and Marsh, H.W. (1990). Choosing a multivariate model : Noncentrality and goodness of fit. Psychological Bulletin, In press.

McDonald, R.P. and Swaminathan, H. (1972). Structural analysis of dispersion matrices based on a very general model with a rapidly convergent procedure for the estimation of parameters. An informal publication of the department of Measurement and Evaluation : The Ontario Institute for studies in Education.

McQuitty, L.L. (1956). Agreement analysis : Classifying persons by predominant pattern of response. British Journal of Statistical Psychology, 9, 5-16.

Mentz, R.P. (1976). On the inverse of some covariance matrices of Toeplitz type. SIAM Journal of Applied Mathematics, 31, 426-437.

Mentz, R.P. (1977). Estimation in the first-order moving average model through the finite autoregressive approximation : Some asymptotic results. Journal of Econometrics, 6, 225-236.

Morf, M. (1980). Doubling algorithms for Toeplitz and related equations. Proceedings of 1980 International Conference on Acoustics, Speech and Signal Processing, Denver, Colorado, April 9-13, 1980, p. 954-959.

Morrison, D.F. (1967). Multivariate Statistical Methods. New York : McGraw-Hill.

Muirhead, R. (1984). Aspects of Multivariate Statistical Theory. 2nd Edn., New York : Wiley.

Mukherjee, B.N. (1963). Derivation of likelihood ratio tests for Guttman quasi-simplex covariance structures. Unpublished Ph.D. dissertation, Univ. of North Carolina, Chapel Hill, N.C., USA.

Mukherjee, B.N. (1966). Derivation of likelihood ratio tests for the Guttman quasi-simplex covariance structures, Psychometrika, 31, 97-123.

Mukherjee, B.N. (1969a). Covariance structure analysis in psychological research. Paper read at the annual convention of the Canadian Psychological Association held at Toronto, Ont.

Mukherjee, B.N. (1969b). Invariance of the Guttman quasi-simplex linear model under selection. British Journal of Mathematical and Statistical Psychology, 22, 1-18.

Mukherjee, B.N. (1970). Likelihood ratio tests of statistical hypotheses associated with patterned covariance matrices in Psychology. British Journal of Mathematical and Statistical Psychology, 23, 89-120.

Mukherjee, B.N. (1973). Analysis of covariance structures and exploratory factor analysis. British Journal of Mathematical and Statistical Psychology, 26, 125-154.

Mukherjee, B.N. (1976). Techniques of covariance structural analysis. Australian Journal of Statistics, 18, 131-150.

Mukherjee, B.N. (1981). A simple approach to testing of hypothesis regarding a class of covariance structures. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on Statistics : Application and New Directions, Calcutta, 442-465.

Mukherjee, B.N. (1982). Measure of multivariate relationships for the uniform covariance case. Behaviormetrika, 12, 27-46.

Mukherjee, B.N. (1984). Explicit maximum likelihood solution in the multi normal estimation problem in analysis of covariance structures. Paper presented in the Divisional Seminar on 11 April, 1984. Division of Applied Statistics, Indian Statistical Institute, Calcutta.

Mukherjee, B.N. and Maiti, S.S. (1984). Explicit maximum likelihood estimators in the analysis of covariance structures. Technical Report No. ASC/84/6, Applied Statistics, Surveys and Computing Division, Indian Statistical Institute, Calcutta.

Mukherjee, B.N. and Maiti, S.S. (1986). Convergence of maximum likelihood solution for structural analysis of patterned covariance matrices. Tech. Report No. ASC/86/10, Applied Statistics, Survey and Computing Division, Indian Statistical Institute, Calcutta.

Mukherjee, B.N. and Maiti, S.S. (1988a). Q-procedure for solving likelihood equations in the analysis of covariance structures. Computational Statistics Quarterly, 4, 105-128.

Mukherjee, B.N. and Maiti, S.S. (1988b). Iterative maximum likelihood estimation of parameters of the Toeplitz correlation structure. British Journal of Mathematical and Statistical Psychology, 41, 63-99.

Mukherjee, B.N. and Maiti, S.S. (1988c). On some properties of positive definite Toeplitz matrices and their possible applications. Linear Algebra and its Applications, 102, 211-240.

Mukherjee, B.N. and Maiti, S.S. (1988d). On conditions for equality of OLSE, GLSE and MLE in analysis of covariance structures. Calcutta Statistical Association Bulletin, 37, 171-192.

Mulaik, S.A., James, L.R., van Alstine, J., Bennett, N., Lind, S. and Stilwell, C.D. (1989). Evaluation of goodness-of-fit indices for structural equation models. Psychological Bulletin, 105, 430-445.

Mustafi, C.K. (1967). On iterated tests of hypotheses. Journal of the Americal Statistical Association. 62, 520-524.

Muthén, B. (1987). LISCOMP : Analysis of linear structural equations using a comprehensive measurement model - User's Guide. Mooresville. Indiana : Scientific Software.

Nagao, H. (1973). On some test criteria for covariance matrix. Annals of Statistics, 1, 700-709.

Narain, R.D. (1950). On the completely unbiased character of tests of independence in multivariate normal systems. Annals of Mathematical Statistics, 21, 293-298.

Nehorai, A. and Morf, M. (1985). A unified derivation for fast estimation algorithms by the conjugate direction dethod. Linear Algebra and Its Applications, 72, 119-143.

Nicholls, D.F., and Hall, A.D. (1979). The exact maximum likelihood function of multivariate autoregressive moving average models. Biometrika, 66, 259-264.

Oberhofer, W. and Kmenta, J. (1974). A general procedure for obtaining maximum likelihood estimator in generalized regression models. Econometrics, 42, 579-590.

Olkin, I. (1973). Testing and estimation for structures which are circularly symmetric in blocks. In D.G. Kabe and R.P. Gupta (Eds.), Multivariate Statistical Inference. Amsterdam : North-Holland.

Olkin, I. and Press, S.J. (1969). Testing and estimation for a circular stationary model. Annals of Mathematical Statistics, 40, 1358-1373.

Osborn, D.R. (1977). Exact and approximate likelihood estimators for vector moving average processes. Journal of the Royal Statistical Society, Ser. B, 39, 114-118.

Ouellet, D.V. (1981). Schur complements and statistics. Linear Algebra and Its Applications, 36, 187-195.

Pagano, M. (1972). An algorithm for fitting auto-regressive schemes. Applied Statistics, 21, 274-281.

Parzen, E. (1971). Efficient estimation of stationary time series mixed schemes. Bulletin of the International Statistical Institute, 44, 315-319.

Powell, D.R. and Macdonald, J.R. (1972). A rapidly convergent iterative method for the solution of the generalized nonlinear least squares problem. The Computer Journal, 15,

Pratt, W.K. (1978). Digital Image Processing. New York : Wiley.

Press, S.J. (1972). Applied Multivariate Analysis. New York : Holt, Rinehart and Winston.

Pullman, N.J. (1976). Matrix Theory and Its Applications, New York and Basel : Marcel Dekker, Inc.

Rao, C.R. (1947). Large sample tests of statistical hypotheses concerning several parameters with applications to problem of estimation. Proc. Camb. Phil. Soc., 44, 50-57.

Rao, C.R. (1969). A decomposition theorem for vector variables with a linear structure. Annals of Mathematical Statistics, 40, 1845-1849.

Rao, C.R. (1970). Estimation of heteroscedastic variances in a linear model. Journal of American Statistical Association, 65, 161-172.

Rao, C.R. (1974). Linear statistical inference and its applications, 2nd Ed., New Delhi : Wiley Eastern.

Rao, C.R. and Mitra, S.K. (1971). Generalized Inverse and its Applications. New York : Wiley.

Rissanen, J. (1973). Algorithms for triangular decompositions of block Hankel and Toeplitz matrices with applications to factoring positive matrix polynomials. Mathematics of Computation, 7, 147-154.

Robinson, P.M. (1974). Identification, estimation and large sample theory for regressions containing unobserable variables. International Economic Review, 15, 680-692.

Robinson, P.M. (1977). The estimation of a multivariate linear relation. Journal of Multivariate Analysis, 7, 409-423.

Rock, D.A. and Werts, C.E. (n.d.). An analysis of time related increments and/or decrements for GRE repeaters across ability and sex, unpublished manuscript. Educational Testing Service, New Jersey.

Rock, D.A., Werts, C.E. and Flaugher, R.I. (1978). The use of covariance structures for comparing the psychometric properties of multiple variable across populations. Multivariate Behavioral Research, 13, 403-418.

Rogers, G.S. and Young, D.L. (1977). Explicit maximum likelihood estimators for certain patterned covariance matrices, Communication in Statistics : Theory and Method, A6, 121-133.

Rosenfled, A. and Kak, A.C. (1982). Digital Picture Processing. New York : Academic.

Roy, J. (1951). The distribution of certain likelihood criteria useful in multivariate analysis. Bulletin of International Statistical Institute. 33, 219-230.

Roy, J. (1954). On some tests of significance in samples from bipolar normal distributions. Sankhya, 14, 203-210.

Roy, S.N., Greenberg, B.G. and Sarhan, A.E. (1960). Evaluation of determinants, characteristic equations and their roots for a class of patterned matrices. Journal of Royal Statistical Society, Ser. B, 22, 348-359.

Roy, S.N. and Sarhan, A.E. (1956). On inverting a class of patterned matrices. Biometrika, 43, 227-231.

Saris, W.E., de Pijper, W.M. and Zegwaart, P. (1979). Detection of Specification errors in the linear structural equation models. In K. Schnessler (Ed.), Sociological Methodology, p. 151-171. San Francisco: Jossey-Bass.

Satorra, A. (1989). Alternative test criteria in covariance structure analysis : A unified approach. Psychometrika, 51, 131-151.

Satorra, A., and Saris, W.E. (1985). Power of the likelihood ratio test in covariance structure analysis. Psychometrika, 50, 83-90.

Savin, N.E. (1976). Conflict among testing procedures in a linear regression model with auto-regressive disturbances. Econometrica, 44, 1303-1315.

Schoenberg, R. (1982). MILS : A Computer Program to Estimate the Parameters of Multiple Indicator Linear Structure Models. Bethesda, Md. : National Institute of Health.

Schwartz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6, 461-464.

Seber, G.A.F. (1984). Multivariate Observations. New York : Wiley.

Seely, J. (1971). Quadratic subspaces and completeness. Annals of Mathematical Statistics, 42, 710-721.

Siddiqui, M.M. (1958). On the inversion of the sample covariance matrix in a stationary autoregressive process. Annals of Mathematical Statistics, 29, 585-588.

Singh, B.K. (1975). Path analysis in social science research. In B.K. Singh (Ed.), Recent Trends in Social Analysis. Special Volume of Indian Journal of Extension Education.

Silvey, S.D. (1959). The Lagrange multiplier test, Annals of Mathematical Statistics, 30, 389-407.

Sörbom, D. (1989). Model modification. Psychometrika, 54, 371-384.

Sörbom, D. and Jöreskog, K.G. (1982). The use of structural equation models in evaluation research. In C. Forenell (Ed.), A Second Generation of Multivariate Analysis (Vol. 2). New York : Praegar, p. 381-418.

Srivastava, J.N. (1966). On testing of hypotheses regarding a class of covariance structures. Psychometrika, 31, 147-164.

Srivastava, J.N. and Maik, R.L. (1967). On a new property of partially balanced association schemes useful in psychometric structural analysis. Psychometrika, 32, 279-289.

Srivastava, M.S. (1965). Some tests for the intra-class correlation model. Annals of Mathematical Statistics, 36, 1802-1806.

Steiger, J.H. (1980). Testing, pattern hypotheses on correlation matrices : alternative statistics and some empirical results. Multivariate Behavioral Research, 15, 335-352.

Steiger, J.H. (1989). EzPATH, Version 1.0, July 1, 1989. SYSTAT, Inc. 1800 Sherman Avenue, Evanston, Illinois 60201.

Styan, G.P.H. (1969). Inference in multivariate normal populations with structure. Part II : Inference when correlations have structure. Technical Report No. 115, University of Minnesota, Minneapolis.

Swain, A.J. (1975). A class of factor analysis estimation procedures with common asymptotic sampling properties. Psychometrika, 40, 315-335.

Szatrowski, T.H. (1979). Asymptotic non-null distributions for likelihood ratio statistics in the multivariate normal patterned mean and covariance matrix testing. Annals of Statistics, 7, 823-837.

Szatrowski, T.H. (1980). Necessary and sufficient conditions for explicit solutions in the multivariate normal estimation problem for patterned means and covariances. Annals of Statistics, 9, 802-808.

Szatrowski, T.H. (1982). Relative efficiencies of estimates using patterned covariances and correlations in the multivariate normal estimation problem. Annals of the Institute of Statistical Mathematics, 34, Part A, 299-307.

Takeuchi, K., Yanai, H. and Mukherjee, B. N. (1982). The Foundations of Multivariate Analysis, New Delhi : Wiley Eastern.

Tanaka, J.S. (1984). Some results on the estimation of covariance structure models (Doctoral dissertation, University of California, Los Angeles, 1984). Dissertation Abstracts International, 45, 924B.

Tanaka, J.S. (1987). " How big is big enough ?" : Sample size and goodness of fit in structural equation models with latent variables. Child Development, 58, 134-146.

Tanaka, J.S. and Huba, G.J. (1985). A fit index for covariance structure models under arbitrary GLS estimation, British Journal of Mathematical and Statistical Psychology, 38, 197-201.

Tanaka, J.S. and Huba, G.J. (1987). Assessing the stability of depression in college students. Multivariate Behavioral Research, 22, 5-19.

Tanaka, J.S., Panter, A.T., Winborne, W.C. and Huba, G.J. (1989). Theory testing in personality and social psychology with structural equation models : A Primer in 20 questions. Review of Personality and Social Psychology, 11, (to appear).

Tanaka, J.S., Panter, A.T. and MacCallum, R. (1990). Personal communication.

Tanaka, Y. and Odaka, Y. (1989). Sensitivity analysis in maximum likelihood factor analysis. Communication in Statistics: Theory and Methods, 18, 4067-4084.

Thurstone L.L. (1947). Multiple Factor Analysis. Chicago : Univ. of Chicago Press.

Timm, N.H. and Carlson, J.E. (1976). Part and bipartial canonical analysis. Psychometrika, 41, 159-176.

Trench, W.F. (1964). An algorithm for the inversion of finite Toeplitz matrices. SIAM Jour. of Applied Mathematics, 12, p. 515-522.

Trench, W.F. (1967). Weighing coefficients for the prediction of stationary time series from the finite past. SIAM Jour. of Applied Mathematics, 15, 1502-1510.

Trench, W.F. (1985). On the eigenvalue problem for Toeplitz matrices generated by rational functions. Linear and Multilinear Algebra, 17, 337-353.

Trench, W.F. (1988). Numerical solution of the eigenvalue problem for symmetric rationally generated Toeplitz matrices. SIAM Journal of Matrix Analysis and Application, 10, 135-146.

Tucker, L.R. and Lewis, C. (1973). A reliability co-efficient for maximum likelihood factor analysis. Psychometrika, 38, 1-10.

Tukey, J.W. (1964). Causation, regression and path analysis. In O. Kempthorne, T.A. Bancroft, J.W. Gower and J.L. Lush (Eds.), Statistics and Mathematics in Biology, New York : Hafner.

Vinod, H.D. and Ullah, A. (1981). Recent Advances in Regression Methods. New York : Marcel Dekker.

Votaw, D.F. (1948). Testing compound symmetry in a normal multivariate distribution. Annals of Mathematical Statistics, 19, 447-448.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the Americal Mathematical Society, 54, 426-482.

Ware, J.H. (1985). Linear models for the analysis of longitudinal studies. The American Statistician, 39, 95-101.

Wendroff, B. (1966). Theoretical Numerical Analysis, New York : Academic Press.

Werts, C.E., Breland, H.M., Grandy, J. and Rock, D.R. (1980). Using longitudinal data to estimate reliability in the presence of correlated measurement errors. Educational and Psychological Measurement, 40, 19-29.

Werts, C.E., Jöreskog, K.G. and Linn, R.L. (1973). Identification and estimation in path analysis with unmeasured variables. American Journal of Sociology, 78, 1469-1484.

Werts, C.E. and Linn, R.L. (1972). Correlation for attenuation Education and Psychological Measurement, 32, 117-127.

Werts, C.E., Linn, R.L. and Jöreskog, K.G. (1974). Intraclass reliability estimates : Testing structural assumptions. Educational and Psychological Measurement, 34, 25-33.

Werts, C.E., Linn, R.L. and Jöreskog, K.G. (1978). The reliability of college grades from longitudinal data. Educational and Psychological Measurement, 38, 89-95.

Werts, C.E., Pike, L., Rock, D.A. and Grandey, J. (1981). Applications of quasi-Markov simplex models across populations. Educational and Psychological Measurement, 41, 296-307.

Werts, C.E., Rock, D.R., Linn, R.L. and Jöreskog, K.G. (1978). A general method of estimating the reliability of a composite. Educational and Psychological Measurement, 38, 933-938.

Whittle, P. (1951). Hypotheses testing in time series analysis. Uppsala, Sweden : Almquist and Wiksells.

Whittle, P. (1963). On the fitting of multivariate autoregressive and the approximate canonical factorization of a spectral density matrix, Biometrika, 50, 129-134.

Widom, H. (1958). On the eigenvalues of certain Hermitian operators. Transactions of American Mathematical Society, 88, 491-522.

Widom, H. (1960). Inversion of Toeplitz matrices II. Illinois Journal of Mathematics, 4, 88-89.

Wise, J. (1955). The autocorrelation function and the spectral density function. Biometrika, 42, 151-159.

Wiley, D.E., Schmidt, W.H. and Bramble, W.J. (1973). Studies of a class of covariance structure models. Journal of American Statistical Association, 68, 317-323.

Wilks, D.M. and Hayes, M.H. (1987). An eigenvalue recursion for Toeplitz matrices. IEEE Transaction on Acoustics, Speech, Signal Processing, Vol. ASSP - 35, 907-909.

Wilks, S.S. (1946). Sample criteria for testing equality of means, equality of variance and equality of covariance in a normal multivariate distribution. Annals of Mathematical Statistics, 17, 257-281.

Wilson, G.A. and Martin, S.A. (1983). An empirical comparison of two methods for testing the significance of a correlation matrix. Educational and Psychological Measurement, 43, 11-14.

Wright, S. (1934). The method of path coefficients, Annals of Mathematical Statistics, 5, 161-215.

Wright, S. (1960a). Path coefficients and path regressions : Alternative or complimentary concepts. Biometrics, 16, 189-202.

Wright, S. (1960b). The treatment of reciprocal interaction, with or without lag in path analysis. Biometrics, 16, 423-445.

Zellini, P. (1979). On the optimal computation of a set of symmetric and persymmetric biolinear forms. Linear Algebra and its Applications, 23, 101-119.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressors and tests for aggregation bias. Journal of the American Statistical Association, 57, 348-368.

Zohar, S. (1969). The algorithm of W.F. Trench. Jour. of Association of Computing Machine, 16, 592-601.

Zohar, S. (1974). The solution of a Toeplitz set of linear equations. Jour. of Association of Computing Machine, 21, 272-276.

ERRATA SHEET

to accompany the Ph.D. Dissertation entitled

" ON SOME PROBLEMS IN ANALYSIS OF COVARIANCE STRUCTURES''

by

Sadhan Samar Maiti

Sl. no.	Page and Location	Incorrect portion	Corrected version
1.	p.2,line 11	Jöreskög	Jöreskog
2.	p.29, line 12	are	is
3.	p.40, line 18	tools especially	tools, especially
4.	p. 54, line 7	simultoneous	simultaneous
5.	p.55, line 3	problem	problems
6.	p.58, line 8	Life example	Live example
7.	p.58, line 10	is reported	are reported
8.	p.58, line 16	exphasis on	emphasis on its
9.	p.77, last line	=	≡
10.	p.86, line 6 from bottom	slove	solve
11.	p.89, line 3	breakes	breaks
12.	p.97, line 5	as	an
13.	p.105, last line	$[J(\theta)]^{-1}$	$[\underset{\sim}{J}(\underset{\sim}{\theta})]^{-1}$or by $\underset{\sim}{E}_2=-[\underset{\sim}{I}(\underset{\sim}{\theta})]^{-1}$.
14.	p.132, line 9	]	].
15.	p.167, line 4	$GLSE(\hat{\Sigma}_R)$	$GLSE(\hat{\underset{\sim}{\Sigma}}_R)$ of $\underset{\sim}{\Sigma}$
16.	p. 170, last line	easy and	easy
17.	p. 172, line 6	method	methods
18.	p. 172, line 9	$(tr\ S^{-1})^2$	$[tr(\underset{\sim}{S}^{-1})]^2$
19.	p. 183, line 11 from bottom	that S has	that when $\underset{\sim}{S}$ has
20.	p. 187, line 2	K = 1	k = 2
21.	p. 190, line 2	casual	causal
22.	p. 195, line 15	1989a	1989c
23.	p. 208, line 12 from bottom	Thus	Thus,

(Cont. 2)

24.	p. 213, line 3	area	areas
25.	p. 233, line 3 from bottom	$\underset{\sim}{U}_m^{-1} = \underset{\sim}{U}'_m =$	$\underset{\sim}{U}'_m =$
26.	p.248, line 1	$[\rho_m^{-1}]_{i+1,j+1} = [\rho_m^{-1}]_{i,j} +$	$[\underset{\sim}{\rho}_m^{-1}]_{i+1,j+1} = [\underset{\sim}{\rho}_m^{-1}]_{i,j} +$
27.	p. 254, line 8	ρ_i^*	$\underset{\sim}{\rho}_i^*$
28.	p. 254 line 8	ρ_{i-1}	$\underset{\sim}{\rho}_{i-1}$
29.	p. 267, line 6 from bottom	K_k	$\underset{\sim}{K}_k$
30.	p. 384, line 4	Listing	testing
31.	p. 388, line 4 from bottom	A note on	A Monte-Carlo study of
32.	p. 388, last line	30, 198-203	32, 121-132.
33.	p.389, line 7	department	Department
34.	p. 389, line 3	In press	107, 247-255.
35.	p. 390, line 7 from bottom	Measure	Measures
36.	p. 392, line 11	dethod.	method.
37.	p. 396, line 5 from bottom	Testing,	Testing
38.	p. 401, line 9 from bottom	biolinear	bilinear