ESTIMATION AND TESTING FOR BLOCK COMPOUND SYMMETRY
AND OTHER PATTERNED COVARIANCE MATRICES WITH LINEAR
AND NON-LINEAR STRUCTURE

BY

TED HOWARD SZATROWSKI

TECHNICAL REPORT NO. 107
JANUARY 31, 1976

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT
MPS 7509450-000

Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ESTIMATION AND TESTING FOR BLOCK COMPOUND SYMMETRY
AND OTHER PATTERNED COVARIANCE MATRICES WITH LINEAR
AND NON-LINEAR STRUCTURE

by

TED HOWARD SZATROWSKI

TECHNICAL REPORT NO. 107
JANUARY 31, 1976

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT MPS 7509450-000

INGRAM OLKIN, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ACKNOWLEDGMENTS

I would like to express my appreciation to my advisor, Professor Ingram Olkin, for suggesting this area for investigation and for many helpful discussions. I am also grateful to Professors Rosedith Sitgreaves and Paul Switzer for reading the final manuscript and to Professor Dennis Young for reading and making suggestions on the first chapter while the research was in progress. Thanks are also due to Professor John J. Miller for useful comments and conversations. I wish to express my appreciation to my wife, Mira Herrmann, for her encouragement and patience.

For the excellent and speedy preparation of the final manuscript, I thank Carolyn Knutsen.

I am grateful to the National Science Foundation for its financial support, both as a Competitive Fellow and as a research associate under Contract Number MPS 7509450-000, in the preparation of this dissertation.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ACKNOWLEDGMENT</td>
<td>iii</td>
</tr>
<tr>
<td></td>
<td>INTRODUCTION</td>
<td>vii</td>
</tr>
<tr>
<td>I</td>
<td>TESTING AND ESTIMATION FOR BLOCK COMPOUND SYMMETRY IN A NORMAL MULTIVARIATE DISTRIBUTION</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2.</td>
<td>Types of Block Compound Symmetry and the Statement of the Hypotheses</td>
<td>9</td>
</tr>
<tr>
<td>3.</td>
<td>Canonical Forms and Estimation</td>
<td>16</td>
</tr>
<tr>
<td>4.</td>
<td>Type I One-population Hypotheses: LRT and Approximate Null and Asymptotic Non-null Distributions</td>
<td>32</td>
</tr>
<tr>
<td>5.</td>
<td>Type I k-population Hypotheses: LRT and Approximate Null and Asymptotic Non-null Distributions</td>
<td>45</td>
</tr>
<tr>
<td>6.</td>
<td>Type II One-population Hypotheses: LRT and Approximate Null and Asymptotic Non-null Distributions</td>
<td>55</td>
</tr>
<tr>
<td>7.</td>
<td>Type II k-population Hypotheses: LRT and Approximate Null and Asymptotic Non-null Distributions</td>
<td>65</td>
</tr>
<tr>
<td>II</td>
<td>TESTING AND ESTIMATION FOR A (4 \times 4) CIRCULAR STATIONARY MODEL WHEN (\rho_2 \neq 0)</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>Introduction</td>
<td>74</td>
</tr>
<tr>
<td>2.</td>
<td>Canonical Forms and Estimation When (\rho_2 \neq 0)</td>
<td>76</td>
</tr>
<tr>
<td>3.</td>
<td>MLE When (\rho_2 = 0)</td>
<td>77</td>
</tr>
<tr>
<td>4.</td>
<td>Roots of the MLE</td>
<td>80</td>
</tr>
<tr>
<td>5.</td>
<td>LRT and Its Asymptotic Distribution When (\rho_2 = 0)</td>
<td>83</td>
</tr>
<tr>
<td>6.</td>
<td>The Asymptotic Non-null Distribution of (-(2/n)\log \lambda)</td>
<td>84</td>
</tr>
<tr>
<td>7.</td>
<td>Asymptotic Efficiencies</td>
<td>92</td>
</tr>
</tbody>
</table>
III. OTHER SPECIFIC COVARIANCE MATRICES OF THE FORM

\[ \Sigma = \sigma^2 \mathbf{\Psi} = \sigma^2 \left( \mathbf{I} + \sum_{g=1}^{m} \rho_g \mathbf{g}_g \right) \]

1. Introduction

2. A Special \(4 \times 4\) Circular Stationary Model With \( \rho_1 = \rho^4 \)

3. Hypothesis Testing for Two Different Forms of \( \Sigma \), Both Functions of \( \sigma^2 \) and \( \rho \)

4. A General Stationary Model

5. Examples of Covariance Matrices Without Explicit MLE

IV. COMPUTATIONAL PROCEDURES

1. Introduction

2. The Newton-Raphson and Scoring Methods

3. A Third Algorithm

4. Simplifications When the \( \mathbf{g}_g \) Are Simultaneously Diagonalizable

5. Discussion of Computation Procedures

6. A Method of Monitoring Step Size

V. "AVERAGING" AND MAXIMUM LIKELIHOOD ESTIMATION

1. Introduction

2. Preliminary Remarks and Definitions

3. Main Theorem

4. Applications

5. Discussion of Open Questions

APPENDIX

I. SAMPLE CALCULATIONS FOR \( \widetilde{\mathbf{h}}_1(\text{mve}) \)

1. MLE Calculations Leading to IRT
# Appendix

2. Moments of the IRT Under the Null Distribution 144
3. Approximate Null Distribution Calculations 147
4. Non-null Asymptotic Distribution 148
5. Non-null Distribution: k-population Case 153

## II Tables of Asymptotic Means and Variances

1. Table A 164
2. Table B 169
3. Table C 170

# Bibliography 175
INTRODUCTION

This study deals with several problems in the area of testing and estimation of patterned covariance matrices in multivariate analysis. Applications to the fields of psychology, education, medicine and genetics arise naturally when comparisons of certain classes of multivariate data are desired. Maximum likelihood estimators (MLE) are derived for the estimation problems and likelihood ratio tests (LRT) are used for testing problems. Approximate or asymptotic distributions of the LRT are given using the Box (1949) approximation or standard delta method (Cramer, 1946). Problems under consideration include cases where explicit MLE are found and where the MLE are solutions to quadratic, cubic or higher degree polynomials. For cases where the MLE can not be explicitly found, computational procedures are suggested and difficulties in implementation are discussed. Finally, sufficient conditions for explicit MLE, single iteration convergence and a method of averaging yielding the MLE for patterned covariance matrices are given.

In chapter I, we consider testing and estimation problems that arise under assumptions of block compound symmetry. This generalizes the work of Votaw (1948). The hypotheses under consideration include exchangeability of subsets of variables. Such exchangeability is of interest when one is trying to determine if several variates are exchangeable with respect to some outside criterion as in educational testing, for example. Estimates and tests are derived for twelve hypotheses as well as approximate null and asymptotic non-null distributions of the LRT.
In chapter II, we test the hypothesis that the second order correlation is zero in a $4 \times 4$ circulant covariance matrix. The MLE under this restriction are solutions to a cubic polynomial. The multiple root case is discussed and the global maximum is found. Asymptotic null and non-null distributions are obtained.

Cases where the MLE are solutions to quadratic, cubic and higher order polynomials are discussed in chapter III. This leads to a discussion of computational procedures in chapter IV. The method of scoring applied to patterned covariance matrices with linear structures of Anderson (1973) is reviewed. Computational procedures for finding the MLE of covariance matrices whose elements are a common constant times a polynomial in the correlation coefficient with known coefficients are derived. Initial starting points and convergence of the procedures are also discussed.

In chapter V, sufficient conditions are given for the cases in which patterned covariance matrices have explicit solutions and when these explicit solutions are obtained in a single iteration from any positive definite starting point using the method of scoring. Also included are sufficient conditions for cases when the MLE may be obtained by averaging elements of the sample covariance matrix. These results are applied to some well-known patterns including the cases of block and non-block complete, compound and circular symmetry.
I. TESTING AND ESTIMATION FOR BLOCK COMPOUND
SYMmetry IN A NORMAL Multivariate DISTRIBUTION

1. Introduction.

Among the models considered for testing and estimation of multivariate normal distributions have been models arising from observations where an underlying (mean and/or covariance) structure is assumed. Various symmetric structures have been proposed and appropriate hypothesis tests derived. We begin with a brief review of some of the literature in this area. Mauchly (1940) investigates the case of spherical symmetry. The case of complete symmetry in the intraclass correlation model is first studied by Wilks (1946) and extended by Votaw (1948) to the case of compound symmetry. Both the Wilks and Votaw papers are motivated by applications to psychological testing problems. The distributions of sums of squares and cross-products of multivariate normal variables with compound symmetric covariance matrices are found by Morrison (1962). The exact distribution of Votaw's likelihood criteria have been considered by Consul (1968, 1969), and Mathai (1970).

Geisser (1963) considers the problem of testing the hypothesis that the mean vector \( \mu \) is specified, \( \mu = \mu_0 \), versus the alternative that \( \mu \) is unspecified, under the assumption that the covariance matrix conforms to the intraclass correlation model. This is a modified Hotelling's \( T^2 \) problem, the modification being that the covariance matrix has special structure. Olkin and Shrikhande (1970) describe an extension of Wilks' test for the equality of means. Morrison (1972) also studies the problem of testing means with structured covariance matrix and compares the
performance of several of his proposed estimators for problems in this area with Hotelling's $T^2$.

Another class of problems where structure can be exploited is the comparison of more than one population. M.S. Srivastava (1965) derives several tests used in the comparison of several populations assumed to conform to the intraclass correlation model. An alternative procedure to this model is suggested by Krishnaiah and Pathak (1967). Olkin and Press (1969) study the circularly symmetric model, deriving tests and distributions of these tests for several nested hypotheses involving structure on the means and covariance matrices. Gleser and Olkin (1972) consider testing problems with mean and covariance structure that arise out of testing the hypothesis that $k$ psychological tests are parallel forms of the same test.

The simplex model was introduced by Guttman (1954, 1957). Mukherjee (1966) discusses maximum likelihood estimates (MLE) of several special cases of the simplex model. A general discussion of MLE for simplex models can be found in Jöreskog (1970). Other special cases of the simplex model are considered in discussions of the model where the covariance matrix $\Sigma$ is assumed to be of the form:

$$\Sigma = a_1 G_1 + \cdots + a_m G_m,$$

where the $a$'s are unknown scalars and the $G$'s known symmetric matrices. J.N. Srivastava (1966) considers a special case of the assumption that the $G$'s are simultaneously diagonalizable in the above model. Mukherjee (1970) considers a more general case when the $G$'s are diagonalizable. Anderson (1970, 1973) proposes two algorithms for finding the MLE for $G$'s not restricted to being diagonalizable.
These ideas can also be applied to covariance matrices with block structures. Previous studies on models with special block structures include those by Fleiss (1966), Arnold (1970, 1973), John (1971), and Olkin (1972, 1973). Fleiss considers a case of block complete symmetry; Arnold investigates the block cases of the models of complete and compound symmetry; John reports on a general set of hypotheses for models in which the covariance structure may be block diagonalized, and Olkin (1972) generalizes the circularly symmetric model to the block circularly symmetric case. Olkin (1973) gives a general discussion concerning structured covariance matrices including some comments on block structures.

Votaw presents two types of compound symmetry together with the tests of six hypotheses for each type. Let \( x_1, \ldots, x_N \) be a sample from a \( p \)-variate normal distribution with mean vector \( \mu \) and covariance matrix \( \Sigma \). Suppose the \( p \)-variate vector is divided into several mutually exclusive subsets. A typical hypothesis considered by Votaw, denoted \( H_1(mvc) \), is that within each subset of variates the means are equal, the variances are equal and the covariances are equal and between two subsets of variates the covariances are equal. Votaw tests this hypothesis against the alternative that \( \mu \) and \( \Sigma \) are unrestricted.

In this chapter, we consider the problems of estimation and testing for these hypotheses in the case of block compound symmetry. We also include a study of the additional hypothesis that \( \mu = \mu_0 \) and \( \Sigma \) has block compound symmetric structure. This is tested against the alternative that \( \mu \neq \mu_0 \) and \( \Sigma \) has block compound symmetric structure. It is shown in this chapter that one can reduce these problems to simple canonical forms for which solutions are readily available.
As an example of a problem involving type I compound symmetry, Votaw (1948) considers the case of testing whether three examinations are inter-changeable with respect to some outside criterion under an assumption of normality. Let $X$ be a 4-variate multivariate normal row vector whose first component is the score of an individual on the outside criterion and second, third and fourth components are the individual's scores on the three examinations. Votaw considers the hypothesis that $X$, a 4-variate normal vector, has mean

$$
\mu = (\mu_1 ; \mu_2, \mu_2, \mu_2),
$$

and covariance matrix

$$
\Sigma = \begin{bmatrix}
A & C & C & C \\
C' & B & D & D \\
C' & D & B & D \\
C' & D & D & B
\end{bmatrix},
$$

versus the alternative hypothesis that $X$ has a general mean and covariance matrix. In this case $A$, $B$, $C$ and $D$ are all scalars.

This example can be extended naturally to an example of block compound symmetry. Suppose there are $b$ outside criteria and each examination consists of $n$ subtests. Then a score for each subtest of the examination could be reported rather than one overall score on the examination. In this case, partition $X$ into:

$$
X = (X_1, X_2, X_3, X_4), \quad X_1 = (X_{11}, \ldots, X_{1b}),
$$

$$
X_i = (X_{i1}, \ldots, X_{in}), \quad i = 2, 3, 4.
$$
We can again test the hypothesis that $X$ has mean vector $\mu$, 

$$\mu = (\mu_1, \mu_2, \mu_2, \mu_2),$$

where $\mu_1(b \times b)$ and $\mu_2(1 \times n)$ are general vectors and $\Sigma$ is given by (1.2) where $A$ is $b \times b$, $C$ is $b \times n$ and $B$ and $D$ are $n \times n$ matrices. Again the alternative hypothesis is a general mean and covariance matrix.

Fleiss (1966) considers the following example of block complete symmetry involving a test of reliability. In his problem, $k$ observers each grade the same $N$ subjects on $p$ tests. Fleiss wishes to test the agreement of the $k$ observers. On each subject, he has $p$ scores for each of the $k$ observers, i.e., a $pk$ component row vector $X$ containing $k$ subvectors, with $p$ elements in each subvector, $X = (X_1', \ldots, X_k')$ where $X_i = (X_{i1}, \ldots, X_{ip})$, $i = 1, \ldots, k$. Assuming a covariance matrix $\Sigma$ of the form:

$$\Sigma = \begin{pmatrix}
A & B & \cdots & B \\
B & & & \\
\vdots & & & \\
B & & & A
\end{pmatrix},$$

and a mean vector $\mu$ of the form:

$$\mu = (\mu_1, \mu_2, \ldots, \mu_k),$$

Fleiss tests the hypothesis that the $k$ $p$-component subvectors of $\mu$ are equal, i.e., that $\mu_1 = \mu_2 = \cdots = \mu_k$. 

5
This example can be extended to an example of block compound symmetry. Suppose a later study of the same $N$ subjects is made by $r$ different observers on $q$ tests. The data on each subject could be represented by a $kp + rq$ component row vector $X$, divided into two groups, $X_1: 1 \times kp$ and $X_2: 1 \times rq$. Assuming a covariance matrix $\Sigma$ of the form:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

$$\Sigma_{11} = \begin{pmatrix} A & B & \cdots & B \\ B & A & \cdots & B \\ \vdots & \vdots & \ddots & \vdots \\ B & B & \cdots & A \end{pmatrix}, \quad \Sigma_{22} = \begin{pmatrix} C & D & \cdots & D \\ D & C & \cdots & D \\ \vdots & \vdots & \ddots & \vdots \\ D & D & \cdots & C \end{pmatrix},$$

$$\Sigma_{12} = \begin{pmatrix} E \\ \vdots \\ E \end{pmatrix}, \quad \Sigma_{21} = \Sigma_{12}.$$

$\Sigma_{11}: pk \times pk$; $\Sigma_{22}: qr \times qr$; $\Sigma_{12}: pk \times qr$,

and a $\mu$ vector divided into two groups, $\mu_1: 1 \times kp$ and $\mu_2: 1 \times qr$, $\mu = (\mu_1, \mu_2)$ of the form:

$$\mu_1 = (\mu_{11}, \ldots, \mu_{1k}), \quad \mu_1: 1 \times p,$$

$$\mu_2 = (\mu_{21}, \ldots, \mu_{2r}), \quad \mu_2: 1 \times q,$$

we could test $\mu_{11} = \cdots = \mu_{1k}$ and $\mu_{21} = \cdots = \mu_{2r}$. We could also consider testing that the means are equal within each group and that the covariance matrix has the above block compound symmetry versus the alternative of an
unstructured (unrestricted) mean and covariance matrix.

Other applications to problems involving complete or compound symmetry in the area of testing and psychology include examples by Gulliksen (1950), Quereshi (1969), Gleser and Olkin (1973) and Olkin (1973). Applications to medical data include those of Votaw (1950) and Beeken (1962).

Type I compound symmetry involves several subsets of variates with the assumptions that within each subset the means are equal, variances are equal and covariances are equal and between subsets the covariances are equal. Type II compound symmetry involves several subsets of variates with equal numbers of elements in each subset. The assumption of equal numbers of elements in each subset means that the matrix of covariances between two subsets is square. We can then assume that between subsets, the diagonal covariances are equal and the off-diagonal covariances are equal. Thus the type II assumptions on these variates are that within each subset of variates the means are equal, variances are equal and covariances are equal and between subsets the diagonal covariances are equal and the off-diagonal covariances are equal.

In Section 2, the two types of block compound symmetry and six hypotheses plus a modified Hotelling's $T^2$ hypothesis for each type are explained. Section 3 contains two theorems which show how to transform the problems under consideration into canonical form. Also, the methods used to find the maximum likelihood estimators (MLE) under the various hypotheses, and to derive the likelihood ratio tests, (LRT), their approximate null distributions and their asymptotic non-null distributions are described in Section 3. Section 4 contains the results for the type I one-population
hypotheses and modified Hotelling's $T^2$ hypothesis. Section 5 contains
the results for the type I k-population hypotheses. Sections 6 and 7
contains the same information as do 4 and 5 pertaining to the type II
hypotheses. Finally Appendix I contains detailed calculations illustrating
the methods described at the end of Section 3 for calculating MLE, LRT,
moments of the LRT under the null distribution, an approximate distribution
of the LRT under the null hypothesis and the asymptotic non-null distri-
bution of the LRT for one of the hypotheses under consideration.
2. Types of Block Compound Symmetry and the Statement of the Hypotheses.

In this section the two types of block compound symmetry (denoted BCS-I and BCS-II) and six hypotheses for each type considered in the non-block case by Votaw (1948) are presented. The six hypotheses break into two groups of three, depending on whether we are examining the internal structure of one population or we are comparing \( k \) populations with block compound symmetric structure. We also test a 1-population modified Hotelling's \( T^2 \) hypothesis for each type.

First it is necessary to introduce some notation. In this chapter, all vectors are row vectors. The \( 1 \times k \) vector of all 1's is denoted by \( e_k \); \( J_{ij} = e_i'e_j \) is the \( i \times j \) matrix of all 1's, and \( I_n \) is the \( n \times n \) identity matrix.

The Kronecker product of two matrices is defined by:

\[
\text{(2.1)} \quad A \otimes B = \begin{pmatrix} a_{ij}B \end{pmatrix}.
\]

If \( A \) is an \( m \times n \) matrix and \( B \) is a \( p \times q \) matrix, then \( A \otimes B \) is a matrix of order \( mp \times nq \). We need several well-known facts concerning Kronecker products:

\[
\text{(2.2)} \quad (A_1 \otimes B_1)(A_2 \otimes B_2) = A_1A_2 \otimes B_1B_2, \\
\text{(2.3)} \quad A \otimes (B+C) = (A \otimes B) + (A \otimes C), \\
\text{(2.4)} \quad (A + B) \otimes C = (A \otimes C) + (B \otimes C), \\
\text{(2.5)} \quad (A \otimes B)' = A' \otimes B'.
\]
Let \( X_1, \ldots, X_N \) be a sample from a \( p \)-variate normal population with mean \( \mu \) and covariance matrix \( \Sigma \).

**Block Compound Symmetry, Type I (BCS-I):**

Under BCS-I, we assume that the \( p \) variates are grouped into \( b+q \) subgroups. The first \( b \) subgroups each contain one \( 1 \times m_i \) row vector, \( i = 1, \ldots, b \), while the next \( q \) groups each contain \( n_j \ 1 \times t_j \) vectors, \( j = 1, \ldots, q \). Thus a \( p \)-variate row vector is decomposed as

\[
P = (m_1, m_2, \ldots, m_b, t_1, \ldots, t_1, t_2, \ldots, t_2, \ldots, t_q, \ldots, t_q).
\]

Let

\[
m = \sum_{i=1}^{b} m_i, \quad t = \sum_{j=1}^{q} t_j, \quad g = \sum_{j=1}^{q} n_j t_j.
\]

Then \( p = m + g \). Under the assumption of BCS-I, the vectors within each subgroup are exchangeable, i.e., the distribution of \( X \) is assumed invariant over permutations of vectors within the same subgroup. These assumptions impose a certain structure on \( \mu \) and \( \Sigma \). Thus with respect to the vectors within each subgroup, the vectors of means are equal, the blocks of variances are equal and the blocks of covariances are equal within each subgroup. Between two distinct subgroups, the block covariances are equal.

More formally, we assume \( X_1, \ldots, X_N \) are of the form:

\[
X = (x_0, x_1, \ldots, x_q),
\]

where \( x_0 \) contains the first \( b \) subgroups, each subgroup consisting of
a $1 \times m_i$ vector, $i = 1, \ldots, b,$

$$x_0 = (x_{01}, \ldots, x_{0b}) \equiv (x_{011}, \ldots, x_{01m_1}, \ldots, x_{0bl}, \ldots, x_{0bm_b}),$$

and $x_1, \ldots, x_q$ are the remaining $q$ subgroups, each containing $n_j$ $t_j$-component row vectors, $j = 1, \ldots, q,$

$$x_j = (x_{j1}, \ldots, x_{jn_j}) \equiv (x_{j11}, \ldots, x_{j1t_j}, \ldots, x_{jn_j1}, \ldots, x_{jn_jt_j}), \quad j = 1, \ldots, q.$$  

Under the BCS-I, $\mu$ is of the form:

$$(2.8) \quad \mu = (\mu_0, \mu_1, \ldots, \mu_q),$$

where $\mu_0$ is an unrestricted $m$-component vector,

$$\mu_0 = (\mu_{01}, \ldots, \mu_{0b}) \equiv (\mu_{011}, \ldots, \mu_{01m_1}, \ldots, \mu_{0bl}, \ldots, \mu_{0bm_b}),$$

but $\mu_1, \ldots, \mu_q$ have the restriction:

$$\mu_j = (\delta_j, \ldots, \delta_j) = e_{n_j} \otimes \delta_j, \quad \delta_j = (\delta_{j1}, \ldots, \delta_{jt_j}), \quad j = 1, \ldots, q,$$

where each $\delta_j$ is an unrestricted $t_j$-component vector, $j = 1, \ldots, q.$

Under BCS-I, $\Sigma$ is of the form:
\[
\Sigma_0 = \left( \begin{array}{c|ccc}
\Sigma_{00} & \Sigma_{01} & \cdots & \Sigma_{0q} \\
\hline
\Sigma_{10} & \Sigma_{11} & \cdots & \Sigma_{1q} \\
\vdots & \vdots & \ddots & \vdots \\
\Sigma_{q0} & \Sigma_{q1} & \cdots & \Sigma_{qq}
\end{array} \right),
\]

\[
\Sigma_{ij} = \begin{bmatrix} C_{ij} & C_{ij} \\ C_{ij} & C_{ij} \end{bmatrix} = \begin{bmatrix} e_{ij} \otimes C_{ij} \\ e_{ij} \otimes C_{ij} \end{bmatrix}, \quad C_{ij}: \mathbb{M}_i \times t_j, \quad i=1, \ldots, b, \quad j=1, \ldots, q,
\]

\[
\Sigma_{00} = \left[ \left( \Sigma_{00} \right)_{ij} \right] = \left[ E_{ij} \right]: \mathbb{M}_i \times \mathbb{M}_j, \quad i,j=1, \ldots, b,
\]

\[
\Sigma_0 = \Sigma_{0j}, \quad j=1, \ldots, q,
\]

\[
\Sigma_{ii} = \begin{bmatrix} A_i & B_i & \cdots & B_i \\ B_i & A_i & \cdots & B_i \\ \vdots & \vdots & \ddots & \vdots \\ B_i & B_i & \cdots & A_i \end{bmatrix} = I_n \otimes (A_i - B_i) + J_n, \quad A_i: t_i \times t_i, \quad B_i: t_i \times t_i, \quad i=1, \ldots, q,
\]

\[
\Sigma_{ij} = \begin{bmatrix} D_{ij} & D_{ij} \\ \vdots & \vdots \\ D_{ij} & D_{ij} \end{bmatrix} = J_{n_i n_j} \otimes D_{ij}, \quad D_{ij}: t_i \times t_j, \quad i,j=1, \ldots, q, \quad i \neq j.
\]
BCS-I and the hypotheses:

We first note various type I hypotheses for testing the internal structure of one population:

\[ H_1^{(mvc)}: \mu \text{ and } \Sigma \text{ have BCS-I}, \]
\[ H_1^{(vc)}: \Sigma \text{ has BCS-I}, \]
\[ H_1^{(m|vc)}: \mu \text{ has BCS-I given } \Sigma \text{ has BCS-I}, \]
\[ H_1^{(\mu_0|vc)}: \mu = \mu_0 \text{ given } \Sigma \text{ has BCS-I}. \]

The first two hypotheses, \( H_1^{(mvc)} \) and \( H_1^{(vc)} \) are tested against the alternative \( H_{1,\Lambda}^{(mvc)}: \mu \text{ and } \Sigma \text{ unstructured.} \) The hypotheses \( H_1^{(m|vc)} \) and \( H_1^{(\mu_0|vc)} \) are tested against the alternative \( H_1^{(vc)} \).

The three type I hypotheses for comparing \( k \)-populations, each of which is assumed to have BCS-I in both its mean vector \( \mu_i \) and covariance matrix \( \Sigma_i, i=1,\ldots,k \), are:

\[ H_k^{(MVC|mvc)}: \mu_1 = \mu_2 = \cdots = \mu_k \text{ and } \Sigma_1 = \cdots = \Sigma_k \text{ given } \mu_i \text{ and } \Sigma_i \text{ have BCS-I, } i=1,\ldots,k; \]
\[ H_k^{(VC|mvc)}: \Sigma_1 = \cdots = \Sigma_k \text{ given } \mu_i \text{ and } \Sigma_i \text{ have BCS-I, } i=1,\ldots,k; \]
\[ H_k^{(M|MVC)}: \mu_1 = \cdots = \mu_k \text{ given } \mu_i \text{ and } \Sigma_i \text{ have BCS-I, } i=1,\ldots,k, \text{ and } \Sigma_1 = \cdots = \Sigma_k. \]

The first two hypotheses, \( H_k^{(MVC|mvc)} \) and \( H_k^{(VC|mvc)} \) are tested against the alternative:

\[ H_{k,\Lambda}^{(mvc)}: \mu_i \text{ and } \Sigma_i \text{ have BCS-I, } i=1,\ldots,k. \]

\( H_k^{(M|MVC)} \) is tested against the alternative \( H_k^{(VC|mvc)}. \)
Block Compound Symmetry, Type II (BCS-II):

Under BCS-II, we assume that the \( p \) variates are grouped into \( h \) subgroups. Each subgroup has \( n \) \( r \)-component row vectors. Thus, \( p = nrh \). Under the assumption of BCS-II, as in the case of BCS-I, the vectors within each subgroup are exchangeable, i.e., the distribution of \( X \) is assumed invariant over permutations of vectors within the same group. These assumptions impose a certain structure on \( \mu \) and \( \Sigma \). Thus with respect to the vectors within each subgroup, (as is the case for BCS-I), the vectors of means are equal, the blocks of variances are equal and the blocks of covariances are equal within each subgroup.

BCS-I and BCS-II differ in the condition on the block covariances between distinct subgroups. Under BCS-II, between two distinct subgroups the diagonal block covariances are equal and the off-diagonal block covariances are equal. We can make this further assumption under BCS-II because each subgroup has an equal number of \( n \) \( r \)-component row vectors.

More formally, we assume \( X_1, \ldots, X_N \) are of the form:

\[
(2.10) \quad X = (x_1, \ldots, x_h) ,
\]

decomposing \( X \) into \( h \) subgroups. Each subgroup contains \( n \) \( r \)-component row vectors, i.e.,

\[
x_i = (x_{i1}, \ldots, x_{in}) = (x_{i11}, \ldots, x_{i1r}, \ldots, x_{in1}, \ldots, x_{inr}), \quad i = 1, \ldots, h .
\]

Under BCS-II, \( \mu \) is of the form:

\[
(2.11) \quad \mu = (\mu_1, \ldots, \mu_h) ,
\]

where \( \mu_1, \ldots, \mu_h \) have the restriction:
\[ \mu_i = (\delta_i, \ldots, \delta_i) = e_n \otimes \delta_i, \quad \delta_i = (\delta_{i1}, \ldots, \delta_{ir}), \quad i=1, \ldots, h, \]

where each \( \delta_i \) is an unrestricted \( r \)-component vector, \( i=1, \ldots, h \).

Under BCS-II, \( \Sigma \) is of the form

(2.12) \[
\Sigma = \begin{pmatrix}
\Sigma_{11} & - & - & \Sigma_{1h} \\
- & \ddots & - & - \\
- & - & \ddots & - \\
- & - & - & \Sigma_{hh}
\end{pmatrix},
\]

\[
\Sigma_{ij} = \begin{pmatrix}
A_i \\
B_i \\
B_i \\
A_i
\end{pmatrix} = I_n \otimes (A_i - B_i) + J_{nn} \otimes B_i,
\]

\( A_i: r \times r, \quad B_i: r \times r, \quad i=1, \ldots, h, \)

\[
\Sigma_{ij} = \begin{pmatrix}
C_{ij} \\
D_{ij} \\
D_{ij} \\
C_{ij}
\end{pmatrix} = I_n \otimes (C_{ij} - D_{ij}) + J_{nn} \otimes D_{ij},
\]

\( C_{ij}: r \times r, \quad D_{ij}: r \times r, \quad i,j=1, \ldots, h, \quad i \neq j. \)

**BCS-II and the hypotheses:**

The hypotheses considered under BCS-II are the same as those considered under BCS-I with the obvious modification of all BCS-I structures to BCS-II structures. The BCS-II hypotheses will be denoted \( \overline{H}_i(mvc), \overline{H}_i(vc), \overline{H}_i(m|vc), \overline{H}_i(\mu_0|vc), \overline{H}_i(A), \overline{H}(MVC|mvc), \) etc.

Thus, as an example, \( \overline{H}_i(mvc) \) is the hypothesis that \( \mu \) and \( \Sigma \) have BCS-II.
3. **Canonical Forms and Estimation.**

In this section two theorems are introduced concerning the existence of orthogonal matrices which are used to transform BCS-I and BCS-II into canonical forms. We then work with the canonical forms to obtain maximum likelihood estimates (MLE) under the various models assumed. In the second part of this section, the problems are transformed to canonical forms and techniques used to find MLE, LRT, moments of the LRT, approximate null distribution of the LRT and asymptotic non-null distribution of the LRT, are discussed.

Theorem 3.1 guarantees the existence of an orthogonal matrix $C_I$.

This orthogonal matrix is used to transform a mean vector $\mu$ and covariance matrix $\Sigma$ which have BCS-I into the canonical forms, $v$ and $\Xi$, where $v = (v_0, 0) = N^{1/2} \mu C_I$ and $\Xi = C_I^T \Sigma C_I$ is block diagonal.

**Theorem 3.1.** If $\mu$ and $\Sigma$ have BCS-I structure, there exist:

1. An orthogonal $p \times p$ matrix, $C_I$, independent of $\mu$ and $\Sigma$,

2. $\Xi = \text{diag}(\Xi_0, I_{n_1} \otimes \Xi_{1}, \ldots, I_{n_q} \otimes \Xi_q)$, where $\Xi$ is a $p \times p$ matrix, $\Xi_0$ is an $r \times r$ matrix, $\Xi_j$ is a $t_j \times t_j$ matrix, $j=1, \ldots, q$,

(with $m = \sum_{i=1}^{b} m_i$, $t = \sum_{j=1}^{q} t_j$, $r = m + t$),

3. $v = (v_0, 0)$ where $v$ is $l \times p$ and $v_0$ is $l \times r$, such that
   a. $\Xi = C_I^T \Sigma C_I$,
   b. $\Sigma > 0$ if and only if $\Xi_i > 0$, $i=0, \ldots, q$,
   c. $v = N^{1/2} \mu C_I$.

If $\mu$ does not have BCS-I structure, but $\Sigma$ does, the above theorem remains true with condition 3 and conclusion c omitted.
Remark: The development in terms of the Kronecker product is due to Fleiss (1966) and Arnold (1970). Theorem 3.1 and its proof are essentially a theorem and proof of Arnold (1970, Chapter 4), and Arnold's techniques are used in the proof of Theorem 3.2. Both proofs are included in this section for completeness.

Before proving Theorem 3.1, some notation must be introduced. Let $U_k$ be a $k \times k$ orthogonal matrix with $k^{-1/2}e_k'$ as its first column. (Recall from Section 2, $e_k = (1, \ldots, 1)$). Let $F_{jk}$ be a $j \times k$ matrix with $\sqrt{jk}$ in the $(1,1)$ position, zeros elsewhere. The notation introduced in the beginning of Section 2 and properties of Kronecker products listed there are also used in the proofs of the theorems. In particular, the BCS-I forms of $\mu$ and $\Sigma$ are defined in Section 2, (2.8) and (2.9).

In addition, the following properties of $F_{jk}$ and the orthogonal matrix $U_k$ are needed:

\begin{equation}
U_k'U_k = I_k,
\end{equation}

\begin{equation}
e_{k'k} = \sqrt{k} (1,0,\ldots,0),
\end{equation}

\begin{equation}
U'_{m}e_{k}U = U'_{m}e'e_{n}U = (e_{m}e'_{m})' (e_{n}e_{n}) = F_{mn}.
\end{equation}

Proof of theorem 3.1.

Let $\Gamma = \text{diag}(I_m, U_1 \otimes I_{t_1}, U_2 \otimes I_{t_2}, \ldots, U_q \otimes I_{t_q})$ and $\Sigma = (\Sigma_{ij}) = \Gamma \Sigma \Gamma'$, where $\Sigma_{ij}$ is the same dimension as $\Sigma_{ij}$, $i,j = 0, \ldots, q$.

Then using the properties of Kronecker products and the properties listed in (3.1), (3.2) and (3.3), multiplication yields the relationships,
\[ \Sigma_{00} = \Sigma_{00}' = 1 \]

\[ \Sigma_{0j} = \Sigma_{0j}' = \Sigma_{0j}(U_{n_j} \otimes I_{t_j}) = \Sigma_{0j}(U_{n_j} \otimes I_{t_j}) \]

\[ = \left( \begin{array}{c} e_{n_j} \otimes C_{l_j} \\ \vdots \\ e_{n_j} \otimes C_{b_j} \end{array} \right) \left( \begin{array}{c} e_{n_j} U_{n_j} \otimes C_{l_j} I_{t_j} \\ \vdots \\ e_{n_j} U_{n_j} \otimes C_{b_j} I_{t_j} \end{array} \right) \]

\[ = \left( \sqrt{n_j}(1,0,\ldots,0) \otimes C_{l_j} \right) \left( \begin{array}{c} \sqrt{n_j} C_{l_j} 0 \\ \vdots \\ \sqrt{n_j} C_{b_j} 0 \end{array} \right), \quad j=1,\ldots,q, \]

\[ \Sigma_{j0} = \Sigma_{0j}', \quad j=2,\ldots,q, \]

\[ \Sigma_{jj} = (U_{n_j} \otimes I_{t_j})' \Sigma_{jj}' (U_{n_j} \otimes I_{t_j}) \]

\[ = (U_{n_j} \otimes I_{t_j})' \left( I_{n_j} \otimes (A_j-B_j) + n_j B_j \right)(U_{n_j} \otimes I_{t_j}) \]

\[ = I_{n_j} \otimes (A_j-B_j) + n_j B_j \]

\[ = \begin{pmatrix} A_j + (n_j-1)B_j & 0 \\ 0 & I_{n_j-1} \otimes (A_j-B_j) \end{pmatrix}, \quad j=1,\ldots,q, \]

\[ \Sigma_{ij} = (U_{n_i} \otimes I_{t_i})' \Sigma_{ij}' (U_{n_j} \otimes I_{t_j}) \]

\[ = (U_{n_i} \otimes I_{t_i})' \left( J_{n_i} I_{n_j} \otimes D_{ij} \right)(U_{n_j} \otimes I_{t_j}) \]

\[ = I_{n_i} D_{ij}, \quad i=1,\ldots,q, \quad i \neq j. \]
Look at the first $m^{n_1} t_1^{+n_2} t_2$ rows and $m^{n_1} t_1^{+n_2} t_2$ columns of $\bar{\Sigma}$,

$$
\begin{pmatrix}
\Sigma_{00} & \Sigma_{01} & \Sigma_{02} \\
\Sigma_{10} & \Sigma_{11} & \Sigma_{12} \\
\Sigma_{20} & \Sigma_{21} & \Sigma_{22}
\end{pmatrix}
$$

with $\bar{c}_{ij} = \sqrt{n_j} c_{ij}$, $i = 1, \ldots, b$, $j = 1, \ldots, q$, and $\bar{d}_{ij} = \sqrt{n_1 n_j} d_{ij}$, $i, j = 1, \ldots, q$, $i \neq j$. We seek an orthogonal rearrangement of these first $m^{n_1} t_1^{+n_2} t_2$ rows and columns so that they are in block diagonal form. There is a permutation matrix $P_1$ such that $P_1 \bar{\Sigma} P_1^t$ leaves all but the above corner (the first $m^{n_1} t_1^{+n_2} t_2$ rows and $m^{n_1} t_1^{+n_2} t_2$ columns of $\bar{\Sigma}$) unchanged and the above corner becomes $\text{diag}(\Lambda_0, \Lambda_1, \Lambda_2)$, where
\[
\Lambda_0 = \begin{pmatrix}
E & \overline{C}_{11} & \overline{C}_{12} \\
\vdots & \vdots & \vdots \\
\overline{C}_{b1} & \overline{C}_{b2} \\
\overline{C}_{11} & \cdots & \overline{C}_{b1} \\
\overline{C}_{12} & \cdots & \overline{C}_{b2} \\
A_1(n_1-1)B_1 & \overline{D}_{12} \\
\overline{D}_{21} & A_2(n_2-1)B_2
\end{pmatrix},
\]

\[\Lambda_i = I_{n_i-1} \otimes (A_i - B_i), \quad i=1,2.\]

In a similar way, we can find a permutation matrix \(P_I\) to rearrange all of \(\Xi\) so that:

\[P'_I \Gamma_I \Sigma_I P_I = P'_I \Sigma\Gamma_I P_I = \text{diag}(\Xi_0, I_{n_1-1} \otimes \Xi_1, \ldots, I_{n_q-1} \otimes \Xi_q),\]

\[\Xi_0 = (\xi_{i,j}), \quad i,j=0,\ldots,q,
\]

\[\xi_{00} = E, \quad \xi_{j0} = (\overline{C}_{lj}, \ldots, \overline{C}_{bj}), \quad \xi_{0j} = \xi_{j0}, \quad j=1,\ldots,q,
\]

\[\xi_{ij} = A_1(n_1-1)B_1, \quad i=1,\ldots,q, \quad \xi_{ij} = \overline{D}_{ij}, \quad i,j=1,\ldots,q, \quad i \neq j,
\]

\[\Xi_i = A_i - B_i, \quad i=1,\ldots,q.
\]

Let \(C_I = \Gamma_I P_I\). This completes the proof of \(a\) and \(b\).

To complete the proof, we must verify result \(c\). To do this, it is straightforward to investigate \(N^{1/2} \mu C_I\). First, let

\[
\phi = \mu C_I = (\mu_0, \ldots, \mu_q) \Gamma_I = (\phi_0, \ldots, \phi_q),
\]

\[\phi_0 = \mu_0 I_m.\]

\[\phi_1 = \mu_1 (U_{n_1} \otimes I_{t_1}) = (e_{n_1} \otimes \delta_i)(U_{n_1} \otimes I_{t_1})\]

\[= e_{n_1} U_{n_1} \otimes \delta_i I_{t_1} = (\sqrt{n_1} \delta_i, 0), \quad i=1,\ldots,q.
\]
Consider the first $m_1 t_1 + n_2 t_2$ components of $\phi$,

$$(\phi_0, \phi_1, \phi_2) = (\mu_0, \sqrt{n_1 s_1}, 0, \sqrt{n_2 s_2}, 0).$$

Observe that the permutation matrix $P_1$ used in the first part of the proof leaves all but these first $m_1 t_1 + n_2 t_2$ components unchanged under the transformation $\phi P_1$ and these first $m_1 t_1 + n_2 t_2$ components become $(\mu_0, \sqrt{n_1 s_1}, \sqrt{n_2 s_2}, 0, 0)$. Similarly, the $P_I$ matrix considered above transforms $\phi$ to:

$$v = N^{1/2} \mu C_I = N^{1/2} \mu P_1 P_I = N^{1/2} \phi P_I = (v_0', 0),$$

$$v_0 = N^{1/2}(\mu_0, \sqrt{n_1 s_1}, \ldots, \sqrt{n_q s_q}), v_0: 1 \times r, (r = \sum_{i=1}^b m_i + \sum_{j=1}^q t_j).$$

This completes the proof of Theorem 3.1.

The following theorem shows that $\mu$ and $\Sigma$ with BCS-II structure can be transformed to canonical form by proving the existence of an orthogonal matrix $C_{II}$ with the property that $v = (v_1', 0) = N^{1/2} \mu C_{II}$ and $\Xi = C_{II}' \Sigma C_{II}$ is block diagonal.

**Theorem 3.2.** If $\mu$ and $\Sigma$ have BCS-II structure, there exist:

1. An orthogonal $p \times p$ matrix, $C_{II}$, independent of $\mu$ and $\Sigma$.

2. $\Xi = \text{diag}(\Xi_1, \ldots, \Xi_2)$, where $\Xi$ is a $p \times p$ matrix and $\Xi_1$ and $\Xi_2$ are $rh \times rh$ matrices,

3. $v = (v_1, 0)$ where $v$ is $1 \times p$ and $v_1$ is $1 \times rh$, such that:
   a. $\Xi = C_{II}' \Sigma C_{II}$,
   b. $\Sigma > 0$ if and only if $\Xi_1 > 0$ and $\Xi_2 > 0$,
   c. $v = N^{1/2} \mu C_{II}$.

If $\mu$ does not have BCS-II structure, but $\Sigma$ does, the above theorem remains true with condition 3 and conclusion c omitted.
Note: The BCS-II forms of $\mu$ and $\Sigma$ are defined in Section 2, (2.11) and (2.12).

**Proof of theorem 3.2.**

Let $\Gamma_{II} = I_h \otimes (U_n \otimes I_r)$ and $\Sigma_{ij} = \Gamma_{II} \Sigma_{II}$, $i, j = 1, \ldots, h$. Then using the properties of Kronecker products and the properties listed in (3.1), (3.2) and (3.3), multiplication yields the relationships:

$$
\Sigma_{ii} = (U_n \otimes I_r)' \Sigma_{ii} (U_n \otimes I_r) \\
= (U_n' \otimes I_r') [I_n \otimes (A_i - B_i) + J_{n,n} \otimes B_i] (U_n \otimes I_r) \\
= I_n \otimes (A_i - B_i) + F_{n,n} \otimes B_i \\
= \begin{pmatrix}
A_i + (n-1)B_i & 0 \\
0 & I_{n-1} \otimes (A_i - B_i)
\end{pmatrix}, \quad i = 1, \ldots, h,
$$

$$
\Sigma_{ij} = (U_n \otimes I_r)' \Sigma_{ij} (U_n \otimes I_r) \\
= (U_n' \otimes I_r') [I_n \otimes (C_{ij} - D_{ij}) + J_{n,n} \otimes D_{ij}] (U_n \otimes I_r) \\
= \begin{pmatrix}
C_{ij} + (n-1)D_{ij} & 0 \\
0 & I_{n-1} \otimes (C_{ij} - D_{ij})
\end{pmatrix}, \quad i, j = 1, \ldots, h, \quad i \neq j.
$$

Now look at the first $2rn$ rows and $2rn$ columns of $\Sigma$:

$$
\begin{pmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{pmatrix} = 
\begin{bmatrix}
A_1 + (n-1)B_1 & 0 & C_{12} + (n-1)D_{12} & 0 \\
0 & I_{n-1} \otimes (A_1 - B_1) & 0 & I_{n-1} \otimes (C_{12} - D_{12}) \\
C_{21} + (n-1)D_{21} & 0 & A_2 + (n-1)B_2 & 0 \\
0 & I_{n-1} \otimes (C_{21} - D_{21}) & 0 & I_{n-1} \otimes (A_2 - B_2)
\end{bmatrix}.
$$
We seek an orthogonal rearrangement of these first $2rn$ rows and $2rn$ columns of $\Sigma$ so that they are in block diagonal form. There is a permutation matrix $P_2$ such that $P_2^T \Sigma P_2$ leaves all but the $2rn \times 2rn$ upper left corner of $\Sigma$ fixed and that corner becomes $\text{diag}(\Lambda_1, \Lambda_2)$ where:

$$\Lambda_1 = \begin{pmatrix} A_1 + (n-1)B_1 & C_{12} + (n-1)D_{12} \\ C_{21} + (n-1)D_{21} & A_2 + (n-1)B_2 \end{pmatrix},$$

$$\Lambda_2 = I_{n-1} \otimes \begin{pmatrix} A_1 - B_1 & C_{12} - D_{12} \\ C_{21} - D_{21} & A_2 - B_2 \end{pmatrix}. $$

In a similar way, we can find a permutation matrix $P_{II}$ to rearrange all of $\Sigma$ so that:

$$P_{II}^T \Sigma_P P_{II} = P_{II}^T \Sigma P_{II} = \text{diag}(\Xi_1, I_{n-1} \otimes \Xi_2),$$

$$\Xi_k = (\xi_{ij}^k), \ i,j=1,\ldots,h, \ k=1,2,$$

$$\xi_{ii}^k = A_i + (n-1)B_i, \ i=1,\ldots,h,$$

$$\xi_{ij}^k = C_{ij} + (n-1)D_{ij}, \ i,j=1,\ldots,h, \ i \neq j,$$

$$\xi_{ii}^2 = A_i - B_i, \ i=1,\ldots,h,$$

$$\xi_{ij}^2 = C_{ij} - D_{ij}, \ i,j=1,\ldots,h, \ i \neq j.$$

Letting $C_{II} = \Gamma_{II} P_{II}^T$, the proof of conditions a and b of the theorem is completed.

To prove part c, we must investigate $N^{1/2} \mu C_{II}$. Define $\phi$ by:

$$\phi = \mu_{II} = (e_n \otimes \delta_1, \ldots, e_n \otimes \delta_h)[I_h \otimes (U_n \otimes I_r)]$$

$$= (\phi_1, \ldots, \phi_h),$$

$$\phi_i = (e_n \otimes \delta_i)(U_n \otimes I_r) = (e_n \otimes \delta_i \otimes I_r) = (\sqrt{n} \delta_i, 0), \ i=1,\ldots,h.$$

23
Consider the first $2\eta\eta$ components of $\phi$:

$$(\phi_1, \phi_2) = (\sqrt{\eta} \delta_1, 0, \sqrt{\eta} \delta_2, 0).$$

Observe that the permutation matrix $P_2$ used above leaves all but the first $2\eta\eta$ components of $\phi$ unchanged under the transformation $\phi P_2$ and these $2\eta\eta$ components become $(\sqrt{\eta} \delta_1, \sqrt{\eta} \delta_2, 0, 0)$. The permutation matrix $P_{II}$ used above results in:

$$\nu = N^{1/2} \mu_{II} = N^{1/2} \mu_{II} P_{II} = N^{1/2} \phi P_{II} = (\nu_1, 0),$$

$$\nu_1 = N^{1/2} \sqrt{\eta} (\delta_1, \ldots, \delta_k), \nu_1: 1 \times h, \delta_i: 1 \times r, i=1, \ldots, h.$$

This completes the proof of Theorem 3.2.

Next we use the results of these theorems to reduce the problems of estimation and hypothesis testing to easily manageable forms. We do this by using the principle of sufficiency and by orthogonal rotations.

Given a sample $x_1, \ldots, x_N$ from a $p$-variate $N(\mu, \Sigma)$ distribution, we may (by sufficiency) consider the mean vector $\bar{x} = \frac{1}{N} \sum_{\alpha=1}^{N} x_\alpha$ and the sample cross-product matrix $A = \sum_{\alpha=1}^{N} (x_\alpha - \bar{x}) (x_\alpha - \bar{x})^t$, $m = N-1$, as our starting point. Denoting $\mathcal{L}(Z)$ the law of the random matrix (or vector), $Z$, we note that $\bar{x}$ and $A$ are independently distributed, with

$$\chi^2(N^{1/2} \bar{x}) = \eta(N^{1/2} \mu, \Sigma), \mathcal{L}(A) = W(\Sigma, p, m), m = N-1,$$

i.e., $A$ has a density function:

$$p(A) = c(p, m)|\Sigma|^{-m/2}|A|^{-\frac{m-p-1}{2}} \exp(-\frac{1}{2} \text{tr} \Sigma^{-1} A), A > 0, \Sigma > 0,$$
(3.6) \[ c(p,m) = 2^{-p/2 \pi^{-p(p-1)/4}} \prod_{i=1}^{p} \frac{1}{i! \left( \frac{1}{2}(m-i+1) \right)}^{-1}. \]

We now transform to canonical forms, combining the analysis for both BCS-I and BCS-II. Let:

(3.7) \[ y = N^{1/2 \mu} \mathbf{c}, \quad V = C'AC, \]

where \( C \) is \( C_I \) in Theorem 3.1 (BCS-I) or \( C_{II} \) in Theorem 3.2 (BCS-II). Then \( y \) and \( V \) are independently distributed with:

(3.8) \[ \mathcal{L}(y) = \eta(v, \Xi), \quad \mathcal{L}(V) = \mathcal{W}(\Xi, p, m), \]
\[ v = N^{1/2 \mu} \mathbf{c}, \quad \Xi = C'EC. \]

The forms of \( v \) and \( \Xi \) under BCS structure have been given in theorems 3.1 and 3.2.

The hypotheses to be tested are also transformed. The assumptions that \( \mu \) and/or \( \Sigma \) have BCS-I(II) are the same as \( v \) and/or \( \Xi \) having the form given in Theorem 3.1 (3.2). If \( \mu \) is assumed to be an unstructured or specified mean vector, then \( v \) is also an unstructured or specified mean vector. If \( \Sigma \) is an unstructured covariance matrix, then \( \Xi \) is an unstructured covariance matrix. Making these changes to the hypotheses in Section 2, we may solve either of the equivalent problems using \( \mathbf{x}, A, \mu, \Sigma \) and the hypotheses in Section 2 or \( y, V, v, \Xi \) and the hypotheses of Section 2 transformed as just described.

For the remainder of this paper, we work only with the canonical forms of the problem. Maximum likelihood estimates (MLE) and likelihood ratio tests (LRT) are all functions of \( y \) and \( V \). To get their forms
in terms of $\bar{X}$ and $A$, one need only use the transformations $\bar{X} = N^{-1/2} \gamma C'$ and $A = C V C'$. For any distributional results involving $\nu$ and $\Sigma$ one may substitute $\mu = N^{-1/2} \gamma C'$ and $\Sigma = C \Sigma C'$ to get their forms in terms of $\mu$ and $\Sigma$.

Using the canonical forms we find the MLE in four regions:

$\omega_1$: $\nu$ and $\Sigma$ structured as in Theorem 3.1 (3.2),

$\omega_2$: $\Sigma$ structured as in Theorem 3.1 (3.2), $\nu$ unstructured,

$\omega_3$: $\Sigma$ structured as in Theorem 3.1 (3.2), $\nu = 0$

$\omega_4$: $\nu$ and $\Sigma$ unstructured.

Note that we have reduced the modified Hotelling's $T^2$ problem from $H_1(\mu_0|\nu C)$ to $\Sigma$ structured as in Theorem 3.1 (3.2), $\nu = \nu_0 = N^{1/2} \mu_0 C I(II)$. We make the further transformation to the sufficient statistics $y$ and $V$ by setting $z = y - \nu_0$ and leaving $V$ unchanged. This testing problem is then equivalent to the one with $y$ and $V$ where it is assumed $\nu_0 = 0$. We thus assume $\nu_0 = 0$ and test $\omega_3$ versus $\omega_4$ as listed above.

To find the MLE, we use the following lemmas which contain standard multivariate normal results. (See Appendix I for an application of these lemmas).

**Lemma 3.1.** Let $x_1, \ldots, x_N$ be a sample of size $N$ from a $p$-variate normal population, $N(\mu, \Sigma)$. Then:

1. $\bar{x} = \frac{1}{N} \sum_{\alpha=1}^{N} x_\alpha$ and $A = \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})'(x_\alpha - \bar{x})$ are independent.

2. $\mathcal{L}(N^{1/2}(\bar{x})) = \mathcal{C}(N^{1/2} \mu, \Sigma)$, $\mathcal{L}(A) = \mathcal{W}(\Sigma, p, N-1)$

26
3. The MLE of $\mu$ and $\Sigma$, denoted $\hat{\mu}$ and $\hat{\Sigma}$ are $\hat{\mu} = \bar{x}$ and $\hat{\Sigma} = A/N$.

**Lemma 3.2.** If $y$ and $A$ are independently distributed, $L(y) = \eta(y, \Sigma)$, $L(A) = W(\Sigma, p, n)$, then the MLE are $\hat{\nu} = y$ and $\hat{\Sigma} = A/(n+1)$.

**Lemma 3.3.** If $L(A) = W(\Sigma, p, n)$, then the MLE of $\Sigma$ is $\hat{\Sigma} = A/n$.

Once we have the MLE's, we can form the likelihood ratio test (LRT) for testing $H: \omega_i$ vs. $A: \omega_j$ with $\omega_i \subsetneq \omega_j$ by:

$$
\lambda_{ij} = \frac{\sup_{\nu, \Xi \in \omega_i} L(\nu, \Xi)}{\sup_{\nu, \Xi \in \omega_j} L(\nu, \Xi)}
$$

The moments of the LRT can be found under the null hypothesis, here assumed to be the more restricted parameter space $(\omega_i \subsetneq \omega_j)$, by taking advantage of the structure under the null hypothesis. The techniques used are similar to those in Anderson (1958, p. 235). An example for the case of testing $H_0$ (mvv) appears in Appendix I.

Once the moments of the LRT have been found one can use the result of Box (1949) to form an approximate distribution under the null hypothesis of $-2 \log \lambda$ in terms of a linear combination of central chi-squared variates. Anderson (1958, p. 203) has expressed this result of Box as follows:

**Lemma 3.4.** Consider the random variable $W(0 \leq W \leq 1)$ with $\eta$-th moment:

27
\[ \mathcal{E}_W^h = K \left( \prod_{k=1}^{a} x_k \prod_{j=1}^{b} y_j \right)^h \frac{\prod_{k=1}^{a} \Gamma(x_k(1+h)+\epsilon_k)}{\prod_{k=1}^{a} \Gamma(x_k(1+h)+\epsilon_k)} , \quad h=0,1, \ldots, \]

where \( K \) is a constant (such that \( \mathcal{E}_W^0 = 1 \)), \( \Gamma \) is the gamma function,

\[ \sum_{k=1}^{a} x_k = \sum_{j=1}^{b} y_j . \]

Then,

\[ P(-2 \log W \leq z) = (1-\omega) P\left\{ x_f^2 \leq \rho z \right\} + \omega P\left\{ x_{f+4}^2 \leq \rho z \right\} + O(N^{-3}) , \]

\[ f = -2[\sum_{k \leq 1} \sum_{j \leq 4} (a-b)/2] , \]

\( \rho \) is the solution of:

\[ \sum_k \frac{B_2(\beta_k+\epsilon_k)}{x_k} = \sum_j \frac{B_2(\epsilon_j+\eta_j)}{y_j} , \]

\[ \beta_k = (1-\rho)x_k , \quad \epsilon_j = (1-\rho)y_j , \]

\( B_2 \) is the Bernoulli polynomial of degree 2, \( B_2(h) = h^2 - h + 1/6 \), and

\[ -6\omega = \sum_k \frac{B_3(\beta_k+\epsilon_k)}{(\rho x_k)^2} - \sum_j \frac{B_3(\epsilon_j+\eta_j)}{(\rho y_j)^2} , \]

where \( B_3 \) is the Bernoulli polynomial of degree 3, \( B_3(h) = h^3 - (3/2)h^2 + \frac{1}{2} h \).

It can be shown that, under the null hypothesis, as the sample size increases, \( \rho \) approaches one and \( \omega \) approaches zero, i.e.,

\[ \lim_{N \to \infty} P\left\{ -2 \log \lambda \leq z \right\} = P\left\{ X_f^2 \leq z \right\} . \]
A detailed example in which the calculations are done is included in Appendix I for $\overline{H}_1(\text{mvc})$.

To obtain the asymptotic distribution of $-(2/N)\log \lambda$, where $\lambda$ is a LRT, under the alternative hypothesis, we use the standard delta method. (See Cramér 1946, pp. 366). Note $\lambda = \lambda(y, V)$ where under the alternative hypothesis $\mathcal{L}(y) = \mathcal{N}(\sqrt{N} \varphi, \Sigma)$, $\mathcal{L}(V) = \mathcal{W}(\Xi, \rho, n)$, with the mean vector $\varphi$ and covariance matrix $\Sigma$ unrestricted. First, make the transformations, $y = \sqrt{N} (z \pm \frac{1}{2} + \varphi)$, $V = N \Xi \pm \frac{1}{2} \rho^{-1} \Xi \pm \frac{1}{2} \Xi^{-1}$. Then $\lambda = \lambda(z, W)$ where $\mathcal{L}(\sqrt{N} z) = \mathcal{N}(0, I)$, $\mathcal{L}(NW) = \mathcal{W}(I, \rho, n)$. Apply the standard delta method to $h(z, W) = -(2/N)\log \lambda(z, W)$.

Under suitable regularity conditions, which are satisfied in this problem, we use the result of Madansky and Olkin (1969)

$$\mathcal{L}[\sqrt{N} (h(z, W) - h(0, I))] \rightarrow \mathcal{N}(0, 2 \text{ tr } H^2 + dd') ,$$

as $N \rightarrow \infty$, where

$$H = (h_{ij}) = \frac{1}{2} \frac{\partial^2 h}{\partial w_{ij}} \bigg|_{W=I, z=0} \quad \text{for } i \neq j; \quad h_{ii} = \frac{\partial h}{\partial w_{ii}} \bigg|_{W=I, z=0}$$

and the row vector

$$d = (d_1) = \frac{\partial h}{\partial z_1} \bigg|_{W=I, z=0} .$$

The following lemma is used in the evaluation of these matrix derivatives. (See Appendix I for an application of this lemma).

**Lemma 3.5.** Let $R = AXB + CX'D + EX'XFY$ where $A, B, C, D, E, F$ and $Y$ are matrices independent of the matrix $X$. Then,

$$\frac{\partial \log |R|}{\partial X} = A'R^{-1}B' + DR^{-1}C' + XFR^{-1}E + XB'R^{-1}F' .$$
The proof of this lemma uses three well-known facts:

(3.11) \[ \frac{\partial \log |X|}{\partial x} = (X^{-1})' \]

ii. If \( w = w(Y(X)) \), \( w \) a real-valued function, \( Y \) a matrix function of the matrix \( X \), then

(3.12) \[ \frac{\partial w}{\partial x_{\alpha \beta}} = \sum_{i,j} \frac{\partial w}{\partial y_{ij}} \frac{\partial y_{ij}}{\partial x_{\alpha \beta}} \]

iii. If \( Y \) is a matrix function of the matrix \( X \), and \( A \) and \( B \) are independent of \( X \), then

(3.13) \[ \text{tr} \frac{\partial (AYB)}{\partial x_{\alpha \beta}} = \text{tr} BA \frac{\partial Y}{\partial x_{\alpha \beta}} \]

Proof of Lemma 3.5.

Let \( R^{-1} = (r_{ij}) \). Then using (3.11) and (3.12),

(3.14) \[ \frac{\partial \log |R|}{\partial x_{\alpha \beta}} = \sum_{i,j} \frac{\partial \log |R|}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial x_{\alpha \beta}} = \sum_{i,j} r_{ij} \frac{\partial r_{ij}}{\partial x_{\alpha \beta}} = \text{tr} R^{-1} \frac{\partial R}{\partial x_{\alpha \beta}} \]

(3.15) \[ \frac{\partial R}{\partial x_{\alpha \beta}} = \frac{\partial}{\partial x_{\alpha \beta}} (AXB + CX'D + EX'XF + Y) \]

Substituting (3.15) into (3.14), noting that \( \frac{\partial X}{\partial x_{\alpha \beta}} = 0 \) we obtain

(3.16) \[ \frac{\partial \log |R|}{\partial x_{\alpha \beta}} = \text{tr} R^{-1} \frac{\partial (AYB)}{\partial x_{\alpha \beta}} + \text{tr} R^{-1} \frac{\partial (DX'D)}{\partial x_{\alpha \beta}} + \text{tr} R^{-1} \frac{\partial (EX'XF)}{\partial x_{\alpha \beta}} \]

\[ = \text{tr} BR^{-1} A \frac{\partial X}{\partial x_{\alpha \beta}} + \text{tr} DR^{-1} C \frac{\partial X'}{\partial x_{\alpha \beta}} + \text{tr} FR^{-1} E \frac{\partial (X'X)}{\partial x_{\alpha \beta}} \]
The second term may be rewritten as

\[(3.17) \quad \text{tr} \, DR^{-1}C \, \frac{\partial x'}{\partial x_{\alpha \beta}} = \text{tr}(DR^{-1}C)^T \, \frac{\partial x}{\partial x_{\alpha \beta}}.\]

Using \( \frac{\partial (x'X)}{\partial x_{\alpha \beta}} = (\frac{\partial x'}{\partial x_{\alpha \beta}})x + x'(\frac{\partial x}{\partial x_{\alpha \beta}}) \), the third term may be rewritten as

\[(3.18) \quad \text{tr} \, FR^{-1}E \, \frac{\partial (X'X)}{\partial x_{\alpha \beta}} = \text{tr}[XFR^{-1}E + XE'R^{-1}F'], \frac{\partial x}{\partial x_{\alpha \beta}}.\]

Substituting the results of (3.17) and (3.18) into (3.16) yields

\[(3.19) \quad \frac{\partial \log |R|}{\partial x_{\alpha \beta}} = \text{tr}[A'R^{-1}B' + DR^{-1}C + XFR^{-1}E + XE'R^{-1}F'], \frac{\partial x}{\partial x_{\alpha \beta}}\]

\[= (A'R^{-1}B' + DR^{-1}C + XFR^{-1}E + XE'R^{-1}F')_{\alpha \beta}.\]

The matrix result follows immediately from this elementwise result.\]

Note that it follows easily from that proof of Lemma 3.5 that if

\[(3.20) \quad R = \sum_i A_i X B_i + \sum_j C_j X' D_j + \sum_k E_k X' F_k,\]

then

\[(3.21) \quad \frac{\partial \log |R|}{\partial x} = \sum_i A_i'R^{-1}B'_i + \sum_j D_j R^{-1}C'_j + \sum_k (XFR^{-1}E + XE'R^{-1}F').\]
4. **Type I One-population Hypotheses: LRT and Approximate Null and Asymptotic Non-null Distributions.**

We start with the canonical forms introduced in Section 3, (3.7) and (3.8). The canonical forms \( y \) and \( V \) are independently distributed with their distributions given by:

\[
(4.1) \quad \mathcal{L}(y) = \mathcal{N}(\nu, \Sigma), \quad \mathcal{L}(V) = \mathcal{W}(\Sigma, p, n-1).
\]

Under BCS-I, \( \nu \) and \( \Sigma \) have the following special structure (see Theorem 3.1):

\[
(4.2) \quad m = \sum_{i=1}^{b} m_i, \quad t = \sum_{j=1}^{g} t_j, \quad g = \sum_{j=1}^{g} t_j n_j, \quad p = m + g, \quad r = m + t,
\]

we have,

\[
(4.3) \quad \nu = (\nu_0, 0), \quad \nu_0: 1 \times (m + t),
\]

\[
(4.4) \quad \Sigma = \text{diag}(\Sigma_0, I_{n_1-1} \otimes \Sigma_1, \ldots, I_{n_q-1} \otimes \Sigma_q),
\]

\[
\Sigma_0: r \times r, \quad \Sigma_j: t_j \times t_j, \quad j = 1, \ldots, q.
\]

Next, we identify special blocks of \( y \) and \( V \) consistent with the special structure of \( \nu \) and \( \Sigma \) under BCS-I. Decompose the \( p \)-variate row vector \( y \) by

\[
(4.5) \quad y = (y_0, y_1, \ldots, y_q),
\]

\( y_0: 1 \times (m + t), \)

\( y_1 = (y_{i1}, \ldots, y_{in_1-1}), \quad y_i: 1 \times (n_i - 1) t_i, \quad i = 1, \ldots, q. \)
Decompose the \( p \times p \) matrix \( V \) into \( V = (V_{ij}) \), \( i,j=0,\ldots,q \), where

\[(4.6) \quad \begin{align*}
V_{00} & \text{ is } (m+t) \times (m+t) , \\
V_{0j} & \text{ is } (m+t) \times (n_j-1)t_j , \quad j=1,\ldots,q , \\
V_{j0} & \text{ is } (n_j-1)t_j \times (m+t) , \quad j=1,\ldots,q , \\
V_{ij} & \text{ is } (n_i-1)t_i \times (n_j-1)t_j , \quad i,j=1,\ldots,q .
\end{align*} \]

Also decompose \( V_{ii} \) into \( (V_{ii})_{jj} \), \( i=1,\ldots,q \), \( j=1,\ldots,n_i-1 \), where
\( (V_{ii})_{jj} \), \( j=1,\ldots,n_i-1 \) are \( t_i \times t_i \) blocks on the diagonal of \( V_{ii} \), \( i=1,\ldots,q \).

Because under BCS-I, the mean vector \( \nu \) has some elements identically zero, we use \( y \) in the estimation of \( \Xi_i \), \( i=1,\ldots,q \), and in the modified Hotelling's \( T^2 \) problem. Thus define \( D \) and \( H_i \) by

\[(4.7) \quad D = y'y , \quad H_i = \sum_{j=1}^{n_i-1} y_{ij}'y_{ij} , \quad i=1,\ldots,q .\]

Decompose \( D \) similarly to the decomposition of \( V \).

Maxima of the likelihood functions.

Using Lemma 3.1, we can find the maxima of the likelihood functions \( L(y,V) \) over the regions

\[(4.8) \quad \omega_1: \nu \text{ and } \Xi \text{ have BCS-I structure}, \]

\[(4.9) \quad \omega_2: \Xi \text{ has BCS-I structure, } \nu \text{ unrestricted}, \]

\[(4.10) \quad \omega_3: \Xi \text{ has BCS-I structure, } \nu = 0 ,\]

33
\[ \omega_q: \ \nu \text{ and } \Xi \text{ unrestricted} . \]

\[ \text{Let } b(p, N) = N^{pN/2} \left( 2\pi e \right)^{-pN/2} . \]

Then the maxima of the likelihood functions are

\[ \sup_{\nu, \Xi \in \omega_1} L(y, V) = \frac{b(p, N)}{|V_{oo}|^{N/2}} \prod_{i=1}^{q} \left\{ \frac{(n_i - 1)^{t_i}}{|H_i + \sum_{j=1}^{n_i-1} (V_{ii})_{jj}|} \right\}^{(n_i - 1)N/2} \]

\[ \sup_{\nu, \Xi \in \omega_2} L(y, V) = \frac{b(p, N)}{|V_{oo}|^{N/2}} \prod_{i=1}^{q} \left\{ \frac{(n_i - 1)^{t_i}}{\sum_{j=1}^{n_i-1} (V_{ii})_{jj}} \right\}^{(n_i - 1)N/2} \]

\[ \sup_{\nu, \Xi \in \omega_3} L(y, V) = \frac{b(p, N)}{|D_{oo} + V_{oo}|^{N/2}} \prod_{i=1}^{q} \left\{ \frac{(n_i - 1)^{t_i}}{\sum_{j=1}^{n_i-1} [(D_{ii})_{jj} + (V_{ii})_{jj}]} \right\}^{(n_i - 1)N/2} \]

\[ \sup_{\nu, \Xi \in \omega_4} L(y, V) = \frac{b(p, N)}{|V|^{N/2}} \]

We may now generate the LRT for the type I one-population hypotheses using:

\[ \lambda_{ij} = \frac{\sup_{\nu, \Xi \in \omega_i} L(y, V)}{\sup_{\nu, \Xi \in \omega_j} L(y, V)} . \]

The LRT, moments of the LRT under the null distribution, and asymptotic non-null distributions follow for each of the one-population type I hypotheses. Methods used to derive them are explained at the end of Section 3 and are illustrated in Appendix I for $\overline{H}_1(mvc)$. The moments
of the LRT and its approximate null distribution do not depend on \( \nu \) or \( \Xi \). The asymptotic non-null distributions do depend on \( \nu \) and \( \Xi \). We therefore need to define the following structures for \( \nu \) and \( \Xi \) used in the non-null distributions.

\[
(4.18) \quad \nu = (\nu_{01}; \nu_{11}, \ldots, \nu_{1l-1}; \nu_{21}, \ldots, \nu_{2n_1-1}; \ldots; \nu_{q1}, \ldots, \nu_{q_n-1}) ;
\]
\[
\nu: \ 1 \times p ; \nu_{01}: 1 \times r ;
\]
\[
\nu_{ij}: 1 \times t_i , \ i=1, \ldots, q; \ j=1, \ldots, n_i - 1 ,
\]
\[
\tilde{\nu} = N^{-1/2} \nu .
\]

The decomposition of \( \Xi \) is the same as that of \( \nu \):

\[
(4.19) \quad \Xi = [\Xi_{ij}], \ i,j=0,1, \ldots, q ,
\]
\[
\Xi: p \times p , \Xi_{00}: r \times r ,
\]
\[
\Xi_{0i}: r \times (n_i - 1)t_i ; \Xi_{i0}: (n_i - 1)t_i \times r , \ i=1, \ldots, q ,
\]
\[
\Xi_{ij}: (n_i - 1)t_i \times (n_j - 1)t_j , \ i,j=1, \ldots, q .
\]

We further decompose each of the blocks \( \Xi_{ij} \) defined in (4.19) into the following subblocks:

\[
(4.20) \quad \Xi_{ij} = [(\Xi_{ij})_{uv}] , \ i,j=0, \ldots, q; \ u=1, \ldots, n_i - 1 ; \ v=1, \ldots, n_j - 1 ;
\]
\[
(\Xi_{00})_{11} = \Xi_{00} : r \times r
\]
\[
(\Xi_{0j})_{1v} : r \times t_j ; (\Xi_{j0})_{1v} : t_j \times r ; \ j=1, \ldots, q ; \ v=1, \ldots, n_j - 1 ;
\]
\[
(\Xi_{ij})_{uv} : t_i \times t_j , \ i,j=1, \ldots, q; \ u=1, \ldots, n_i - 1 ; \ v=1, \ldots, n_j - 1 .
\]
Recall that the original data \( x_1, \ldots, x_N \) is distributed 
\( \mathcal{L}(x) = N(\mu, \Sigma) \). Under the non-null one population hypotheses, \( \mu \) and \( \Sigma \) are unrestricted. The values of \( \nu \) and \( \Xi \) corresponding to specific \( \mu \) and \( \Sigma \) alternatives may be found by using

\[
(4.21) \quad \nu = N^{1/2} \mu C, \quad \Xi = C' \Sigma C,
\]

where \( C \) is found in Section 3, Theorem 3.1.

**Test for \( H_1(mvc) \) versus \( H_{1,A} \):**

The results needed for testing \( H_1(mvc) \), the hypothesis that the mean, variances and covariances have BCS-I versus \( H_{1,A} \), the alternative hypothesis that the mean, variances and covariances are unstructured follow below.

From (4.13) and (4.16), the LRT \( \lambda_{14} \) is given by:

\[
(4.22) \quad \lambda_{14}^{2/N} = \frac{|V|}{|V_{00}|} \left( \sum_{i=1}^{q} \left( \frac{(n_i - 1) t_{i}^{(n_i - 1)}}{n_i - 1} \right) \right)^{(n_i - 1)}.
\]

The moments used for deriving the approximate null distribution are given by:

\[
(4.23) \quad \mathcal{E}_{14}^{S} = \frac{K}{\Gamma(N(1+S) - \frac{1}{2})} \left[ \sum_{i=1}^{m} \frac{t_i^{N(n_i - 1)}}{\Gamma(N(1+S) - \frac{1}{2})} \right] \left[ \sum_{j=1}^{l} \frac{t_j^{N(n_j - 1)}}{\Gamma(N(1+S) - \frac{1}{2})} \right],
\]

with \( K \) chosen so \( \mathcal{E}_{14}^{0} = 1. \)
Using the Box method, the approximate null distribution of $-2 \log \lambda_{14}$ is

\begin{equation}
(4.24) \quad P(-2 \log \lambda_{14} \leq z) = (1-\omega) P(\chi^2_F \leq \rho z) + \omega P(\chi^2_{F+4} \leq \rho z) + O(N^{-3}),
\end{equation}

where $f$, $\rho$ and $\omega$ are defined by

$$f = \frac{[p(p+3)-r(r+3)-t-d_0]/2}{},$$

\begin{equation}
(4.25) \quad r = m+t, \quad d_0 = \sum_{i=1}^{q} t_i^2,
\end{equation}

$$\rho = 1 - \frac{1}{12Nf} \left[d_1 - \sum_{i=1}^{q} \frac{t_i}{(n_i-1)} (2t_i^2 + 3t_i - 1)\right],$$

\begin{equation}
(4.26) \quad d_1 = p(2p^2 + 9p + 11) - r(2r^2 + 9r + 11),
\end{equation}

$$\rho^2 \omega = (1-\rho)^2 f/4 - \left(\frac{1-\rho}{24N}\right) \left[p(2p^2 + 9p + 11) - r(2r^2 + 9r + 11) - \sum_{i=1}^{q} \frac{t_i(2t_i^2 + 3t_i - 1)}{(n_i-1)}\right]$$

$$+ \frac{1}{48N^2} \left[p(p+1)(p^2+5p+6) - r(r+1)(r^2+5r+6) - \sum_{i=1}^{q} \frac{t_i(t_i+1)(t_i-1)(t_i+2)}{(n_i-1)^2}\right].$$

It is important to note that as the sample size goes to infinity, $\rho \to 1$ and $\omega \to 0$, i.e.,

$$\lim_{N \to \infty} P(-2 \log \lambda_{14} \leq z) = P(\chi^2_F \leq z).$$

Using the standard delta method to find the asymptotic non-null distribution of $(2/N)\log \lambda_{14}$ under the assumption that $\nu$ and $\Xi$ are unrestricted, we find that
(4.27) \[ x^2 \left[ \frac{2}{N} \log \lambda_{1i} \cdot \sum_{1 \leq i \leq q} \left( \frac{(\Xi_{01})^{-1}}{A_1} \right)^{n_i-1} \right] \rightarrow n(0, \nu_\infty), \]

where

\[ A_1 = \sum_{j=1}^{n_i-1} (\tilde{\nu}_{ij} \tilde{\nu}_{ij} + \Xi_{ij})_{jj} \],

\[ \nu_{\infty} = 2(p-r) + 4 \sum_{i=1}^{q} (n_i-1) \text{tr} \left( \sum_{j=1}^{n_i-1} \left( (\Xi_{01})^{-1} - (\Xi_{ij})_{jj} \right) A_1^{-1} \right) \]

\[ + 2 \sum_{i,j=1}^{q} \sum_{m=1}^{n_i-1} (n_i-1)(n_j-1) \text{tr} \left( A_1^{-1} (\Xi_{ij})_{mm} A_1^{-1} (\Xi_{jj})_{nn} \right) \]

\[ + 2 \sum_{i,j=1}^{q} \sum_{m=1}^{n_i-1} (\tilde{\nu}_{ij}) \text{tr} \left( A_1^{-1} (\Xi_{ij})_{mm} A_1^{-1} \right). \]

Note that under the null hypothesis, \( \nu_{\infty} = 0 \). This is explained by the fact that the standard delta method is a second order result and thus is not valid for the null hypothesis. In fact, as the sample size goes to infinity, the distribution of \(-2 \log \lambda\) is chi-squared distributed, (see approximate null distribution), not normally distributed.

The asymptotic non-null distribution has a very complicated appearance due to the variance term. The non-null distribution is most often used in order to compute a sample size under an alternative hypothesis so that the power is at a desired level. Thus, under a specific alternative, one need only substitute its parameters directly into the variance term.

If one does not have a specific alternative in mind, one may consider a class of alternatives and choose a sample size so that the minimum power over alternatives in this class is above a desired level. Another possibility
is to consider a set of alternatives that are in some sense "close" to
the null hypothesis parameters, i.e., an "indifference zone" and to set
the sample size so that the minimum power of alternatives not in this
"indifference zone" is above a desired level.

Test for $H_1(\nu c)$ versus $H_1,A$:

The results needed for testing $H_1(\nu c)$, the hypothesis that the
variances and covariances have BCS-I versus $H_1,A$, the alternative
hypothesis that the mean, variances and covariances are unstructured
follow below.

From (4.14) and (4.16), the LRT $\lambda_{24}$ is given by:

$$\lambda_{24}^2 / N = \frac{|v|}{|v_{00}|} \frac{1}{q} \left\{ \sum_{i=1}^{n_1-1} \left( \frac{n_i}{n_1-1} \right)^t \left( \frac{n_i}{n_1-1} \right) \left( \frac{1}{\sum_{j=1}^{n_i} (v_{ij})^2} \right) \right\}^{n_1-1}.$$

The moments used for deriving the approximate null distribution are
given by:

$$\mathcal{E}^S_{\lambda_{24}} = \frac{K \left[ \sum_{i=1}^{q} \left( \frac{t_i}{n_1-1} \right)^{\frac{n_1}{2}-1} \left( \frac{N(n_1-1)}{2} \right)^{\frac{1}{2}} \right]}{\left[ \sum_{i=1}^{1} \frac{N(n_1-1)}{2} \right]^{\frac{1}{2}} \left( \frac{N(n_1-1)}{2} \right)^{\frac{1}{2}} \left( \frac{2-n_1-j}{2} \right)}}, \quad \text{s=0,1,2,} \ldots,$$

with $K$ chosen so $\mathcal{E}^0_{\lambda_{24}} = 1$.

Using the Box method, the approximate null distribution of $-2 \log \lambda_{24}$
is
\( P[-2 \log \lambda_{24} \leq z] = (1 - \omega) P(X_1^2 \leq \rho z) + \omega P(X_1^2 + X_{r+1}^2 \leq \rho z) + O(N^{-3}) \),

where \( f, \rho \) and \( \omega \) are defined by

\[
\begin{align*}
 f &= [p(p+1)-r(r+1)-d_0-t]/2, \\
 \rho &= 1 - \frac{1}{12Nf} \left[ d_1 - 6(p-r+d_0+t) - \sum_{i=1}^{q} \frac{t_i}{(n_i-1)} \right], \\
 \rho^2 \omega &= (1-\rho)^2 f/4 \\
 - \frac{(1-\rho)}{24N} &\left[ p(2p^2+9p+5) - r(2r^2+9r+5) - 6d_0 + t \right] \\
 + \frac{1}{48N} &\left[ p(3p^2+9p+5) - r(3r^2+9r+5) - 6d_0 + t \right] \\
 - 2 \sum_{i=1}^{q} \frac{t_i(2t_i+3t_i-1)}{(n_i-1)} - \sum_{i=1}^{q} \frac{t_i(t_i+1)(t_i+2)(t_i-1)}{(n_i-1)^2}. 
\end{align*}
\]

Note \( r \) and \( d_0 \) are defined by (4.25), \( d_1 \) by (4.26).

Using the standard delta method to find the asymptotic non-null distribution of \( (2/N) \log \lambda_{24} \) under the assumption that \( v \) and \( \Xi \) are unrestricted, we find that

\[
(4.31) \quad \mathcal{L} \left\{ \sqrt{N} \left[ \frac{2}{N} \log \lambda_{24} - \log \left( \frac{\Xi}{\Xi_{00}} \right) \right] \right\} \to \mathcal{N}(0, \nu_{\infty}),
\]

where

\[
\begin{align*}
 B_i &= \sum_{j=1}^{q} \left( \Xi_{ij} \right)_{jj}, \\
 \nu_{\infty} &= -2(p-r) + 4 \sum_{i=1}^{q} \sum_{j=1}^{n_i-1} \left( n_i-1 \right) \text{tr} \left( \Xi_{00} \right)^{-1} O_i \left( \Xi_{ij} \right)_{jj} B_i^{-1} \\
 &+ 2 \sum_{i,j=1}^{q} \sum_{m=1}^{n_i-1} \sum_{n=1}^{n_j-1} \left( n_i-1 \right) \left( n_j-1 \right) \text{tr} B_i^{-1} \left( \Xi_{ij} \right)_{mn} B_j^{-1} \left( \Xi_{ij} \right)_{nm}.
\end{align*}
\]

40
Test for $H_1(m|vc)$ versus $H_1(vc)$:

The results needed for testing $H_1(m|vc)$, the hypothesis that the mean has BCS-I given that the variances and covariances have BCS-I versus $H_1(vc)$, the alternative hypothesis that the variances and covariances have BCS-I follow below.

From (4.13) and (4.14), the LRT $\lambda_{12}$ is given by:

$$\lambda_{12}^{2/N} = \prod_{i=1}^{q} \left\{ \frac{n_i^{-1}}{H_1 + \sum_{j=1}^{n_i} (V_{ij}^2)_{jj}} \right\}^{(n_i-1)}$$

(4.32)

The moments used for deriving the approximate null distribution are given by:

$$\xi^{S}_{\lambda_{12}} = K \prod_{i=1}^{t_i} \prod_{j=1}^{n_i} \left\{ \frac{n_i^{-1}N}{\Gamma\left(\frac{n_i}{2}\right)(1+S)} + \frac{(2-n_i-j)}{2} \right\}$$

(4.33)

with $K$ chosen so $\xi^{S}_{\lambda_{12}} = 1$.

Using the Box method, the approximate null distribution of $-2 \log \lambda_{12}$ is:

$$\text{P}\left(-2 \log \lambda_{12} \leq z\right) = (1-\omega) \text{P}\left(\chi^2_f \leq \rho z\right) + \omega \text{P}\left(\chi^2_{f+4} \leq \rho z\right) + O(N^{-1})$$

(4.34)

where $f$, $\rho$ and $\omega$ are defined by

$$f = p-r,$$

$$\rho = 1 - \frac{1}{2Nf} \left[ f+a_0+t \right],$$

$$\omega = 1 - \frac{1}{2Nf} \left[ a_0+t \right].$$
\[ \rho^2_\omega = \left[ \frac{(1-\rho)^2r}{4} - \frac{(1-\rho)}{4N} \right] (p-r+d_0+t) \]

\[ + \frac{1}{24N^2} [2(p-r)+3(d_0+t) + \sum_{i=1}^{q} \frac{t_i}{(n_i-1)} (2t_i^2+3t_i^2-1)] . \]

Note that \( d_0 \) and \( r \) are defined by (4.25).

Using the standard delta method to find the asymptotic non-null distribution of \((2/N) \log \lambda_{12}\) under the assumption that \( \nu \) is unrestricted, and \( \Xi = \text{diag}(\Xi_0, I_{n_1-1} \otimes \Xi_1, \ldots, I_{n_q-1} \otimes \Xi_q) \), we find that:

\[ (4.35) \quad \sqrt[N]{\frac{2}{N} \log \lambda_{12} - \log \left[ \prod_{i=1}^{p} \frac{|\Xi_i| (n_i-1)^{t_i}}{|A_i|} \right]} \sim \eta(0, \nu) , \]

where

\[ \nu_\infty = 2(p-r) - 4 \sum_{i=1}^{q} (n_i-1)^2 \text{tr} A_i^{-1} \Xi_i + 2 \sum_{i=1}^{q} (n_i-1)^3 \text{tr}(A_i^{-1} \Xi_i)^2 \]

\[ + 4 \sum_{i=1}^{q} \sum_{j=1}^{n_i-1} (n_i-1)^2 \gamma_{ij} A_i^{-1} \Xi_i - A_i^{-1} \gamma_{ij} - A_i^{-1} \gamma_{ij} , A_i = (n_i-1) \Xi_i + \sum_{j=1}^{n_i-1} \gamma_{ij} \gamma_{ij} , i = 1, \ldots, q. \]

Test for \( H_1(\mu | \nu) \) versus \( H_1(\nu) \):

The results needed for testing \( H_1(\mu | \nu) \), the hypothesis that the mean is \( \mu_0 \) and the variances and covariances have BGS-I versus \( H_1(\nu) \), the alternative hypothesis that the variances and covariances have BGS-I, follow below. This is the modified Hotelling's \( T^2 \) problem of testing \( \mu = \mu_0 \) versus \( \mu \neq \mu_0 \) when the covariance matrix has structure.

From (4.15) and (4.14), the LRT \( \lambda_{32} \) is given by:

\[ (4.36) \quad \lambda_{32}^{2/N} = \frac{|\nu_{00}|}{D_{00} + V_{00}} \prod_{i=1}^{q} \left\{ \frac{n_i-1}{\sum_{i=1}^{n_i} (V_{ii})_{jj}} \right\}^{n_i-1} \left\{ \sum_{j=1}^{n_i-1} [(D_{ii})_{jj} + (V_{ii})_{jj}] \right\} . \]
The moments used for deriving the approximate null distribution are given by:

\[
\mathcal{E}_\lambda S = K \prod_{i=1}^{m+t} \left[ \frac{\Gamma\left(\frac{N}{2} (1+S) - \frac{i-1}{2}\right)}{\Gamma\left(\frac{N}{2} (1+S) + \frac{i-1}{2}\right)} \right] \prod_{i=1}^{q} \frac{\Gamma\left(\frac{N}{2} (1+S) + \frac{2-n_i-j}{2}\right)}{\Gamma\left(\frac{N}{2} (1+S) + \frac{1-i}{2}\right)} \]

\[
= 0, 1, 2, \ldots, \text{ with } K \text{ chosen so } \mathcal{E}_\lambda^0 S \lambda^{32} = 1.

Using the Box method, the approximate null distribution of \(-2 \log \lambda^{32}\) is:

\[
P\{-2 \log \lambda^{32} \leq z\} = (1-\omega) P[\chi^2_p \leq \omega z] + \omega P[\chi^2_{p+q} \leq \rho z] + O(N^{-1})
\]

where \(f, \rho\) and \(\omega\) are defined by

\[
f = \frac{1}{p}, \quad \rho = 1 - \frac{k}{2N} \left[ r(r+1) + p + d_0 + t \right],
\]

\[
\omega = (1-\rho)^2 - (1-\rho) \left( r(r+1) + p + d_0 + t \right) + \frac{r(r+1)(r+2)}{12N^2} + \frac{2(p-r) + 3(d_0 + t)}{24N^2} + \frac{1}{24N^2} \sum_{i=1}^{q} \frac{t_i(2t_i^2 + 3t_i - 1)}{(n_i - 1)}.
\]

Note \(r\) and \(d_0\) are defined by (4.25).

Using the standard delta method to find the asymptotic non-null distribution of \((2/N)\log \lambda^{32}\) under the assumption that \(v\) is unrestricted, and \(\Xi = \text{diag}(\Xi_0, I_{n_1 - 1} \otimes \Xi_1, \ldots, I_{n_q - 1} \otimes \Xi_q)\), we find that...
$$(4.39) \quad \mathcal{L} \left\{ \sqrt{N} \left[ \frac{2}{N} \log \lambda_{\Sigma_{B}^{2}} - \log \left| \Xi_{0} \right| \frac{q}{1} \left\{ \frac{\left| \Xi_{i} \right| \left( n_{i}-1 \right) t_{i}}{\left| A_{1} \right|} \right\} ^{n_{i}-1} \right] \right\} \rightarrow n(0, \nu_{\infty}),$$

where

$$A_{i} = (n_{i}-1) \Xi_{i} + \sum_{j=1}^{n_{i}-1} \tilde{\nu}_{i,j} \tilde{\nu}_{i,j}, \quad i=1, \ldots, q,$$

$$B = \Xi_{0} + \tilde{\nu}_{0} \tilde{\nu}_{0},$$

$$\nu_{\infty} = 2p - 4 \quad \text{tr} \quad \Xi_{0} B^{-1} + 2 \quad \text{tr} \left( \Xi_{0} B^{-1} \right)^{2} - 4 \sum_{i=1}^{q} (n_{i}-1)^{2} \text{tr} \Xi_{1} A_{1}^{-1}$$

$$+ 2 \sum_{i=1}^{q} (n_{i}-1)^{3} \text{tr} (\Xi_{1} A_{1}^{-1})^{2} + 4 \tilde{\nu}_{0} B^{-1} \Xi_{0} B^{-1} \nu_{0}$$

$$+ 4 \sum_{i=1}^{q} (n_{i}-1)^{2} \sum_{j=1}^{n_{i}-1} \tilde{\nu}_{i,j} A_{1}^{-1} \Xi_{i} A_{1}^{-1} \tilde{\nu}_{i,j}.$$
5. **Type I k-population Hypotheses: LRT and Approximate Null and Asymptotic Non-null Distributions.**

We start with the canonical forms introduced in Section 3, (3.7), and (3.8) for each of the k populations. Denote these forms \( y^d \) and \( v^d \), \( d=1,\ldots,k \). We assume that we have a sample of size \( p_d \) from each of the populations and let \( N = \sum_{d=1}^{k} p_d \). The canonical forms \( y_1^1,\ldots,y_k^1,v_1^1,\ldots,v_k^1 \) are independently distributed with their distributions given by

\[
\mathcal{L}(y^d) = \mathcal{N}(v^d, \Xi^d), \quad \mathcal{L}(v^d) = \mathcal{W}(\Xi^d_p, p_d^{-1}),
\]

where \( v^d \) and \( \Xi^d \) have the following special structure (see Theorem 3.1):

\[
m = \sum_{i=1}^{b} m_i, \quad t = \sum_{j=1}^{q} t_j, \quad g = \sum_{j=1}^{q} t_j n_j, \quad p = m + g, \quad r = m + t,
\]

\[
v^d = (v^d_0, 0), \quad v^d_0: 1 \times (m + t), \quad v^d: 1 \times p, \quad v^d = \sqrt{v^d} \sqrt{p_d},
\]

\[
\Xi^d = \text{diag}(\Xi^d_0, I_{n_1-1} \otimes \Xi^d_1, \ldots, I_{n_q-1} \otimes \Xi^d_q),
\]

\[
\Xi^d_0: r \times r, \quad \Xi^d_j: t_j \times t_j, \quad j=1,\ldots,q.
\]

All the notation introduced in Section 4 remains the same except that a superscript is now included to indicate which population is being referenced. Examples include \( y^d, v^d_{ii}, (v^d_{ii})_{jj}, H^d_{ii}, y^d_i, y^d_{ij} \), where the superscript "d" indicates the population being referenced.

In addition, the following new notation is introduced. Let \( \bar{y} \) be the overall mean, i.e.,
(5.5) \[ \bar{y} = \frac{1}{N} \sum_{d=1}^{k} \sqrt{p_d} y_d. \]

When all population means are assumed equal, \( \bar{\nu}^1 = \cdots = \bar{\nu}^k \), and all covariance matrices are assumed equal \( \bar{\Sigma}^1 = \cdots = \bar{\Sigma}^k \), we may use the sample means to estimate the covariances. To this end we define:

(5.6) \[ D_{00} = \sum_{d=1}^{k} (y_0^d - \sqrt{p_d} \bar{y}_0)(y_0^d - \sqrt{p_d} \bar{y}_0) = \sum_{d=1}^{k} y_0^d y_0^d - \bar{y}_0 y_0^d. \]

Maxima of the likelihood functions.

Using Lemma 3.1, we can find the maxima of the likelihood functions

\( L(y^1, \ldots, y^k; \nu^1, \ldots, \nu^k) \) over the regions:

(5.7) \( \omega_1: \bar{\nu}^d = \cdots = \bar{\nu}^k; \bar{\Sigma}^1 = \cdots = \bar{\Sigma}^k \),

(5.8) \( \omega_2: \bar{\Sigma}^1 = \cdots = \bar{\Sigma}^k \),

(5.9) \( \omega_j: \nu_0^1, \ldots, \nu_0^k \text{ and } (\bar{\Sigma}_0^d, \ldots, \bar{\Sigma}_0^d), \ d = 1, \ldots, k \) are unrestricted.

Let \( a(p, N) = (2\pi)^{-pN/2} \). Then the maxima of the likelihood functions are

(5.10) \[ \sup_{\omega_1} L(y^1, \ldots, y^k; \nu^1, \ldots, \nu^k) = \frac{a(p, N) N^{pN/2}}{|D_{00} + \sum_{d=1}^{k} y_0^d y_0^d|^{N/2}} \times \]

\[ \left( \begin{array}{c}
\prod_{i=1}^{g} \frac{(n_i - 1)}{\sum_{d=1}^{k} (n_i - 1)} \\
\prod_{i=1}^{g} \frac{\sum_{d=1}^{k} (n_i - 1)}{\sum_{j=1}^{k} (n_i - 1)} \\
\prod_{i=1}^{g} \frac{\sum_{j=1}^{k} (n_i - 1)}{\sum_{d=1}^{k} (n_i - 1)} \\
\prod_{i=1}^{g} \frac{\sum_{j=1}^{k} (n_i - 1)}{\sum_{d=1}^{k} (n_i - 1)} \\
\end{array} \right)^{(n_i - 1)N/2}. \]

46
\[ \begin{align*}
(5.11) \quad \sup_{\omega_2} L(y^1, \ldots, y^k; v^1, \ldots, v^k) &= \left. \frac{a(p, N) N^{p d/2}}{1 \sum_{d=1}^k v^d \ln N/2} \prod_{i=1}^q \left\{ \frac{(n_i - 1)^{t_i}}{n_i - 1} \right\} \right\} \left( \frac{1}{\sum_{d=1}^k (H^d_i + \sum_{j=1}^q (v^d_{i1})_{jj})} \right) \left( \frac{1}{(n_i - 1)^{p d/2}} \right) \\
(5.12) \quad \sup_{\omega_3} L(y^1, \ldots, y^k; v^1, \ldots, v^k) &= \left. \frac{a(p, N) \prod_{d=1}^k \frac{p^d d^{p d/2}}{v^d \ln p d/2} \prod_{i=1}^q \left\{ \frac{(n_i - 1)^{t_i}}{n_i - 1} \right\} \right\} \left( \frac{1}{\sum_{d=1}^k (H^d_i + \sum_{j=1}^q (v^d_{i1})_{jj})} \right) \left( \frac{1}{(n_i - 1)^{p d/2}} \right) \\
\end{align*} \]

We may now generate the LRT for the three type I \( k \)-population hypotheses using

\[ \lambda_{ij} = \frac{\sup_{\omega_i} L(y^1, \ldots, y^k; v^1, \ldots, v^k)}{\sup_{\omega_j} L(y^1, \ldots, y^k; v^1, \ldots, v^k)} . \]

The LRT, moments of the LRT under the null distribution, approximate null distribution and asymptotic non-null distributions follow for each of the three \( k \)-population type I hypotheses. Methods used to derive them are explained at the end of Section 3 and are illustrated in Appendix I for \( \bar{H}_1(\text{mwc}) \). The moments of the LRT and the approximate null distributions do not depend on \( v^d \) and \( \bar{v}^d \), \( d=1, \ldots, k \). The asymptotic non-null distributions do depend on \( v^d \) and \( \bar{v}^d \), \( d=1, \ldots, k \).
Test for $H_k(MVC|mvc)$ versus $F_{k,A}$.

The results needed for testing $H_k(MVC|mvc)$, the hypothesis that the means and covariance matrices for the $k$ populations are the same given that each population's mean and covariance matrix have BCS-I versus $H_{k,A}$, the alternative hypothesis that the mean and covariance matrix of each population have BCS-I but are not all the same, follow below.

From (5.10) and (5.12), the LRT $\lambda_{13}$ is given by

\begin{equation}
\lambda_{13}^2/N = \frac{\sum_{d=1}^{k} P_d/N}{\prod_{d=1}^{k} P_d} \times \frac{|V_0^d| P_d/N \sum_{i=1}^{a} H_{i,j}^{d} + \sum_{j=1}^{b} (V_1^d)_{i,j} (n_{i,j} - 1) P_d/N |D_0 + \sum_{d=1}^{k} V_0^d | - \sum_{d=1}^{k} (H_{i,j}^{d} + \sum_{j=1}^{b} (V_1^d)_{i,j} n_{i,j} - 1)| n_{i,j} - 1}}
\end{equation}

The moments used for deriving the approximate null distribution are given by

\begin{equation}
\xi_{13}^S = K \frac{N^{PSN/2}}{\prod_{d=1}^{k} P_d} \times \frac{\left(\prod_{d=1}^{m+t} \frac{P_d(l+S)}{2} + \frac{1}{2}\right)\left(\prod_{d=1}^{m+t} \prod_{i=1}^{q} \prod_{j=1}^{t} \frac{(n_{i,j} - 1) P_d}{2} + \frac{1}{2}\right)\left(\prod_{i=1}^{m+t} \prod_{j=1}^{N(l+S)} \prod_{j=1}^{N} \prod_{j=1}^{N-1} \prod_{i=1}^{q} \prod_{j=1}^{t} \frac{N(n_{i,j} - 1)}{2} + \frac{1}{2}\right)}
\end{equation}

$S=0,1,2,\ldots$, with $K$ chosen so $\xi_{13}^0 = 1$. 

48
Using the Box method, the approximate null distribution of \(-2\log \lambda_{13}\) is

\[
(5.16) \quad P\{-2\log \lambda_{13} \leq z\} = (1-\omega) P\{\chi_{r}^{2} \leq \rho z\} + \omega P\{\chi_{r+4}^{2} \leq \rho z\} + O(\min_{d=1, \ldots, k} p_{d}^{-3}),
\]

where

\[
(5.17) \quad r = \frac{(k-1)}{2}[r(r+3) + d_{0} + t],
\]

\(r\) and \(d_{0}\) defined by (4.25); \(\rho\) and \(\omega\) are defined by

\[
(5.18) \quad \rho = 1 - \frac{(N{s}_{i}^{-1})}{12N^{f}}[r(2r^{2}+9r+11) + \sum_{i=1}^{q} \frac{t_{i}(2t_{i}^{2}+3t_{i}+1)}{(n_{i}^{-1})}],
\]

\[
(5.19) \quad s_{1} = \sum_{d=1}^{k} p_{d}^{-1},
\]

\[
\rho^{2}\omega = \frac{(1-\rho)^{2}r}{4} - \frac{(1-\rho)(N{s}_{i}^{-1})}{24N}[r(2r^{2}+9r+11) + \sum_{i=1}^{q} \frac{t_{i}(2t_{i}^{2}+3t_{i}+1)}{(n_{i}^{-1})}]
\]

\[
+ \frac{N^{2}s_{2}^{-1}}{48N^{2}}[r(r+1)(r+2)(r+3) + \sum_{i=1}^{q} \frac{t_{i}(t_{i}+1)(t_{i}^{2}+t_{i}+2)}{(n_{i}^{-1})^{2}}],
\]

\[
(5.20) \quad s_{2} = \sum_{d=1}^{k} p_{d}^{-2}.
\]

Using the standard delta method to find the asymptotic non-null distribution of \((2/N)\log \lambda_{13}\) under the assumption that
\(\nu_0', ..., \nu_k', (\Xi_0', ..., \Xi_d), \ d=1, ..., k,\) are unrestricted, we find that

\[
(5.21) \quad \mathcal{L} \left\{ \sqrt{N} \log \lambda_{1^5} - \log \prod_{d=1}^{k} \left| \mathbf{A} \right| \prod_{i=1}^{q} \left| \mathbf{B}_i \right| \right\} \rightarrow n(0, \nu_\infty),
\]

where

\[
(5.22) \quad A = \sum_{d=1}^{k} f_d (\Xi_0 + \Delta_{0,d}) - \sum_{i,j=1}^{k} \nu_i \nu_j \tilde{r}_i \tilde{r}_j,
\]

\[
\nu_\infty = 2p + 2 \sum_{d=1}^{k} f_d [2c_d A_{0,d}^{-1} - 2tr(\Xi_0 A^{-1}) + tr(\Xi_0 A^{-1})^2
\]

\[+ \sum_{i=1}^{q} (n_{i-1})^2 \{ (n_{i-1}) tr(\Xi_{i,i}^{-1})^2 - 2tr(\Xi_{i,i}^{-1}) \}]
\]

\[
(5.23) \quad B_i = \sum_{d=1}^{k} f_d \Xi_{i,d}, \quad i=0, ..., q,
\]

\[
(5.24) \quad f_d = p_d / N, \quad c_d = \nu_0 - \sum_{r=1}^{k} \tilde{r}_r \nu_0, \quad d=1, ..., k.
\]

Test for \(H_k(VC|mvc)\) versus \(H_{k,A}\).

The results needed for testing \(H_k(VC|mvc)\), the hypothesis that the covariance matrices for the \(k\) populations are the same given that each population's mean and covariance matrix have BCS-I versus \(H_{k,A}\), the alternative that the mean and covariance matrix of each population have BCS-I but are not all the same, follow below.
From (5.11) and (5.12), the LRT $\lambda_{23}$ is given by

\[
\frac{2/N}{\lambda_{23}} = \frac{N^p}{\prod_{d=1}^{k} \prod_{i=1}^{2d} \prod_{j=1}^{d} (V_{ij}^{d})^{-1/2}} \times \frac{\sum_{d=1}^{k} p_d/N}{\sum_{d=1}^{k} V_{00}^{d} + \sum_{i=1}^{q} (V_{ii}^{d})^{-1/2} \sum_{j=1}^{n_i-1} (H_i^{d})^{-1/2} + \sum_{j=1}^{n_i-1} (V_{ii}^{d})^{-1/2} (n_i-1)p_d/N}.
\]

The moments used for deriving the approximate null distribution are given by

\[
\mathcal{E} \lambda_{23}^S = k \frac{N^{PSN/2}}{\prod_{d=1}^{k} p_d S/2} \times \left[ \prod_{d=1}^{k} \prod_{i=1}^{a} \prod_{j=1}^{t_i} \Gamma\left( \frac{n_i-1)p_d(1+S)}{2} \right) \right]^{-1/2} \left[ \prod_{i=1}^{m+t} \Gamma\left( \frac{N(1+S)}{2} + \frac{1-k-i}{2} \right) \prod_{i=1}^{a} \prod_{j=1}^{t_i} \Gamma\left( \frac{N(1-S)}{2} + \frac{1-i}{2} \right) \right],
\]

$S=0,1,2,\ldots$, with $K$ chosen so $\mathcal{E} \lambda_{23}^0 = 1$.

Using the Box method, the approximate null distribution of $-2 \log \lambda_{23}$ is:

\[
P\{ -2 \log \lambda_{23} \leq z \} = (1-\omega) P\{ x_{f}^{2} \leq \rho z \} + \omega P\{ x_{f+t}^{2} \leq \rho z \} + O(\min_{d=1, \ldots, k} p_d^{-3}),
\]

\[
f = \frac{k-1}{2} [r(r+1)+d_0+t],
\]

\[51\]
\[ \rho = 1 - \frac{N_s - 1}{2N^2} \left[ r(2r^2 + 9r + 11) + \sum_{i=1}^{q} \frac{t_i(2t_i^2 + 3t_i - 1)}{(n_i - 1)} \right] + \frac{(k-1)r(r+k+2)}{2N^2}, \]

\[ \rho^2 = \frac{(1-\rho)^2}{4} - \frac{(1-\rho)}{2N} \left[ (N_s - 1)(r(2r^2 + 9r + 11) + \sum_{i=1}^{q} \frac{t_i(2t_i^2 + 3t_i - 1)}{(n_i - 1)} \right] \]

\[ - 6(k-1)r(k+r+2) + \frac{N^2 s_2 - 1}{48N^2} \left[ r(r+1)(r+2)(r+3) + \sum_{i=1}^{q} \frac{t_i(t_i+1)(t_i^2 + t_i - 2)}{(n_i - 1)^2} \right] \]

\[ - \frac{(k-1)r}{24N^2} \left[ 2(k-1)^2 + 3(k-1)(r+3) + 2r^2 + 9r + 11 \right], \]

where \( r \) and \( d_0 \) are defined by (4.25), \( s_1 \) by (5.19) and \( s_2 \) by (5.20).

Using the standard delta method to find the asymptotic non-null distribution of \( (2/N)\log \lambda_{23} \) under the assumption that \( v_0^1, \ldots, v_0^k, \{v_0^d, \ldots, v_q^d\}, d=1, \ldots, k, \) are unrestricted, we find that

\[ \mathcal{L} \left\{ \sqrt{\frac{2}{N}} \log \lambda_{23} - \log \left( \prod_{d=1}^{k} \frac{f_d}{|\mathbb{E}_d|} \prod_{i=1}^{q} \frac{|v_i^d|}{|\mathbb{E}_i|} \right) \right\} \rightarrow \mathcal{N}(0, \nu_\infty), \]

where

\[ \nu_\infty = 2p + 2 \sum_{d=1}^{k} f_d [-2\text{tr}(\mathbb{E}_0 B_0^{-1}) + \text{tr}(\mathbb{E}_0 B_0^{-1})^2 \]

\[ + \sum_{i=1}^{q} (n_i - 1)^2 [(n_i - 1)\text{tr}(\mathbb{E}_i B_i^{-1})^2 - 2\text{tr}(\mathbb{E}_i B_i^{-1})]], \]

\( B_1 \) defined by (5.25).
Test for $H_k(M|\text{mVC})$ versus $H_k(VC|\text{mVC})$.

The results needed for testing $H_k(M|\text{mVC})$, the hypothesis that the means for the $k$ populations are the same given that the covariance matrices of the populations are the same and that the mean and covariance matrix of each population have BCS-I versus $H_k(VC|\text{mVC})$, the alternative hypothesis that the covariance matrices for the $k$ populations are the same given that the mean and covariance matrix of each population have BCS-I, but the means are not all the same, follow below.

From (5.10) and (5.11), the LRT $\lambda_{12}$ is given by:

$$\lambda_{12}^{2/N} = \frac{\left| \sum_{d=1}^{k} v_{d0}^2 \right|}{\left| D_{00} + \sum_{d=1}^{k} v_{d0}^2 \right|}.$$  

(5.29)

The moments used for deriving the approximate null distribution are given by

$$\phi_{\lambda_{12}^S} = k \prod_{i=1}^{m+t+1} \left\{ \frac{\Gamma \left( \frac{N}{2} (1+S) + \frac{1-k-i}{2} \right)}{\Gamma \left( \frac{N}{2} (1+S) - \frac{i}{2} \right)} \right\}, \quad S = 0, 1, 2, \ldots,$$

(5.30)

with $K$ chosen so $\phi_{\lambda_{12}^0} = 1$.

Using the Box method, the approximate null distribution of $-2 \log \lambda_{12}$ is:

$$P(-2 \log \lambda_{12} \leq z) = (1-\omega) P\left( \chi_{f}^{2} \leq \rho z \right) + \omega P\left( \chi_{f+k}^{2} \leq \rho z \right) + O\left( \min_{d=1,\ldots,k} \left\{ \frac{f+1}{f} \right\} \right),$$

(5.31)  

$$f = (k-1)r, \quad \rho = 1 - \frac{r+k+2}{2N}.$$
\[
\rho^2 \omega = \frac{(1-\rho)^2 f}{4} - \frac{(1-\rho) f(k^2+2)}{4N} + \frac{f}{24N^2} \left[2r^2 + 3r(k+2) + 2k^2 + 5k + 4\right].
\]

Using the standard delta method to find the asymptotic non-null distribution of \( (2/N) \log \lambda_{12} \) under the assumption that \( \nu_0, \ldots, \nu_0^k \) are unrestricted, and \( \Xi_0 = \Xi_0^2 = \cdots = \Xi_0^k = \Xi_0^* \), we find that

\[
\sqrt{N} \left[ \frac{2}{N} \log \lambda_{12} - \log \frac{\Xi_0^*}{E} \right] \rightarrow \mathcal{N}(0, \nu_\infty),
\]

where

\[
\nu_\infty = 2r + 2 \sum_{d=1}^k f_d \left[ 2c_d E^{-1} \Xi_0^{-1} \Xi_0^* \Xi_0^{-1} + 2 \text{tr}(\Xi_0^{-1} E^{-1}) + \text{tr}(\Xi_0^* E^{-1}) \right],
\]

\[
E = \Xi_0^* + \sum_{d=1}^k f_d \sum_{d=0}^{d'} \nu_0^{-d'} E^{-1} \nu_0^d + \sum_{i,j=1}^k \nu_0^i \nu_0^j f_i f_j,
\]

\( c_d \) defined by \((5.24)\).
6. **Type II One-population Hypotheses: LRT and Approximate Null and Asymptotic Non-null Distributions.**

We start with the canonical forms introduced in Section 3, (3.7) and (3.8). The canonical forms \( y \) and \( V \) are independently distributed with their distributions given by:

\[
\mathcal{L}(y) = \mathcal{N}(v, \Xi), \quad \mathcal{L}(V) = \mathcal{W}(\Xi, p, N-1).
\]

Under BCS-II, \( v \) and \( \Xi \) have the following special structure (see Theorem 3.2):

\[
(6.2) \quad v = (v_1, 0), \quad v_1: l \times rh,
\]

\[
(6.3) \quad \Xi = \text{diag}(\Xi_1, I_{n-1} \otimes \Xi_2),
\]

\( \Xi: p \times p, \quad \Xi_1 \) and \( \Xi_2: rh \times rh \).

Next, we identify special blocks of \( y \) and \( V \) consistent with the special structure of \( v \) and \( \Xi \) under BCS-II.

Decompose the p-variate row vector \( y \) by \( y = (y_1, \ldots, y_n) \) where \( y_i \) is \( l \times rh, i=1,\ldots,n \). The \( p \times p \) matrix \( V \) is decomposed as

\( V = (V_{ij}) \) where \( V_{ij} \) is \( rh \times rh, i,j=1,\ldots,n \).

Because under BCS-II the mean vector \( v \) has some elements identically 0, we use \( y \) in the estimation of \( \Xi_2 \). To this end, we define \( \mathbf{H} = \sum_{i=2}^{n} y_i'y_i \).

Also define \( D = y'y \). This is used in the estimation of \( \Xi \) in the modified Hotelling's \( T^2 \) problem where under the null hypothesis, \( \mathcal{L}(y) = \mathcal{N}(0, \Xi) \).

\( D \) is decomposed the same as \( V \).
It is interesting to compare these structures of $v$ and $\Xi$ under BCS-II with those of BCS-I. Recall that under BCS-I,

$$v = (v_0, 0), \quad v_0: 1 \times m + t,$$

$$\Xi = \text{diag}(\Xi_0, I_{n_1 - 1} \otimes \Xi_1, \ldots, I_{n_q - 1} \otimes \Xi_q),$$

$$\Xi_0: (m + t) \times (m + t), \quad t_j \times t_j, \quad j = 1, \ldots, q.$$

We observe that the BCS-II structure appears, in canonical form, to be a simpler case of BCS-I structure, in canonical form.

Due to the extensive substitutions that are necessary to get from the type I to the type II results, the type II results are presented in their entirety. A study of the type II results may help in the understanding of the type I results. This may be especially true when it is necessary to interpret the non-null asymptotic distributions.

**Maxima of the likelihood functions.**

Using Lemma 3.1, we can find the maxima of the likelihood functions $L(y, V)$ over the regions:

$$\omega_1: \quad v \text{ and } \Xi \text{ have BCS-II structure},$$

$$\omega_2: \quad \Xi \text{ has BCS-II structure, } \quad v \text{ unrestricted},$$

$$\omega_3: \quad \Xi \text{ has BCS-II structure, } \quad v = 0,$$

$$\omega_4: \quad v \text{ and } \Xi \text{ unrestricted}.$$
Let \( b(p,N) = N^{pN/2} (2\pi e)^{-pN/2} \). Then the maxima of the likelihood functions are

\[ \text{(6.10)} \quad \sup_{v, \Xi \in \omega_1} L(y,V) = \frac{b(p,N)(n-1) \text{Re} (n-1)N/2}{|V_{11}|^{N/2} |V_{11} + \sum_{j=2}^{n} V_{jj}|^{(n-1)N/2}}, \]

\[ \text{(6.11)} \quad \sup_{v, \Xi \in \omega_2} L(y,V) = \frac{b(p,N)(n-1) \text{Re} (n-1)N/2}{|V_{11}|^{N/2} + \sum_{j=2}^{n} V_{jj}|^{(n-1)N/2}}, \]

\[ \text{(6.12)} \quad \sup_{v, \Xi \in \omega_3} L(y,V) = \frac{b(p,N)(n-1) \text{Re} (n-1)N/2}{|D_{11} + V_{11} + \sum_{j=2}^{n} (D_{jj} + V_{jj})|^{(n-1)N/2}}, \]

\[ \text{(6.13)} \quad \sup_{v, \Xi \in \omega_4} L(y,V) = b(p,N)/|V|^{N/2}. \]

We may now generate the LRT for the type II one-population hypotheses using:

\[ \text{(6.14)} \quad \lambda_{ij} = \frac{\sup_{v, \Xi \in \omega_i} L(y,V)}{\sup_{v, \Xi \in \omega_j} L(y,V)}. \]

The LRT, moments of the LRT under the null distribution, approximate null distribution and asymptotic non-null distribution follow for each of the three one-population type II hypotheses. Methods used to derive them are explained at the end of Section 3 and are illustrated in Appendix I.
for $\overline{H}(mvc)$. The moments of the LRT and the approximate null distributions do not depend on $\nu$ or $\Xi$. The asymptotic non-null distributions do depend on $\nu$ and $\Xi$. Thus we define special structures of $\nu$ and $\Xi$ by

\begin{equation}
(6.15) \quad \nu = (\nu_1, \ldots, \nu_n), \quad \nu_i: 1 \times rh, \quad i = 1, \ldots, n, \quad \nu = \sqrt{N} \tilde{\nu},
\end{equation}

\begin{equation}
(6.16) \quad \Xi = (\Xi_{ij}), \quad \Xi_{ij}: rh \times rh, \quad i, j = 1, \ldots, n.
\end{equation}

Test for $\overline{H}_1(mvc)$ versus $\overline{H}_{1,A}$.

The results needed for testing $\overline{H}_1(mvc)$, the hypothesis that the mean, variances and covariances have BCS-II versus $\overline{H}_{1,A}$, the alternative hypothesis that the mean, variances and covariances are unstructured, follow below.

From (6.10) and (6.13), the LRT $\lambda_{1A}$ is given by

\begin{equation}
(6.17) \quad \chi^2_{14}/N = \frac{|V|(n-1)^{rh(n-1)}}{|V_{11}|H + \sum_{j=2}^{n} V_{jj}^{n-1}j^{rh(n-1)}}.
\end{equation}

The moments used for deriving the approximate null distribution are given by

\begin{equation}
(6.18) \quad \xi_{14}^S = K(n-1)^{rhS(n-1)N/2} \frac{(n-1)^{rh}}{i=1} \frac{r}{i} \left( \frac{N}{2} (1+S) - \frac{rh+i}{2} \right), \quad S = 0, 1, 2, \ldots,
\end{equation}

\begin{equation}
\frac{r}{i=1} \left( \frac{N}{2} (1+S) + \frac{1-i}{2} \right)
\end{equation}

with $K$ chosen so $\xi_{14}^0 = 1$. 

58
Using the Box method, the approximate null distribution of \(-2 \log \lambda_{14}\) is:

\[(6.19) \quad P[-2 \log \lambda_{14} \leq z] = (1-\omega) P(\chi_1^2 \leq \rho z) + \omega P(\chi_{t+4}^2 \leq \rho z) + o(N^{-1}),\]

where \(t, \rho, \omega\) are defined by

\[f = \frac{rh}{2} [rh(n^2-2)+3n-4] = \frac{1}{2}(p(p+3)-2rh(rh+2)),\]

\[\rho = r - \frac{rh}{12Nt} (d_4 + d_5),\]

\[(6.20) \quad d_4 = (n-1)[2(rh)^2(n^2+n+1)+9rh(n+1)+1l],\]

\[(6.21) \quad d_5 = - \frac{2(rh)^2+3rh-1}{(n-1)},\]

\[\rho^2 \omega = \frac{(1-\rho)^2 f}{4} - \frac{(1-\rho)rh}{24N} [(n-1)[2(rh)^2(n^2+n+1)+9rh(n+1)+1l]
\]

\[\quad - \frac{2(rh)^2+3rh-1}{n-1}] + \frac{rh}{48N^2} [(n-1)((rh)^3(n^2+n+1)
\]

\[\quad + 6(rh)^2(n^2+n+1)+11rh(n+1)+6) - \frac{(rh+1)(rh-1)(rh+2)}{(n-1)^2}] .\]

Using the standard delta method to find the asymptotic non-null distribution of \((2/N)\log \lambda_{14}\) under the assumption that \(\nu\) and \(\Xi\) are unrestricted, we find that
\[(6.22) \quad \mathcal{L} \left\{ \sqrt{N} \left[ \frac{2}{N} \log \lambda_{14} - \log \frac{|\Sigma|}{|A|^{n-1}} \right] \right\} \to \eta(0, \nu), \]

where

\[\lambda_{14} = \sum_{k=2}^{n} (\Sigma_{kk} + \nu_{kk}), \]

\[\nu = 2\rho(n-1)^{-4(n-1)} \text{tr} A^{-1} \sum_{k=2}^{n} (\Sigma_{kk} - \Sigma_{k1}\Sigma_{1k}), \]

\[+ 2(n-1)^{2} \sum_{i,j=2}^{n} \text{tr} (\Sigma_{ij} + 2\nu_{ij}) A_{ij}^{-1}, \]

Test for \( H_1 (\nu c) \) versus \( H_{1A} \).

The results needed for testing \( H_1 (\nu c) \), the hypothesis that the variances and covariances have BCS-II versus \( H_{1A} \), the alternative hypothesis that the mean, variances and covariances are unstructured, follow below.

From (6.11) and (6.13), the LRT \( \lambda_{24} \) is given by

\[(6.23) \quad \mathcal{L}^{2/N} = \frac{|V|(n-1)^{rh(n-1)}}{|V|^{1/2} \sum_{j=2}^{n} \nu_{jj}|^{n-1}}. \]

The moments used for deriving the approximate null distribution are given by

\[(6.24) \quad \mathcal{E}^{S} = K(n-1)^{rhS(n-1)N/2} \frac{\Gamma_{i=1}^{n}(1+S) - \frac{rh+1}{2}}{\prod_{i=1}^{rh} \Gamma_{i=1}^{N(n-1)}/(1+S) + \frac{2-n-i}{2}}, \quad S=0,1,2,\ldots, \]

with \( K \) chosen so \( \mathcal{E}^{0} = 1. \)

Using the Box method, the approximate null distribution of \(-2 \log \lambda_{24}\) is
(6.25) \[ P\{-2\log \lambda_{24} \leq z\} = (1-\omega) P(\chi_{r}^{2} \leq \rho z) + \omega P(\chi_{r+4}^{2} \leq \rho z) + o(N^{-3}), \]

\[ f = \frac{rh}{2} [rh(n^2-2)+n-2] = \frac{p(p+1)-2rh(rh+1)}{2}, \]

\[ \rho = 1 - \frac{rh}{12Nf} \left[ d_{4} + d_{5} - 6rh(n+rh) \right], \]

d_{4} and \( d_{5} \) defined by (6.20) and (6.21); \( \omega \) is defined by

\[ \rho^{2}\omega = \frac{(1-p)^{2}r}{4} - \frac{(1-p)rh}{24N} \left( (n-1)\frac{2(rh)^{2}(n^{2}+n+1) + 9rh(n+1)+5}{(n-1)} \right) \]

\[ -6(nh+1) - \frac{2(rh)^{2}+3rh-1}{(n-1)} \]

\[ + \frac{rh}{48N} \left( (n-1)((rh)^{3}(n^{3}+n^{2}+n+1) + 6(rh)^{2}(n^{2}+n+1) + 11rh(n+1)+2) \right) \]

\[ - 6(nh+1) - \frac{2(2(rh)^{2}+3rh-1)}{(n-1)} - \frac{(rh+1)(rh+2)(rh-1)}{(n-1)^{2}}. \]

Using the standard delta method to find the asymptotic non-null distribution of \( (2/N)\log \lambda_{24} \) under the assumption that \( \nu \) and \( \Xi \) are unrestricted, we find that

(6.26) \[ \mathcal{L}\left\{ \sqrt{N} \left[ \frac{2}{N} \log \lambda_{24} - \log \frac{|\Xi|}{|B|^{n-1}} \right] \right\} \rightarrow \eta(0, \nu_{\infty}), \]

\[ B = \sum_{k=2}^{n} \Xi_{kk}, \]

\[ \nu_{\infty} = -2p+4(n-1)\text{tr} B^{-1} \left( \sum_{k=2}^{n} \Xi_{kk} \Xi_{kk}^{-1} \right) + 2rh \]

\[ + 2(n-1)^{2} \text{tr} B^{-1} \left( \sum_{i,j=2}^{n} \Xi_{ij} B^{-1} \Xi_{ij} \right). \]
Test for \( \bar{H}_1(m|vc) \) versus \( \bar{H}_1(vc) \).

The results needed for testing \( \bar{H}_1(m|vc) \), the hypothesis that the mean has BCS-II given that the variances and covariances have BCS-II versus \( H_1(vc) \), the alternative hypothesis that the variances and covariances have BCS-II follow below.

From (6.10) and (6.11), the LRT \( \lambda_{12} \) is given by

\[
\lambda_{12}^{2/N} = (\frac{1}{\sum_{j=2}^{n} V_{jj}} / |H + \frac{1}{\sum_{j=2}^{n} V_{jj}}|)^{(n-1)}. \tag{6.27}
\]

The moments used for deriving the approximate null distribution are given by

\[
\mathcal{E}\lambda_{12}^S = K \prod_{i=1}^{r} \frac{r! \frac{N(n-1)}{2} \frac{1}{2} (l+i) + \frac{2-n}{2}}{r! \frac{N(n-1)}{2} \frac{1}{2} (l+i) + \frac{1-i}{2}}, \quad S=0,1,2,\ldots,
\tag{6.28}
\]

with \( K \) chosen so \( \mathcal{E}\lambda_{12}^0 = 1 \).

Using the Box method, the approximate null distribution of \(-2 \log \lambda_{12}\) is

\[
P{-2 \log \lambda_{12} \leq z} = (1-\omega) P\{X_{1f}^2 \leq \rho z\} + \omega P\{X_{1f+4}^2 \leq \rho z\} + O(N^{-3}), \tag{6.29}
\]

where \( f, \rho \) and \( \omega \) are defined by

\[
f = (n-1)rh, \quad \rho = 1 - \frac{1}{2N} \left( \frac{rh+n}{n-1} \right),
\]

\[
\rho^2\omega = \frac{(1-\rho)^2 f}{4} - \frac{(1-\rho)rh(n+rh)}{4N} + \frac{rh}{24N^2} [(3rh+2n+1) + \frac{2(rh)^2+n+3}{(n-1)}].
\]

62
Using the standard delta method to find the asymptotic non-null distribution of \((2/N)\log \lambda_{12}\) under the assumption that \(\nu\) is unrestricted, and \(\Xi = \text{diag}(\Xi_1, \mathbb{I}_{n-1} \otimes \Xi_2)\), we find that

\[
(6.30) \quad \mathcal{L} \left\{ \sqrt{N} \left[ \frac{2}{N} \log \lambda_{12} - \log \left( \frac{|\Xi_2|}{|A|} \right) (n-1)^{\frac{\rho h}{2}} n^{-1} \right] \right\} \rightarrow \mathcal{N}(0, \nu_{\infty}),
\]

\[
\nu_{\infty} = 2(n-1)[\rho h - 2(n-1)tr \Xi_2 A^{-1} + (n-1)^2 tr(\Xi_2 A^{-1})^2] + 4(n-1)^2 \sum_{k=2}^{n} \overline{\Xi}_k A^{-1} \Xi_2 A^{-1} \overline{\Xi}_k,
\]

\[
A = (n-1)\Xi_2 + \sum_{k=2}^{n} \overline{\Xi}_k \overline{\Xi}_k.
\]

Test for \(\overline{H}_1(\mu_0 | \nu c)\) versus \(\overline{H}(\nu c)\).

The results needed for testing \(\overline{H}_1(\mu_0 | \nu c)\), the hypothesis that the mean is \(\mu_0\) and the variances and covariances have BCS-II, versus \(\overline{H}_1(\nu c)\), the alternative hypothesis that the variances and covariances have BCS-II follow below. This is the modified Hotelling's \(T^2\) problem of testing \(\mu = \mu_0\) versus \(\mu \neq \mu_0\) when the covariance matrix has structure.

From (6.12) and (6.11), the LRT \(\lambda_{32}^2\) is given by

\[
(6.31) \quad \lambda_{32}^{2/N} = \frac{|V_{11}|}{|D_{11}^+ V_{11}|} \left( \frac{|\sum_{j=2}^{n} V_{jj}|}{|\sum_{j=2}^{n} (D_{jj}^+ V_{jj})|} \right)^{n-1}.
\]

The moments used for deriving the approximate null distribution are given by

63
(6.32) \[ \mathcal{E} \lambda_{32}^2 = K \frac{rh}{i=1} \left[ \frac{\Gamma \left( \frac{N}{2} (1+S) - \frac{1}{2} \right)}{\Gamma \left( \frac{N}{2} (1+S) + \frac{1}{2} \right)} \cdot \frac{\Gamma \left( \frac{N(n-1)}{2} (1+S) + \frac{2-n-1}{2} \right)}{\Gamma \left( \frac{N(n-1)}{2} (1+S) + \frac{1}{2} \right)} \right] \]

S=0,1,2,\ldots, \text{ with } K \text{ chosen so } \mathcal{E} \lambda_{32}^0 = 1.

Using the Box method, the approximate null distribution of \(-2 \log \lambda_{32}\)

is

(6.33) \[ P[-2 \log \lambda_{32} \leq z] = (1-\omega) P(X_f^2 \leq \rho z) + \omega P(X_{f+4}^2 \leq \rho z) + O(\frac{1}{n-3}) \]

where \(f, \rho\) and \(\omega\) are defined by

\[ f = \frac{p}{4}, \quad \rho = 1 - \frac{1}{2N^2} [2rh+2n] \]

\[ \rho^2 \omega = \frac{(1-\rho)^2}{4} - \frac{(1-\rho)rh(2rh+2n)}{4N} + \left[ \frac{(rh+1)(rh+2)}{12N^2} \right. \]

\[ + \frac{2(n-1)+3(rh+1)}{24N^2} + \frac{2(rh)^2+3rh-1}{24N^2(n-1)} \] \[ \left. \right] rh . \]

Using the standard delta method to find the asymptotic non-null distribution of \((2/N) \log \lambda_{32}\) under the assumption that \(\nu\) is unrestricted, and \(\Xi = \text{diag}(\Xi_1, \ldots, \Xi_{n-1} \otimes \Xi_2)\), we find that

(6.34) \[ \mathcal{L} \left\{ \sqrt{N} \left[ \frac{2(2/N) \log \lambda_{32} - \log |\Xi| - \log |B|}{|A|} \left( \frac{|\Xi_2| (n-1)^{rh} - n-1}{|A|} \right)^{n-1} \right] \right\} \rightarrow \mathcal{N}(0, \nu_\infty) , \]

\[ A = (n-1)\Xi_2 + \sum_{j=2}^{n-1} \tilde{\nu}_j \tilde{\nu}_j , \quad B = \Xi_1 + \tilde{\nu}_1 \tilde{\nu}_1 , \]

\[ v_\infty = 2p-4tr \Xi_1 B^{-1} + 2tr(\Xi_1 B^{-1})^2 - 4(n-1)^2 tr \Xi_2 A^{-1} \]

\[ + 2(n-1)^3 tr(\Xi_2 A^{-1})^2 + 4\tilde{\nu}_1 B^{-1} \Xi_1 B^{-1} \tilde{\nu}_1 + 4(n-1) \sum_{j=2}^{n-1} \tilde{\nu}_j A^{-1} \Xi_2 A^{-1} \tilde{\nu}_j . \]

64
7. Type II k-population Hypotheses: LRT and Approximate Null and Asymptotic Non-null Distributions.

We start with the canonical forms introduced in Section 3, (3.7) and (3.8) for each of the k populations. Denote these forms $y^d$ and $V^d$, $d=1,\ldots,k$. We assume that we have a sample of size $p_d$ from each of the populations and let $N = \sum_{d=1}^{k} p_d$. The canonical forms $y_1^d,\ldots,y_k^d, V_1^d,\ldots,V_k^d$ are independently distributed with their distributions given by

$$\mathcal{L}(y^d) = h(y^d, \Xi^d), \quad \mathcal{L}(V^d) = W(\Xi^d, p,p_d,1),$$

where $v^d$ and $\Xi^d$ have special structure (see Theorem 3.2) under which

$$v^d = (v_1^d,0), \quad v_1^d: 1 \times rh, \quad v^d: 1 \times p, \quad v^d = \sqrt[p_d]{p} \sqrt[p_d]{v_1^d},$$

$$\Xi^d = \text{diag}(\Xi_1^d, I_{n-1} \otimes \Xi_2^d), \quad \Xi_1^d: rh \times rh, \quad \Xi_2^d: rh \times rh.$$

All the notation introduced in Section 4 remains the same except that a superscript is now included to indicate which population is being referenced. Examples include $y^d, V_1^d, H^d, y_i^d$, where the superscript "d" indicates the population being referenced.

In addition, the following new notation is introduced. Let $\overline{y}$ be the overall mean, i.e.,

$$\overline{y} = \frac{1}{N} \sum_{d=1}^{k} \sqrt[p_d]{p} y^d.$$
When all population means are assumed equal, \( \nu^1 = \ldots = \nu^k \), and all covariance matrices are assumed equal, \( \Sigma_1 = \ldots = \Sigma_k \), we may use the sample means to estimate the covariance. To this end we define:

\[
(7.5) \quad D_{11} = \sum_{d=1}^{k} (y^d_1 - \sqrt{p_d} \overline{y}_1) (y^d_1 - \sqrt{p_d} \overline{y}_1) = \sum_{d=1}^{k} y^d_1 y^d_1 - N \overline{y}_1 \overline{y}_1.
\]

Maxima of the likelihood functions.

Using Lemmas 3.1-3.3, we can find the maxima of the likelihood functions \( L(y^1, \ldots, y^k, \nu^1, \ldots, \nu^k) \) over the regions:

\[
(7.6) \quad \omega_1: \nu^1 = \ldots = \nu^k, \quad \Sigma^1 = \ldots = \Sigma^k,
\]

\[
(7.7) \quad \omega_2: \Sigma^1 = \ldots = \Sigma^k,
\]

\[
(7.8) \quad \omega_3: \nu^1_1, \ldots, \nu^k_1 \text{ and } (\Xi^d_1, \Xi^d_2), \text{ } d=1, \ldots, k \text{ are unrestricted.}
\]

Let \( a(p,N) = (2\pi)^{-pN/2} \). Then the maxima of the likelihood functions are

\[
(7.9) \quad \sup_{\omega_1} L(y^1, \ldots, y^k, \nu^1, \ldots, \nu^k) = \frac{a(p,N) N^{pN/2} (n-1)^{(n-1) \text{rhN/2}}}{|D_{11} + \sum_{d=1}^{k} v^d_{11} |N/2| \sum_{d=1}^{k} (\tilde{H}_d + \sum_{j=2}^{n} \tilde{V}_d^j) | (N/2)(n-1)|},
\]

\[
(7.10) \quad \sup_{\omega_2} L(y^1, \ldots, y^k, \nu^1, \ldots, \nu^k) = \frac{a(p,N) N^{pN/2} (n-1)^{(n-1) \text{rhN/2}}}{| \sum_{d=1}^{k} v^d_{11} |N/2| \sum_{d=1}^{k} (\tilde{H}_d + \sum_{j=2}^{n} \tilde{V}_d^j) | (N/2)(n-1)|},
\]

66
\[(7.11) \quad \sup_{\omega_j} L(y^1, \ldots, y^k, v^1, \ldots, v^k) = \frac{1}{a(p, N)} \prod_{d=1}^{k} \left\{ \frac{p_d^{p_d/2}}{n-1} \frac{(n-1) \gamma p_d^2}{\sum_{j=2}^{n} v_{d}^{(n-1)p_d^2/2}} \right\} \]

We may now generate the LRT for the three type II k-population hypotheses using

\[\lambda_{ij} = \frac{\sup_{\omega_i} L(y^1, \ldots, y^k, v^1, \ldots, v^k)}{\sup_{\omega_j} L(y^1, \ldots, y^k, v^1, \ldots, v^k)} \quad (7.12)\]

The LRT, moments of the LRT under the null distribution, approximate null distribution and asymptotic non-null distributions follow for each of the three k-population type II hypotheses. Methods used to derive them are explained at the end of Section 3 and are illustrated in Appendix I for \(\bar{H}_i(mvc)\). The moments of the LRT and the approximate null distributions do not depend on \(v^d\) and \(\bar{z}^d\), \(d=1, \ldots, k\). The asymptotic non-null distributions do depend on \(v^d\) and \(\bar{z}^d\), \(d=1, \ldots, k\).

Test for \(\bar{H}_k(MVC|mvc)\) versus \(\bar{H}_{k,A}\).

The results needed for testing \(\bar{H}_k(MVC|mvc)\), the hypothesis that the means and covariance matrices for the \(k\) populations are the same given that each population's mean and covariance matrix have BCS-II, versus \(\bar{H}_{k,A}\), the alternative hypothesis that the mean and covariance matrix of each population have BCS-II but are not all the same, follow below.
From (7.9) and (7.11), the LRT $\lambda_{13}$ is given by

\[
(7.13) \quad \lambda_{13}^{2/N} = \frac{n^2 |H^d|}{\frac{k}{d=1} \prod_{d=1}^{k} \left( \frac{p_d}{p_d} \right)^{N^d}} \cdot \frac{\frac{k}{d=1} \sum_{d=1}^{k} \frac{n^d}{V^d^d} \left( \frac{1}{n-1} \right)^d}{\left(D_{11} + \sum_{d=1}^{k} \sum_{j=2}^{k} v_{jj}^d \right)}.
\]

The moments used for deriving the approximate null distribution are given by:

\[
(7.14) \quad \mathcal{S}_{13}^S = K \cdot \frac{n^{pSN/2}}{\frac{k}{d=1} \prod_{d=1}^{k} \frac{p_d}{p_d^S/2}} \cdot \frac{\prod_{d=1}^{k} \prod_{j=1}^{rh} \left[ \Gamma\left( \frac{p_d}{2} (1+S) - \frac{1}{2} \right) \Gamma\left( \frac{p_d}{2} (1+S) + \frac{1}{2} \right) \right]}{\prod_{j=1}^{rh} \left[ \Gamma\left( \frac{N}{2} (1+S) - \frac{1}{2} \right) \Gamma\left( \frac{N}{2} (1+S) + \frac{1}{2} \right) \right]},
\]

where $S=0,1,2,\ldots$, with $K$ chosen so $\mathcal{S}_{13}^0 = 1$.

Using the Box method, the approximate null distribution of $-2 \log \lambda_{13}$ is:

\[
(7.15) \quad P\left(-2 \log \lambda_{13} \leq z\right) = (1-\omega) P\left(\chi^2_f \leq \rho z\right) + \omega P\left(\chi^2_{f+4} \leq \rho z\right) + o\left(\prod_{d=1,\ldots,k} \frac{p_d}{p_d^S/2}\right),
\]

where $f$, $\rho$ and $\omega$ are defined by

\[
f = (k-1)rh(rh+2), \quad \rho = 1 - \frac{(NS_{1,1}^2 - 1)rh}{12NF} d_6,
\]
\[(7.16)\quad d_6 = 2(\rho h)^2 + 9\rho h + 11 + \frac{2(\rho h)^2 + 3\rho h - 1}{n-1},\]
\[
\rho_2 = \frac{(1-\rho)^2}{4} + \frac{(1-\rho)(nS_1-1)\rho}{24N} d_6
\]
\[
+ \frac{(N^2S_2-1)\rho h (h+1)}{48N^2} \left[\frac{(h+2)(h+3) + (\rho h^2 + rh - 2)}{(n-1)^2}\right],
\]

\[(7.17)\quad S_1 = \sum_{d=1}^{k} p_d^{n-1},
\]

\[(7.18)\quad S_2 = \sum_{d=1}^{k} p_d^{n-2}.
\]

Using the standard delta method to find the asymptotic non-null distribution of \((2/N)\log \lambda_{13}\) under the assumption that \(v_1^d, \ldots, v_k^d ; \Xi_{1,d}, \Xi_{2,d} ; d=1, \ldots, k\) are unrestricted, we find that:

\[(7.19)\quad \mathcal{L}\left\{ \sqrt{\frac{2}{N}} \log \lambda_{13} - \log \frac{\prod_{d=1}^{k} \left[ \frac{1}{1} \right]}{|A|} \sum_{d=1}^{k} \frac{f_d \Xi_{1,d}^{d} (n-1)}{\Xi_{2,d}^{d} n} \right\} \rightarrow \mathcal{N}(0, \infty),
\]

\[
A = \sum_{d=1}^{k} f_d (\Xi_{1,d}^{d} + v_1^{d}) - \sum_{d=1}^{k} \sum_{i,j=1}^{n} v_i^{d} v_j^{d} f_i f_j ; \quad B = \sum_{d=1}^{k} f_d \Xi_{2,d}^{d} (n-1);
\]

\[
v_\infty = 2p + 2 \sum_{d=1}^{k} \left[ 2c_d A_{1,d}^{d} - c_d^{d} - 2tr(\Xi_{1,d}A_{1,d}^{d}) + tr(\Xi_{1,d}A_{1,d}^{d})^2 \right] +
\]
\[(n-1)^2((n-1)\text{tr}(\Sigma B^{-1})^2 - 2\text{tr}(\Sigma B^{-1}))],
\]
\[f_d = \frac{p_d}{N}, c_d = \sum_{d=1}^{k} f_r v_1^r, d = 1, \ldots, k.\]

\[S_2\] defined by (7.18).

**Test for \(H_k(VC|mvc)\) versus \(H_{k,A}\).**

The results needed for testing \(H_k(VC|mvc)\), the hypothesis that the covariance matrices for the \(k\) populations are the same given that each population's mean and covariance matrix have BCS-II versus \(H_{k,A}\), the alternative that the mean and covariance matrix of each population have BCS-II but are not all the same, follow below.

From (7.10) and (7.11), the LRT \(\chi_{23}^2\) is given by

\[\frac{(7.20)}{2/N} = \frac{\sum_{d=1}^{k} (\sum_{j=1}^{p} |v_{d,1j}| p_{d/N})}{\sum_{d=1}^{k} \sum_{j=1}^{p} (\sum_{d=1}^{k} |v_{d,1j}| (n-1)p_{d/N})}.\]

The moments used for deriving the approximate null distribution are given by

\[\frac{(7.21)}{\xi \chi_{23}^S} = \frac{K}{k} \frac{pp_{d}S/2}{p_{d}} \times
\]
\[\frac{rh}{j=1} \left[ (\Gamma(p_{d}/2) - \frac{1}{2}) \Gamma\left(\frac{p_{d}(n-1)}{2} (1+S) + \frac{1-2j}{2}\right) \right]
\]
\[\left[ \Gamma\left(\frac{N}{2} (1+S) + \frac{1-k-1}{2}\right) \Gamma\left(\frac{N(n-1)}{2} (1+S) + \frac{1-1}{2}\right) \right].\]

\[S=0,1,2,\ldots, \text{with } K \text{ chosen so } \xi \chi_{23}^0 = 1.\]
Using the Box method, the approximate null distribution of \(-2 \log \lambda_{23}\) is:

\begin{equation}
(7.22) \quad P(-2 \log \lambda_{23} \leq z) = (1-\omega) \cdot P(X_{1}^{2} \leq \rho z) + \omega \cdot P(X_{f+4}^{2} \leq \rho z) + O\left(\min_{d=1, \ldots, k} p_{d}^{-3}\right),
\end{equation}

where \(f, \rho\) and \(\omega\) are defined by

\[ f = (k-1)rh(rh+1), \]

\[ \rho = 1 - \frac{(NS_{0}-1)rh}{12Nf} d_{6} + \frac{(k-1)rh(rh+k+2)}{2Nf}, \]

\[ \rho^{2} \omega = \frac{(1-\rho)^{2} f}{4} - \frac{(1-\rho)rh}{24N} \left\{(NS_{0}-1)d_{6} - 6(k-1)(rh+k+2)\right\} + \frac{(N^{2}S_{0}-1)rh(rh+1)}{48N^{2}} \left\{ (rh+2)(rh+3) + \frac{(rh)^{2}+rh-2}{(n-1)^{2}} \right\} + \frac{(k-1)rh}{16N^{2}} \left\{ 2(k-1)^{2}+3(k-1)(rh+3)+2(rh)^{2}+9rh+11 \right\}, \]

\(d_{6}\) defined by (7.16), \(S_{0}\) by (7.17), \(S_{2}\) by (7.18).

Using the standard delta method to find the asymptotic non-null distribution of \((2/N) \log \lambda_{23}\) under the assumption that

\[ v_{1}, \ldots, v_{k}(z_{1}, z_{2}), d=1, \ldots, k, \] are unrestricted, we find that:

\begin{equation}
(7.23) \quad \sqrt{N} \left[ \frac{2}{N} \log \lambda_{23} - \log \frac{k \left| \sum_{d=1}^{k} p_{d}/\xi_{d} \right| \xi_{d}^{(n-1)p_{d}/N}}{\left| B_{1} \right| \sum_{d=1}^{k} f_{d} \xi_{d}^{n-1}} \right] \rightarrow \mathcal{N}(0, \Sigma_{\infty})
\end{equation}
\[ B_i = \frac{k}{d=1} f_{d1}^{d1}, \quad i = 1, 2, \]

\[ v_\infty = 2p + 2 \sum_{d=1}^{k} f_{d1}^{d1} \left[ -2\text{tr}(\Sigma_{1B_{1}}^{-1}) + \text{tr}(\Sigma_{1B_{1}}^{-1})^2 + (n-1)^2 (n-1) \text{tr}(\Sigma_{2B_{2}}^{-1})^2 - 2\text{tr}(\Sigma_{2B_{2}}^{-1}) \right]. \]

Test for \( H_k(M|mVC) \) versus \( H_k(VC|mVC) \).

The results needed for testing \( H(M|mVC) \), the hypothesis that the means for the \( k \) populations are the same given that the covariance matrices of the populations are the same and that the mean and covariance matrix of each population have BCS-II, versus \( H_k(VC|mVC) \), the alternative hypothesis that the covariance matrices for the \( k \) populations are the same given that the mean and covariance matrix of each population have BCS-II, but the means are not all the same, follow below.

From (7.9) and (7.10), the LRT \( \lambda_{12} \) is given by

\[ (7.24) \quad \lambda_{12}^{2/N} = \left| \sum_{d=1}^{k} v_{11}^{d} \right| / \left| D_{11} + \sum_{d=1}^{k} v_{11}^{d} \right|. \]

The moments used for deriving the approximate null distribution are given by:

\[ (7.25) \quad \xi_{12}^S = K \frac{r^h}{\prod_{i=1}^{r^h}} \left[ \frac{\Gamma(\frac{N}{2}(1+S) + \frac{1-k-i}{2})}{\Gamma(\frac{N}{2}(1+S) - \frac{i}{2})} \right], \quad S = 0, 1, 2, \ldots, \]

with \( K \) chosen so \( \xi_{12}^0 = 1 \).

Using the Box method, the approximate null distribution of \(-2 \log \lambda_{12}\) is
\[(7.26) \quad P(-2\log \lambda_{12} \leq z) = (1-\omega) P(x_f^2 \leq \rho z) + \omega P(x_{f+4}^2 \leq \rho z) + O([\min_{d=1, \ldots, k} p_d]^{-3}),\]

where \( f, \rho \) and \( \omega \) are defined by
\[
f = rh(k-1), \quad \rho = 1 - \frac{rh+k+2}{2N},
\]

\[
\rho^2 \omega = \frac{(1-\rho)^2 f}{4} - \frac{(1-\rho)f(rh+k+2)}{4N} + \frac{f}{24N^2} \left[ 2(rh)^2 + 3rh(k+2) + 4k^2 + 5k + 4 \right].
\]

Using the standard delta method to find the asymptotic distribution of \((2/N)\log \lambda_{12}\) under the assumption that \(v_1^l, \ldots, v_k^l\) are unrestricted and \(\Xi^*_1 = \cdots = \Xi^*_k = \Xi^{*}_1\), \(i=1,2\), we find that

\[(7.27) \quad \sqrt{N} \left\{ \frac{2}{N} \log \lambda_{12} - \log \left| \frac{A_1^*}{A} \right| \right\} \rightarrow \mathcal{N}(0, \nu_\infty),
\]

\[
A = \Xi^*_1 + \sum_{d=1}^k f_d v_1^d v_1^d - \sum_{i,j=1}^k f_{ij} v_1^i v_1^j,
\]

\[
\nu_\infty = 2rh + 2 \sum_{d=1}^k f_d [2c_1 A^{-1} \Xi^*_1 A^{-1} c_1 - 2\text{tr}(\Xi^*_1 A^{-1}) + \text{tr}(\Xi^*_1 A^{-1})^2].
\]
II. TESTING AND ESTIMATION FOR A $4 \times 4$ CIRCULAR STATIONARY MODEL WHEN $\rho_2 = 0$

1. Introduction.

In this chapter we consider the problem of testing and estimation for a $4 \times 4$ circular stationary model when $\rho_2 = 0$. Specifically, suppose we have an independent sample of 4-variate row vectors, $x_1, \ldots, x_n$ where $\mathbf{z}(x_i) = \mathbf{n}(0, \Sigma)$, $i = 1, \ldots, n$ where the covariance matrix $\Sigma$ is a circulant, i.e.,

$$
\Sigma = \sigma^2 \begin{bmatrix}
1 & \rho_1 & \rho_2 & \rho_1 \\
\rho_1 & 1 & \rho_1 & \rho_2 \\
\rho_2 & \rho_1 & 1 & \rho_1 \\
\rho_1 & \rho_2 & \rho_1 & 1
\end{bmatrix}.
$$

We are interested in estimating $\sigma^2$ and $\rho_1$ under the assumption that $\rho_2 = 0$ and testing the null hypothesis $H_0: \rho_2 = 0$ versus the alternative hypothesis $H_A: \rho_2 \neq 0$.

Olkin and Press (1969) have found the maximum likelihood estimates of a general $p \times p$ circulant and have considered the problem of testing the null hypothesis that $\Sigma$ is an intraclass correlation matrix versus the alternative hypothesis that $\Sigma$ is a circulant. In our simple $4 \times 4$ model, this is equivalent to testing the null hypothesis $H_0: \rho_2 = \rho_1$ versus the alternative hypothesis $H_A: \rho_2 \neq \rho_1$. Olkin and Press explain that the circulant model arose from a physical model in which observations are made at the vertices of a regular polygon. Our model consists
of observations being made at 4 vertices of a square where we assume there is non-zero correlation between adjacent vertices and zero correlation between non-adjacent vertices.

Rao (1948), exhibits data of this form in a problem concerned with the weights of cork borings for 28 trees. The weight of the cork borings on the north, east, south and west sides of each tree is reported. Other examples of a circulant pattern in which one may want to test that certain higher order correlations are zero may occur in a problem where a directional signal in the center of a regular polygon is transmitted in a random direction to receivers at the vertices of the polygon.

In section 2 we use an orthogonal matrix suggested by Olkin and Press to rotate the circulant matrix $E$ given in (1.1) to a diagonal form. This along with the principle of sufficiency puts the problem into a canonical form which we use for testing and estimation in both the circulant case and the case where $\rho_2 = 0$. The known estimates for the parameters of the circulant are easily derived at the end of section 2. The problem of estimating the parameters when $\rho_2 = 0$ is considered in section 3 and the maximum likelihood estimates (MLE) are given at the end of this section. Section 4 contains additional discussion on the nature of the roots of the MLE. In section 5 the likelihood ratio test (LRT) for testing $H_0: \rho_2 = 0$ versus $H_1: \rho_2 \neq 0$ is derived and its asymptotic null distribution is stated. Section 6 contains the derivation of the asymptotic non-null distribution of the LRT and explains the table of asymptotic means and variances that appear.
in Appendix II. Section 7 contains a discussion of the asymptotic efficiency of the MLE derived for the case \( \rho_2 = 0 \) and compares them with another set of estimators which are easily calculated but less efficient.

2. Canonical Forms and Estimation When \( \rho_2 \neq 0 \).

Consider again the problem of estimating the parameters of \( \Sigma \) given the sample \( x_1, \ldots, x_n \) when \( \mathcal{L}(x_i) = \eta(0, \Sigma), \ i = 1, \ldots, n \), where \( \Sigma \) is of the form

\[
\Sigma = \sigma^2 \begin{bmatrix}
1 & \rho_1 & \rho_2 & \rho_1 \\ 
\rho_1 & 1 & \rho_1 & \rho_2 \\ 
\rho_2 & \rho_1 & 1 & \rho_1 \\ 
\rho_1 & \rho_2 & \rho_1 & 1 \\
\end{bmatrix}.
\]

(2.1)

This is equivalent to the problem of estimating \( \Sigma \) in the problem where we are given \( A = \sum_{i=1}^{n} x_i' x_i \), \( \mathcal{L}(A) = W(\Sigma, 4, n) \). Using the orthogonal matrix \( \Gamma \),

\[
\Gamma = \frac{1}{2} \begin{bmatrix}
1 & 1 & 1 & 1 \\ 
1 & 1 & -1 & -1 \\ 
1 & -1 & -1 & 1 \\ 
1 & -1 & 1 & -1 \\
\end{bmatrix},
\]

(2.2)

we define \( B = \Gamma A \Gamma' \). We then know that \( \mathcal{L}(B) = W(\Sigma^*, 4, n) \) where \( \Sigma^* = \Gamma \Sigma \Gamma' \) and \( \Sigma^* \) is a diagonal matrix of the form
(2.3) \[ \Sigma^* = \sigma^2 \text{diag}(1+2\rho_1+\rho_2, 1-\rho_2, 1-\rho_2, 1-2\rho_1+\rho_2) \]

Let \( u = b_{11}/n, v = b_{44}/n \) and \( w = (b_{22}+b_{33})/n \). We note that \( u, v \) and \( w \) are independent and by sufficiency the problem is reduced to the problem of estimation of the parameters \( \sigma^2, \rho_1 \) and \( \rho_2 \) given \( u, v \) and \( w \) where

\[
(2.4) \quad \mathcal{L} \left( \frac{nu}{\sigma^2(1+2\rho_1+\rho_2)} \right) = \chi^2_n \quad \mathcal{L} \left( \frac{nv}{\sigma^2(1-2\rho_1+\rho_2)} \right) = \chi^2_n \quad \mathcal{L} \left( \frac{nw}{\sigma^2(1-\rho_2)} \right) = \chi^2_{2n}.
\]

We first consider the problem of estimation when \( \rho_2 \neq 0 \). Letting \( \tau_1 = \sigma^2(1+2\rho_1+\rho_2), \tau_2 = \sigma^2(1-2\rho_1+\rho_2) \) and \( \tau_3 = \sigma^2(1-\rho_2) \), we find the MLE of \( \hat{\tau}_i, i = 1, 2, 3 \), to be \( \hat{\tau}_1 = u, \hat{\tau}_2 = v \) and \( \hat{\tau}_3 = w/2 \). Thus the MLE of \( \sigma^2, \rho_1 \) and \( \rho_2 \) are given by \( \hat{\sigma}^2 = (u+v+w)/4, \hat{\rho}_1 = (u-v)/(u+v+w) \) and \( \hat{\rho}_2 = (u+v-w)/(u+v+w) \). In the next section we work from the canonical form given in (2.4) and obtain MLE for \( \sigma^2 \) and \( \rho_1 \) when \( \rho_2 = 0 \).

3. **MLE When \( \rho_2 = 0 \).**

In this section, we start with the canonical form given by (2.4), setting \( \rho_2 = 0 \). We omit the subscript on \( \rho_1 \) and denote \( \rho_1 \) by \( \rho \). We write down the likelihood function \( L(\sigma^2, \rho) \) and the log-likelihood function \( \log L(\sigma^2, \rho) \) and investigate the solutions of the maximum likelihood equations.
We find that one of the maximum likelihood equations can be transformed into a third degree polynomial in \( \rho \), independent of \( \sigma^2 \). This can be solved explicitly and its solution is discussed at the end of this section.

If \( a \) is a real constant such that \( a > 0 \), and \( Z \) is a random variable with the property that \( \mathcal{L}(az) = \chi^2_n \), then the density of \( Z \) is given by

\[
(3.2) \quad f_z(z) = \frac{a^{n/2} \Gamma(n/2-1)}{\Gamma(n/2)2^{n/2}} \exp\left(\frac{-az}{2}\right), \quad z > 0.
\]

The likelihood function is derived using the canonical form in (2.4) and the density (3.2) and is given by

\[
(3.3) \quad L(\sigma^2, \rho) = \frac{n^{n/2} \Gamma(n/2-1)}{\Gamma(n/2)2^{n/2}} [\sigma^2(2\rho)]^{-n/2} \exp\left(\frac{-nu}{2\sigma^2(1+2\rho)}\right) \times \frac{n^{n/2} \Gamma(n/2-1)}{\Gamma(n/2)2^{n/2}} [\sigma^2(1-2\rho)]^{-n/2} \exp\left(\frac{-nv}{2\sigma^2(1-2\rho)}\right) \times \frac{n^{n/2} \Gamma(n/2-1)}{\Gamma(n/2)2^{n/2}} [\sigma^2]^n \exp\left(\frac{-nw}{2\sigma^2}\right).
\]

Taking logarithms of the likelihood function (3.3), we find the log-likelihood function \( \log L(\sigma^2, \rho) \) is.
\[(3.4) \quad \log L(\sigma^2, \rho) = \log K - 2n \log (\sigma^2) - (n/2) \log (1-4\rho^2) \]
\[- \frac{nu}{2\sigma^2(1+2\rho)} - \frac{nv}{2\sigma^2(1-2\rho)} - \frac{nw}{2\sigma^2}, \]

where \(K\) is a constant independent of \(\sigma^2\) and \(\rho\). Taking the first partial derivatives of the log-likelihood function with respect to \(\sigma^2\) and \(\rho\) and setting them to zero as in \((3.1)\), we find the maximum likelihood equations:

\[(3.5) \quad - \frac{2n}{\sigma^2} + \frac{nu}{2\sigma^4(1+2\rho)} + \frac{nv}{2\sigma^4(1-2\rho)} + \frac{nw}{2\sigma^4} = 0, \]

\[(3.6) \quad \frac{4n\rho}{1-4\rho^2} + \frac{nu}{\sigma^2(1+2\rho)^2} - \frac{nv}{\sigma^2(1-2\rho)^2} = 0. \]

Simplifying these equations, we find that

\[(3.7) \quad 4\sigma^2 = \frac{u}{1+2\rho} + \frac{v}{1-2\rho} + w, \]

\[(3.8) \quad 4\sigma^2 \rho(1-4\rho^2) = v(1+2\rho)^2 - u(1-2\rho)^2. \]

Multiplying equations \((3.7)\) by \(\rho(1-4\rho^2)\), and then equating the left-hand side of equations \((3.7)\) and \((3.8)\), we arrive at the cubic polynomial in \(\rho\), independent of \(\sigma^2\)

\[(3.9) \quad 4w\rho^3 + 2(v-u)\rho^2 + (3u+3v-w)\rho + (v-u) = 0. \]
Since a cubic polynomial has explicit solutions, there exist explicit solutions to the maximum likelihood equations. We need only find the three solutions to equation (3.9), rejecting those that are not real or do not satisfy $|\rho| < 1/2$. The condition $|\rho| < 1/2$ is needed for positive definiteness. If more than one solution remains, we choose the solution, which along with the corresponding values of $\sigma^2$ obtained from equation (3.7), maximizes the likelihood function, $L(\sigma^2, \rho)$, given by (3.3). Denote the solution by $\hat{\rho}$ and $\hat{\sigma}^2$.

In the next section, we investigate the possibility of three positive definite roots of the MLE. In this case, we find that two roots are relative maxima and one is a saddle point. In section 5, we return to the problem of finding the IRT and its asymptotic distribution.

4. Roots of the MLE.

In this section, we characterize the roots of the MLE. First we concentrate on the cubic polynomial in $\rho$ given in (3.9),

$$f(\rho) = 4w\rho^3 + 2(v-u)\rho^2 + (3v+3v-w)\rho + (v-u) = 0. \tag{4.1}$$

After studying the three possible roots of the cubic polynomial, we substitute the values of $\rho$ into the MLE (3.7) of $\sigma^2$ as a function of $\rho$. We then examine these three possible roots of the MLE and characterize them in terms of maximizing the likelihood equation (3.3).

First, we examine the roots of the cubic polynomial (4.1). We note that without loss of generality, we can fix $w$. Also, for fixed $w$,
interchanging $u$ and $v$ results in reversing the sign of all roots. While we can solve explicitly for the roots of this cubic polynomial, this procedure becomes too complicated. Instead the following characterization of the roots can be obtained by looking at certain extreme cases. With $w$ fixed, these cases include the cases of $u = v$ and $u = 0$.

In figure (4.1), the first quadrant of the $u$-$v$ plane with $w$ fixed is shown. The first quadrant is divided up depending on the various characteristics of the roots of the cubic polynomial in $\rho$. Note that for $|\rho| < 1/2$, necessary for positive definiteness, $\sigma^2$ obtained from (3.9) is always positive. See figure (4.1).

![Figure 4.1](image_url)
In the non-shaded area A, there is only one real root. In the shaded areas, there are three real roots. In regions \( B_1 \) and \( B_2 \), however, only one of these three real roots satisfies \(|\rho| < 1/2\). In region C, there are three real roots, all of which satisfy \(|\rho| < 1/2\). In this case, we would evaluate all three pairs of \((\sigma^2, \rho)\) values which satisfy the maximum likelihood equations and choose the pair which maximizes the likelihood function \((3.3)\). The arrows in Figure 4.1 indicate what happens to these regions as \( w \) increases. Region C increases in area and regions \( B_1 \) and \( B_2 \) move away from the origin. The regions are symmetric about the line \( w = v \).

The case of three real roots satisfying the positive definite condition \(|\rho| < 1/2\), (region C), deserves further attention. By looking at the second partial derivatives of the likelihood function with respect to \( \rho \) and/or \( \sigma^2 \), we discover that these three positive definite roots of the MLE consist of two relative maxima and one saddle point. If we order the three roots depending on the parameter \( \rho \), we find the middle root is the saddle point. Of the remaining two roots, each a relative maxima, we find the global maximum by substitution into the likelihood function. We observe if \( u > v \), the global maximum occurs at the root with most positive \( \rho \), if \( u < v \), at the most negative \( \rho \). In the case where \( u = v \), an event of probability zero, we may select either extreme root to maximize the likelihood function.

In the next section, we return to the problem of finding the IRT and its asymptotic distribution.
5. LRT and Its Asymptotic Distribution When $\rho_2 = 0$.

Define the parameter space $\Omega$ for the circulant model and the parameter space $\omega$ for the circulant model with $\rho_2 = 0$ as follows:

(5.1) \[ \Omega = \{(\sigma^2, \rho_1, \rho_2) : \sigma^2 > 0, \ |\rho_1| < (1+\rho_2)/2, \ |\rho_2| < 1\} , \]

(5.2) \[ \omega = \{(\sigma^2, \rho_1, \rho_2) : \sigma^2 > 0, \ |\rho_1| < 1/2, \ \rho_2 = 0\} . \]

These values of $(\sigma^2, \rho_1, \rho_2)$ are the values which keep $\Sigma$ in (2.1) positive definite which is equivalent to $\Sigma^*$ in (2.3) being positive definite.

Substitution of the MLE found at the end of section 2 for $\Omega$, into $L(\sigma^2, \rho_1, \rho_2)$ yields

(5.3) \[ \sup_{\Omega} L(\sigma^2, \rho_1, \rho_2) = K \left( \frac{u \nu^2}{4} \right)^{-n/2} \exp(-2n) . \]

Substitution of the MLE found at the end of section 3 for $\omega$, into $L(\sigma^2, \rho_1, \rho_2)$, along with equation (3.7) used to simplify the exponential term, gives

(5.4) \[ \sup_{\omega} L(\sigma^2, \rho_1, \rho_2) = K \left[ \sigma^8 (1-4\rho_1^2) \right]^{-n/2} \exp(-2n) . \]

The LRT $\lambda$ is defined by
\( (5.5) \quad \lambda = \frac{\pi \sup_{\Omega} L(\sigma^2, \rho_1, \rho_2)}{\sup_{\Omega} L(\sigma^2, \rho_1, \rho_2)} . \)

Substitution of (5.3) and (5.4) into (5.5) results in

\( (5.6) \quad \lambda = \left( \frac{uvw}{4\sigma^2 \left( 1-4\rho^2_1 \right)} \right)^{\frac{n}{2}} . \)

We reject \( H_0 : \rho_2 = 0 \) if \( \lambda \) is too small. Because we do not have an explicit form for \( \lambda \), it is not as easy to calculate the moments of \( \lambda \) so we use the asymptotic distribution of \(-2 \log \lambda\), i.e.,

\( (5.7) \quad \mathcal{L}(-2 \log \lambda) \rightarrow \chi^2_1 , \)

and reject \( H_0 : \rho_2 = 0 \) for large values of \(-2 \log \lambda\).

In section 6, we obtain the asymptotic distribution of \(-\left(\frac{2}{n}\right) \log \lambda\) under the alternative hypothesis.

6. **The Asymptotic Non-null Distribution of \(-\left(\frac{2}{n}\right) \log \lambda\).**

In this section, the asymptotic non-null distribution of \(-\left(\frac{2}{n}\right) \log \lambda\) is obtained using the standard delta method. This result is stated in theorem 6.1 for completeness. It is applied to our problem to yield asymptotic means and variances which appear in table A of appendix II. The section concludes with a short discussion of these tabulated values.

Anderson (1958, p. 76) states the standard delta method as follows:
Theorem 6.1. Let \( Q(n) \) be an \( m \)-component random vector and \( b \) a fixed vector. Assume \( n^{1/2}(Q(n)-b) \) is asymptotically distributed according to \( N(0,T) \). Let \( w = f(q) \) be a scalar function of a vector \( q \) with first and second derivatives existing in the neighborhood of \( u = b \).

Let \( \frac{\partial f}{\partial q_1} \bigg|_{q=b} \) be the \( i \)-th component of \( \mathbf{g}_b \). Then the limiting distribution of \( n^{1/2}[f(Q(n))-f(b)] \) is \( N(0,\mathbf{g}_b^T T \mathbf{g}_b) \).

We recall from (5.6) that

\[
-(2/n) \log \lambda = \log [(4 \hat{\sigma}^2 (1-4 \hat{\rho}_1^2))/(uvw^2)]
\]

\[
= \log 4 + 4 \log \hat{\sigma}^2 + \log (1-4 \hat{\rho}_1^2) - \log u - \log v - 2 \log w.
\]

Observe that \( -(2/n) \log \lambda \) is not an explicit function of \( u, v \) and \( w \) since \( \hat{\sigma}^2 \) and \( \hat{\rho}_1 \) are both functions of \( u, v \) and \( w \).

We let \( Q(n) \) in the theorem be given by \( Q(n) = (u,v,w) \). Since \( u, v \) and \( w \) are independent with distributions under the alternative hypothesis given by

\[
\mathcal{L} \left( \frac{nu}{\sigma^2(1+2\rho_1+\rho_2)} \right) = \chi^2_n, \quad \mathcal{L} \left( \frac{nv}{\sigma^2(1-2\rho_1+\rho_2)} \right) = \chi^2_n, \quad \mathcal{L} \left( \frac{nw}{\sigma^2(1-\rho_2)} \right) = \chi^2_{2n},
\]

we know that \( u, v \) and \( w \) have the following means and variances

\[
\mathcal{E}(u) = \sigma^2(1+2\rho_1+\rho_2) = b_1, \quad \text{Var}(u) = 2[\sigma^2(1+2\rho_1+\rho_2)]^2/n = t_{11}/n,
\]

\[
\mathcal{E}(v) = \sigma^2(1-2\rho_1+\rho_2) = b_2, \quad \text{Var}(v) = 2[\sigma^2(1-2\rho_1+\rho_2)]^2/n = t_{22}/n,
\]

\[
\mathcal{E}(w) = 2\sigma^2(1-\rho_2) = b_3, \quad \text{Var}(w) = 4[\sigma^2(1-\rho_2)]^2/n = t_{33}/n.
\]
Thus, by the standard univariate central limit theorem, we know \( n^{1/2}(q(n) - b) \) is asymptotically distributed as \( N(0, T) \) where 
\[
T = \text{diag}(t_{11}, t_{22}, t_{33}).
\]

Let \( f(q) = -(2/n)\log \lambda \). We next calculate \( \frac{\partial f}{\partial q_i} \), \( i = 1, 2, 3 \), as

\[
\frac{\partial f}{\partial q_1} = \frac{\partial f}{\partial u} = \frac{4}{\sigma^2} \frac{\partial \sigma^2}{\partial u} - \frac{8 \hat{\beta}_1}{1-4\hat{\beta}_1} \frac{\partial \hat{\beta}_1}{\partial u} - \frac{1}{u},
\]

\[
\frac{\partial f}{\partial q_2} = \frac{\partial f}{\partial v} = \frac{4}{\sigma^2} \frac{\partial \sigma^2}{\partial v} - \frac{8 \hat{\beta}_1}{1-4\hat{\beta}_1} \frac{\partial \hat{\beta}_1}{\partial v} - \frac{1}{v},
\]

\[
\frac{\partial f}{\partial q_3} = \frac{\partial f}{\partial w} = \frac{4}{\sigma^2} \frac{\partial \sigma^2}{\partial w} - \frac{8 \hat{\beta}_1}{1-4\hat{\beta}_1} \frac{\partial \hat{\beta}_1}{\partial w} - \frac{2}{w}.
\]

Using the maximum likelihood equation (3.7)

\[
\sigma^2 = \frac{1}{4} \left( \frac{u}{1+2\hat{\beta}_1} + \frac{v}{1-2\hat{\beta}_1} + w \right),
\]

we calculate \( \frac{\partial \sigma^2}{\partial q_i} \), \( i = 1, 2, 3 \), to be

\[
\frac{\partial \sigma^2}{\partial u} = \frac{1}{4} \left( \frac{1}{1+2\hat{\beta}_1} \right) + \frac{1}{2} \left( \frac{v}{(1-2\hat{\beta}_1)^2} - \frac{u}{(1+2\hat{\beta}_1)^2} \right) \frac{\partial \hat{\beta}_1}{\partial u},
\]

\[
\frac{\partial \sigma^2}{\partial v} = \frac{1}{4} \left( \frac{1}{1-2\hat{\beta}_1} \right) + \frac{1}{2} \left( \frac{v}{(1-2\hat{\beta}_1)^2} - \frac{u}{(1+2\hat{\beta}_1)^2} \right) \frac{\partial \hat{\beta}_1}{\partial v},
\]

\[
\frac{\partial \sigma^2}{\partial w} = \frac{1}{4} + \frac{1}{2} \left( \frac{v}{(1-2\hat{\beta}_1)^2} - \frac{u}{(1+2\hat{\beta}_1)^2} \right) \frac{\partial \hat{\beta}_1}{\partial w}.
\]

Using the maximum likelihood equation (3.9)
\[(6.13) \quad 4w\rho_1^3 + 2(v-u)\rho_1^2 + (3u+3v-w)\rho_1 + (v-u) = 0, \]

we can calculate \( \frac{\partial \rho_1}{\partial q_1}, \) \( i = 1, 2, 3. \) To find \( \frac{\partial \rho_1}{\partial u}, \) we take the partial derivative \((\partial/\partial u)\) on both sides of equation \((6.13),\) yielding

\[(6.14) \quad \frac{12w\rho_1^2}{\partial u} \frac{\partial \rho_1}{\partial u} - 2\rho_1^2 + 4(v-u)\rho_1 \frac{\partial \rho_1}{\partial u} + 3\rho_1 + (3u+3v-w) \frac{\partial \rho_1}{\partial u} + 1 = 0. \]

Solving \((6.14)\) for \( \frac{\partial \rho_1}{\partial u} \) yields

\[(6.15) \quad \frac{\partial \rho_1}{\partial u} = \frac{2\rho_1^2 - 3\rho_1 + 1}{12w\rho_1^2 + 4(v-u)\rho_1 + (3u+3v-w)}. \]

Similarly, we calculate \( \frac{\partial \rho_1}{\partial v} \) and \( \frac{\partial \rho_1}{\partial w} \) to be

\[(6.16) \quad \frac{\partial \rho_1}{\partial v} = -\frac{2\rho_1^2 + 3\rho_1 + 1}{12w\rho_1^2 + 4(v-u)\rho_1 + (3u+3v-w)}, \]

\[(6.17) \quad \frac{\partial \rho_1}{\partial w} = -\frac{\rho_1(4\rho_1^2 - 1)}{12w\rho_1^2 + 4(v-u)\rho_1 + (3u+3v-w)}. \]

Next we begin to evaluate these expressions at \( q = b. \) First we evaluate the maximum likelihood equation in \( \hat{\rho}_1 \) \((6.13)\) which becomes

\[(6.18) \quad 2(1-\rho_2)\tilde{\rho}_1^3 - 2\rho_1\tilde{\rho}_1^2 + (1+2\rho_2)\tilde{\rho}_1\rho_1 - \rho_1 = 0, \]

independent of \( \sigma^2. \) Note that \( \rho_1 \) and \( \rho_2 \) are parameters under the
alternative hypothesis, $\rho_2 \neq 0$ and $\tilde{\rho}_1$ is a solution of (6.18), an asymptotic maximum likelihood equation.

Continuing with the evaluation at $q = b$, we find

\begin{equation}
\frac{\partial \hat{\rho}_1}{\partial u} \bigg|_b = \frac{2\tilde{\rho}_1^2 - 3\tilde{\rho}_1 + 1}{g_1(\rho_1, \rho_2, \tilde{\rho}_1)} \left( \frac{1}{\sigma^2} \right),
\end{equation}

\begin{equation}
\frac{\partial \hat{\rho}_1}{\partial v} \bigg|_b = -\frac{2\tilde{\rho}_1^2 + 3\tilde{\rho}_1 + 1}{g_1(\rho_1, \rho_2, \tilde{\rho}_1)} \left( \frac{1}{\sigma^2} \right),
\end{equation}

\begin{equation}
\frac{\partial \hat{\rho}_1}{\partial w} \bigg|_b = -\frac{\tilde{\rho}_1(4\tilde{\rho}_1^2 - 1)}{g_1(\rho_1, \rho_2, \tilde{\rho}_1)} \left( \frac{1}{\sigma^2} \right),
\end{equation}

\begin{equation}
g_1(\rho_1, \rho_2, \tilde{\rho}_1) = 4[6(1-\rho_2)\tilde{\rho}_1^2 - 4\rho_1\tilde{\rho}_1 + (1+2\rho_2)].
\end{equation}

Evaluating the equations in $\hat{\sigma}^2$ and in $\frac{\partial \sigma^2}{\partial \rho_i}$, $i = 1, 2, 3$, we find that

\begin{equation}
\hat{\sigma}^2 \bigg|_b = \frac{\sigma^2}{4} \left( \frac{1+2\rho_1+\rho_2}{1+2\tilde{\rho}_1} + \frac{1-2\rho_1+\rho_2}{1-2\tilde{\rho}_1} + 2(1-\rho_2) \right),
\end{equation}

\begin{equation}
\frac{\partial \sigma^2}{\partial u} \bigg|_b = \frac{1}{4} \left( \frac{1}{1+2\tilde{\rho}_1} \right) + g_2(\rho_1, \rho_2, \tilde{\rho}_1)\sigma^2 \left( \frac{\partial \hat{\rho}_1}{\partial u} \bigg|_b \right),
\end{equation}

\begin{equation}
\frac{\partial \sigma^2}{\partial v} \bigg|_b = \frac{1}{4} \left( \frac{1}{1-2\rho_1} \right) + g_2(\rho_1, \rho_2, \tilde{\rho}_1)\sigma^2 \left( \frac{\partial \hat{\rho}_1}{\partial v} \bigg|_b \right),
\end{equation}

\begin{equation}
\frac{\partial \sigma^2}{\partial w} \bigg|_b = \frac{1}{4} + g_2(\rho_1, \rho_2, \tilde{\rho}_1)\sigma^2 \left( \frac{\partial \hat{\rho}_1}{\partial w} \bigg|_b \right),
\end{equation}

\begin{equation}
g_2(\rho_1, \rho_2, \tilde{\rho}_1) = \frac{1}{2} \left( \frac{1-2\rho_1+\rho_2}{(1-2\tilde{\rho}_1)^2} - \frac{1-2\rho_1+\rho_2}{(1+2\tilde{\rho}_1)^2} \right).
\end{equation}
Noting that \( \hat{\rho}_1 \) is independent of \( \sigma^2 \) (see 6.18), and that the partial derivatives of \( \hat{\rho}_1 \) are of the form

\[
(6.28) \quad \frac{\partial \hat{\rho}_1}{\partial q_i} \bigg|_b = \frac{h_1(\rho_1, \rho_2, \tilde{\rho}_1)}{\sigma^2}
\]

we observe that \( \frac{\partial \sigma^2}{\partial q_i} \bigg|_b \) is independent of \( \sigma^2 \). Evaluating \( \frac{\partial r}{\partial q_i} \bigg|_b \), \( i = 1,2,3 \), at \( q = b \), we obtain

\[
(6.29) \quad \frac{\partial r}{\partial u} \bigg|_b = \frac{4}{\sigma^2} \left| \frac{\partial \sigma^2}{\partial u} \right|_b - \frac{5\tilde{\rho}_1}{1-4\tilde{\rho}_1^2} \frac{\partial \hat{\rho}_1}{\partial u} \bigg|_b - \frac{1}{\sigma^2(1+2\rho_1 \rho_2)}
\]

\[
(6.30) \quad \frac{\partial r}{\partial v} \bigg|_b = \frac{4}{\sigma^2} \left| \frac{\partial \sigma^2}{\partial v} \right|_b - \frac{5\tilde{\rho}_1}{1-4\tilde{\rho}_1^2} \frac{\partial \hat{\rho}_1}{\partial v} \bigg|_b - \frac{1}{\sigma^2(1-2\rho_1 \rho_2)}
\]

\[
(6.31) \quad \frac{\partial r}{\partial w} \bigg|_b = \frac{4}{\sigma^2} \left| \frac{\partial \sigma^2}{\partial w} \right|_b - \frac{5\tilde{\rho}_1}{1-4\tilde{\rho}_1^2} \frac{\partial \hat{\rho}_1}{\partial w} \bigg|_b - \frac{1}{\sigma^2(1-\rho_2)}
\]

Combining (6.29)-(6.31) with the facts that

(i) \( \tilde{\rho}_1 \) is independent of \( \sigma^2 \),

(ii) \( \frac{\partial \hat{\rho}_1}{\partial q_i} \bigg|_b = \frac{h_1(\rho_1, \rho_2, \tilde{\rho}_1)}{\sigma^2} \), \( i = 1,2,3 \)

(iii) \( \frac{\partial \sigma^2}{\partial q_i} \bigg|_b \) is independent of \( \sigma^2 \), \( i = 1,2,3 \),

(iv) \( \sigma^2 \bigg|_b = \sigma^2_m(\rho_1, \rho_2, \tilde{\rho}_1) \),

we observe that
\[ \frac{\partial r}{\partial q_1} \bigg|_b = \frac{r_1(\rho_1, \rho_2, \tilde{\rho}_1)}{\sigma^2}, \ i = 1, 2, 3. \]

Thus, using theorem 6.1, we find that

\[ F\{n^{1/2}[-(2/n) \log \lambda - (-2/n)\log \lambda \big|_b]\} \to \mathcal{N}(0, \nu_\infty), \]

where \( \nu_\infty \) is given by

\[ \nu_\infty = \sum_{i=1}^{3} \left( \frac{\partial r}{\partial q_1} \bigg|_b \right)^2 t_{1i} \]

\[ = \left( \frac{r_1}{\sigma} \right)^2 \left[ 2\sigma^2(1+2\rho_1+\rho_2) \right]^2 + \left( \frac{r_2}{\sigma} \right)^2 \left[ 2\sigma^2(1-2\rho_1+\rho_2) \right]^2 \]

\[ + \left( \frac{r_3}{\sigma} \right)^2 4\sigma^2(1-\rho_2)^2, \]

\[ = 2(r_1^2(\rho_1, \rho_2, \tilde{\rho}_1)(1+2\rho_1+\rho_2)^2 + r_2^2(\rho_1, \rho_2, \tilde{\rho}_1)(1-2\rho_1+\rho_2)^2 \]

\[ + 2r_3^2(\rho_1, \rho_2, \tilde{\rho}_1)(1-\rho_2)^2. \]

Note that \( \nu_\infty \) is independent of \( \sigma^2 \), depending only on \( \rho_1 \) and \( \rho_2 \).

Also, it is easily seen that \( -(2/n)\log \lambda \big|_b \) is independent of \( \sigma^2 \) where

\[ (6.35) \quad -(2/n)\log \lambda \big|_b = \log \left\{ \frac{\sigma^2(1+2\rho_1+\rho_2)(1-2\rho_1+\rho_2)(1-\rho_2)^2}{(\sigma^2 \big|_b)^2(1-4\rho_1^2)} \right\}, \]

by recalling that \( \sigma^2 \big|_b = \sigma^2(m(\rho_1, \rho_2, \tilde{\rho}_1)). \)

90
It is easy to calculate \((-\frac{2}{n})\log \lambda_b\) and \(v_\infty\) at any point in the parameter space \(\Psi\), where

\[
(6.36) \quad \Psi = \{(\rho_1, \rho_2): |\rho_2| < 1, \rho_2 \neq 0, |\rho_1| < \frac{1+\rho_2}{2}\}.
\]

For a given point \((\rho_1, \rho_2) \in \Psi\), one calculates \((\rho_1, \sigma^2_b)\) from equations (6.18) and (6.23). In the case of more than one real root for \(\tilde{\rho}_1\) satisfying \(|\tilde{\rho}_1| < 1/2\), use the point \((\tilde{\rho}_1, \sigma^2_b)\) which minimizes \(-\frac{2}{n}\log \lambda_b\) given by (6.35). Note this is equivalent to using the point which maximizes the likelihood function. Since we know \(-\frac{2}{n}\log \lambda_b\) and \(v_\infty\) are independent of \(\sigma^2\), we can set \(\sigma^2 = 1\) in all expressions used to calculate these quantities. The asymptotic variance, \(v_\infty\), is symmetric about \(\rho_1 = 0\) since

\[
(6.37) \quad v_\infty(\rho_1, \rho_2) = v_\infty(-\rho_1, \rho_2).
\]

Appendix II contains a table of the values of \(-\frac{2}{n}\log \lambda_b\) (denoted "MEAN") and \(v_\infty\) (denoted "VAR") for values of \(\rho_2\) (denoted P-2) from -.99 to +.99 by increments of .03 (excluding \(\rho_2 = 0\)) and eleven equally spaced values of \(\rho_1\) (denoted P-1) starting at 0 and ending at \(\frac{10}{11} \left(\frac{1+\rho_2}{2}\right)\). Note that the value \(\rho_2 = 0\) does not belong to the alternative hypothesis parameter space and that when the asymptotic variance is obtained by the standard delta method at \(\rho_2 = 0\), one finds \(v_\infty = 0\). This occurs because in solving for the MLE under the null hypothesis we have set \(\frac{\partial f}{\partial q_i} = 0\), \(i = 1, 2, 3\). The same phenomenon occurs
in the asymptotic non-null distributions in Chapter I when \( v_\infty \) is evaluated at a point under the null hypothesis. The asterisks "\(*" in the table indicate values of \( (\rho_1, \rho_2) \) where three real roots occur in the solution to equation (6.18). This occurs when \( \rho_2 \) is close to 1 or -1. (See Figure 6.1).

![Region A: One real root](image)

Region B: Three real roots, only one satisfies \(|\tilde{\rho}_1| < 1/2\).

Region C: Three real roots, all satisfy \(|\tilde{\rho}_1| < 1/2\).

Figure 6.1

When \( \rho_2 \) is close to 1, only one of the three real roots satisfies \(|\tilde{\rho}_1| < 1/2\). When \( \rho_2 \) is close to -1, all three real roots satisfy \(|\tilde{\rho}_1| < 1/2\) so one must choose the pair \((\tilde{\rho}_1, \hat{\sigma}^2)\) which minimizes \(-(2/n)\log \lambda\)_{b}.

7. **Asymptotic Efficiencies.**

In this section, we compare the efficiencies of two estimates of \( \sigma^2 \) and \( \rho_1 \) in the canonical problem (2.4) when \( \rho_2 = 0 \). In this problem, we are given the independent random variables \( u, v \) and \( w \) with distributions given by
(7.1) \[ \mathcal{L} \left( \frac{nu}{\sigma^2 (1+2\rho_1)} \right) = \chi^2_n, \quad \mathcal{L} \left( \frac{nv}{\sigma^2 (1-2\rho_1)} \right) = \chi^2_n, \]

(7.2) \[ \mathcal{L} \left( \frac{nv}{\sigma^2} \right) = \chi^2_{2n}, \]

At the end of Section 3, we found the MLE for this problem, \( \hat{\sigma}^2 \) and \( \hat{\rho}_1 \). These estimates are consistent, asymptotically normal, and asymptotically efficient.

Suppose we are not given the random variable \( w \). In this case, we easily obtain the MLE for \( \sigma^2 \) and \( \rho_1 \) using only (7.1) to be

(7.3) \[ \hat{\sigma}^2 = \frac{u+v}{2}, \quad \hat{\rho}_1 = \frac{u-v}{2(u+v)}. \]

In this reduced problem, these MLE are consistent, asymptotically normal, and asymptotically efficient. They have the advantage over \( \hat{\sigma}^2 \) and \( \hat{\rho}_1 \) in that they are easily derived. However, they are only 1/3 to 1/2 as efficient as \( \sigma^2 \) and \( \rho_1 \) depending upon the actual value of \( \rho_1 \), a fact we now prove.

To compare these two estimators, we calculate their efficiency by using concentration ellipsoids as suggested by Anderson (1958, p. 57). A discussion from Anderson (1958) of concentration ellipsoids and their relationship to efficiency is presented here for completeness. First we give the definition of the concentration ellipsoid for a p-component random vector \( Y \) as
**Definition 7.1.** If $Y$ is a $p$-component vector, where $Y = \nu$ and $\xi(Y-\nu)'(Y-\nu) = \psi$, then $(Y-\nu)\psi^{-1}(Y-\nu)' = p+2$ is called the concentration ellipsoid of $Y$.

The following two lemmas give the motivation for the definition of efficiency.

**Lemma 7.1.** Let $t$ be an unbiased estimator of a $p$-component vector of parameters of a distribution. Then the ellipsoid

$$
(7.4) \quad (t-\theta)A(t-\theta)' = p+2,
$$

$$
(7.5) \quad A = \xi\left[\frac{\partial \log L}{\partial \theta}\right]'\left[\frac{\partial \log L}{\partial \theta}\right],
$$

lies entirely within the concentration ellipsoid of $t$. $L$ denotes the likelihood function of a sample of $n$ points from the distribution under consideration.

**Lemma 7.2.** If $B$ is a $p \times p$, positive definite, symmetric matrix, then the characteristic roots of $B$ are proportional to the squares of the reciprocals of the lengths of the principal axes of the ellipsoid $xAx' = 1$.

Lemma 7.1 is an extension of the discussion in Cramér (1946, p. 495) for scalar observations, and Lemma 7.2 is given in a discussion of matrix theory in Anderson (1958, p. 340). The definition of efficiency of an estimator follows.
Definition 7.2. The efficiency of the unbiased estimator \( t \) is given by the square of the ratio of the volumes of the ellipsoid defined in Lemma 7.1 in (7.4) to that of the ellipsoid of concentration of \( t \).

Lemma 7.2 along with the fact that the determinant of a matrix is a product of its characteristic roots provides us with the following method for calculating efficiency, stated as Lemma 7.3.

Lemma 7.3. The efficiency of the unbiased estimator \( t \) of \( \theta \) is defined as the ratio of the determinant of the inverse of the covariance matrix of \( t \) to the determinant of \( A \) defined in Lemma 7.1 in (7.5).

In comparing two unbiased estimators of the same parameters, we calculate the efficiency as the ratio of the determinants of the inverses of their respective covariance matrices. Using the fact that \( (\hat{\sigma}^2, \hat{\rho}) \) and \( (\hat{\theta}_2, \hat{\theta}_3) \) are consistent and asymptotically efficient estimators of \( (\sigma^2, \rho) \) in their respective problems, we need only compute

\[
(7.6) \quad \Lambda_i = \frac{\partial \log L_i}{\partial \left( \begin{array}{c} \sigma^2 \\ \rho \end{array} \right)} \left( \frac{\partial \log L_i}{\partial \left( \begin{array}{c} \sigma^2 \\ \rho \end{array} \right)} \right)^{-1}, \quad i = 1, 2,
\]

where \( L_1 \) and \( L_2 \) are the likelihood functions of \( u, v \) and \( w \) and of \( u \) and \( v \), respectively. The ratio of the determinants of \( \Lambda_2 \) to \( \Lambda_1 \) is used for the comparison of efficiencies of the two estimators.

Performing the straightforward calculation of \( \Lambda_1 \) and \( \Lambda_2 \) yields the following results:
\[ \Lambda_1 = \begin{pmatrix} \frac{n}{\sigma^4} + \frac{n}{\sigma^4} & -\frac{4n \rho_1}{\sigma^2 (1 - 4 \rho_1^2)} \\ -\frac{4n \rho_1}{\sigma^2 (1 - 4 \rho_1^2)} & \frac{4n (1 + 4 \rho_1^2)}{(1 - 4 \rho_1^2)^2} \end{pmatrix} \]

(7.7)

\[ \Lambda_1 = \begin{pmatrix} \frac{n}{\sigma^4} & -\frac{4n \rho_1}{\sigma^2 (1 - 4 \rho_1^2)} \\ -\frac{4n \rho_1}{\sigma^2 (1 - 4 \rho_1^2)} & \frac{4n (1 + 4 \rho_1^2)}{(1 - 4 \rho_1^2)^2} \end{pmatrix} \]

(7.8)

Note that \( \Lambda_1 \) and \( \Lambda_2 \) are identical except for the (1,1) element, in which case \( (\Lambda_1)_{11} = 2(\Lambda_2)_{11} \).

Calculating the efficiency as \( |\Lambda_2|/|\Lambda_1| \), which corresponds to the square of the ratio of the volume of the ellipsoid (7.4) involving \( \Lambda_1 \) to the ellipsoid (7.4) involving \( \Lambda_2 \), we find that

\[ \frac{|\Lambda_2|}{|\Lambda_1|} = \frac{1}{2(1 + 2 \rho_1^2)} , \quad |\rho_1| < 1/2 . \]

(7.9)

Thus, the MLE of \( \sigma^2 \) and \( \rho_1 \) are calculated by ignoring the information in \( w \) are 1/3 to 1/2 as efficient of the MLE calculated using \( u, v \) and \( w \).
III. OTHER SPECIFIC COVARIANCE MATRICES OF THE FORM:

\[ \Sigma = \sigma^2 \psi = \sigma^2 (I + \sum_{g=1}^{m} \rho G_g) \]

1. Introduction.

In chapter II we considered one specific case of the above covariance structure, i.e. Σ of the form

\[ (1.1) \quad \Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & 0 & \rho \\ \rho & 1 & \rho & 0 \\ 0 & \rho & 1 & \rho \\ \rho & 0 & \rho & 1 \end{bmatrix} = \sigma^2 \left( I + \rho \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \right). \]

In this chapter, we study several other specific cases of this form for which MLE can be explicitly obtained as solutions to quadratic or cubic polynomials. For some of these problems, the problem of hypothesis testing is considered using the LRT.

Several matrices of this general form which do not have explicit MLE are also introduced. The need for a computational algorithm thus arises. The subject of computational algorithms is treated in chapter IV as well as specific suggestions for solving for the MLE of a covariance matrix of the general form:

\[ (1.2) \quad \Sigma = \sigma^2 \psi = \sigma^2 (I + \sum_{g=1}^{m} \rho G_g) \]

where the \( G_g \) are known symmetric matrices. In section 2 we consider the problem of testing:
\[(1.3) \quad H_0: \Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho & \rho^2 & \rho & 1 \end{bmatrix} \quad \text{versus} \quad H_A: \Sigma = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_1 \\ \rho_1 & 1 & \rho_2 & \rho_1 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_1 & \rho_2 & \rho_1 & 1 \end{bmatrix}.\]

In section 3, the problem of testing:

\[(1.4) \quad H_0: \Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & 0 & \rho \\ \rho & 1 & \rho & 0 \\ 0 & \rho & 1 & \rho \\ \rho & 0 & \rho & 1 \end{bmatrix} \quad \text{versus} \quad H_A: \Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho & \rho^2 & \rho & 1 \end{bmatrix}.\]

is considered. A matrix arising from a stationary Markov is studied in section 4. In this case, \( \Sigma \) has the form:

\[(1.5) \quad \Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{r-1} \\ 1 & \rho & \cdots & \rho^{r-2} & \cdots & \cdots & \cdots & \cdots & 1 \end{bmatrix}.\]

Sampson (1970) studies the case where \( \sigma^2 \) is known. He finds the MLE are solutions of a cubic polynomial. In the case of \( \sigma^2 \) unknown, the MLE are again solutions of a cubic polynomial. Section 5 contains examples of \( \Sigma \) of the form (1.2) for which we can not find explicit MLE. These cases are treated by computational algorithms discussed in chapter IV. Results and canonical forms used in chapter II are used in this chapter.
2. A Special $4 \times 4$ Circular Stationary Model With $\rho_1=\rho^2$.

In this section we consider the problem of estimation and testing under the hypotheses:

\[
H_0: \Sigma = \sigma^2 \begin{bmatrix}
1 & \rho & \rho^2 & \rho \\
\rho & 1 & \rho & \rho^2 \\
\rho^2 & \rho & 1 & \rho \\
\rho & \rho^2 & \rho & 1
\end{bmatrix}
\quad \text{versus} \quad H_A: \Sigma = \sigma^2 \begin{bmatrix}
1 & \rho_1 & \rho_2 & \rho \\
\rho_1 & 1 & \rho_1 & \rho_2 \\
\rho_2 & \rho_1 & 1 & \rho_1 \\
\rho_1 & \rho_2 & \rho_1 & 1
\end{bmatrix}.
\]

Using the notation and canonical forms of chapter II, we note the MLE under the alternative hypothesis are given at the end of section 2 in chapter II. In the canonical form of chapter II, (II.2.4), we set $\rho_i = \rho^i$, $i=1,2$, thus reducing the problem to one of estimating $\sigma^2$ and $\rho$ given independent random variables $u$, $v$ and $w$ with distributions:

\[
(2.2) \quad \mathcal{L} \left( \frac{nu}{\sigma^2 (1+\rho)^2} \right) = \chi^2_n, \quad \mathcal{L} \left( \frac{nv}{\sigma^2 (1-\rho)^2} \right) = \chi^2_n, \\
\quad \mathcal{L} \left( \frac{nw}{\sigma^2 (1-\rho^2)} \right) = \chi^2_{2n}.
\]

From (2.2), we see the loglihood function $\log L(\sigma^2, \rho)$ is of the form

\[
(2.3) \quad \log L(\sigma^2, \rho) = \log K - 2n \log \sigma^2 - 2n \log (1-\rho^2) - \frac{nu}{2\sigma^2 (1+\rho)^2} - \frac{nv}{2\sigma^2 (1-\rho)^2} - \frac{nw}{2\sigma^2 (1-\rho^2)}. 
\]
Taking the partial derivatives of \( \log L(\sigma^2, \rho) \) with respect to \( \sigma^2 \) and \( \rho \), and setting these derivatives equal to zero, we find the maximum likelihood equations:

\[
\sigma^2 = \frac{1}{4} \left( \frac{u}{(1+\rho)^2} + \frac{v}{(1-\rho)^2} + \frac{w}{(1-\rho^2)} \right),
\]

\[
\rho^2 = \frac{1}{4\rho} \left( \frac{u}{(1+\rho)^3} - \frac{v}{(1-\rho)^3} - \frac{wp}{(1-\rho^2)^2} \right).
\]

Combining (2.4) and (2.5) yields the quadratic polynomial in \( \rho \):

\[
(u-v)\rho^2 - 2(u+v)\rho + (u-v) = 0,
\]

with two solutions

\[
\rho = \frac{1}{2} \left( \frac{1}{u-v} \right), \quad \frac{1}{u+v} \frac{1}{u-v}.
\]

We reject the second root since it does not satisfy \( |\rho| < 1 \). Thus the MLE of \( \rho \), denoted \( \hat{\rho} \), is

\[
\hat{\rho} = \frac{1}{2} \left( \frac{1}{u-v} \right).
\]

Substitution of this value into (2.4) gives us the MLE, \( \hat{\sigma}^2 \), of \( \sigma^2 \):

\[
\hat{\sigma}^2 = \left( \frac{1}{2} \frac{1}{u^2 + v^2} \right) \frac{1}{16(uv)^2}.
\]
We use the LRT for testing these hypotheses. Let

\[(2.10) \quad \Omega = \{(\sigma^2, \rho_1, \rho_2): \sigma^2 > 0, \ |\rho_1| < (1 + \rho_2)/2, \ |\rho_2| < 1\},\]

\[(2.11) \quad \omega = \{(\sigma^2, \rho): \sigma^2 > 0, \ |\rho| < 1\} .\]

From the results of chapter II, (II.5.3), we know that

\[(2.12) \quad \sup_{\Omega} L(\sigma^2, \rho_1, \rho_2) = K(\frac{uvw^2}{4})^{-n/2} \exp(-2n).\]

Substituting the MLE under \(H_0\) derived in this section into the likelihood function, we find that

\[(2.13) \quad \sup_\omega L(\sigma^2, \rho) = K[\frac{1}{4} (w+2(\rho v^{1/2}))]^{-2n} \exp(-2n).\]

The LRT \(\lambda\) is given by

\[(2.14) \quad \lambda = \frac{\sup_\omega L(\sigma^2, \rho)}{\sup_{\Omega} L(\sigma^2, \rho_1, \rho_2)} = \left( \frac{64uvw^2}{(w+2(\rho v^{1/2}))^4} \right)^{n/2}.\]

Finding the distribution of \(\lambda\) under the null hypothesis appears to be fairly intractable, so instead we present the standard asymptotic distribution under \(H_0\),

\[(2.15) \quad \lim_{n \to \infty} \mathcal{L}(-2 \log \lambda) = \chi^2_1 .\]
To derive the asymptotic non-null distribution, we use the standard delta method, as described in chapter II, section 6, on the function 

\[-(2/n)\log \lambda. \text{ Here } -(2/n)\log \lambda \text{ is of the form} \]

\[(2.16) \quad -(2/n)\log \lambda = -6 \log 2 + 4 \log \left(w + 2(1 + 2 \rho_1 + \rho_2)^{1/2}\right) - \log(uv)^2.\]

Letting \(f(u, v, w) = -(2/n)\log \lambda\), we find that

\[(2.17) \quad \frac{\partial f}{\partial u} = \frac{4(v/u)^{1/2}}{w + 2(uv)^{1/2}} - \frac{1}{u},\]

\[(2.18) \quad \frac{\partial f}{\partial v} = \frac{4(u/v)^{1/2}}{w + 2(uv)^{1/2}} - \frac{1}{v},\]

\[(2.19) \quad \frac{\partial f}{\partial w} = \frac{4}{w + 2(uv)^{1/2}} - \frac{2}{w}.\]

Using the means and variances of \(u, v\) and \(w\) under \(H_A\) given in chapter II, (II.6.3)-(II.6.5), we evaluate \(f(u, v, w)\) and the partial derivatives at the point \(b = (\xi u, \xi v, \xi w)\), yielding

\[(2.20) \quad f(b) = -(2/n)\log \lambda|_b = \log \left(\frac{(1 - \rho_2 + (1 + 2 \rho_1 + \rho_2)(1 - \rho_1 - \rho_2)^{1/2})}{16(1 + 2 \rho_1 + \rho_2)(1 - 2 \rho_1 + \rho_2)(1 - \rho_2)^2}\right)^{1/4},\]

\[(2.21) \quad \frac{\partial f}{\partial u}|_b = \frac{r}{\sigma^2(1 + 2 \rho_1 + \rho_2)},\]
\[ (2.22) \quad \frac{\partial r}{\partial w} \bigg|_b = \frac{r}{\sigma^2 (1-2\rho_1 + \rho_2)} , \]
\[ (2.23) \quad \frac{\partial r}{\partial w} \bigg|_b = \frac{-r}{\sigma^2 (1-\rho_2)} , \]
\[ (2.24) \quad r = \frac{[ (1-2\rho_1 + \rho_2)(1+2\rho_1 + \rho_2) ]^{1/2} - (1-\rho_2)}{[ (1-2\rho_1 + \rho_2)(1+2\rho_1 + \rho_2) ]^{1/2} + (1-\rho_2)} . \]

Using these terms in the standard delta method, the asymptotic variance \( \nu_\infty \) is
\[ (2.25) \quad \nu_\infty = 8r^2 . \]

Thus, under the alternative hypothesis, the asymptotic distribution of \(- (2/n) \log \lambda\) is given by
\[ (2.26) \quad \lim_{n \to \infty} \sqrt{n} ( - (2/n) \log \lambda - (-2/n) \log \lambda \bigg|_b ) \sim \eta(0, 8r^2) . \]

3. Hypothesis Testing for Two Different Forms of \( \Sigma \), Both Functions of \( \sigma^2 \) and \( \rho \).

In this section we consider testing the hypothesis
\[(3.1) \quad H_0: \Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & 0 & \rho \\ \rho & 1 & \rho & 0 \\ 0 & \rho & 1 & \rho \\ \rho & 0 & \rho & 1 \end{bmatrix} \quad \text{versus} \quad H_A: \Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho & \rho^2 & \rho & 1 \end{bmatrix}.\]

We have already derived the MLE under the null and alternative hypothesis. Using the MLE we form the LRT \( \lambda \). In this case, we cannot use the standard asymptotic results for the distribution of \(-2 \log \lambda\) under the null hypothesis. Instead, we use the standard delta method to derive asymptotic distributions under the null and alternative hypotheses.

The MLE under \( H_0 \) are given in chapter II as solutions to a cubic polynomial. The MLE under \( H_A \) are derived in the previous section of this chapter, section 2. Using these MLE, we form the LRT \( \lambda \),

\[(3.2) \quad \lambda = \frac{\sup_{H_0} L(\sigma^2, \rho)}{\sup_{H_1} L(\sigma^2, \rho)} = \left\{ \frac{\left[ \frac{1}{4} (w + 2(uv)^{1/2}) \right]^4}{\tilde{\sigma}^2 (1 - \tilde{\rho}_1^2)} \right\}^{n/2},\]

where \( \tilde{\sigma} \) and \( \tilde{\sigma}^2 \) are the MLE under \( H_0 \) given in section 3 of chapter II.

We now derive the asymptotic distribution of \(-2/n \log \lambda\) using the standard delta method. It is interesting to note that we do not have the usual asymptotic \( X^2 \) results under \( H_0 \). We relate the partial derivatives used by the standard delta method for this problem to the one detailed in chapter II. Then a short description of the tables of asymptotic means and variances is given. These tables are similar to the one for the asymptotic non-null distribution that is given in Appendix II for the chapter II results.
Letting \( f(u,v,w) = -(2/n)\log \lambda \), the partial derivatives of \( f(u,v,w) \) with respect to \( u, v \) and \( w \) are given by:

\[
\frac{\partial f}{\partial u} = \frac{4}{\sigma^2} \frac{\partial^2}{\partial u} - \frac{8\hat{\rho}_1 \hat{\rho}_2}{1-4\hat{\rho}_1^2} - \frac{4(v/u)^{1/2}}{w+2(uv)^{1/2}},
\]

\[
\frac{\partial f}{\partial v} = \frac{4}{\sigma^2} \frac{\partial^2}{\partial v} - \frac{8\hat{\rho}_1 \hat{\rho}_2}{1-4\hat{\rho}_1^2} - \frac{4(u/v)^{1/2}}{w+2(uv)^{1/2}},
\]

\[
\frac{\partial f}{\partial w} = \frac{4}{\sigma^2} \frac{\partial^2}{\partial w} - \frac{8\hat{\rho}_1 \hat{\rho}_2}{1-4\hat{\rho}_1^2} - \frac{4}{w+2(uv)^{1/2}}.
\]

Combining (3.3)-(3.5) to the similar expressions (II.6.6)-(II.6.8) in chapter II, we observe that they are the same except that in chapter II, the last term in (3.3)-(3.5) is replaced by \( 1/u, 1/v \) and \( 1/w \) respectively. We know \( u, v \) and \( w \) have means and variances given in chapter II, (II.6.3)-(II.6.5) where under \( H_0 \), \( \rho_1=0 \) and \( \rho_2=0 \) and under \( H_A \), \( \rho_1^1=\rho^1 \), \( i=1,2 \). Thus we may use the expression in chapter II with only slight modification to obtain the asymptotic mean and variance of \( -(2/n)\log \lambda \) as a function of \( \rho \) under \( H_0 \) and \( H_A \). These means and variances are given in table B, Appendix II, under \( H_0 \) and \( H_A \).

Only positive values of \( \rho \) are considered since the means and variances for \( \rho \) are the same as for \( -\rho \). Thus, as in chapter II, (II.6.33), we have

\[
\mathcal{L}^{1/2}(-(2/n)\log \lambda - (2/n)\log \lambda_0) \rightarrow N(0,v_\infty).
\]

105
Table C, appearing in Appendix II, contains the asymptotic mean and variance of \(-(2/n)\log \lambda\) under the general \(4 \times 4\) circulant alternative,

\[
\Sigma = \sigma^2 \begin{bmatrix}
1 & \rho_1 & \rho_2 & \rho_3 \\
\rho_1 & 1 & \rho_1 & \rho_2 \\
\rho_2 & \rho_1 & 1 & \rho_1 \\
\rho_3 & \rho_2 & \rho_1 & 1
\end{bmatrix},
\]

(3.7)

Only positive values of \(\rho_1\) are given since the means and variances for \((\rho_1, \rho_2)\) are the same as those for \((-\rho_1, \rho_2)\).


In this section we show that the MLE for the covariance matrix \(\Sigma = \sigma^2 (\rho \mid 1-j \mid)\) can be found as solutions to a cubic polynomial in \(\rho\). By noting that this model can be generated by a first-order stochastic difference equation: (see Anderson, 1969)

\[
x_t - u_t = \rho (x_{t-1} - u_{t-1}) + u_t, \quad t=2, \ldots, r,
\]

(4.1)

[where \(u_2, \ldots, u_r\) are independently distributed according to \(N(0, (1-\rho^2)\sigma^2)\) and \(x_1 - u_1\) is distributed as \(N(0, \sigma^2)\), we know that \(\Sigma\) is positive definite for all \(\sigma^2 > 0\) and \(|\rho| < 1\). In section 5, we consider a similar form where

\[
(\Sigma)_{ij} = \begin{cases} 
\sigma^2 \rho \mid i-j \mid, & \mid i-j \mid \leq s, \\
0, & \text{otherwise}
\end{cases}
\]

(4.2)
This is one of the forms of \( \Sigma = \sigma^2 (I + \sum_{g=1}^{m} \rho^g G_g) \) in which we are not able to find explicit solutions.

Returning to the problem of finding MLE for \( \Sigma \) of the form

\[
(4.3) \quad \Sigma = \sigma^2 \rho^{i-j} = \sigma^2 \begin{bmatrix}
1 & \rho & \rho^2 & \ldots & \rho^{r-1} \\
\rho & 1 & \rho & \ldots & \rho^{r-2} \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
\rho^{r-1} & \vdots & \ddots & \ddots & 1
\end{bmatrix},
\]

we observe that this is a form of a Green's matrix. \([G = (g_{i,j}), g_{i,j} = a_i b_j \text{ for } i \leq j \text{ where } a_i = 1/\rho^i, b_j = \rho^j].\) The inverse of \( \Sigma \) has the form

\[
(4.4) \quad \Sigma^{-1} = \frac{1}{(1-\rho^2)\sigma^2} \begin{bmatrix}
1 & -\rho & 0 \\
-\rho & 1+\rho^2 & -\rho & 0 \\
\vdots & \vdots & \ddots & \ddots \\
\rho^{r-1} & \vdots & \ddots & \ddots & 1
\end{bmatrix}
\]

and the determinant of \( \Sigma \) is given by

\[
(4.5) \quad |\Sigma| = (\sigma^2)^r (1-\rho^2)^{r-1}.
\]

Suppose we have a sample of column vectors \( x_1, \ldots, x_N \) from a normal \( r \)-variate population, \( \mathcal{N}(\mu, \Sigma) \), where \( \mu \) is unstructured and \( \Sigma \) is given by (4.3). Sufficient statistics for \( \mu \) and \( \Sigma \) are \( \bar{x} \) and \( \mbox{C} \).
\( (4.6) \quad \bar{x} = \frac{1}{N} \sum_{\alpha=1}^{N} x_{\alpha}, \quad C = \frac{1}{N} \sum_{\alpha=1}^{N} (x_{\alpha}-\bar{x})(x_{\alpha}-\bar{x})' \).

Since \( \mu \) is unstructured, the MLE of \( \mu \) is \( \bar{x} \). All the information about \( \Sigma \) comes from \( C \). (This is the form that is used in chapter IV.)

Using (4.4) and (4.5), we can express the likelihood function as (see Sampson (1970)),

\( (4.7) \quad (2/N) \log L = -r \log(2\pi) - \log|\Sigma| - \text{tr} \Sigma^{-1} C \)

\[ = -r \log(2\pi) - r \log \sigma^2 - (r-1) \log(1-\rho^2) - \frac{(\text{tr}(C-2\rho V+\rho^2 T))}{\sigma^2(1-\rho^2)}, \]

\[ V = \sum_{i=1}^{r-1} C_{i,i+1}, \quad T = \sum_{i=2}^{r-1} C_{ii}. \]

Taking partial derivatives of \( (2/N) \log L \) with respect to \( \sigma^2 \) and \( \rho \), setting these equal to zero, yields the maximum likelihood equations

\( (4.8) \quad \sigma^2 = \frac{(\text{tr}(C-2\rho V+\rho^2 T))}{r(1-\rho^2)}, \)

\( (4.9) \quad \sigma^2 = \frac{(-\rho^2 V+\rho(\text{tr}(C+T))-V)}{(r-1)\rho(1-\rho^2)}. \)

Combining (4.8) and (4.9) yields a cubic polynomial in \( \rho \)

\( (4.10) \quad \rho^3 - \frac{V}{T} \frac{(r-2)}{(r-1)} \rho^2 - (\frac{\text{tr} C}{T} + r) \frac{\rho}{(r-1)} + \frac{Vr}{T(r-1)} = 0. \)

To study the behavior of (4.10), we need to know the possible
values of \((W_1,W_2)\), where \(W_1 = V/T, \ W_2 = tr \ C/T\). We know these possible values of \((W_1,W_2)\) are generated by all positive definite sample covariance matrices \(C\). Obviously \(W_2\) is greater than 1. This is a necessary condition. We can derive sufficient conditions by looking at other patterned covariance matrices. For example, the \(r \times r\) intraclass correlation matrix, \(\Sigma\), given by

\[
\Sigma = \sigma^2 \begin{bmatrix}
1 & \rho & \cdots & \rho \\
\rho & 1 & \cdots & \vdots \\
\vdots & \ddots & \ddots & \rho \\
\rho & \cdots & \rho & 1
\end{bmatrix},
\]

(4.11)
is positive definite for \(1 > \rho > -1/(r-1)\). From this we know \(W_1\) can assume the values in the interval \((-1/(r-2),(r-1)/(r-2))\). Thus to make a study of the roots of the polynomial in \(\rho\) (4.10) similar to the one in chapter II, we would have to characterize the possible values of \((W_1,W_2)\). This might be important if more than one real root occurs in the allowable parameter region.

However, to find the MLE, we can solve explicitly for the three roots of the cubic (4.10) using the method of Appendix II, and then choose as the MLE the one in the region \(\sigma^2 > 0, \ |\rho| < 1\) which maximizes the likelihood function. The only question remaining is one concerning the possibility of no maximum in the positive definite region. We are guaranteed a maximum for a covariance matrix \(\Sigma\) for which there is at least one value of the parameters for which \(\Sigma\) is positive definite.
Anderson (1969) points out that "the likelihood function \( L \rightarrow 0 \) for \( \Sigma \) approaching a singular matrix or for one or more elements of \( \Sigma \) approaching \( \infty \) and/or \(-\infty\)." Thus, there exists at least one relative maximum.

5. **Examples of Covariance Matrices Without Explicit MLE.**

Several examples of covariance matrices of the form

\[
(5.1) \quad \Sigma = \sigma^2 \Psi = \sigma^2 \left( I + \sum_{g=1}^{m} \rho^g G_g \right)
\]

are given in this section with the property that their MLE can not be explicitly found. The MLE are solutions to polynomials of high order. Interest in these cases motivates the study of computational algorithms appearing in chapter IV for their solution.

One case of interest is the \( r \times r \) circular stationary model where \( \rho_1 = \rho^1 \). The \( r \times r \) circular stationary model is given by

\[
(5.2) \quad \Sigma = \sigma^2 \begin{bmatrix}
1 & \rho_1 & \rho_2 & \cdots & \rho_2 & \rho_1 \\
\rho_1 & 1 & \rho_1 & \cdots & \rho_3 & \rho_2 \\
\rho_2 & \cdots & \cdots & \cdots & \cdots & \rho_1 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\rho_1 & \cdots & \cdots & \cdots & \cdots & 1
\end{bmatrix}
\]

The general \( r \times r \) circular stationary model (5.2) has explicit MLE (see Olkin and Press (1969)). We saw in section 2 of this chapter that in the \( 4 \times 4 \) case, this model restricted by \( \rho_1 = \rho^1, 1 = 1,2 \) has an explicit MLE as a root of a quadratic polynomial. In general, for \( r > 4 \), the MLE appears to be a root of a high degree polynomial.
We are also interested in MLE of $\Sigma$ of the circular stationary form (5.2) where several of the highest order $\rho_1$ are set equal to zero. Explicit MLE are no longer available. We can reparametrize $\Sigma$ so it is of the linear form $\Sigma = \sum_{g=0}^{m} \sigma_{G} G_{g}$, algorithms for whose solution have been suggested by Anderson (1969, 1975). If we set $\rho_1 = \rho^1$ for the lower order $\rho_1$, again we have a form which does not have explicit MLE and which again cannot be reparameterized in Anderson's linear form.

Another model without explicit MLE which is of the form $\Sigma = \sigma^2(I + \sum_{g=1}^{m} \rho G_{g})$ is the one mentioned in section 4 in conjunction with the stationary model. This model is of the form

$$\Sigma_{ij} = \begin{cases} \sigma^2 \rho^{|i-j|}, & |i-j| \leq s, \\ 0, & \text{otherwise}. \end{cases}$$

(5.3)

It is derived from the stationary model, $\Sigma = \sigma^2 \rho^{|i-j|}$ by setting all terms with powers of $\rho$ greater than $s$ equal to zero.

These examples motivate the interest in computational algorithms for MLE for the parameters of $\Sigma$ of the form $\Sigma = \sigma^2(I + \sum_{g=1}^{m} \rho G_{g})$, where $G_{g}$ are known symmetric matrices. It is important to note that in general (for $m > 1$), this form of $\Sigma$ can not be expressed in Anderson's linear form $\Sigma = \sum_{g=0}^{m} \sigma_{G} G_{g}$. A discussion of these computational algorithms appears in chapter IV.
IV. COMPUTATIONAL PROCEDURES

1. **Introduction.**

In this chapter, we consider the problem of estimating $\sigma^2$ and $\rho$ for a patterned covariance matrix of the general form

$$\Sigma = \sigma^2 \psi, \quad \psi = I + \sum_{g=1}^{m} \rho G_g,$$

where the $G_g$ are known symmetric matrices. We have considered specific examples of this form in chapters II and III. These specific examples have the property that their MLE can be explicitly found. In the more general form considered here, this is not possible. Instead, iterative procedures are needed to find the MLE of $\sigma^2$ and $\rho$ for a covariance matrix of the form (1.1). Examples of specific cases of this general form for which explicit MLE can not be found have been given in chapter III, section 5 for special forms of the circular stationary and stationary models.

A brief review of the literature on patterned covariance matrices appears in the introduction to chapter I. Anderson (1969, 1973) has considered the problem of estimation of a patterned covariance matrix of the form

$$\Sigma = \sum_{g=0}^{m} \sigma G_g,$$

where the $G_g$ are known symmetric matrices and the $\sigma_g$ are the unknown parameters. Anderson's 1969 paper uses the Newton-Raphson method and
his 1973 paper uses the method of scoring to find the MLE by iterative methods. Miller (1975a,b) has shown that when explicit closed form solutions to the MLE of the $\sigma_g$ exist, then the method of scoring used by Anderson (1973) gives this closed form solution in one iteration from any initial positive definite starting point. Note that except in the case $m \leq 1$, the patterned covariance matrix (1.1) can not be expressed in the linear form of (1.2).

Barnett (1966) has studied the problem of several roots of the ML equations and points out the possibility that the method of scoring may not converge or may converge to a relative minimum. Vandaele and Chowdhury (1971) propose a revised method of scoring. Miller (1973,1975a,b) considers some problems of slow convergence due to a slowly damped oscillation of the iterates around the maximum. These problems and others that have arisen in obtaining the MLE for the patterned covariance matrix (1.1) are discussed in this chapter. Their applicability is not limited to this particular problem.

In section 2, Newton-Raphson and scoring methods are applied to find the MLE of $\sigma^2$ and $\rho$ for the covariance matrix $\Sigma$ of the form (1.1). In section 3, we present another iterative method, motivated by Anderson's 1973 paper and apply it to our problem. The problem may be reduced to a simpler form in the special case in which the $G_g$ matrices are simultaneously diagonalizable. This simplification is considered in section 4. A discussion of computational difficulties, initial starting points and variable step sizes appears in section 5 and 6.
2. The Newton-Raphson and Scoring Methods.

In this section, a short review of the Newton-Raphson and scoring methods is presented. These methods are then applied to the problem of finding MLE for $\theta$ of the form (1.1). A good short review of some iterative procedures appears in Appendix II of Lawley and Maxwell's (1971) book.

Consider the problem of finding a relative minimum of a function $f(\theta)$, where $\theta$ is a $q$-dimensional vector of parameters. We denote the $i^{th}$ iterate by $\theta^{(i)}$. At any point, there is a gradient vector defined by

$$g(s) = \frac{\partial f}{\partial \theta} \bigg|_{\theta=\theta(s)}.$$  

(2.1)

In addition, there is a matrix of second partials, $G(s)$, defined by

$$G_{ij}^{(s)} = \frac{\partial^2 f(\theta)}{\partial \theta_i \partial \theta_j} \bigg|_{\theta=\theta(s)}.$$  

(2.2)

In the cases that we study, $f(\theta)$ is proportional to the negative loglikelihood function and $\theta$ is a 2-component vector with $\theta_1 = \sigma^2$ and $\theta_2 = \rho$. This function $f(\theta)$ is also a function of the sample. Thus, we can define $\Gamma$ as $G\theta$,

$$\Gamma^{(s)} = (\theta' G^{(s)} \theta).$$  

(2.3)

We assume that $\theta^{(0)}$ is the initial point. The choice of the initial point is discussed in section 5.

The Newton-Raphson method defines a new iterate, $\theta^{(s+1)}$, from the previous iterate $\theta^{(s)}$ by
(2.4) \[ \theta^{(s+1)} = \theta^{(s)} - (G^{(s)})^{-1} g^{(s)}. \]

The method of scoring uses

(2.5) \[ \theta^{(s+1)} = \theta^{(s)} - \Gamma^{(s)} g^{(s)}. \]

Since we are looking for a minimum, we are only interested in cases where \( G^{(s)} \) is positive definite. In general, for the search for a minimum, \( \Gamma^{(s)} \) is also positive definite. The Newton-Raphson procedure is more sensitive to the data than the method of scoring. This can be a disadvantage if the initial point \( \theta^{(0)} \) is not close to the minimum point. The method of scoring being less sensitive to the data, may do better than the Newton-Raphson method when \( \theta^{(0)} \) is not close to the minimum but may not do as well as the Newton-Raphson method when near the minimum. The method of scoring may be easier to compute. Lawley and Maxwell (1971) suggest that for a large number of parameters, the method of scoring may converge much slower than the Newton-Raphson method. At this point, let this suffice to show that one method is not universally superior to the other method. Further comparisons are given in section 5 and 6.

Suppose we have a sample of column vectors \( x_1, \ldots, x_N \) from a normal p-variate population, \( N(\mu, \Sigma) \) where \( \mu \) is unstructured. Sufficient statistics for \( \mu \) and \( \Sigma \) are \( \bar{x} \) and \( C \),

(2.6) \[ \bar{x} = \frac{1}{N} \sum_{\alpha=1}^{N} x_\alpha, \quad C = \frac{1}{N} \sum_{\alpha=1}^{N} (x_\alpha - \bar{x})(x_\alpha - \bar{x})'. \]
Since \( \mu \) is unstructured, the MLE of \( \mu \) is \( \bar{x} \). All the information about \( \Sigma \) comes from \( C \).

We are interested in the MLE of \( \sigma^2 \) and \( \rho \) for \( \Sigma \) of the form (1.1). These MLE are the values of \( \sigma^2 \) and \( \rho \) for which \( \Sigma \) remains positive definite and for which the function \( f(\theta) = f(\sigma^2, \rho) \) is minimized. Note that \( f(\theta) \) is given by

\[
(2.7) \quad f(\sigma^2, \rho) = -(2/N)\log L = p \log(2\pi) + \log |\Sigma| + \text{tr} \Sigma^{-1} C,
\]

\[
(2.8) \quad \Sigma = \sigma^2 \psi, \quad \psi = I + \sum_{g=1}^{m} \rho G_g,
\]

where \( L \) is the likelihood function (evaluated at \( \mu = \bar{x} \)) and \( G_g \) are known symmetric matrices.

The first partial derivatives, needed for the gradient are given by:

\[
(2.9) \quad g_1 = \frac{\partial f}{\partial \sigma^2} = \frac{p}{\sigma^2} - \frac{1}{\sigma^4} \text{tr} \psi^{-1} C,
\]

\[
(2.10) \quad g_2 = \frac{\partial f}{\partial \rho} = \text{tr} \psi^{-1} A - (\text{tr} \psi^{-1} A \psi^{-1} C)/\sigma^2,
\]

\[
(2.11) \quad A = \frac{\partial \psi}{\partial \rho} = g_1 + \sum_{g=2}^{m} \rho g^{-1} G_g.
\]

The elements of the matrix of second partial derivatives are given by

\[
(2.12) \quad g_{11} = \frac{\partial^2 f}{\partial \sigma^2 \partial \sigma^2} = \frac{1}{\sigma^6} (2\text{tr} \psi^{-1} C - \rho \sigma^2),
\]

116
\( G_{12} = G_{21} = \frac{\partial^2 f}{\partial \sigma^2} = \frac{1}{4} \text{tr} \, \psi^{-1} A \psi^{-1} C , \)

\( G_{22} = \frac{\partial^2 f}{\partial \sigma^2} = \text{tr} \, \psi^{-1} B - \text{tr}(\psi^{-1} A)^2 \)

\[ + \left[ 2 \text{tr}(\psi^{-1} A)^2 \psi^{-1} C - \text{tr} \, \psi^{-1} C \psi^{-1} B \right]/\sigma^2 , \]

\( B = \frac{\partial \psi}{\partial \sigma^2} = 2C_2 + \sum_{g=3}^{m} g (g-1)/(g-2) C_g^2 . \)

Using the fact that \( \Sigma = \sigma^2 \psi , \) we find the elements of \( \Gamma \) are

\( \Gamma_{11} = \varepsilon G_{11} = \frac{p}{\sigma^4} , \)

\( \Gamma_{12} = \Gamma_{21} = \varepsilon G_{12} = \text{tr} \, \psi^{-1} A /\sigma^2 \)

\( \Gamma_{22} = \varepsilon G_{22} = \text{tr}(\psi^{-1} A)^2 . \)

Thus we have solved for the coefficients needed for the iterative equations (2.4) and (2.5). In the following section, we suggest a third algorithm. These three algorithms are compared in sections 5 and 6.

3. A Third Algorithm.

In this section, a third iterative scheme is given. First, Anderson's (1973) result for an iterative algorithm used to solve for MLE of \( \Sigma \) of the linear form,
(3.1) \[ \Sigma = \sum_{g=0}^{m} \sigma \Sigma^g \Sigma^g \]

is presented. Looking at the ML equations, Anderson suggests an iterative procedure. We then derive the ML equations for our special form of \( \Sigma \) and use the same method on these equations. While Anderson's (1973) method was pointed out by Rao (1973) to be the method of scoring, the method derived here for our problem appears to be different from the method of scoring. It also takes advantage of the special structure in our problem. Comparisons of the procedure derived in this section, and those derived in section 2, are discussed in sections 5 and 6.

Using the notation of section 2, Anderson (1973) derives the ML equations for the parameters \( \sigma \) in the patterned covariance matrix, \( \Sigma = \sum_{g=0}^{m} \sigma \Sigma^g \Sigma^g \), known symmetric matrices. These ML equations are

(3.2) \[ \text{tr} \Sigma^{-1} G = \text{tr} \Sigma^{-1} G \Sigma^{-1} C, \quad g=0,1,\ldots,m. \]

Since \( I = \Sigma^{-1} \Sigma \), we rewrite these equations as

(3.3) \[ \text{tr} \Sigma^{-1} G (\Sigma^{-1} \Sigma) = \text{tr} \Sigma^{-1} G \Sigma^{-1} C, \quad g=0,1,\ldots,m. \]

Rewriting (3.3) and substituting for \( \Sigma \), we find that

(3.4) \[ \sum_{h=0}^{m} \sigma_h (\text{tr} \Sigma^{-1} G \Sigma^{-1} G_h) = \text{tr} \Sigma^{-1} G \Sigma^{-1} C, \quad g=0,1,\ldots,m. \]
Anderson notes that the ML equations, written in this form (3.4) are suggestive of a linear algorithm in which we estimate \( \Sigma \) by the previous iterative values of \( \sigma_0, \ldots, \sigma_m \) and solve (3.4) for the values of \( \sigma_0, \ldots, \sigma_m \), which then become the new iterate.

For our problem, \( \Sigma = \sigma^2 \psi = \sigma^2 (I + \sum_{g=1}^{m} \rho g G_g) \), the ML equations are easily obtained from the results of section 2 by setting the gradient vector equal to zero. Thus from (2.9) and (2.10), we find the ML equations to be

\[
(3.5) \quad \sigma^2 = \frac{1}{p} \text{tr} \Sigma^{-1} C,
\]

\[
(3.6) \quad \text{tr} \psi^{-1} A = \text{tr}(\psi^{-1} A \psi^{-1} C)/\sigma^2,
\]

\[
(3.7) \quad A = G_1 + \sum_{g=2}^{m} \rho g \psi^{-1} G_g.
\]

Since \( I \neq \psi^{-1} \psi \), we rewrite (3.6) as

\[
(3.8) \quad \text{tr}(\psi^{-1} A \psi^{-1}) \psi = \text{tr}(\psi^{-1} A \psi^{-1} C)/\sigma^2.
\]

Substituting \( \psi = I + \sum_{g=1}^{m} \rho g G_g \) in (3.8) yields

\[
(3.9) \quad \text{tr} \psi^{-1} A \psi^{-1} (\sigma^2 I - C) + \sum_{g=1}^{m} (\sigma^2 \text{tr} \psi^{-1} A \psi^{-1} G_g) \rho g = 0.
\]

Since \( \sigma^2 \) is an explicit function of \( \rho \) (see 3.5), (equation 3.9) can be thought of as an equation only in \( \rho \). Thus, if we use
the previous value of $\rho$ to estimate $\psi$, $A$ and $\sigma^2$, (3.9) is just a polynomial in $\rho$, independent of $\sigma^2$. The coefficients of this polynomial depend on the previous value of $\rho$. Define $h(\rho)$ as the left side of (3.9),

$$h(\rho) = \text{tr} \psi^{-1} A \psi^{-1} (\sigma^2 I - C) + \sum_{g=1}^{m} (\sigma^2 \text{tr} \psi^{-1} A \psi^{-1} G_g) \rho^g.$$  

(3.10)

We define $h_s(\rho)$ as the polynomial $h(\rho)$ whose coefficients are evaluated at $\rho_s$. We wish to solve $h_s(\rho) = 0$, using a root of this polynomial for the next approximation to the MLE of $\rho$, denoting this next approximation by $\rho_{s+1}$.

At one extreme, we could solve $h_s(\rho) = 0$ for all roots and select the one which maximizes the likelihood function and keeps $\psi$ positive definite. Then we would use this new value, $\rho_{(s+1)}$, to create the new coefficients of $h(\rho)$, and thus give us a new polynomial $h_{s+1}(\rho)$. At the other extreme, we could use a linear or quadratic approximation of $h_s(\rho)$ around the point $\rho_s$ by approximating $h_s(\rho)$ by the first few terms of a Taylor expansion of $h_s(\rho)$ about $\rho_s$. In the special cases that I have considered on the computer, I have used a linear approximation. Thus $h_s(\rho)$ is approximated by

$$h_s(\rho) \approx h_s(\rho_s) + h'_s(\rho_s)(\rho - \rho_s).$$  

(3.11)

We want a value of $\rho$ such that $h_s(\rho) = 0$. Solving for $\rho$ in (3.11) with this condition, we find that
\[(3.12) \quad \rho_{s+1} = \rho = \rho_s - \frac{h_s(\rho_s)}{h'_s(\rho_s)},\]

\[(3.13) \quad h'_s(\rho) = \sigma^2 \text{tr} \psi^{-1} A\psi^{-1} G_1 + \sum_{g=2}^{m} (\sigma^2 \text{tr} \psi^{-1} A\psi^{-1} G_g) \rho^{g-1}.\]

Comparisons of the performance of this algorithm with those suggested in section 2, appear in section 6. First, a brief note on the simplification in these algorithms that appears when the $G_g$ are simultaneous diagonalizable is given in section 4.

4. Simplifications When the $G_g$ Are Simultaneously Diagonalizable.

Anderson (1969,1973) notes that when the $G_g$, $g = 0,1,\ldots,m$, commute, they are simultaneously diagonalizable. Using this fact, Anderson observes that there exists an orthogonal matrix $P$ and diagonal matrices $\Lambda_g$, $g = 0,1,\ldots,m$, so that the diagonal $\Lambda_g$ can be expressed by

\[(4.1) \quad \Lambda_g = P \rho_g P', \quad g=0,1,\ldots,m.\]

Thus $P \Psi P' = I + \sum_{g=1}^{m} \rho_g \Lambda_g = \Xi$, where $\Xi$ is a diagonal matrix. Letting $E = PCP'$, we can simplify the procedures in sections 2 and 3 somewhat by replacing $G_g$ by $\Lambda_g$ and $C$ by $E$, and noting that with these changes $\Psi$, $A$ and $B$ are diagonal. This greatly simplifies the calculation of $\psi^{-1}$. Also, we need only retain the diagonal elements of $E$.  

121
5. Discussion of Computation Procedures.

This section contains a discussion of various aspects of actually using computation procedures and convergence of such procedures. The reader hoping to find a proof of convergence for an algorithm will not find one. However, the reader with a problem in which convergence is lacking may find some helpful advice. While some parts of the section may apply only to the specific covariance structure (1.1) being considered in this chapter, most parts have general applicability. Using the covariance matrix (1.1), combined with an understanding of some specific cases in chapter II and III, we are able to test various algorithms and find difficulties with them. Thus no Monte Carlo study appears here. Instead, several pathological cases have been studied, and they indicate difficulties with the algorithms for which we here suggest corrective modification.

The initial starting point for an algorithm is often very important. An initial point that results in a covariance matrix not being positive definite occasionally occurs. As a simple example of such an occurrence, suppose one wants initial values for $\Sigma$ of the form:

\[
(5.1) \quad \Sigma = \begin{bmatrix}
    a & b & c & d \\
    b & a & d & d \\
    c & d & a & c \\
    d & d & c & a
\end{bmatrix}.
\]

Suppose the sample covariance matrix $C$, positive definite with probability one is given by
If a method of averaging is used, the following initial value of $\Sigma$, denoted $\Sigma_0$, is obtained

$$\Sigma_0 = \begin{bmatrix} .7 & .8 & 0 & 0 \\ .8 & .7 & 0 & 0 \\ 0 & 0 & .7 & .4 \\ 0 & 0 & .4 & .7 \end{bmatrix}.$$  

Clearly $\Sigma_0$ is not positive definite.

If one continues to iterate once a non-positive definite point is reached, one may step back into the positive definite region, converge to a non-positive definite solution, or not converge at all. Coefficient matrices may no longer be positive definite. Finally, the procedure for taking the inverse is simplified under the assumption that a matrix is positive definite symmetric. Such a procedure should be used in these problems to save on computation time. However, this procedure does not work when the positive definite symmetric assumption is violated.

Assuming that the structured covariance matrix under consideration has at least one set of values of the parameters for which it is positive definite, one can usually use the data in some way to help find an initial point. The main general advice one can give is to consider not using all the data in the sample covariance matrix to set the initial value unless a good understanding of the space of positive definite parameters is known.
Sometimes, one is able to recognize a patterned covariance matrix, with known space of positive definite parameters, as a special case of the matrix under consideration. For example, one of the parameterizations of the intraclass correlation matrix is given by

\[ \Sigma_I = \begin{bmatrix} a & b & b \\ b & a & b \\ b & b & a \end{bmatrix}. \]

(5.4)

Here \( \Sigma_I \) is a special case of \( \Sigma_1 \) and \( \Sigma_2 \) of the form

\[ \Sigma_1 = \begin{bmatrix} a & b & c \\ b & a & b \\ c & b & a \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} a_1 & b & c \\ b & a_2 & b \\ c & b & a_2 \end{bmatrix}. \]

(5.5)

We know that \( \Sigma_I \) is positive definite if \( a > 0 \) and \( -0.5 < b/a < 1 \).

Thus if we find initial estimates for the parameters of \( \Sigma_I \) (the MLE are explicitly known), then they may be used as initial estimates of \( \Sigma_1 \) and \( \Sigma_2 \). For \( \Sigma_1 \), \( \Sigma_I \) is the special case where \( b = c \). For \( \Sigma_2 \), \( \Sigma_I \) is the special case where \( a_1 = a_2 = a_3 \) and \( b = c \).

In the general form under consideration in this chapter, \( \Sigma = \sigma^2 \Psi = \sigma^2 (I + \sum_{g=1}^{m} \rho^g G_g) \), we know setting \( \rho = 0 \) results in the initial value of \( \sigma^2 \), \( \sigma^2 = (\text{tr} \ C)/p > 0 \). Other better values exist, but with no understanding of the positive definite region for \( \rho \), this value has the virtue of being positive definite, regardless of the form of the \( G_g \). We note that in chapter II, we had an example where \( |\rho| < 1/2 \). For a general \( p \times p \) intraclass correlation matrix, \( \rho \) must satisfy the inequality, \( -1/(p-1) < \rho < 1 \) for positive definiteness to be maintained.
If one decides to use an initial value without knowing if it is positive definite, it is easy enough to have the computer check for this property. Suppose one does start with a positive definite starting point. Must the remaining iterates be positive definite? I have started all three algorithms at positive definite starting points and had them iterate to a non-positive definite point. This tended to happen very quickly, thus indicating that the first few iterations are often the most critical. One can always select a new starting point when this happens. However, if careful consideration has already been given to the initial starting point, then a second good initial starting point may not be as easy to find. One may then ask, how did one iterate to a non-positive definite starting point and could one change the iterative procedure to prevent this? Basically the problem is not one of choosing the wrong direction for the Newton-Raphson or scoring method as long as the G and Γ matrices are still positive definite. [The Γ matrix being proportional to a covariance matrix, is always positive definite when calculated at parameters in the positive definite region. (See Kendall and Stuart, 1967, pp. 35-74).] The problem is that one has stepped too far in the correct direction. The solution is to introduce a variable step size. This topic is discussed in section 6.

Mention has been made of the possibility that the G matrix used by the Newton-Raphson method might not be positive definite. This may occur when one is near a local maximum when looking for a minimum point. If this occurs at the initial point, a new initial value must be chosen. If this phenomenon occurs after the initial value, one may use the procedure presented in section 6 for monitoring step size.
Another difficulty that occurs is one in which convergence is very slow due to the iterates occurring as a slowly damped oscillation around the solution. Miller (1973,1975a,b) mentions this possibility and suggests a form of averaging. The specifics of this method are left for section 6 when monitoring step size is discussed. I have experienced this difficulty in my studies quite frequently. Averaging in these cases resulted in much faster convergence. There is no guarantee, especially as the number of parameters increases, that this averaging method will always quickly dampen oscillations. However, if one finds that slow convergence is occurring, even with the method of averaging, a careful inspection of the iterates may allow one to pick a new starting value which speeds up convergence.

6. **A Method of Monitoring Step Size.**

A method of monitoring step size has been suggested by Crockett and Chernoff (1955) for the method of steepest ascent. A revised method for the method of scoring is used by Vandaele and Chowdhury (1971). Their method is presented here, a slight modification is made, and it is applied to all three algorithms suggested in sections 2 and 3.

We can rewrite the Newton-Raphson and scoring methods, respectively, as

\begin{align}
\theta^{(s+1)} &= \theta^{(s)} - w_s G^{(s)} \Gamma^{-1} g^{(s)}, \\
\theta^{(s+1)} &= \theta^{(s)} - w_s r^{(s)} \Gamma^{-1} g^{(s)},
\end{align}

(6.1) (6.2)
where these methods use \( w_s = 1 \). The method of section 3 can be modified by replacing (3.12) by

\[
\rho_{s+1} = \rho_s - w_s \frac{h_s(\rho_s)}{h_s'(\rho_s)}.
\]

In all three cases, we let \( w_s \) be a positive weight, possibly different for each iteration. Vandaele and Chowbury suggest monitoring the likelihood function. If the likelihood function increases from \( L(\theta^{(s)}) \) to \( L(\theta^{(s+1)}) \) for \( w_s = 1 \), they keep doubling \( w_s \), recalculating \( \theta^{(s+1)} \), until the likelihood function decreases. At that point, they halve \( w_s \), recalculate \( \theta^{(s+1)} \) and take this as their new iterate. If, however, they find \( L(\theta^{(s+1)}) \) evaluated at \( w_s = 1 \) is lower than \( L(\theta^{(s)}) \), they keep halving \( w_s \) and recalculating \( L(\theta^{(s+1)}) \), until \( L(\theta^{(s+1)}) > L(\theta^{(s)}) \). They use the value of \( \theta^{(s+1)} \) at this point as their new iterate. Since \( \Gamma \) is always positive definite, they claim that such monitoring guarantees convergence to a relative maximum.

This method as applied to the Newton-Raphson procedure may fail if \( G \) is not positive definite. Otherwise, in all three methods suggested in this chapter, we may monitor the likelihood function and monitor positive definiteness. If we step too far, we take a smaller step to maintain positive definiteness. In the case of the Newton-Raphson procedure, we can also monitor the positive definiteness of \( G \). The method of averaging suggested by Miller (1973, 1975a, b) uses \( w_s = 1/2 \) under certain conditions.

Rather than worry about bigger steps than those suggested by the iterative procedures, I suggest using \( w_s = 1 \), checking if \( \theta^{(s+1)} \)
is a positive definite point, in the Newton-Raphson method, checking if $G$ is positive definite, and checking $L(\theta^{(s+1)}) > L(\theta^{(s)})$. If any of these conditions does not hold, continue halving $w_s$ until they all hold. We assume these conditions hold at the initial point. If all these conditions hold at $w_s = 1$, check if the likelihood function increases at $w_s = 1/2$, i.e. if

$$(6.4) \quad L(\theta^{(s+1)}(w_s=1/2)) > L(\theta^{(s+1)}(w_s=1)).$$

If there is no increase, use $\theta^{(s+1)}$ calculated at $w_s = 1$. If there is an increase at $w_s = 1/2$, use $\theta^{(s+1)}$ calculated at $w_s = 1/2$. In a typical calculation, most of the iterates use $w_s = 1$. Occasionally, especially during early iterations when I step into non-positive definite values, I keep halving the step size until I am in the positive definite region and am at a higher value of the likelihood function. During the final iterations, I expect to use $w_s = 1/2$, a form of averaging, to reduce the number of iterations until convergence is reached.

Applying this method of monitoring step size I was able to get convergence with all three procedures derived in sections 2 and 3. In general the method of scoring and the method proposed in section 3 using a linear approximation to the root of the polynomial, performed about the same. Both these methods tended to do better than the Newton-Raphson method. These methods of monitoring are not limited to the specific covariance matrix (1.1) considered in this chapter. They should work well in other problems including Anderson's linear form (1.2).
While we may get convergence to a relative maximum, we do not necessarily have the maximum of the relative maxima within the positive definite region. Barnett (1966) studies the multiple root case. In general, one must use different starting points to obtain all maxima. There is an example considered in chapter II showing that in certain hypothesis testing problems the null hypothesis is very unlikely when multiple roots occur. Thus one would expect the null hypothesis would be rejected regardless of the root obtained. This indicates that there are cases in which it is not important to find all the roots.
V. "AVERAGING" AND MAXIMUM LIKELIHOOD ESTIMATION

1. Introduction.

In chapter IV, section 5, we gave an example in which averaging elements of a sample covariance matrix to obtain an initial estimate of the parameters in a patterned covariance matrix resulted in an initial estimate covariance matrix that is not positive definite. Thus we know that averaging does not always result in a positive definite matrix, much less the MLE. In section 3 of this chapter, we present sufficient conditions for the situation in which the explicit MLE for a patterned covariance matrix may be obtained by averaging and for the situation in which the method of scoring (Anderson, 1973) converges in one iteration to the explicit MLE from any positive definite starting point. We conclude in section 4 from these results that averaging is applicable to many patterned covariance matrices including the block and non-block forms of complete, compound and circular symmetry. We conclude the chapter with a brief discussion of unsolved problems in this area in section 5.

2. Preliminary Remarks and Definitions.

We start by defining the term "patterned covariance matrix" and what is meant by "averaging." We consider the problem of estimating a covariance matrix $\Sigma$ from the sample covariance matrix $C$ constructed from an independent, identically distributed multivariate normal sample.
We assume the mean \( \mu \) is unstructured and estimated by \( \bar{x} \). (See chapter IV, section 2).

**Definition 2.1.** \( \Sigma \) is a **patterned covariance matrix** (PCM) if

\[
\Sigma = \sum_{g=0}^{m} \sigma_g G_g,
\]

where \( G_0, \ldots, G_m \) are known, symmetric, linearly independent matrices and \( \sigma_0, \ldots, \sigma_m \) are unknown scalars with the property that there exists at least one set of values of \( \sigma_0, \ldots, \sigma_m \) such that \( \Sigma \) is positive definite.

**Definition 2.2.** The method of averaging applied to a PCM involves the solution of the linear equations

\[
\text{tr} \, \Sigma \, G_g = \text{tr} \, C \, G_g, \quad g = 0, 1, \ldots, m,
\]

or equivalently,

\[
\sum_{h=0}^{m} \sigma_h \left( \text{tr} \, G_g G_h \right) = \text{tr} \, C \, G_g, \quad g = 0, 1, \ldots, m.
\]

These averaging equations are the solution to a "least squares" problem of the form

\[
\min_{\sigma_0, \ldots, \sigma_m} \text{tr}(\Sigma - C)^2.
\]

The averaging equations result in estimates of the \( \sigma \)'s which may also be obtained by averaging over elements in the sample covariance matrix that are in the same positions as those in the PCM when the
positive elements of any $G_g$ matrix are the same. This last condition is equivalent to being able to rewrite the PCM into a form where the $G$ matrices consist only of zeroes and ones.

Examples where the averaging equations result in a solution different from one obtained by averaging are easy to find. We construct a PCM with at least one $G_g$ having ones and twos. For example, let $\Sigma$ be of the form

$$
(2.5) \quad \Sigma = \begin{bmatrix} b & 0 \\ 0 & 2b \end{bmatrix} = b \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}.
$$

The averaging equations result in the estimate of $b$ given by $(c_{11} + 2c_{22})/5$. Regular averaging would lead one to use $(c_{11} + c_{22})/3$. In such cases, however, neither of these solutions is the MLE given by $(2c_{11} + c_{22})/4$. Note the $c$'s are elements of the $2 \times 2$ sample covariance matrix $C$. In this particular case, it is interesting to note that the method of scoring gives us the MLE in one iteration from a positive definite starting point.

We now continue with several additional preliminary remarks in which we describe the MLE, method of scoring and clarifying what is meant by explicit solutions of the MLE.

Anderson (1973) gives the MLE

$$
(2.6) \quad \text{tr} \Sigma^{-1} G_g = \text{tr} \Sigma^{-1} G_g \Sigma^{-1} C, \quad g = 0, \ldots, m,
$$

and rewrites these equations in the form
\[(2.7) \quad \sum_{h=0}^{m} \sigma_h (\text{tr} \Sigma^{-1} G_h \Sigma^{-1} G_h) = \text{tr} \Sigma^{-1} G \Sigma^{-1} C), \quad g = 0, 1, \ldots, m, \]

which suggest a linear iterative method that is the same as the method of scoring. The parameters in parentheses are estimated by some initial positive definite version of the patterned covariance matrix \( \Sigma \) and the new values of the \( \sigma \)'s are obtained by solving the linear equations.

Before stating the main theorem of this chapter, we note that by saying that a PCM has explicit solutions, we mean that the maximum likelihood equations have one and only one solution. A PCM whose MLE have more than one solution even though only one solution results in a positive definite matrix does not have explicit solutions. (For examples, see chapter II).

3. Main Theorem.

Theorem 3.1. Let \( \Sigma \) be a PCM. If there exists an orthogonal matrix \( P \) with the following properties

\[(3.1) \quad \bar{\Sigma} = P \Sigma P' = \text{diag}(D_1 \otimes U_1, \ldots, D_r \otimes U_r), \]

where

i. \( D_1, \ldots, D_r \) are diagonal matrices with positive diagonal elements of dimensions \( n_1 \times n_1, \ldots, n_r \times n_r \), respectively,

ii. \( U_1, \ldots, U_r \) are symmetric,

iii. there exists a one-to-one transformation between \( \sigma_0, \ldots, \sigma_m \) and the parameters of \( U_1, \ldots, U_r \).
(Note: If $U_i$ is $p_i \times p_i$, then by the parameters of $U_1, \ldots, U_r$, we
mean $s = \sum_{i=1}^{r} p_i(p_i+1)/2$ independent parameters consisting of the
upper right hand triangles of the $U$'s),

then:

a. $\Sigma$ has explicit MLE;

b. the method of scoring converges in one iteration to the explicit
MLE from any positive definite starting point;

c. in the case where the $D$ matrices may be expressed as

\[ D_i = \alpha_i I_{n_i}, \quad i = 1, \ldots, r, \]

where the $\alpha$'s are known positive scalars,

the explicit MLE may be obtained by the method of averaging.

Before beginning the proof of this theorem, we note Miller (1975)
proved part b in the case where the $D$ matrices are identity matrices
and $U_1, \ldots, U_r$ are scalars, $r = m+1 = s$.

**Proof.** Let $\Lambda_g = P g P'$, $g = 0, \ldots, m$. We then write out $\hat{v}$ as

\[ \hat{v} = P \Sigma P' = \sum_{g=0}^{m} \sigma(g P g') = \sum_{g=0}^{m} \sigma \Lambda_g = \text{diag}(D_1 \otimes U_1, \ldots, D_r \otimes U_r). \]

From this last equality, we observe $\Lambda_g$ has the form

\[ \Lambda_g = \text{diag}(D_1 \otimes A_{g_1}, \ldots, D_r \otimes A_{g_r}), \quad g = 0, \ldots, m, \]

where $A_{g_i}$ is of the same dimension as $U_i$, $i = 1, \ldots, r$. Thus, we
we can relate the $\sigma$'s and $U$'s by
\[(3.4) \quad U_i = \sum_{g=0}^{m} \sigma^g A^g_i, \quad i = 1, \ldots, r.\]

Next, let \(E = PCP'\) be the sample covariance matrix for the transformed problem and let \(E_{11}^j, \ldots, E_{nj}^j\) be the blocks of matrices along the diagonal of \(E\) corresponding to the positions of the \(U_j, j = 1, \ldots, r\) in \(\Psi\). Finally, let \(D_i = \text{diag}(d_{i1}, \ldots, d_{in_i}), i = 1, \ldots, r.\)

With this new notation, we note the MLE, \(\text{tr} \Sigma^{-1}G_g = \text{tr} \Sigma^{-1}G_{\Psi} \Sigma^{-1}C_g, g = 0, \ldots, m\) may be rewritten as

\[(3.5) \quad \text{tr} \Psi^{-1}A_g = \text{tr} \Psi^{-1}A_g \Psi^{-1}E_g, \quad g = 0, \ldots, m.\]

This last expression can be further simplified by taking advantage of the structure of the \(A'\)s and \(\Psi\) to yield

\[(3.6) \quad \sum_{i=1}^{r} n_i U_i^{-1} A_g = \sum_{i=1}^{r} \text{tr} U_i^{-1} A_g U_i^{-1} F_i, \quad g = 0, \ldots, m,\]

\[(3.7) \quad F_i = \sum_{k=1}^{n_i} E_i^k / d_{ik}, \quad i = 1, \ldots, r.\]

We note that \(U_i = F_i / n_i, i = 1, \ldots, r\) is the solution of the MLE, thus proving part a.

To prove part b, we repeat the above substitutions into the scoring equations, (2.7), yielding
\[(3.8) \quad \sum_{i=1}^{r} n_i \sum_{h=0}^{m} \sigma_h \text{tr} U_i^{-1} A_{i} U_i^{-1} A_{i} = \sum_{i=1}^{r} \text{tr} U_i^{-1} A_{i} U_i^{-1} F_i, \quad g = 0, \ldots, m. \]

Note in this form, the \( U \)'s are obtained from the previous estimate of the \( \sigma \)'s (or the initial estimate at the first iteration). From \((3.4)\), we make the substitution \( U_i^* = \sum_{h=0}^{m} \sigma_h A_{hi} \), \( i = 1, \ldots, r \) into \((3.8)\) yielding

\[(3.9) \quad \sum_{i=1}^{r} n_i \text{tr} U_i^{-1} A_{i} U_i^{-1} U_i^* = \sum_{i=1}^{r} \text{tr} U_i^{-1} A_{i} U_i^{-1} F_i, \quad g = 0, \ldots, m. \]

We see that \( U_i^* = F_i / n_i \), \( i = 1, \ldots, r \), is a solution to these linear equations. To show that this solution is unique, we need only show the matrix \( \text{tr}(\Sigma^{-1} g \Sigma^{-1} g_h) \) is positive definite in the original scoring equations:

\[(3.10) \quad \sum_{h=0}^{m} \sigma_h \text{tr}(\Sigma^{-1} g \Sigma^{-1} g_h) = \text{tr} \Sigma^{-1} g \Sigma^{-1} g, \quad g = 0, \ldots, m, \]

under the condition that \( \Sigma \) is positive definite. Actually, in the proof, we see that if \( \Sigma^{-1} \) exists and is symmetric, then \( \text{tr} \Sigma^{-1} g \Sigma^{-1} g_h \) is positive definite. The proof follows from Anderson (1969).

Let \( Y = (Y_0, \ldots, Y_m) \) be any non-zero row vector. We must show
\( Y[\text{tr} \Sigma^{-1} g \Sigma^{-1} g_h] Y' > 0. \) We show this by rewriting this term.

\[(3.11) \quad Y[\text{tr} \Sigma^{-1} g \Sigma^{-1} g_h] Y' = \sum_{g, h=0}^{m} Y_g \text{tr}(\Sigma^{-1} g \Sigma^{-1} g_h) Y_h
= \text{tr} \Sigma^{-1} \left( \sum_{g=0}^{m} Y_g g \right) \Sigma^{-1} \left( \sum_{h=0}^{m} Y_h g_h \right). \]
Since the $G'$s are from a PCM, they are linearly independent. Thus

\[ \sum_{g=0}^{m} Y g g' \neq 0. \]

Since $\Sigma^{-1}$ is symmetric, we can rewrite the above expression as $\text{tr} AA'$ where $A = \Sigma^{-1}(\sum_{g=0}^{m} Y g g')$. Since $A \neq 0$, $\text{tr} AA' > 0$. This completes the proof of part b.

The proof to part c follows directly from part b. The averaging equations are given by

\[ (3.12) \quad \sum_{h=0}^{m} \sigma_h \text{tr} G_h G_h = \text{tr} G C, \quad g = 0, \ldots, m. \]

The scoring equations are given by

\[ (3.13) \quad \sum_{h=0}^{m} \sigma_h \text{tr} \Sigma^{-1} G_h \Sigma^{-1} G_h = \text{tr} \Sigma^{-1} G \Sigma^{-1} G \quad g = 0, \ldots, m \]

where if $\Sigma^{-1}$ is any positive definite matrix, we know that we get the MLE as a solution to the linear equations. Noting that $\Sigma = I$ is one such matrix, and substituting this value into the scoring equations (3.13) yields the averaging equations, (3.12).

The necessity of restricting the D matrices to be known multiples of an identity matrix that is used in part c may be seen by rewriting the averaging equations using the techniques of the proof of parts a and b of theorem 3.1 to yield the equivalent equations

\[ (3.14) \quad \sum_{i=1}^{r} \beta_i \text{tr} U_i A g_i = \sum_{i=1}^{r} \text{tr} F_i A g_i, \quad g = 0, \ldots, m, \]
\[ (3.15) \quad \tilde{F}_i = \frac{1}{\beta_i} \sum_{k=1}^{n_i} d_{ik} E_{kk}^i, \quad \beta_i = \sum_{k=1}^{n_i} d_{ik}^2, \quad i = 1, \ldots, r. \]

The unique solution of these averaging equations is given by \( U_i = \tilde{F}_i / \tilde{\beta}_i \), \( i = 1, \ldots, r \). However, the MLE solution is given by \( U_i = F_i / n_i \), where \( F_i = \sum_{k=1}^{n_i} E_{kk}^i / d_{ik} \), \( i = 1, \ldots, r \). Unless the \( D \)'s are known multiples of identity matrices,

\[ (3.16) \quad \tilde{\tilde{F}_i} = \sum_{k=1}^{n_i} \left( -\frac{1}{\tilde{\beta}_i} \right) E_{kk}^i \neq \frac{1}{n_i} \sum_{k=1}^{n_i} \left( \frac{1}{d_{ik}} \right) E_{kk}^i = \frac{F_i}{n_i}, \quad i = 1, \ldots, r. \]

4. Applications.

As an application of this theorem, we note that in the problem of estimating PCM of the block and non-block complete, compound and circular symmetry where the mean is assumed unstructured, we may average the sample covariance elements to get the MLE for the covariance matrix elements. To see this, we note that in the case of complete symmetry, \( \Sigma_1 = (a-b)I_p + bJ_{p,p} \) and \( \Sigma_2 = I_p \otimes (A-B) + J_{p,p} \otimes B \) are the non-block (a and b scalars) and block (A and B symmetric matrices) forms respectively of the PCM. By techniques used in chapter I these PCM can be put in the form

\[ (4.1) \quad \tilde{\tilde{Y}_1} = \text{diag}(a+(p-1)b, I_{p-1} \otimes (a-b)) = (u_1, I_{p-1} \otimes u_2), \]

\[ (4.2) \quad \tilde{\tilde{Y}_2} = \text{diag}(A+(p-1)B, I_{p-1} \otimes (A-B)) = (U_1, I_{p-1} \otimes U_2). \]
The cases of block and non-block compound symmetry follow similarly. (See chapter I, theorems 3.1 and 3.2). That the non-block and block forms of circular symmetry are of this form have been shown by Olkin and Press (1969) and Olkin (1972) respectively.

5. Discussion of Open Questions.

We conclude this chapter by discussing several unsolved questions. It would be useful to know when a PCM has explicit solutions, when the method of scoring yields these solutions in one iteration, and when the averaging method also yields the MLE. Theorem 5.1 gives sufficient conditions. Are these conditions necessary? In addition, how does one go about determining if a PCM does or does not have explicit MLE? Can one recommend a method of analysis that is easy to use and guarantees an answer to this last question? A discussion of partial answers to these questions follows.

First, we consider the problem of diagonalizing the PCM into the block diagonal form of theorem 5.1, \( \mathbf{I} = \text{diag}(D_1 \otimes U_1, \ldots, D_r \otimes U_r) \) omitting for the moment the consideration of a one-to-one transformation between the \( \sigma \)'s and \( U \)'s. We first note a well-known result on necessary and sufficient conditions for the existence of an orthogonal matrix \( P \) independent of the values of \( \sigma_0, \ldots, \sigma_m \) such that \( \mathbf{I} = P \Sigma P' \) is diagonal.

**Lemma 5.1.** Let \( \Sigma \) be a PCM. The following statements are equivalent:

a. There exists a matrix \( P \) orthogonal, independent of the \( \sigma \)'s such that \( \mathbf{I} = P \Sigma P' \) is diagonal.
b. the $G$ matrices commute.

c. $G_{ij}$ is symmetric, $i,j = 0, \ldots, m$.

Proof. See, for example, Rogers (1975). We note that this lemma does not mean that if the $G$'s do not commute, that there does not exist an orthogonal matrix $P$ such that $PEP = \mathbf{I}$ is of the block diagonal form of theorem 5.1. (See compound symmetry, for such a counterexample.) If the $G$'s do commute, this lemma allows us to conclude that $\Sigma$ may be put in this diagonal form.

To extend the power of this lemma, we note that every PCM can be put in the form $\Sigma = \sum_{g=0}^{n} G_{g} \bigotimes \Sigma_{gg}$ where the $\Sigma_{gg}$ are block symmetric matrices of unknown coefficients. Thus if a PCM can be rewritten in this form, one can again apply lemma 5.1. This is a useful notational form for PCM which have patterns of blocks. Note that this new form is equivalent to the definitional form of a PCM in the sense that it is an alternative form but does not extend the definition of a PCM.

Applying lemma 5.1 to the PCM forms $\Sigma = \sum_{g=0}^{m} \sigma G_{g}$ or $\Sigma = \sum_{g=0}^{n} G_{g} \bigotimes \Sigma_{gg}$ in the case where the $G$ matrices commute allows one to find an orthogonal matrix $P$ such that $\mathbf{I} = PEP$ is of the block diagonal form of theorem 5.1. The MLE have explicit solutions in terms of the $U$'s, but only when the $\sigma$'s and $U$'s are one-to-one will these solutions in terms of $U$'s translate to explicit solutions in terms of the $\sigma$'s. This gives us a constructive procedure for checking if PCM's with commuting $G$ matrices have explicit solutions.

Lacking a definitive method of analysis to apply to the problem of easily establishing whether a PCM has explicit solutions, I conclude this
chapter with a brief description of techniques I would use to decide whether a PCM has explicit solutions in a specific case. First, I would see if the $G_\ell$ matrices commute. If not, then I would see if the PCM could be written in the form $\Sigma = \sum_{g=0}^{n} G_{g} \otimes \Sigma_{gg}$ and again check if the $G_\ell$'s commute. Assuming the $G_\ell$'s do not commute in either form, I would then proceed to attempt to put the PCM into the block diagonal form of theorem 5.1 by using a number of orthogonal transformations on the PCM, using orthogonal matrices of the form used in the compound symmetry problem (see chapter I) and/or orthogonal matrices used to diagonalize the circulant, (see Olkin and Press (1969)). Putting any part of the PCM into a block diagonal form would help simplify any further analysis.
APPENDIX I

SAMPLE CALCULATIONS FOR $\bar{\Pi}_1(mvc)$

In this appendix, sample calculations are done for $\bar{\Pi}_1(mvc)$. This appendix is divided into five sections. In the first section, MLE are found using Lemmas 3.1, 3.2, and 3.3. These MLE are used to find the LRT. In the second section, the moments of the LRT are found under the null hypothesis. The third section uses these moments and the Box method to find an approximate null distribution of the LRT. In the fourth and fifth sections the asymptotic non-null distribution of the LRT is found.

We start with $y$ and $V$ independently distributed, $\mathcal{L}(y) = N(\bar{v}, \bar{\Xi}), \mathcal{L}(V) = W(\Xi, p, N-1)$. Under BCS-II, $\bar{v}$ and $\bar{\Xi}$ have the special structure (see Theorem 3.2):

$$\bar{v} = (\bar{v}_1, 0), \quad \bar{v}_1: 1 \times rh,$$

$$\bar{\Xi} = \text{diag}(\Xi_1, I_{n-1} \otimes \Xi_2).$$

We derive some of the $\bar{\Pi}_1(mvc)$ results appearing in Section 6.

I.1. MLE Calculations Leading to LRT.

In order to derive the LRT $\lambda_{14}$ we need to evaluate $\sup_{\omega_4} L(y, V)$ for $i=1,4$ where $L(y, V)$ is the likelihood function for $y$ and $V$, $\omega_1$ is the restriction of $\nu$ and $\Xi$ to the special structure under BOS-II, and $\omega_4$ is the unrestricted case of $\nu$ and $\Xi$. (See (6.6) and (6.9)). It is sufficient to find the MLE under the restrictions $\omega_1$ and $\omega_4$.

142
Using Lemma 3.2 for the case \( \omega_1 \), we find the MLE are \( \hat{\nu} = y \) and \( \hat{\Xi} = V/N \).

The \( \omega_1 \) case is not quite as straightforward. Under the \( \omega_1 \) restrictions, \( y_1, \ldots, y_n, V_{11}, \ldots, V_{nn} \) are independently distributed as:

\[
(I.1.1) \quad \mathcal{L}(y_1) = \mathcal{N}(\nu_1, \Xi_1), \quad \mathcal{L}(V_{11}) = \mathcal{W}(\Xi_1, RH, N-1),
\]

\[
(I.1.2) \quad \mathcal{L}(y_j) = \mathcal{N}(0, \Xi_2), \quad \mathcal{L}(V_{jj}) = \mathcal{W}(\Xi_2, RH, N-1), \quad j = 2, \ldots, n.
\]

The problem of finding MLE under the \( \omega_1 \) restriction separates into two independent problems. Problem I, (I.1.1), is easily solved by Lemma 3.2 which given the MLE:

\[
(I.1.3) \quad \hat{\nu}_1 = y_1, \quad \hat{\Xi}_1 = V_{11}/N.
\]

For the solution of Problem II, (I.1.2), we note that \( \mathcal{L}(\sum_{j=2}^{n} y_j y_j) = \mathcal{W}(\Xi_2, RH, n-1) \) and \( \mathcal{L}(\sum_{j=2}^{n} y_j y_j + \sum_{j=2}^{n} V_{jj}) = \mathcal{W}(\Xi_2, RH, N(N-1)) \). Using Lemma 3.3 we find that

\[
(I.1.4) \quad \hat{\Xi}_2 = \sum_{j=2}^{n} (y_j y_j + V_{jj})/N(n-1).
\]

The MLE of \( \nu \) and \( \Xi \) are thus given by

\[
(I.1.5) \quad \hat{\nu} = (\hat{\nu}_1, 0), \quad \hat{\Xi} = (\hat{\Xi}_1, I_{n-1} \otimes \hat{\Xi}_2).
\]
under the restriction \( \omega_1 \). Substitution of the MLE into \( L(y,V) \) gives the results for \( \sup_{\omega_1} L(y,V) \), (6.10) and (6.13) after simplification.

1.2. Moments of the LRT Under the Null Distribution.

The LRT \( \lambda_{14} \) is:

\[
\lambda_{14}^{2/N} = \frac{(n-1)^{\frac{p}{2} (n-1)}}{|V_{11}|^H + \sum_{j=2}^{n} V_{jj}^{n-1}}.
\]

(I.2.1)

It is sufficient to find the moments of

\[
\phi = \frac{|V|}{|V_{11}|^H + \sum_{j=2}^{n} V_{jj}^{n-1}},
\]

since \( \phi \lambda_{14}^{S} = (n-1)^{\frac{p}{2} (n-1)NS/2} \phi^{NS/2} \).

First, we recall that under the null hypotheses, \( v = (v_1, 0) \) and \( \Xi = \text{diag}(\Xi_1, I_{n-1} \otimes \Xi_2) \). Then we have

(I.2.3) \( \mathcal{L}(V) = W(\Xi, p, N-1) \), \( \mathcal{L}(V_{11}) = W(\Xi_1, rh, N-1) \),

(I.2.4) \( \mathcal{L}(V_{jj}) = W(\Xi_2, rh, N-1) \), \( \mathcal{L}(H) = W(\Xi_2, rh, n-1), j=2, \ldots, n \),

(see previous section), where \( V \) and \( H \) are independent. We use the following form of the Wishart density. If \( \mathcal{L}(A) = W(\Sigma, p, n) \),

144
\[ \text{(I.2.5)} \quad p(A) = K(\Sigma, p, n) \left| A \right|^{(n-p-1)/2} \exp\left(-\frac{1}{2} \text{tr} \Sigma^{-1} A \right), \quad A > 0, \Sigma > 0, \]

\[ \text{(I.2.6)} \quad [K(\Sigma, p, n)]^{-1} = 2^{np/2} p^{(p-1)/4} |\Sigma|^{n/2} \prod_{i=1}^{p} \Gamma[(n+1-i)/2]. \]

In the first step we combine \(|V|^t\) with the density \(p(V)\) as follows.

\[ \mathcal{E}^t = K(\Xi, p, N-1) \int_{V>0, H>0} \frac{|V|^t |V|^{(N-1-p-1)/2}}{|V_{11}|^t |H + \sum_{j=2}^{n} V_{jj}|^{(t-n+1)/2}} \exp\left(-\frac{1}{2} \Xi^{-1} V \right) dV dH \]

\[ \text{(I.2.7)} \]

\[ = K(\Xi, p, N-1+2t) \int_{V>0, H>0} \frac{|V|^{(N-1+2t-p-1)/2}}{|V_{11}|^t |H + \sum_{j=2}^{n} V_{jj}|^{(t-n+1)/2}} \exp\left(-\frac{1}{2} \Xi^{-1} V \right) dV dH. \]

We may rewrite (I.2.7) as

\[ \text{(I.2.8)} \quad \mathcal{E}^t = \frac{K(\Xi, p, N-1)}{K(\Xi, p, N-1+2t)} \mathcal{E}\left( |V_{11}|^{-t} |H + \sum_{j=2}^{n} V_{jj}|^{-t(n-1)} \right), \]

where the expectation is taken with respect to \(\mathcal{L}(V) = W(\Xi, p, N-1+2t)\)

and \(\mathcal{L}(H) = W(\Xi_2, rh, n-l)\). Because \(\Xi = \text{diag}(\Xi_1, I_{n-1} \otimes \Xi_2)\), we note

that \(V_{11}, V_{22}, \ldots, V_{nn}, H\) are independent. Thus

\[ \text{(I.2.9)} \quad \mathcal{E}(|V_{11}|^{-t} |H + \sum_{j=2}^{n} V_{jj}|^{-t(n-1)}) = (\mathcal{E} |V_{11}|^{-t})(\mathcal{E} |H + \sum_{j=2}^{n} V_{jj}|^{-t(n-1)}). \]
We note that we have proved, in the first calculation, the lemma:

**Lemma A.1.** If $\mathcal{L}(A) = W(\Sigma, p, n)$, then

(I.2.10) \[ \mathcal{E}^{|A|^t} = K(\Sigma, p, n)/K(\Sigma, p, n+2t). \]

Observing that $\mathcal{L}(V_{11}) = W(\Xi_1, rh, N-1+2t)$ and applying Lemma A.1, we have

(I.2.11) \[ \mathcal{E}^{|V_{11}|^{-t}} = K(\Xi_1, rh, N-1+2t)/K(\Xi_1, rh, N-1). \]

To evaluate the other expectation in (I.2.9), we have $\mathcal{L}(H) = W(\Xi_2, rh, n-1)$ and $\mathcal{L}(V_{jj}) = W(\Xi_2, rh, N-1+2t)$. Thus setting $F = H + \sum_{j=2}^{n} V_{jj}$ yields:

(I.2.12) \[ \mathcal{L}(F) = \mathcal{L}(H + \sum_{j=2}^{n} V_{jj}) = W(\Xi_2, rh, N(n-1)+2t(n-1)). \]

Applying Lemma A.1 to $\mathcal{E}^{|F|^{-t(n-1)}}$, we obtain

(I.2.13) \[ \mathcal{E}^{|F|^{-t(n-1)}} = K(\Xi_2, rh, N(n-1)+2t(n-1))/K(\Xi_2, rh, N(n-1)). \]

Combining (I.2.9), (I.2.11) and (I.2.13), we have

(I.2.14) \[ \mathcal{E}^{|t} = \frac{K(\Xi_2, rh, N-1+2t)K(\Xi_1, rh, (n-1)(N+2t))}{K(\Xi_2, rh, N(n-1)+2t(n-1))} \]

\[ = \frac{K(\Xi_1, rh, N-1+2t)K(\Xi_2, rh, (n-1)(N+2t))}{K(\Xi, p, N-1+2t)}. \]
where "a" is chosen so $\phi^\circ = 1$. Using the relationship between the moments of $\lambda_{14}$ and $\phi$, and the definition of $K(\Sigma, \rho, h)$, (I.2.6), we arrive at the result listed in Section 6, (6.18).

I.3. Approximate Null Distribution Calculations.

In this section we comment on the evaluation of $r$, $\rho$, and $\omega$ in the Box expansion, (see 3.10 and 6.19),

(I.3.1) \[ P\{-2 \log \lambda_{14} \leq z\} = (1-\omega) P(x_1^2 \leq \rho z) + \omega P(x_{14}^2 \leq \rho z) + O(N^{-3}) \]

Using Lemma 3.4, the moments of the LRT under the null hypothesis are:

(I.3.2) \[ \phi_{\lambda_{14}}^S = K(n-1)^{rS(n-1)N/2} \frac{(n-1)rh}{\prod_{k=1}^{(n-1)rh} r\left(\frac{N}{2}(1+S) - \frac{rh+k}{2}\right)} \]

\[ \prod_{j=1}^{rh} r\left(\frac{N(n-1)}{2}(1+S) + \frac{1-l_j}{2}\right) \]

with $K$ chosen so $\phi_{\lambda_{14}}^0 = 1$. We next identify the constants in the general formula below with the corresponding constants in I.3.2.

(I.3.3) \[ \phi_{W_S} = K \left( \begin{array}{c} b \\ \sum_{j=1}^{y_j} \end{array} \right)^S \frac{a}{\prod_{k=1}^{x_k} \Gamma(x_k(1+S)+\xi_k)} \left( \begin{array}{c} a \\ \sum_{k=1}^{x_k} \end{array} \right)^S \frac{b}{\prod_{j=1}^{y_j} \Gamma(y_j(1+S)+\eta_j)} \]

We find $a = (n-1)rh$, $x_k = \frac{N}{2}$, $\xi_k = -\frac{rh+k}{2}$, $b = rh$, $y_j = \frac{N(n-1)}{2}$, and $\eta_j = \frac{1-l_j}{2}$.
In calculating $f$, $\rho$, and $\omega$, we use the following three well-known formulas:

\begin{align*}
(1.3.4) \quad \sum_{j=1}^{n} j &= \frac{n(n+1)}{2}, \quad \sum_{j=1}^{n} j^2 = \frac{n(n+1)(2n+1)}{6}, \quad \sum_{j=1}^{n} j^3 = \frac{n^2(n+1)^2}{4}.
\end{align*}

Substitution into the formulas of Lemma 3.4 and lengthy calculations yield the values of $f$, $\rho$, and $\omega$.


In this section we evaluate the non-null asymptotic distribution of

\begin{align*}
(1.4.1) \quad \frac{2}{N} \log \lambda_{14} &= \log \left( \frac{(n-1)^{\frac{n(n-1)}{2}} |V|}{|V_{11}| |H + \sum_{j=2}^{n} V_{j} | n^{-1}} \right).
\end{align*}

We note that $H = \sum_{j=2}^{n} y_{j}^{'y_{j}}$ and $V$ are independently distributed,

\begin{align*}
\mathcal{L}(y) &= \mathcal{N}(\sqrt{N} \tilde{y}, \Xi), \quad \mathcal{L}(V) = \mathcal{W}(\Xi, p, N-1).
\end{align*}

Under the non-null distribution $V$ and $\Xi$ are unrestricted. First, we make the transformations

\begin{align*}
(1.4.2) \quad \sqrt{N} x = (y - V) \Xi^{-1/2}, \quad NZ = \Xi^{-1/2} V \Xi^{-1/2}.
\end{align*}

$Z$ and $x$ are independent with distributions:

\begin{align*}
(1.4.3) \quad \mathcal{L}(\sqrt{N} x) &= \mathcal{N}(0, I), \quad \mathcal{L}(NZ) = \mathcal{W}(I, p, N-1).
\end{align*}
We make the substitution \( y = \sqrt{N} (x^{-1/2} + \tilde{v}) \) and \( V = N x^{-1/2} z^{-1/2} \) in (I.4.1), yielding:

\[
\frac{2}{N} \log \lambda_{14} = rh(n-1)\log(n-1) + \log|\Xi| + \log|Z| = -n \log \left( \frac{1}{|\Xi|^n} \right) - \left| \log \left( z^{-1/2} \right)^{-n} \right| - (n-1) \log \left[ \sum_{j=2}^{n} (x^{-1/2} + \tilde{v})_j \left( x^{-1/2} + \tilde{v} \right)_j + (z^{-1/2} z^{-1/2})_{jj} \right].
\]

Into this expression we substitute

\[
(\Xi z^{-1/2})_{jj} = \sum_{r,s=1}^{n} (\Xi z^{-1/2})_{jr} z_{rs} (\Xi z^{-1/2})_{sj},
\]

\[
(x z^{-1/2})_j = \sum_{r=1}^{n} x_r (x z^{-1/2})_{rj},
\]

resulting in

\[
\frac{2}{N} \log \lambda_{14} = C + \log|Z| - \log|T| - (n-1) \log|R|,
\]

\[
C = \text{constants, independent of } x \text{ and } Z,
\]

\[
T = \sum_{r,s=1}^{n} (\Xi z^{-1/2})_{jr} z_{rs} (\Xi z^{-1/2})_{sl},
\]

\[
R = S + \sum_{j=2}^{n} \sum_{r,s=1}^{n} (\Xi z^{-1/2})_{jr} z_{rs} (\Xi z^{-1/2})_{sj},
\]

\[
S = \sum_{j=2}^{n} \left[ \sum_{r,s=1}^{n} (\Xi z^{-1/2})_{jr} x_r x_s (\Xi z^{-1/2})_{sj} + \sum_{r=1}^{n} (\Xi z^{-1/2})_{jr} x_r \tilde{v}_j \right.
\]

\[
+ \sum_{s=1}^{n} \tilde{v}_j x_s (\Xi z^{-1/2})_{sj} + \tilde{v}_j \tilde{v}_j \right].
\]
Note

\[
T\bigg|_{x=0, z=I} = \sum_{r=1}^{n} (\Xi^{1/2})_{rr} I_{rr} (\Xi^{1/2})_{rl} = \Xi_{ll},
\]

\[
A = R\bigg|_{x=0, z=I} = \sum_{j=2}^{n} (\tilde{\gamma}_j \tilde{\gamma}_j + \Xi_{jj}).
\]

Evaluating the derivatives of \( \frac{2}{N} \log \lambda_{14} \) with respect to \( x \) and \( z \) using Lemma 3.5, we have

\[
\frac{\partial}{\partial z} \log |Z|\bigg|_{x=0, z=I} = Z^{-1}\bigg|_{x=0, z=I} = I,
\]

\[
E_{tu} = \frac{\partial}{\partial z_{tu}} \log |T|\bigg|_{x=0, z=I} = (\Xi^{1/2})_{tt} (\Xi_{ll})^{-1} (\Xi^{1/2})_{u1} = (\Xi^{1/2})_{tt} (\Xi_{ll})^{-1} (\Xi^{1/2})_{1u},
\]

\[
F_{tu} = \frac{\partial}{\partial z_{tu}} \log |R|\bigg|_{x=0, z=I} = \sum_{j=2}^{n} (\Xi^{1/2})_{tj} A^{-1} (\Xi^{1/2})_{ju},
\]

\[
\frac{d}{(n-1)} \frac{\partial}{\partial x_t} \log |R|\bigg|_{x=0, z=I} = 2 \sum_{j=2}^{n} \tilde{\gamma}_j A^{-1} (\Xi^{1/2})_{jt},
\]

using the fact that \( A \) is symmetric.
Let \( H_{tu} = \frac{\partial}{\partial Z_{tu}} \frac{\partial}{\partial N} \log \lambda_{14} \bigg|_{x=0} = I_{tu} - E_{tu} - (n-1)F_{tu} \). Note we have ignored the fact that \( Z \) is symmetric while taking derivatives, thus we have omitted the factor \( \frac{1}{2} \) in the definition of \( H = (h_{ij}) = \frac{1}{2} \frac{\partial h}{\partial w_{ij}} \bigg|_{w=I} \) for \( i \neq j \) that appears at the end of Section 3.

To evaluate the asymptotic variance, we use (Madansky and Olkin, 1969)

\[
(I.4.12) \quad v_{\omega} = 2 \text{tr} H^{2} + dd' = 2\text{tr} \sum_{t,u=1}^{n} H_{tu}H_{ut} + \sum_{t=1}^{n} d_{t}d'_{t} .
\]

First examine the terms in \( \text{tr} \sum_{t,u=1}^{n} H_{tu}H_{ut} \):

\[
(I.4.13) \quad \text{tr} \sum_{t,u=1}^{n} H_{tu}H_{ut} = \text{tr} \left( \sum_{t=1}^{n} I_{tt} - 2 \sum_{t=1}^{n} E_{tt} - 2(n-1) \sum_{t=1}^{n} F_{tt} \right.

+ \left. \sum_{t,u=1}^{n} E_{tu}E_{ut} + 2(n-1) \sum_{t,u=1}^{n} E_{tu}F_{ut} + (n-1)^{2} \sum_{t,u=1}^{n} F_{tu}F_{ut} \right) .
\]

Evaluating each term in this expression, we get

\[
(I.4.14) \quad \sum_{j=1}^{n} \text{tr} I_{tt} = nr \rho ,
\]

\[
(I.4.15) \quad \sum_{t=1}^{n} \text{tr} E_{tt} = \sum_{t=1}^{n} \text{tr}(\xi_{11}^{1/2})_{t1}(\xi_{11}^{-1})_{1t}^{1/2} = \text{tr}(\xi_{11}^{-1})_{11} = \text{tr} I_{11} = r \rho .
\]
(I.4.16) \[ \sum_{n=1}^{N} \text{tr} F_{tt} = \sum_{j=2}^{N} \sum_{t=1}^{n} \text{tr}(\Xi^{1/2})_{tj} A^{-1}(-\Xi^{1/2})_{jt} = \sum_{j=2}^{n} \text{tr} A^{-1}(-\Xi)_{jj}, \]

(I.4.17) \[ \sum_{t, u=1}^{n} \text{tr} E_{tu} E_{ut} = \]
\[ = \sum_{t, u=1}^{n} \text{tr}(\Xi^{1/2})_{t1} (\Xi^{1/2})_{1u}^{-1}(\Xi^{1/2})_{u1} (\Xi^{1/2})_{1t}^{-1}(\Xi^{1/2})_{tt} \]
\[ = \text{tr}(\Xi)_{11}^{-1} \left( \sum_{u=1}^{n} (\Xi^{1/2})_{u1}^{-1}(\Xi^{1/2})_{1u} \right) \left( \sum_{t=1}^{n} (\Xi^{1/2})_{1t}^{-1}(\Xi^{1/2})_{tt} \right) \]
\[ = \text{tr} I_{11} = r \]

(I.4.18) \[ \sum_{t, u=1}^{n} \text{tr} E_{tu} F_{ut} = \]
\[ = \sum_{j=2}^{n} \sum_{t, u=1}^{n} \text{tr}(\Xi^{1/2})_{t1} (\Xi^{1/2})_{1u}^{-1}(\Xi^{1/2})_{u1} (\Xi^{1/2})_{1j} A^{-1}(\Xi^{1/2})_{jt} \]
\[ = \sum_{j=2}^{n} \text{tr}(\Xi)_{11}^{-1}(\Xi)_{1j} A^{-1}(\Xi)_{jj} \]

(I.4.19) \[ \sum_{t, u=1}^{n} \text{tr} F_{tu} F_{ut} = \sum_{i, j=2}^{n} \sum_{t, u=1}^{n} \text{tr}(\Xi^{1/2})_{ti} A^{-1}(\Xi^{1/2})_{iu} (\Xi^{1/2})_{uj} A^{-1}(\Xi^{1/2})_{jt} \]
\[ = \sum_{i, j=2}^{n} \text{tr} A^{-1}(\Xi)_{ij} A^{-1}(\Xi)_{ji}, \]

(I.4.20) \[ \sum_{t=1}^{n} d_{t} d_{t} = 4(n-1)^{2} \sum_{i, j=2}^{n} \tilde{v} A^{-1} \left( \sum_{t=1}^{n} (\Xi^{1/2})_{tt} A^{-1} \tilde{v} \right) \]
\[ = 4(n-1)^{2} \sum_{i, j=2}^{n} \tilde{v} A^{-1}(\Xi)_{ij} A^{-1} \tilde{v}. \]

Substitution of these calculations in (I.4.13) and (I.4.12) yields the desired result for \( v_{\infty}. \)

The k-population case calculations follow the same procedures as those of the 1-population case. However, in the k-population case, in computing the asymptotic variance, we must contend with a set of weights depending on the sample sizes, \( p_d \), \( d=1, \ldots, k \) for each population and \( N = p_1 + \cdots + p_k \), the total sample size. In this section, we briefly illustrate the calculations involved in determining the asymptotic mean and variance in the case \( H_k(MVC|mvc) \). (See section 5 of chapter I).

In the case of \( H_k(MVC|mvc) \), the LRT is a function of \( y^1, y^d, \ldots, y^k \), where

\[
L(y^d) = n(v^d, z^d), \quad L(V^d) = W(z^d, p, p_k^d - 1), \quad d=1, \ldots, k.
\]

Here the \( y \)'s and \( V \)'s have the special structure under the alternative hypothesis given by

\[
y^d = (v^d_0, 0) = (\sqrt{p_d}, v^d_0, 0)
\]

\[
z^d = diag(z^d_0, I_{n_1-1} \otimes z^d_1, \ldots, I_{n_q-1} \otimes z^d_q).
\]

\( \lambda \) may be observed to be a function of \( y^1_0, \ldots, y^k_0, v^1_0, \ldots, v^k_0 \) and \( F^d_i, d=1, \ldots, k; \ i=1, \ldots, q \) where \( F^d_i \) is defined by

\[
F^d_i = \sum_{j=1}^{n_i-1} [(v^d_{i1})_j + (y^d_i)_j(y^d_i)_j], \quad i=1, \ldots, q.
\]

The distributions of these variables under the alternative hypothesis are given by
\( \mathcal{L}(y^d_0) = n(\sqrt{p_d} v^d_0, \varepsilon^d_0) \), \( \mathcal{L}(v^d_0) = \mathcal{N}(\varepsilon^d_0, r_p, p_d - 1) \),

\( \mathcal{L}(F^d_i) = \mathcal{W}(\varepsilon^d_i, t_i, p_d(n_i - 1)) \), \( i = 1, \ldots, q; d = 1, \ldots, k \).

We then make the following substitutions into \( \lambda \):

\( y^d_0 = \sqrt{N} \frac{d}{d} \varepsilon^d_0(\varepsilon^d_0)^{1/2} + \sqrt{p_d} v^d_0, v^d_0 = (p_d - 1)(\varepsilon^d_0)^{1/2} \mathcal{W}(\varepsilon^d_0)^{1/2} ,

\( F^d_i = p_d(n_i - 1)(\varepsilon^d_i)^{1/2} g^d_i(\varepsilon^d_i)^{1/2} ,

\)

and observe that \( \sqrt{N} \varepsilon^d_0 \), \( \mathcal{W}^d_0 \) and \( \mathcal{G}^d_i \) have the distributions:

\( \mathcal{L}(\sqrt{N} \varepsilon^d_0) = n(0, I) \); \( \mathcal{L}(\mathcal{W}^d_0) = \mathcal{W}(I/(p_d - 1)r_p, p_d - 1) \),

\( \mathcal{L}(\mathcal{G}^d_i) = \mathcal{W}(I/[p_d(n_i - 1)], t_i, p_d(n_i - 1)), i = 1, \ldots, q; d = 1, \ldots, k \).

We next define the partial derivatives of \( (2/N) \log \lambda \) with respect to \( \varepsilon^d_0 \), \( \mathcal{W}^d_0 \) and \( \mathcal{G}^d_i \) by

\( \frac{\partial}{\partial \varepsilon^d_0} \frac{(2)}{N} \log \lambda \), \( \frac{\partial}{\partial \mathcal{W}^d_0} (2/N) \log \lambda \),

\( \frac{\partial}{\partial \mathcal{G}^d_i} (2/N) \log \lambda \), \( i = 1, \ldots, q; d = 1, \ldots, k \).

The asymptotic mean and variance for the asymptotic non-null distribution of \( (2/N) \log \lambda \) of the form

\( \mathcal{L}(\sqrt{N}(2/N) \log \lambda - (2/N) \log \lambda \mid b) \Rightarrow n(0, \nu_0) \).
are obtained by evaluating \((2/N)\log \lambda\) at the point \(b\) where

\[ z_0^d = 0, \quad W_{00}^d = I \quad \text{and} \quad G_i^d = I \quad \text{and} \]

\[(I.5.10) \quad v_\infty = d_0 d_0^\dagger + 2 \sum_{d=1}^k \frac{N}{p_d} \{ \text{tr}(H_0^d)^2 + \sum_{i=1}^q \frac{1}{(n_i - 1)} \text{tr}(G_i^d)^2 \} .\]
APPENDIX II

TABLES OF ASYMPTOTIC MEANS AND VARIANCES

This appendix contains tables A, B and C referenced in chapters II and III. Before presenting the tables, we give a short description of the material in each table. Suggestions for the calculation of power and sample sizes under specified alternative hypotheses are also made.

In each table asymptotic means and variances are given for the asymptotic normal distribution

\[
(II.1) \quad \lim_{n \to \infty} \frac{1}{\sqrt{n}} \left( -\frac{2}{n} \log \lambda - \left( -\frac{2}{n} \log \lambda \right)_b \right) = \mathcal{N}(0, \nu_\infty).
\]

Specifically the mean and variance are given where \( \text{MEAN} = \left( -\frac{2}{n} \log \lambda \right)_b \) and \( \text{VAR} = \nu_\infty \). Thus equation (II.1) may be rewritten as

\[
(II.2) \quad \lim_{n \to \infty} \frac{1}{\sqrt{n}} \left( -\frac{2}{n} \log \lambda - \text{MEAN} \right) = \mathcal{N}(0, \text{VAR}).
\]

\( \lambda \) in each of these expressions is the likelihood ratio statistic and \( n \) is the sample size.

To use these results for a practical problem, one uses the approximate (3) expression

\[
(II.3) \quad \frac{1}{\sqrt{n}} \left( -\frac{2}{n} \log \lambda \right) \sim \mathcal{N}(\text{MEAN}, \text{VAR}/n).
\]
This approximation becomes more exact as \( n \) gets large. Expression (II.3) can also be rewritten as

\[
(II.4) \quad \mathcal{L}(-2 \log \lambda) \equiv \mathcal{N}(\text{mean}, \text{VAR}).
\]

Table A.

Table A contains asymptotic means and variances for the distribution of the LRT used in chapter II under the alternative hypothesis of a general 4 \( \times \) 4 circulant model when \( \rho_2 \neq 0 \). (See chapter II, section 6).

One uses the distribution of a test statistic under the alternative hypothesis to calculate, for a specified alternative, either the power at a given sample size or the sample size necessary to achieve a given power, all at a fixed probability of type I error.

Suppose we use the statistic \( \psi = -2 \log \lambda \). We know (see chapter II, (5.7)) that \( \psi \) is asymptotically distributed as \( \chi^2_1 \) and that one rejects the null hypothesis \( H_0: \rho_2 = 0 \) for large values of \( \psi \). The power of a test is the probability that we reject the null hypothesis under a specified alternative. Let \( a = a(\alpha) \) be the upper \( (1-\alpha) \) percentile of a \( \chi^2_1 \) distribution. The power of the test

\[
(II.5) \quad \pi = \pi(\alpha) = \pi(\alpha, \rho_1, \rho_2, n) = P[\psi \geq a(\alpha)|\alpha, n, \rho_1, \rho_2]
\]

may be approximated for large \( n \) using the asymptotic non-null distribution, (II.4). Thus we find that the power is approximately given by
\[ \pi = \pi(\alpha, \rho_1, \rho_2, n) = \Phi \left( \frac{n(\text{MEAN}) - a(\alpha)}{\sqrt{n(\text{VAR})}} \right), \]

where \( \Phi \) is the standard cumulative normal distribution function. The \text{MEAN} and \text{VAR} are given in table A for various \((\rho_1, \rho_2)\) alternatives. Note this expression is independent of the parameter \( \sigma^2 \). For given \( \rho_1, \rho_2, \) and \( \alpha \), we may either approximate the power \( \pi \) at a given sample size \( n \) or solve for the sample size \( n \) at a given power \( \pi \).

Since power and sample size vary depending upon the value of \((\rho_1, \rho_2)\) under consideration, one may create an "indifference zone" consisting of parameters near to the null hypothesis parameters. Then, one may choose a sample size such that all values of the power calculated on these border values exceed a certain value. In calculating power for a certain sample size, one may state the minimum power achieved by these border values.

\textbf{Table B}.

Table B contains asymptotic normal distribution means and variances for the LRT used in chapter III, section 3. The problem being considered is one of choosing between two different forms of a patterned covariance matrix, each form involving \( \sigma^2 \) and \( \rho \). The exact forms of the null and alternative hypothesis are given in chapter III, (3.1).

If we let \( \Psi = -2 \log \lambda \), we have an approximate distribution for \( \Psi \) under the null or alternative hypothesis given by (II.4). The critical point for a size \( \alpha \) test is determined by using the appropriate \text{MEAN} and \text{VAR} terms from the null hypothesis part of table B. Power
and sample size calculations may then be made using the appropriate MEAN and VAR terms for the alternative hypothesis part of table B.

Table C.

Table C contains asymptotic normal means and variances for the LRT used in chapter III, section 3 for the case of a $4 \times 4$ circulant alternative. It may be used to study the results of testing when one has a model that is in neither the null or alternative hypothesis parameter region.

**Interpolation.**

For most power and sample size calculations, linear interpolation should be adequate. Other methods of interpolation such as fitting a polynomial are not worth using since we have given, in chapters II and III, explicit expressions for the asymptotic normal means and variances once the appropriate root of an asymptotic cubic polynomial (chapter II, (6.18), for example) has been found. Thus if an entry not in the tables is needed to more accuracy than obtained by linear interpolation, it may be found by direct evaluation.

**Numerical Examples.**

**Problem A.** Suppose we have four signal receivers located on the vertices of a square and a transmitter at the center of the square. A characteristic of the transmitted signals is measured at each of the receiver stations. Measurements taken on $N$ independent, identically distributed transmissions at each of the four receivers provides us with $N$ independent, identically distributed four-variate row vectors $x_1, \ldots, x_N$. We assume that these
observations have a multivariate normal distribution \( f(x_i) = \mathcal{N} (\mu, \Sigma) \), \( i = 1, \ldots, N \). We are interested in the structure of the covariance matrix and thus concentrate only on the sample covariance matrix

\[
C = \sum_{i=1}^{N} (x_i - \bar{x})'(x_i - \bar{x}).
\]

Assume that the covariance matrix \( \Sigma \) is a \( 4 \times 4 \) circulant,

\[
\Sigma = \sigma^2 \begin{bmatrix}
1 & \rho_1 & \rho_2 & \rho_4 \\
\rho_1 & 1 & \rho_1 & \rho_2 \\
\rho_2 & \rho_1 & 1 & \rho_1 \\
\rho_4 & \rho_2 & \rho_1 & 1
\end{bmatrix}.
\]

This assumption is motivated by the physical model. One expects each receiver to have a first order correlation, \( \rho_1 \), with the two adjacent receivers and a second order correlation, \( \rho_2 \), with the non-adjacent receiver.

We are interested in testing the null hypothesis \( H_0: \rho_2 = 0 \), i.e. that non-adjacent receivers are not correlated, versus the alternative \( H_A: \rho_2 \neq 0 \).

The problem under consideration is to perform a 5% test of significance for this hypothesis and also to determine the power of this test under the alternative \( (\rho_1, \rho_2) = (.40, .30) \) for several sample sizes \( n = N-1 \).

Solution. The solution follows directly from chapter II and the earlier discussion of Table A in this appendix. To test the hypothesis, we form the LRT \( \lambda \) given in chapter II, (5.6). We form the test statistic
\( \psi = -2 \log \lambda \). For large samples \((n)\), we know that \( \psi \) is approximately distributed as a chi-squared variable with one degree of freedom under the null hypothesis. The 5\% significance critical value, \( a(.05) \), is approximated by the 95\% percentile of a \( \chi^2_1 \) distribution. Thus \( a(.05) \approx 3.841 \). We reject the null hypothesis if \( \psi \geq a(.05) \).

To determine the power of this test under the alternative \((\rho_1, \rho_2) = (.40, .30)\), we use Table A and apply the formula for power given in (II.6). We enter Table A at the value \((\rho_1, \rho_2) = (.40, .30)\) and note that this value is between the tabled values, \((\rho_1, \rho_2) = (3.54545, .30)\) and \((.413636, .30)\). We find the MEAN and VAR values for \((\rho_1, \rho_2) = (.40, .30)\) by linear interpolation.

<table>
<thead>
<tr>
<th>( \rho_1 )</th>
<th>( \rho_2 )</th>
<th>MEAN</th>
<th>VAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>.354545</td>
<td>.30</td>
<td>.17201</td>
<td>.68896</td>
</tr>
<tr>
<td>.40</td>
<td>.30</td>
<td>.16746</td>
<td>.67925</td>
</tr>
<tr>
<td>.413636</td>
<td>.30</td>
<td>.16610</td>
<td>.67634</td>
</tr>
</tbody>
</table>

The power \( \pi \), as approximated by (II.6), is given by

\[
(II.7) \quad \pi = \pi(.05, .40, .30, n) = \Phi \left( \frac{n(.16746) - 3.841}{\sqrt{n(.67925)}} \right).
\]

From (II.7), we may generate the following set of power values for different sample sizes.

<table>
<thead>
<tr>
<th>Sample size ((n))</th>
<th>40</th>
<th>60</th>
<th>80</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power ((\pi))</td>
<td>.71</td>
<td>.83</td>
<td>.90</td>
<td>.94</td>
</tr>
</tbody>
</table>
**Problem B.** We continue with the problem described in Problem A testing a set of hypotheses considered in chapter III, section 3. We are still interested in testing hypotheses involving the covariance structure of $\Sigma$ and again assume that $\Sigma$ has the structure of a $4 \times 4$ circulant.

Under the null hypothesis, we assume that the second order correlation is zero, $H_0: \rho_2 = 0$; but for the alternative hypothesis, we assume that the first and second order correlation are related, $H_A: \rho_1 = \rho^1, \rho_2 = \rho^2$.

The problem under consideration is to perform a 5% test of significance for these hypotheses when $\rho_1 = .40$ and also to determine the power of this test under the alternative $\rho = .40$ for a sample size $n = 100$.

**Solution.** The solution follows directly from chapter III, section 3 and the discussion of Table B in this appendix. To test these hypotheses, we form the LRT $\lambda$ given in chapter III, (3.2). We form the test statistic $\psi = -2\log \lambda$. For large samples ($n$), the 5% significance critical value, $a(.05)$, may be approximated by the 95th percentile of a normal distribution, $\hat{n}(n(\text{MEAN}), n(\text{VAR}))$. We find the MEAN and VAR values by entering Table B at $\rho = .40$ in the null hypothesis section. Using the fact that $\Phi(1.645) = .95$ and the normal approximation (II.4), we find that

$$a(.05) = n(\text{MEAN}) + 1.645 \sqrt{n(\text{VAR})} = 100(-.129) + 1.645 \sqrt{100(.500)}$$

$$= -1.27.$$

To find the power of this test, we again use the normal approximation given in (II.4). The MEAN and VAR values for the alternative $\rho = .40$
are given in Table B under the alternative hypothesis section. We find the power \( \pi \) is given by

\[
\pi = \Phi\left( \frac{n(\text{MEAN}) - a(\alpha)}{\sqrt{n(\text{VAR})}} \right) = \Phi\left( \frac{100(0.034) + 1.27}{\sqrt{100(0.179)}} \right)
\]

\[
= \Phi(0.31) = 0.62 .
\]

**Problem C.** Suppose in Problem B we are actually sampling from a \( 4 \times 4 \) circulant with \( \rho_1 = .20 \) and \( \rho_2 = .18 \). What is the probability that we accept the null hypothesis?

**Solution.** We again use the normal approximation in (II.4) and seek the probability that we select a value below \( a(.05) = -1.27 \) when sampling from a normal distribution, \( \mathcal{N}(n(\text{MEAN}), n(\text{VAR})) \). Here the MEAN and VAR terms are found by entering Table C under \( \rho_1 = .20 \) and \( \rho_2 = .18 \). The probability of accepting the null hypothesis is given by

\[
\Phi\left( \frac{a(\alpha) - n(\text{MEAN})}{\sqrt{n(\text{VAR})}} \right) = \Phi\left( \frac{-1.27 - 100(0.0175)}{\sqrt{100(0.0153)}} \right)
\]

\[
= \Phi(-2.44) = .007 .
\]
<table>
<thead>
<tr>
<th>P-2</th>
<th>p-1</th>
<th>P-1</th>
<th>P-1</th>
<th>P-1</th>
<th>P-1</th>
<th>P-1</th>
<th>P-1</th>
<th>P-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.69</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>MEAN</td>
<td>7.8139</td>
<td>7.8139</td>
<td>7.8139</td>
<td>7.8139</td>
<td>7.8139</td>
<td>7.8139</td>
<td>7.8139</td>
<td>7.8139</td>
</tr>
</tbody>
</table>

**TABLE A (See Appendix II)**
<table>
<thead>
<tr>
<th>P</th>
<th>P-1</th>
<th>P-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.57</td>
<td>0.00</td>
<td>0.071364</td>
</tr>
<tr>
<td>0.59</td>
<td>0.00</td>
<td>0.070000</td>
</tr>
<tr>
<td>0.61</td>
<td>0.00</td>
<td>0.068636</td>
</tr>
<tr>
<td>0.63</td>
<td>0.00</td>
<td>0.067273</td>
</tr>
<tr>
<td>0.65</td>
<td>0.00</td>
<td>0.065909</td>
</tr>
<tr>
<td>0.67</td>
<td>0.00</td>
<td>0.064545</td>
</tr>
<tr>
<td>0.69</td>
<td>0.00</td>
<td>0.063182</td>
</tr>
<tr>
<td>0.71</td>
<td>0.00</td>
<td>0.061818</td>
</tr>
<tr>
<td>0.73</td>
<td>0.00</td>
<td>0.060455</td>
</tr>
<tr>
<td>0.75</td>
<td>0.00</td>
<td>0.059081</td>
</tr>
<tr>
<td>0.77</td>
<td>0.00</td>
<td>0.057727</td>
</tr>
<tr>
<td>0.79</td>
<td>0.00</td>
<td>0.056364</td>
</tr>
<tr>
<td>0.81</td>
<td>0.00</td>
<td>0.055000</td>
</tr>
<tr>
<td>0.83</td>
<td>0.00</td>
<td>0.053636</td>
</tr>
<tr>
<td>0.85</td>
<td>0.00</td>
<td>0.052262</td>
</tr>
<tr>
<td>0.87</td>
<td>0.00</td>
<td>0.050887</td>
</tr>
<tr>
<td>0.89</td>
<td>0.00</td>
<td>0.049512</td>
</tr>
<tr>
<td>0.91</td>
<td>0.00</td>
<td>0.048137</td>
</tr>
<tr>
<td>0.93</td>
<td>0.00</td>
<td>0.046762</td>
</tr>
<tr>
<td>0.95</td>
<td>0.00</td>
<td>0.045387</td>
</tr>
<tr>
<td>0.97</td>
<td>0.00</td>
<td>0.044012</td>
</tr>
<tr>
<td>0.99</td>
<td>0.00</td>
<td>0.042637</td>
</tr>
</tbody>
</table>

**TABLE A**

**TABLE OF ASYMPTOTIC NON-NULL MEANS AND VARIANCES**

(* Indicates 3 REAL ROOTS*)
<table>
<thead>
<tr>
<th>P-2</th>
<th>P-1</th>
<th>TABLE OF ASYMPTOTIC NON-NULL MEANS AND VARIANCES (* INDICATES 3 REAL ROOTS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.15</td>
<td>0.15 0.05273 0.0585 0.05489 0.0590 0.0536 0.0524 0.0520 0.0516 0.0512 0.0508 0.0504</td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>0.04551 0.04537 0.0450 0.0448 0.0446 0.0444 0.0442 0.0440 0.0438 0.0436 0.0434</td>
<td></td>
</tr>
<tr>
<td>VAP.</td>
<td>0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.6 0.5909 0.618 0.625 0.627 0.629 0.631 0.633 0.635 0.637 0.639 0.641</td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>0.0529 0.0527 0.0525 0.0523 0.0521 0.0519 0.0517 0.0515 0.0513 0.0511 0.0509</td>
<td></td>
</tr>
<tr>
<td>VAP.</td>
<td>0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520</td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>0.15 0.05273 0.0585 0.05489 0.0590 0.0536 0.0524 0.0520 0.0516 0.0512 0.0508 0.0504</td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>0.04551 0.04537 0.0450 0.0448 0.0446 0.0444 0.0442 0.0440 0.0438 0.0436 0.0434</td>
<td></td>
</tr>
<tr>
<td>VAP.</td>
<td>0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.6 0.5909 0.618 0.625 0.627 0.629 0.631 0.633 0.635 0.637 0.639 0.641</td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>0.0529 0.0527 0.0525 0.0523 0.0521 0.0519 0.0517 0.0515 0.0513 0.0511 0.0509</td>
<td></td>
</tr>
<tr>
<td>VAP.</td>
<td>0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520</td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td>0.15 0.05273 0.0585 0.05489 0.0590 0.0536 0.0524 0.0520 0.0516 0.0512 0.0508 0.0504</td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>0.04551 0.04537 0.0450 0.0448 0.0446 0.0444 0.0442 0.0440 0.0438 0.0436 0.0434</td>
<td></td>
</tr>
<tr>
<td>VAP.</td>
<td>0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>0.6 0.5909 0.618 0.625 0.627 0.629 0.631 0.633 0.635 0.637 0.639 0.641</td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>0.0529 0.0527 0.0525 0.0523 0.0521 0.0519 0.0517 0.0515 0.0513 0.0511 0.0509</td>
<td></td>
</tr>
<tr>
<td>VAP.</td>
<td>0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520 0.0520</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE A**
<table>
<thead>
<tr>
<th>P-2</th>
<th>P-1</th>
<th>P-1</th>
<th>P-1</th>
<th>P-1</th>
<th>P-1</th>
<th>P-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>MEAN</td>
<td>0.10862</td>
<td>0.1863</td>
<td>0.1863</td>
<td>0.1863</td>
<td>0.1863</td>
<td>0.1863</td>
</tr>
<tr>
<td>VAR.</td>
<td>0.07200</td>
<td>0.07062</td>
<td>0.07062</td>
<td>0.07062</td>
<td>0.07062</td>
<td>0.07062</td>
</tr>
<tr>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>MEAN</td>
<td>0.23860</td>
<td>0.23860</td>
<td>0.23860</td>
<td>0.23860</td>
<td>0.23860</td>
<td>0.23860</td>
</tr>
<tr>
<td>VAR.</td>
<td>0.04102</td>
<td>0.04102</td>
<td>0.04102</td>
<td>0.04102</td>
<td>0.04102</td>
<td>0.04102</td>
</tr>
<tr>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
<td>0.66</td>
</tr>
<tr>
<td>MEAN</td>
<td>0.27761</td>
<td>0.27761</td>
<td>0.27761</td>
<td>0.27761</td>
<td>0.27761</td>
<td>0.27761</td>
</tr>
<tr>
<td>VAR.</td>
<td>0.03680</td>
<td>0.03680</td>
<td>0.03680</td>
<td>0.03680</td>
<td>0.03680</td>
<td>0.03680</td>
</tr>
<tr>
<td>-0.33</td>
<td>-0.33</td>
<td>-0.33</td>
<td>-0.33</td>
<td>-0.33</td>
<td>-0.33</td>
<td>-0.33</td>
</tr>
<tr>
<td>MEAN</td>
<td>0.32999</td>
<td>0.32999</td>
<td>0.32999</td>
<td>0.32999</td>
<td>0.32999</td>
<td>0.32999</td>
</tr>
<tr>
<td>VAR.</td>
<td>0.12160</td>
<td>0.12160</td>
<td>0.12160</td>
<td>0.12160</td>
<td>0.12160</td>
<td>0.12160</td>
</tr>
<tr>
<td>-0.66</td>
<td>-0.66</td>
<td>-0.66</td>
<td>-0.66</td>
<td>-0.66</td>
<td>-0.66</td>
<td>-0.66</td>
</tr>
<tr>
<td>MEAN</td>
<td>0.38814</td>
<td>0.38814</td>
<td>0.38814</td>
<td>0.38814</td>
<td>0.38814</td>
<td>0.38814</td>
</tr>
<tr>
<td>VAR.</td>
<td>0.14120</td>
<td>0.14120</td>
<td>0.14120</td>
<td>0.14120</td>
<td>0.14120</td>
<td>0.14120</td>
</tr>
<tr>
<td>-1.0</td>
<td>-1.0</td>
<td>-1.0</td>
<td>-1.0</td>
<td>-1.0</td>
<td>-1.0</td>
<td>-1.0</td>
</tr>
<tr>
<td>MEAN</td>
<td>0.45255</td>
<td>0.45255</td>
<td>0.45255</td>
<td>0.45255</td>
<td>0.45255</td>
<td>0.45255</td>
</tr>
<tr>
<td>VAR.</td>
<td>0.46200</td>
<td>0.46200</td>
<td>0.46200</td>
<td>0.46200</td>
<td>0.46200</td>
<td>0.46200</td>
</tr>
<tr>
<td>-1.33</td>
<td>-1.33</td>
<td>-1.33</td>
<td>-1.33</td>
<td>-1.33</td>
<td>-1.33</td>
<td>-1.33</td>
</tr>
<tr>
<td>MEAN</td>
<td>0.52537</td>
<td>0.52537</td>
<td>0.52537</td>
<td>0.52537</td>
<td>0.52537</td>
<td>0.52537</td>
</tr>
<tr>
<td>VAR.</td>
<td>1.84320</td>
<td>1.84320</td>
<td>1.84320</td>
<td>1.84320</td>
<td>1.84320</td>
<td>1.84320</td>
</tr>
<tr>
<td>-1.66</td>
<td>-1.66</td>
<td>-1.66</td>
<td>-1.66</td>
<td>-1.66</td>
<td>-1.66</td>
<td>-1.66</td>
</tr>
<tr>
<td>MEAN</td>
<td>0.60321</td>
<td>0.60321</td>
<td>0.60321</td>
<td>0.60321</td>
<td>0.60321</td>
<td>0.60321</td>
</tr>
<tr>
<td>VAR.</td>
<td>2.02649</td>
<td>2.02649</td>
<td>2.02649</td>
<td>2.02649</td>
<td>2.02649</td>
<td>2.02649</td>
</tr>
<tr>
<td>MEAN</td>
<td>0.66506</td>
<td>0.66506</td>
<td>0.66506</td>
<td>0.66506</td>
<td>0.66506</td>
<td>0.66506</td>
</tr>
<tr>
<td>VAR.</td>
<td>2.10390</td>
<td>2.10390</td>
<td>2.10390</td>
<td>2.10390</td>
<td>2.10390</td>
<td>2.10390</td>
</tr>
<tr>
<td>-2.33</td>
<td>-2.33</td>
<td>-2.33</td>
<td>-2.33</td>
<td>-2.33</td>
<td>-2.33</td>
<td>-2.33</td>
</tr>
<tr>
<td>MEAN</td>
<td>0.77180</td>
<td>0.77180</td>
<td>0.77180</td>
<td>0.77180</td>
<td>0.77180</td>
<td>0.77180</td>
</tr>
<tr>
<td>VAR.</td>
<td>2.17834</td>
<td>2.17834</td>
<td>2.17834</td>
<td>2.17834</td>
<td>2.17834</td>
<td>2.17834</td>
</tr>
<tr>
<td>MEAN</td>
<td>0.86305</td>
<td>0.86305</td>
<td>0.86305</td>
<td>0.86305</td>
<td>0.86305</td>
<td>0.86305</td>
</tr>
<tr>
<td>VAR.</td>
<td>2.25000</td>
<td>2.25000</td>
<td>2.25000</td>
<td>2.25000</td>
<td>2.25000</td>
<td>2.25000</td>
</tr>
<tr>
<td>MEAN</td>
<td>0.95958</td>
<td>0.95958</td>
<td>0.95958</td>
<td>0.95958</td>
<td>0.95958</td>
<td>0.95958</td>
</tr>
<tr>
<td>VAR.</td>
<td>2.31902</td>
<td>2.31902</td>
<td>2.31902</td>
<td>2.31902</td>
<td>2.31902</td>
<td>2.31902</td>
</tr>
<tr>
<td>MEAN</td>
<td>1.06238</td>
<td>1.06238</td>
<td>1.06238</td>
<td>1.06238</td>
<td>1.06238</td>
<td>1.06238</td>
</tr>
<tr>
<td>VAR.</td>
<td>2.38554</td>
<td>2.38554</td>
<td>2.38554</td>
<td>2.38554</td>
<td>2.38554</td>
<td>2.38554</td>
</tr>
<tr>
<td>MEAN</td>
<td>1.17266</td>
<td>1.17266</td>
<td>1.17266</td>
<td>1.17266</td>
<td>1.17266</td>
<td>1.17266</td>
</tr>
<tr>
<td>VAR.</td>
<td>2.44976</td>
<td>2.44976</td>
<td>2.44976</td>
<td>2.44976</td>
<td>2.44976</td>
<td>2.44976</td>
</tr>
</tbody>
</table>

**TABLE A**
<table>
<thead>
<tr>
<th>P-2</th>
<th>P-1</th>
<th>P-1</th>
<th>P-1</th>
<th>P-1</th>
<th>P-1</th>
<th>P-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.72</td>
<td>0.0</td>
<td>0.012727</td>
<td>0.025455</td>
<td>0.038182</td>
<td>0.050909</td>
<td>0.063636</td>
</tr>
<tr>
<td>Mean</td>
<td>1.29204</td>
<td>1.32311</td>
<td>1.35419</td>
<td>1.38526</td>
<td>1.41633</td>
<td>1.44740</td>
</tr>
<tr>
<td>Var</td>
<td>2.51163</td>
<td>2.51163</td>
<td>2.51163</td>
<td>2.51163</td>
<td>2.51163</td>
<td>2.51163</td>
</tr>
<tr>
<td>-0.75</td>
<td>0.0</td>
<td>0.013468</td>
<td>0.027772</td>
<td>0.041078</td>
<td>0.054384</td>
<td>0.067690</td>
</tr>
<tr>
<td>Mean</td>
<td>1.42266</td>
<td>1.45366</td>
<td>1.48466</td>
<td>1.51566</td>
<td>1.54666</td>
<td>1.57766</td>
</tr>
<tr>
<td>Var</td>
<td>2.57714</td>
<td>2.57714</td>
<td>2.57714</td>
<td>2.57714</td>
<td>2.57714</td>
<td>2.57714</td>
</tr>
<tr>
<td>-0.74</td>
<td>0.0</td>
<td>0.010000</td>
<td>0.020000</td>
<td>0.030000</td>
<td>0.040000</td>
<td>0.050000</td>
</tr>
<tr>
<td>Mean</td>
<td>1.56749</td>
<td>1.56749</td>
<td>1.56749</td>
<td>1.56749</td>
<td>1.56749</td>
<td>1.56749</td>
</tr>
<tr>
<td>Var</td>
<td>2.62821</td>
<td>2.62821</td>
<td>2.62821</td>
<td>2.62821</td>
<td>2.62821</td>
<td>2.62821</td>
</tr>
<tr>
<td>-0.70</td>
<td>0.0</td>
<td>0.006364</td>
<td>0.006364</td>
<td>0.006364</td>
<td>0.006364</td>
<td>0.006364</td>
</tr>
<tr>
<td>Mean</td>
<td>1.71081</td>
<td>1.71081</td>
<td>1.71081</td>
<td>1.71081</td>
<td>1.71081</td>
<td>1.71081</td>
</tr>
<tr>
<td>Var</td>
<td>2.68808</td>
<td>2.68808</td>
<td>2.68808</td>
<td>2.68808</td>
<td>2.68808</td>
<td>2.68808</td>
</tr>
<tr>
<td>-0.68</td>
<td>0.0</td>
<td>0.006364</td>
<td>0.006364</td>
<td>0.006364</td>
<td>0.006364</td>
<td>0.006364</td>
</tr>
<tr>
<td>Mean</td>
<td>1.87081</td>
<td>1.87081</td>
<td>1.87081</td>
<td>1.87081</td>
<td>1.87081</td>
<td>1.87081</td>
</tr>
<tr>
<td>Var</td>
<td>2.73913</td>
<td>2.73913</td>
<td>2.73913</td>
<td>2.73913</td>
<td>2.73913</td>
<td>2.73913</td>
</tr>
<tr>
<td>-0.67</td>
<td>0.0</td>
<td>0.006364</td>
<td>0.006364</td>
<td>0.006364</td>
<td>0.006364</td>
<td>0.006364</td>
</tr>
<tr>
<td>Mean</td>
<td>1.92857</td>
<td>1.92857</td>
<td>1.92857</td>
<td>1.92857</td>
<td>1.92857</td>
<td>1.92857</td>
</tr>
<tr>
<td>Var</td>
<td>2.79546</td>
<td>2.79546</td>
<td>2.79546</td>
<td>2.79546</td>
<td>2.79546</td>
<td>2.79546</td>
</tr>
<tr>
<td>-0.66</td>
<td>0.0</td>
<td>0.006364</td>
<td>0.006364</td>
<td>0.006364</td>
<td>0.006364</td>
<td>0.006364</td>
</tr>
<tr>
<td>Mean</td>
<td>1.98571</td>
<td>1.98571</td>
<td>1.98571</td>
<td>1.98571</td>
<td>1.98571</td>
<td>1.98571</td>
</tr>
<tr>
<td>Var</td>
<td>2.84211</td>
<td>2.84211</td>
<td>2.84211</td>
<td>2.84211</td>
<td>2.84211</td>
<td>2.84211</td>
</tr>
<tr>
<td>-0.64</td>
<td>0.0</td>
<td>0.006364</td>
<td>0.006364</td>
<td>0.006364</td>
<td>0.006364</td>
<td>0.006364</td>
</tr>
<tr>
<td>Mean</td>
<td>2.03929</td>
<td>2.03929</td>
<td>2.03929</td>
<td>2.03929</td>
<td>2.03929</td>
<td>2.03929</td>
</tr>
<tr>
<td>Var</td>
<td>2.89719</td>
<td>2.89719</td>
<td>2.89719</td>
<td>2.89719</td>
<td>2.89719</td>
<td>2.89719</td>
</tr>
<tr>
<td>-0.62</td>
<td>0.0</td>
<td>0.006364</td>
<td>0.006364</td>
<td>0.006364</td>
<td>0.006364</td>
<td>0.006364</td>
</tr>
<tr>
<td>Mean</td>
<td>2.10393</td>
<td>2.10393</td>
<td>2.10393</td>
<td>2.10393</td>
<td>2.10393</td>
<td>2.10393</td>
</tr>
<tr>
<td>Var</td>
<td>2.95774</td>
<td>2.95774</td>
<td>2.95774</td>
<td>2.95774</td>
<td>2.95774</td>
<td>2.95774</td>
</tr>
</tbody>
</table>

**TABLE A**
<table>
<thead>
<tr>
<th>P</th>
<th>MEAN</th>
<th>VAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>-0.000009020</td>
<td>0.000000000</td>
</tr>
<tr>
<td>0.02</td>
<td>-0.000003311</td>
<td>0.000002124</td>
</tr>
<tr>
<td>0.03</td>
<td>-0.000009136</td>
<td>0.000006053</td>
</tr>
<tr>
<td>0.04</td>
<td>-0.000004515</td>
<td>0.000002642</td>
</tr>
<tr>
<td>0.05</td>
<td>-0.000012626</td>
<td>0.000012500</td>
</tr>
<tr>
<td>0.06</td>
<td>-0.000016256</td>
<td>0.000025799</td>
</tr>
<tr>
<td>0.07</td>
<td>-0.000019978</td>
<td>0.000026200</td>
</tr>
<tr>
<td>0.08</td>
<td>-0.000023587</td>
<td></td>
</tr>
<tr>
<td>0.09</td>
<td>-0.000027105</td>
<td>0.000026382</td>
</tr>
<tr>
<td>0.10</td>
<td>-0.000028730</td>
<td>0.000031360</td>
</tr>
<tr>
<td>0.11</td>
<td>-0.000030762</td>
<td>0.000031047</td>
</tr>
<tr>
<td>0.12</td>
<td>-0.000041782</td>
<td>0.000031497</td>
</tr>
<tr>
<td>0.13</td>
<td>-0.000042125</td>
<td>0.000031486</td>
</tr>
<tr>
<td>0.14</td>
<td>-0.000031171</td>
<td>0.000044500</td>
</tr>
<tr>
<td>0.15</td>
<td>-0.000018653</td>
<td>0.000053284</td>
</tr>
<tr>
<td>0.16</td>
<td>-0.000018091</td>
<td>0.000075086</td>
</tr>
<tr>
<td>0.17</td>
<td>-0.000018976</td>
<td>0.000074824</td>
</tr>
<tr>
<td>0.18</td>
<td>-0.000019200</td>
<td>0.000074525</td>
</tr>
<tr>
<td>0.19</td>
<td>-0.000019576</td>
<td>0.000074216</td>
</tr>
<tr>
<td>0.20</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.21</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.22</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.23</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.24</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.25</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.26</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.27</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.28</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.29</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.30</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.31</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.32</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.33</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.34</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.35</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.36</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.37</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.38</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.39</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.40</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.41</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.42</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.43</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.44</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.45</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.46</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.47</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.48</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
<tr>
<td>0.49</td>
<td>-0.000019554</td>
<td>0.000074125</td>
</tr>
</tbody>
</table>

Null Hypothesis

Alternative Hypothesis

Table B (See Appendix II)
<table>
<thead>
<tr>
<th>P-2</th>
<th>P-1</th>
<th>P-1</th>
<th>P-1</th>
<th>P-1</th>
<th>P-1</th>
<th>P-1</th>
<th>P-1</th>
<th>P-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50</td>
<td>0.002917</td>
<td>0.156333</td>
<td>0.2687750</td>
<td>0.336167</td>
<td>0.504930</td>
<td>0.475000</td>
<td>0.500017</td>
<td>0.633333</td>
</tr>
<tr>
<td>0.99</td>
<td>0.01961</td>
<td>0.029269</td>
<td>0.07685</td>
<td>0.12964</td>
<td>0.19401</td>
<td>0.29705</td>
<td>0.41243</td>
<td>0.59289</td>
</tr>
<tr>
<td>0.01</td>
<td>0.01218</td>
<td>0.04117</td>
<td>0.11238</td>
<td>0.20437</td>
<td>0.32199</td>
<td>0.49300</td>
<td>0.70254</td>
<td>0.98085</td>
</tr>
<tr>
<td>0.50</td>
<td>0.561667</td>
<td>0.161333</td>
<td>0.295000</td>
<td>0.365667</td>
<td>0.400000</td>
<td>0.400000</td>
<td>0.576667</td>
<td>0.633333</td>
</tr>
<tr>
<td>0.99</td>
<td>0.00499</td>
<td>0.01920</td>
<td>0.04185</td>
<td>0.07577</td>
<td>0.12514</td>
<td>0.19048</td>
<td>0.27857</td>
<td>0.39966</td>
</tr>
<tr>
<td>0.01</td>
<td>0.00417</td>
<td>0.04083</td>
<td>0.12120</td>
<td>0.21616</td>
<td>0.31264</td>
<td>0.46645</td>
<td>0.71995</td>
<td>0.92689</td>
</tr>
<tr>
<td>0.93</td>
<td>0.080417</td>
<td>0.106083</td>
<td>0.261250</td>
<td>0.45893</td>
<td>0.45893</td>
<td>0.45893</td>
<td>0.562917</td>
<td>0.643333</td>
</tr>
<tr>
<td>0.67</td>
<td>0.020426</td>
<td>0.01724</td>
<td>0.03761</td>
<td>0.07258</td>
<td>0.11812</td>
<td>0.17942</td>
<td>0.26162</td>
<td>0.37361</td>
</tr>
<tr>
<td>0.01</td>
<td>0.01239</td>
<td>0.04191</td>
<td>0.05565</td>
<td>0.17363</td>
<td>0.27892</td>
<td>0.41626</td>
<td>0.59282</td>
<td>0.81994</td>
</tr>
<tr>
<td>0.67</td>
<td>0.07917</td>
<td>0.155833</td>
<td>0.233750</td>
<td>0.311667</td>
<td>0.36583</td>
<td>0.467500</td>
<td>0.54517</td>
<td>0.623333</td>
</tr>
<tr>
<td>0.01</td>
<td>0.00892</td>
<td>0.03960</td>
<td>0.06036</td>
<td>0.16390</td>
<td>0.26308</td>
<td>0.39222</td>
<td>0.55793</td>
<td>0.77068</td>
</tr>
<tr>
<td>0.64</td>
<td>0.076667</td>
<td>0.151333</td>
<td>0.230000</td>
<td>0.306667</td>
<td>0.38133</td>
<td>0.460000</td>
<td>0.536667</td>
<td>0.613333</td>
</tr>
<tr>
<td>0.67</td>
<td>0.01024</td>
<td>0.01627</td>
<td>0.03735</td>
<td>0.06836</td>
<td>0.11017</td>
<td>0.16332</td>
<td>0.24466</td>
<td>0.34875</td>
</tr>
<tr>
<td>0.01</td>
<td>0.00956</td>
<td>0.03175</td>
<td>0.05818</td>
<td>0.15442</td>
<td>0.24769</td>
<td>0.36895</td>
<td>0.52429</td>
<td>0.72349</td>
</tr>
<tr>
<td>0.61</td>
<td>0.075417</td>
<td>0.150833</td>
<td>0.226250</td>
<td>0.310667</td>
<td>0.37708</td>
<td>0.452500</td>
<td>0.52971</td>
<td>0.603333</td>
</tr>
<tr>
<td>0.72</td>
<td>0.007167</td>
<td>0.014333</td>
<td>0.025000</td>
<td>0.05867</td>
<td>0.10396</td>
<td>0.15756</td>
<td>0.22766</td>
<td>0.32215</td>
</tr>
<tr>
<td>0.50</td>
<td>0.071667</td>
<td>0.148333</td>
<td>0.222500</td>
<td>0.298667</td>
<td>0.37083</td>
<td>0.485000</td>
<td>0.51167</td>
<td>0.543333</td>
</tr>
<tr>
<td>0.69</td>
<td>0.073454</td>
<td>0.01730</td>
<td>0.03276</td>
<td>0.06500</td>
<td>0.10568</td>
<td>0.16592</td>
<td>0.21060</td>
<td>0.29645</td>
</tr>
<tr>
<td>0.66</td>
<td>0.071717</td>
<td>0.02890</td>
<td>0.06681</td>
<td>0.11190</td>
<td>0.17061</td>
<td>0.23828</td>
<td>0.34081</td>
<td>0.55443</td>
</tr>
<tr>
<td>0.69</td>
<td>0.070417</td>
<td>0.014833</td>
<td>0.021250</td>
<td>0.05867</td>
<td>0.10568</td>
<td>0.16592</td>
<td>0.21060</td>
<td>0.29645</td>
</tr>
<tr>
<td>0.66</td>
<td>0.069167</td>
<td>0.138333</td>
<td>0.207500</td>
<td>0.276667</td>
<td>0.34833</td>
<td>0.415000</td>
<td>0.48167</td>
<td>0.563333</td>
</tr>
<tr>
<td>0.63</td>
<td>0.067917</td>
<td>0.158333</td>
<td>0.237500</td>
<td>0.271667</td>
<td>0.33983</td>
<td>0.407500</td>
<td>0.475017</td>
<td>0.543333</td>
</tr>
<tr>
<td>0.61</td>
<td>0.066667</td>
<td>0.133333</td>
<td>0.200000</td>
<td>0.266667</td>
<td>0.33333</td>
<td>0.400000</td>
<td>0.466667</td>
<td>0.533333</td>
</tr>
<tr>
<td>0.60</td>
<td>0.06530</td>
<td>0.01227</td>
<td>0.02866</td>
<td>0.05104</td>
<td>0.08222</td>
<td>0.12315</td>
<td>0.17617</td>
<td>0.24887</td>
</tr>
</tbody>
</table>

**TABLE C** (See Appendix II)
<table>
<thead>
<tr>
<th>P-2</th>
<th>P-1</th>
<th>TABLE OF ASYMPTOTIC NON-NULL MEANS AND VARIANCES (# INDICATES 3 REAL ROOTS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.27</td>
<td>0.030417</td>
<td>0.060333</td>
</tr>
<tr>
<td>-0.30</td>
<td>0.029167</td>
<td>0.058333</td>
</tr>
<tr>
<td>-0.33</td>
<td>0.027917</td>
<td>0.055417</td>
</tr>
<tr>
<td>-0.36</td>
<td>0.026667</td>
<td>0.051333</td>
</tr>
<tr>
<td>-0.39</td>
<td>0.025533</td>
<td>0.047333</td>
</tr>
<tr>
<td>-0.42</td>
<td>0.024667</td>
<td>0.043333</td>
</tr>
<tr>
<td>-0.45</td>
<td>0.023917</td>
<td>0.039333</td>
</tr>
<tr>
<td>-0.48</td>
<td>0.023217</td>
<td>0.035333</td>
</tr>
<tr>
<td>-0.51</td>
<td>0.022617</td>
<td>0.031333</td>
</tr>
<tr>
<td>-0.54</td>
<td>0.022067</td>
<td>0.027333</td>
</tr>
<tr>
<td>-0.57</td>
<td>0.021567</td>
<td>0.023333</td>
</tr>
<tr>
<td>-0.60</td>
<td>0.021167</td>
<td>0.019333</td>
</tr>
<tr>
<td>-0.63</td>
<td>0.020717</td>
<td>0.015333</td>
</tr>
<tr>
<td>-0.66</td>
<td>0.020317</td>
<td>0.011333</td>
</tr>
<tr>
<td>MEAN</td>
<td>0.020917</td>
<td>0.017333</td>
</tr>
<tr>
<td>VARP</td>
<td>0.020517</td>
<td>0.013333</td>
</tr>
</tbody>
</table>

**TABLE C**


