DIMENSION REDUCTION IN PCA:
LIKELIHOOD-BASED METHODS

by
Kamolchanok Choochaow

A dissertation submitted in partial fulfillment
of the requirements for the degree
of
Doctor of Philosophy
in
Statistics

MONTANA STATE UNIVERSITY
Bozeman, Montana

July 2002
APPROVAL

of a dissertation submitted by

Kamolchanok Choochaow

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Robert J. Boik  
(Signature)  
Date

Approved for the Department of Statistics

Kenneth L. Bowers  
(Signature)  
Date

Approved for the College of Graduate Studies

Bruce McLeod  
(Signature)  
Date
STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with “fair use” as prescribed in the U. S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to Bell & Howell Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted “the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part.”

Signature ________________________________

Date ________________________________
ACKNOWLEDGEMENTS

I would like to thank my adviser, Professor Robert J. Boik, for his invaluable assistance and guidance during the preparation of this thesis. I truly appreciate his encouragement and patience.

I am also grateful for John Borkowski, Jim Robinson-Cox, Steve Cherry, and Bill Quimby’s advice throughout my graduate studies.

Lastly, I would like to thank my dad, mom, and husband for their love and support.
TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................ vii

LIST OF FIGURES ..................................................................................................... x

1. INTRODUCTION .................................................................................................... 1
   Principal Component Analysis ............................................................................... 2
   Reduction in the Number of Components .......................................................... 4
   Reduction in the Number of Original Variables ............................................... 12
   Example 1 ........................................................................................................... 30
   Example 2 ........................................................................................................... 32

2. PARAMETERIZATION .......................................................................................... 35
   Spectral Model ................................................................................................... 35
   Parameterization of Eigenvalues ....................................................................... 36
      Reduction in the Number of Variables .......................................................... 37
      Reduction in the Number of Components ..................................................... 41
   Parameterization of Eigenvectors ................................................................... 52
   Solving for Implicit Parameters ....................................................................... 70
      Solving for $\mathbf{\tau}$ ....................................................................................... 70
      Solving for $\mathbf{\eta}$ ....................................................................................... 71

3. ESTIMATING PARAMETERS AND CONSTRUCTING ASYMPTOTIC DISTRIBUTIONS ................................................................. 72
   Loglikelihood Function ..................................................................................... 73
   Fisher Scoring Algorithm ................................................................................... 75
   Solving the Likelihood Equation ....................................................................... 76
   Finding the Initial Guesses ............................................................................... 77
   Asymptotic Distributions of Estimators ........................................................... 81
      Normal Population ....................................................................................... 81
      Non-normal Population .............................................................................. 85

4. TESTING HYPOTHESIS AND BARTLETT CORRECTION .................. 87
   Hypotheses ......................................................................................................... 87
   Likelihood Ratio Test ......................................................................................... 89
   Bartlett Correction ............................................................................................. 92
      Third and Fourth Order Moments ................................................................ 92
      Expectation of LRT Statistic ......................................................................... 98
   Under a Restricted $\mathbf{H}_a$ .............................................................................. 100
   Under an Unrestricted $\mathbf{H}_a$ ........................................................................ 100
   Matrix Expressions for $\mathbf{Z}_i$ and $\mathbf{K}_i$ ...................................................... 101
5. SIMULATION .................................................................................. 108
   Simulation Study of Variable Reduction (Redundancy) Tests .......... 108
   The Comparison Test .................................................................. 109
   Variance Reduction Procedure ................................................... 111
   Discussion ................................................................................. 132
   Simulation Study of Component Reduction Tests ....................... 133
   Numerical Examples .................................................................... 135
   Example 1. (continued) ............................................................... 135
   Example 2. (continued) ............................................................... 136

6. CONCLUSION .............................................................................. 138

REFERENCES ................................................................................. 140

APPENDICES .................................................................................. 144
   APPENDIX A – Notation ............................................................. 145
   APPENDIX B – List of Matlab Subprograms ............................... 147
   APPENDIX C – Programming Codes .......................................... 151
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Eigenvalues of Covariance Matrix: Photographic Example</td>
<td>31</td>
</tr>
<tr>
<td>2. Eigenvectors of Covariance Matrix: Photographic Example</td>
<td>31</td>
</tr>
<tr>
<td>3. Eigenvalues of Covariance Matrix: Women’s Track Example</td>
<td>34</td>
</tr>
<tr>
<td>4. Eigenvectors of Covariance Matrix: Women’s Track Example</td>
<td>34</td>
</tr>
<tr>
<td>5. Probabilities of Type I error when ((p, k, q) = (10, 2, 4)) and (\lambda = (1^t, 10, 25)_s^{(1)})</td>
<td>116</td>
</tr>
<tr>
<td>6. Probabilities of Type I error when ((p, k, q) = (10, 2, 4)) and (\lambda = (1^t, 5, 10)_s^{(1)})</td>
<td>116</td>
</tr>
<tr>
<td>7. Probabilities of Type I error when ((p, k, q) = (10, 2, 4)) and (\lambda = (1^t, 2, 10)_s^{(1)})</td>
<td>117</td>
</tr>
<tr>
<td>8. Probabilities of Type I error when ((p, k, q) = (10, 2, 4)) and (\lambda = (1^t, 1.5, 10)_s^{(1)})</td>
<td>118</td>
</tr>
<tr>
<td>9. Probabilities of Type I error when ((p, k, q) = (10, 2, 1)) and (\lambda = (1^t, 10, 25)_s^{(1)})</td>
<td>118</td>
</tr>
<tr>
<td>10. Probabilities of Type I error when ((p, k, q) = (10, 2, 1)) and (\lambda = (1^t, 5, 10)_s^{(1)})</td>
<td>119</td>
</tr>
<tr>
<td>11. Probabilities of Type I error when ((p, k, q) = (10, 2, 1)) and (\lambda = (1^t, 1.5, 10)_s^{(1)})</td>
<td>119</td>
</tr>
<tr>
<td>12. Probabilities of Type I error when ((p, k, q) = (10, 4, 2)) and (\lambda = (1^t, 10, 10, 25)_s^{(1)})</td>
<td>120</td>
</tr>
<tr>
<td>13. Probabilities of Type I error when ((p, k, q) = (10, 4, 2)) and (\lambda = (1^t, 5, 5, 10, 10)_s^{(1)})</td>
<td>120</td>
</tr>
<tr>
<td>14. Probabilities of Type I error when ((p, k, q) = (10, 4, 2)) and (\lambda = (1^t, 2, 5, 10, 10)_s^{(1)})</td>
<td>121</td>
</tr>
</tbody>
</table>
16. Probabilities of Type I error when \((p, k, q) = (10, 4, 2)\) and \(\lambda = \left(1_1', 5, 5, 10, 10\right)'\) ................................................................. 121

17. Probabilities of Type I error when \((p, k, q) = (10, 4, 5)\) and \(\lambda = \left(1_1', 10, 10, 10, 25\right)'\) ................................................................. 122

18. Probabilities of Type I error when \((p, k, q) = (10, 4, 5)\) and \(\lambda = \left(1_1', 5, 5, 10, 10\right)'\) ................................................................. 122

19. Probabilities of Type I error when \((p, k, q) = (10, 4, 5)\) and \(\lambda = \left(1_1', 2, 5, 10, 10\right)'\) ................................................................. 123

20. Probabilities of Type I error when \((p, k, q) = (10, 4, 5)\) and \(\lambda = \left(1_1', 10, 10, 10, 10\right)'\) ................................................................. 123

21. Probabilities of Type I error when \((p, k, q) = (15, 2, 1)\) and \(\lambda = \left(1_1', 15, 35\right)'\) ................................................................. 124

22. Probabilities of Type I error when \((p, k, q) = (15, 2, 1)\) and \(\lambda = \left(1_1', 10, 15\right)'\) ................................................................. 124

23. Probabilities of Type I error when \((p, k, q) = (15, 2, 1)\) and \(\lambda = \left(1_1', 2, 15\right)'\) ................................................................. 125

24. Probabilities of Type I error when \((p, k, q) = (15, 2, 1)\) and \(\lambda = \left(1_1', 15, 15\right)'\) ................................................................. 125

25. Probabilities of Type I error when \((p, k, q) = (15, 4, 2)\) and \(\lambda = \left(1_1', 15, 15, 15, 35\right)'\) ................................................................. 126

26. Probabilities of Type I error when \((p, k, q) = (15, 4, 2)\) and \(\lambda = \left(1_1', 10, 10, 15, 15\right)'\) ................................................................. 126

27. Probabilities of Type I error when \((p, k, q) = (15, 4, 2)\) and \(\lambda = \left(1_1', 2, 10, 15, 15\right)'\) ................................................................. 127

28. Probabilities of Type I error when \((p, k, q) = (15, 4, 2)\) and \(\lambda = \left(1_1', 15, 10, 15, 15\right)'\) ................................................................. 127

29. Power of test when \(\lambda = \left(1_1', 10, 25\right)'\), \(k = 2\), \(q = 1\) .................................. 128

30. Power of test when \(\lambda = \left(1_1', 2, 10\right)'\), \(k = 2\), \(q = 1\) .................................. 129
31. Power of test when $\lambda = (1'_6, 10, 10, 10, 25)'$, $k = 4$, $q = 2$ .......... 130
32. Power of test when $\lambda = (1'_5, 2, 5, 10, 10)'$, $k = 4$, $q = 2$ ................. 131
33. Percentage coverage of CI when $(p, k) = (8, 2)$ and $\lambda = (1'_6, 10, 25)'$ .... 134
34. Percentage coverage of CI when $(p, k) = (8, 2)$ and $\lambda = (1'_6, 2, 10)'$...... 134
35. Percentage coverage of CI when $(p, k) = (8, 4)$ and $\lambda = (1'_4, 10, 10, 10, 25)'$ 134
36. Percentage coverage of CI when $(p, k) = (8, 4)$ and $\lambda = (1'_4, 2, 10, 10, 10)'$ 134
37. P-value of Testing Hypothesis: Women’s Track Example ............... 136
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Ideal Scree Graph</td>
<td>6</td>
</tr>
<tr>
<td>2.</td>
<td>A Profile Plot of the Expected Mean</td>
<td>33</td>
</tr>
<tr>
<td>3.</td>
<td>Power of test when $\lambda = (1^t_8, 10, 25)^t$, $k = 2$, $q = 1$</td>
<td>128</td>
</tr>
<tr>
<td>4.</td>
<td>Power of test when $\lambda = (1^t_8, 2, 10)^t$, $k = 2$, $q = 1$</td>
<td>129</td>
</tr>
<tr>
<td>5.</td>
<td>Power of test when $\lambda = (1^t_6, 10, 10, 10, 25)^t$, $k = 4$, $q = 2$</td>
<td>130</td>
</tr>
<tr>
<td>6.</td>
<td>Power of test when $\lambda = (1^t_6, 2, 5, 10, 10)^t$, $k = 4$, $q = 2$</td>
<td>131</td>
</tr>
</tbody>
</table>
ABSTRACT

The main objective of this thesis is to develop procedures for making inferences about the eigenvalues and eigenvectors of a covariance matrix. Specifically, new procedures for examining dimension reduction in principal component analysis (PCA) are developed. The dimension reduction consists of the following two aspects: reduction in the number of components and reduction in the number of original variables. The procedures are based on a likelihood approach. Parameterizations of eigenvalues and eigenvectors are presented. The parameterizations allow arbitrary eigenvalue multiplicities. The use of the Fisher scoring algorithm for computing maximum likelihood estimates of the covariance parameters subject to multiplicity and other constraints is discussed. Asymptotic distributions of estimators of covariance parameters are derived under normality and non-normality. Likelihood ratio tests and Bartlett corrections are described. Simulation studies show the effectiveness of Bartlett corrections. The new procedures are demonstrated to give better overall results than some existing methods.
CHAPTER 1

INTRODUCTION

The main objective of this thesis is to develop procedures for making inferences about the eigenvalues and eigenvectors of a covariance matrix. Specifically, new procedures for examining dimension reduction in principal component analysis (PCA) are developed. The procedures are based on a likelihood approach.

The first chapter of this thesis briefly describes principal component analysis followed by a discussion of dimensionality reduction methods. Two examples are given. In Chapter 2, parameterizations of eigenvalues and eigenvectors are presented. Two eigenvalue parameterizations are adopted under different objectives of use. The parameterizations allow arbitrary eigenvalue multiplicities. The Newton-Raphson algorithm is used to solve implicit parameter equations. The hypothesis test of redundancy is described and an eigenvector parameterization under redundancy is proposed.

Chapter 3 describes the use of the Fisher scoring algorithm for computing maximum likelihood estimates of the covariance parameters subject to multiplicity and other constraints. Asymptotic distributions of estimators of covariance parameters are derived under normality and non-normality. In Chapter 4, likelihood ratio tests and Bartlett corrections are described. The results of simulations that examine the
effectiveness of Bartlett corrections are presented in Chapter 5. Two numerical examples also are illustrated.

**Principal Component Analysis**

Investigators often measure or make observations on a large number of variables. There are several useful techniques to reduce the dimensionality of data without the loss of much information such as factor analysis and cluster analysis. Principal component analysis also is one such technique and is one of the most widely-used multivariate techniques.

Typically, principal component analysis is used to reduce the dimensionality of a data set, while retaining as much of the original information as possible. This is achieved by transforming the original set of variables into a smaller set of linear combinations called principal components. These new variables are uncorrelated and ordered so that the first principal component accounts for the largest proportion of the variation present in the original set of variables. The usual objective of the analysis is to determine whether the first few components account for most of the variation in the original data. If they do, then these components can be used to summarize the data with little loss of information. This dimension reduction can be useful in simplifying subsequent analyses.

Before defining principal components, some notation will be described. Matrices will be denoted by boldface upper case letters. Vectors will be denoted by boldface
lower case letters. A diagonal matrix having diagonal elements $a_1, a_2, \ldots, a_n$ is denoted by $\text{diag}(a_1, a_2, \ldots, a_n)$ and the trace of matrix $A$ is denoted by $\text{tr}(A)$. The $p \times p$ identity matrix is denoted by $I_p$.

Suppose that the vector of original variables $y = (y_1, y_2, \ldots, y_p)'$ has a positive definite covariance matrix $\Sigma$. For mathematical convenience and without loss of generality, assume that the mean of $y_i$ is zero for all $i = 1, 2, \ldots, p$. Because $\Sigma$ is symmetric and positive definite, all eigenvalues of $\Sigma$ are real and positive. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$ be the ordered eigenvalues of $\Sigma$ and let $\Gamma = (\gamma_1 \gamma_2 \cdots \gamma_p)$ be a $p \times p$ orthogonal matrix such that

$$\Sigma = \Gamma \Lambda \Gamma',$$  \hspace{1cm} (1.1)

where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$. That is, $\gamma_i$ is an eigenvector of $\Sigma$ corresponding to the eigenvalue $\lambda_i$. The principal components of $y$ are the entries of the $p$-vector $z$, where $z$ is the linear combination that can be written as

$$z = \Gamma'y.$$ \hspace{1cm} (1.2)

The covariance matrix of $z$, $\Sigma_z$, can be written as

$$\Sigma_z = \Gamma'\Sigma \Gamma.$$  

Substituting $\Sigma$ from (1.1) yields

$$\Sigma_z = \Gamma' (\Gamma \Lambda \Gamma') \Gamma = \Lambda.$$
Hence the components \( z_1 = \gamma_1 y, z_2 = \gamma_2 y, \ldots, z_p = \gamma_p y \) are uncorrelated. The variance of \( z_i \) is \( \lambda_i \). The maximum value of variance of \( \gamma_i' y \) satisfying \( \gamma_i' \gamma_1 = 1 \) is equal to \( \lambda_1 \), the largest eigenvalue of \( \Sigma \). This maximum occurs when \( \gamma_1 \) is an eigenvector of \( \Sigma \) corresponding to \( \lambda_1 \). That is,

\[
\max_{i' \in 1} l' \Sigma l = \gamma_1' \Sigma \gamma_1 = \lambda_1.
\]

Then, the component \( z_1 \) has the largest variance \( \lambda_1 \), \( z_2 \) has the second largest variance \( \lambda_2 \), and so on. The total variance of the \( p \) principal components is equal to the total variance of the original variables so that

\[
\sum_{i=1}^{p} \lambda_i = \text{tr}(\Sigma).
\]

Consequently, the \( j^{th} \) principal component accounts for a proportion

\[
t = \frac{\lambda_j}{\text{tr}(\Sigma)}
\]

of the total variance of the original variables.

This thesis focuses on two major aspects of dimensionality reduction:

1. Reduction in the number of components.
2. Reduction in the number of original variables.

**Reduction in the Number of Components**

The major problem in reducing the number of components is deciding how many principal components should be retained in order to account for most of the variability
in the original data. Some rules of thumb have been suggested. Kaiser (1958) suggested a cut off point to retain the components whose sample variances are greater than the average, i.e. greater than 1 for a correlation matrix. It can be argued that if one variable is nearly independent of all others variables, it will dominate a component with variance slightly less than 1 assuming that variables have been standardized. Since this variable provides independent information from the other variables, there is no reason to discard it. Kaiser’s rule is arbitrary and tends to retain too few components (Jolliffe, 1986). Rencher (1995) suggested that investigators should retain sufficient components to account for a specified percentage of the total variance. Figures between 70 and 90 percent are generally suggested.

Three methods are in common use for determining the number of components to be retained. The first method is graphical and is called a scree graph. This technique was discussed and named by Cattell (1966). The scree graph consists of a plot of the ordered eigenvalues against eigenvalue numbers. With this plot, the components to be retained correspond to the eigenvalues plotted in the steep curve above the straight line formed by the smaller eigenvalues. Thus in Figure 1, the first two components would be retained. Jolliffe (1986) suggested plotting \( \log(\lambda) \) rather than \( \lambda \). This plot is known as a log-eigenvalue diagram.

The second method is testing the hypothesis that the last \( m = p - k \) eigenvalues of \( \Sigma \) are equal. Anderson (1963) derived the likelihood ratio test for \( H_0_k: \lambda_{k+1} = \cdots = \lambda_p = \lambda \) against the alternative that at least two of the last \( m \) eigenvalues are different.
Let $y_1, y_2, \ldots, y_N$ be a random sample of size $N = n + 1$ from a $p$-variate normal distribution with mean $\mu$ and covariance matrix $\Sigma$. If $\mathbf{y}$ can be written as $\mathbf{Cw} + \varepsilon$, where $\mathbf{C}$ is a $p \times k$ matrix with rank $k$, $\mathbb{E}(\varepsilon) = 0$, $\text{var}(\varepsilon) = \sigma^2 I_p$, and $\text{var}(\mathbf{w}) = I_k$, then $\text{var}(\mathbf{y}) = \Sigma = \mathbf{CC}^t + \sigma^2 \mathbf{I}_p$. Write $\mathbf{CC}^t$ in diagonal form as $\mathbf{CC}^t = \Gamma_1 \mathbf{\Lambda}_1 \Gamma_1^t$ and let $\Gamma_2$ be an orthogonal complement to $\Gamma_1$. That is, $\mathbf{I}_p = \Gamma_1 \Gamma_1^t + \Gamma_2 \Gamma_2^t$ and $(\Gamma_1 \quad \Gamma_2) \in \mathcal{O}(p)$, where $\mathcal{O}(p)$ denotes group of orthogonal $p \times p$ matrices. The covariance matrix of $\mathbf{y}$ can be rewritten as

$$
\Sigma = \Gamma_1 \mathbf{\Lambda}_1 \Gamma_1^t + \sigma^2 \mathbf{I}_p \\
= \Gamma_1 \mathbf{\Lambda}_1 \Gamma_1^t + \sigma^2 (\Gamma_1 \Gamma_1^t + \Gamma_2 \Gamma_2^t) \\
= \Gamma_1 (\mathbf{\Lambda}_1 + \sigma^2 \mathbf{I}) \Gamma_1^t + \Gamma_2 \sigma^2 \Gamma_2^t.
$$
Therefore, the first \( k \) eigenvalues are \( \lambda_i + \sigma^2, \ i = 1, 2, \ldots, k \), where \( \Lambda_1 = \text{diag}(\lambda_i), \ i = 1, 2, \ldots, k \), and the last \( m = p - k \) eigenvalues are \( \sigma^2 \). Anderson derived the test for testing the equality of the last \( m \) eigenvalues. The test statistic is

\[
\Lambda_k = \left\{ \prod_{i=k+1}^{p} \frac{l_i}{\left( \sum_{i=k+1}^{p} \frac{l_i}{m} \right)^m} \right\}^{n/2},
\]

where \( l_i, \ i = 1, 2, \ldots, p \), are eigenvalues of the sample covariance matrix, \( S \). If \( H_{ok} \) is true, then the limiting distribution of \( -2 \ln \Lambda_k \) is \( \chi^2_{\nu} \), where \( \nu = \frac{1}{2}(m + 2)(m - 1) \).

These results can be summarized as

\[
T_1 = n \left[ m \ln \bar{l} - \sum_{i=k+1}^{p} \ln l_i \right] \sim \chi^2_{\nu} + O_p(n^{-\frac{1}{2}}),
\]

where \( \bar{l} = \sum_{i=k+1}^{p} l_i/m \).

Bartlett (1954) improved Anderson’s test by using a multiplying factor. Bartlett’s test is to reject \( H_{ok} \) whenever

\[
T_2 = \left( n - k - \frac{2m^2 + m + 2}{6m} \right) \left[ m \ln \bar{l} - \sum_{i=k+1}^{p} \ln l_i \right] \geq \chi^2_{1-\alpha,\nu},
\]

where \( \chi^2_{1-\alpha,\nu} \) is the 100(1-\( \alpha \)) percentile of the \( \chi^2_{\nu} \) distribution.

Lawley (1956) proposed a further improvement in Bartlett’s multiplying factor. Lawley’s test is to reject \( H_{ok} \) whenever

\[
T_3 = \left( n - k - \frac{2m^2 + m + 2}{6m} + \sum_{i=1}^{k} \frac{l_i^2}{(l_i - \bar{l})} \right) \left[ m \ln \bar{l} - \sum_{i=k+1}^{p} \ln l_i \right] \geq \chi^2_{1-\alpha,\nu}.
\]

Acceptance of \( H_{ok} \) suggests that each of the last \( m \) components contain the same amount of information. If these eigenvalues are very small, then little information is
lost by discarding the corresponding principal components. The first $k$ components should be retained for further analysis. A sequence of tests can be conducted starting with $k = 0$ and increasing $k$ until the null hypothesis is accepted.

Bentler and Yuan (1996) developed a method of testing the linear trend in the last $m = p - k$ eigenvalues of the covariance matrix. The hypothesis of interest is $H_{0m}: \lambda_{k+i} = \alpha + \beta x_i$, $x_i = m - i$, $i = 1, 2, \ldots, m$. The likelihood ratio test statistic is

$$
\Lambda_m = \frac{\operatorname{Sup}_{\mu, \Sigma} \ell(\mu, \Sigma)}{\operatorname{Sup}_{\mu, \Sigma_0} \ell(\mu, \Sigma)},
$$

where $\Sigma_0 = \Gamma_0 \Lambda_0 \Gamma_0'$ and $\Lambda_0 = \operatorname{diag}(\lambda_1, \ldots, \lambda_k, \alpha + \beta x_1, \alpha + \beta x_2, \ldots, \alpha + \beta x_m)$. For $\Sigma_0$ and $\Sigma$, $\ell(\mu, \Sigma)$ is maximized when $\hat{\mu} = \hat{\gamma}$. To maximize $\ell(\hat{\gamma}, \Sigma_0)$, Bentler and Yuan showed that $\hat{\lambda}_i = \frac{n + i}{n + 1} l_i$ for $i = 1, 2, \ldots, k$ are the maximum likelihood estimators (MLEs) of $\lambda_1, \ldots, \lambda_k$ under $H_{0m}$. Maximum likelihood estimates of $\alpha$ and $\beta$ were numerically obtained by solving the following equations:

$$
\sum_{j=1}^{m} \frac{n l_{k+j} - (n + 1)(\alpha + \beta x_j)}{(\alpha + \beta x_j)^2} = 0, \\
\sum_{j=1}^{m} \frac{n l_{k+j} - (n + 1)(\alpha + \beta x_j) x_j}{(\alpha + \beta x_j)^2} = 0.
$$

The asymptotic distribution of $-2 \ln \Lambda_m$ is $\chi^2$ with $\nu = m - 2$. If all $x_i$ are equal to zero, then the hypothesis $H_{0m}$ is equivalent to $H_{0k}$ as considered by Anderson (1963).

Boik (2002a) proposed a method for modeling the eigen-structure of several covariance matrices simultaneously. The proposed model for the $i^{th}$ covariance matrix
is as follows:

\[ \Sigma_i = \Gamma_i \Lambda_i \Gamma_i', \quad i = 1, 2, \ldots, g, \]

where the eigenvector matrices \( \{\Gamma_i\} \) and eigenvalue matrices \( \{\Lambda_i\} \) may share certain properties. The model extends common principal components (Flury, 1988) and subsumes Anderson’s (1963) model, Lawley’s (1956) model and Bentler and Yuan’s (1996) model as special cases when \( g=1 \). Boik presented an algorithm to compute maximum likelihood estimates of covariance parameters and also gave likelihood ratio tests including Bartlett corrections. Several of his results are referred to throughout this thesis.

The third method is based on the cumulative proportion of total variance. If \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \) are the ordered eigenvalues, then the cumulative proportions of the first \( k \) eigenvalues, \( \delta_k \), and the last \( m = p - k \) eigenvalues, \( \delta_m \), are as follows:

\[
\delta_k = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{p} \lambda_i} \quad \text{and} \quad \delta_m = 1 - \delta_k = \frac{\sum_{i=k+1}^{p} \lambda_i}{\sum_{i=1}^{p} \lambda_i}.
\]

Let \( \hat{\delta}_k = \frac{\sum_{i=1}^{k} l_i}{\sum_{i=1}^{p} l_i} \) and \( \hat{\delta}_m = \frac{\sum_{i=k+1}^{p} l_i}{\sum_{i=1}^{p} l_i} \), where \( l_i, \ i = 1, 2, \ldots, p, \) are defined in (1.3).

Anderson (1963) proposed a test statistic for the hypothesis,

\[ H_{0m}: \delta_m = \delta_{m,0}, \]  

(1.4)

where \( \delta_{m,0} \) is a known constant. If \( \delta_m \) is small, then little information is lost by discarding the corresponding principal components. Anderson showed that if \( \lambda_k >
If \( \lambda_{k+1} \), then \( \delta_m \) is asymptotically distributed as

\[
\sqrt{n}(\delta_m - \delta_m) \overset{d}{\to} N\left(0, \frac{2\delta_m^2 \sum_{i=1}^{k} \lambda_i^2 + 2(1 - \delta_m)^2 \sum_{i=k+1}^{p} \lambda_i^2}{(\sum_{i=1}^{n} \lambda_i)^2}\right).
\] (1.5)

The test rejects \( H_{0m} \) whenever \(|z^*| \geq z_{\alpha/2}\), where

\[
z^* = \frac{\delta_m - \delta_{m,0}}{\sqrt{\hat{V}(\delta_m)}},
\] (1.6)

\( z_{\alpha/2} \) is the upper \( 100\alpha/2 \) percentage point of the standard normal distribution, and \( \hat{V}(\delta_m) \) is an estimator of the variance of \( \delta_m \) which can be computed by replacing \( \lambda_i \) by \( l_i \) in the variance of \( \hat{\delta}_m \) in (1.5).

Sugiyama and Tong (1976) studied an approximate distribution of the cumulative proportion of the first \( k \) eigenvalues, \( \delta_k \). Sugiyama discussed that this quantity is interpreted as a measure of the amount of information in the \( k \) retained components. Assuming that the eigenvalues are distinct, he derived a perturbation expansion for the distribution of \( \hat{\delta}_k \) that is accurate to \( O(n^{-3/2}) \).

Huang and Tseng (1992) devised the following method of selecting the number of components to be retained. Let \( \sum_{i=1}^{k} \lambda_i / \sum_{i=1}^{p} \lambda_i = \delta_k \) and

\[\Omega_q(\omega) = \{ \lambda; \delta_q \geq \omega \},\]

where \( \omega \) is a fixed constant and \( q \) is an integer less than \( p \). Their decision rule is to select \( k \) components, where \( k \) is the smallest integer that satisfies

\[
\hat{\delta}_k \geq c,
\]
where \( c \) is a fixed constant to be determined. They discussed how to determine \( N \) and \( c \) such that the minimum probability of retaining the important components is at least some specified level \( P^* \), \( 0 \leq P^* \leq 1 \). That is,

\[
\min_k \left\{ \inf_{\lambda \in \Omega_k(\omega)} P_k(\delta_k \geq c) \right\} \geq P^*. \tag{1.7}
\]

Huang and Tseng showed that if \( \omega \geq 0.5 \), \( k \leq \frac{n}{2} \), and \( N \) and \( c \) satisfy

\[
c \leq \omega \{1 - (1 - \omega)\sqrt{2p/(p-1)(N-1)}\Phi^{-1}(P^*)\},
\]

where \( \Phi(x) \) is the c.d.f of the standard normal distribution, then (1.7) holds.

In this thesis, a likelihood approach for making inferences about the cumulative proportion of total variance is given. The hypothesis of interest is

\[
H_0: \frac{C_1^t \Lambda}{\text{tr}(\Sigma)} = c_0, \tag{1.8}
\]

where \( C_1 \) is a matrix of known constants and \( c_0 \) is a vector of known constants.

This hypothesis is more flexible and general than the hypothesis in (1.4). If \( C_1 \) is \((0_k', 1_{p-k}')'\) and \( c_0 \) is a scalar, then the hypothesis in (1.4) is subsumed as a special case of the hypothesis in (1.8). Specific goals are to obtain maximum likelihood estimates (MLEs) of covariance parameters, to construct a likelihood ratio test of the hypothesis, and to construct confidence intervals for \( \frac{C_1^t \Lambda}{\text{tr}(\Sigma)} \). The MLEs of parameters and a likelihood ratio test for the hypothesis in (1.8) are given using one of the parameterizations of eigenvalues presented in Chapter 2. The parameterizations allow for arbitrary eigenvalues multiplicities. A Bartlett correction is derived to improve the
likelihood ratio test. The tests are inverted to obtained confidence intervals. The like-
lihood ratio intervals and Bartlett-corrected intervals are compared with Anderson’s intervals inverted from (1.6) by means of simulations. The results of the simulations are described in Chapter 5.

**Reduction in the Number of Original Variables**

A principal component is a linear combination of all of the original variables. Thus, interpretation and subsequent data analysis involve all of the variables even if some components are discarded. Not only is it useful to reduce the number of components but it also is useful to reduce the number of the original variables. If the first $k$ principal components have zero coefficients on a set of original variables, then those variables are redundant. These redundant variables can be eliminated from the study and need not be collected in any subsequent studies. The term “redundancy” has been employed in other contexts, e.g. multivariate regression (Lazraq, 2001), discriminant analysis (Fujikoshi, 1985), canonical correlation (Van Den Wollenberg, 1977; Gleason, 1976), and growth curve analysis (Fujikoshi and Rao, 1991). The term takes a different meaning in each of these contexts. In this thesis, the principal components of $y$ are defined as in (1.2). It follows that

$$y = \Gamma z.$$  

The coefficient for obtaining the $i^{th}$ variable from the $j^{th}$ principal component is $\gamma_{ji}$ and is called a loading. A redundant variable is defined to be an original variable
whose loadings on the first $k$ principal components are zero.

Two approaches for determining the number of variables and which variables to retain will be discussed, namely the rejection approach and the hypothesis testing approach. A rejection approach is a set of cut-off rules for choosing how many and which variables to be retained. A hypothesis test is a method of inference to decide which of two complementary hypotheses is true. The rejection approach will be described first.

Beale, Kendall and Mann (1967) developed cut-off rules that specify the number of variables to be discarded. However, they focused primarily on multiple regression analysis rather than on principal component analysis. In multiple regression with $p$ independent variables, they suggested selecting the set of $r$ variables that maximizes the multiple correlation of the dependent variable with the $r$ independent variables. An extension of this rule to principal components is to retain set of $r$ variables which maximizes the minimum multiple correlation between $r$ selected variables and any of $p - r$ discarded variables. They also mentioned another rejection approach for use in principal component analysis. With all $p$ variables, $p$ principal components are obtained. If $p_1$ eigenvalues are smaller than some number, $\lambda_0$, then the last $p_1$ components are inspected. Starting with the one corresponding to the smallest eigenvalue, then the next component corresponding to the second smallest eigenvalue and so forth, the variable that has the largest coefficient in the component and has not already been deleted by a previously considered component is deleted and the
number of variables is reduced from \( p \) to \( p - p_1 \). Another principal component analysis is done on the remaining \( p - p_1 \) variables. Similarly, if \( p_2 \) eigenvalues are smaller than the \( \lambda_0 \), \( p_2 \) variables associated with the largest coefficients in the last \( p_2 \) components are rejected. The process continues until all eigenvalues are greater than the \( \lambda_0 \). The number of variables is reduced from \( p \) to 
\[
r = (p - p_1 - \cdots - p_k).
\]
The value of \( r \) is determined by choice of \( \lambda_0 \). Jolliffe (1972) discussed the appropriate \( \lambda_0 \) based on simulation studies.

Jolliffe (1972) discussed eight rejection methods for deciding which variables to be discarded. The eight rejection methods were divided into three groups: namely multiple correlation methods (A1 and A2), principal component methods (B1, B2, B3, and B4), and cluster methods (C1 and C2). These methods are briefly described below.

- **Multiple Correlation Methods**
  - Method A1 is the method of Beale, Kendall and Mann (1967).
  - Method A2 is a stepwise method. In multiple regression with \( p \) independent variables, a subset of independent variables is selected, in each step, such that the variable having maximum multiple correlation with the remaining variables is deleted. The process is repeated until \( r \) variables are retained.

- **Principal Component Methods**
  - Method B1 also is based on Beale, Kendall and Mann (1967) using PCA.
– Method B2 is the same as Method B1 except that only one principal component analysis is done. If \( k \) variables are to be retained, then the last \( p-k \) components are inspected. For each of these \( p-k \) components, starting with the last component, the variable that has the largest coefficient in the component and has not already been deleted by a previously considered component is deleted.

– Method B3 uses the last \( p-k \) components. The sum of squares of coefficients of each of the \( p \) variables in the last \( p-k \) components are computed. The \( p-k \) variables that have the largest sum of squares are rejected.

– Method B4 uses the first \( k \) components. For each of \( k \) components, starting with the first component, the variable that has the largest coefficient in the component and has not already been selected by a previously considered component is selected. Variables that are not selected are rejected.

• Cluster Methods

– Method C1 is a single-linkage method (Seber, 1984). A measure of similarity between two clusters of variables \( X \) and \( Y \) is defined by \( r_{xy} \) such that

\[
r_{xy} = \max r_{ij},
\]

where \( r_{ij} \) is the correlation coefficient between variable \( i \) and \( j \). For each of \( p(p-1)/2 \) pairs of \( p \) variables, \( r_{xy} \) are computed. Two clusters having the maximum \( r_{xy} \) are combined into a single cluster. For the new set of
clusters, the process returns to calculate $r_{xy}$ and so forth. The process continues until $k$ clusters of variables remain. The principal component analysis is based on $k$ variables, one chosen from each cluster. Jolliffe described 3 ways to select a variable from each cluster.

- Method C2 is an average-linkage method (Seber, 1984). It follows the same steps as Method C1 except that $r_{xy}$ in (1.9) is replaced with

$$r_{xy} = \frac{\sum_{i \in X} \sum_{j \in Y} r_{ij}}{p_1 p_2},$$

where $p_1, p_2$ are the numbers of variables in $X$ and $Y$, respectively.

These rejection methods were tested on artificial data. The artificial data were constructed so that certain variables were linear combination of other variables. For example, if $y_1 = y_2 + e$, where $e$ is a random disturbance, then either $y_1$ or $y_2$ may be discarded without loss of information. It was shown that several rejection methods eliminated precisely those certain variables. In general, Method A2 is the most consistent method because it only retained good or best subsets. Method B4 retained best subsets more often than Method A2 but it retained moderate or bad sets fairly frequently.

Jolliffe (1973) discussed the five methods, A1, B2, B4, C1, and C2 which had been successfully used on artificial data in Jolliffe (1972). He applied these methods on four sets of real data. The methods were equally effective with real and artificial data, although none of rejection methods was apparently better than the others.

McCabe (1984) proposed methods to select a subset of variables called principal
variables that contains as much information as possible. McCabe gave four optimality criteria to evaluate all possible subsets. These criteria are based on the conditional covariance matrix of variables not selected, given those selected. Let $\Sigma$ be partitioned as

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

where $\Sigma_{11}$ is the covariance matrix of the selected variables and $\Sigma_{22}$ is the covariance matrix of the discarded variables. The conditional covariance matrix of the variables not selected given those selected can be written as

$$\Sigma_{221} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}.$$

Let $\theta_i, \ i = 1, 2, \ldots, p - k$, be the eigenvalues of $\Sigma_{221}$ and $\lambda_i, \ i = 1, 2, \ldots, p$, be the eigenvalues of $\Sigma$. McCabe proposed the following four criteria:

1. $\min |\Sigma_{221}| = \min \prod_{i=1}^{p-k} \theta_i,$

2. $\min \text{tr}(\Sigma_{221}) = \min \sum_{i=1}^{p-k} \theta_i,$

3. $\min \|\Sigma_{221}\|^2 = \min \sum_{i=1}^{p-k} \theta_i^2,$

4. $\max \sum_{i=1}^{k} \rho_i^2 = \max \text{tr}(\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}),$

where $\rho_i$ are canonical correlations between the variables not selected and those selected. The percentage of variance explained by a set of $k$ principal variables is

$$P = \left(1 - \frac{\sum_{i=1}^{p-k} \theta_i}{\sum_{i=1}^{p} \lambda_i}\right)100\%.$$
This percentage can be compared to the variation explained by the first few principal components.

Jolliffe and Cadima (1995) showed that discarding variables that have a small loading is not appropriate in various respects. They examined the effects on a single component if one or more variables are eliminated. They did not examine the effects on a subspace spanned by several components simultaneously. Jolliffe and Cadima implied that the correlation between the $i^{th}$ variable and the $j^{th}$ principal component, $\hat{\rho}_{ij}$, is an appropriate quantity to examine. This quantity can be written as

$$\hat{\rho}_{ij} = \frac{\hat{\gamma}_{ij}}{\sqrt{l_j}}$$

where $\hat{\gamma}_{ij}$ is loading for the $i^{th}$ variable in the $j^{th}$ component, $l_j$ is the eigenvalue associated with that component, and $s_i^2$ is the variance of the $i^{th}$ variable. They used an example to show that similar coefficients, $\hat{\gamma}_{ij}$, even very large ones, may translate into very different correlations between those variables and principal components or very different coefficients may be associated with similarly correlated variables and principal components. Jolliffe and Cadima examined the distance between the principal component and the truncated principal component. The truncated principal component is the principal component in which variables with small loading are ignored. The distance between the principal component and the truncated principal component can be expressed as

$$d = \left( \frac{\|X\hat{\gamma}_i - X_k\hat{\gamma}_k^k\|}{\|X\hat{\gamma}_i\|} \right)^{1/2} = \left( 1 - 2\hat{\gamma}_{ij}^2 \hat{\gamma}_{ij}^k + \frac{\hat{\gamma}_{ij}^k S_k \hat{\gamma}_{ij}^k}{l_j} \right)^{1/2},$$
where $\hat{\gamma}_j^k$ is the sub-vector of $\hat{\gamma}_j$ which results from retaining only the $k$ coefficients associated with the variables that were retained, $S_k$ is the sample covariance of the $k$ selected variables, and $X_k$ is the sub-matrix of a $n \times p$ column-centered data matrix, $X$, obtained by retaining only $k$ of its columns. The last term of $d$ shows that selecting the $k$ variables with the largest loadings is not guaranteed to yield the smallest distance.

The second approach for determining the number of variables is hypothesis testing. Anderson (1963) gave the asymptotic distribution of the sample eigenvalues and eigenvectors when the population eigenvalues are equal in sets. Anderson considered the hypothesis $H_0: \gamma_1 = \gamma_0$, where $\gamma_1$ is the eigenvector of the first principal component and $\gamma_0$ is a specified vector such that $\gamma_0^T\gamma_0 = 1$. If the hypothesis is true, then

$$n[l_1\gamma_0^T S^{-1} \gamma_0 + l_1^{-1}\gamma_0^T S\gamma_0 - 2]$$

has the limiting Chi-square distribution with degree of freedom $p - 1$, where $l_1$ is the first eigenvalue of the sample covariance matrix, $S$. Similar results can be applied for any other eigenvector corresponding to an eigenvalue with multiplicity 1.

Tyler (1981) modified and extended Anderson’s results. Tyler derived asymptotic procedures for testing more general hypotheses concerning eigenvectors. Let $H$ be a $p \times p$ matrix and $W$ be a $p \times p$ positive definite symmetric matrix such that $WH$ is symmetric. These conditions ensure that the eigenvalues and eigenvectors of $H$ are real. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ be the ordered eigenvalues of $H$ and let $A$ be a
\( p \times q \) matrix with rank \( q \). Lastly, let \( w \) be a subset of \( m \) integers \((1 \leq m < p)\) from \( \{1, 2, \ldots, p\} \) such that \( \lambda_i \neq \lambda_j \) for all \( i \in w, j \notin w \). Tyler considered the following two hypotheses:

for \( q \leq m \),

\[
H_0: \quad \text{The columns of } A \text{ lie in the subspace generated by the set of eigenvectors of } H \text{ associated with } m \text{ eigenvalues, } \lambda_i, \ i \in w, \text{ and}
\]

for \( q \geq m \),

\[
H_0^*: \quad \text{The eigenvectors of } H \text{ associated with the } m \text{ eigenvalues, } \lambda_i, \ i \in w, \lie in the subspace generated by the columns of } A.
\]

Let \( \beta_1, \beta_2, \ldots, \beta_p \) be the eigenvectors of \( H \) and let \( b_1, b_2, \ldots, b_p \) be the eigenvectors of \( H_n \), where

\[
\sqrt{n \text{ vec}(H_n - H)} \overset{L}{\longrightarrow} N(0, \Xi).
\]

Denote the \( p \times m \) matrix whose columns are \( \beta_i \) for \( i \in w \) by \( \beta_w \) and denote the \( p \times m \) matrix whose columns are \( b_i \) for \( i \in w \) by \( B_w \). Let \( d_1 \geq d_2 \geq \cdots \geq d_p \) be the ordered eigenvalues of \( H_n \), \( P_0 = \beta_w (\beta'_w \beta_w)^{-1} \beta'_w \), and \( \hat{P}_0 = \sum_{i \in w} \hat{P}_i = B_w (B'_w B_w)^{-1} B'_w \).

Then the test statistic for \( H_0 \) is

\[
T_n(A) = n \left\{ \text{vec} \left[ (I_p - \hat{P}_0) A \right] \right\}' \left[ \hat{\Xi}(A) \right]^{-1} \text{vec} \left[ (I_p - \hat{P}_0) A \right], \quad (1.10)
\]
where

\[ \hat{\Xi}(A) = (A' \otimes I_p)C^t_w \Xi C_w (A \otimes I_p), \]

\[ C_w = \sum_{i \in w} \sum_{j \notin w} (d_i - d_j)^{-1} \hat{P}_i \otimes \hat{P}_j'. \]

Tyler showed that, under $H_0$, $T_n(A) \xrightarrow{L} \chi^2_{(p-m)q}$. The test is to reject $H_0$ if $r = \text{rank}(\hat{P}_0A) < q$ or if $r = q$ and $T_n(A) > \chi^2_{1-\alpha,(p-m)q}$. The hypothesis $H_0^*$ can be tested by using the following approach. Let $B$ be a fixed $p \times (p-q)$ matrix with rank $p-q$ whose columns are orthogonal to $A$, i.e. $A'B = 0$. The hypothesis $H_0^*$ can be rephrased as

\[ H_0 : \text{ The columns of } B \text{ lie in the subspace generated by the set of eigenvectors of } H' \text{ associated with } p - m \text{ eigenvalues, } \lambda_i, \ i \notin w, \]

and can be treated as $H_0$.

The test statistic can be simplified if a random sample has been drawn from an elliptical distribution. Denote the covariance matrix by $H$. Let $\hat{\kappa}$ be an estimator of $\kappa$, the kurtosis parameter. An estimator of the covariance matrix of vec($H_n$) can be written as

\[ \hat{\Xi} = (1 + \hat{\kappa})(I_p \otimes I_{(p,p)}) (H_n \otimes H_n) + \hat{\kappa} \text{vec}(H_n)[\text{vec}(H_n)]'. \]

The test statistic in (1.10) can be simplified as

\[ T_n(A) = n(1 + \hat{\kappa}) \sum_{j \notin w} d_j^{-1} \text{tr}\left\{ A' \hat{P}_i A [A' B_w D_j B_w A]^{-1} \right\}, \quad (1.11) \]
where $D_j$ is an $m \times m$ diagonal matrix with entries $d_i/(d_i - d_j)^2$, $j \notin w$. For $q = m = 1$, $T_n(A)$ is asymptotically equivalent to the statistic given by Anderson (1963).

Flury (1986) discussed the test for redundancy of variables in the comparison of two covariance matrices. Let $\Sigma_1$ and $\Sigma_2$ be the covariance matrices of two independent $p$-vectors. Flury analyzed the eigenvectors of $\Sigma_1^{-1}\Sigma_2$. Let $\{\beta_j\}_{j=1}^p$ be a set of eigenvectors of $\Sigma_1^{-1}\Sigma_2$ and partition $\{\beta_j\}_{j=1}^p$ into the first $p - q$ and the last $q$ rows as follows:

$$\{\beta_j\}_{j=1}^p = \left\{ \begin{array}{c} \beta_j^1 \\ \beta_j^2 \end{array} \right\}_{j=1}^p.$$

The hypothesis of interest is

$$H_0(p, q): \beta_{j2} = 0 \text{ for all } j \in w. \quad (1.12)$$

This hypothesis can be formulated in the form $H_0^*$ of Tyler (1981), by putting $A = (I_{p-q} \ 0)'$.

Then by using

$$B = \begin{pmatrix} 0 \\ I_q \end{pmatrix}, \quad (1.13)$$

the hypothesis can be rephrased as

$$H_0(p, q) : \text{ The columns of } B \text{ lie in the subspace generated by the eigenvector } \alpha_i, \quad (i \notin w) \text{ of } \Sigma_2\Sigma_1^{-1}.$$
The test statistic in (1.10) can be applied. Let \( \{b_j\}_{j=1}^p \) be a set of eigenvectors of \( S_1^{-1}S_2 \) and partition \( \{b_j\}_{j=1}^p \) into the first \( p - q \) and the last \( q \) rows as follows:

\[
\{b_j\}_{j=1}^p = \left\{ \begin{array}{c} b_{j1} \\ b_{j2} \end{array} \right\}_{j=1}^p.
\]

The test statistic is

\[
R(p, q) = n \sum_{j \in w} b_{j2}' \left( \sum_{i \notin w} \frac{k_i d_i^2 + k_2 d_i d_j}{(d_i - d_j)^2} b_{i2} b_{j2}' \right)^{-1} b_{j2},
\]

where \( k_i = \lim_{n_i \to \infty} \frac{n_i}{n_i} \) for \( i = 1, 2 \) and \( d_j, j = 1, 2, \ldots, p, \) are the eigenvalues of \( S_2 S_1^{-1}. \)

Under \( H_0(p, q), R(p, q) \xrightarrow{L} \chi^2_{mq}. \)

Fujikoshi (1989) reviewed the problem of testing the hypothesis of redundancy in various multivariate situations including principal component analysis. Let \( \{\gamma_j\}_{j=1}^p \) be a set of eigenvectors of \( \Sigma \) and partition \( \{\gamma_j\}_{j=1}^p \) into the first \( p - q \) and the last \( q \) rows as follow:

\[
\{\gamma_j\}_{j=1}^p = \left\{ \begin{array}{c} \gamma_{j1} \\ \gamma_{j2} \end{array} \right\}_{j=1}^p.
\]

The hypothesis of redundancy of the last \( q \) variables in the first \( k \) components can be written as \( H_0(k, q): \gamma_{j2} = 0 \) for all \( j \in w, \ w = \{1, 2, \ldots, k\}, \) which is equivalent to

\[
H_0(k, q): \text{ The columns of } B \text{ lie in the subspace generated by the eigenvectors } \\
\gamma_i, (i \notin w) \text{ of } \Sigma,
\]

where \( B \) is defined in (1.13). Fujikoshi used the test statistic in (1.11) to test the hypothesis. Let \( l_i \) and \( \tilde{l}_i, i = 1, 2, \ldots, p, \) be the eigenvalues and eigenvectors of \( S, \)
respectively. The test statistic can be rewritten as

\[ T_n(\mathbf{B}) = n \sum_{j \in w} \tilde{\gamma}_{j2} \left( \sum_{i \in w} \frac{l_i l_j}{(l_i - l_j)^2} \tilde{\gamma}_{i2} \tilde{\gamma}_{i2}^\prime \right)^{-1} \tilde{\gamma}_{j2}, \]  

(1.15)

where under \( H_0(k, q) \), \( T_n(\mathbf{B}) \xrightarrow{L} \chi^2_{kq} \).

Schott (1991) also employed the test statistic in (1.11) to test the hypothesis that in each of the first \( k \) components have zero loadings on the last \( q \) original variables. This hypothesis is equivalent to (1.14). The test is to reject \( H_0(k, q) \) if \( r = \text{rank}(\mathbf{T}'\mathbf{B}) \leq q \), or if \( r = q \) and

\[ T_{k,q} = n \sum_{j \notin w} l_j^{-1} \text{tr} \left\{ \mathbf{B}' \tilde{\gamma}_j \tilde{\gamma}_j^\prime \mathbf{B}' \mathbf{D}_j \mathbf{D}_j' \mathbf{B}^{-1} \right\} \geq \chi^2_{k-q} \]

(1.16)

where \( \mathbf{B} \) is defined in (1.13) and \( \mathbf{D}_j \) is an \( m \times m \) diagonal matrix with entries \( l_i/(l_i - l_j)^2, \ j \notin w, \ w = \{1, 2, \ldots, k\} \). The statistic \( T_{k,q} \) in (1.16) is a matrix expression for \( T_n(\mathbf{B}) \) in (1.15). A Bartlett adjustment was proposed based on the idea that if a test statistic \( T \) has a mean which can be expressed as

\[ \text{E}(T) = a \left\{ 1 + \frac{c}{n} + O(n^{-3/2}) \right\}, \]

(1.17)

where \( a \) is the asymptotic mean and \( c \) is a constant, then the mean of adjusted statistic,

\[ T^* = \left( 1 - \frac{c}{n} \right) T, \]

(1.18)

approaches \( a \). It follows from \( T_{k,q} \xrightarrow{L} \chi^2_{k-q} \) that the asymptotic mean of \( T_{k,q} \) in (1.16) is \( a = kq \). An expression for \( c \) can be found in Schott (1991). Schott compared the unadjusted statistic, \( T_{k,q} \), and adjusted statistic, \( T^* \), under several conditions using
simulated data. He showed that, in general, the Bartlett adjusted statistic performs better than $T_{k,q}$ with respect to Type I error.

Dümbgen (1995) proposed a likelihood ratio test for principal components of a matrix $\Sigma$. Let $\Omega$ be the set of all symmetric matrices in $\mathbb{R}^{p \times p}$ and $\mathcal{O}(p)$ be the set of all orthogonal matrices in $\mathbb{R}^{p \times p}$. For $\Sigma \in \Omega$, let $\Gamma \in \mathcal{O}(p)$ such that $\Gamma' \Sigma \Gamma = \Lambda$, where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. Let $m = (m_1, m_2, \ldots, m_{\bar{a}})$ be a partition of $\{1, 2, \ldots, p\}$ into $\bar{a} > 1$ sets. The hypothesis of interest is

$$H_0: \Sigma = \Gamma \Lambda \Gamma',$$

where $\Gamma = \bigoplus_{i=1}^{\bar{a}} \Gamma_{ii}$, $\Lambda = \bigoplus_{i=1}^{\bar{a}} \Lambda_{ii}$, $\Lambda_{ii} = \text{diag}(\lambda_i)$ for $i \in m_i$, $\Gamma_{ii}$ is an $m_i \times m_i$ matrix, $\Gamma_{ii} \in \mathcal{O}(m_i)$, $\sum_{i=1}^{\bar{a}} m_i = p$, and $\oplus$ stands for the direct sum. Let $l_1 \geq l_2 \geq \cdots \geq l_p$ be the eigenvalues of $S$ and define

$$\mu = \arg \min_{\nu \in \Lambda^{(1)}} \| \nu - \lambda_m(S) \|^2,$$

where $\lambda(S)$ is the vector of the ordered eigenvalues of $S$, and

$$\lambda_m(S) = \begin{pmatrix} \lambda(S_{ij})_{i,j \in m_1} \\ \lambda(S_{ij})_{i,j \in m_2} \\ \vdots \\ \lambda(S_{ij})_{i,j \in m_{\bar{a}}} \end{pmatrix}.$$

The exact computation of $\mu$ is described in Dümbgen (1995). The likelihood ratio test is to reject $H_0$ whenever $t_k(S) \leq c(\alpha)$, where

$$t_k(S) = \sum_{i=1}^{p} \ln \left( \frac{\mu_i}{l_i} \right).$$
The test statistic requires a simulation to obtain critical values.

In this thesis, a likelihood approach for making inferences about the covariance parameters is proposed and is applied to redundancy test. Let \( \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p) \) be the diagonal matrix of the eigenvalues of \( \Sigma \) and \( \Gamma^* \) be a set of the corresponding eigenvectors. The subspace spanned by the columns of a matrix, \( T \), is denoted by \( \mathcal{R}(T) \). Let \( A \) be a known \( p \times q \) semi-orthogonal matrix with rank \( q \), and \( M^* \) be a \( p \times m \) semi-orthogonal matrix such that \( \mathcal{R}(M^*) \) spans the subspace generated by \( m \) specific columns of \( I_p \). The hypotheses of interest are

\[
H_0 : \quad A \in \mathcal{R}(\Gamma^* M^*) \quad \text{for} \quad q \leq m, \quad \text{and} \\
H_0^* : \quad \Gamma^* M^* \in \mathcal{R}(A) \quad \text{for} \quad q > m.
\]

(1.19) (1.20)

It follows from (1.20), see Tyler (1981), that

\[
\Gamma^* M^* \in \mathcal{R}(A) \iff A^c \in \mathcal{R}(\Gamma^* M^{*c}),
\]

where \( A^c \) and \( M^{*c} \) are orthogonal complement to \( A \) and \( M^* \), respectively. That is, \( A^c A'^c = I_p - AA' \) and \( M^{*c} M'^{*c} = I_p - M^* M'^* \). Accordingly, the hypothesis \( H_0^* \) in (1.20) can be rewritten and treated as the hypothesis \( H_0 \) in (1.19). Only hypothesis \( H_0 \) in (1.19) is employed in the remainder of this thesis.

To match the theoretical setup in the later chapters and without loss of generality, \( M^* \) can always be equated to the first \( m \) columns of \( I_p \). This simplification is possible because the columns of \( \Gamma^* \) and the rows of \( M^* \) can be permuted to satisfy

\[
\mathcal{R}(M^*) = \mathcal{R} \left( \begin{array}{c} I_m \\ 0 \end{array} \right).
\]
If $\mathcal{R}(M^*) \neq \mathcal{R}\left( \begin{pmatrix} I_m \\ 0 \end{pmatrix} \right)$, then replace $\Gamma^* M^*$ by $\Gamma M$, where $\Gamma = \Gamma^* P$ and $M = P' M^*$, where $P$ is a $p \times p$ permutation matrix. Choose $P$ so that

$$
\mathcal{R}(P' M^*) = \mathcal{R}\left( \begin{pmatrix} I_m \\ 0 \end{pmatrix} \right).
$$

In the remainder of this thesis, hypothesis (1.19) will be written as

$$
H_0: A \in \mathcal{R}(\Gamma M), \quad \text{where } \mathcal{R}(M) = \mathcal{R}\left( \begin{pmatrix} I_m \\ 0 \end{pmatrix} \right).
$$

(1.21)

This hypothesis is general and the hypothesis of redundancy is subsumed as a special case. In Chapter 3, an algorithm for computing MLEs of the covariance parameters satisfying multiplicity and other constraints is proposed under $H_0$ in (1.21). A likelihood ratio test and a Bartlett correction for the test of (1.21) are also developed in Chapter 4.

The hypothesis of redundancy can be written as (1.21) in the following manner. First, the eigenvalues will be reordered so that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_p$. This order is used throughout the remainder of the thesis. The rational for reordering the eigenvalues is described in Chapter 2. Let $\Gamma$ be a matrix whose columns are the corresponding eigenvectors. Then, by equating $A$ to $A = (I_q - 0')'$, the hypothesis in (1.21) matches with the hypothesis in (1.14). This hypothesis states that each of the first $k$ components has zero loadings on the last $q$ original variables. The redundancy hypothesis is identical to that tested by Fujikoshi (1989) and Schott (1991) but it differs from the variable selection approaches of Jolliffe (1972) and McCabe (1984). To understand
the distinction, first note that it follows from (1.2) that \( y = \Gamma z \). Partition \( y \) as

\[
y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},
\]

where \( y_1 \) is \((p - q) \times 1\) and \( y_2 \) is \( q \times 1\). Partition \( z \) as

\[
z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix},
\]

where \( z_1 \) is \( k \times 1 \) and \( z_2 \) is \((p - k) \times 1\). The last \( q \) variables are redundant if their loadings on the first \( k \) principal components are zero. That is, the last \( q \) variables are redundant if

\[
\Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ 0 & \Gamma_{22} \end{pmatrix},
\] (1.22)

It follows from (1.2) that

\[
\text{var} \left( \begin{pmatrix} y \\ z \end{pmatrix} \right) = \text{var} \left( \begin{pmatrix} I_p \\ \Gamma' \end{pmatrix} \right) y = \begin{pmatrix} \Gamma \Lambda \Gamma' & \Gamma \Lambda \\ \Lambda \Gamma' & \Lambda \end{pmatrix},
\]

where \( \Gamma \) is defined in (1.22). The conditional covariance of \( y_2 \) given \( z_2 \) can be written as

\[
\text{var}(y_2|z_2) = \text{var}(y_2) - \text{cov}(y_2, z_2)\{\text{var}(z_2)\}^{-1}\text{cov}(z_2, y_2).
\]

If the last \( q \) variables are redundant, then

\[
\text{var}(y_2|z_2) = \Gamma_{22} \Lambda_2 \Gamma'_{22} - \Gamma_{22} \Lambda_2 \Lambda_2^{-1} \Lambda_2 \Gamma'_{22} = 0.
\]

That is, if \( y_2 \) is redundant with respect to \( z_1 \) then \( \text{var}(y_2|z_2) = 0 \).

On the other hand, Jolliffe (1972) and McCabe (1984) reduced the number of
variables in such a way that the variables can be discarded if all linear functions of $y$
can be approximately reproduced as linear combination of $y_1$. It can be shown that
satisfaction of the constraint (1.22) does not guarantee that any of McCabe criteria
(see page 17) are optimized. For example, if $\Gamma$ satisfies (1.22), then the conditional
covariance of $y_2$ given $y_1$ can be written as

$$\Sigma_{22,1} = (\Gamma_{22}^{-1} \Gamma_{22}'_{22})^{-1},$$

which can have norm as large as $\| (\Gamma_{22}^{-1} \Gamma_{22}'_{22})^{-1} \|^2 = \text{tr}(\Lambda_2^2) = (p - k)\lambda_2^2$ if all
eigenvalues in $\Lambda_2$ are equal to $\lambda_2$. Conversely, $\| \Sigma_{22,1} \|^2$ can be small even though
redundancy is not satisfied. For example, suppose that $y = \begin{pmatrix} y_1 \\ y_1 + \varepsilon \end{pmatrix}$, where
var($y_1$) = $\Sigma_{11}$, var($\varepsilon$) = $a \Sigma_{11}$, and $a$ is a scalar. The covariance matrix of $y$ can be
written as

$$\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 1 + a \end{pmatrix} \otimes \Sigma_{11}.$$  

Let $\Gamma_{11}$ be the matrix of eigenvectors of $\Sigma_{11}$. Then, the eigenvector matrix of $\Sigma$, $\Gamma$,
can be expressed as

$$\Gamma = \begin{pmatrix} 1 & 1 \\ \frac{a}{2} + \sqrt{1 + \frac{a^2}{4}} & \frac{a}{2} - \sqrt{1 + \frac{a^2}{4}} \end{pmatrix} \otimes \Gamma_{11}.$$  

The conditional covariance of $y_2$ given $y_1$ can be written as

$$\Sigma_{22,1} = (1 + a) \Sigma_{11} - \Sigma_{11} \Sigma_{11}^{-1} \Sigma_{11} = a \Sigma_{11}.$$  

McCabe criteria can be satisfied by choosing $a$ to be a small number. If $a$ is close to
zero, then $\Gamma$ approaches

$$\Gamma = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes \Gamma_{11}.$$
which does not satisfy the redundancy constraint in (1.22). Accordingly, the criteria of Jolliffe and McCabe were not constructed to reduce the number of variables with respect to redundancy.

The likelihood ratio test for redundancy is compared with Schott’s tests in (1.16) and (1.18) using simulated data. A simulation study of the effectiveness of the Bartlett correction is described in Chapter 5.

To illustrate the dimensionality reduction problem, two examples are described below.

Example 1

A large number of variables were observed in a study of the quality of pictures produced by a photographic process. The data were originally presented by Jackson and Morris (1957) and were discussed by Schott (1991). The procedure for a check on the process was as follows: A film strip was given a graded series of exposures to white light and was processed. Optical densities were measured through red, green, and blue filters at the high-density portion of the characteristic curve (shadow areas), at the middle-tone portion of the curve (average picture density) and at the toe portion of the curve (highlights, whites, etc.). There were $p = 9$ measurements: three density levels and three colors at each level. The sample covariance matrix based on $N = 109$ was given in Jackson and Morris (1957). The eigenvalues and cumulative proportions of total variance are given in Table 1. The eigenvectors of the first two
Table 1. Eigenvalues of Covariance Matrix: Photographic Example.

<table>
<thead>
<tr>
<th>Component</th>
<th>Eigenvalue</th>
<th>Cumulative proportion of total variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>878.52</td>
<td>0.6092</td>
</tr>
<tr>
<td>2</td>
<td>196.10</td>
<td>0.7452</td>
</tr>
<tr>
<td>3</td>
<td>128.64</td>
<td>0.8344</td>
</tr>
<tr>
<td>4</td>
<td>103.43</td>
<td>0.9062</td>
</tr>
<tr>
<td>5</td>
<td>81.26</td>
<td>0.9625</td>
</tr>
<tr>
<td>6</td>
<td>37.85</td>
<td>0.9888</td>
</tr>
<tr>
<td>7</td>
<td>6.98</td>
<td>0.9936</td>
</tr>
<tr>
<td>8</td>
<td>5.71</td>
<td>0.9976</td>
</tr>
<tr>
<td>9</td>
<td>3.52</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Table 2. Eigenvectors of the First Two Principal Components.

<table>
<thead>
<tr>
<th>Component</th>
<th>Eigenvector</th>
<th>High density</th>
<th>Middle tone</th>
<th>Toe portion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>red</td>
<td>green</td>
<td>blue</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0.305</td>
<td>0.653</td>
<td>0.483</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>-0.486</td>
<td>-0.151</td>
<td>0.588</td>
</tr>
</tbody>
</table>

Principal components are presented in Table 2. The information in these tables leads to the following two questions:

1. How many components should be retained?

2. How many and which variables can be considered as redundant? Consider the first two principal components. These components account for 75% of the total sample variance. Table 2 reveals that first two components have small coefficients on the last three variables. The question is whether these three variables can be discarded. These two questions have been addressed in several papers referenced.
earlier. This thesis proposes an alternative method for answering these questions. This method is based on a likelihood approach and is discussed in Chapters 2-4.

Example 2

The 1984 Olympic track records dataset for women was examined. The data were first analyzed using PCA by Dawkins (1989). There were $p = 7$ track events; namely, 100 meters, 200 meters, 400 meters, 800 meters, 1500 meters, 3000 meters, and marathon (42,195 meters). There were 55 countries having complete records for all events. The data consist of time to complete each event by athlete from each country. Dawkins normalized the data to have mean 0 and standard deviation 1 and then analyzed it. Naik and Khattree (1996) discussed an alternative choice to compare the athletic performance of nations. Instead of using time as the response, speed defined as distance per unit time was used. Speeds for the women track events and coefficients in the first two principal components are presented in Naik and Khattree (1996). In the present context, suppose that it is believed that variability in speed over track events can be summarized in terms of linear, quadratic, cubic, and higher order functions. A profile plot of the expected mean speed against log(distance) appears in Figure 2. It is of interest to examine the polynomial components of the response profile. In this example, the matrix $C^*$ containing the coefficients of orthogonal polynomials is
Figure 2. A Profile Plot of the Expected Mean.

\[ C^* = \begin{pmatrix} -0.4742 & 0.5426 & -0.4769 & 0.2996 & -0.1380 & 0.0402 \\ -0.3332 & 0.1609 & 0.2878 & -0.5738 & 0.5053 & -0.2298 \\ -0.1921 & -0.1229 & 0.4780 & -0.1101 & -0.5214 & 0.5410 \\ -0.0511 & -0.3087 & 0.2700 & 0.4374 & -0.1839 & -0.6791 \\ 0.0768 & -0.3924 & -0.1152 & 0.3879 & 0.5930 & 0.4264 \\ 0.2179 & -0.3914 & -0.5938 & -0.4752 & -0.2612 & -0.0995 \\ 0.7558 & 0.5120 & 0.1502 & 0.0341 & 0.0062 & 0.0008 \end{pmatrix}. \]

The first column of \( C^* \) contains coefficients of the linear contrast. The second column of \( C^* \) contains coefficients of the quadratic contrast etc. The matrix \( C^* \) can be computed by using the Matlab program given in Appendix C.

The eigenvalues of \( C^* S C^* \) and cumulative proportions of total variance are given in Table 3. The eigenvectors of the first two principal components are presented in Table 4. Two questions of interest are addressed. First, how many principal components should be retained? The first component accounts for 55%. The first two components account for 81%.
Table 3. Eigenvalues of Covariance Matrix: Women’s Track Example.

<table>
<thead>
<tr>
<th>Component</th>
<th>Eigenvalue</th>
<th>Cumulative proportion of total variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1246</td>
<td>0.5509</td>
</tr>
<tr>
<td>2</td>
<td>0.0586</td>
<td>0.8102</td>
</tr>
<tr>
<td>3</td>
<td>0.0240</td>
<td>0.9164</td>
</tr>
<tr>
<td>4</td>
<td>0.0091</td>
<td>0.9566</td>
</tr>
<tr>
<td>5</td>
<td>0.0061</td>
<td>0.9836</td>
</tr>
<tr>
<td>6</td>
<td>0.0037</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Table 4. Eigenvectors of the First Two Principal Components.

<table>
<thead>
<tr>
<th>Component</th>
<th>Eigenvector</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>linear</td>
</tr>
<tr>
<td>1</td>
<td>-0.9784</td>
</tr>
<tr>
<td>2</td>
<td>0.0316</td>
</tr>
</tbody>
</table>

principal component. The data will be analyzed using the approach in Chapters 2-4.

The conclusion of this example will be discussed in Chapter 5.
CHAPTER 2

PARAMETERIZATION

In this chapter, two parameterizations of eigenvalues and one parameterization of eigenvectors are presented. Expressions for the first three derivatives of each of the parameterizations are given in Theorems 1-8. These theorems are proven and the results are employed in Chapter 3 to estimate parameters. To make the theorems distinct from the text, theorems and proofs will be set in italic letters. The notation and terminology used here closely parallel that of Boik (2002a, 2002b).

Spectral Model

Let \( \Sigma \) be the \( p \times p \) covariance matrix with eigenvalues \( \lambda_1, \ldots, \lambda_p \) and suppose that \( \gamma_1, \ldots, \gamma_p \) is a set of orthonormal eigenvectors corresponding to these eigenvalues. Then, the spectral model for the covariance matrix is as follows:

\[
\Sigma = \Gamma \Lambda \Gamma',
\]

(2.1)

where \( \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p) \) and \( \Gamma = (\gamma_1 \gamma_2 \cdots \gamma_p) \).

Let \( \mu \) and \( \varphi \) be vectors of unknown parameters. The quantities, \( \Gamma \) and \( \Lambda \), are parameterized as functions of \( \mu \) and \( \varphi \) such that \( \Gamma = \Gamma(\mu) \) and \( \Lambda = \Lambda(\varphi) \). The parameters are arranged in a vector, \( \theta \):

\[
\theta = \begin{pmatrix} \mu \\ \varphi \end{pmatrix}, \quad \begin{pmatrix} \dim(\mu) \\ \dim(\varphi) \end{pmatrix} = \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix},
\]
where \( \dim(\cdot) \) stands for the dimension of vector and \( \nu_1 + \nu_2 = \nu \).

**Parameterization of Eigenvalues**

Denote the \( p \times 1 \) vector of eigenvalues of \( \Sigma \) by \( \mathbf{\lambda} \). Then,

\[
\mathbf{\Lambda} = \sum_{j=1}^{p} \sum_{i=1}^{p} \mathbf{e}_j^p \mathbf{e}_j^p \mathbf{\Lambda} \mathbf{e}_i^p \mathbf{e}_i^p
\]

\[
= \sum_{j=1}^{p} \mathbf{e}_j^p \mathbf{e}_j^p \mathbf{\Lambda} \mathbf{e}_j^p \mathbf{e}_j^p \text{ because } \mathbf{e}_j^p \mathbf{\Lambda} \mathbf{e}_i^p = 0 \text{ if } j \neq i,
\]

\[
= \sum_{j=1}^{p} \mathbf{e}_j^p \mathbf{\lambda}_j \mathbf{e}_j^p,
\]

where \( \mathbf{e}_j^p \) is the \( j \)th column of \( \mathbf{I}_p \).

Taking the vec of both sides yields

\[
\text{vec}(\mathbf{\Lambda}) = \sum_{j=1}^{p} \left( \mathbf{e}_j^p \otimes \mathbf{e}_j^p \right) \mathbf{\lambda}_j
\]

\[
= \sum_{j=1}^{p} \left( \mathbf{e}_j^p \otimes \mathbf{e}_j^p \right) \mathbf{e}_j^p \mathbf{\lambda}.
\]

Accordingly, the relationship between \( \mathbf{\Lambda} \) and \( \mathbf{\lambda} \) is

\[
\text{vec}(\mathbf{\Lambda}) = \mathbf{L} \mathbf{\lambda} \quad \text{and} \quad \mathbf{\lambda} = \mathbf{L}' \text{vec}(\mathbf{\Lambda}), \tag{2.2}
\]

where

\[
\mathbf{L} = \sum_{j=1}^{p} \left( \mathbf{e}_j^p \otimes \mathbf{e}_j^p \right) \mathbf{e}_j^p \mathbf{\lambda} \text{ and } \mathbf{L}' \mathbf{L} = \mathbf{I}_p. \tag{2.3}
\]

Two parameterizations of \( \mathbf{\lambda} \) are used in two applications of the dimension reduction in principal component analysis. These two structures for \( \mathbf{\lambda} \) are described in each of the applications below.
Reduction in the Number of Variables

It is assumed that $\lambda$ is a differentiable function of $\varphi$, where $\varphi$ is a $\nu_2 \times 1$ vector of unknown parameters. Boik (2002a) introduced several possible structures for eigenvalue parameterization. The structure used here is as follows:

$$\lambda = T_1 \exp\{\odot(T_2 \varphi)\},$$

(2.4)

where $T_1$ and $T_2$ are full column-rank eigenvalue design matrices with dimensions $p \times q_2$ and $q_2 \times \nu_2$, respectively and $\odot$ indicates that the operation is to be performed element-wise. For example, if $u$ is a $p \times 1$ vector, then $\exp(\odot u) = (e^{u_1} e^{u_2} \cdots e^{u_p})'$. The following examples show how to construct $T_1$ and $T_2$.

Suppose that $\lambda$ contains $\nu_2$ distinct values. Let $m = (m_1 m_2 \cdots m_{\nu_2})'$ be the vector of eigenvalue multiplicities, where $\sum_{i=1}^{\nu_2} m_i = p$. For example, if $\lambda = (10 \ 10 \ 5 \ 1 \ 1)'$ then $m = (2 \ 1 \ 2)'$. The parameterization in (2.4) allows arbitrary eigenvalue multiplicities. Consider the $p \times 1$ vector of $\lambda$ with multiplicity $m$ and

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ \lambda_2 & \ddots \\ 0 & \cdots & \lambda_p \end{pmatrix} = \begin{pmatrix} I_{m_1} \exp(\varphi_1) \\ I_{m_2} \exp(\varphi_2) \\ \vdots \\ I_{m_{\nu_2}} \exp(\varphi_{\nu_2}) \end{pmatrix}.$$

If the distinct eigenvalues are unordered, then $T_1$ and $T_2$ can be written as

$$T_1 = \bigoplus_{i=1}^{\nu_2} 1_{m_i} \quad \text{and} \quad T_2 = I_{\nu_2}.$$
To order all the eigenvalues from small to large when \( m = 1_p \) and \( p = 4 = \nu_2; T_1 \) and \( T_2 \) can be written as

\[
T_1 = \begin{pmatrix}
1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1
\end{pmatrix} \quad \text{and} \quad T_2 = I_{\nu_2}.
\] (2.5)

To order all the eigenvalues from small to large when \( m = (1 \ 1 \ 2)' \), \( p = 4 \), and \( \nu_2 = 3; T_1 \) and \( T_2 \) can be written as

\[
T_1 = \begin{pmatrix}
1 & 0 & 0 \\
1 & 1 & 0 \\
1 & 1 & 1 \\
1 & 1 & 1
\end{pmatrix} \quad \text{and} \quad T_2 = I_{\nu_2}.
\]

To ensure that the first eigenvalue is the largest when \( m = 1_p \) and \( p = 4 = \nu_2 \), set \( T_1 \) and \( T_2 \) as

\[
T_1 = \begin{pmatrix}
1 & 0 & 0 & 0 \\
1 & -1 & 0 & 0 \\
1 & 0 & -1 & 0 \\
1 & 0 & 0 & -1
\end{pmatrix} \quad \text{and} \quad T_2 = I_{\nu_2}.
\]

To order all the eigenvalues such that \( \lambda_i = \exp\{\alpha + \beta(p - i)\} \), \( i = 1, 2, \ldots, p \), where \( \alpha \) and \( \beta \) are constants; \( T_1 \) and \( T_2 \) can be constructed as

\[
T_1 = I_p \quad \text{and} \quad T_2 = (1_p \quad p),
\]

where \( p \) is a vector with elements \( p - i, \ i = 1, 2, \ldots, p \).

To order all eigenvalues in Example 1 (the study of the quality of pictures) so that the last eigenvalue is the largest, the second last eigenvalue is the second largest, and the first seven eigenvalues have no order but are smaller than the last two; \( T_1 \) and \( T_2 \) can be constructed as follows:

\[
T_1 = \begin{pmatrix}
1_7 & 0_7 & -I_7 \\
1 & 0 & 0_5 \\
1 & 1 & 0_3
\end{pmatrix} \quad \text{and} \quad T_2 = I_9.
\]
Expressions for the first, second, and third derivatives of $\lambda$ in (2.4) with respect to $\varphi'$ are given in Theorem 1. These results are useful for estimating parameters and for constructing likelihood ratio tests including Bartlett corrections as explained in Chapters 3-4.

**Theorem 1 (Boik 2002a, in supplement).** The first three derivatives of $\lambda$ in (2.4) with respect to $\varphi'$ can be written as follows:

$$D^{(1)}_\lambda = \frac{\partial \lambda}{\partial \varphi'} = T_1 \text{diag}\{\exp\{\odot(T_2 \varphi)\}\}T_2,$$

$$D^{(2)}_\lambda = \frac{\partial^2 \lambda}{\partial \varphi' \otimes \partial \varphi'} = \{\exp(\odot T_2 \varphi) \otimes T_1\}'W^{(2)}(T_2 \otimes T_2),$$

$$D^{(3)}_\lambda = \frac{\partial^3 \lambda}{\partial \varphi' \otimes \partial \varphi' \otimes \partial \varphi'} = \{\exp(\odot T_2 \varphi) \otimes T_1\}'W^{(3)}(T_2 \otimes T_2 \otimes T_2),$$

where

$$W^{(2)} = \sum_{i=1}^{q_2} (e_i^{q_2} e_i^{q_3} \otimes e_i^{q_2} e_i^{q_3}) \quad \text{and} \quad W^{(3)} = \sum_{i=1}^{q_2} (e_i^{q_2} e_i^{q_3} \otimes e_i^{q_2} e_i^{q_3} \otimes e_i^{q_3}).$$

**Proof.** Expressing $\lambda$ in (2.4) in term of summation yields

$$\lambda = T_1 \sum_{i=1}^{q_2} e_i^{q_2} \exp\{e_i^{q_3} T_2 \varphi\}. $$

Taking the derivative of $\lambda$ with respect to $\varphi'$ yields

$$D^{(1)}_\lambda = T_1 \sum_{i=1}^{q_2} e_i^{q_2} \exp\{e_i^{q_3} T_2 \varphi\} e_i^{q_3} T_2$$

$$= T_1 \text{diag}\{\exp\{\odot(T_2 \varphi)\}\}T_2,$$

where

$$\text{diag}\{\exp\{\odot(T_2 \varphi)\}\} = \sum_{i=1}^{q_2} e_i^{q_2} \exp\{e_i^{q_3} T_2 \varphi\} e_i^{q_3}.$$
Taking the derivative of (2.6) yields

$$D_\lambda^{(2)} = T_1 \sum_{i=1}^{q_2} e_i^{q_2} \frac{\partial}{\partial \varphi} \left[ \exp \{ e_i^{q_2} T_2 \varphi \} e_i^{q_2} T_2 \right]$$

$$= T_1 \sum_{i=1}^{q_2} e_i^{q_2} \{ \exp \{ e_i^{q_2} T_2 \varphi \} e_i^{q_2} T_2 (I_{q_2} \otimes e_i^{q_2} T_2) \}.$$

Expressing \( \exp \{ e_i^{q_2} T_2 \varphi \} \) as \( e_i^{q_2} \exp \{ \circ (T_2 \varphi) \} \) yields

$$D_\lambda^{(2)} = \sum_{i=1}^{q_2} T_1 e_i^{q_2} e_i^{q_2} \exp \{ \circ (T_2 \varphi) \} (e_i^{q_2} \otimes e_i^{q_2}) (T_2 \otimes T_2)$$

$$= \sum_{i=1}^{q_2} \left( \exp \{ \circ (T_2 \varphi) \} \right)' e_i^{q_2} \otimes T_1 e_i^{q_2} (e_i^{q_2} \otimes e_i^{q_2}) (T_2 \otimes T_2)$$

$$= \sum_{i=1}^{q_2} \left( \exp \{ \circ (T_2 \varphi) \} \right)' \otimes T_1 (e_i^{q_2} \otimes e_i^{q_2}) (e_i^{q_2} \otimes e_i^{q_2}) (T_2 \otimes T_2),$$

where \( (e_i^{q_2} \otimes I_{q_2}) e_i^{q_2} = (e_i^{q_2} \otimes e_i^{q_2}) \).

Simplification yields

$$D_\lambda^{(2)} = \left( \exp \{ \circ (T_2 \varphi) \} \otimes T_1 \right)' \sum_{i=1}^{q_2} (e_i^{q_2} e_i^{q_2} \otimes e_i^{q_2} e_i^{q_2}) (T_2 \otimes T_2)$$

$$= \left( \exp \{ \circ (T_2 \varphi) \} \otimes T_1 \right)' W^{(2)} (T_2 \otimes T_2).$$

Similarly, the third derivative can be obtained as follows:

$$D_\lambda^{(3)} = T_1 \sum_{i=1}^{q_2} e_i^{q_2} \frac{\partial}{\partial \varphi} \left( \exp \{ e_i^{q_2} T_2 \varphi \} (e_i^{q_2} T_2 \otimes e_i^{q_2} T_2) \right)$$

$$= T_1 \sum_{i=1}^{q_2} e_i^{q_2} \{ \exp \{ e_i^{q_2} T_2 \varphi \} e_i^{q_2} T_2 (I_{q_2} \otimes e_i^{q_2} T_2 \otimes e_i^{q_2} T_2) \}$$

$$= \sum_{i=1}^{q_2} T_1 e_i^{q_2} e_i^{q_2} \exp \{ \circ (T_2 \varphi) \} (e_i^{q_2} \otimes e_i^{q_2} \otimes e_i^{q_2}) (T_2 \otimes T_2 \otimes T_2)$$

$$= \sum_{i=1}^{q_2} \left( \exp \{ \circ (T_2 \varphi) \} \right)' e_i^{q_2} \otimes T_1 e_i^{q_2} (e_i^{q_2} \otimes e_i^{q_2} \otimes e_i^{q_2}) (T_2 \otimes T_2 \otimes T_2)$$

$$= \sum_{i=1}^{q_2} \left( \exp \{ \circ (T_2 \varphi) \} \right)' \otimes T_1 (e_i^{q_2} \otimes e_i^{q_2}) (e_i^{q_2} \otimes e_i^{q_2} \otimes e_i^{q_2}) (T_2 \otimes T_2 \otimes T_2).$$
Finally,

\[
D^{(3)}_{\lambda} = (\exp\{(T_2 \varphi)\} \otimes T_1)^j \sum_{i=1}^{q_2} (e_i^{q_2} q_2^{q_2} \otimes e_i^{q_2} q_2^{q_2} \otimes e_i^{q_2} q_2^{q_2})(T_2 \otimes T_2 \otimes T_2)
\]

\[
= (\exp\{(T_2 \varphi)\} \otimes T_1)^j W^{(3)}(T_2 \otimes T_2 \otimes T_2).
\]

\[\square\]

**Reduction in the Number of Components**

To make inferences about the proportion of variance that the smallest \(m\) components account for, the eigenvalues are parameterized to satisfy the following constraints:

\[1_p' \lambda = \varphi_0, \text{ where } \varphi_0 = \text{tr}(\Sigma), \text{ and} \]

\[
\frac{C_1' \lambda}{\varphi_0} = c_0,
\]

where \(C_1\) is a matrix of known constants and \(c_0\) is a vector of known constants. The following parameterization is employed:

\[
\lambda = \frac{\varphi_3 T_3 \exp\{(T_3 \xi)\}}{1_p T_3 \exp\{(T_3 \xi)\}},
\]

where \(\xi = \xi(\varphi)\) is a differentiable function of \(\varphi\), \(T_3\) and \(T_4\) are full column-rank design matrices of known constants with dimensions \(p \times q_3\) and \(q_3 \times q_4\), respectively, \(\dim(\varphi) = \nu_2\), and \(\nu_2 \leq q_4\). It is assumed that (2.8) is consistent. That is, \(\exists \xi\) such that \(\frac{C_1' \lambda}{\varphi_0} = c_0\).

The vector of eigenvalues in (2.9) depends on \(R(T_4)\), the vector space generated by the columns of \(T_4\). If \(1_{q_3} \in R(T_4)\), then \(T_4\) in (2.9) can be reduced to a matrix
that has one fewer column than the original matrix in the following manner. Let

\[ P_1 = \text{ppo}(1_{q_3}) = 1_{q_3} q_3^{-1} 1'_{q_3} \] and write \( T_4 \xi \)

\[ T_4 \xi = (I_{q_3} - P_1) T_4 \xi + P_1 T_4 \xi \]

\[ = (I_{q_3} - P_1) T_4 \xi + 1_{q_3} q_3^{-1} 1'_{q_3} T_4 \xi. \]

Substituting in (2.9) yields

\[ \lambda = \frac{\varphi_4 T_3 \exp\{\circ((I_{q_3} - P_1) T_4 \xi)\}}{1'_{q_3} T_3 \exp\{\circ((I_{q_3} - P_1) T_4 \xi)\}} \]

because \( \exp(q_3^{-1} 1'_{q_3} T_4 \xi) \) is a scalar and can be canceled out. If \( 1_{q_3} \in \mathcal{R}(T_4) \), then \( 1_{q_3} = T_4 b \) for some \( b \). Therefore, if \( 1_{q_3} \in \mathcal{R}(T_4) \), then

\[ \text{rank}[(I_{q_3} - P_1) T_4] = \text{rank} (T_4 - 1_{q_3} q_3^{-1} 1'_{q_3} T_4) \]

\[ = \text{rank} (T_4 - T_4 b q_3^{-1} 1'_{q_3} T_4) \]

\[ = \text{rank} [T_4 (I_{q_4} - b q_3^{-1} 1'_{q_3} T_4)] \]

\[ = \text{rank} (I_{q_4} - b q_3^{-1} 1'_{q_3} T_4) \]

because \( T_4 \) has full column-rank

\[ = \text{tr}(I_{q_4}) - \text{tr}(1_{q_3} q_3^{-1} 1'_{q_3}) \]

because \( b q_3^{-1} 1'_{q_3} T_4 \) is idempotent

\[ = \text{rank}(T_4) - 1. \]

Factor \( (I_{q_3} - P_1) T_4 \) as

\[ (I_{q_3} - P_1) T_4 = U_p V_p', \]

where \( U_p \) and \( V_p \) each have full column-rank. Then

\[ (I_{q_3} - P_1) T_4 \xi = U_p \xi^*, \]

where \( \xi^* = V_p' \xi. \)
Accordingly, $T_4$ should be replaced by $U_p$, a matrix whose columns form a basis set for $\mathcal{R}\left[(I_{q_3} - \mathbf{1}_{q_3}\mathbf{1}_{q_3}^{-1})T_4\right]$. For convenience, the remainder of the thesis still uses symbols $T_4$ and $\xi$ instead of $U_p$ and $\xi^*$. For the example in (2.5), $T_4 = I_{q_2}$ should be replaced by

$$T_4 = \begin{pmatrix} I_{q_2-1}' \\ -1' \end{pmatrix},$$

or any other full column-rank matrix that has column space equal to $\mathcal{R}\left(\begin{pmatrix} I_{q_2-1}' \\ -1' \end{pmatrix}\right)$.

Consider the second constraint. It follows from (2.8) that

$$\frac{C_1'\lambda}{\varphi_0} = c_0 \iff C_2'\exp\{\circ(T_4\xi)\} = 0, \quad (2.10)$$

where $C_2$ is a $q_3 \times r$ matrix with rank $r$ whose columns are a basis set for $\mathcal{R}[T_3'(C_1 - 1_p c_0)]$.

To show the equivalence in equation (2.10), substitute $\lambda$ from (2.9) to obtain

$$\frac{C_1'T_3\exp\{\circ(T_4\xi)\}}{1_p'T_3\exp\{\circ(T_4\xi)\}} = c_0$$

$$\iff C_1'T_3\exp\{\circ(T_4\xi)\} = c_01_p'T_3\exp\{\circ(T_4\xi)\}$$

$$\iff (C_1' - c_01_p')T_3\exp\{\circ(T_4\xi)\} = 0$$

$$\iff C_2'\exp\{\circ(T_4\xi)\} = 0.$$
\[ T_4(V_1 \tau + V_2 \varphi), \] (2.11)

where \( \tau = (V_1' V_1)^{-1} V_1' \xi, \) \( \varphi = (V_2' V_2)^{-1} V_2' \xi, \) and \( \tau \) is an implicit function of \( \varphi. \)

Substituting \( T_4 \xi \) from (2.11) in the constraint (2.10) yields

\[ C_2 \exp\{ \circ(T_4(V_1 \tau + V_2 \varphi)) \} = 0. \]

Taking the derivative of the above constraint with respect to \( \tau' \) yields

\[ \frac{\partial}{\partial \tau'} \{ C_2 \exp\{ \circ(T_4(V_1 \tau + V_2 \varphi)) \} \} = C_2 \text{diag}(\exp\{ \circ(T_4 \xi) \}) T_4 V_1 \]

\[ = C_2' W_1 V_1, \] (2.12)

where

\[ W_1 = \text{diag}(\exp\{ \circ(T_4 \xi) \}) T_4. \] (2.13)

By the implicit function theorem (Taylor and Mann, 1983), if the matrix of derivatives in (2.12) is nonsingular, then \( \tau \) is an implicit function of \( \varphi \) implying that \( \frac{\partial \tau}{\partial \varphi} \) exists.

Theorem 2 chooses a matrix \( V \) such that \( C_2' W_1 V_1 \) in (2.12) is nonsingular. In the theorem, it is assumed that \( C_2' W_1 \) has full row-rank when evaluated in an open neighborhood of \( \xi^* \) and \( \xi^* \) satisfies \( C_2' \exp\{ \circ(T_4 \xi^*) \} = 0. \)

**Theorem 2.** Assuming that \( C_2' W_1 \) has full row-rank when evaluated in an open neighborhood of \( \xi^* \) and \( \xi^* \) satisfies \( C_2' \exp\{ \circ(T_4 \xi^*) \} = 0. \) Write the singular value decomposition of \( C_2' W_1 \) as follows:

\[ C_2' W_1 = U_\xi (D_\xi \ 0) \left( \begin{array}{c} V_1' \\ V_2' \end{array} \right) = U_\xi D_\xi V_1', \] (2.14)
where $D_\xi$ is an $r \times r$ diagonal matrix of singular values, $U_\xi \in O(r)$, $V = (V_1 \ V_2) \in O(q_1)$, and $W_1$ is defined in (2.13). This choice of $V$ ensures that the derivative matrix in (2.12) is nonsingular.

Proof. Assuming that $C'_2 W_1$ has full row-rank. It follows from (2.14) that

$$C'_2 W_1 V_1 = U_\xi D_\xi V'_1 V_1 = U_\xi D_\xi$$

$$\implies (C'_2 W_1 V_1)^{-1} = D_\xi^{-1} U_\xi'.$$

Therefore, $C'_2 W_1 V_1$ is nonsingular for this choice of $V$.

By Theorem 2, $\tau$ is an implicit function of $\varphi$ implying that $\frac{\partial \tau}{\partial \varphi'}$ exists. To find $\frac{\partial \tau}{\partial \varphi'}$, differentiate both sides of the constraint in (2.10) with respect to $\varphi'$ to obtain

$$\frac{\partial}{\partial \varphi'} C'_2 \exp\{\otimes T_4[V_1 \tau + V_2 \varphi]\} = 0$$

$$\implies C'_2 W_1 (V'_1 \frac{\partial \tau}{\partial \varphi'} + V_2) = 0$$

$$\implies \frac{\partial \tau}{\partial \varphi'} = -(C'_2 W_1 V_1)^{-1} C'_2 W_1 V_2 = 0,$$  \hspace{1cm} (2.15)

because $C'_2 W_1 = U_\xi D_\xi V'_1$ and $V'_1 V_2 = 0$.

For fixed $V$ from Theorem 2, the first three derivatives of $\lambda$ in (2.9) with respect to $\varphi'$ are given in Theorem 3. These result are useful for estimating parameters and for constructing likelihood ratio tests including Bartlett correction as discussed in Chapters 3-4.
Theorem 3 (Adapted from Boik 2002b). For fixed $V$ from Theorem 2, the first three derivatives of $\lambda$ in (2.9) with respect to $\varphi^i$ can be written as follows:

\[
D^{(1)}_{\lambda} = \frac{\partial \lambda}{\partial \varphi^i} = w \varphi_0 H_{1,2} W_1 V_2,
\]

\[
D^{(2)}_{\lambda} = \frac{\partial^2 \lambda}{\partial \varphi^i \otimes \partial \varphi^i} = -2w \left( D^{(1)}_{\lambda} \otimes 1_p T_3 W_1 V_2 \right) N_{\nu_2} + w \varphi_0 H_{1,2} H_{\xi} W_2 (V_2 \otimes V_2),
\]

\[
D^{(3)}_{\lambda} = \frac{\partial^3 \lambda}{\partial \varphi^i \otimes \partial \varphi^i \otimes \partial \varphi^i} = -2w \varphi_0 H_{1,2} H_{\xi} W_2 \left[ V_1 D_{\xi} W_2 (V_2 \otimes V_2) \right] N_{\nu_2} + \left\{ \left( D^{(2)}_{\lambda} \otimes 1_p T_3 W_1 V_2 \right) + \left[ 1_p T_3 H_{\xi} W_2 (V_2 \otimes V_2) \otimes D^{(1)}_{\lambda} \right] \right\} N_{\nu_2}
\]

\[+ w \varphi_0 H_{1,2} H_{\xi} W_2 (V_2 \otimes V_2 \otimes V_2), \]

where

\[w = (1_p T_3 \exp \{ \varphi (T_4 \xi) \})^{-1}, \quad H_{1,2} = I_p - \lambda \varphi_0^{-1} 1_p,\]

\[H_{\xi} = I_{q_3} - W_1 V_1 D_{\xi}^{-1} U_{\xi} C_2, \quad N_{\nu_2} = \frac{1}{2} (I_{\nu_2} \otimes I_{\nu_2}),\]

\[W_2 = \sum_{i=1}^{q_3} e_i^{q_3} \exp \{ e_i^{q_3} T_4 \xi \} (e_i^{q_3} T_4 \otimes e_i^{q_3} T_4),\]

\[W_3 = \sum_{i=1}^{q_3} e_i^{q_3} \exp \{ e_i^{q_3} T_4 \xi \} (e_i^{q_3} T_4 \otimes e_i^{q_3} T_4 \otimes e_i^{q_3} T_4),\]

\[e_i^{q_3} \text{ is the } i^{th} \text{ column of } I_{q_3}, \text{ and } I_{(\cdot,\cdot)} \text{ is the commutation matrix.}\]

Proof. The first derivative of $\lambda$ with respect to $\varphi^i$ can be written as

\[
D^{(1)}_{\lambda} = \frac{\partial}{\partial \varphi^i} \left\{ \varphi_0 T_3 \exp \{ \varphi (T_4 \xi) \} [1_p T_3 \exp \{ \varphi (T_4 \xi) \}]^{-1} \right\}
\]

\[= \varphi_0 \left\{ w T_3 W_1 \left( V_1 \frac{\partial \tau}{\partial \varphi^i} + V_2 \right) - w \frac{\lambda}{\varphi_0} 1_p T_3 W_1 \left( V_1 \frac{\partial \tau}{\partial \varphi^i} + V_2 \right) \right\},\]

because \( \frac{\partial \exp \{ \varphi (T_4 \xi) \}}{\partial \varphi^i} = W_1 \frac{\partial \xi}{\partial \varphi^i}, \) where $W_1$ is defined in (2.13).
Therefore,

\[
D^{(1)}_{\lambda} = \varphi_0 \left\{ w T_3 W_1 V_2 - w \frac{\lambda}{\varphi_0} 1'_p T_3 W_1 V_2 \right\} \\
= \varphi_0 w \left( I_p - \frac{\lambda}{\varphi_0} 1'_p \right) T_3 W_1 V_2 \\
= \varphi_0 w H_\lambda T_3 W_1 V_2,
\]

because \( \frac{\partial \tau}{\partial \varphi'} = 0 \) by (2.15). The second derivative of \( \lambda \) with respect to \( \varphi' \) can be written as

\[
D^{(2)}_{\lambda} = \frac{\partial}{\partial \varphi'} \left\{ \varphi_0 w H_\lambda T_3 W_1 \left( V_1 \frac{\partial \tau}{\partial \varphi'} + V_2 \right) \right\} \\
= \varphi_0 \left\{ \frac{\partial w}{\partial \varphi'} (I_{\nu_2} \otimes H_\lambda T_3 W_1 V_2) + w \frac{\partial H_\lambda}{\partial \varphi'} (I_{\nu_2} \otimes T_3 W_1 V_2) \\
+ w H_\lambda T_3 \frac{\partial W_1}{\partial \varphi'} (I_{\nu_2} \otimes V_2) + w H_\lambda T_3 W_1 V_1 \frac{\partial^2 \tau}{\partial \varphi' \otimes \partial \varphi'} \right\}. \tag{2.16}
\]

Applying standard procedures for differentiation yields the following results:

\[
\frac{\partial w}{\partial \varphi'} = -w^2 1'_p T_3 W_1 V_2, \tag{2.17}
\]

\[
\frac{\partial H_\lambda}{\partial \varphi'} = -\frac{1}{\varphi_0} \left( D^{(1)}_{\lambda} \otimes 1'_p \right), \tag{2.18}
\]

\[
\frac{\partial W_1}{\partial \varphi'} = W_2 (V_2 \otimes I_{\nu_2}), \text{ and} \tag{2.19}
\]

\[
\frac{\partial^2 \tau}{\partial \varphi' \otimes \partial \varphi'} = - \left( D^{-1}_{\xi} U_{\xi}' \right) C_2 W_2 (V_2 \otimes V_2). \tag{2.20}
\]

To verify equations (2.19) and (2.20), take the second derivative of (2.15) to obtain

\[
\frac{\partial^2 \tau}{\partial \varphi' \otimes \partial \varphi'} = -\left( C_2 W_1 V_1 \right)^{-1} C_2 \frac{\partial W_1}{\partial \varphi'} (I_{\nu_2} \otimes V_2) \\
= -D^{-1}_{\xi} U_{\xi}' C_2 \frac{\partial W_1}{\partial \varphi'} (I_{\nu_2} \otimes V_2), \tag{2.21}
\]
because $V_1^t V_1 = I_r$. To compute $\frac{\partial W_1}{\partial \varphi'}$, express $W_1$ in term of summation as follows:

$$W_1 = \sum_{i=1}^{q_3} e_i^{\varphi_3} \exp\{e_i^{\varphi_{3d}} T_3 \xi\} e_i^{\varphi_{3d}} T_4.$$

Applying the product rule yields

$$\frac{\partial W_1}{\partial \varphi'} = \sum_{i=1}^{q_3} \{e_i^{\varphi_3} \exp\{e_i^{\varphi_{3d}} T_3 \xi\}(e_i^{\varphi_{3d}} T_4 \varphi e_i^{\varphi_{3d}} T_4)\}(V_2 \otimes I_{\nu_2})$$

$$= W_2(V_2 \otimes I_{\nu_2}). \quad (2.22)$$

Substituting (2.22) in (2.21) yields

$$\frac{\partial^2 \tau}{\partial \varphi' \otimes \partial \varphi'} = -D_{\xi}^{-1} U_{\xi}' C_2 W_2(V_2 \otimes I_{\nu_2})(I_{\nu_2} \otimes V_2)$$

$$= -D_{\xi}^{-1} U_{\xi}' C_2 W_2(V_2 \otimes V_2).$$

Substituting the results (2.17)-(2.20) in (2.16) yields

$$D_\lambda^{(2)} = -w \left( D^{(1)}_\lambda \otimes 1_p T_3 W_1 V_2 \right) I_{(\nu_2, \nu_2)} - w \left( D^{(1)}_\lambda \otimes 1_p T_3 W_1 V_2 \right)$$

$$+ \varphi_0 w H_\lambda T_3 W_2(V_2 \otimes V_2) - \varphi_0 w H_\lambda T_3 W_1 V_1 D_{\xi}^{-1} U_{\xi}' C_2 W_2(V_2 \otimes V_2)$$

$$= -2w \left( D^{(1)}_\lambda \otimes 1_p T_3 W_1 V_2 \right) N_{\nu_2} + \varphi_0 w H_\lambda T_3 H_\xi W_2(V_2 \otimes V_2).$$

To compute the third derivative of $\lambda$ with respect to $\varphi'$, the following additional results are needed:

$$\frac{\partial W_2}{\partial \varphi'} = W_3(V_2 \otimes I_{\nu_2} \otimes I_{\nu_2}) \quad \text{and} \quad (2.23)$$

$$\frac{\partial H_\xi}{\partial \varphi'} = -H_\xi W_2(V_2 \otimes V_1 D_{\xi}^{-1} U_{\xi}' C_2). \quad (2.24)$$

To verify equation (2.24), note that it follows from (2.14) that

$$(C_2 W_1 V_1)^{-1} = D_{\xi}^{-1} U_{\xi}'.$$
\[
\begin{align*}
\frac{\partial H_\xi}{\partial \varphi'} &= \frac{\partial}{\partial \varphi'} \{ I_{q_3} - W_1 V_1 (C_2' W_1 V_1)^{-1} C_2' \} \\
&= - \frac{\partial W_1}{\partial \varphi'} (I_{\nu_2} \otimes V_1 D_\xi^{-1} U_\xi' C_2') - W_1 V_1 \frac{\partial (C_2' W_1 V_1)^{-1}}{\partial \varphi'} (I_{\nu_2} \otimes C_2') \\
&= - W_2 (V_2 \otimes V_1 D_\xi^{-1} U_\xi' C_2') + W_1 V_1 D_\xi^{-1} U_\xi' C_2' W_2 (V_2 \otimes V_1 D_\xi^{-1} U_\xi' C_2') \\
&= - (I_{q_3} - W_1 V_1 D_\xi^{-1} U_\xi' C_2') W_2 (V_2 \otimes V_1 D_\xi^{-1} U_\xi' C_2') \\
&= - H_\xi W_2 (V_2 \otimes V_1 D_\xi^{-1} U_\xi' C_2'),
\end{align*}
\]

by using the property
\[
\frac{\partial X^{-1}}{\partial \theta'} = - X^{-1} \frac{\partial X}{\partial \theta'} (I_{\nu_2} \otimes X^{-1}),
\]

where \( X \) is a nonsingular matrix. The third derivative of \( \lambda \) with respect to \( \varphi' \) can be written as
\[
D^{(3)}_\lambda = \frac{\partial}{\partial \varphi'} \left\{ - 2 w \left[ D^{(1)}_\lambda \otimes 1'_p T_3 W_1 \left( V_1 \frac{\partial \tau}{\partial \varphi'} + V_2 \right) \right] N_{\nu_2} \right. \\
+ \left. \varphi_0 w H_\lambda T_3 H_\xi W_2 \left[ \left( V_1 \frac{\partial \tau}{\partial \varphi'} + V_2 \right) \otimes \left( V_1 \frac{\partial \tau}{\partial \varphi'} + V_2 \right) \right] \right\}. \tag{2.25}
\]

For convenience, the two parts on the right-hand-side of (2.25) are labeled as \( D^{(3)}_\lambda \) (a) and \( D^{(3)}_\lambda \) (b). Simplifying \( D^{(3)}_\lambda \) (a) yields
\[
D^{(3)}_\lambda (a) = -2 \left\{ \frac{\partial w}{\partial \varphi'} \left( I_{\nu_2} \otimes D^{(1)}_\lambda \otimes 1'_p T_3 W_1 V_1 V_2 \right) \right. \\
+ w \left( D^{(2)}_{\lambda \varphi} \otimes 1'_p T_3 W_1 V_2 \right) \\
+ w \left( 1'_p T_3 \frac{\partial W_1}{\partial \varphi'} (I_{\nu_2} \otimes V_2) \otimes D^{(1)}_\lambda \right) \\
+ w \left( 1'_p T_3 W_1 V_1 \frac{\partial^2 \tau}{\partial \varphi' \otimes \partial \varphi'} \otimes D^{(1)}_\lambda \right) \left( I_{\nu_2} \otimes N_{\nu_2} \right). \tag{2.26}
\]
Applying the results (2.17)-(2.20) and (2.23)-(2.24) gives

\[ D_{\lambda}^{(3)}(a) = 2w \left\{ w \left( 1^p T_3 W_1 V_2 \otimes D_{\lambda}^{(1)} \otimes 1^p T_3 W_1 V_2 \right) \\
- \left( D_{\lambda}^{(2)} \otimes 1^p T_3 W_1 V_2 \right) \\
- \left( 1^p T_3 H_\xi W_2 (V_2 \otimes V_2) \otimes D_{\lambda}^{(1)} \right) \right\} (I_{\nu_2} \otimes N_{\nu_2}). \]

(2.27)

Simplifying \( D_{\lambda}^{(3)}(b) \) yields

\[ D_{\lambda}^{(3)}(b) = \varphi_0 \left\{ \frac{\partial w}{\partial \varphi'} (I_{\nu_2} \otimes H_\lambda T_3 H_\xi W_2 (V_2 \otimes V_2)) \right\} \]

\[ + w \frac{\partial H_\lambda}{\partial \varphi'} (I_{\nu_2} \otimes T_3 H_\xi W_2 (V_2 \otimes V_2)) \]

\[ + w H_\lambda T_3 H_\xi W_2 \left( V_1 \frac{\partial^2 \tau}{\partial \varphi' \otimes \partial \varphi'} \otimes V_2 \right) \]

\[ + w H_\lambda T_3 H_\xi W_2 \left( I_{(\nu_4, q_4)} \left( V_1 \frac{\partial^2 \tau}{\partial \varphi' \otimes \partial \varphi'} \otimes V_2 \right) (I_{\nu_2} \otimes I_{(\nu_2, \nu_2)}) \right) \]

\[ = - \varphi_0 w^2 \left( 1^p T_3 W_1 V_2 \otimes H_\lambda T_3 H_\xi W_2 (V_2 \otimes V_2) \right) \]

\[ - w \left( D_{\lambda}^{(1)} \otimes 1^p T_3 H_\xi W_2 (V_2 \otimes V_2) \right) \]

\[ - \varphi_0 w H_\lambda T_3 H_\xi W_2 \left( V_2 \otimes V_1 D_\xi^{-1} U_\xi C_2 W_2 (V_2 \otimes V_2) \right) \]

\[ + \varphi_0 w H_\lambda T_3 H_\xi W_3 (V_2 \otimes V_2 \otimes V_2) \]

\[ + \varphi_0 w H_\lambda T_3 H_\xi W_2 \left( V_1 D_\xi^{-1} U_\xi C_2 W_2 (V_2 \otimes V_2) \otimes V_2 \right) \]

\[ + \varphi_0 w w H_\lambda T_3 H_\xi W_2 \left( V_1 D_\xi^{-1} U_\xi C_2 W_2 (V_2 \otimes V_2) \otimes V_2 \right) \]

\[ \times (I_{\nu_2} \otimes I_{(\nu_2, \nu_2)}). \]

(2.28)
Additional simplifications yield

\[
D^{(3)}_{\lambda}(b) = -\varphi_0 w^2 \left( 1^{\prime}_p T_3 W_1 V_2 \otimes H_\lambda T_3 H_\xi W_2 (V_2 \otimes V_2) \right) \\
- w \left( D^{(1)}_{\lambda} \otimes 1^{\prime}_p T_3 H_\xi W_2 (V_2 \otimes V_2) \right) \\
+ \varphi_0 w H_\lambda T_3 H_\xi W_3 (V_2 \otimes V_2 \otimes V_2) \\
- 2 \varphi_0 w H_\lambda T_3 H_\xi W_2 \left( V_1 D^{-1}_{\xi} U'_\xi C'_2 (V_2 \otimes V_2) \otimes V_2 \right) N_{\nu_2}^{*}, \quad (2.29)
\]

where \( N_{\nu_2}^{*} = \frac{1}{2} I_{(\nu_2, \nu_2)} + (I_{\nu_2} \otimes N_{\nu_2}). \) Combining (2.27) and (2.29) yields

\[
D^{(3)}_{\lambda} = -2w\varphi_0 H_\lambda T_3 H_\xi W_2 \left[ V_1 D^{-1}_{\xi} U'_\xi C'_2 W_2 (V_2 \otimes V_2) \otimes V_2 \right] N_{\nu_2}^{*} \\
- 2w \left\{ \left( D^{(1)}_{\lambda} \otimes 1^{\prime}_p T_3 W_1 V_2 \right) + \left[ 1^{\prime}_p T_3 H_\xi W_2 (V_2 \otimes V_2) \otimes D^{(1)}_{\lambda} \right] \right\} N_{\nu_2}^{*} \\
+ w\varphi_0 H_\lambda T_3 H_\xi W_3 (V_2 \otimes V_2 \otimes V_2).
\]

\[
\square
\]

Corollary 1. If \( C_2 \) has rank = 0, that is there are no additional constraints, then

\( r = 0, \ V_2 = I_{q_4}, \ \nu_2 = q_4, \ \varphi = \xi \) and the derivatives simplify to

\[
D^{(1)}_{\lambda} = w\varphi_0 H_\lambda T_3 W_1, \\
D^{(2)}_{\lambda} = -2w \left( D^{(1)}_{\lambda} \otimes 1^{\prime}_p T_3 W_1 \right) N_{\nu_2}^{*} + w\varphi_0 H_\lambda T_3 W_2, \\
D^{(3)}_{\lambda} = -2w \left\{ \left( D^{(2)}_{\lambda} \otimes 1^{\prime}_p T_3 W_1 \right) + \left[ 1^{\prime}_p T_3 W_2 \otimes D^{(1)}_{\lambda} \right] \right\} N_{\nu_2}^{*} \\
+ w\varphi_0 H_\lambda T_3 W_3.
\]

Since \( \varphi_0 \) defined in (2.7) is a parameter, the first three derivatives of \( \lambda \) in (2.9) with respect to \( \varphi_0 \) are useful for estimating parameters and for constructing the tests. The results are shown in Theorem 4.
Theorem 4. The first derivative of $\lambda$ in (2.9) with respect to $\varphi_0$ can be written as

$$D^{(1)}_{\lambda, \varphi_0} = \frac{\partial \lambda}{\partial \varphi_0} = \frac{T_3 \exp\{\odot(T_i \xi)\}}{1_p' T_3 \exp\{\odot(T_i \xi)\}} = \frac{\lambda}{\varphi_0}$$

and the higher order derivatives are zero.

Parameterization of Eigenvectors

Boik (1998) described a local parameterization of orthogonal and semi-orthogonal matrices. Boik (2002a) applied this parameterization to the eigenvectors of the covariance matrix. Let $\Gamma$ be a $p \times p$ matrix whose columns consist of the eigenvectors of $\Sigma$ and let $\Gamma_0$ be an orthogonal matrix in a small neighborhood of $\Gamma$. That is $\Gamma \approx \Gamma_0$. Boik’s parameterization is based on the trivial equality $\Gamma = I_p \Gamma$. Write the identity matrix as an orthogonal matrix, $\Gamma_0 \Gamma_0'$. Let $G(\mu^*) = \Gamma_0' \Gamma$, where $G(\mu^*) \in O(p)$ and $\mu^*$ is a vector of parameters. Then, $G(\mu^*) \approx I_p$ and the parameterization of eigenvectors can be written as

$$\Gamma = \Gamma_0 G(\mu^*).$$

(2.30)

It is assumed, for now, that $\Gamma_0$ is a known matrix. This assumption is not made in Chapter 3 when estimation is discussed. By orthogonality, $G$ is subject to $p(p+1)/2$ constraints, $GG' = I_p$. If each eigenvalue has multiplicity 1, then $G$ is a function of $p(p-1)/2$ parameters denoted by $\mu^*$. The $p(p-1)/2$ elements of $\mu^*$ correspond to the $p(p-1)/2$ elements in the upper triangle of $G$. Let $\eta^*$ be an implicit function of $\mu^*$, then $\eta^*$ has $p(p+1)/2$ elements. Let $m = (m_1 \ m_2 \ \cdots \ m_{\nu_2})'$ be the vector
of eigenvalue multiplicities and denote the distinct eigenvalues by $(\varphi_1, \varphi_2, \ldots, \varphi_{\nu_2})$ where $\nu_2 = \dim(\varphi)$, $\nu_2 \leq p$, and $\sum_{i=1}^{\nu_2} m_i = p$. The columns of $G$ can be partitioned with respect to the multiplicity vector $m$ as follows:

$$G = \begin{pmatrix} G_1 & G_2 & \cdots & G_{\nu_2} \end{pmatrix}.$$  

The diagonal form of $\Sigma$ can be written as

$$\Sigma = \Gamma \Lambda \Gamma' = \Gamma_0 G \Lambda G' \Gamma_0' = \sum_{i=1}^{\nu_2} \Gamma_0 G_i \varphi_i I_{m_i} G'_i \Gamma_0' = \Gamma_0 \left\{ \sum_{i=1}^{\nu_2} \varphi_i I_{m_i} \right\} \Gamma_0' = \Gamma_0 \left\{ \sum_{i=1}^{\nu_2} \varphi_i I_{m_i} Q_i Q'_i \right\} \Gamma_0'$$

where $Q_i$ is an orthogonal matrix, $Q_i \in O(m_i)$. Partition $G_i$ as $(G'_{i1} \quad G'_{i2} \quad G'_{i3})'$, where $G'_{i1}$ is $\sum_{j=1}^{i-1} m_j \times m_i$, $G_{i2}$ is $m_i \times m_i$, and $G_{i3}$ is $\sum_{j=i+1}^{\nu_2} m_j \times m_i$. Using QR factorization, an orthogonal matrix $Q_i$ can be found such that $G_{i2} Q_i$ is a lower triangular matrix. Then, if $m_i \geq 2$, the columns of $G_i$ can be rotated to annihilate the $m_i(m_i-1)/2$ elements above the main diagonal of $G_{i2}$. The dimension of $\mu^*$ is reduced to $(p^2 - m'm)/2$. For example, if $p = 6$ and $m = (1 \quad 3 \quad 2)'$, then $G$ can be written as follows:

$$G = \begin{pmatrix} \eta_1 & \mu_1^* & \mu_2^* & \mu_3^* & \mu_4^* & \mu_5^* \\
\eta_2 & \eta_7 & 0 & 0 & \mu_6^* & \mu_5^* \\
\eta_3 & \eta_8 & \eta_12 & 0 & \mu_6^* & \mu_{10}^* \\
\eta_4 & \eta_9 & \eta_13 & \eta_16 & \mu_7^* & \mu_{11}^* \\
\eta_5 & \eta_{10} & \eta_{14} & \eta_{17} & \eta_{19} & 0 \\
\eta_6 & \eta_{11} & \eta_{15} & \eta_{18} & \eta_{20} & \eta_{21}^* \end{pmatrix} = \begin{pmatrix} G_1 & G_2 & G_3 \end{pmatrix}.$$  

If $\|\mu^*\|$ is small, then $G \approx I_p$ and given $\mu^*$, $GG' = I_p$ can be solved for $\eta^*$. Solving
for $\eta^*$ is discussed in the last section of this chapter. In general, $G$ can be written as

$$\text{vec } G = A_1 \mu^* + A_2 \eta^*, \quad (2.31)$$

where $A_1$ and $A_2$ are known indicator matrices with dimensions $p^2 \times ((p^2 - m'm)/2)$ and $p^2 \times (p(p + 1)/2)$, respectively. Because $A_1' A_1 = I_{(p^2 - m'm)/2}$, $A_2' A_2 = I_{p(p+1)/2}$, and $A_1' A_2 = 0$, $\text{vec } G$ satisfies $A_1' \text{vec } G = \mu^*$ and $A_2' \text{vec } G = \eta^*$.

Suppose that interest is in testing the hypothesis in (1.21),

$$H_0 : A \in \mathcal{R}(\Gamma M),$$

where $A$ is a known $p \times q$ semi-orthogonal matrix with rank $q$ and $M$ is a $p \times m$ semi-orthogonal matrix such that $\mathcal{R}(M) = \mathcal{R}
\begin{pmatrix}
I_m \\
0
\end{pmatrix}$. The restriction $A \in \mathcal{R}(\Gamma M)$ further reduces the number of entries in $\mu^*$. Theorem 5 describes this reduction. First, however, note that the parameterization in (2.30) is valid for any matrix $\Gamma_0$ in a neighborhood of $\Gamma$. It is most convenient to equate $\Gamma_0$ to $\Gamma$. In this way, the true value of $\mu^*$ becomes zero. That is,

$$\Gamma = \Gamma G(\mu^*)\bigg|_{\mu^* = 0},$$

because $G(0) = I_p$. In the remainder of this thesis, $\Gamma_0$ will be taken to be $\Gamma$.

**Theorem 5.** Denote $C = (M^{cl} \otimes A' \Gamma)$, where $M^{cl} M^{cl} = I_p - MM'$ and $\Gamma = \Gamma G(\mu^*)|_{\mu^* = 0}$. Then,

$$A \in \mathcal{R}(\Gamma M) \iff C \text{vec } G = 0.$$
Proof. It will be first shown that $A \in \mathcal{R}(\Gamma M) \iff A'\Gamma M^e = 0$. The proof consists of the following two steps:

Step 1: proof that $A \in \mathcal{R}(\Gamma M) \implies A'\Gamma M^e = 0$.

Consider $A = (\Gamma MM'\Gamma' + \Gamma M^e M'^e\Gamma')A$.

\[
A \in \mathcal{R}(\Gamma M) \implies \Gamma MM'\Gamma' A = A
\]

\[
\implies \Gamma M^e M'^e\Gamma' A = 0
\]

\[
\implies A'\Gamma M^e = 0.
\]

Step 2: proof that $A'\Gamma M^e = 0 \implies A \in \mathcal{R}(\Gamma M)$.

\[
A'\Gamma M^e = 0 \implies A = \Gamma MM'\Gamma' A
\]

\[
\implies A \in \mathcal{R}(\Gamma M).
\]

Applying the property of the vec,

\[
\text{vec}(XYZ) = (Z' \otimes X) \text{vec} Y,
\]

(2.32)

where $X$, $Y$, and $Z$ are any matrices that are conformable for multiplication, yields

\[
A \in \mathcal{R}(\Gamma M) \iff A'\Gamma M^e = 0
\]

\[
\iff \text{vec}(A'\Gamma M^e) = 0
\]

\[
\iff \text{vec}(A'\Gamma G M^e) = 0
\]

\[
\iff C \text{vec} G = 0.
\]

\[\square\]
Accordingly, \( G(\mu^*) \) is further constrained. It was shown in Theorem 5 that \( A \in \mathcal{R}(\mathbf{M}) \iff C \text{ vec } G = 0 \). These additional constraints reduce the number of parameters by \((p - m)q\) and increase the number of implicit parameters by \((p - m)q\).

Let \( V^* \) be a \((p^2 - m'm)/2 \times (p^2 - m'm)/2\) nonsingular matrix and \( V^* = (V_3 \ V_4) \), where \( V_3 \) has dimension \((p^2 - m'm)/2 \times \nu_1\), \( \nu_1 = (p^2 - m'm)/2 - (p - m)q\), \( V_4 \) has dimension \((p^2 - m'm)/2 \times (p - m)q\), and \( V_4'V_3 = 0 \).

Then, vec \( G \) can be rewritten as

\[
\text{vec } G = A_1 \left( V^*V^{*-1} \right) \mu^* + A_2 \eta^* \\
= A_1 V_3(V_3'V_3)^{-1}V_3'\mu^* + A_1 V_4(V_4'V_4)^{-1}V_4'\mu^* + A_2 \eta^* \\
= A_1 V_3 \mu + A_1 V_3 \eta_1 + A_2 \eta_2,
\]

(2.33)

where \( \eta_1 = (V_4'V_4)^{-1}V_4'\mu^* \), \( \eta_2 = \eta^* \), and \( \mu = (V_3'V_3)^{-1}V_3'\mu^* \). Equivalently,

\[
\text{vec } G = A_1 V_3 \mu + [A_1 V_4 \ A_2] \eta, \text{ where } \eta = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}
\]

(2.34)

and \( \eta \) is an implicit function of \( \mu \). Taking the derivative of the constraints with respect to \( \eta' \) yields

\[
\frac{\partial}{\partial \eta'} \left( \begin{array}{c}
D_p' \text{ vec } GG' \\
C \text{ vec } G
\end{array} \right) \bigg|_{\mu = 0} = \begin{bmatrix}
2D_p'A_1V_4 & 2D_p'A_2 \\
CA_1V_4 & CA_2
\end{bmatrix},
\]

(2.35)

where \( D_p \) is the duplication matrix (Magnus and Neudecker, 1999), \( D_p \) satisfies \( \text{vec } H = D_p \text{ vec } H \), \( H \) is a symmetric matrix. By the implicit function theorem (Taylor and Mann, 1983), if the matrix of derivatives in (2.35) is nonsingular, then \( \eta \) is an implicit function of \( \mu \) implying that \( \frac{\partial \eta}{\partial \mu} \) exists. To verify that (2.35) is
nonsingular, the following results are useful:

\[ D_p^t A_2 = I_{p(p+1)/2} \quad \text{and} \quad A_2 D_p^t A_1 = I_{(p,p)} A_1. \] (2.36) (2.37)

The derivative matrix in (2.35) is nonsingular if and only if its rank = \( p(p+1)/2 + (p-m)q \). The rank of a matrix is unchanged by interchange of columns, therefore

\[
\text{rank} \begin{bmatrix} 2D_p^t A_1 V_4 & 2D_p^t A_2 \\ CA_1 V_4 & CA_2 \end{bmatrix} = \text{rank} \begin{bmatrix} 2D_p^t A_2 & 2D_p^t A_1 V_4 \\ CA_2 & CA_1 V_4 \end{bmatrix}.
\]

Because the matrix \( D_p^t A_2 \) is nonsingular with rank \( p(p+1)/2 \) by (2.36), the derivative matrix in (2.35) is nonsingular if and only if rank \( (CA_1 V_4 - CA_2 D_p^t A_1 V_4) = (p-m)q \) (Ouellette, 1981). That is,

\[ C(I_p^2 - I_{(p,p)}) A_1 V_4 \] (2.38)

is nonsingular. The matrix \( V^* \) should be chosen to ensure that matrix (2.38) is nonsingular. A suitable \( V^* \) and expressions for the first three derivatives of vec \( G \) with respect to \( \mu^t \) for a special choice of \( V^* \) are given in Theorem 6 and Theorem 7.

**Theorem 6.** Define \( W_4 \) as \( W_4 = C(I_p^2 - I_{(p,p)}) A_1 \). Express \( W_4 \) in term of its singular value decomposition as

\[ W_4 = U(D \quad 0) \begin{pmatrix} V_4' \\ V_3' \end{pmatrix} = UD V_4', \] (2.39)

where \( C \) is defined in Theorem 5, \( D \) is an \( (p-m)q \times (p-m)q \) diagonal matrix of singular values, and \( V^* = (V_3 \quad V_4) \in \mathcal{O}((p^2 - m'm)/2) \). This choice of \( V^* \) ensures that the derivative in (2.35) is nonsingular.
Proof. It is necessary to show that $W_4 = C(I_p - I_{(p,p)})A_1$ has full row-rank. The quantity $W_4$ has full row-rank if and only if

$$t'C(I_p - I_{(p,p)})A_1 = 0 \implies t = 0.$$  \hfill (2.40)

It follows from (1.21) that

$$M^c = \begin{pmatrix} 0 \\ R \end{pmatrix},$$

where $R$ is any $(p - m) \times (p - m)$ orthogonal matrix.

Therefore,

$$C = (M^c \otimes A\Gamma) = \begin{bmatrix} (0_{(p-m \times m)} \otimes R') \otimes A\Gamma \end{bmatrix}.$$  \hfill (2.41)

Let $\Gamma = (\Gamma_1 \ \ \Gamma_2)$ which have dimensions $p \times m$ and $p \times (p - m)$, respectively. It also follows from (1.21) that

$$A \in \mathcal{R}(\Gamma M) \implies A \in \mathcal{R}(\Gamma_1) \implies A = \Gamma_1 T \quad \text{for some } T$$

$$\implies A = T\Gamma_1'$$

$$\implies A\Gamma = (T' \ 0),$$

because $\Gamma_1$ and $\Gamma_2$ are orthogonal. Substituting in (2.41) yields

$$C = (M^c \otimes A\Gamma) = \begin{pmatrix} 0 & R' \otimes [T' \ 0] \end{pmatrix}.$$  \hfill (2.42)

Expressing $A_1$ as $\sum_{i=1}^{p} \sum_{j=1}^{p} (e^{p}_j \otimes e^{p}_i) e^{w}_j$, where $j \geq i$ and $w = \frac{p^2 - \frac{m}{2} m}{2}$, the left-hand-side of (2.40) can be written as

$$\sum_{i=1}^{p} \sum_{j=1}^{p} t' \begin{pmatrix} 0 & R' \otimes [T' \ 0] \end{pmatrix} (I_p - I_{(p,p)}) (e^{p}_j \otimes e^{p}_i) e^{w}_j = 0.$$
The following properties of the vec are applied:

\[(X' \otimes Y)a_i = \text{vec}(Y \text{dvec}(a_i)X) \quad \text{and} \quad (4.43)\]

\[\text{dvec} \left( \text{vec}(e_i^p e_j^p), p, p \right) = e_i^p e_j^p, \quad (4.44)\]

where \(X\) and \(Y\) are any matrices.

Accordingly,

\[
\sum_{i=1}^{p} \sum_{j=i}^{p} \left\{ t^i \text{vec} \left( \begin{bmatrix} T' & 0 \end{bmatrix} (e_i^p e_j^p) \begin{bmatrix} 0 \\ R \end{bmatrix} \right) e_j^p \right\} - t^i \text{vec} \left( \begin{bmatrix} T' & 0 \end{bmatrix} (e_j^p e_i^p) \begin{bmatrix} 0 \\ R \end{bmatrix} \right) e_i^p \right\} = 0.
\]

Therefore,

\[
\sum_{i=1}^{p} \sum_{j=i}^{p} t^i \text{vec} \left( \begin{bmatrix} T' & 0 \end{bmatrix} (e_i^p e_j^p) \begin{bmatrix} 0 \\ R \end{bmatrix} \right) = \sum_{i=1}^{p} \sum_{j=i}^{p} t^i \text{vec} \left( \begin{bmatrix} T' & 0 \end{bmatrix} (e_j^p e_i^p) \begin{bmatrix} 0 \\ R \end{bmatrix} \right) \times \begin{bmatrix} 0 \\ R \end{bmatrix}.
\]

Define \(T^*\) so that \(t = \text{vec} T^t\), then

\[
\text{tr} \left( T^* | T' 0 | (e_i^p e_j^p) \begin{bmatrix} 0 \\ R \end{bmatrix} \right) = \text{tr} \left( T^* | T' 0 | (e_j^p e_i^p) \begin{bmatrix} 0 \\ R \end{bmatrix} \right)
\]

\[
\implies e_j^p \begin{bmatrix} 0 \\ R \end{bmatrix} T^* | T' 0 | e_i^p = e_i^p \begin{bmatrix} 0 \\ R \end{bmatrix} T^* | T' 0 | e_j^p,
\]

which implies that \(\begin{bmatrix} 0 \\ R \end{bmatrix} T^* | T' 0\) is symmetric.

Finally,

\[
\begin{bmatrix} 0 \\ R \end{bmatrix} T^* | T' 0 = \begin{bmatrix} 0 \\ RT^* T' 0 \end{bmatrix} \quad \text{and} \quad RT^* T' = 0
\]

\[
\implies T^* = 0 \quad \implies t = 0,
\]
because \( R \) is orthogonal and \( T \) has full column-rank. Hence, \( W_A \) has full row-rank and for fixed \( V^* \) in (2.39), \( W_AV_A = UDV \) is nonsingular.

By Theorem 6, \( \eta \) is an implicit function of \( \mu \) implying that \( \frac{\partial \eta}{\partial \mu} \) exists. Expressions for the first derivative of \( \eta \) with respect to \( \mu' \) can be obtained as follows:

\[
\frac{\partial \eta_1}{\partial \mu'} = 0, \quad (2.45)
\]
\[
\frac{\partial \eta_2}{\partial \mu'} = -D_p A_1 V_3. \tag{2.46}
\]

To verify the above equations, take the derivative of the vec of the first constraint, \( GG' = I_p \), and use property (2.32) to obtain

\[
\frac{\partial \text{vec } GG'}{\partial \mu'} = (I_p^2 + I_{(p,p)})(I_p \otimes G') \frac{\partial \text{vec } G}{\partial \mu'} = 0 \quad \text{and}
\]
\[
\frac{\partial \text{vec } GG'}{\partial \mu'} \bigg|_{\mu=0} = 2 N_p \frac{\partial \text{vec } G}{\partial \mu'} \bigg|_{\mu=0} = 0,
\]

where

\[
N_p = \frac{1}{2}(I_p^2 + I_{(p,p)}). \tag{2.47}
\]

Equivalently,

\[
D_p' \frac{\partial \text{vec } G}{\partial \mu'} \bigg|_{\mu=0} = 0, \tag{2.48}
\]

because \( N_p = D_p (D_p D_p')^{-1} D_p' \).

Simplification yields

\[
D_p' A_1 V_3 + D_p' A_1 V_4 \frac{\partial \eta_1}{\partial \mu'} + D_p' A_2 \frac{\partial \eta_2}{\partial \mu'} = 0
\]
\[ \frac{\partial \eta_2}{\partial \mu'} = -[D_p' A_1 V_3 + D_p' A_1 V_4 \frac{\partial \eta_1}{\partial \mu'}], \quad (2.49) \]

because of the result in (2.36).

Taking the derivative of the second constraint, \( C \) vec \( G = 0 \), yields

\[ C \frac{\partial \text{vec } G}{\partial \mu'} = 0 \]

\[ \implies CA_1 V_3 + CA_1 V_4 \frac{\partial \eta_1}{\partial \mu'} + CA_2 \frac{\partial \eta_2}{\partial \mu'} = 0. \quad (2.50) \]

Replacing (2.49) in (2.50) and using the result in (2.37) gives

\[ C(I_p^2 - I_{(p,p)}) A_1 V_4 \frac{\partial \eta_1}{\partial \mu'} = -C(I_p^2 - I_{(p,p)}) A_1 V_3 \]

\[ \implies W_4 V_4 \frac{\partial \eta_1}{\partial \mu'} = -W_4 V_3 \]

\[ \implies \frac{\partial \eta_1}{\partial \mu'} = 0, \]

because \( W_4 = UDV_4' \) and \( V_4' V_3 = 0 \). Then, it follows from (2.49) that

\[ \frac{\partial \eta_2}{\partial \mu'} = -D_p' A_1 V_3. \]

Expressions for the first, second, and third derivatives of vec \( G \) in (2.33) with respect to \( \mu' \) are given in Theorem 7. These results are useful for estimating parameters and for constructing likelihood ratio tests including Bartlett corrections as explained in Chapters 3-4.
Theorem 7 (Adapted from Boik 2002a). The first three derivatives of $\text{vec} \, G$ in (2.33) with respect to $\mu'$, evaluated at $\mu = 0$, can be written as follows:

$$D_G^{(1)} = \frac{\partial \text{vec} \, G}{\partial \mu'} \bigg|_{\mu=0} = (I_{p^2} - I_{(p,p)}) A_1 V_3,$$

$$D_G^{(2)} = \frac{\partial^2 \text{vec} \, G}{\partial \mu' \otimes \partial \mu'} \bigg|_{\mu=0} = (I_{p^2} - P) A_2 D_p' (I_p \otimes \text{vec} \, I_p \otimes I_p)' \left( D_G^{(1)} \otimes D_G^{(1)} \right),$$

$$D_G^{(3)} = \frac{\partial^3 \text{vec} \, G}{\partial \mu' \otimes \partial \mu' \otimes \partial \mu'} \bigg|_{\mu=0} = (P - I_{p^2}) A_2 D_p' (I_p \otimes \text{vec} \, I_p \otimes I_p)'$$

$$\times \left[ \left( D_G^{(1)} \otimes I_{(p,p)} D_G^{(2)} \right) - \left( D_G^{(2)} \otimes D_G^{(1)} \right) \{ I_{\nu^2} + (I_{\nu_1} \otimes I_{\nu_1,\nu_1}) \} \right],$$

where $P = K_1 (CK_1)^{-1} C$, $K_1 = (I_{p^2} - I_{(p,p)}) A_1 V_4$.

$$CK_1 = W_4 V_4 = UD, \text{ and } C \text{ is defined in Theorem 5.} \quad (2.51)$$

Proof. Denote the $i^{th}$ derivative of $\eta_j$ by $\eta_j^{(i)}$. It follows from (2.33) that

$$D_G^{(1)} = A_1 V_3 + A_1 V_4 \eta_1^{(1)} + A_2 \eta_2^{(1)}$$

$$= A_1 V_3 - A_2 D_p' A_1 V_3 \text{ by (2.45) and (2.46)}$$

$$= (I_{p^2} - I_{(p,p)}) A_1 V_3 \text{ by (2.37)}.$$

The second derivative of $\text{vec} \, G$ can be written as

$$D_G^{(2)} = A_1 V_4 \eta_1^{(2)} + A_2 \eta_2^{(2)}. \quad (2.52)$$

To compute $\eta_1^{(2)}$ and $\eta_2^{(2)}$, first examine the second derivative of $\text{vec} \, G'G$:

$$\frac{\partial^2 \text{vec} \, G'G}{\partial \mu' \otimes \partial \mu'} = \frac{\partial}{\partial \mu'} \left\{ 2N_p (I_p \otimes G') \frac{\partial \text{vec} \, G'}{\partial \mu'} \right\} = 0$$

$$= 2 \, N_p \left\{ \frac{\partial (I_p \otimes G')}{\partial \mu'} \left( I_{\nu^2} \otimes \frac{\partial \text{vec} \, G'}{\partial \mu'} \right) + (I_p \otimes G') \frac{\partial^2 \text{vec} \, G}{\partial \mu' \otimes \partial \mu'} \right\} = 0.$$
Using the properties,

\[(X \otimes Y)Z = I_{(r,p)} (I_r \otimes (\text{vec } I_s)^t \otimes X) (\text{vec } Y^t \otimes I_{(q,s)} Z),\]

where \(X\) is \(p \times q\) and \(Y\) is \(r \times s\), and \(N_p = N_p I_{(p,p)}\), yields

\[
\left. \frac{\partial^2 \text{vec } G^t G}{\partial \mu^t \otimes \partial \mu^t} \right|_{\mu=0} = N_p \left\{ (I_p \otimes (\text{vec } I_p)^t \otimes I_p) \left( D_G^{(1)} \otimes I_{(p,p)} D_G^{(1)} \right) + D_G^{(2)} \right\} = 0
\]

\[
= D_p^t \left\{ - (I_p \otimes (\text{vec } I_p)^t \otimes I_p) \left( D_G^{(1)} \otimes D_G^{(1)} \right) + D_G^{(2)} \right\} = 0
\]

\[
= D_p^t \left\{ -K + D_G^{(2)} \right\} = 0,
\]

(2.53)

where \(K = (I_p \otimes \text{vec } I_p \otimes I_p)^t \left( D_G^{(1)} \otimes D_G^{(1)} \right)\) and \(I_{(p,p)} D_G^{(1)} = - D_G^{(1)}\).

Taking the second derivative of the second constraint, \(C \text{ vec } G = 0\), yields

\[
C \left. \frac{\partial^2 \text{vec } G}{\partial \mu^t \otimes \partial \mu^t} \right|_{\mu=0} = 0
\]

\[
C D_G^{(2)} = 0.
\]

(2.54)

Substituting \(D_G^{(2)}\) from (2.52) in (2.53) yields

\[
D_p^t \left\{ -K + A_1 V_4 \eta_1^{(2)} + A_2 \eta_2^{(2)} \right\} = 0,
\]

then

\[
\eta_2^{(2)} = D_p^t K - D_p^t A_1 V_4 \eta_1^{(2)},
\]

(2.55)

by using result (2.37). Substituting \(D_G^{(2)}\) from (2.52) in (2.54) yields

\[
C \left[ A_1 V_4 \eta_1^{(2)} + A_2 \eta_2^{(2)} \right] = 0.
\]

Substituting \(\eta_2^{(2)}\) from (2.55) yields

\[
C(I_p^2 - I_{(p,p)}) A_1 V_4 \eta_1^{(2)} + CA_2 D_p^t K = 0
\]

\[
\Rightarrow W_4 V_4 \eta_1^{(2)} + CA_2 D_p^t K = 0.
\]
Replacing $W_4 = UDV_4^\dagger$ yields

$$\eta_1^{(2)} = -(U \, D)^{-1} C A_2 D_p^i K.$$  

Then, it follows from (2.55) that

$$\eta_2^{(2)} = D_p^i K + D_p^i A_1 V_4 (U \, D)^{-1} C A_2 D_p^i K.$$  

Consequently,

$$D_G^{(2)} = A_2 D_p^i K - P A_2 D_p^i K$$

$$= (I_p^2 - P) A_2 D_p^i (I_p \otimes \text{vec} \, I_p \otimes I_p)^T \left( D_G^{(1)} \otimes D_G^{(1)} \right).$$

Similarly, $D_G^{(3)}$ can be obtained. It follows from (2.33) that

$$D_G^{(3)} = A_1 V_4 \eta_1^{(3)} + A_2 \eta_2^{(3)}. \tag{2.56}$$

To compute $\eta_1^{(3)}$ and $\eta_2^{(3)}$, the third derivative of $\text{vec} \, G' G$ with respect to $\mu^i$ is obtained as follows:

$$\frac{\partial^3 \text{vec} \, G' G}{\partial \mu^i \otimes \partial \mu^i \otimes \partial \mu^i} \bigg|_{\mu=0} = \mathbf{N}_p \left\{ \begin{array}{l} (I_p \otimes (\text{vec} \, I_p)^T \otimes I_p) \left( I_p^2 \otimes I_{(p,p)} \right) \frac{\partial \left( D_G^{(1)} \otimes D_G^{(1)} \right)}{\partial \mu^i} \bigg|_{\mu=0} \\ + \frac{\partial (I_p \otimes G')}{\partial \mu^i} D_G^{(2)} \bigg|_{\mu=0} \end{array} \right\} = 0. \tag{2.57}$$

For convenience, the two parts on the right-hand-side of (2.57) are labeled as $G_a$ and $G_b$. Using the Kronecker product rule, $G_a$ can be simplified to

$$G_a = -N_p \left( I_p \otimes (\text{vec} \, I_p)^T \otimes I_p \right) \left[ D_G^{(2)} \otimes D_G^{(1)} \right]$$

$$+ I_{(p^2, p^2)} \left( D_G^{(2)} \otimes D_G^{(1)} \right) \left( I_{\nu_1} \otimes I_{(\nu_1, \nu_1)} \right).$$
Express \((\text{vec } I_p)'\) as \(\sum_{i=1}^{p} (e_i^p \otimes e_i^p)\) and repeatedly use the permutation matrix:

\[
(A \otimes B) = I_{(p,r)}(B \otimes A)I_{(c,q)}, \text{ where } A \text{ is } r \times c \text{ and } B \text{ is } p \times q,
\]

to simplify \(G_a\) as follows:

\[
G_a = -N_p (I_p \otimes (\text{vec } I_p)' \otimes I_p) \left( D_G^{(2)} \otimes D_G^{(1)} \right) \left[ I_{\nu_1} + (I_{\nu_1} \otimes I_{(\nu_1, \nu_1)}) \right].
\]

Similarly, \(G_b\) can be simplified to

\[
G_b = N_p \left\{ (I_p \otimes (\text{vec } I_p)' \otimes I_p) \left( D_G^{(1)} \otimes I_{(p,p)} D_G^{(2)} \right) + D_G^{(3)} \right\}.
\]

Combining \(G_a\) and \(G_b\) yields

\[
\frac{\partial^3 \text{vec } G'G}{\partial \mu' \otimes \partial \mu' \otimes \partial \mu'} \bigg|_{\mu=0} = N_p \left\{ (I_p \otimes (\text{vec } I_p)' \otimes I_p) \left[ -\left( D_G^{(2)} \otimes D_G^{(1)} \right) \right] \right. \\
\left. \left[ I_{\nu_1} + (I_{\nu_1} \otimes I_{(\nu_1, \nu_1)}) \right] + \left( D_G^{(1)} \otimes I_{(p,p)} D_G^{(2)} \right) \right\} + D_G^{(3)}
\]

\[
= D_p' \left\{ K^* + D_G^{(3)} \right\} = 0, \quad (2.58)
\]

where

\[
K^* = (I_p \otimes \text{vec } I_p \otimes I_p)' \left\{ \left( D_G^{(1)} \otimes I_{(p,p)} D_G^{(2)} \right) \\
- \left( D_G^{(2)} \otimes D_G^{(1)} \right) \right\} \left[ I_{\nu_1} + (I_{\nu_1} \otimes I_{(\nu_1, \nu_1)}) \right].
\]

It follows from (2.54) that

\[
C D_G^{(3)} = 0. \quad (2.59)
\]

Replacing \(D_G^{(3)}\) from (2.56) in (2.58) yields

\[
D_p' \left\{ K^* + A_1 V_n \eta_1^{(3)} + A_2 \eta_2^{(3)} \right\} = 0.
\]
Therefore,

\[ \eta_2^{(3)} = -D_p'K^* - D_p'A_1V_4\eta_1^{(3)}. \] (2.60)

Substituting \( D_G^{(3)} \) from (2.56) in (2.59) yields

\[ C \left[ A_1V_4\eta_1^{(3)} + A_2\eta_2^{(3)} \right] = 0. \]

Substituting \( \eta_2^{(3)} \) from (2.60) yields

\[ C(I_{p^2} - I_{(p,p)})A_1V_4\eta_1^{(3)} - CA_2D_p'K^* = 0 \]

\[ \implies W_4V_4\eta_1^{(3)} - CA_2D_p'K^* = 0. \]

Replacing \( W_4 = UDV_4' \) yields

\[ \eta_1^{(3)} = (UD)^{-1}CA_2D_p'K^*. \]

It follows from (2.60) that

\[ \eta_2^{(3)} = -D_p'K^* - D_p'A_1V_4(UD)^{-1}CA_2D_p'K^*. \]

Finally, replacing \( \eta_1^{(3)} \) and \( \eta_2^{(3)} \) into (2.56) yields

\[ D_G^{(3)} = (P - I_{p^2})A_2D_p'K^* \]

\[ = (P - I_{p^2})A_2D_p' \left( I_p \otimes \text{vec} I_p \otimes I_p \right)' \left\{ \begin{array}{c} D_G^{(1)} \otimes I_{(p,p)} \ D_G^{(2)} \\ \left( D_G^{(2)} \otimes D_G^{(1)} \right) \left[ I_{v_1} + (I_{v_1} \otimes I_{(v_1,v_1)}) \right] \end{array} \right\}. \]
Corollary 2. The first three derivatives of \( \text{vec} \, G \) in (2.33) with respect to \( \mu' \), evaluated at \( \mu = 0 \), under constraint \( \mathbf{A} \in \mathcal{R}(\Gamma \mathbf{M}) \), where \( \mathcal{R}(\mathbf{M}) = \mathcal{R}\left(\begin{pmatrix} \mathbf{I}_m \\ 0 \end{pmatrix}\right) \) can be written as follows:

\[
D_G^{(1)} = \frac{\partial \text{vec} G}{\partial \mu'} \bigg|_{\mu=0} = (I_p^2 - I_{(p,p)})A_1 V_3,
\]

\[
D_G^{(2)} = \frac{\partial^2 \text{vec} G}{\partial \mu' \otimes \partial \mu'} \bigg|_{\mu=0} = A_2 D_p' (I_p \otimes \text{vec} I_p) (D_G^{(1)} \otimes D_G^{(1)}) + A_2 D_p' (I_p \otimes \text{vec} I_p) (D_G^{(1)} \otimes I_{(p,p)}) D_G^{(2)}
\]

\[
D_G^{(3)} = \frac{\partial^3 \text{vec} G}{\partial \mu' \otimes \partial \mu' \otimes \partial \mu'} \bigg|_{\mu=0} = -A_2 D_p' (I_p \otimes \text{vec} I_p) (D_G^{(1)} \otimes I_{(p,p)}) D_G^{(2)} - (D_G^{(2)} \otimes D_G^{(1)}) \{I_{p_1} + (I_{\mu} \otimes I_{\mu_1})\}.
\]

Proof. This is a special case of Theorem 7 when \( \mathcal{R}(\mathbf{M}) = \mathcal{R}\left(\begin{pmatrix} \mathbf{I}_m \\ 0 \end{pmatrix}\right) \). It is similar to the proof of Theorem 7 when \( \mathbf{C} \mathbf{A}_2 = 0 \) proven in the next theorem. Then, the first three derivatives can be reduced forms because \( \mathbf{P} \mathbf{A}_2 \) is equal to zero. The second and third derivatives of \( \eta \) with respect to \( \mu' \) can be written as

\[
\eta_1^{(2)} = 0, \quad \eta_2^{(2)} = D_p' K, \quad \eta_1^{(3)} = 0, \quad \text{and} \quad \eta_2^{(3)} = -D_p' K'^*.
\]

\hfill \Box

Under constraint \( \mathbf{A} \in \mathcal{R}(\Gamma \mathbf{M}) \), where \( \mathcal{R}(\mathbf{M}) = \mathcal{R}\left(\begin{pmatrix} \mathbf{I}_m \\ 0 \end{pmatrix}\right) \), the implicit parameter \( \eta \) is restricted that \( \eta_1 = 0 \). This condition reduces form of the first three derivatives of \( \text{vec} \, G \) in (2.33) with respect to \( \mu' \) as shown in Corollary 2. The proof is shown in the following Theorem.
**Theorem 8.** For fixed $V^*$ from Theorem 6 and under constraint $A \in \mathcal{R}(\Gamma M)$ yielding $C \text{vec } G = 0$, where $\mathcal{R}(M) = \mathcal{R}
olimits \begin{pmatrix} I_m \\ 0 \end{pmatrix}$, then

$$\eta_1 = 0,$$

where $\eta_1$ has dimension $(p - m)q \times 1$.

**Proof.** Note that

$$\text{vec } G = A_1 V_3 \mu + A_1 V_3 \eta_1 + A_2 \eta_2,$$

where $\mu = V_3^t \mu^*$, $\eta_1 = V_3^t \mu^*$, and $\eta_2 = \eta^*$.

It follows from (2.39) that

$$C (I_p^2 - I_{(p,p)}) A_1 = UDV_4^t$$

$$\implies CA_1 = UDV_4^t + CI_{(p,p)} A_1. \quad \text{(2.61)}$$

Since,

$$C \text{vec } G = CA_1 \mu^* + CA_2 \eta^* = 0,$$

substituting (2.61) in yields

$$UDV_4^t \mu^* + CI_{(p,p)} A_1 \mu^* + CA_2 \eta^* = 0.$$ 

Therefore,

$$UD \eta_1 = -C(I_{(p,p)} A_1 \mu^* + A_2 \eta^*)$$

$$= -CA_2 \left(D_p^t A_1 \mu^* + \eta^* \right). \quad \text{(2.62)}$$
It will now be shown that $CA_2 = (M^{cl} \otimes AT)A_2 = 0$. It follows from (2.42) that

$$(M^{cl} \otimes AT)A_2 = \left( \begin{bmatrix} 0 & R' \otimes [T' \ 0] \end{bmatrix} \right)A_2,$$

where $R$ is any $(p-m)\times(p-m)$ orthogonal matrix. Expressing $A_2$ as $\sum_{j=1}^{p} \sum_{i=j}^{p} (e_j^p \otimes e_i^w)$, where $i \geq j$, $w = p(p+1)/2$, and $l = \frac{(j-1)(2p-j)}{2} + i$ yields

$$(M^{cl} \otimes AT)A_2 = \sum_{j=1}^{p} \sum_{i=j}^{p} \left( \begin{bmatrix} 0 & R' \otimes [T' \ 0] \end{bmatrix} \right)(e_j^p \otimes e_i^w)e_i^{wt}.$$

Accordingly, using (2.43) and (2.44) yields

$$(M^{cl} \otimes AT)A_2 = \sum_{j=1}^{p} \sum_{i=j}^{p} \text{vec} \left( \begin{bmatrix} T' \ 0 \end{bmatrix}(e_i^p e_j^p)^t \begin{bmatrix} 0 \\ R \end{bmatrix} \right)e_i^{wt},$$

where $i \geq j$.

Choosing $j \leq i \leq m$ yields

$$e_j^p \begin{bmatrix} 0 \\ R \end{bmatrix} = 0.$$

Choosing $i \geq j \geq m$ yields

$$[T' \ 0]e_i^p = 0.$$

Consequently,

$$(M^{cl} \otimes AT)A_2 = CA_2 = 0.$$

By replacing in (2.62), the result is

$$UD\eta_1 = 0$$

$$\implies \eta_1 = 0.$$
Solving for Implicit Parameters

Denote by \( \ell(\theta) \) the loglikelihood function of \( \theta \) given an observed \( Y \). The MLE of \( \theta, \hat{\theta} \), is the vector which maximizes \( \ell(\theta) \). To compute this maximization, the Newton-Raphson algorithm is employed. Let \( \hat{\theta}_0 \) be an initial guess and \( \hat{\theta}_i \) be the estimate of \( \theta \) after the \( i \)th iteration. Expand \( \ell(\hat{\theta}) \) in a Taylor series around \( \hat{\theta} = \hat{\theta}_i \), so that,

\[
\ell(\hat{\theta}) = \ell(\hat{\theta}_i) + (\hat{\theta} - \hat{\theta}_i)' \left[ \frac{\partial \ell(\theta)}{\partial \theta} \right]_{\theta=\hat{\theta}_i} + \frac{1}{2}(\hat{\theta} - \hat{\theta}_i)' \left[ \frac{\partial^2 \ell(\theta)}{\partial \theta \otimes \partial \theta'} \right]_{\theta=\hat{\theta}_i} (\hat{\theta} - \hat{\theta}_i) + o(||\hat{\theta} - \hat{\theta}_i||^2)
\]

\[
\approx \ell(\hat{\theta}_i) + (\hat{\theta} - \hat{\theta}_i)' g_{\hat{\theta}_i} + \frac{1}{2}(\hat{\theta} - \hat{\theta}_i)' H_{\hat{\theta}_i} (\hat{\theta} - \hat{\theta}_i),
\]

(2.63)

where

\[
g_{\hat{\theta}_i} = \frac{\partial \ell(\theta)}{\partial \theta} \bigg|_{\theta=\hat{\theta}_i} \quad \text{and} \quad H_{\hat{\theta}_i} = \frac{\partial^2 \ell(\theta)}{\partial \theta \otimes \partial \theta'} \bigg|_{\theta=\hat{\theta}_i}.
\]

Set the derivative of \( \ell(\hat{\theta}) \) to zero and solve for \( \hat{\theta} \). The result is

\[
\hat{\theta} = \hat{\theta}_i - H_{\hat{\theta}_i}^{-1} g_{\hat{\theta}_i}.
\]

(2.64)

Solving for \( \tau \)

In reduction in the number of components, \( T_i \xi \) can be written as (2.11), where \( \tau \) is an implicit function of \( \varphi \). For given \( \varphi \), to solve for \( \tau \) in (2.11) subject to constraint (2.10), \( C_2' \exp\{\odot(T_i \xi)\} = 0 \), the equation (2.64) is employed as follows:

\[
\hat{\tau} = \hat{\tau}_i - H_{\hat{\tau}_i}^{-1} g_{\hat{\tau}_i},
\]

(2.65)
where

\[ g_{\hat{\eta}_i} = C_2^t \exp \{ \theta (T_i \xi^{(i)}) \} \]  and 

\[ H_{\hat{\eta}_i} = C_2^t \text{diag} \left( \exp \{ \theta (T_i \xi^{(i)}) \} \right) T_i V_1^{(i)}. \]

Solving for \( \eta \)

In reduction in the number of variables, \( \text{vec} \ G \) can be written as (2.34), where \( \eta \) is an implicit function of \( \mu \). For given \( \mu \), to solve for \( \eta \) in (2.34) subject to constraints, \( GG' = I_p \) and \( C \text{vec} \ G = 0 \), the equation (2.64) is employed as follows:

\[ \hat{\eta} = \hat{\eta}_i - H_{\hat{\eta}_i}^{-1} g_{\hat{\eta}_i}, \]

where \( \eta = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \),

\[ g_{\hat{\eta}_i} = \begin{pmatrix} D_p \text{vec} (G^{(i)} G^{(i)} - I_p) \\ C \text{vec} \ G^{(i)} \end{pmatrix}, \quad \text{and} \]

\[ H_{\hat{\eta}_i} = \begin{pmatrix} 2D_p (G^{(i)} \otimes I_p) [A_1 V_1 \quad A_2] \\ C [A_1 V_1 \quad A_2] \end{pmatrix}. \]

In the special case described in Theorem 8, \( \text{vec} \ G \) can be reduced to

\[ \text{vec} \ G = A_1 V_2 \mu + A_2 \eta_2. \]  \hspace{1cm} (2.66)

To solve for \( \eta_2, g_{\hat{\eta}_i} \) and \( H_{\hat{\eta}_i} \) can be simplified to

\[ g_{\hat{\eta}_i} = D_p \text{vec} (G^{(i)} G^{(i)} - I_p) \quad \text{and} \]

\[ H_{\hat{\eta}_i} = 2D_p (G^{(i)} \otimes I_p) A_2. \]
CHAPTER 3

ESTIMATING PARAMETERS AND CONSTRUCTING ASYMPTOTIC DISTRIBUTIONS

In this chapter, the likelihood function will be shown to depend on the unknown parameter \( \Sigma \) and this parameter will then be estimated. The method of maximum likelihood is the most popular technique for deriving estimators. It has been very successful in finding suitable estimators of parameters in many problems. A maximum likelihood estimator, MLE, can be obtained for any multivariate density function whose likelihood function can be written down and maximized with respect to the unknown parameter. This chapter deals with finding the MLE of \( \Sigma \) when the likelihood function is tractable but there is no closed form solution. Then, maximum likelihood estimates are obtained numerically.

Suppose interest is in estimating \( \Sigma \) under the model in (2.1). To find the maximum likelihood estimator of \( \Sigma \), say \( \hat{\Sigma} \), one common method is to use the invariance property of the maximum likelihood estimator in statistical estimation. The invariance property is described in Cox and Hinkley (1974). Briefly, the invariance property of MLE says that if \( \hat{\theta} \) is the maximum likelihood estimator of \( \theta \), then \( \tau(\hat{\theta}) \) is the maximum likelihood estimator of \( \tau(\theta) \), where \( \tau(\theta) \) is a function of \( \theta \). Using this property, the MLE of \( \Sigma \) is given by

\[
\hat{\Sigma} = \hat{\Gamma} \hat{\Lambda} \hat{\Gamma}',
\]
where \( \hat{\Gamma} \) and \( \hat{\Lambda} \) are the MLEs of \( \Gamma \) and \( \Lambda \), respectively. Specific goals of this chapter are to develop procedures for estimating \( \Gamma \), \( \Lambda \), and \( \Sigma \) and to construct asymptotic distributions of \( \hat{\Gamma} \) and \( \hat{\Lambda} \).

**Loglikelihood Function**

Assume that the sample covariance matrix, \( S \), follows a Wishart distribution with covariance \( \Sigma \) and \( n \) degrees of freedom. The distribution of \( S \) is denoted by

\[
nS \sim W_p(n, \Sigma).
\]

The loglikelihood function can be written as

\[
\ell(\theta) = \ell(\theta; S) = -\frac{n}{2} \text{tr} \left( S \Sigma^{-1} \right) - \frac{n}{2} \ln |\Sigma| + \text{constants}, \tag{3.1}
\]

where \( \Sigma = \Gamma \Lambda \Gamma' \), \( \Gamma = \Gamma G \), and \( \theta = (\mu' \quad \varphi')' \) with dimension \( (\nu_1 \quad \nu_2)' \). By using the following results:

\[
\frac{\partial \Sigma^{-1}}{\partial \theta_j} = -\Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_j} \Sigma^{-1},
\]

\[
\frac{\partial \ln |\Sigma|}{\partial \theta_j} = \text{tr} \left( \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_j} \right),
\]

\[
\text{tr}(AB) = (\text{vec } A')' \text{ vec } B,
\]

the first derivative of \( \ell(\theta) \) with respect to \( \theta_j \) can be written as

\[
\frac{\partial \ell(\theta)}{\partial \theta_j} = \frac{n}{2} \text{tr} \left( S \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_j} \Sigma^{-1} \right) - \frac{n}{2} \text{tr} \left( \Sigma \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_j} \Sigma^{-1} \right)
\]

\[
= \frac{n}{2} \left( \text{vec}(S - \Sigma) \right)' (\Sigma^{-1} \otimes \Sigma^{-1}) \frac{\partial \text{vec } \Sigma}{\partial \theta_j}.
\]
Accordingly, the first derivative of $\ell(\theta)$ with respect to $\theta'$ can be written as

$$\frac{\partial \ell(\theta)}{\partial \theta'} = \left( \frac{\partial \ell(\theta)}{\partial \theta_1} \frac{\partial \ell(\theta)}{\partial \theta_2} \ldots \frac{\partial \ell(\theta)}{\partial \theta_\nu} \right)$$

$$= \frac{n}{2} \{ \text{vec}(S - \Sigma) \}' \left( \Sigma^{-1} \otimes \Sigma^{-1} \right) \frac{\partial \text{vec} \Sigma}{\partial \theta'}.$$

Finally, the first derivative of $\ell(\theta)$ with respect to $\theta$ evaluated at $\mu = 0$ can be simplified to

$$\frac{\partial \ell(\theta)}{\partial \theta} \bigg|_{\mu = 0} = \frac{n}{2} F^{(1)} \left( \Sigma^{-1} \otimes \Sigma^{-1} \right) \text{vec}(S - \Sigma), \quad (3.2)$$

where

$$F^{(1)} = \frac{\partial \{ \text{vec} \Sigma \}}{\partial \theta'} \bigg|_{\mu = 0} = \left( \frac{\partial \text{vec} \Sigma}{\partial \mu'} \bigg|_{\mu = 0} \frac{\partial \text{vec} \Sigma}{\partial \phi'} \bigg|_{\mu = 0} \right). \quad (3.3)$$

To compute $F^{(1)}$, $\Sigma$ is expressed as

$$\Sigma = \Gamma G\Lambda G'\Gamma'.$$

Using property (2.32) yields

$$\text{vec} \Sigma = (\Gamma G \otimes \Gamma G) \text{vec} \Lambda.$$

Therefore,

$$\frac{\partial \text{vec} \Sigma}{\partial \phi'} \bigg|_{\mu = 0} = (\Gamma \otimes \Gamma)L D^{(1)}_\lambda,$$

where $D^{(1)}_\lambda = \frac{\partial \lambda}{\partial \phi'}$ and $L$ is defined in (2.3). To compute the derivative of $\text{vec} \Sigma$ with respect to $\mu'$, $\text{vec} \Sigma$ can be expressed as

$$\text{vec} \Sigma = (\Gamma G\Lambda \otimes \Gamma) \text{vec} G \quad \text{or} \quad \text{vec} \Sigma = (\Gamma \otimes \Gamma G\Lambda) \text{vec} G' = I_{(p,p)}(\Gamma G\Lambda \otimes \Gamma) \text{vec} G.$$
Therefore,
\[
\frac{\partial \text{vec } \Sigma}{\partial \mu'} = (I_p + I_{(p,p)})((\Gamma G \Lambda \otimes \Gamma)) \frac{\partial \text{vec } G}{\partial \mu'} \quad \text{and} \quad (3.4)
\]
\[
\frac{\partial \text{vec } \Sigma}{\partial \mu'} \bigg|_{\mu=0} = 2N_p(\Gamma \Lambda \otimes \Gamma) D_G^{(1)},
\]
where \(N_p = (I_p + I_{(p,p)})/2\) and \(D_G^{(1)} = (I_p - I_{(p,p)})A_1V_3\).

Boik (1998) showed that the maximum likelihood estimate of \(\Sigma, \hat{\Sigma},\) can be obtained by solving
\[
\frac{\partial \ell(\theta)}{\partial \theta} \bigg|_{\mu=0} = 0. \quad (3.5)
\]
There is no simple, closed form solution for \(\hat{\Sigma}.\) Therefore, the solution, \(\hat{\Sigma},\) to this equation has to be obtained numerically. Details are explained in the next sections.

**Fisher Scoring Algorithm**

The Fisher Scoring algorithm is an iterative procedure that can solve equation (3.5). The method is based on the derivative of the loglikelihood function and the expected value of the Hessian matrix. Denote by \(\ell(\theta)\) the loglikelihood function of \(\theta\) given an observed \(Y.\) The MLE of \(\theta, \hat{\theta},\) is the vector that maximizes \(\ell(\theta).\) Let \(\hat{\theta}_0\) be an initial guess and \(\hat{\theta}_i\) be the estimate of \(\theta\) after the \(i^{th}\) iteration and let \(I_{\hat{\theta}_i}\) be Fisher’s information at the value \(\theta = \hat{\theta}_i.\) That is,
\[
I_{\hat{\theta}_i} = -E \left[ \frac{\partial^2 \ell(\theta)}{\partial \theta' \partial \theta} \bigg|_{\theta = \hat{\theta}_i} \right]. \quad (3.6)
\]
Expanding \(\ell(\hat{\theta})\) in a Taylor series around \(\hat{\theta} = \hat{\theta}_i\) yields
\[
\ell(\hat{\theta}) \approx \ell(\hat{\theta}_i) + (\hat{\theta} - \hat{\theta}_i)'g_{\hat{\theta}_i} - \frac{1}{2}(\hat{\theta} - \hat{\theta}_i)'I_{\hat{\theta}_i}(\hat{\theta} - \hat{\theta}_i). \quad (3.7)
\]
Set the derivative of $\ell(\hat{\theta})$, expressed as (3.7), to zero and solve for $\hat{\theta}$. The result is

$$\hat{\theta} = \hat{\theta}_i + I^{-1}_{\hat{\theta}_i} g_{\hat{\theta}_i},$$

(3.8)

where $g_{\hat{\theta}_i}$ is defined in (2.63). The right-hand-side in (3.8) becomes the new guess and the procedure is repeated until convergence.

The above description is a conventional application of the Fisher scoring algorithm. The present application is not conventional. Details are given below.

**Solving the Likelihood Equation**

Because there is no closed form solution for the MLE, the solution must be obtained numerically. To compute the maximum likelihood estimates numerically under parameterization of $\lambda$ in (2.4), Boik (2002a)'s algorithm is adopted as follows:

**Step 1** Compute initial guesses for $\hat{\Gamma}$ and $\hat{\Lambda}$ (discussed in the next section). Denote these guesses by $\hat{\Gamma}_0$ and $\hat{\Lambda}_0$, respectively.

**Step 2** Denote the maximum likelihood estimate of $\Sigma$ after the $i^{th}$ iteration by

$$\hat{\Sigma}_i = \hat{\Gamma}_i \hat{\Lambda}_i \hat{\Gamma}_i'.$$

**Step 3** Set $\mu_i = 0$ then $\hat{\theta}_i = (0^{'} \ \varphi_i')'$. Use one iteration of the Fisher scoring algorithm to update $\hat{\theta}_i$, that is,

$$\hat{\theta}_{i+1} = \hat{\theta}_i + I^{-1}_{\hat{\theta}_i} g_{\hat{\theta}_i},$$
where $g_{\hat{\theta}_i}$ is obtained from (3.2) evaluated at $\theta = \hat{\theta}_i$ and $I_{\hat{\theta}_i}$ is defined in (3.6). This is not the conventional Fisher scoring algorithm because $\hat{\mu}_i$ is not updated from iteration to iteration. Instead, $\hat{\mu}_i$ is set to zero and updated from zero in each iteration.

**Step 4** By using the updated $\hat{\theta}_{i+1}$ from step 3, compute $G(\hat{\mu}_{i+1})$ from model (2.31) or (2.33), and $\hat{\Lambda}_{i+1} = \Lambda(\hat{\varphi}_{i+1})$.

**Step 5** Update $\hat{\Gamma}_{i+1} = \hat{\Gamma}_i G(\hat{\mu}_{i+1})$.

**Step 6** Iterate steps 2-5 until convergence.

To numerically compute the MLE under parameterization of $\lambda$ in (2.9), step 4 of the above algorithm is replaced by

**Step 4** By using the updated $\hat{\theta}_{i+1}$ from step 3, solve for $\hat{\xi}_{i+1} = V_1 \hat{\tau}_{i+1} + V_2 \hat{\varphi}_{i+1}$, where $V_1$ and $V_2$ are given in Theorem 2. For given $\hat{\varphi}_{i+1}$, $\hat{\tau}_{i+1}$ is obtained from (2.65). Compute $\hat{\Lambda}_{i+1} = \Lambda(\hat{\xi}_{i+1})$ and $G(\hat{\mu}_{i+1})$ from model (2.31).

**Finding the Initial Guesses**

In using an iterative procedure, the original initial guess is very important. The algorithm may not converge to the MLE if the original initial guess is far from the MLE. The following manner is used for making initial estimates of $\hat{\Gamma}$ and $\hat{\Lambda}$. 
To find an initial guess for \( \hat{\Gamma} \) under the hypothesis (1.21), \( H_0: A \in \mathcal{R}(\Gamma M) \), the singular value decomposition of the sample covariance matrix is employed as follows:

\[
S = \hat{\Gamma}^* \hat{\Lambda}^* \hat{\Gamma}'^*.
\]

Assume that the diagonal entries of \( \hat{\Lambda}^* \) are sorted from small to large and \( \hat{\Gamma}^* \) is a matrix of the corresponding eigenvectors. The quantity \( \hat{\Gamma}^* \) can be partitioned as

\[
\begin{bmatrix}
\hat{\Gamma}^*_1 & \hat{\Gamma}^*_2
\end{bmatrix},
\]

where \( \hat{\Gamma}^*_1 \) and \( \hat{\Gamma}^*_2 \) have dimensions \( p \times m \) and \( p \times (p - m) \), respectively. Denote the original initial guesses for \( \hat{\Gamma} \) and \( \hat{\Lambda} \) by \( \hat{\Gamma}_0 \) and \( \hat{\Lambda}_0 \), respectively. Let \( A^* \) be a \( p \times m \) semi-orthogonal matrix, where \( A^* = (A \quad A^c) \) and \( A^c \) is orthogonal to \( A \). Let \( Q \) be a block diagonal \( p \times p \) orthogonal matrix which, when applied to \( [A^* \quad A^{*c}] \), results in a matrix which is as similar as possible to \( \hat{\Gamma}^* \). To find \( \hat{\Gamma}_0 \), choose \( Q \) such that

\[
\hat{\Gamma}_0 = [A^* \quad A^{*c}] \ Q \approx \hat{\Gamma}^*,
\]

where \( A^{*c} \) is an orthogonal complement to \( A^* \). Writing \( Q \) as a block diagonal matrix yields

\[
[A^* \quad A^{*c}] \begin{pmatrix}
Q_1 & 0 \\
0 & Q_2
\end{pmatrix} \approx \begin{pmatrix}
\hat{\Gamma}^*_1 & \hat{\Gamma}^*_2
\end{pmatrix},
\]

where \( Q_1 \in \mathcal{O}(m) \) and \( Q_2 \in \mathcal{O}(p - m) \). Accordingly,

\[
A^*Q_1 \approx \hat{\Gamma}^*_1 \quad \text{and} \quad (3.10)
\]

\[
A^{*c}Q_2 \approx \hat{\Gamma}^*_2. \quad (3.11)
\]

To solve for the orthogonal matrices \( Q_1 \) and \( Q_2 \), the orthogonal rotation to congruence by Cliff (1966) is repeatedly applied. This method minimizes \( \|A^*Q_1 - \hat{\Gamma}^*_1\|^2 \)
with respect to $Q_1$ for (3.10) and $|A^*Q_2 - \hat{A}^*|^2$ with respect to $Q_2$ for (3.11) and maximizes $\text{tr}(\hat{\Gamma}^+_1 Q'_1 A^{*c})$ and $\text{tr}(\hat{\Gamma}^+_2 Q'_2 A^{*c})$, respectively. The results are shown in the algorithm below.

It follows from (3.10) that

$$[A \ A^c] = \hat{\Gamma}^+_1 Q'_1 = [B_1 \ B_2],$$  \hspace{1cm} (3.12)

where $B_1$ and $B_2$ are the first $q$ and the last $m-q$ columns of $\hat{\Gamma}^+_1 Q'_1$, respectively.

The quantity $A^c$ must be orthogonal to $A$. Accordingly, let $A^c = (I_p - AA')F$ for some $F$, where $F$ is orthogonal by columns. Thus,

$$(I_p - AA')F = B_2. \hspace{1cm} (3.13)$$

The following steps give an algorithm for computing $\hat{\Gamma}_0$.

**Step 1** Begin with any orthogonal matrix $Q_1$.

**Step 2** By using $Q_1$ from step 1, compute $B_2$ from (3.12). Solve for $F$ (Cliff, 1966) in (3.13) as follows: write the singular value decomposition of $B'_2(I_p - AA')$ as

$$B'_2(I_p - AA') = U_1 D_1 J'_1 \hspace{1cm} \Rightarrow F = J_1 U'_1.$$  

**Step 3** Compute $A^c$ and $A^*$ by using $F$ from step 2. Solve for $Q_1$ (Cliff, 1966) in (3.10) as follows: write the singular value decomposition of $\hat{\Gamma}^+_1 A^*$ as

$$\hat{\Gamma}^+_1 A^* = U_2 D_2 J'_2 \hspace{1cm} \Rightarrow Q_1 = J_2 U'_{2p}.$$
Step 4 Repeat steps 2-3 until $Q_1$ converges.

Step 5 Solve for $Q_2$ (Cliff, 1966) in (3.11) as follows: write the singular value decomposition of $\tilde{\Gamma}'' A''$ as

$$\tilde{\Gamma}'' A'' = U_2 D_3 J_3',$$

$$\Rightarrow Q_2 = J_3'U_3'.$$

Step 6 Obtain $\tilde{\Gamma}_0$ from (3.9) by using $Q_1$ and $Q_2$ from steps 4-5, respectively.

To find $\hat{\lambda}_0$ or $\hat{\varphi}_0$ that satisfies model (2.4), let $\hat{\lambda} = \text{diag}(\hat{\lambda}^*)$. Differentiating $\|\hat{\lambda} - T_1 \exp\{\circ T_2 \varphi_0\}\|^2$ or minimizing $\left(\hat{\lambda} - T_1 \exp\{\circ T_2 \varphi_0\}\right)' \left(\hat{\lambda} - T_1 \exp\{\circ T_2 \varphi_0\}\right)$ with respect to $\exp\{\circ T_2 \varphi_0\}$ yields

$$\frac{\partial}{\partial \exp\{\circ T_2 \varphi_0\}} \left(\hat{\lambda} - T_1 \exp\{\circ T_2 \varphi_0\}\right)' \left(\hat{\lambda} - T_1 \exp\{\circ T_2 \varphi_0\}\right).$$

Equating the derivative to zero yields

$$T_1' T_1 \exp\{\circ T_2 \varphi_0\} = T_1' \hat{\lambda}^*$$

$$\Rightarrow \exp\{\circ T_2 \varphi_0\} = (T_1' T_1)^{-1} T_1' \hat{\lambda}^*.$$

If $(T_1' T_1)^{-1} T_1' \hat{\lambda}^*$ contains positive numbers only, then taking the log of both sides yields

$$\hat{\varphi}_0 = (T_2' T_2)^{-1} T_2' \ln\left\{(T_1' T_1)^{-1} T_1' \hat{\lambda}^*\right\}.\quad (3.14)$$

If $(T_1' T_1)^{-1} T_1' \hat{\lambda}^*$ contains non-positive numbers, a way to find $\hat{\varphi}_0$ is to minimize $\|\hat{\lambda} - T_1 \exp\{\circ T_2 \varphi_0\}\|^2$ with respect to $\varphi_0$ which can be done numerically by using
the Newton-Raphson algorithm. To compute \( \hat{\lambda}_0, \varphi \) in (2.4) is replaced by \( \hat{\varphi}_0 \).

To find an initial guess for \( \hat{\xi} \), say \( \hat{\xi}_0 \), that satisfies the constraint (2.10), the following algorithm is used.

**Step 1** Set \( \hat{\varphi}_0 = \text{tr}(S) \) and \( \hat{\lambda}^* = \text{diag}(\hat{\lambda}^*) \). Let \( X = [1_q, T_i] \) and if \( (T_3' T_3)^{-1} T_3' \hat{\lambda}^* \)

contains positive numbers only, define

\[
\left( \begin{array}{c}
\hat{\xi}_0 \\
d
\end{array} \right) = (X' X)^{-1} X' \ln \left[ (T_3' T_3)^{-1} T_3' \hat{\lambda}^* \right],
\]

where \( d \) is a scalar. If \( (T_3' T_3)^{-1} T_3' \hat{\lambda}^* \) contains non-positive numbers, \( \hat{\xi}_0 \) can be obtained numerically by minimizing \( \| \hat{\lambda}^* - \lambda \|^2 \) with respect to \( \xi_0 \), where \( \lambda \) is defined in (2.9).

**Step 2** Compute \( V_1 \) and \( V_2 \) from Theorem 2 and obtain \( \hat{\varphi}_0 = (V_2' V_2)^{-1} V_2 \hat{\xi}_0 \).

**Step 3** Set \( \hat{\xi}_0 = V_1 \hat{\tau}_0 + V_2 \hat{\varphi}_0 \). For fixed \( \hat{\varphi}_0 \), solve for \( \hat{\tau}_0 \) to satisfy the constraint (2.10) by using the Newton-Raphson algorithm in (2.65). Use \( \hat{\tau} = (V_1' V_1)^{-1} V_1' \hat{\xi}_0 \) to be an initial guess.

**Step 4** Obtain new \( \hat{\xi}_0 \) from

\[
\hat{\xi}_0 = V_1 \hat{\tau}_0 + V_2 \hat{\varphi}_0.
\]

**Step 5** Repeat steps 2-4 to obtain \( \hat{\xi}_0 \) and \( \hat{\varphi}_0 \) that satisfy the constraint (2.10).

Asymptotic Distributions of Estimators
Normal Population

Assume that $nS \sim W_p(n, \Sigma)$. Let $\hat{\theta} = (\hat{\mu}' \quad \hat{\varphi}')'$ be the maximum likelihood estimator of $\theta = (\mu' \quad \varphi')'$. A useful property of the MLE is that as $n \to \infty$,

$$
\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{L} N \left( 0, \bar{I}_\theta^{-1} \right), \tag{3.15}
$$

where $\bar{I}_\theta$ is the average Fisher’s information and is defined as

$$
\bar{I}_\theta = \frac{1}{n} E[S(\theta)S(\theta)', \tag{3.16}
$$

where $S(\theta)$ is a score function, $\frac{\partial h(\theta)}{\partial \theta}$, given in (3.2). It follows from (3.16) that

$$
\bar{I}_\theta = \frac{n}{4} F^{(1)'} \left[ \Sigma^{-1} \otimes \Sigma^{-1} \right] E[\text{vec}(S - \Sigma)\{\text{vec}(S - \Sigma)'\}'] \left[ \Sigma^{-1} \otimes \Sigma^{-1} \right] F^{(1)}

= \frac{n}{4} F^{(1)'} \left[ \Sigma^{-1} \otimes \Sigma^{-1} \right] \text{var}(\text{vec} S) \left[ \Sigma^{-1} \otimes \Sigma^{-1} \right] F^{(1)},
$$

where $F^{(1)}$ is defined in (3.3). Magnus and Neudecker (1979) showed that, under normality, $\text{var}(\text{vec} S) = \frac{1}{n^2} N_p(\Sigma \otimes \Sigma)$, where $N_p$ is defined in (2.47). Therefore, the average Fisher’s information can be expressed as

$$
\bar{I}_\theta = \frac{1}{2} F^{(1)'} \left[ \Sigma^{-1} \otimes \Sigma^{-1} \right] F^{(1)}. \tag{3.17}
$$

Partition $\bar{I}_\theta$ conformably to $\theta = (\mu' \quad \varphi')'$ to obtain

$$
\bar{I}_\theta = \begin{pmatrix}
\bar{I}_{\theta,\mu\mu} & \bar{I}_{\theta,\mu\varphi} \\
\bar{I}_{\theta,\varphi\mu} & \bar{I}_{\theta,\varphi\varphi}
\end{pmatrix}, \tag{3.18}
$$

**Theorem 9.** Let $\bar{I}_\theta$ be the average Fisher’s information defined in (3.17). Partition $\bar{I}_\theta$ conformably to $\theta$ as (3.18). The parameters $\mu$ and $\varphi$ are orthogonal (Severini, 2000). That is, $\bar{I}_\theta$ is a block diagonal matrix and can be written as

$$
\bar{I}_\theta = \begin{pmatrix}
\bar{I}_{\theta,\mu\mu} & 0 \\
0 & \bar{I}_{\theta,\varphi\varphi}
\end{pmatrix}. \tag{3.18}
$$
Proof. To show that $\bar{I}_0$ is a block diagonal matrix, $\bar{I}_{\theta,\mu\varphi}$ and $\bar{I}_{\theta,\varphi\mu}$ must be zeros. It follows from (3.18) that

$$\bar{I}_{\theta,\mu\varphi} = -\frac{1}{n} \mathbb{E} \left[ \frac{\partial^2 \ell(\theta)}{\partial \mu \otimes \partial \varphi'} \right].$$

Because

$$\frac{\partial \ell(\theta)}{\partial \varphi'} = \frac{n}{2} \{ \text{vec}(S - \Sigma) \}' \left( \Sigma^{-1} \otimes \Sigma^{-1} \right) \frac{\partial \text{vec} \Sigma}{\partial \varphi'},$$

it follows that

$$-\frac{1}{n} \mathbb{E} \left[ \frac{\partial^2 \ell(\theta)}{\partial \mu \otimes \partial \varphi'} \right]_{\mu = 0} = \frac{1}{2} \left( D_G^{(1)}(\Lambda \Gamma' \otimes \Gamma') \right) \left( \Sigma^{-1} \otimes \Sigma^{-1} \right) \left( \Gamma \otimes \Gamma \right) L D_{\lambda}^{(1)} = \frac{1}{2} D_G^{(1)}(I_p \otimes \Lambda^{-1}) L D_{\lambda}^{(1)}.$$ 

Expressing $L$ as in (2.3) gives

$$\bar{I}_{\theta,\mu\varphi} = \frac{1}{2} D_G^{(1)} \sum_{j=1}^p \left( e_j^p \otimes e_j^p \right) e_j^p \lambda_i^{-1} D_{\lambda}^{(1)}$$

$$= \frac{1}{2} V_3 A_1 \sum_{j=1}^p (I_{p^2} - I_{(p,p)}) \left( e_j^p \otimes e_j^p \right) e_j^p \lambda_i^{-1} D_{\lambda}^{(1)} = 0,$$

because $I_{(p,p)}(e_j^p \otimes e_j^p) = (e_j^p \otimes e_j^p)$.

Similarly, it can be shown that

$$\bar{I}_{\theta,\varphi\mu} = -\frac{1}{n} \mathbb{E} \left[ \frac{\partial^2 \ell(\theta)}{\partial \varphi \otimes \partial \mu'} \right]$$

$$= \frac{1}{2} D_{\lambda}^{(1)} L' \left( I_p \otimes \Lambda^{-1} \right) D_G^{(1)} = 0.$$

Hence, $\bar{I}_0$ is a block diagonal matrix. \qed
Due to the block diagonal structure of $\hat{I}_\theta$, $\hat{\mu}$ is asymptotically independent of $\hat{\phi}$ which implies that $\hat{\Gamma}$ also is asymptotically independent of $\hat{\Lambda}$. Accordingly, it follows from (3.15) that the joint asymptotic distribution of $\hat{\mu}$ and $\hat{\phi}$ is as follows:

$$
\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{L} N \left( 0, \hat{I}_{\theta,\mu}^{-1} \right), \quad (3.19)
$$

$$
\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{L} N \left( 0, \hat{I}_{\theta,\phi\phi}^{-1} \right), \quad (3.20)
$$

and $\hat{\mu} \perp \hat{\phi}$.

Asymptotic distributions of functions of $\hat{\theta}$ can be obtained by the delta method. Let $\hat{\theta}$ be a random $p$-vector with asymptotic distribution $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{L} N(0, \hat{I}_\theta^{-1})$. Suppose that $g(\hat{\theta})$ is a differentiable function. It follows that the asymptotic distribution of $g(\hat{\theta})$ is

$$
\sqrt{n}|g(\hat{\theta}) - g(\theta)| \xrightarrow{L} N \left[ 0, D(\theta)\hat{I}_\theta^{-1}D(\theta)' \right],
$$

where

$$
D(\theta) = \frac{\partial g(\hat{\theta})}{\partial \hat{\theta}} \bigg|_{\hat{\theta} = \theta}.
$$

The proof can be found in Sen and Singer (1993).

To find the asymptotic distributions of eigenvector estimators, consider that

$$
\hat{\Gamma} = \Gamma \hat{G}
$$

$$
\implies \text{vec} \hat{\Gamma} = (I_p \otimes \Gamma) \text{vec} \hat{G}.
$$
The Taylor expansion of $\text{vec} \hat{\Gamma}$ around $\hat{\mu} = 0$ can be expressed as

$$\text{vec} \hat{\Gamma} = \text{vec} \Gamma + (I_p \otimes \Gamma) D^{(1)}_G (\hat{\mu} - \mu) + O_p(n^{-1})$$

and

$$\sqrt{n} \text{vec}(\hat{\Gamma} - \Gamma) = \sqrt{n} (I_p \otimes \Gamma) D^{(1)}_G (\hat{\mu} - \mu) + O_p(n^{-1/2}).$$

Using the delta method and (3.19) yields

$$\sqrt{n} \text{vec}(\hat{\Gamma} - \Gamma) \overset{L}{\rightarrow} N \left[ 0, (I_p \otimes \Gamma) D^{(1)}_G I_{\theta,\mu}^{-1} D^{(1)\prime}_G (I_p \otimes \Gamma') \right]. \quad (3.21)$$

If $\gamma_i$ is identifiable, it follows from (3.21) that

$$\sqrt{n}(\hat{\gamma}_i - \gamma_i) \overset{L}{\rightarrow} N \left[ 0, (I_p \otimes e_i' \Gamma) D^{(1)}_G I_{\theta,\mu}^{-1} D^{(1)\prime}_G (I_p \otimes \Gamma' e_i) \right]. \quad (3.22)$$

Similarly, the relationship between $\hat{\lambda}$ and $\hat{\Lambda}$ is

$$\hat{\lambda} = L \text{vec} \hat{\Lambda}.$$ 

The Taylor expansion of $\hat{\lambda}$ can be written as

$$\hat{\lambda} = \lambda + D^{(1)}_\lambda (\hat{\varphi} - \varphi) + O_p(n^{-1})$$

and

$$\sqrt{n}(\hat{\lambda} - \lambda) = \sqrt{n} D^{(1)}_\lambda (\hat{\varphi} - \varphi) + O_p(n^{-1/2}).$$

Using the delta method and (3.20) yields

$$\sqrt{n}(\hat{\lambda} - \lambda) \overset{L}{\rightarrow} N \left[ 0, D^{(1)}_\lambda I_{\varphi,\phi}^{-1} D^{(1)\prime}_\lambda \right]. \quad (3.23)$$

**Non-normal Population**

Assume that $S$ no longer follows a Wishart distribution. Let $y_i$ for $i = 1, 2, \ldots, N$, $N = n + 1$ be i.i.d. random $p$-vectors with covariance matrix of $\Sigma$. Assume that the fourth
moments of $y$ are finite. Then, as $n \to \infty$,

$$
\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{L} N \left(0, \bar{I}_\theta^{-1} \Omega \bar{I}_\theta^{-1} \right),
$$

where $\Omega = \lim_{n \to \infty} \frac{1}{4} F^{(1)'} \left( \Sigma^{-1} \otimes \Sigma^{-1} \right) \Psi_n \left( \Sigma^{-1} \otimes \Sigma^{-1} \right) F^{(1)}$ and $\Psi_n = n \var\left(\text{vec} \, S\right)$.

Boik (1998) provided an unbiased estimator of $\Psi_n$. Asymptotic distributions of $\hat{\Gamma}$, $\hat{\gamma}_i$, and $\hat{\lambda}$ can be obtained by substituting $\bar{I}_\theta^{-1} \Omega \bar{I}_\theta^{-1}$ for $\bar{I}_\theta^{-1}$ in equation (3.21)-(3.23).
CHAPTER 4

TESTING HYPOTHESIS AND BARTLETT CORRECTION

In principal component analysis, it is often of interest to test hypotheses concerning eigenvalues and eigenvectors of the covariance matrix. The matrix $\Sigma$ is rarely known and tests of hypotheses must be based on an appropriate estimate of $\Sigma$ (discussed in the previous chapter). This chapter deals with procedures for testing dimension reduction. The procedures are based on the standard likelihood ratio test. The hypotheses of interest concern the dimension reduction in the number of components and in the number of original variables. For each hypothesis, the likelihood ratio test statistic is derived and improved by multiplication by a Bartlett correction factor which provides a better approximation to a $\chi^2$ distribution.

**Hypotheses**

Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_p$ be the eigenvalues of $\Sigma$ and $\Gamma = (\gamma_1 \quad \gamma_2 \cdots \gamma_p)$ be a matrix of the corresponding eigenvectors.

In testing dimension reduction in the number of components, the eigenvalues give information about the dimension of the data. If the cumulative proportion of total variance accounted for by the $p - k$ smallest eigenvalues is small, then little information is lost by discarding the corresponding components. The general null
hypothesis of interest is defined in (1.8) as

$$H_0: \frac{C_1^\prime \mathbf{A}}{\text{tr}(\Sigma)} = \mathbf{c}_0,$$

where $C_1$ is a matrix of known constants and $\mathbf{c}_0$ is a vector of known constants.

This hypothesis is tested against the alternative $H_a$ which says that $H_0$ is not true. Specifically, to focus on dimension reduction in the number of components, $C_1$ is defined as $(\mathbf{1}_{p-k} \ 0_k)^\prime$ and $\mathbf{c}_0$ is a small number recommended by an expert in the area of study. The test of this hypothesis is useful for constructing a confidence interval for proportion of variance.

In testing dimension reduction in the number of original of variables, the general null hypothesis of interest is defined in (1.21) as

$$H_0: \mathbf{A} \in \mathcal{R}(\Gamma \mathbf{M}), \text{ where } \mathcal{R}(\mathbf{M}) = \mathcal{R} \left( \begin{pmatrix} I_m \\ 0 \end{pmatrix} \right),$$

where $\mathbf{A}$ is a fixed $p \times q$ semi-orthogonal matrix with rank $q$ and $\mathbf{\Gamma}$ is a matrix of eigenvectors. The alternative hypothesis is the general $H_a$ which says $H_0$ is not true. Suppose $\mathbf{A}$ is equated to $\mathbf{A} = (\mathbf{I}_q \ 0^\prime)^\prime$. Then, the hypothesis to be tested is that the first $k$ components have zero loadings in the last $q$ positions. This test is the redundancy test. If the hypothesis is not rejected, then the last $q$ original variables have zero coefficients in each of the first $k$ components. Hence, these variables are redundant and can be discarded from the study.

In deriving suitable tests for testing these hypotheses, the well-known likelihood ratio tests are used.
Likelihood Ratio Test

This section introduces the likelihood ratio test (LRT) related to the maximum likelihood estimates discussed in Chapter 3. The general strategy of the LRT is to maximize the likelihood function under the null hypothesis, and also to maximize the likelihood function under the alternative hypothesis.

For the loglikelihood function in (3.1) and the maximum likelihood estimates from Chapter 3, the likelihood ratio test statistic is

\[ \Lambda^* = \frac{L_0(\hat{\theta}_0)}{L_a(\hat{\theta}_a)}, \]  

(4.1)

where \( L_0(\hat{\theta}_0) \) and \( L_a(\hat{\theta}_a) \) are the likelihood functions maximized under \( H_0 \) and \( H_a \), and \( \hat{\theta}_0 \) and \( \hat{\theta}_a \) are the maximum likelihood estimators under \( H_0 \) and \( H_a \) respectively. Equivalently, it follows from (4.1) that \( H_0 \) is rejected for large values of

\[ Q = -2 \ln \Lambda^* = 2 \left\{ \ell_a(\hat{\theta}_a) - \ell_0(\hat{\theta}_0) \right\} = D(\hat{\theta}_a; \theta) - D(\hat{\theta}_0; \theta), \]  

(4.2)

where \( D(\hat{\theta}; \theta) \) is defined as \( 2 \left\{ \ell(\hat{\theta}) - \ell(\theta) \right\} \) and \( \ell(\cdot) = \ln L(\cdot) \). Denote the dimensions of \( \theta_0 \) and \( \theta_a \) by \( \nu_0 \) and \( \nu_a \), respectively. The test statistic \( Q \) asymptotically follows

\[ Q \xrightarrow{L} \chi^2_{\nu_\Delta}, \]  

(4.3)

where \( \nu_\Delta = \nu_a - \nu_0 \). The above general result may be used to obtained a test statistic in variety of situations including hypotheses of interest in (1.8) and (1.21). Boik (2002a) discussed a way to compute \( D(\hat{\theta}; \theta) \) in the following manner.
Denote the $i^{th}$ derivative of the loglikelihood function, evaluated at $\mu = 0$, by $\ell_i(\theta)$. That is,

$$
\ell_1(\theta) = \frac{\partial \ell(\theta)}{\partial \theta},
$$

$$
\ell_2(\theta) = \frac{\partial^2 \ell(\theta)}{\partial \theta \otimes \partial \theta'},
$$

$$
\ell_3(\theta) = \frac{\partial^3 \ell(\theta)}{\partial \theta \otimes \partial \theta' \otimes \partial \theta''},
$$

and so forth. Let

$$
K_i = n^{-1} \text{E}(\ell_i(\theta)) \quad \text{and}
$$

$$
Z_i = \sqrt{n} \left( \frac{\ell_i(\theta)}{n} - K_i \right),
$$

where $\text{E}(Z_i) = 0$. Therefore,

$$
\ell_i(\theta) = \sqrt{n}Z_i + nK_i. \quad (4.4)
$$

Let $\hat{\theta}$ be a consistent root of the likelihood equation such that $\ell_1(\hat{\theta}) = 0$. Expanding $\ell_1(\hat{\theta})$ in a Taylor series around $\hat{\theta} = \theta$, substituting $\ell_i(\theta)$ from (4.4), and dividing both sides by $\sqrt{n}$ yields

$$
0 = Z_1 + \left(n^{-\frac{1}{2}}Z_2 + K_2\right)g + \frac{1}{2} \left(n^{-1}Z_3 + n^{-\frac{3}{2}}K_3\right)(g \otimes g)
$$

$$
+ \frac{1}{6} \left(n^{-\frac{3}{2}}Z_4 + n^{-1}K_4\right)(g \otimes g \otimes g) + O_p \left(n^{-\frac{3}{2}}\right), \quad (4.5)
$$

where $g = \sqrt{n}(\hat{\theta} - \theta)$ and $K_1 = 0$. Let

$$
g = \sqrt{n}(\hat{\theta} - \theta) = \delta_0 + n^{-\frac{1}{2}}\delta_1 + n^{-1}\delta_2 + O_p \left(n^{-\frac{3}{2}}\right). \quad (4.6)
To solve for \( \delta_i \) in terms of \( Z_i \), replace \( g \) in (4.5) by (4.6) to obtain polynomial functions of \( n \) which have zero coefficients. Collecting each of coefficients and solving for \( \delta_i \) yields the following results:

\[
\begin{align*}
\delta_0 &= I_\theta^{-1}Z_1, \\
\delta_1 &= I_\theta^{-1} \left\{ Z_2\delta_0 + \frac{1}{2} K_3 \left( \delta_0 \otimes \delta_0 \right) \right\}, \\
\delta_2 &= I_\theta^{-1} \left\{ Z_2\delta_1 + \frac{1}{2} Z_3 \left( \delta_0 \otimes \delta_0 \right) + K_3 \left( \delta_0 \otimes \delta_1 \right) + \frac{1}{6} K_4 \left( \delta_0 \otimes \delta_0 \otimes \delta_0 \right) \right\},
\end{align*}
\]

where \( K_2 = -I_\theta \).

To express \( D(\hat{\theta}; \theta) \) in terms of \( Z_i \), \( i = 1, 2, 3 \), expand \( D(\hat{\theta}; \theta) \) in (4.2) in a Taylor series around \( \hat{\theta} = \theta \) and substitute \( \ell_i(\theta) \) as (4.4) to obtain

\[
D(\hat{\theta}; \theta) = 2g^T Z_1 + g^T \left( n^{-\frac{1}{2}} Z_2 + K_2 \right) g + \frac{1}{3} g^T \left( n^{-1} Z_3 + n^{-\frac{3}{2}} K_3 \right) (g \otimes g)
+ \frac{1}{12} g^T \left( n^{-\frac{3}{2}} Z_4 + n^{-1} K_4 \right) (g \otimes g \otimes g) + O_p \left( n^{-\frac{5}{2}} \right).
\]

Substituting \( g \) from (4.6) and collecting terms of \( O_p(1) \), \( O_p(n^{-\frac{1}{2}}) \) and \( O_p(n^{-1}) \) yields

\[
D(\hat{\theta}; \theta) = Z_1^T \tilde{I}_\theta Z_1 + n^{-\frac{1}{2}} \left\{ Z_1^T K_3 (Z_1^* \otimes Z_1^*) / 3 + Z_1^T Z_2 Z_1 \right\}
+ n^{-1} \left[ \left\{ Z_2 Z_1^* + K_3(Z_1^* \otimes Z_1^*) / 2 \right\} \tilde{I}_\theta^{-1} \left\{ Z_2 Z_1^* + K_3(Z_1^* \otimes Z_1^*) / 2 \right\}
+ Z_1^T K_4 (Z_1^* \otimes Z_1^* \otimes Z_1^*) / 12 + Z_1^T Z_3 (Z_1^* \otimes Z_1^*) / 3 \right] + O_p \left( n^{-\frac{5}{2}} \right),
\]

where \( Z_1^* = \tilde{I}_\theta^{-1} Z_1 \). Because it can be shown that

\[
Z_1 \overset{L}{\rightarrow} N \left( 0, \tilde{I}_\theta \right),
\]
the $O_p(1)$ term in (4.7), $\mathbf{Z}_1^t \mathbf{I}_0 \mathbf{Z}_1^*$, is asymptotically distributed as $\chi^2_{\nu}$ and $E(\mathbf{Z}_1^t \mathbf{I}_0 \mathbf{Z}_1^*) = \nu$. The $O_p\left(n^{-\frac{1}{2}}\right)$ and $O_p\left(n^{-1}\right)$ are useful for constructing Bartlett corrections. Matrix expressions for $\mathbf{Z}_i$ and $\mathbf{K}_i$ for $\ell(\theta)$ in (3.1) are shown in the last section in this chapter.

**Bartlett Correction**

It follows from (4.3) that

$$Q \sim \chi^2_{\nu_{\Delta}} + O_p\left(n^{-\frac{1}{2}}\right).$$

The Bartlett corrected test statistic (Boik, 2002a) is

$$Q_c = \frac{Q_{\nu_{\Delta}}}{\widehat{E}(Q|H_0)}$$  \hspace{1cm} (4.8)

which has null distribution $Q_c \sim \chi^2_{\nu_{\Delta}} + O_p\left(n^{-\frac{3}{2}}\right)$, where

$$\widehat{E}(Q|H_0) = \widehat{E}\left\{D(\hat{\theta}_a; \theta_a)|H_0\right\} - \widehat{E}\left\{D(\hat{\theta}_0; \theta_0)|H_0\right\}.$$  \hspace{1cm} (4.9)

The approximate expectations can be obtained by evaluation of the expectation of equation (4.7). To compute the expectation of $D(\hat{\theta}; \theta)$ in (4.7), the third and fourth order moments of $\text{vec}(\mathbf{S} - \Sigma)$ are needed.

**Third and Fourth Order Moments**

Boik (2002a) discussed an easy way to get these moments directly from the cumulant generating function in the following manner. Let $\mathbf{S} = \sum_{i=1}^{N} \frac{u_i u_i'}{N}$, where $u_i \overset{iid}{\sim} N(\mathbf{0}, \Sigma)$, then $\mathbf{S} \sim W_p(n, \Sigma/n)$. Note that $D_p$ and $H_p$ satisfy $D_p \text{vech} \mathbf{X} = \ldots$
vec $X$, $H_p \ vec X = vech X$, and $D_p H_p = N_p$, where $X$ is any $p \times p$ symmetric matrix and $N_p$ is defined in (2.47). The moment generating function of $vech(S)$ can be written as

$$M_{vech(S)}(t) = E[\exp(t \ vech(S))]$$

$$= E[\exp(t H_p \ vech(S))]$$

$$= \frac{\mid \Sigma^* \mid^{n/2}}{\mid \Sigma \mid^{n/2}},$$

where $vech \Sigma^{-1} = vech \Sigma^{-1} - \frac{2}{n} (D_p D_p)^{-1} t$. Then, the cumulant generating function of $vech(S)$ is

$$C^{(0)}(t) = \frac{n}{2} \ln |\Sigma^*| - \frac{n}{2} \ln |\Sigma|.$$

To find the first four derivatives of the cumulant generating function, the following results are repeatedly used:

$$\frac{\partial \ vec \Sigma^{-1}}{\partial t} = -\frac{2}{n} H_p' \ and \quad (4.10)$$

$$\frac{\partial \ vec \Sigma^*}{\partial t} = \frac{2}{n} (\Sigma^* \otimes \Sigma^*) H_p'. \quad (4.11)$$

The first derivative of the cumulant generating function can be written as

$$C^{(1)}(t) = \frac{\partial C^{(0)}(t)}{\partial t} = \frac{n}{2} \frac{\partial \ln |\Sigma^*|}{\partial t}$$

$$\Rightarrow \frac{\partial C^{(0)}(t)}{\partial t_i} = \frac{n}{2} \text{tr} \left( \Sigma^{-1} \frac{\partial \Sigma}{\partial t_i} \right) = -\frac{n}{2} \text{tr} \left( \Sigma^* \frac{\partial \Sigma^{-1}}{\partial t_i} \right)$$

$$= -\frac{n}{2} (\vec \Sigma^*)' \frac{\partial \vec \Sigma^{-1}}{\partial t_i}$$

$$\Rightarrow \frac{\partial C^{(0)}(t)}{\partial t'} = (\vec \Sigma^*)' H_p'.$$
Therefore,

\[ C^{(1)}(t) = H_p \text{vec} \Sigma^* \text{ or vec } \Sigma^*. \]  

(4.12)

The second derivative of the cumulant generating function can be written as

\[ C^{(2)}(t) = \frac{\partial^2 C^{(0)}(t)}{\partial t^\prime \otimes \partial t} = H_p \frac{\partial \text{vec } \Sigma^*}{\partial t^\prime} \\
= 2 H_p (\Sigma^* \otimes \Sigma^*) H_p^\prime. \]

(4.13)

The third derivative of the cumulant generating function can be written as

\[ C^{(3)}(t) = \frac{\partial^3 C^{(0)}(t)}{\partial t^\prime \otimes \partial t^\prime \otimes \partial t} = \frac{2}{n} H_p \frac{\partial(\Sigma^* \otimes \Sigma^*)}{\partial t^\prime} \frac{\partial \text{vec } \Sigma^*}{\partial t^\prime} \\
= \frac{2}{n} H_p \left\{ 2 (I_p \otimes (\text{vec } I_p)^\prime \otimes \Sigma^*) \left( \frac{\partial \text{vec } \Sigma^*}{\partial t^\prime} \otimes H_p^\prime \right) \right\} \\
= \frac{8}{n^2} H_p (\Sigma^* \otimes (\text{vec } \Sigma^*)^\prime \otimes \Sigma^*) (H_p \otimes H_p)^\prime, \]

(4.14)

by \((\text{vec } I_p)^\prime (\Sigma^* \otimes I_p) = (\text{vec } \Sigma^*)^\prime\).

To compute the fourth derivative of the cumulant generating function, the following results are useful:

\[ E(ABC \otimes D) = \{(\text{vec } B)^\prime \otimes E\} [(I_r \otimes \text{vec } A)C \otimes D], \]

where \(B\) is \(q \times r\), \(A, C, D,\) and \(E\) are any matrices that are conformable for multiplication, and

\[ E(A \otimes BCD) = \{(\text{vec } C)^\prime \otimes E\} I_{(t_p,qr)} [A \otimes (\text{vec } B)^\prime \otimes D], \]

where \(A\) is \(t \times s\), \(B\) is \(p \times q\), \(C\) is \(q \times r\), and \(D\) and \(E\) are any matrices that are conformable for multiplication. The fourth derivative of the cumulant generating
function can be written as

\[
C^{(4)}(t) = \frac{\partial^4 C^{(0)}(t)}{\partial t \otimes \partial t' \otimes \partial t'' \otimes \partial t'''} = \frac{8}{n^2} H_p \left\{ \left( \frac{(\text{vec } \Sigma^*)'}{\partial t} \otimes H_p \right)(I_p \otimes \text{vec } I_p \otimes (\text{vec } \Sigma^*)' \otimes \Sigma^*) \\
+ \left( \frac{(\text{vec } \Sigma^*)'}{\partial t} \otimes H_p \right)(I_{p^2} \otimes (\text{vec } \Sigma^*)' \otimes \text{vec } I_p \otimes I_p) \\
+ \left( I_{\frac{n(n-1)}{2}} \otimes H_p \right) \left( \frac{(\text{vec } \Sigma^*)'}{\partial t} \otimes \Sigma^* \otimes \Sigma^* \right)(I_{p^2} \otimes I_p) \right\} (H_p \otimes H_p)'.
\]

Substituting (4.11) and simplifying yields

\[
C^{(4)}(t) = \frac{32}{n^3} (H_p \otimes H_p) N_p^2 \{ \text{vec } \Sigma^* \otimes (\text{vec } \Sigma^*)' \otimes \text{vec } \Sigma^* \otimes \Sigma^* \}(H_p \otimes H_p)'.
\]

(4.15)

The matrix $\Sigma^*$, evaluated at $t = 0$, simplifies to $\Sigma$.

Suppose $w$ is a random vector with $E(w) = \sigma$. It can be shown that

\[
C^{(1)}(0) = E(w) \quad \text{and} \quad C^{(2)}(0) = E \{ (w - \sigma)(w - \sigma)' \}.
\]

Equivalently,

\[
C^{(2)}(0) = H_p E[\text{vec}(S - \Sigma)\{\text{vec}(S - \Sigma)\}'] H_p'.
\]

(4.16)

where $w = \text{vech } S = H_p \text{vec } S$. Pre-multiplying (4.13) by $D_p$ and post-multiplying by $D_p$ and using (4.16) yield

\[
E \{ \text{vec}(S - \Sigma)\{\text{vec}(S - \Sigma)\}' \} = \frac{2}{n} N_p \Sigma \otimes \Sigma = \text{var}(\text{vec } S),
\]

(4.17)
because $\mathbf{N}_p \text{vec}(\mathbf{S} - \Sigma) = \text{vec}(\mathbf{S} - \Sigma)$ and $\mathbf{N}_p(\Sigma \otimes \Sigma) = (\Sigma \otimes \Sigma) \mathbf{N}_p$. The third cumulant of $\mathbf{w}$ can be written as

$$C^{(3)}(0) = \mathbb{E} \{ (\mathbf{w} - \mathbf{\sigma})' \otimes (\mathbf{w} - \mathbf{\sigma})(\mathbf{w} - \mathbf{\sigma})' \}.$$  

Equivalently,

$$C^{(3)}(0) = \mathbf{H}_p \mathbb{E} \left[ \{ \text{vec}(\mathbf{S} - \Sigma) \}' \otimes \text{vec}(\mathbf{S} - \Sigma) \{ \text{vec}(\mathbf{S} - \Sigma) \}' \right] (\mathbf{H}_p \otimes \mathbf{H}_p)',$$  

where $\mathbf{w} = \text{vech} \mathbf{S} = \mathbf{H}_p \text{vec} \mathbf{S}$. Pre-multiplying (4.14) by $\mathbf{D}_p$ and post-multiplying by $(\mathbf{D}_p \otimes \mathbf{D}_p)'$ and using (4.18) yield

$$\mathbb{E} \left[ \{ \text{vec}(\mathbf{S} - \Sigma) \}' \otimes \text{vec}(\mathbf{S} - \Sigma) \{ \text{vec}(\mathbf{S} - \Sigma) \}' \right] = \frac{8}{n^2} \mathbf{N}_p \{ \Sigma \otimes (\text{vec} \Sigma)' \otimes \Sigma \} \times (\mathbf{N}_p \otimes \mathbf{N}_p).$$  

Similarly, the fourth cumulant of vech $\mathbf{S}$ can be written as

$$C^{(4)}(0) = (\mathbf{H}_p \otimes \mathbf{H}_p) \left[ \mathbb{E} \left\{ \{ \text{vec}(\mathbf{S} - \Sigma) \{ \text{vec}(\mathbf{S} - \Sigma) \}' \otimes \text{vec}(\mathbf{S} - \Sigma) \{ \text{vec}(\mathbf{S} - \Sigma) \}' \right\} ight. \\
- 2 \mathbf{N}_p^2 \left[ \mathbb{E} \{ \text{vec}(\mathbf{S} - \Sigma) \{ \text{vec}(\mathbf{S} - \Sigma) \}' \} \otimes \mathbb{E} \{ \text{vec}(\mathbf{S} - \Sigma) \{ \text{vec}(\mathbf{S} - \Sigma) \}' \} \right] \\
\left. - \mathbb{E} \{ \text{vec}(\mathbf{S} - \Sigma) \otimes \text{vec}(\mathbf{S} - \Sigma) \} \mathbb{E} \{ \{ \text{vec}(\mathbf{S} - \Sigma) \}' \otimes \{ \text{vec}(\mathbf{S} - \Sigma) \}' \} \right] \\
\times (\mathbf{H}_p \otimes \mathbf{H}_p)'.$$  

Pre-multiplying (4.15) by $(\mathbf{D}_p \otimes \mathbf{D}_p)$ and post-multiplying by $(\mathbf{D}_p \otimes \mathbf{D}_p)'$ and using (4.20) yields
\[
E \left[ \text{vec}(S - \Sigma) \{\text{vec}(S - \Sigma)\}' \otimes \text{vec}(S - \Sigma) \{\text{vec}(S - \Sigma)\}' \right] = \\
\frac{32}{n^3} N_p^2 (N_p \otimes N_p) \{\Sigma \otimes (\text{vec } \Sigma)' \otimes \text{vec } \Sigma \otimes \Sigma \} (N_p \otimes N_p) + \frac{16}{n^3} (N_p \otimes N_p) (I_p \otimes I_{(p,p)} \otimes I_p) \{\Sigma \otimes \Sigma \otimes \Sigma \otimes \Sigma \} (N_p \otimes N_p) \\
+ 2N_p \left[ E \{\text{vec}(S - \Sigma) \{\text{vec}(S - \Sigma)\}' \} \otimes E \{\text{vec}(S - \Sigma) \{\text{vec}(S - \Sigma)\}' \} \right] \\
+ \left[ E \{\text{vec}(S - \Sigma) \otimes \text{vec}(S - \Sigma) \} \right] \left[ E \{\text{vec}(S - \Sigma) \}' \otimes \{\text{vec}(S - \Sigma) \}' \right]. \tag{4.21}
\]

To compute the last part of (4.21), the second derivative of \( C^{(0)}(t) \) with respect to \( t \) evaluated at \( t = 0 \) can be obtained as

\[
\frac{\partial^2 C^{(0)}(t)}{\partial t \otimes \partial t} = \frac{2}{n} (H_p \otimes H_p) \text{vec } (\Sigma \otimes \Sigma). \tag{4.22}
\]

Algebraically,

\[
\frac{\partial^2 C^{(0)}(t)}{\partial t \otimes \partial t} = E \{\text{vec}(S - \Sigma) \otimes \text{vec}(S - \Sigma) \}. \tag{4.23}
\]

Pre-multiplying (4.22) by \((D_p \otimes D_p)\) and using (4.23) yields

\[
E \{\text{vec}(S - \Sigma) \otimes \text{vec}(S - \Sigma) \} = \frac{2}{n} (N_p \otimes N_p) \text{vec } (\Sigma \otimes \Sigma). \tag{4.24}
\]

Similarly,

\[
E \left[ \{\text{vec}(S - \Sigma)\}' \otimes \{\text{vec}(S - \Sigma)\}' \right] = \frac{2}{n} \{\text{vec } (\Sigma \otimes \Sigma)\}' (N_p \otimes N_p). \tag{4.25}
\]

Substituting (4.24) and (4.25) in (4.21) yields

\[
E \left[ \text{vec}(S - \Sigma) \{\text{vec}(S - \Sigma)' \otimes \text{vec}(S - \Sigma) \{\text{vec}(S - \Sigma)' \right] = \\
\frac{32}{n^3} N_p^2 (N_p \otimes N_p) \{\Sigma \otimes (\text{vec } \Sigma)' \otimes \text{vec } \Sigma \otimes \Sigma \} (N_p \otimes N_p)
\]
\[
\begin{align*}
&+ \frac{16}{n^3} (N_p \otimes N_p) (I_p \otimes I_{(p,p)} \otimes I_p) (\Sigma \otimes \Sigma \otimes \Sigma \otimes \Sigma) (N_p \otimes N_p) \\
&+ \frac{8}{n^2} N_p^2 (N_p \otimes N_p) (\Sigma \otimes \Sigma \otimes \Sigma \otimes \Sigma) (N_p \otimes N_p) \\
&+ \frac{4}{n^2} (N_p \otimes N_p) \text{vec}(\Sigma \otimes \Sigma) \{\text{vec}(\Sigma \otimes \Sigma)\}'(N_p \otimes N_p). \quad (4.26)
\end{align*}
\]

The results of the third and fourth order moments of vec(\(S - \Sigma\)) are used for computing the expectation of \(D(\hat{\theta}; \theta)\) in (4.7).

**Expectation of LRT Statistic**

Partition \(\theta\) as \(\theta = (\mu', \varphi')'\) with dimension \((\nu_1, \nu_2)'\), where \(\nu = \nu_1 + \nu_2\). In particular, \(k = 2\). Denote the derivatives of vec \(\Sigma\) with respect to \(\theta\) as follows:

\[
\begin{align*}
F^{(1)} &= \frac{\partial \text{vec} \Sigma}{\partial \theta'} = \sum_{s=1}^{k} \frac{\partial \text{vec} \Sigma}{\partial \theta''_s} E_{s,\nu}; \\
F^{(2)} &= \frac{\partial^2 \text{vec} \Sigma}{\partial \theta' \otimes \theta'} = \sum_{s=1}^{k} \sum_{t=1}^{k} \frac{\partial^2 \text{vec} \Sigma}{\partial \theta''_s \otimes \partial \theta''_t} (E_{s,\nu} \otimes E_{t,\nu}); \\
F^{(11)} &= \frac{\partial^2 \text{vec} \Sigma}{\partial \theta' \otimes \theta'} = \text{dvec} \left( F^{(2)}, p^2 \nu, \nu \right); \\
F^{(3)} &= \frac{\partial^3 \text{vec} \Sigma}{\partial \theta' \otimes \partial \theta' \otimes \partial \theta'} = \sum_{s=1}^{k} \sum_{t=1}^{k} \sum_{u=1}^{k} \frac{\partial^3 \text{vec} \Sigma}{\partial \theta''_s \otimes \partial \theta''_t \otimes \partial \theta''_u} (E_{s,\nu} \otimes E_{t,\nu} \otimes E_{u,\nu}); \\
F^{(111)} &= \text{dvec} \left( F^{(3)}, p^2 \nu^2, \nu \right).
\end{align*}
\]

The matrix \(E_{i,\nu}\) with dimension \(\nu_i \times \nu\) for \(1 \leq i \leq k\) is defined as

\[
E_{i,\nu} = \begin{bmatrix} 0_{\nu_i \times a_i} & I_{\nu_i} & 0_{\nu_i \times (\nu-a_i - \nu_j)} \end{bmatrix}.
\]

where \(a_i = \sum_{j=1}^{i} \nu_j - \nu_i\). For convenience, define \(\tilde{F}^{(1)}, \tilde{F}^{(2)}, \ldots^{(1)}\), and \(\ldots^{(2)}\) as

\[
\begin{align*}
\tilde{F}^{(1)} &= (\Sigma^{-1} \otimes \Sigma^{-1}) F^{(1)}; \\
\tilde{F}^{(2)} &= (\Sigma^{-1} \otimes \Sigma^{-1}) F^{(2)}; \\
\ddots^{(1)} &= (\Sigma^{-1} \otimes \text{vec} \Sigma^{-1} \otimes \Sigma^{-1}) F^{(1)}; \quad \text{and} \quad \ddots^{(2)} = (\Sigma^{-1} \otimes \text{vec} \Sigma^{-1} \otimes \Sigma^{-1}) F^{(2)}.
\end{align*}
\]
Boik (2002a, in a supplement) derived the expectation of the right-hand-side of the equation (4.7) as follows:

\[
E\{D(\tilde{\theta}; \theta)\} = \dim(\theta) + \frac{1}{n} \sum_{i=1}^{10} \zeta_i + O(n^{-2}),
\]

(4.27)

where

\[
\zeta_1 = \text{tr}\left\{ \left( \bar{I}_{\theta}^{-1} \tilde{F}^{(1)} \otimes N_p \right) \left( I_p \otimes \text{vec} \Sigma \otimes I_p \right) F^{(1)} \bar{I}_{\theta}^{-1} U^{(1)} \right\};
\]

\[
U^{(1)} = F^{(1)} \left( I_\nu \otimes \Sigma^{-1} \otimes \Sigma^{-1} \right) - 2 \tilde{F}^{(1)} \left( F^{(1)} \otimes I_p^2 \right);
\]

\[
\zeta_2 = \frac{1}{3} \text{tr}\left\{ \left( F^{(1)} \bar{I}_{\theta}^{-1} \otimes F^{(1)} \bar{I}_{\theta}^{-1} \right) \tilde{F}^{(1)} \bar{I}_{\theta}^{-1} K^*_3 \right\};
\]

\[
K^*_3 = 2 \tilde{F}^{(1)} \left( F^{(1)} \otimes F^{(1)} \right) - \text{dvec} \left( \tilde{F}^{(1)} \left( F^{(2)} \left( I_\nu, I_\nu \right), I_\nu, I_\nu \right)' - \frac{1}{2} \tilde{F}^{(1)} \left( F^{(2)} \right) \right);
\]

\[
\zeta_3 = \frac{1}{2} \text{tr}\left\{ \left[ I_{\theta}^{-1} \otimes N_p \left( \Sigma \otimes \Sigma \right) \right] U^{(1)} \bar{I}_{\theta}^{-1} U^{(1)} \right\};
\]

\[
\zeta_4 = L_1' \bar{I}_{\theta}^{-1} L + L_2' \left( I_\nu \otimes \bar{I}_{\theta}^{-1} \right) L_2;
\]

\[
L_1 = \frac{1}{2} U^{(1)} \text{vec} \left( F^{(1)} \bar{I}_{\theta}^{-1} \right); \quad L_2 = \frac{1}{2} \text{vec} \left\{ U^{(1)} \left( I_\nu \otimes F^{(1)} \bar{I}_{\theta}^{-1} \right) \right\};
\]

\[
\zeta_5 = \frac{1}{2} \text{tr} \left\{ N_\nu \left( \bar{I}_{\theta}^{-1} \otimes \bar{I}_{\theta}^{-1} \right) K^*_3 \bar{I}_{\theta}^{-1} K^*_3 \right\} + \frac{1}{4} \left( \text{vec} \bar{I}_{\theta}^{-1} \right)' K^*_3 \bar{I}_{\theta}^{-1} K^*_3 \text{vec} \bar{I}_{\theta}^{-1};
\]

\[
\zeta_6 = \text{tr} \left\{ N_\nu \left( \bar{I}_{\theta}^{-1} \otimes \bar{I}_{\theta}^{-1} F^{(1)} \right) U^{(1)} \bar{I}_{\theta}^{-1} K^*_3 \right\};
\]

\[
\zeta_7 = \frac{1}{2} \left( \text{vec} \bar{I}_{\theta}^{-1} \right)' K^*_3 \bar{I}_{\theta}^{-1} U^{(1)} \text{vec} \left( F^{(1)} \bar{I}_{\theta}^{-1} \right);
\]

\[
\zeta_8 = \frac{1}{3} \text{tr} \left\{ N_\nu \left( \bar{I}_{\theta}^{-1} F^{(1)} \otimes \bar{I}_{\theta}^{-1} \right) \tilde{U}^{(2)} \right\};
\]

\[
\tilde{U}^{(2)} = -6 \text{dvec} \left( \tilde{F}^{(2)} , p^2 \nu, p^2 \right) F^{(2)}
\]

\[
+ \text{dvec} \left[ \text{dvec} \left\{ \left( \Sigma^{-1} \otimes \Sigma^{-1} \right) F^{(3)} , p^2 \nu , p^2 \nu \right\} , p^2 \nu , p^2 \right] + 2 \left( I_p \otimes F^{(1)} \right)' \left( I_p \otimes \text{vec} \Sigma^{-1} \otimes I_p \right) \left( 2 N_p + I_{(p,p)} \right) L_3;
\]
\[ L_3 = \left( \Sigma^{-1} \otimes \text{vec} \Sigma^{-1} \otimes I_p \right)^\prime \left( F^{(1)} \otimes F^{(1)} \right); \]

\[ \zeta_9 = \frac{1}{6} \left\{ \text{vec} \left( I_\theta^{-1} F^{(1)} \right)^\prime \tilde{U}^{(2)} \left( \text{vec} I_\theta^{-1} \right) \right\}; \]

\[ \zeta_{10} = \frac{1}{6} \text{tr} \left\{ N_v \left( I_\theta^{-1} \otimes I_\theta^{-1} \right) K_4^* \right\} + \frac{1}{12} \left( \text{vec} I_\theta^{-1} \right)^\prime K_4^* \text{vec} I_\theta^{-1}; \quad \text{and} \]

\[ K_4^* = 12 \left( F^{(1)} \otimes F^{(1)} \right)^\prime \tilde{F}^{(2)} - 9 L_3^* I_{(p,p)} L_3 - 2 \text{dvec} \left( \tilde{F}^{(1)y} F^{(3)}, \nu^2, \nu^2 \right) - \frac{3}{2} F^{(2)} \tilde{F}^{(2)}. \]

**Under a Restricted H_a**

To compute the Bartlett-corrected test statistic in (4.8), the expectation in (4.27) is used twice to obtain the right-hand-side of (4.9).

**Under an Unrestricted H_a**

To compute the Bartlett-corrected test statistic in (4.8), the expectation in (4.27) is used once to obtain \( \hat{E} \left\{ D(\hat{\theta}; \theta_0) | H_0 \right\} \) in (4.9). To obtain \( \hat{E} \left\{ D(\hat{\theta}_a; \theta_a) | H_0 \right\} \), Boik (2002a) showed that if \( \Sigma \) is unrestricted under H_a, then the asymptotic distribution of Q in (4.2) is \( \chi^2_{\nu_\Delta} \), where \( \nu_\Delta = \frac{\nu_{p+1}}{2} - \nu_0 \). Furthermore,

\[ \hat{E} \left\{ D(\hat{\theta}_a; \theta_a) | H_0 \right\} = \frac{p(p+1)}{2} + \frac{p(2p^2 + 3p - 1)}{12n} + O(n^{-2}). \quad (4.28) \]

To verify equation (4.28), it follows from (4.2) that

\[ \hat{E} \left\{ D(\hat{\theta}_a; \theta_a) | H_0 \right\} = 2 \hat{E} \left\{ \ell(\hat{\theta}_a) - \ell(\theta_a) \right\} \]

\[ = 2 \hat{E} \left\{ \frac{n}{2} \ln |S| + \frac{n}{2} \ln |\Sigma| \right\} \]

\[ = 2 \hat{E} \left\{ \frac{n}{2} \ln |n S| + \frac{np}{2} \ln(n) + \frac{n}{2} \ln |\Sigma| \right\}. \]
Seber (1984) showed that $|nS| \sim |\Sigma| \prod_{i=1}^{p} \chi_{n-i+1}^2$. Accordingly,

$$
\hat{E} \left\{ D(\hat{\theta}; \theta_a) | H_0 \right\} = np \ln(n) - n \sum_{i=1}^{p} \hat{E} \left( \ln \chi_{n-i+1}^2 \right).
$$

(4.29)

The expectation of $\ln(y)$, $E\{\ln(y)\}$, assuming $y \sim \chi_n^2$, can be obtained by differentiating both sides of

$$
\Gamma \left( \frac{n}{2} \right) = \int_0^\infty y^{\frac{3}{2}-1} \exp\left\{-\frac{y}{2}\right\} \frac{dy}{2^\frac{3}{2}}
$$

(4.30)

with respect to $n$, where $\Gamma(\cdot)$ is the gamma function. The useful result concerning the digamma ($\psi$) function is as follows:

$$
\psi(n) = \frac{\partial \ln \Gamma(n)}{\partial n} = \frac{\Gamma'(n)}{\Gamma(n)}.
$$

Then, the result of differentiating both sides of (4.30) yields

$$
\psi \left( \frac{n}{2} \right) = \int_0^\infty \ln(y) \frac{y^{\frac{3}{2}-1} \exp\left\{-\frac{y}{2}\right\}}{2^{\frac{3}{2}}} \frac{dy}{\Gamma \left( \frac{n}{2} \right)} - \ln(2).
$$

Therefore,

$$
E\{\ln(y)\} = \psi \left( \frac{n}{2} \right) + \ln(2).
$$

Expand the digamma function to obtain

$$
\psi \left( \frac{n}{2} \right) + \ln(2) = \ln(n) - \frac{1}{n} - \frac{1}{3n^2} + \cdots
$$

(4.31)

Applying result (4.31) to (4.29), simplification yields the result (4.28).
Matrix Expressions for $Z_i$ and $K_i$

To obtain matrix expressions for $Z_i$ and $K_i$, the second and third derivatives of vec $\Sigma$ with respect to $\theta'$ evaluated at $\mu = 0$ are needed. The following properties are repeatedly used:

$$
(ABC \otimes D)E = [A \otimes \{\text{vec}(C')\}]' \otimes D\{\text{vec}(B') \otimes E\}
$$

and

$$
(A \otimes BCD)E = \{A \otimes \text{vec}(D') \otimes B\}(E \otimes \text{vec} C),
$$

where $A$, $B$, $C$, $D$, and $E$ are any matrices that are conformable for multiplication.

It follows from (3.4) that

$$
\frac{\partial^2 \text{vec} \Sigma}{\partial \mu' \otimes \partial \mu'} = 2 N_p \left[ (\Gamma G \Lambda \otimes \Gamma) \frac{\partial^2 \text{vec} G}{\partial \mu' \otimes \partial \mu'} + \frac{\partial}{\partial \mu'} \{ \Gamma G \Lambda \otimes \Gamma \} \times \left( I_{\nu_1} \otimes \frac{\partial \text{vec} G}{\partial \mu'} \right) \right].
$$

Applying (4.33) yields

$$
\frac{\partial^2 \text{vec} \Sigma}{\partial \mu' \otimes \partial \mu'} = 2 N_p \left[ (\Gamma G \Lambda \otimes \Gamma) \frac{\partial^2 \text{vec} G}{\partial \mu' \otimes \partial \mu'} + \{ \Gamma \otimes (\text{vec} \Lambda)' \otimes \Gamma \} \left( I_{(p,p)} \frac{\partial \text{vec} G}{\partial \mu'} \otimes \frac{\partial \text{vec} G}{\partial \mu'} \right) \right].
$$

Therefore, evaluated at $\mu = 0$,

$$
\left. \frac{\partial^2 \text{vec} \Sigma}{\partial \mu' \otimes \partial \mu'} \right|_{\mu=0} = 2 N_p \left[ (\Gamma \Lambda \otimes \Gamma) \, D_G^{(2)} + \{ \Gamma \otimes (\text{vec} \Lambda)' \otimes \Gamma \} \left( I_{(p,p)} \, D_G^{(1)} \otimes D_G^{(1)} \right) \right].
$$
Also, it follows from (3.4) that
\[
\frac{\partial^2 \text{vec } \Sigma}{\partial \varphi' \otimes \partial \mu'} = 2 N_p \frac{\partial}{\partial \varphi'} \left\{ (\Gamma G \Lambda \otimes \Gamma) \frac{\partial \text{vec } G}{\partial \mu'} \right\}
\]
\[
= 2 N_p \frac{\partial}{\partial \varphi'} \left\{ (\Gamma \otimes (\text{vec } I_p)' \otimes \Gamma) \left( \text{vec } (G \Lambda)' \otimes \frac{\partial \text{vec } G}{\partial \mu'} \right) \right\} \text{ by (4.32)}.
\]
Therefore, evaluated at \( \mu = 0 \),
\[
\frac{\partial^2 \text{vec } \Sigma}{\partial \varphi' \otimes \partial \mu'} \bigg|_{\mu = 0} = 2 N_p (\Gamma \otimes (\text{vec } I_p)' \otimes \Gamma) \left\{ LD^{(1)}_\lambda \otimes D^{(1)}_G \right\}.
\]
(4.35)

The second derivative of \( \text{vec } \Sigma \) with respect to \( \varphi' \) can be written as
\[
\frac{\partial^2 \text{vec } \Sigma}{\partial \varphi' \otimes \partial \varphi'} \bigg|_{\mu = 0} = (\Gamma \otimes \Gamma) L D^{(2)}_\lambda.
\]

Similarly, the third derivative of \( \text{vec } \Sigma \) with respect to \( \varphi' \) can be written as
\[
\frac{\partial^3 \text{vec } \Sigma}{\partial \varphi' \otimes \partial \varphi' \otimes \partial \varphi'} \bigg|_{\mu = 0} = (\Gamma \otimes \Gamma) L D^{(3)}_\lambda.
\]

Next, it follows from (4.35) that
\[
\frac{\partial^3 \text{vec } \Sigma}{\partial \varphi' \otimes \partial \varphi' \otimes \partial \mu'} \bigg|_{\mu = 0} = 2 N_p \{ \Gamma \otimes (\text{vec } I_p)' \otimes \Gamma \} \left\{ LD^{(2)}_\lambda \otimes D^{(1)}_G \right\}.
\]

It follows from (4.34) that
\[
\frac{\partial^3 \text{vec } \Sigma}{\partial \varphi' \otimes \partial \mu' \otimes \partial \mu'} = 2 N_p \left[ \frac{\partial}{\partial \varphi'} \left\{ (\Gamma G \Lambda \otimes \Gamma) \frac{\partial^2 \text{vec } G}{\partial \mu' \otimes \partial \mu'} \right\} \right.
\]
\[
+ \frac{\partial}{\partial \varphi'} \{ \Gamma \otimes (\text{vec } A)' \otimes \Gamma \} \left( I_{(p,p)} \frac{\partial \text{vec } G}{\partial \mu'} \otimes \frac{\partial \text{vec } G}{\partial \mu'} \right) \].
\]
(4.36)

For convenience, the two parts on the right-hand-side of (4.36) are labeled as \( E_a \) and \( E_b \). Using (4.33) and evaluating at \( \mu = 0 \), \( E_a \) can be simplified to
\[
E_a = (\Gamma \otimes (\text{vec } I_p)' \otimes \Gamma) \left\{ LD^{(1)}_\lambda \otimes D^{(2)}_G \right\}.
\]
(4.37)
The quantity $E_b$, evaluated at $\mu = 0$, can be written as

$$
E_b = \frac{\partial}{\partial \varphi'} \left\{ (\Gamma \otimes (\text{vec } \Lambda)') \otimes \Gamma \right\} \left( I_{(p^1, p^2)} I_{(p, p^2)} \otimes I_{p} \right) \left( I_{(p, p)} \mathbf{D}^{(1)}_G \otimes \mathbf{D}^{(1)}_G \right)
$$

$$
= \frac{\partial}{\partial \varphi'} \left\{ ((\text{vec } \Lambda)' \otimes \Gamma \otimes \Gamma) \left( I_{(p, p^2)} \otimes I_{p} \right) \left( I_{(p, p)} \mathbf{D}^{(1)}_G \otimes \mathbf{D}^{(1)}_G \right) \right\}
$$

$$
= \left( \frac{\partial (\text{vec } \Lambda)'}{\partial \varphi'} \otimes \Gamma \otimes \Gamma \right) \left[ I_{\nu_2} \otimes \left( I_{(p, p^2)} \otimes I_{p} \right) \left( I_{(p, p)} \mathbf{D}^{(1)}_G \otimes \mathbf{D}^{(1)}_G \right) \right]
$$

$$
= \left( \left\{ \text{vec } (\Lambda \mathbf{D}^{(1)}_\lambda) \right\}' \otimes \Gamma \otimes \Gamma \right) \left[ I_{\nu_2} \otimes \left( I_{p} \otimes I_{(p, p)} \otimes I_{p} \right) \left( \mathbf{D}^{(1)}_G \otimes \mathbf{D}^{(1)}_G \right) \right].
$$

Combining (4.37) and (4.38) yields

$$
\frac{\partial^3 \text{vec } \Sigma}{\partial \varphi' \otimes \partial \mu' \otimes \partial \mu'} \bigg|_{\mu = 0} = 2 N_p \left\{ (\Gamma \otimes (\text{vec } I_p)') \otimes \Gamma \right\} \left\{ L \mathbf{D}^{(1)}_\lambda \otimes \mathbf{D}^{(2)}_G \right\}
$$

$$
+ \left\{ \left( \text{vec } L \mathbf{D}^{(1)}_\lambda \right) \otimes \Gamma \otimes \Gamma \right\}
$$

$$
\times \left[ I_{\nu_2} \otimes \left( I_{p} \otimes I_{(p, p)} \otimes I_{p} \right) \left( \mathbf{D}^{(1)}_G \otimes \mathbf{D}^{(1)}_G \right) \right].
$$

Next, it follows from (4.34) that

$$
\frac{\partial^3 \text{vec } \Sigma}{\partial \mu' \otimes \partial \mu' \otimes \partial \mu'} = 2 N_p \left\{ \frac{\partial}{\partial \mu'} \left\{ (\Gamma G \Lambda \otimes \Gamma) \frac{\partial^2 \text{vec } G}{\partial \mu' \otimes \partial \mu'} \right\} \right\}
$$

$$
+ \frac{\partial}{\partial \mu'} \left\{ (\Gamma \otimes (\text{vec } \Lambda)' \otimes \Gamma) \left( I_{(p, p)} \frac{\partial \text{vec } G}{\partial \mu'} \otimes \frac{\partial \text{vec } G}{\partial \mu'} \right) \right\}.
$$

(4.39)

For convenience, the two parts on the right-hand-side of (4.39) are labeled as $E_c$ and $E_d$. Using the product rule and (4.33) and evaluating at $\mu = 0$ yield

$$
E_c = \{ \Gamma \otimes (\text{vec } \Lambda)' \otimes \Gamma \} \left( I_{(p, p)} \mathbf{D}^{(1)}_G \otimes \mathbf{D}^{(2)}_G \right) + (\Gamma \Lambda \otimes \Gamma) \mathbf{D}^{(3)}_G.
$$

(4.40)

Using the Kronecker product rule, $E_d$, evaluated at $\mu = 0$, simplifies to

$$
E_d = 2 N_p \left\{ \Gamma \otimes (\text{vec } \Lambda)' \otimes \Gamma \right\} \left( I_{(p, p)} \mathbf{D}^{(2)}_G \otimes \mathbf{D}^{(1)}_G \right)
$$

$$
+ I_{(p^1, p^2)} \left( \mathbf{D}^{(2)}_G \otimes I_{(p, p)} \mathbf{D}^{(1)}_G \right) \left( I_{\nu_1} \otimes I_{(\nu_1, \nu_1)} \right).
$$
Using the following results repeatedly:

\[(A \otimes B) = I_{(p,r)}(B \otimes A)I_{(c,q)},\]

where \(A\) is \(r \times c\) and \(B\) is \(p \times q\), yields

\[
E_d = 2N_p \left\{ (\Gamma \otimes (\text{vec} \Lambda)' \otimes \Gamma) \left( I_{(p,p)} D^{(2)}_G \otimes D^{(1)}_G \right) \left( I_{\nu_1^0} + (I_{\nu_1} \otimes I_{(p_1,p_1)}) \right) \right\}
\]

\[
= 4N_p \left\{ (\Gamma \otimes (\text{vec} \Lambda)' \otimes \Gamma) \left( I_{(p,p)} D^{(2)}_G \otimes D^{(1)}_G \right) \left( I_{\nu_1} \otimes N_{\nu_1} \right) \right\}, \quad (4.41)
\]

where \(N_{\nu_1} = \frac{1}{2} (I_{\nu_1^2} + I_{(p_1,p_1)})\). Combining (4.40) and (4.41) yields

\[
\frac{\partial^3 \text{vec} \Sigma}{\partial \mu' \otimes \partial \mu' \otimes \partial \mu'} \bigg|_{\mu=0} = 2N_p \left[ (\Gamma \Lambda \otimes \Gamma) D^{(3)}_G + \{ \Gamma \otimes (\text{vec} \Lambda)' \otimes \Gamma \} \right.
\]

\[
\times \left\{ \left( I_{(p,p)} D^{(1)}_G \otimes D^{(2)}_G \right) + 2 \left( I_{(p,p)} D^{(2)}_G \otimes D^{(1)}_G \right) \left( I_{\nu_1} \otimes N_{\nu_1} \right) \right\}. \]

The other order derivatives can be rearranged by post-multiplying by commutation matrices.

\[
\frac{\partial^2 \text{vec} \Sigma}{\partial \theta_i \otimes \partial \theta_s} = \frac{\partial^2 \text{vec} \Sigma}{\partial \theta_i' \otimes \partial \theta_s'} \text{vec} \left( I_{(p,p)} \right)
\]

\[
\frac{\partial^3 \text{vec} \Sigma}{\partial \theta_i' \otimes \partial \theta_s' \otimes \partial \theta_u'} = \frac{\partial^3 \text{vec} \Sigma}{\partial \theta_i' \otimes \partial \theta_s' \otimes \partial \theta_u'} \text{vec} \left( I_{(p,p)} \right)
\]

The following properties are useful for simplifying expressions.

\[
F^{(1)} = I_{(p,p)} F^{(1)}
\]

\[
F^{(2)} = I_{(p,p)} F^{(2)} = F^{(2)} I_{(p,p)}
\]

\[
F^{(11)} = \left( I_{\nu} \otimes I_{(p,p)} \right) F^{(11)}
\]
106

\[ F^{(3)} = I_{(p,p)} F^{(3)} = F^{(3)} I_{(\nu,\nu^2)} = F^{(3)} I_{(\nu^2,\nu)} \]

\[ = F^{(3)} (I_{\nu} \otimes I_{(\nu,\nu)}) = F^{(3)} (I_{(\nu,\nu)} \otimes I_{\nu}) \]

\[ F^{(111)} = (I_{(p,p)} \otimes I_{(\nu,\nu^2)}) F^{(111)} = (I_{p^2} \otimes I_{(\nu,\nu)}) F^{(111)} \]

\[ F^{(1)} (\Sigma^{-1} \otimes \Sigma^{-1}) F^{(1)} = 2 I_\theta \]

\[ \ddot{F}^{(1)} (\Sigma \otimes \Sigma) \ddot{F}^{(1)} = 2 \ddot{I}_\theta \]

\[ (\Sigma^{-1} \otimes \Sigma^{-1}) \text{vec} \Sigma = \text{vec} \Sigma^{-1} \]

\[ \frac{\partial \text{vec} \Sigma^{-1}}{\partial \theta} = - (\Sigma^{-1} \otimes \Sigma^{-1}) F^{(1)} = \ddot{F}^{(1)} \]

It follows from (4.4) that \( K_1 = 0 \). Then

\[ Z_1 = \frac{\sqrt{n}}{2} F^{(1)} (\Sigma^{-1} \otimes \Sigma^{-1}) \text{vec}(S - \Sigma) \]

\[ = \frac{\sqrt{n}}{2} \ddot{F}^{(1)} \text{vec}(S - \Sigma). \]

The second derivatives of \( \ell(\theta) \) can be written as

\[ \ell_2(\theta) = \frac{n}{2} F^{(1)} \left\{ I_{\nu} \otimes \left( \Sigma^{-1} \otimes \Sigma^{-1} \right) \text{vec}(S - \Sigma) \right\} \]

\[ + n F^{(1)} \left\{ I_{p} \otimes \left( \text{vec} I_{p} \right)' \otimes \Sigma^{-1} \right\} \left( \frac{\partial \text{vec} \Sigma^{-1}}{\partial \theta} \otimes \text{vec}(S - \Sigma) \right) \]

\[ - \frac{n}{2} F^{(1)} \left( \Sigma^{-1} \otimes \Sigma^{-1} \right) F^{(1)} \]

Accordingly,

\[ \ell_2(\theta) = \frac{n}{2} F^{(1)} \left\{ I_{\nu} \otimes \left( \Sigma^{-1} \otimes \Sigma^{-1} \right) \text{vec}(S - \Sigma) \right\} \]

\[ - n \left( \Sigma^{-1} \otimes \left( \text{vec} \Sigma^{-1} \right)' \otimes \Sigma^{-1} \right) \left( F^{(1)} \otimes \text{vec}(S - \Sigma) \right) \]

\[ - \frac{n}{2} F^{(1)} \left( \Sigma^{-1} \otimes \Sigma^{-1} \right) F^{(1)}. \]
Taking the expectation of $\ell_2(\theta)$ and dividing by $n$ yields

$$K_2 = -\frac{1}{2} F^{(1)'} (\Sigma^{-1} \otimes \Sigma^{-1}) F^{(1)} = -\frac{1}{2} \tilde{F}^{(1)'} F^{(1)} = -\tilde{I}_\theta.$$ 

Therefore,

$$Z_2 = \frac{\sqrt{n}}{2} F^{(1)'} \{ I_\nu \otimes (\Sigma^{-1} \otimes \Sigma^{-1}) \text{vec}(S - \Sigma) \} - \sqrt{n} \tilde{F}^{(1)'} \{ F^{(1)} \otimes \text{vec}(S - \Sigma) \}. $$

Boik (2002a, in supplement) gave expressions for $Z_3$, $K_3$, and $K_4$ as follows:

$$Z_3 = \sqrt{n} \left( \text{vec} \tilde{F}^{(1)} \otimes \tilde{F}^{(1)} \right)' \left( I_{p\nu} \otimes I_{(p,p)} \otimes I_p \right) \left\{ I_\nu \otimes F^{(1)} \otimes \text{vec}(S - \Sigma) \right\}$$

$$+ 2\sqrt{n} \frac{\tilde{F}^{(1)}}{2} \left( F^{(1)} \otimes N_\nu (I_p \otimes \text{vec} \Sigma^{-1} \otimes I_p) \right) \left\{ F^{(1)} \otimes \text{vec}(S - \Sigma) \right\}$$

$$+ \sqrt{n} \frac{\tilde{F}^{(1)\prime}}{2} \left\{ I_{\nu^2} \otimes (\Sigma^{-1} \otimes \Sigma^{-1}) \text{vec}(S - \Sigma) \right\}$$

$$- 2\sqrt{n} \tilde{F}^{(1)} \left( I_\nu \otimes \Sigma^{-1} \otimes \text{vec} \Sigma^{-1} \otimes \Sigma^{-1} \right)' \left\{ (I_\nu \otimes F^{(1)}) N_\nu \otimes \text{vec}(S - \Sigma) \right\}$$

$$- \sqrt{n} \tilde{F}^{(1)\prime} \left\{ F^{(2)} \otimes \text{vec}(S - \Sigma) \right\};$$

$$K_3 = 2 \tilde{F}^{(1)\prime} \left( F^{(1)} \otimes F^{(1)} \right) N_\nu - \tilde{F}^{(1)\prime} \left( I_\nu \otimes \tilde{F}^{(1)} \right) N_\nu - \frac{1}{2} \tilde{F}^{(1)\prime} \tilde{F}^{(2)};$$

$$K_4 = \tilde{F}^{(1)\prime} \left\{ I_\nu \otimes (\Sigma^{-1} \otimes \text{vec} \Sigma^{-1} \otimes \Sigma^{-1})' (F^{(1)} \otimes F^{(1)}) \right\}$$

$$\times \left\{ 4(N_\nu \otimes I_\nu) + 2 I_{(\nu^2,\nu)} \right\} + 4 \tilde{F}^{(1)\prime} \left( F^{(2)} \otimes F^{(1)} \right) \left\{ (I_\nu \otimes N_\nu) + \frac{1}{2} I_{(\nu,\nu^2)} \right\}$$

$$- F^{(1)\prime} \left( \Sigma^{-1} \otimes \text{vec} \Sigma^{-1} \otimes \text{vec} \Sigma^{-1} \otimes \Sigma^{-1} \right) \left( F^{(1)} \otimes F^{(1)} \otimes F^{(1)} \right)$$

$$\times \left\{ 2 I_{(\nu^2,\nu)} + 3 I_{(\nu^2,\nu)} + 4(N_\nu \otimes I_\nu) \right\}$$

$$- \tilde{F}^{(1)\prime} \left( I_{\nu^2} \otimes \tilde{F}^{(1)} \right) \left\{ (I_\nu \otimes N_\nu) + \frac{1}{2} I_{(\nu,\nu^2)} \right\}$$

$$- \tilde{F}^{(1)\prime} \left( I_\nu \otimes \tilde{F}^{(2)} \right) \left\{ (N_\nu \otimes I_\nu) + \frac{1}{2} I_{(\nu^2,\nu)} \right\} - \frac{1}{2} \tilde{F}^{(1)\prime} \tilde{F}^{(3)};$$

where $N_\nu = \frac{1}{2}(I_{\nu^2} + I_{(\nu,\nu)})$. 
CHAPTER 5

SIMULATION

In this chapter, the performance of the likelihood ratio test and Bartlett correction (see Chapter 4) are evaluated and compared to competing tests. Simulation studies were conducted under 2 applications described in Chapter 2.

Simulation Study of Variable Reduction (Redundancy) Tests

The evaluation is based on test size and power of the test under different conditions. The test statistics examined were

- Likelihood ratio test (LRT)
- Bartlett correction of LRT (BART)
- Schott’s (1991) test \( T_{k,q} \) in (1.16)
- Bartlett correction of \( T_{k,q} \) (\( T^* \)) in (1.18)
- Bartlett correction of \( T_{k,q} \) (\( T^{**} \)), where \( T^{**} = \left( \frac{T_{k,q}}{1+c} \right) \). An expression for \( c \) can be found in Schott (1991).

The conditions under which the simulations were conducted are described in terms of the number of original variables \( (p) \), the number of retained components \( (k) \), the number of variables to be tested for redundancy \( (q) \), and the degrees of freedom of a Wishart distribution \( (n) \). The following conditions were considered:

- \( p = 10, \ k = 2 \) and \( q = 1, 4 \)
• $p = 10$, $k = 4$ and $q = 2,5$

• $p = 15$, $k = 2$ and $q = 1$

• $p = 15$, $k = 4$ and $q = 2$.

The degrees of freedom, $n$, varied from 25 to 75 in the following manner,

\[ n \in \{20, 25, 30, 35, 50, 75\}. \]

The nominal test size used was 0.05. For each simulation condition, Type I error probability was computed from 1000 data sets. The eigenvalues were ordered as follows: $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_p$. For simplicity, the covariance matrices examined were diagonal and have $\lambda_1 = \cdots = \lambda_{p-k} = 1$.

The Comparison Test

Partition $\Gamma$ as

\[
\Gamma = \begin{pmatrix}
\Gamma_{11} & \Gamma_{12} \\
\Gamma_{21} & \Gamma_{22} \\
\Gamma_{31} & \Gamma_{32}
\end{pmatrix},
\]

Consider the special case where $\Gamma_{31} = 0$. In this section a test of hypothesis (1.21), where $A = \begin{pmatrix} I_q \\ 0 \end{pmatrix}$ and $M$ is the first $m$ columns of $I_p$, will be developed. Although this test is not useful in practice (because $\Gamma_{31} = 0$ is not likely to be satisfied), it is useful in improving the efficiency of the simulations by constructing an estimator of test size whose variance is less than or equal the usual estimator.

Suppose that it is known that $\Gamma_{31} = 0$, $A = \begin{pmatrix} I_q \\ 0 \end{pmatrix}$, and $M = \begin{pmatrix} I_m \\ 0 \end{pmatrix}$. Then,

\[
A \in \mathcal{R}(\Gamma M) \iff \Gamma' A \in \mathcal{R}(M),
\]
\[
\left( \begin{array}{ccc}
\Gamma'_{11} & \Gamma'_{21} & \Gamma'_{31} \\
\Gamma'_{12} & \Gamma'_{22} & \Gamma'_{32} \\
0 & 0 & 0
\end{array} \right) \in \mathcal{R} \left( \begin{array}{c} I_q \\ 0 \\
0 \end{array} \right)
\]

\[
\iff 
\left( \begin{array}{c} \Gamma'_{11} \\
\Gamma'_{12} \\
0
\end{array} \right) \in \mathcal{R} \left( \begin{array}{c} I_m \\
0 \\
0
\end{array} \right)
\]

\[
\iff \Gamma_{12} = 0.
\]

Let \( Y \) be an \( n \times p \) matrix with distribution \( \text{vec} Y \sim N(0, \Sigma \otimes I_n) \). Then, \( S = \frac{Y^T Y}{n} \sim W_p(n, \Sigma) \) and \( \Sigma = \Gamma \Lambda \Gamma' \). Partition \( Y \) as \( Y = \begin{pmatrix} Y_1 & Y_2 & Y_3 \end{pmatrix} \). If it is known that \( \Gamma_{31} = 0 \), then

\[
A \in \mathcal{R}(\Gamma M) \iff \Sigma = \begin{pmatrix} \Gamma_{11} & 0 & 0 \\
\Gamma_{21} & \Lambda_1 & \Gamma_{11}' \\
0 & 0 & \Lambda_2 \\
\end{pmatrix} \begin{pmatrix} \Lambda_1 & 0 & \Gamma_{21}' \\
0 & \Lambda_2 & \Gamma_{32}' \\
\end{pmatrix}
\]

\[
\iff \Sigma = \begin{pmatrix} \Gamma_{11} \Lambda_1 \Gamma_{11}' \\
\Gamma_{21} \Lambda_1 \Gamma_{11}' + \Gamma_{22} \Lambda_1 \Gamma_{21}' + \Gamma_{32} \Lambda_2 \Gamma_{32}' \\
0 \\
\end{pmatrix}
\]

\[
\iff Y_1 \perp Y_3.
\]

Accordingly, a test of independence between \( Y_1 \) and \( Y_3 \) is a test of \( H_0: A \in \mathcal{R}(\Gamma M) \)

if \( \Gamma_{31} = 0 \), \( A = \begin{pmatrix} I_q \\
0 \end{pmatrix} \), and \( M = \begin{pmatrix} I_m \\
0 \end{pmatrix} \). Partition \( S \) as

\[
S = \left( \begin{array}{ccc}
S_{11} & S_{12} & S_{13} \\
S_{21} & S_{22} & S_{23} \\
S_{31} & S_{32} & S_{33}
\end{array} \right)
\]

The test statistic is

\[
\Lambda_w = \frac{|S_{11} S_{13}|}{|S_{11}| |S_{33}|},
\]

which is distributed as Wilk’s \( \Lambda \), \( \Lambda_{q,k,n-k} \) where \( k = p - m \). An approximate \( F \)-statistic (Rencher, 1995) is given by

\[
F = \frac{(1 - \Lambda_w^{1/k}) df_2}{\Lambda_w^{1/k} df_1},
\]

\text{(5.1)}
which has an approximate \( F \)-distribution with \( df_1 \) and \( df_2 \) degrees of freedom, where

\[
df_1 = qk,
\]

\[
df_2 = wt - \frac{1}{2}qk + 1,
\]

\[
w = n - \frac{1}{2}(q + k + 1),
\]

\[
t = \sqrt{\frac{q^2k^2 - 4}{q^2 + k^2 - 5}}, \quad \text{and}
\]

when \( qk = 2 \) then \( t \) is set equal to 1. If \( \min(q, k) \) is equal to either 1 or 2, then the \( F \) approximation in (5.1) has an exact \( F \)-distribution. Anderson (1984) gave the asymptotic distribution of \( M = -s \log(A_{q,k,n-k}) \), where \( s = (n - k) - \frac{1}{2}(q - k + 1) \).

The error is of the order \( O(N^{-6}) \). Box (1949) provided the coefficients to be used in term of \( O(N^{-6}) \). The cdf of \( M \) can be written as

\[
P(M \leq m_0) = P_{qk} + \frac{\gamma_2}{s^2} (P_{qk+4} - P_{qk})
\]

\[
+ \frac{1}{s^4} \left[ \gamma_4 (P_{qk+8} - P_{qk}) - \frac{\gamma_2^2}{2} (P_{qk+4} - P_{qk}) \right]
\]

\[
+ \frac{1}{s^6} \left[ \gamma_6 (P_{qk+12} - P_{qk}) - \frac{\gamma_2^2}{2} (P_{qk+8} - P_{qk+4}) \right] + O(N^{-8}),
\]

where \( P_f \) is \( P \left( \chi_f^2 \leq m_0 \right) \),

\[
\gamma_2 = \alpha_2 = \frac{qk}{48} \left( q^2 + k^2 - 5 \right),
\]

\[
\gamma_4 = \frac{\alpha_2^2}{2} + \frac{qk}{1920} \left\{ 3q^4 + 3k^4 + 10q^2k^2 - 50 (q^2 + k^2) + 159 \right\}, \quad \text{and}
\]

\[
\gamma_6 = \frac{\alpha_2^3}{6} + \frac{qk}{16,128} \left\{ (q^6 + k^6) - 105 (q^4 + k^4) + 1,113 (q^2 + k^2) 
\]

\[
+ (21q^2 - 350 + 21k^2) q^2k^2 - 2,995 \right\}. 
\]
Variance Reduction Procedure

Consider two test statistics, namely the target test, $L$, and the comparison test, $F$. Suppose there are two outcomes of testing any hypothesis: rejection of the null hypothesis denoted by 1 and failure to reject the null hypothesis denoted by 0. Let $n_{ij}$ denote the number of observations that belong to the outcomes in the $i^{th}$ row and the $j^{th}$ column of the $L$ and $F$ tests. The $2 \times 2$ table of the number of observations can be written as follows.

<table>
<thead>
<tr>
<th></th>
<th>$F$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>$n_{11}$</td>
<td>$n_{12}$</td>
</tr>
<tr>
<td>0</td>
<td>$n_{21}$</td>
<td>$n_{22}$</td>
</tr>
<tr>
<td>Total</td>
<td>$n_{-1}$</td>
<td>$n_{-2}$</td>
</tr>
</tbody>
</table>

Let $P_{ij}$ be the probability of an observation in the $i^{th}$ row and the $j^{th}$ column. Let $\alpha$ and $\alpha^*$ be the sizes of the $F$ and $L$ tests, respectively. Because the comparison test is exact (or nearly so), it is known that $E(n_{-1}) = \alpha N$, whereas the expectation of $n_{1.}$ is unknown. The likelihood function, conditional on $n_{-1}$, can be written as

$$L(P|n_{-1}) = \left( \frac{n_{-1}}{n_{11}} \right) \left( \frac{n_{-2}}{n_{12}} \right) \left( \frac{P_{11}}{\alpha} \right)^{n_{11}} \left( 1 - \frac{P_{11}}{\alpha} \right)^{n_{1.}-n_{11}} \left( \frac{P_{22}}{1 - \alpha} \right)^{n_{-2}-n_{12}} \left( 1 - \frac{P_{22}}{1 - \alpha} \right)^{n_{12}},$$

where $\alpha = P_{11} + P_{21}$ and $1 - \alpha = P_{12} + P_{22}$.
Then, the loglikelihood function, conditional on \( n_{-1} \), is

\[
\ell(P|n_{-1}) = n_{11} \ln \left( \frac{P_{11}}{\alpha} \right) + (n_{-1} - n_{11}) \ln \left( 1 - \frac{P_{11}}{\alpha} \right) + (n_{2} - n_{12}) \ln \left( \frac{P_{22}}{1 - \alpha} \right) + n_{12} \ln \left( 1 - \frac{P_{22}}{1 - \alpha} \right) + \text{constants.}
\]

The MLE of \( P_{11} \), \( \hat{P}_{11} \), can be obtained by solving

\[
\frac{\partial \ell(P|n_{-1})}{\partial P_{11}} = 0.
\]

Therefore,

\[
\hat{P}_{11} = \frac{n_{11}}{n_{-1}} \alpha.
\]

Similarly, by solving \( \frac{\partial \ell(P|n_{-1})}{\partial P_{12}} = 0 \), the MLE of \( P_{12} \), \( \hat{P}_{12} \) is

\[
\hat{P}_{12} = \frac{n_{12}}{n_{-2}} (1 - \alpha).
\]

Accordingly, an estimator of \( \alpha^* \), say \( \tilde{\alpha} \), is

\[
\tilde{\alpha} = \hat{P}_{11} + \hat{P}_{12} = \frac{n_{11}}{n_{-1}} \alpha + \frac{n_{12}}{n_{-2}} (1 - \alpha).
\]

The expectation and variance of \( \tilde{\alpha} \), conditional on \( n_{-1} \), are

\[
E (\tilde{\alpha}|n_{-1}) = P_{11} + P_{12} = \alpha^* \quad \text{and}
\]

\[
\text{var} (\tilde{\alpha}|n_{-1}) = \frac{\alpha}{n_{-1}} P_{11} \left( 1 - \frac{P_{11}}{\alpha} \right) + \frac{1 - \alpha}{n_{-2}} P_{12} \left( 1 - \frac{P_{12}}{1 - \alpha} \right).
\]

Accordingly,

\[
\text{\( \widehat{\text{var}} \)} (\tilde{\alpha}) = E [\text{var} (\tilde{\alpha}|n_{-1})].
\]
Because \( n_1 \sim B(N, \alpha) \) and \( n_2 \sim B(N, 1 - \alpha) \), then using

\[
E(n_1^{-1}) \approx \frac{1}{\alpha N} \quad \text{and} \quad E(n_2^{-1}) \approx \frac{1}{(1 - \alpha)N}
\]
yields

\[
\hat{\text{var}}(\hat{\alpha}) \approx \frac{\alpha^*(1 - \alpha^*)}{N} (1 - \varphi^2),
\]

where \( \alpha^* = P_{11} + P_{12} \) and

\[
\varphi^2 = \frac{(n_{11}n_{22} - n_{12}n_{21})^2}{n_{11}n_{22}n_{12}n_{21}}.
\]

Let \( \hat{\alpha} = \frac{n}{N} \) be the usual estimator of \( \alpha^* \). Then,

\[
\hat{\text{var}}(\hat{\alpha}) = \frac{\alpha^*(1 - \alpha^*)}{N}.
\]

The relative efficiency of \( \tilde{\alpha} \) to \( \hat{\alpha} \) is given by

\[
\frac{\hat{\text{var}}(\hat{\alpha})}{\text{var}(\tilde{\alpha})} = \frac{1}{(1 - \varphi^2)}.
\]

Therefore, the effective sample size, \( N^* \), is

\[
N^* = \frac{N}{(1 - \varphi^2)}.
\]

All computations were executed using Matlab programs. Details including the programming code are given in Appendices. The test statistic \( T^* \) in (1.18) was programmed using Schott’s (1991) expression for \( c \). However, the value of \( c \), using the data from Example 1, is not identical to his result for the same data set and this
leads to different value of $T^*$. The program was checked thoroughly and no errors were found. Professor Schott was asked about the discrepancy and he responded that he have been unable to locate his code and thus this thesis should proceed under the assumption that the program used is correct.

The following tables report the results of simulations. A cell in each table presents the following quantities:

$$
\begin{pmatrix}
\hat{\alpha} \\
\tilde{\alpha} \\
N^*
\end{pmatrix}
$$

for a test with nominal $\alpha = 0.05$.  

Table 5. Probabilities of Type I error when \((p, k, q) = (10, 2, 4)\) and \(\lambda = (1', 10, 25)'\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>(LRT)</th>
<th>(BART)</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.0610</td>
<td>0.0440</td>
<td>0.0470</td>
<td>0.2690</td>
<td>0.0700</td>
<td>0.1170</td>
</tr>
<tr>
<td></td>
<td>0.0625</td>
<td>0.0452</td>
<td></td>
<td>0.2713</td>
<td>0.0728</td>
<td>0.1198</td>
</tr>
<tr>
<td></td>
<td>1263</td>
<td>1205</td>
<td></td>
<td>1154</td>
<td>2479</td>
<td>1592</td>
</tr>
<tr>
<td>30</td>
<td>0.0820</td>
<td>0.0540</td>
<td>0.0590</td>
<td>0.1780</td>
<td>0.0750</td>
<td>0.0970</td>
</tr>
<tr>
<td></td>
<td>0.0763</td>
<td>0.0497</td>
<td></td>
<td>0.1701</td>
<td>0.0662</td>
<td>0.0884</td>
</tr>
<tr>
<td></td>
<td>1420</td>
<td>1339</td>
<td></td>
<td>1407</td>
<td>4410</td>
<td>2402</td>
</tr>
<tr>
<td>50</td>
<td>0.0560</td>
<td>0.0480</td>
<td>0.0520</td>
<td>0.1100</td>
<td>0.0580</td>
<td>0.0620</td>
</tr>
<tr>
<td></td>
<td>0.0546</td>
<td>0.0468</td>
<td></td>
<td>0.1081</td>
<td>0.0560</td>
<td>0.0600</td>
</tr>
<tr>
<td></td>
<td>1895</td>
<td>1629</td>
<td></td>
<td>1797</td>
<td>9164</td>
<td>5877</td>
</tr>
<tr>
<td>75</td>
<td>0.0510</td>
<td>0.0430</td>
<td>0.0450</td>
<td>0.0860</td>
<td>0.0560</td>
<td>0.0590</td>
</tr>
<tr>
<td></td>
<td>0.0546</td>
<td>0.0465</td>
<td></td>
<td>0.0908</td>
<td>0.0609</td>
<td>0.0639</td>
</tr>
<tr>
<td></td>
<td>1829</td>
<td>2045</td>
<td></td>
<td>2003</td>
<td>4861</td>
<td>4024</td>
</tr>
</tbody>
</table>

Table 6. Probabilities of Type I error when \((p, k, q) = (10, 2, 4)\) and \(\lambda = (1', 5, 10)'\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>(LRT)</th>
<th>(BART)</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.0680</td>
<td>0.0410</td>
<td>0.0470</td>
<td>0.3050</td>
<td>0.0470</td>
<td>0.1190</td>
</tr>
<tr>
<td></td>
<td>0.0695</td>
<td>0.0421</td>
<td></td>
<td>0.3072</td>
<td>0.0491</td>
<td>0.1216</td>
</tr>
<tr>
<td></td>
<td>1201</td>
<td>1198</td>
<td></td>
<td>1126</td>
<td>1896</td>
<td>1494</td>
</tr>
<tr>
<td>30</td>
<td>0.0750</td>
<td>0.0490</td>
<td>0.0590</td>
<td>0.2050</td>
<td>0.0610</td>
<td>0.0990</td>
</tr>
<tr>
<td></td>
<td>0.0694</td>
<td>0.0441</td>
<td></td>
<td>0.1974</td>
<td>0.0535</td>
<td>0.0907</td>
</tr>
<tr>
<td></td>
<td>1450</td>
<td>1539</td>
<td></td>
<td>1321</td>
<td>3096</td>
<td>2120</td>
</tr>
<tr>
<td>50</td>
<td>0.0570</td>
<td>0.0460</td>
<td>0.0520</td>
<td>0.1170</td>
<td>0.0640</td>
<td>0.0730</td>
</tr>
<tr>
<td></td>
<td>0.0556</td>
<td>0.0447</td>
<td></td>
<td>0.1151</td>
<td>0.0621</td>
<td>0.0710</td>
</tr>
<tr>
<td></td>
<td>1863</td>
<td>1966</td>
<td></td>
<td>1706</td>
<td>4340</td>
<td>3295</td>
</tr>
<tr>
<td>75</td>
<td>0.0560</td>
<td>0.0490</td>
<td>0.0510</td>
<td>0.0960</td>
<td>0.0620</td>
<td>0.0630</td>
</tr>
<tr>
<td></td>
<td>0.0553</td>
<td>0.0483</td>
<td></td>
<td>0.0950</td>
<td>0.0611</td>
<td>0.0621</td>
</tr>
<tr>
<td></td>
<td>1837</td>
<td>1881</td>
<td></td>
<td>2024</td>
<td>3944</td>
<td>3756</td>
</tr>
</tbody>
</table>
Table 7. Probabilities of Type I error when \((p, k, q) = (10, 2, 4)\) and \(\lambda = (1^*, 2, 10)^*\).

<table>
<thead>
<tr>
<th></th>
<th>(LRT)</th>
<th>(BART)</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.1520</td>
<td>0.0460</td>
<td>0.0550</td>
<td>0.5460</td>
<td>0.0320</td>
<td>0.1920</td>
</tr>
<tr>
<td></td>
<td>0.1495</td>
<td>0.0441</td>
<td></td>
<td>0.5437</td>
<td>0.0311</td>
<td>0.1894</td>
</tr>
<tr>
<td></td>
<td>1108</td>
<td>1199</td>
<td></td>
<td>1046</td>
<td>1056</td>
<td>1103</td>
</tr>
<tr>
<td>30</td>
<td>0.1200</td>
<td>0.0490</td>
<td>0.0580</td>
<td>0.4140</td>
<td>0.0370</td>
<td>0.1530</td>
</tr>
<tr>
<td></td>
<td>0.1156</td>
<td>0.0458</td>
<td></td>
<td>0.4093</td>
<td>0.0351</td>
<td>0.1481</td>
</tr>
<tr>
<td></td>
<td>1185</td>
<td>1238</td>
<td></td>
<td>1083</td>
<td>1092</td>
<td>1183</td>
</tr>
<tr>
<td>50</td>
<td>0.0910</td>
<td>0.0520</td>
<td>0.0520</td>
<td>0.2630</td>
<td>0.0550</td>
<td>0.1080</td>
</tr>
<tr>
<td></td>
<td>0.0898</td>
<td>0.0511</td>
<td></td>
<td>0.2614</td>
<td>0.0541</td>
<td>0.1066</td>
</tr>
<tr>
<td></td>
<td>1289</td>
<td>1257</td>
<td></td>
<td>1181</td>
<td>1236</td>
<td>1306</td>
</tr>
<tr>
<td>75</td>
<td>0.0670</td>
<td>0.0420</td>
<td>0.0450</td>
<td>0.1760</td>
<td>0.0580</td>
<td>0.0770</td>
</tr>
<tr>
<td></td>
<td>0.0703</td>
<td>0.0447</td>
<td></td>
<td>0.1802</td>
<td>0.0610</td>
<td>0.0808</td>
</tr>
<tr>
<td></td>
<td>1411</td>
<td>1446</td>
<td></td>
<td>1264</td>
<td>1378</td>
<td>1530</td>
</tr>
</tbody>
</table>

Table 8. Probabilities of Type I error when \((p, k, q) = (10, 2, 4)\) and \(\lambda = (1^*, 1.5, 10)^*\).

<table>
<thead>
<tr>
<th></th>
<th>(LRT)</th>
<th>(BART)</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.1690</td>
<td>0.0390</td>
<td>0.0470</td>
<td>0.6550</td>
<td>0.0400</td>
<td>0.2530</td>
</tr>
<tr>
<td></td>
<td>0.1701</td>
<td>0.0395</td>
<td></td>
<td>0.6560</td>
<td>0.0401</td>
<td>0.2540</td>
</tr>
<tr>
<td></td>
<td>1048</td>
<td>1031</td>
<td></td>
<td>1023</td>
<td>1001</td>
<td>1027</td>
</tr>
<tr>
<td>30</td>
<td>0.1560</td>
<td>0.0490</td>
<td>0.0440</td>
<td>0.5680</td>
<td>0.0620</td>
<td>0.2200</td>
</tr>
<tr>
<td></td>
<td>0.1590</td>
<td>0.0513</td>
<td></td>
<td>0.5706</td>
<td>0.0630</td>
<td>0.2226</td>
</tr>
<tr>
<td></td>
<td>1087</td>
<td>1146</td>
<td></td>
<td>1032</td>
<td>1022</td>
<td>1048</td>
</tr>
<tr>
<td>50</td>
<td>0.1350</td>
<td>0.0450</td>
<td>0.0470</td>
<td>0.4500</td>
<td>0.0750</td>
<td>0.1860</td>
</tr>
<tr>
<td></td>
<td>0.1366</td>
<td>0.0457</td>
<td></td>
<td>0.4516</td>
<td>0.0758</td>
<td>0.1873</td>
</tr>
<tr>
<td></td>
<td>1119</td>
<td>1065</td>
<td></td>
<td>1054</td>
<td>1052</td>
<td>1057</td>
</tr>
<tr>
<td>75</td>
<td>0.1160</td>
<td>0.0450</td>
<td>0.0440</td>
<td>0.3860</td>
<td>0.0740</td>
<td>0.1400</td>
</tr>
<tr>
<td></td>
<td>0.1194</td>
<td>0.0474</td>
<td></td>
<td>0.3893</td>
<td>0.0762</td>
<td>0.1427</td>
</tr>
<tr>
<td></td>
<td>1152</td>
<td>1190</td>
<td></td>
<td>1056</td>
<td>1094</td>
<td>1075</td>
</tr>
</tbody>
</table>
Table 9. Probabilities of Type I error when \((p, k, q) = (10, 2, 1)\) and \(\lambda = (1, 8, 10, 25)\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>LRT</th>
<th>BART</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.0660</td>
<td>0.0480</td>
<td>0.0470</td>
<td>0.0870</td>
<td>0.0600</td>
<td>0.0650</td>
</tr>
<tr>
<td></td>
<td>0.0681</td>
<td>0.0498</td>
<td></td>
<td>0.0899</td>
<td>0.0630</td>
<td>0.0679</td>
</tr>
<tr>
<td></td>
<td>1528</td>
<td>1537</td>
<td></td>
<td>2072</td>
<td>4398</td>
<td>3441</td>
</tr>
<tr>
<td>30</td>
<td>0.0640</td>
<td>0.0530</td>
<td>0.0480</td>
<td>0.0760</td>
<td>0.0540</td>
<td>0.0560</td>
</tr>
<tr>
<td></td>
<td>0.0655</td>
<td>0.0545</td>
<td></td>
<td>0.0779</td>
<td>0.0560</td>
<td>0.0580</td>
</tr>
<tr>
<td></td>
<td>1804</td>
<td>1953</td>
<td></td>
<td>2584</td>
<td>8568</td>
<td>6664</td>
</tr>
<tr>
<td>50</td>
<td>0.0530</td>
<td>0.0500</td>
<td>0.0510</td>
<td>0.0680</td>
<td>0.0560</td>
<td>0.0560</td>
</tr>
<tr>
<td></td>
<td>0.0522</td>
<td>0.0492</td>
<td></td>
<td>0.0670</td>
<td>0.0550</td>
<td>0.0550</td>
</tr>
<tr>
<td></td>
<td>2340</td>
<td>2369</td>
<td></td>
<td>3796</td>
<td>7614</td>
<td>7614</td>
</tr>
<tr>
<td>75</td>
<td>0.0520</td>
<td>0.0500</td>
<td>0.0500</td>
<td>0.0580</td>
<td>0.0520</td>
<td>0.0520</td>
</tr>
<tr>
<td></td>
<td>0.0520</td>
<td>0.0500</td>
<td></td>
<td>0.0580</td>
<td>0.0520</td>
<td>0.0520</td>
</tr>
<tr>
<td></td>
<td>2967</td>
<td>3241</td>
<td></td>
<td>6887</td>
<td>24700</td>
<td>24700</td>
</tr>
</tbody>
</table>

Table 10. Probabilities of Type I error when \((p, k, q) = (10, 2, 1)\) and \(\lambda = (1, 8, 5, 10)\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>LRT</th>
<th>BART</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.0590</td>
<td>0.0400</td>
<td>0.0530</td>
<td>0.0860</td>
<td>0.0580</td>
<td>0.0620</td>
</tr>
<tr>
<td></td>
<td>0.0575</td>
<td>0.0388</td>
<td></td>
<td>0.0833</td>
<td>0.0555</td>
<td>0.0595</td>
</tr>
<tr>
<td></td>
<td>1316</td>
<td>1258</td>
<td></td>
<td>2098</td>
<td>2785</td>
<td>2476</td>
</tr>
<tr>
<td>30</td>
<td>0.0650</td>
<td>0.0550</td>
<td>0.0570</td>
<td>0.0700</td>
<td>0.0540</td>
<td>0.0540</td>
</tr>
<tr>
<td></td>
<td>0.0598</td>
<td>0.0502</td>
<td></td>
<td>0.0635</td>
<td>0.0480</td>
<td>0.0480</td>
</tr>
<tr>
<td></td>
<td>1988</td>
<td>1947</td>
<td></td>
<td>3505</td>
<td>4310</td>
<td>4310</td>
</tr>
<tr>
<td>50</td>
<td>0.0560</td>
<td>0.0540</td>
<td>0.0490</td>
<td>0.0650</td>
<td>0.0580</td>
<td>0.0580</td>
</tr>
<tr>
<td></td>
<td>0.0568</td>
<td>0.0548</td>
<td></td>
<td>0.0660</td>
<td>0.0590</td>
<td>0.0590</td>
</tr>
<tr>
<td></td>
<td>2670</td>
<td>2617</td>
<td></td>
<td>3863</td>
<td>5024</td>
<td>5024</td>
</tr>
<tr>
<td>75</td>
<td>0.0560</td>
<td>0.0490</td>
<td>0.0520</td>
<td>0.0580</td>
<td>0.0560</td>
<td>0.0560</td>
</tr>
<tr>
<td></td>
<td>0.0542</td>
<td>0.0474</td>
<td></td>
<td>0.0561</td>
<td>0.0541</td>
<td>0.0541</td>
</tr>
<tr>
<td></td>
<td>3935</td>
<td>2806</td>
<td></td>
<td>6890</td>
<td>6704</td>
<td>6704</td>
</tr>
</tbody>
</table>
Table 11. Probabilities of Type I error when \((p, k, q) = (10, 2, 1)\) and \(\lambda = (1^t, 2, 10)^t\).

<table>
<thead>
<tr>
<th>n</th>
<th>LRT</th>
<th>BART</th>
<th>(F)</th>
<th>(T_{kq})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.0970</td>
<td>0.0560</td>
<td>0.0530</td>
<td>0.1250</td>
<td>0.0730</td>
<td>0.0790</td>
</tr>
<tr>
<td></td>
<td>0.0960</td>
<td>0.0552</td>
<td></td>
<td>0.1231</td>
<td>0.0717</td>
<td>0.0775</td>
</tr>
<tr>
<td></td>
<td>1069</td>
<td>1080</td>
<td></td>
<td>1236</td>
<td>1168</td>
<td>1202</td>
</tr>
<tr>
<td>30</td>
<td>0.0890</td>
<td>0.0580</td>
<td>0.0470</td>
<td>0.1020</td>
<td>0.0660</td>
<td>0.0710</td>
</tr>
<tr>
<td></td>
<td>0.0909</td>
<td>0.0594</td>
<td></td>
<td>0.1044</td>
<td>0.0675</td>
<td>0.0727</td>
</tr>
<tr>
<td></td>
<td>1270</td>
<td>1226</td>
<td></td>
<td>1432</td>
<td>1234</td>
<td>1286</td>
</tr>
<tr>
<td>50</td>
<td>0.0790</td>
<td>0.0540</td>
<td>0.0560</td>
<td>0.0780</td>
<td>0.0710</td>
<td>0.0720</td>
</tr>
<tr>
<td></td>
<td>0.0751</td>
<td>0.0508</td>
<td></td>
<td>0.0727</td>
<td>0.0665</td>
<td>0.0674</td>
</tr>
<tr>
<td></td>
<td>1451</td>
<td>1408</td>
<td></td>
<td>2336</td>
<td>1849</td>
<td>1905</td>
</tr>
<tr>
<td>75</td>
<td>0.0570</td>
<td>0.0480</td>
<td>0.0510</td>
<td>0.0630</td>
<td>0.0560</td>
<td>0.0580</td>
</tr>
<tr>
<td></td>
<td>0.0563</td>
<td>0.0473</td>
<td></td>
<td>0.0622</td>
<td>0.0553</td>
<td>0.0572</td>
</tr>
<tr>
<td></td>
<td>2003</td>
<td>1919</td>
<td></td>
<td>2393</td>
<td>2043</td>
<td>2208</td>
</tr>
</tbody>
</table>

Table 12. Probabilities of Type I error when \((p, k, q) = (10, 2, 1)\) and \(\lambda = (1^t, 1.5, 10)^t\).

<table>
<thead>
<tr>
<th>n</th>
<th>LRT</th>
<th>BART</th>
<th>(F)</th>
<th>(T_{kq})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.1210</td>
<td>0.0710</td>
<td>0.0520</td>
<td>0.1520</td>
<td>0.0850</td>
<td>0.0940</td>
</tr>
<tr>
<td></td>
<td>0.1204</td>
<td>0.0705</td>
<td></td>
<td>0.1509</td>
<td>0.0844</td>
<td>0.0934</td>
</tr>
<tr>
<td></td>
<td>1049</td>
<td>1040</td>
<td></td>
<td>1130</td>
<td>1050</td>
<td>1057</td>
</tr>
<tr>
<td>30</td>
<td>0.1220</td>
<td>0.0700</td>
<td>0.0530</td>
<td>0.1270</td>
<td>0.0880</td>
<td>0.1050</td>
</tr>
<tr>
<td></td>
<td>0.1205</td>
<td>0.0686</td>
<td></td>
<td>0.1252</td>
<td>0.0870</td>
<td>0.1038</td>
</tr>
<tr>
<td></td>
<td>1138</td>
<td>1199</td>
<td></td>
<td>1197</td>
<td>1071</td>
<td>1087</td>
</tr>
<tr>
<td>50</td>
<td>0.1080</td>
<td>0.0740</td>
<td>0.0490</td>
<td>0.1220</td>
<td>0.0920</td>
<td>0.089</td>
</tr>
<tr>
<td></td>
<td>0.1086</td>
<td>0.0745</td>
<td></td>
<td>0.1226</td>
<td>0.0924</td>
<td>0.0894</td>
</tr>
<tr>
<td></td>
<td>1224</td>
<td>1206</td>
<td></td>
<td>1203</td>
<td>1085</td>
<td>1101</td>
</tr>
<tr>
<td>75</td>
<td>0.0910</td>
<td>0.0640</td>
<td>0.0400</td>
<td>0.0920</td>
<td>0.0670</td>
<td>0.0660</td>
</tr>
<tr>
<td></td>
<td>0.0971</td>
<td>0.0698</td>
<td></td>
<td>0.0989</td>
<td>0.0728</td>
<td>0.0716</td>
</tr>
<tr>
<td></td>
<td>1207</td>
<td>1280</td>
<td></td>
<td>1275</td>
<td>1261</td>
<td>1238</td>
</tr>
</tbody>
</table>
Table 13. Probabilities of Type I error when \((p, k, q) = (10, 4, 2)\) and \(\lambda = (1, 10, 10, 10, 25)'\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>(LRT)</th>
<th>(BART)</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.1350</td>
<td>0.0610</td>
<td>0.0480</td>
<td>0.2910</td>
<td>0.0700</td>
<td>0.1240</td>
</tr>
<tr>
<td></td>
<td>0.1362</td>
<td>0.0620</td>
<td></td>
<td>0.2925</td>
<td>0.0718</td>
<td>0.1258</td>
</tr>
<tr>
<td></td>
<td>1151</td>
<td>1228</td>
<td></td>
<td>1140</td>
<td>2397</td>
<td>1553</td>
</tr>
<tr>
<td>30</td>
<td>0.0870</td>
<td>0.0530</td>
<td>0.0430</td>
<td>0.1850</td>
<td>0.0550</td>
<td>0.0750</td>
</tr>
<tr>
<td></td>
<td>0.0920</td>
<td>0.0570</td>
<td></td>
<td>0.1910</td>
<td>0.0614</td>
<td>0.0818</td>
</tr>
<tr>
<td></td>
<td>1354</td>
<td>1374</td>
<td></td>
<td>1246</td>
<td>2960</td>
<td>2243</td>
</tr>
<tr>
<td>50</td>
<td>0.0710</td>
<td>0.0400</td>
<td>0.0390</td>
<td>0.1150</td>
<td>0.0500</td>
<td>0.0580</td>
</tr>
<tr>
<td></td>
<td>0.0787</td>
<td>0.0460</td>
<td></td>
<td>0.1251</td>
<td>0.0606</td>
<td>0.0688</td>
</tr>
<tr>
<td></td>
<td>1385</td>
<td>1409</td>
<td></td>
<td>1454</td>
<td>3703</td>
<td>2933</td>
</tr>
<tr>
<td>75</td>
<td>0.0640</td>
<td>0.0450</td>
<td>0.0400</td>
<td>0.0920</td>
<td>0.0490</td>
<td>0.0510</td>
</tr>
<tr>
<td></td>
<td>0.0717</td>
<td>0.0521</td>
<td></td>
<td>0.1015</td>
<td>0.0586</td>
<td>0.0609</td>
</tr>
<tr>
<td></td>
<td>1604</td>
<td>1812</td>
<td></td>
<td>1698</td>
<td>4286</td>
<td>4450</td>
</tr>
</tbody>
</table>

Table 14. Probabilities of Type I error when \((p, k, q) = (10, 4, 2)\) and \(\lambda = (1, 5, 5, 10, 10)'\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>(LRT)</th>
<th>(BART)</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.1620</td>
<td>0.0630</td>
<td>0.0480</td>
<td>0.3380</td>
<td>0.0620</td>
<td>0.1400</td>
</tr>
<tr>
<td></td>
<td>0.1631</td>
<td>0.0639</td>
<td></td>
<td>0.3394</td>
<td>0.0635</td>
<td>0.1418</td>
</tr>
<tr>
<td></td>
<td>1124</td>
<td>1173</td>
<td></td>
<td>1109</td>
<td>1857</td>
<td>1418</td>
</tr>
<tr>
<td>30</td>
<td>0.0950</td>
<td>0.0490</td>
<td>0.0430</td>
<td>0.2130</td>
<td>0.0510</td>
<td>0.0810</td>
</tr>
<tr>
<td></td>
<td>0.0999</td>
<td>0.0522</td>
<td></td>
<td>0.2188</td>
<td>0.0566</td>
<td>0.0870</td>
</tr>
<tr>
<td></td>
<td>1309</td>
<td>1228</td>
<td></td>
<td>1199</td>
<td>2175</td>
<td>1700</td>
</tr>
<tr>
<td>50</td>
<td>0.0710</td>
<td>0.0460</td>
<td>0.0390</td>
<td>0.1290</td>
<td>0.0540</td>
<td>0.0670</td>
</tr>
<tr>
<td></td>
<td>0.0784</td>
<td>0.0519</td>
<td></td>
<td>0.1390</td>
<td>0.0639</td>
<td>0.0774</td>
</tr>
<tr>
<td></td>
<td>1346</td>
<td>1330</td>
<td></td>
<td>1377</td>
<td>2500</td>
<td>2148</td>
</tr>
<tr>
<td>75</td>
<td>0.0550</td>
<td>0.0490</td>
<td>0.0400</td>
<td>0.1000</td>
<td>0.0510</td>
<td>0.0530</td>
</tr>
<tr>
<td></td>
<td>0.0727</td>
<td>0.0558</td>
<td></td>
<td>0.1094</td>
<td>0.0606</td>
<td>0.0626</td>
</tr>
<tr>
<td></td>
<td>1588</td>
<td>1610</td>
<td></td>
<td>1600</td>
<td>3773</td>
<td>3398</td>
</tr>
</tbody>
</table>
Table 15. Probabilities of Type I error when \((p, k, q) = (10, 4, 2)\) and \(\lambda = (1'_{0.6}, 2, 5, 10, 10')\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>(LRT)</th>
<th>(BART)</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.2140</td>
<td>0.0650</td>
<td>0.0440</td>
<td>0.4570</td>
<td>0.0820</td>
<td>0.1810</td>
</tr>
<tr>
<td></td>
<td>0.2171</td>
<td>0.0670</td>
<td></td>
<td>0.4603</td>
<td>0.0842</td>
<td>0.1844</td>
</tr>
<tr>
<td></td>
<td>1070</td>
<td>1084</td>
<td></td>
<td>1052</td>
<td>1080</td>
<td>1102</td>
</tr>
<tr>
<td>30</td>
<td>0.1340</td>
<td>0.0550</td>
<td>0.0460</td>
<td>0.3230</td>
<td>0.0630</td>
<td>0.1170</td>
</tr>
<tr>
<td></td>
<td>0.1365</td>
<td>0.0568</td>
<td></td>
<td>0.3258</td>
<td>0.0648</td>
<td>0.1199</td>
</tr>
<tr>
<td></td>
<td>1179</td>
<td>1199</td>
<td></td>
<td>1112</td>
<td>1184</td>
<td>1282</td>
</tr>
<tr>
<td>50</td>
<td>0.0820</td>
<td>0.0430</td>
<td>0.0390</td>
<td>0.1870</td>
<td>0.0580</td>
<td>0.0890</td>
</tr>
<tr>
<td></td>
<td>0.0884</td>
<td>0.0478</td>
<td></td>
<td>0.1963</td>
<td>0.0644</td>
<td>0.0965</td>
</tr>
<tr>
<td></td>
<td>1202</td>
<td>1208</td>
<td></td>
<td>1214</td>
<td>1300</td>
<td>1273</td>
</tr>
<tr>
<td>75</td>
<td>0.0850</td>
<td>0.0590</td>
<td>0.0500</td>
<td>0.1520</td>
<td>0.0700</td>
<td>0.0840</td>
</tr>
<tr>
<td></td>
<td>0.0850</td>
<td>0.0590</td>
<td></td>
<td>0.1520</td>
<td>0.0700</td>
<td>0.0840</td>
</tr>
<tr>
<td></td>
<td>1344</td>
<td>1520</td>
<td></td>
<td>1415</td>
<td>2018</td>
<td>1764</td>
</tr>
</tbody>
</table>

Table 16. Probabilities of Type I error when \((p, k, q) = (10, 4, 2)\) and \(\lambda = (1'_{0.6}, 1.5, 5, 10, 10')\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>(LRT)</th>
<th>(BART)</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.2320</td>
<td>0.0650</td>
<td>0.0470</td>
<td>0.5070</td>
<td>0.0760</td>
<td>0.1980</td>
</tr>
<tr>
<td></td>
<td>0.2336</td>
<td>0.0661</td>
<td></td>
<td>0.5085</td>
<td>0.0769</td>
<td>0.1994</td>
</tr>
<tr>
<td></td>
<td>1078</td>
<td>1103</td>
<td></td>
<td>1045</td>
<td>1060</td>
<td>1064</td>
</tr>
<tr>
<td>30</td>
<td>0.1610</td>
<td>0.0640</td>
<td>0.0450</td>
<td>0.3940</td>
<td>0.0890</td>
<td>0.1570</td>
</tr>
<tr>
<td></td>
<td>0.1642</td>
<td>0.0661</td>
<td></td>
<td>0.3972</td>
<td>0.0911</td>
<td>0.1599</td>
</tr>
<tr>
<td></td>
<td>1153</td>
<td>1146</td>
<td></td>
<td>1078</td>
<td>1102</td>
<td>1122</td>
</tr>
<tr>
<td>50</td>
<td>0.1260</td>
<td>0.0590</td>
<td>0.0630</td>
<td>0.3280</td>
<td>0.1070</td>
<td>0.1420</td>
</tr>
<tr>
<td></td>
<td>0.1164</td>
<td>0.0517</td>
<td></td>
<td>0.3192</td>
<td>0.1020</td>
<td>0.1366</td>
</tr>
<tr>
<td></td>
<td>1425</td>
<td>1510</td>
<td></td>
<td>1151</td>
<td>1106</td>
<td>1095</td>
</tr>
<tr>
<td>75</td>
<td>0.0940</td>
<td>0.0500</td>
<td>0.0400</td>
<td>0.2190</td>
<td>0.0750</td>
<td>0.1070</td>
</tr>
<tr>
<td></td>
<td>0.1011</td>
<td>0.0547</td>
<td></td>
<td>0.2269</td>
<td>0.0786</td>
<td>0.1124</td>
</tr>
<tr>
<td></td>
<td>1293</td>
<td>1216</td>
<td></td>
<td>1161</td>
<td>1079</td>
<td>1132</td>
</tr>
</tbody>
</table>
Table 17. Probabilities of Type I error when \((p,k,q) = (10,4,5)\) and \(\lambda = (1^*,10,10,10,25)^\prime\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>(LRT)</th>
<th>(BART)</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.1560</td>
<td>0.0740</td>
<td>0.0800</td>
<td>0.7100</td>
<td>0.0840</td>
<td>0.3190</td>
</tr>
<tr>
<td></td>
<td>0.1370</td>
<td>0.0593</td>
<td></td>
<td>0.7005</td>
<td>0.0590</td>
<td>0.2968</td>
</tr>
<tr>
<td></td>
<td>1287</td>
<td>1347</td>
<td></td>
<td>1036</td>
<td>2968</td>
<td>1227</td>
</tr>
<tr>
<td>30</td>
<td>0.1120</td>
<td>0.0600</td>
<td>0.0650</td>
<td>0.4460</td>
<td>0.0890</td>
<td>0.1770</td>
</tr>
<tr>
<td></td>
<td>0.1019</td>
<td>0.0518</td>
<td></td>
<td>0.4371</td>
<td>0.0746</td>
<td>0.1638</td>
</tr>
<tr>
<td></td>
<td>1378</td>
<td>1469</td>
<td></td>
<td>1094</td>
<td>3202</td>
<td>1477</td>
</tr>
<tr>
<td>50</td>
<td>0.0750</td>
<td>0.0550</td>
<td>0.0500</td>
<td>0.2470</td>
<td>0.0660</td>
<td>0.0890</td>
</tr>
<tr>
<td></td>
<td>0.0750</td>
<td>0.0550</td>
<td></td>
<td>0.2470</td>
<td>0.0660</td>
<td>0.0890</td>
</tr>
<tr>
<td></td>
<td>1605</td>
<td>1905</td>
<td></td>
<td>1191</td>
<td>3918</td>
<td>2167</td>
</tr>
<tr>
<td>75</td>
<td>0.0660</td>
<td>0.0510</td>
<td>0.0430</td>
<td>0.1750</td>
<td>0.0570</td>
<td>0.0700</td>
</tr>
<tr>
<td></td>
<td>0.0720</td>
<td>0.0568</td>
<td></td>
<td>0.1810</td>
<td>0.0639</td>
<td>0.0768</td>
</tr>
<tr>
<td></td>
<td>1950</td>
<td>2346</td>
<td></td>
<td>1268</td>
<td>3896</td>
<td>2481</td>
</tr>
</tbody>
</table>

Table 18. Probabilities of Type I error when \((p,k,q) = (10,4,5)\) and \(\lambda = (1^*,5,5,10,10)^\prime\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>(LRT)</th>
<th>(BART)</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.1560</td>
<td>0.0460</td>
<td>0.0710</td>
<td>0.7980</td>
<td>0.0410</td>
<td>0.3500</td>
</tr>
<tr>
<td></td>
<td>0.1433</td>
<td>0.0397</td>
<td></td>
<td>0.7934</td>
<td>0.0333</td>
<td>0.3356</td>
</tr>
<tr>
<td></td>
<td>1224</td>
<td>1155</td>
<td></td>
<td>1019</td>
<td>1288</td>
<td>1157</td>
</tr>
<tr>
<td>30</td>
<td>0.1170</td>
<td>0.0560</td>
<td>0.0640</td>
<td>0.5180</td>
<td>0.0670</td>
<td>0.1880</td>
</tr>
<tr>
<td></td>
<td>0.1078</td>
<td>0.0489</td>
<td></td>
<td>0.5108</td>
<td>0.0568</td>
<td>0.1759</td>
</tr>
<tr>
<td></td>
<td>1337</td>
<td>1412</td>
<td></td>
<td>1067</td>
<td>2041</td>
<td>1419</td>
</tr>
<tr>
<td>50</td>
<td>0.0760</td>
<td>0.0500</td>
<td>0.0500</td>
<td>0.2890</td>
<td>0.0650</td>
<td>0.0980</td>
</tr>
<tr>
<td></td>
<td>0.0760</td>
<td>0.0500</td>
<td></td>
<td>0.2890</td>
<td>0.0650</td>
<td>0.0980</td>
</tr>
<tr>
<td></td>
<td>1591</td>
<td>1701</td>
<td></td>
<td>1148</td>
<td>2967</td>
<td>1939</td>
</tr>
<tr>
<td>75</td>
<td>0.0620</td>
<td>0.0480</td>
<td>0.0440</td>
<td>0.2070</td>
<td>0.0590</td>
<td>0.0730</td>
</tr>
<tr>
<td></td>
<td>0.0670</td>
<td>0.0525</td>
<td></td>
<td>0.2120</td>
<td>0.0649</td>
<td>0.0788</td>
</tr>
<tr>
<td></td>
<td>2034</td>
<td>2123</td>
<td></td>
<td>1214</td>
<td>3760</td>
<td>2406</td>
</tr>
</tbody>
</table>
Table 19. Probabilities of Type I error when \((p,k,q) = (10,4,5)\) and \(\lambda = (1_5', 2, 5, 10, 10')\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>(LRT)</th>
<th>(BART)</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.1800</td>
<td>0.0340</td>
<td>0.0790</td>
<td>0.9010</td>
<td>0.0460</td>
<td>0.4900</td>
</tr>
<tr>
<td></td>
<td>0.1637</td>
<td>0.0271</td>
<td>0.8979</td>
<td>0.0387</td>
<td>0.4763</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1183</td>
<td>1143</td>
<td>1009</td>
<td>1118</td>
<td>1069</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.1340</td>
<td>0.0430</td>
<td>0.0640</td>
<td>0.7620</td>
<td>0.0730</td>
<td>0.3220</td>
</tr>
<tr>
<td></td>
<td>0.1243</td>
<td>0.0385</td>
<td>0.7584</td>
<td>0.0680</td>
<td>0.3123</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1327</td>
<td>1176</td>
<td>1021</td>
<td>1126</td>
<td>1150</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.0930</td>
<td>0.0460</td>
<td>0.0500</td>
<td>0.4940</td>
<td>0.0740</td>
<td>0.1590</td>
</tr>
<tr>
<td></td>
<td>0.0930</td>
<td>0.0460</td>
<td>0.4940</td>
<td>0.0740</td>
<td>0.1590</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1492</td>
<td>1413</td>
<td>1057</td>
<td>1430</td>
<td>1295</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>0.0690</td>
<td>0.0470</td>
<td>0.0460</td>
<td>0.3230</td>
<td>0.0780</td>
<td>0.1200</td>
</tr>
<tr>
<td></td>
<td>0.0724</td>
<td>0.0500</td>
<td>0.3258</td>
<td>0.0810</td>
<td>0.1233</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1927</td>
<td>2215</td>
<td>1112</td>
<td>1547</td>
<td>1402</td>
<td></td>
</tr>
</tbody>
</table>

Table 20. Probabilities of Type I error when \((p,k,q) = (10,4,5)\) and \(\lambda = (1_5', 1.5, 5, 10, 10')\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>(LRT)</th>
<th>(BART)</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.2020</td>
<td>0.0250</td>
<td>0.0630</td>
<td>0.9450</td>
<td>0.0560</td>
<td>0.5740</td>
</tr>
<tr>
<td></td>
<td>0.1936</td>
<td>0.0220</td>
<td>0.9442</td>
<td>0.0541</td>
<td>0.5694</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1182</td>
<td>1143</td>
<td>1003</td>
<td>1023</td>
<td>1031</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.1620</td>
<td>0.0260</td>
<td>0.0680</td>
<td>0.8490</td>
<td>0.0990</td>
<td>0.4540</td>
</tr>
<tr>
<td></td>
<td>0.1515</td>
<td>0.0214</td>
<td>0.8461</td>
<td>0.0955</td>
<td>0.4454</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1189</td>
<td>1196</td>
<td>1013</td>
<td>1027</td>
<td>1061</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.1080</td>
<td>0.0220</td>
<td>0.0490</td>
<td>0.7180</td>
<td>0.1260</td>
<td>0.3170</td>
</tr>
<tr>
<td></td>
<td>0.1087</td>
<td>0.0223</td>
<td>0.7183</td>
<td>0.1264</td>
<td>0.3175</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1338</td>
<td>1285</td>
<td>1020</td>
<td>1066</td>
<td>1057</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>0.1040</td>
<td>0.0420</td>
<td>0.0520</td>
<td>0.5620</td>
<td>0.1330</td>
<td>0.2540</td>
</tr>
<tr>
<td></td>
<td>0.1022</td>
<td>0.0408</td>
<td>0.5611</td>
<td>0.1321</td>
<td>0.2527</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1763</td>
<td>1720</td>
<td>1044</td>
<td>1103</td>
<td>1121</td>
<td></td>
</tr>
</tbody>
</table>
Table 21. Probabilities of Type I error when \((p,k,q) = (15,2,1)\) and \(\lambda = (1'_{13},15,35)'\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>(LRT)</th>
<th>(BART)</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.0610</td>
<td>0.0470</td>
<td>0.0580</td>
<td>0.0650</td>
<td>0.0450</td>
<td>0.0470</td>
</tr>
<tr>
<td></td>
<td>0.0606</td>
<td>0.0468</td>
<td></td>
<td>0.0647</td>
<td>0.0455</td>
<td>0.0465</td>
</tr>
<tr>
<td></td>
<td>1001</td>
<td>1000</td>
<td></td>
<td>1001</td>
<td>1004</td>
<td>1004</td>
</tr>
<tr>
<td>35</td>
<td>0.0490</td>
<td>0.0390</td>
<td>0.0520</td>
<td>0.0590</td>
<td>0.0410</td>
<td>0.0420</td>
</tr>
<tr>
<td></td>
<td>0.0489</td>
<td>0.0389</td>
<td></td>
<td>0.0580</td>
<td>0.0409</td>
<td>0.0419</td>
</tr>
<tr>
<td></td>
<td>1002</td>
<td>1004</td>
<td></td>
<td>1001</td>
<td>1001</td>
<td>1001</td>
</tr>
<tr>
<td>50</td>
<td>0.0380</td>
<td>0.0310</td>
<td>0.0460</td>
<td>0.0500</td>
<td>0.0370</td>
<td>0.0370</td>
</tr>
<tr>
<td></td>
<td>0.0381</td>
<td>0.0311</td>
<td></td>
<td>0.0503</td>
<td>0.0373</td>
<td>0.0373</td>
</tr>
<tr>
<td></td>
<td>1001</td>
<td>1000</td>
<td></td>
<td>1006</td>
<td>1007</td>
<td>1007</td>
</tr>
<tr>
<td>75</td>
<td>0.0400</td>
<td>0.0370</td>
<td>0.0560</td>
<td>0.0410</td>
<td>0.0380</td>
<td>0.0380</td>
</tr>
<tr>
<td></td>
<td>0.0396</td>
<td>0.0366</td>
<td></td>
<td>0.0405</td>
<td>0.0374</td>
<td>0.0374</td>
</tr>
<tr>
<td></td>
<td>1007</td>
<td>1008</td>
<td></td>
<td>1010</td>
<td>1012</td>
<td>1012</td>
</tr>
</tbody>
</table>

Table 22. Probabilities of Type I error when \((p,k,q) = (15,2,1)\) and \(\lambda = (1'_{13},10,15)'\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>(LRT)</th>
<th>(BART)</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.0620</td>
<td>0.0460</td>
<td>0.0580</td>
<td>0.0610</td>
<td>0.0480</td>
<td>0.0490</td>
</tr>
<tr>
<td></td>
<td>0.0616</td>
<td>0.0458</td>
<td></td>
<td>0.0606</td>
<td>0.0475</td>
<td>0.0485</td>
</tr>
<tr>
<td></td>
<td>1001</td>
<td>1000</td>
<td></td>
<td>1001</td>
<td>1004</td>
<td>1003</td>
</tr>
<tr>
<td>35</td>
<td>0.0470</td>
<td>0.0380</td>
<td>0.0520</td>
<td>0.0580</td>
<td>0.0440</td>
<td>0.0450</td>
</tr>
<tr>
<td></td>
<td>0.0469</td>
<td>0.0370</td>
<td></td>
<td>0.0579</td>
<td>0.0439</td>
<td>0.0449</td>
</tr>
<tr>
<td></td>
<td>1003</td>
<td>1005</td>
<td></td>
<td>1001</td>
<td>1001</td>
<td>1001</td>
</tr>
<tr>
<td>50</td>
<td>0.0370</td>
<td>0.0310</td>
<td>0.0460</td>
<td>0.0480</td>
<td>0.0380</td>
<td>0.0380</td>
</tr>
<tr>
<td></td>
<td>0.0361</td>
<td>0.0394</td>
<td></td>
<td>0.0618</td>
<td>0.0520</td>
<td>0.0520</td>
</tr>
<tr>
<td></td>
<td>1001</td>
<td>1001</td>
<td></td>
<td>1004</td>
<td>1006</td>
<td>1006</td>
</tr>
<tr>
<td>75</td>
<td>0.0560</td>
<td>0.0480</td>
<td>0.0460</td>
<td>0.0680</td>
<td>0.0600</td>
<td>0.0600</td>
</tr>
<tr>
<td></td>
<td>0.0563</td>
<td>0.0482</td>
<td></td>
<td>0.0684</td>
<td>0.0602</td>
<td>0.0602</td>
</tr>
<tr>
<td></td>
<td>1005</td>
<td>1001</td>
<td></td>
<td>1005</td>
<td>1002</td>
<td>1002</td>
</tr>
</tbody>
</table>
Table 23. Probabilities of Type I error when \((p, k, q) = (15, 2, 1)\) and \(\lambda = \left(1_{13}, 2, 15\right)'\).

<table>
<thead>
<tr>
<th>n</th>
<th>LRT</th>
<th>BART</th>
<th>F</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.1000</td>
<td>0.0740</td>
<td>0.0580</td>
<td>0.0830</td>
<td>0.0630</td>
<td>0.0650</td>
</tr>
<tr>
<td></td>
<td>0.0995</td>
<td>0.0735</td>
<td></td>
<td>0.0822</td>
<td>0.0625</td>
<td>0.0644</td>
</tr>
<tr>
<td></td>
<td>1002</td>
<td>1003</td>
<td></td>
<td>1006</td>
<td>1003</td>
<td>1005</td>
</tr>
<tr>
<td>35</td>
<td>0.0800</td>
<td>0.0570</td>
<td>0.0520</td>
<td>0.0720</td>
<td>0.0540</td>
<td>0.0570</td>
</tr>
<tr>
<td></td>
<td>0.0798</td>
<td>0.0568</td>
<td></td>
<td>0.0719</td>
<td>0.0539</td>
<td>0.0569</td>
</tr>
<tr>
<td></td>
<td>1009</td>
<td>1013</td>
<td></td>
<td>1000</td>
<td>1001</td>
<td>1003</td>
</tr>
<tr>
<td>50</td>
<td>0.0900</td>
<td>0.0650</td>
<td>0.0450</td>
<td>0.0670</td>
<td>0.0640</td>
<td>0.0660</td>
</tr>
<tr>
<td></td>
<td>0.0901</td>
<td>0.0650</td>
<td></td>
<td>0.0668</td>
<td>0.0638</td>
<td>0.0658</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>1000</td>
<td></td>
<td>1001</td>
<td>1001</td>
<td>1001</td>
</tr>
<tr>
<td>75</td>
<td>0.0660</td>
<td>0.0550</td>
<td>0.0560</td>
<td>0.0470</td>
<td>0.0400</td>
<td>0.0410</td>
</tr>
<tr>
<td></td>
<td>0.0655</td>
<td>0.0548</td>
<td></td>
<td>0.0466</td>
<td>0.0397</td>
<td>0.0407</td>
</tr>
<tr>
<td></td>
<td>1005</td>
<td>1001</td>
<td></td>
<td>1004</td>
<td>1003</td>
<td>1003</td>
</tr>
</tbody>
</table>

Table 24. Probabilities of Type I error when \((p, k, q) = (15, 2, 1)\) and \(\lambda = \left(1_{13}, 1.5, 15\right)'\).

<table>
<thead>
<tr>
<th>n</th>
<th>LRT</th>
<th>BART</th>
<th>F</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.1230</td>
<td>0.0850</td>
<td>0.0580</td>
<td>0.0890</td>
<td>0.0650</td>
<td>0.0710</td>
</tr>
<tr>
<td></td>
<td>0.1221</td>
<td>0.0844</td>
<td></td>
<td>0.0881</td>
<td>0.0645</td>
<td>0.0701</td>
</tr>
<tr>
<td></td>
<td>1005</td>
<td>1003</td>
<td></td>
<td>1007</td>
<td>1003</td>
<td>1009</td>
</tr>
<tr>
<td>35</td>
<td>0.1000</td>
<td>0.0620</td>
<td>0.0520</td>
<td>0.0910</td>
<td>0.0730</td>
<td>0.0790</td>
</tr>
<tr>
<td></td>
<td>0.0998</td>
<td>0.0618</td>
<td></td>
<td>0.0909</td>
<td>0.0728</td>
<td>0.0788</td>
</tr>
<tr>
<td></td>
<td>1007</td>
<td>1008</td>
<td></td>
<td>1002</td>
<td>1005</td>
<td>1004</td>
</tr>
<tr>
<td>50</td>
<td>0.1060</td>
<td>0.0780</td>
<td>0.0530</td>
<td>0.0860</td>
<td>0.0860</td>
<td>0.0820</td>
</tr>
<tr>
<td></td>
<td>0.1059</td>
<td>0.0778</td>
<td></td>
<td>0.0859</td>
<td>0.0860</td>
<td>0.0820</td>
</tr>
<tr>
<td></td>
<td>1001</td>
<td>1002</td>
<td></td>
<td>1001</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>75</td>
<td>0.0930</td>
<td>0.0640</td>
<td>0.0560</td>
<td>0.0740</td>
<td>0.0620</td>
<td>0.0700</td>
</tr>
<tr>
<td></td>
<td>0.0923</td>
<td>0.0635</td>
<td></td>
<td>0.0736</td>
<td>0.0616</td>
<td>0.0694</td>
</tr>
<tr>
<td></td>
<td>1007</td>
<td>1006</td>
<td></td>
<td>1004</td>
<td>1004</td>
<td>1007</td>
</tr>
</tbody>
</table>
Table 25. Probabilities of Type I error when \((p,k,q) = (15,4,2)\) and \(\lambda = (1_{11},15,15,15,35)'\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>(LRT)</th>
<th>(BART)</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.1360</td>
<td>0.0700</td>
<td>0.1520</td>
<td>0.2080</td>
<td>0.0690</td>
<td>0.0930</td>
</tr>
<tr>
<td></td>
<td>0.1247</td>
<td>0.0626</td>
<td></td>
<td>0.1919</td>
<td>0.0623</td>
<td>0.0820</td>
</tr>
<tr>
<td></td>
<td>1013</td>
<td>1010</td>
<td></td>
<td>1020</td>
<td>1008</td>
<td>1018</td>
</tr>
<tr>
<td>35</td>
<td>0.1050</td>
<td>0.0640</td>
<td>0.0830</td>
<td>0.1440</td>
<td>0.0710</td>
<td>0.0860</td>
</tr>
<tr>
<td></td>
<td>0.1005</td>
<td>0.0602</td>
<td></td>
<td>0.1379</td>
<td>0.0675</td>
<td>0.0809</td>
</tr>
<tr>
<td></td>
<td>1015</td>
<td>1016</td>
<td></td>
<td>1021</td>
<td>1013</td>
<td>1024</td>
</tr>
<tr>
<td>50</td>
<td>0.0740</td>
<td>0.0440</td>
<td>0.0660</td>
<td>0.0113</td>
<td>0.0580</td>
<td>0.0620</td>
</tr>
<tr>
<td></td>
<td>0.0722</td>
<td>0.0419</td>
<td></td>
<td>0.0110</td>
<td>0.0564</td>
<td>0.0602</td>
</tr>
<tr>
<td></td>
<td>1012</td>
<td>1025</td>
<td></td>
<td>1009</td>
<td>1011</td>
<td>1013</td>
</tr>
<tr>
<td>75</td>
<td>0.0770</td>
<td>0.0630</td>
<td>0.0640</td>
<td>0.0910</td>
<td>0.0610</td>
<td>0.0620</td>
</tr>
<tr>
<td></td>
<td>0.0744</td>
<td>0.0604</td>
<td></td>
<td>0.0882</td>
<td>0.0586</td>
<td>0.0597</td>
</tr>
<tr>
<td></td>
<td>1024</td>
<td>1035</td>
<td></td>
<td>1030</td>
<td>1030</td>
<td>1029</td>
</tr>
</tbody>
</table>

Table 26. Probabilities of Type I error when \((p,k,q) = (15,4,2)\) and \(\lambda = (1_{11},10,10,15,15)'\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>(LRT)</th>
<th>(BART)</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.1350</td>
<td>0.0710</td>
<td>0.1520</td>
<td>0.2100</td>
<td>0.0710</td>
<td>0.0940</td>
</tr>
<tr>
<td></td>
<td>0.1243</td>
<td>0.0629</td>
<td></td>
<td>0.1933</td>
<td>0.0637</td>
<td>0.0839</td>
</tr>
<tr>
<td></td>
<td>1012</td>
<td>1012</td>
<td></td>
<td>1021</td>
<td>1010</td>
<td>1014</td>
</tr>
<tr>
<td>35</td>
<td>0.1060</td>
<td>0.0680</td>
<td>0.0830</td>
<td>0.1530</td>
<td>0.0710</td>
<td>0.0820</td>
</tr>
<tr>
<td></td>
<td>0.1016</td>
<td>0.0639</td>
<td></td>
<td>0.1468</td>
<td>0.0671</td>
<td>0.0776</td>
</tr>
<tr>
<td></td>
<td>1014</td>
<td>1018</td>
<td></td>
<td>1021</td>
<td>1016</td>
<td>1018</td>
</tr>
<tr>
<td>50</td>
<td>0.0750</td>
<td>0.0440</td>
<td>0.0660</td>
<td>0.1120</td>
<td>0.0580</td>
<td>0.0630</td>
</tr>
<tr>
<td></td>
<td>0.0729</td>
<td>0.0419</td>
<td></td>
<td>0.1100</td>
<td>0.0564</td>
<td>0.0612</td>
</tr>
<tr>
<td></td>
<td>1015</td>
<td>1025</td>
<td></td>
<td>1009</td>
<td>1011</td>
<td>1013</td>
</tr>
<tr>
<td>75</td>
<td>0.0740</td>
<td>0.0620</td>
<td>0.0640</td>
<td>0.0880</td>
<td>0.0610</td>
<td>0.0610</td>
</tr>
<tr>
<td></td>
<td>0.0714</td>
<td>0.0594</td>
<td></td>
<td>0.0853</td>
<td>0.0586</td>
<td>0.0586</td>
</tr>
<tr>
<td></td>
<td>1031</td>
<td>1036</td>
<td></td>
<td>1027</td>
<td>1030</td>
<td>1030</td>
</tr>
</tbody>
</table>
Table 27. Probabilities of Type I error when \((p,k,q) = (15,4,2)\) and \(\lambda = (1_{11},2,10,15,15)^t\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>(LRT)</th>
<th>(BART)</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.1970</td>
<td>0.0990</td>
<td>0.1520</td>
<td>0.2950</td>
<td>0.0550</td>
<td>0.1110</td>
</tr>
<tr>
<td></td>
<td>0.1843</td>
<td>0.0880</td>
<td></td>
<td>0.2830</td>
<td>0.0490</td>
<td>0.1014</td>
</tr>
<tr>
<td></td>
<td>1012</td>
<td>1017</td>
<td></td>
<td>1008</td>
<td>1008</td>
<td>1011</td>
</tr>
<tr>
<td>35</td>
<td>0.1450</td>
<td>0.0880</td>
<td>0.0830</td>
<td>0.2320</td>
<td>0.0540</td>
<td>0.0820</td>
</tr>
<tr>
<td></td>
<td>0.1407</td>
<td>0.0851</td>
<td></td>
<td>0.2273</td>
<td>0.0520</td>
<td>0.0784</td>
</tr>
<tr>
<td></td>
<td>1010</td>
<td>1007</td>
<td></td>
<td>1008</td>
<td>1005</td>
<td>1011</td>
</tr>
<tr>
<td>50</td>
<td>0.1030</td>
<td>0.0590</td>
<td>0.0660</td>
<td>0.1650</td>
<td>0.0510</td>
<td>0.0740</td>
</tr>
<tr>
<td></td>
<td>0.1009</td>
<td>0.0572</td>
<td></td>
<td>0.1619</td>
<td>0.0493</td>
<td>0.0724</td>
</tr>
<tr>
<td></td>
<td>1012</td>
<td>1015</td>
<td></td>
<td>1071</td>
<td>1015</td>
<td>1008</td>
</tr>
<tr>
<td>75</td>
<td>0.0960</td>
<td>0.0600</td>
<td>0.0640</td>
<td>0.1240</td>
<td>0.0650</td>
<td>0.0690</td>
</tr>
<tr>
<td></td>
<td>0.0937</td>
<td>0.0580</td>
<td></td>
<td>0.1209</td>
<td>0.0636</td>
<td>0.0672</td>
</tr>
<tr>
<td></td>
<td>1019</td>
<td>1020</td>
<td></td>
<td>1026</td>
<td>1009</td>
<td>1015</td>
</tr>
</tbody>
</table>

Table 28. Probabilities of Type I error when \((p,k,q) = (15,4,2)\) and \(\lambda = (1_{11},1.5,10,15,15)^t\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>(LRT)</th>
<th>(BART)</th>
<th>(F)</th>
<th>(T_{k,q})</th>
<th>(T^*)</th>
<th>(T^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.2290</td>
<td>0.1160</td>
<td>0.1520</td>
<td>0.3460</td>
<td>0.0650</td>
<td>0.1300</td>
</tr>
<tr>
<td></td>
<td>0.2162</td>
<td>0.1054</td>
<td></td>
<td>0.3386</td>
<td>0.0617</td>
<td>0.1227</td>
</tr>
<tr>
<td></td>
<td>1011</td>
<td>1013</td>
<td></td>
<td>1003</td>
<td>1002</td>
<td>1005</td>
</tr>
<tr>
<td>35</td>
<td>0.1790</td>
<td>0.0870</td>
<td>0.0830</td>
<td>0.2970</td>
<td>0.0660</td>
<td>0.1080</td>
</tr>
<tr>
<td></td>
<td>0.1746</td>
<td>0.0854</td>
<td></td>
<td>0.2912</td>
<td>0.0653</td>
<td>0.1054</td>
</tr>
<tr>
<td></td>
<td>1009</td>
<td>1002</td>
<td></td>
<td>1011</td>
<td>1000</td>
<td>1005</td>
</tr>
<tr>
<td>50</td>
<td>0.1230</td>
<td>0.0700</td>
<td>0.0660</td>
<td>0.2340</td>
<td>0.0760</td>
<td>0.0910</td>
</tr>
<tr>
<td></td>
<td>0.1210</td>
<td>0.0683</td>
<td></td>
<td>0.2323</td>
<td>0.0752</td>
<td>0.0894</td>
</tr>
<tr>
<td></td>
<td>1009</td>
<td>1010</td>
<td></td>
<td>1003</td>
<td>1002</td>
<td>1007</td>
</tr>
<tr>
<td>75</td>
<td>0.1240</td>
<td>0.0780</td>
<td>0.0640</td>
<td>0.1940</td>
<td>0.0750</td>
<td>0.0980</td>
</tr>
<tr>
<td></td>
<td>0.1216</td>
<td>0.0761</td>
<td></td>
<td>0.1911</td>
<td>0.0743</td>
<td>0.0967</td>
</tr>
<tr>
<td></td>
<td>1015</td>
<td>1015</td>
<td></td>
<td>1017</td>
<td>1002</td>
<td>1006</td>
</tr>
</tbody>
</table>
The following tables report power of the tests when \( p = 10, k = 2, 4, \) and \( q = 1, 2. \) Figures compare the power of the \( BART \) and \( T^* \) tests both of which are controlled for Type I error.

Figure 3. Power of test when \( \lambda = (1_8, 10, 25)^t, k = 2, q = 1. \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>Power of test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( LRT )</td>
</tr>
<tr>
<td>25</td>
<td>0.3330</td>
</tr>
<tr>
<td>50</td>
<td>0.5890</td>
</tr>
<tr>
<td>75</td>
<td>0.7480</td>
</tr>
<tr>
<td>100</td>
<td>0.8580</td>
</tr>
<tr>
<td>150</td>
<td>0.9620</td>
</tr>
</tbody>
</table>

Table 29. Power of test when \( \lambda = (1_8, 10, 25)^t, k = 2, q = 1. \)
Figure 4. Power of test when $\lambda = (1_8, 2, 10)'$, $k = 2$, $q = 1$.

Table 30. Power of test when $\lambda = (1_8, 2, 10)'$, $k = 2$, $q = 1$.

<table>
<thead>
<tr>
<th>n</th>
<th>Power of test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$LRT$</td>
</tr>
<tr>
<td>25</td>
<td>0.3150</td>
</tr>
<tr>
<td>50</td>
<td>0.4580</td>
</tr>
<tr>
<td>75</td>
<td>0.5960</td>
</tr>
<tr>
<td>100</td>
<td>0.6930</td>
</tr>
<tr>
<td>150</td>
<td>0.8750</td>
</tr>
</tbody>
</table>
Figure 5. Power of test when $\lambda = (1_6', 10, 10, 10, 25)'$, $k = 4$, $q = 2$.

![Figure 5](image_url)

Table 31. Power of test when $\lambda = (1_6', 10, 10, 10, 25)'$, $k = 4$, $q = 2$.

<table>
<thead>
<tr>
<th>n</th>
<th>Power of test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LRT</td>
</tr>
<tr>
<td>25</td>
<td>0.2750</td>
</tr>
<tr>
<td>50</td>
<td>0.3420</td>
</tr>
<tr>
<td>75</td>
<td>0.4990</td>
</tr>
<tr>
<td>100</td>
<td>0.6130</td>
</tr>
<tr>
<td>150</td>
<td>0.8230</td>
</tr>
</tbody>
</table>
Figure 6. Power of test when $\lambda = (1_6^t, 2, 5, 10)^t$, $k = 4$, $q = 2$.

Table 32. Power of test when $\lambda = (1_6^t, 2, 5, 10)^t$, $k = 4$, $q = 2$.

<table>
<thead>
<tr>
<th>n</th>
<th>Power of test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LRT</td>
</tr>
<tr>
<td>50</td>
<td>0.2900</td>
</tr>
<tr>
<td>75</td>
<td>0.3770</td>
</tr>
<tr>
<td>100</td>
<td>0.4840</td>
</tr>
<tr>
<td>150</td>
<td>0.6240</td>
</tr>
<tr>
<td>175</td>
<td>0.7020</td>
</tr>
<tr>
<td>200</td>
<td>0.7890</td>
</tr>
</tbody>
</table>
Discussion

The results of the simulation study of variable reduction test reported in Table 5 - Table 28 show that the likelihood ratio test \((LRT)\) can be improved using a Bartlett correction, especially when \(\lambda_{(p-k)+1}/\lambda_{p-k}\) is quite small (see Tables 7, 8, 11, 12, 15, 16, 19, 20, 23, 24, 27, and 28). In these cases, the \(LRT\) becomes liberal whereas the \(BART\) is still conservative. With a few exceptions in Tables 8 and 19-22, the \(BART\) gives probabilities of Type I error between 0.04 to 0.10. In general, the \(LRT\) performs better than \(T_{k,q}\), especially when \(k + q\) is large, say greater than or equal to \(\frac{1}{2}\mu.\) In these cases, the \(LRT\) continues to perform well whereas \(T_{k,q}\) becomes liberal (see Tables 17-20). The \(T_{k,q}\) test gives probabilities of Type I error up to 0.945 (see Table 20). The \(T^{**}\) test does not perform nearly as well as \(T^{*}\) and \(BART\) unless \(k + q\) is small (see Tables 9-12 and 21-24) or the sample size is quite large. Under most conditions, the \(BART\) results are better than \(T^{*}\). The effective sample size, \(N^{*}\), is larger as the sample size is larger. The comparison test is useful when \(q\) is not far from \(m = p - k\). If \(q\) is far from \(m\), \(N^{*}\) is close to \(N\) and the relative efficiency is almost 1 (see Tables 21-28).

With respect to the power of test, it is sufficient to compare the \(BART\) with the \(T^{*}\) both of which are controlled for Type I error. The results are reported in Figure 3 - Figure 6. Only a few cases were performed. These figures show that the Bartlett-corrected test has a better power than \(T^{*}\), especially when \(\lambda_{(p-k)+1}/\lambda_{p-k}\) is small (see Figures 4 and 6).
Simulation Study of Component Reduction Tests

The evaluation is based on the percentage coverage of confidence interval for \( \frac{C^1 \lambda}{\phi_0} \) under different conditions. The intervals examined are

- Likelihood ratio interval (\( LRI \))
- Bartlett-corrected interval (\( BARI \))
- Anderson’s interval (\( ZI \)) by inverting from (1.6).

To obtain \( LRI \) and \( BARI \), the confidence interval of \( \frac{C^1 \lambda}{\phi_0} \) can be constructed by inverting the LRT and BART of \( H_0: \frac{C^1 \lambda}{\phi_0} = c_0 \) against \( H_a: \frac{C^1 \lambda}{\phi_0} \neq c_0 \). The interval consists of all values \( c_0 \) for which \( H_0: \frac{C^1 \lambda}{\phi_0} = c_0 \) cannot be rejected. That is, a (1 - \( \alpha \))100\% CI for \( \frac{C^1 \lambda}{\phi_0} \) can be obtained as

\[
CI \left( \frac{C^1 \lambda}{\phi_0} \right) = \left\{ c_0; Q \leq \chi^2_{1-\alpha,r} \right\},
\]

where \( Q \) is the LRT or BART statistic.

In this simulation study, the \( LRI \) and \( BARI \) are obtained by applying the fminsearch command in Matlab program (see Matlab subprogram in Appendices). To minimize the quantity \((p\text{-value} - \alpha)^2\), this subprogram finds a local minimizer \( c_0 \).

The interval \((c_{01}, c_{02})\) is found such that the \( p \)-values for testing \( H_0: \frac{C^1 \lambda}{\phi_0} = c_{01} \) and \( H_0: \frac{C^1 \lambda}{\phi_0} = c_{02} \) are each equal to \( \alpha \). The endpoints of Anderson’s intervals are used as initial guesses.

The number of original variables, \( p \), was 8. The number of retained components, \( k \), was 2 and 4. For each simulation condition, 10000 data sets were generated. For
simplicity, the covariance matrices examined were diagonal and have \( \lambda_1 = \cdots = \lambda_{p-k} = 1 \) and values for the remaining eigenvalue are listed in the tables below. The vector \( \mathbf{C}_1 \) was chosen so that \( \frac{C_i^\Delta}{\phi_0} = \sum_{i=1}^{m} \frac{\lambda_i}{\text{tr}(\Sigma)} \). The nominal test size used was 0.05. Tables with the percentage coverage of confidence interval for \( \frac{C_i^\Delta}{\phi_0} \) are shown below.

Table 33. Percentage coverage of CI when \((p,k) = (8,2)\) and \(\lambda = (1_6',10,25)'\).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( LRI )</th>
<th>( BARI )</th>
<th>( ZI )</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>90.70</td>
<td>92.40</td>
<td>88.90</td>
</tr>
<tr>
<td>75</td>
<td>94.10</td>
<td>94.90</td>
<td>93.90</td>
</tr>
</tbody>
</table>

Table 34. Percentage coverage of CI when \((p,k) = (8,2)\) and \(\lambda = (1_6',2,10)'\).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( LRI )</th>
<th>( BARI )</th>
<th>( ZI )</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>84.70</td>
<td>90.10</td>
<td>84.90</td>
</tr>
<tr>
<td>75</td>
<td>92.30</td>
<td>93.50</td>
<td>92.60</td>
</tr>
</tbody>
</table>

Table 35. Percentage coverage of CI when \((p,k) = (8,4)\) and \(\lambda = (1_4',10,10,10,25)'\).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( LRI )</th>
<th>( BARI )</th>
<th>( ZI )</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>78.80</td>
<td>93.90</td>
<td>72.50</td>
</tr>
<tr>
<td>75</td>
<td>90.40</td>
<td>94.90</td>
<td>89.90</td>
</tr>
</tbody>
</table>

Table 36. Percentage coverage of CI when \((p,k) = (8,4)\) and \(\lambda = (1_4',2,10,10,10)'\).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( LRI )</th>
<th>( BARI )</th>
<th>( ZI )</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>62.30</td>
<td>91.00</td>
<td>60.10</td>
</tr>
<tr>
<td>75</td>
<td>87.40</td>
<td>93.60</td>
<td>87.40</td>
</tr>
</tbody>
</table>

The results of the simulation of component reduction reported in Table 33 - Table 36 suggest that the Bartlett-corrected intervals perform better than the other intervals. The percentage coverage of CI for \( \frac{C_i^\Delta}{\phi_0} \) of the Bartlett-corrected intervals
are closer to 95% than that of the likelihood ratio intervals and Anderson's intervals in all conditions.

**Numerical Examples**

**Example 1. (continued)**

(The study of the quality of pictures) In Chapter 1, the following two questions were addressed.

1. How many components should be retained?

2. How many and which variables can be considered as redundant?

To answer these questions, a confidence interval for $\frac{C_1^\Lambda}{\varphi_0}$ was computed, where $\mathbf{C}_1 = (0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 1)'$. That is the CI for the cumulative proportion of total variance accounted for by the first two components. The 95% confidence interval is (0.7422, 0.7690). It can be concluded that the first two components are sufficient to explain the variability.

The focus, now, is on the first two components. From Table 2, the first two components have small coefficients on the last three variables. The question is whether these three variables can be eliminated. Consider hypothesis (1.21), where $k = 2$, $q = 3$ and $n = 108$. That is, the last three variables have zero coefficients in each of the first two components. The likelihood ratio test statistic is 3.36 which asymptotically follows a $\chi^2_6$ and the corresponding $p$-value is 0.7613. The hypothesis could not be rejected at either the 0.01 or 0.05 significant levels. Then, these three variables are
redundant and can be discarded. The Bartlett-corrected test statistic is 3.41 and the corresponding p-value is 0.7549. This leads to the same conclusion.

Example 2. (continued)

(Women’s track records) In Chapter 1, two questions were addressed: how many components should be retained and which polynomials contribute to the largest component? To answer these questions, the confidence interval for \( \frac{C_1 \lambda}{\nu_0} \) was computed, where \( C_1 = (0 \ 0 \ 0 \ 0 \ 1)' \). The 95% CI for the cumulative proportion of total variance accounted for by the first two components is (0.7823, 0.8282). Therefore, the first two components should be retained.

To reduce the dimension in the number of variables, hypothesis (1.21) was examined, where \( k = 1, 2 \) and \( q = 1, 2, 3, 4, 5 \). That is, the first \( k \) components have zero loadings on the last \( q \) variables.

Table 37. P-value of Testing Hypothesis: Women’s Track Example.

<table>
<thead>
<tr>
<th>( k )</th>
<th>( q )</th>
<th>( LRT )</th>
<th>p-value</th>
<th>( BART )</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.5877</td>
<td>0.4433</td>
<td>0.6082</td>
<td>0.4355</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2.5382</td>
<td>0.2811</td>
<td>2.5549</td>
<td>0.2787</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>8.7912</td>
<td>0.0322*</td>
<td>8.5538</td>
<td>0.0359*</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>9.7598</td>
<td>0.0447*</td>
<td>9.3310</td>
<td>0.0533</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>14.8112</td>
<td>0.0112*</td>
<td>13.9096</td>
<td>0.0162*</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.6040</td>
<td>0.7393</td>
<td>0.5991</td>
<td>0.7411</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>8.5104</td>
<td>0.0746</td>
<td>8.2538</td>
<td>0.0827</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>31.3926</td>
<td>&lt;0.0001**</td>
<td>30.1616</td>
<td>&lt;0.0001**</td>
</tr>
</tbody>
</table>

Note:* denote rejection of \( H_0 \) at 0.05

** denote rejection of \( H_0 \) at 0.01
For each of the hypotheses, the likelihood ratio test statistics, Bartlett-corrected test statistics which asymptotically follow a $\chi^2_{kq}$ and the corresponding $p$-values are presented in Table 37. Failure to reject $H_0$ means that, in the first $k$ components, all variability in speed over track events is related to the first $p - q$ polynomials. The last $q$ polynomials are redundant.
CHAPTER 6

CONCLUSION

In this thesis, new procedures for examining dimension reduction in principal component analysis were developed in Chapters 2, 3 and 4. The Bartlett-corrected test of redundancy appears to be promising. The tabled results show an improvement in Type I error control for the Bartlett-corrected test when compared to the likelihood ratio test and other competing tests. Figures 3-6 show that the Bartlett-corrected test has as good or better power than the competing test in all conditions that were examined.

The Bartlett-corrected test also appears to be promising in respect to the percentage coverage of confidence interval for $\frac{C_1\lambda}{\phi_0}$. Results show that the Bartlett-corrected intervals substantially improve the control over the percentage coverage of CI for $\frac{C_1\lambda}{\phi_0}$ in these conditions.

There are some limitations of this thesis. First, the procedures for estimating MLEs, for constructing LRTs, and for evaluating Bartlett corrections are derived under normality. The results do not apply to non-normal data. Second, the proposed model is based on the covariance matrix. The model does not apply to the correlation matrix. A limitation of the procedures is computational intensity. The computation of MLEs and the construction of LRTs are fast, but to compute the Bartlett correction, the procedure is computationally intensive. Computer time increases as $p$
increases. This may not be noticeable when computing on just one data set. However, a simulation study may take a long time. Efficient programming is needed to shorten the computer time.

For future research, it would be useful to construct one-sided confidence intervals for $\frac{C_i \lambda}{\gamma_0}$. Deriving Edgeworth approximations and saddlepoint approximations for the distribution of estimators of eigenvectors and eigenvalues also could be an interesting topic. To shorten the computer time, algebraic simplification of Bartlett equations could be done. Under non-normality, adopting robust methods such as M estimators could be an important extension. Another interesting topic would be to examine dimension reduction in a correlation matrix.

Finally, it would be useful to develop guidelines for a general set of methods for exploratory and confirmatory principal component analysis. Especially, in confirmatory PCA consisting of several types of hypothesis testings, the future research could be driven by the demands of applications. This present research is one part of the confirmatory PCA.
REFERENCES


APPENDICES
APPENDIX A

Notation
Notation

\text{diag}(a) \quad \text{diagonal matrix having diagonal elements as a vector } a \\
\text{det} \quad \text{determinant} \\
\text{tr} \quad \text{trace} \\
\exp \quad \text{exponential function} \\
\text{dim} \quad \text{dimension} \\
\mathcal{O}(p) \quad \text{group of orthogonal } p \times p \text{ matrices} \\
\text{vec}(A) \quad \text{stacking the columns of } A \text{ into a single column vector} \\
\text{dvec}(a, p, p) \quad \text{converging a vector } a \text{ to a } p \times p \text{ matrix} \\
\otimes \quad \text{direct or Kronecker product} \\
\oplus \quad \text{direct sum} \\
\odot \quad \text{the operation is to be performed element-wise} \\
\mathcal{R}(T) \quad \text{subspace spanned by the columns of a matrix } T \\
N_p(\mu, \Sigma) \quad p\text{-variate normal distribution with mean } \mu \text{ and covariance matrix } \Sigma \\
W_p(n, \Sigma) \quad p \times p \text{ matrix variate Wishart distribution with } n \text{ degrees of freedom and covariance matrix } \Sigma \\
L(\cdot) \quad \text{likelihood function} \\
\ell(\cdot) \quad \text{loglikelihood function} \\
\|X\|^2 \quad \text{norm of } X \text{ i.e } \sum_{i=1}^{p} X_i^2 \\
L \quad \text{the matrix } p^2 \times p \text{ of } \sum_{i=1}^{p} (e_j^p \otimes e_j^p) e_j^p \\
I_p \quad \text{the } p \times p \text{ identity matrix} \\
e_j^p \quad \text{the } j^{th} \text{ column of } I_p \\
I_{(p,p)} \quad \text{the } p^2 \times p^2 \text{ commutation matrix} \\
D_G^{(i)} \quad \text{the } i^{th} \text{ derivative of vec } G \text{ with respect to } \mu' \text{ evaluated at } \mu = 0 \\
D_\lambda^{(i)} \quad \text{the } i^{th} \text{ derivative of } \lambda \text{ with respect to } \varphi' \\
D_p \quad \text{the duplication matrix} \\
N_p \quad \text{the matrix } (I_{p^2} + I_{(p,p)}) / 2 \\
I_0 \quad \text{Fisher's information} \\
\bar{I}_0 \quad \text{the average Fisher's information} \\
\psi(n) \quad \text{digamma function}
APPENDIX B

List of Matlab Subprograms
The following list describes the Matlab subprograms used in the simulation.

1. commute.m: Computes the permutation matrix, \( I_{(a,b)} \).

2. dup.m: Computes the duplication matrix, \( D_p \).

3. kron.m: Computes Kronecker products of three or more input matrices.

4. make_E.m: Constructs the elementary matrix, \( E_{ip} \).

5. oplus.m: Computes the direct sum of two or more matrices.

6. vec.m: Stacks the columns of a matrix into a single column vector.

7. dvec.m: Converges a column vector into a matrix.

8. Pvechv2.m: Constructs the matrices of \( A_1, A_2, L \), and \( (I_{p^2} - I_{(p,p)})A_1 \).

9. solve_eta.m: Given parameter \( \mu \), solves for the implicit parameter \( \eta \) under model (2.31).

10. solve_eta1.m: Given parameter \( \mu \), solves for the implicit parameter \( \eta \) under model (2.33).

11. solve_eta2.m: Given parameter \( \mu \), solves for the implicit parameter \( \eta_z \) under model (2.66).

12. solve_\tau.m: Given parameter \( \varphi \), solves for the implicit parameter \( \tau \) and computes \( \xi \) under model (2.11).
13. find_guess.m: Computes initial guesses for $\Gamma_0$ and $\Lambda_0$.

14. fisher_scoring1.m: Updates the parameters $\mu$ and $\varphi$ under the parameterization of $\lambda$ in (2.4).

15. fisher_scoring2.m: Updates the parameters $\mu$ and $\varphi$ under the parameterization of $\lambda$ in (2.9).

16. find_Vstar.m: Computes $V^* = (V_3 \ V_4)$ ensured that the matrix of derivatives in (2.35) is nonsingular. This subprogram is used for estimating parameters.

17. find_Vstar_bart.m: Computes $V^* = (V_3 \ V_4)$ by using the row reduced echelon form to increase the number of zero entries in the matrices. This subprogram is used to compute a Bartlett correction to help reduce memory and run faster.

18. solve_mle.m: Computes the maximum likelihood estimates of $\Gamma$, $\Lambda$, and $\Sigma$ under model (2.31) and the parameterization of $\lambda$ in (2.4).

19. solve_mle1.m: Computes the maximum likelihood estimates of $\Gamma$, $\Lambda$, and $\Sigma$ under model (2.33) and the parameterization of $\lambda$ in (2.4).

20. solve_mle2.m: Computes the maximum likelihood estimates of $\Gamma$, $\Lambda$, and $\Sigma$ under model (2.31) and the parameterization of $\lambda$ in (2.9).

21. Bartlett.m: Computes $E \left( D(\hat{\theta}; \theta) \right)$ under model (2.31) and the parameterization of $\lambda$ in (2.4).
22. Bartlett1.m: Computes \( E\left(D(\hat{\theta}; \theta)\right) \) under model (2.33) and the parameterization of \( \lambda \) in (2.4).

23. Bartlett2.m: Computes \( E\left(D(\hat{\theta}; \theta)\right) \) under model (2.31) and the parameterization of \( \lambda \) in (2.9).

24. Exp_Da.m: Computes \( E\left(D(\hat{\theta}; \theta)\right) \) under an unrestricted \( H_a \).

25. schott_func.m: Computes \( T_{k,q}, T^*, T^{**} \), and their \( p \)-values.

26. CI function: Computes a local minimizer \( c_0 \) such that the \( p \)-value for testing

\[
H_0: \frac{C_l^\lambda}{\nu_0} = c_0 \text{ is equal to } \alpha \text{ by minimizing } (p\text{-value} - \alpha)^2.
\]
APPENDIX C

Programming Codes
The following codes for Matlab version 5.3 and 6.1 (Mathworks Inc., 2000) can be used to compute the maximum likelihood estimates of $\Gamma$, $\Lambda$, and $\Sigma$, likelihood ratio test statistics, and Bartlett-corrected test statistics, including $p$-values. Matrices can be stored as sparse matrices to reduce memory. The programming codes numbered 1-9 from the preceding list can be found on Boik's webpage:


**solve_eta2.m function.**

```matlab

% given a solution for nu, solve for the implicit parameter eta
% eta1 must be zero then solve for eta2

[p,q]= size(A);Ip = speye(p);Dp = dup(p);

% find Mc which is Ip taken off M's column

b = M'*[1:p]'; Mc = Ip ; Mc(:,b) = 0;

% initial guess of eta2

eta2 = vech(Ip);
C = kron(Mc',A'*Gam);
VG = (A1*V3*nu)+(A2*eta2);
G = dvec(VG,p,p);

% find eta2 using Newton's method

while norm(G*G'-Ip,'fro')+ norm(C*vec(G),'fro') > 1e-8
    F = Dp'*vec(G*G'-Ip);
    J = 2*Dp'*kron(G,Ip)*A2;
    Jinv = inv(J);
    check = Jinv*F;
    eta2 = eta2-check;
    VG = (A1*V3*nu)+(A2*eta2);
    G = dvec(VG,p,p);
end
```

solve_tau.m function.

function [Xi,tau] = solve_tau(phi,Xi,T4,V1,V2,C2)
  
  tau = inv(V1'*V1)*V1'*Xi;

find_guess.m function.

function [Gam0,lam0] = find_guess(A,U,S,p,q,m,order)
  
  lam0 = diag(lhat);

if order == 2
  lam0 = lam0([p:-1:1]);
  Ghat = Ghat(:,[p:-1:1]);
  Lam0 = diag(lam0);
end;

Ghat1 = Ghat(:,1:m);Ghat2 = Ghat(:,m+1:p);
%%% begin with any orthogonal matrix %%

Q01 = orth(randn(m,m));
check = Q01; Q1=ones(m,m);

%%% do loop until Q1 converges %

while (norm(Q1-check)) > 1e-4
  check = Q01;
  B2 = Ghat1*Q01';
  B2 = B2(:,q+1:m);
  [U3,D3,V3] = svd(B2'*U);
  h = sum(diag(D3)>1e-20);
  F = V3(:,1:h)*U3';
  Astar = [ A U*F];
  [U4,D4,V4] = svd(Ghat1'*Astar);
  h = sum(diag(D4)>1e-20);
  Q1 = V4(:,1:h)*U4';
%
%%% check if tr close to m %%
  tr1 = trace(Q1'*Astar'*Ghat1);
%
Q01 = Q1;
end; % end while %

Ac = null(full(Astar)');
[U5,D5,V5] = svd(Ghat2'*Ac);
  h = sum(diag(D5)>1e-20);
  Q2 = V5(:,1:h)*U5';
%
%%% check if tr close to p-m %%
  tr2 = trace(Q2'*Ac'*Ghat2);
%
Q = oplus(Q1,Q2);
Gam0 = [Astar Ac]*Q;

end function %%

fisher_scoring1.m function.

function [mu,newphi,check_conv] = fisher_scoring1(n,phi,Gam,S,T1,T2,Dg1,
L,check_conv0)

%%% one iteration of Fisher_scoring algorithm %%

[p,p]=size(Gam);w= size(Dg1,2);
Ip2 = speye(p^2);Ipp = commute(p,p);
Np = (Ip2+Ipp)/2;
check_eig = 1;
%%% set mu to zero %%
%%% Zeta0 = [zeros(w,1); phi];
%%% construct matrix of F1 %%
    lam = T1*exp(T2*phi);
    Lamda = dvec(L*lam,p,p);
    Sigma = Gam*Lamda*Gam';
    invsigma = inv(Sigma);
    F11 = 2*Np*kron(Gam*Lamda,Gam)*Dg1;
    D11 = T1*diag(exp(T2*phi))*T2;
    F12 = kron(Gam,Gam)*L*D11;
    F1 = [ F11 F12 ];
%%% construct information matrix %%
%%% pre = (n/2)*F1'*kron(invsigma,invsigma);
    ihat = pre*F1;
    invihat = inv(ihat);
%%% construct score function %%
%%% Szeta = pre*vec(S-Sigma);
    delta = invihat*Szeta;
    Zeta = Zeta0 + delta;
%%% check if it is on the right direction but too far %%
%%% the magnitude is reduced %%
%%% while norm(delta(1:w,:)) > 0.25
    delta = delta/2;
    Zeta = Zeta0 + delta;
end; % end while %
check_conv = Szeta'*delta;
while check_conv > check_conv0
    delta = delta/2;
    Zeta = Zeta0 + delta;
    check_conv = Szeta'*delta;
end; % end while %
%%% obtain mu and newphi %%
%%% mu = Zeta(1:w,:);
    newphi = Zeta(w+1:end,:);
%% end function 

fishe_scoring2.m function.

function [mu,newphi,check_conv] = fishe_scoring2(n,phi,Gam,S,Dg1,
L,check_conv0,Xi,V2,T3,T4)

%%
%% one iteration of Fisher scoring algorithm 

[p,p]=size(Gam);w= size(Dg1,2);
Ip = speye(p); Ipp = speye(p^2);
Ipp = commute(p,p);
Np = (Ip2+Ipp)/2;
check_eig = 1;

%%
%% set mu to zero 

Zeta0 = [zeros(w,1); phi];
onep = ones(p,1);

%%
%% construct F1 

wi = inv(onep'gT3*gexp(T4*X));
lam = wi*phi(1)gT3*gexp(T4*X);
Lambda = dvec(L*lamp,p,p);
Sigma = Gam*Lambda*Gam';
invSigma = inv(Sigma);
F11 = 2*Np*kron(Gam*Lambda,Gam)*Dg1;
Hp = sparse(Ip - (lam/phi(1))*onep');
W1 = sparse(diag(exp(T4*X))*T4);

%%
%% Dl1 with respect to phi0 and phi 

Dl1 = [wi*T3*gexp(T4*X) wi*phi(1)gHp*T3*gW1*V2];
F12 = kron(Gam,Gam)*L*Dl1;
F1 = [ F11 F12];

%%
%% construct information matrix 

pre = (n/2)*F1'gkron(invSigma,invSigma);
ihat = pre*F1;
invihat = inv(ihat);

%%
%% score function 

Szeta = pre*vec(S-Sigma);
delta = invihatgSzeta;
Zeta = Zeta0 + delta;

%%
%% check if is on the right direction but too far %%
%% the magnitude is reduced %%
%%
while norm(delta(1:w,:)) > 0.25
    delta = delta/2;
    Zeta = Zeta0 + delta;
end; % end while 
check_conv = Szeta'*delta;
while check_conv >= check_conv0
    delta = delta/2;
    Zeta = Zeta0 + delta;
    check_conv = Szeta'*delta;
end; % end while 
%%
%% obtain mu and newphi %%
%%
mu = Zeta(1:w,:);
newphi = Zeta(w+1:end,:);
%% end function %%

find_Vstar.m function.

function [V3,V4] = find_Vstar(A,A1,M,Gam)

%%
%% find V=V3 V4) ensured that C*(Ip2-Ipp)*A1 is invertable %%
%%
[p,q]= size(A);
Ip = speye(p);
Ip2 = speye(p^2);
Ipp = commute(p,p);

%%
%% M is the first m columns of Ip %%
%% find Mc which is Ip taken off M's column %%
%%
b = M'*[1:p]'; Mc = Ip ; Mc(:,b) = [];

%%
%% find V3 and V4 using the SVD %%
%% V3'V3= I, V4'V4 = I and V3'V4 = 0 %%
%%
C = kron(Mc',A'*Gam);
Cin = C*(Ip2-Ipp)*A1;
[U,D,V]=svd(full(Cin));
h = sum(diag(D)>1e-20);
%%
%% obtain V4 and V3 %%
%%
V4 = V(:,1:h);
V3 = null(V4);  
V3 = sparse(V3);
V4 = sparse(V4);
%% end function %

find_Vstar_bart.m function.

function [V3,V4] = find_Vstar_bart(A,A1,M,Gam)
%%
%% find V=(V3 V4) %%
%% using rref to increase zero entries %%
%%
[p,q]= size(A);
Ip = speye(p);
Ip2 = speye(p^2);
Ipp = commute(p,p);
%%
%% Mc is the last p-m columns of Ip %%
%%
b = M'*[1:p]'; Mc = Ip ; Mc(:,b) = []; 
C = kron(Mc',A'*Gam);
Cin = C*(Ip2-Ipp)*A1*A1';
AV4 = rref(Cin)';
[U,D,V]=svd(full(Cin));
h = sum(diag(D)>1e-20);
AV4 = AV4(:,1:h);
%%
%% obtain V4 b/c A1'A1V4=V4 %%
%%
V4 = A1'*AV4;
%%
%% obtain V3 which is orthogonal to V4 %%
%%
V3 = null(full(V4)');
AV3 = rref(V3'*A1')';
V3 = A1'*AV3;
V3 = sparse(V3);
V4 = sparse(V4);
%% end function %%

solve_mle1.m function.

function [Gamma,Lamda,phi,La,loglk,conv] = solve_mle1(n,A,A1,A2,L,M, 
Gam,lm,S,T1,T2,Dgs) 
%%
%% compute maximum likelihood estimates of Gamma and Lamda

[p,q]= size(A); Ip = speye(p);
%%
%% initial guess of zeta

phi = inv(T2'*T2)*T2'*log(inv(T1'*T1)*T1'*lam); G = ones(p,p); loglk = []; i = 0;
Gamma = Gam;
check_conv = 1; check_conv0 = 1e8;
 conv = 1;
%%
%% check_conv = Szeta*sinvihat*Szeta

La = -(n*p/2)-(n/2)*log(det(S)); S0 = S;
while sqrt(check_conv)> 1e-6
  S0 = Gam'*S0*Gam;
  A = rref(A'*Gam)'
  Gam = Ip;
  [V3,V4] = find_Vstar(A,A1,M,Gam);
  Dg1 = Dgs*V3;
  [mu,newphi,check_conv] = fisher_scoring1(n,phi,Gam,S0,T1,T2,Dg1,
L,check_conv0);
  check_conv0 = check_conv;
%%
%% check convergence

if max(abs(T2*newphi)) > 14
    conv = 0;
    return;
end % end if

[G,eta2] = solve_eta2(A,A1,A2,V3,M,mu,Gam);
phi = newphi;
Gam = G;
Gamma = Gamma*G;
%%
%% computer log likelihood function without constant

lamda = T1*exp(T2*phi);
Lamda = dvec(L*lamda,p,p);
Sigma = Gam*Lambda*Gam';
invsigma = inv(Sigma);
loglk = [loglk ; -(n*trace(S0*invsigma))/2-(n*log(det(Sigma)))/2
check_conv];
end;  % end of while loop %
%% end function

solve_mle2.m function

function [Gam,Lamda,phi0,Xi,La,loglk,conv] = solve_mle2(n,A1,A2,L,Gam, lam,S,Dgs,T3,T4,C2)
  %
  % compute maximum likelihood estimate of Gamma,Lamda,phi0 

[p,p]= size(Gam);  Ip = speye(p);
[p,q3]= size(T3); oneq3 = ones(q3,1); onep = ones(p,1);
% initial guess of Xi 

X = [T4 oneq3 ];
Xi = inv(X'*X)*X'*log(inv(T3'*T3)*T3'*lam);
Xi(end) = []; 
G = ones(p,p); loglk = []; i = 0;
check_conv = 1; check_conv0 = 1e8;
conv = 1;
  La = -((n*p)/2)-(n/2)*log(det(S)); 
  trs = trace(S);
  W = diag(exp(T4*Xi))*T4;
  [U,D,V] = svd(C2'*W);
  h = sum(diag(D)>1e-20);
  V1 = V(:,1:h);
  V2 = V(:,h+1:end);
  phi1 = inv(V2'*V2)*V2'*Xi;
  [Xi,tau] = solve_tau(phi1,Xi,T4,V1,V2,C2);
  phi = [trs;phi1];
%
% check_conv = Szeta'*invihat*Szeta 
%  while sqrt(check_conv)> 1e-6
    W = diag(exp(T4*Xi))*T4;
    [U,D,V] = svd(C2'*W);
    h = sum(diag(D)>1e-20);
    V1 = V(:,1:h);
    V2 = V(:,h+1:end);
    phi(2:end) = inv(V2'*V2)*V2'*Xi;
    [Xi,tau] = solve_tau(phi(2:end),Xi,T4,V1,V2,C2);
    [mu,newphi,check_conv] = fisher_scoring2(n,phi,Gam,S,Dgs,L,check_conv0,Xi,V2,T3,T4);
    check_conv0 = check_conv;
    [Xi,tau] = solve_tau(newphi(2:end),Xi,T4,V1,V2,C2);
%
% check convergence
```matlab
%%
if max(abs(T4*Xi)) > 14
    conv = 0;
    return;
end % end if %
eta = solve_eta(A1,A2,mu);
VG = A1*mu+A2*eta;
G = dvec(VG,p,p);
Gam = Gam*G;
phi(1) = newphi(1);
%%
%% computer log likelihood function without constant %%
%%
lamda = phi(1)*T3*exp(T4*Xi)/(onep'*T3*exp(T4*Xi));
Lamda = dvec(L*lamda,p,p);
Sigma = Gam*Lamda*Gam';
invsigma = inv(Sigma);
loglk = [loglk ; [ -(n*trace(S*invsigma))/2-(n*log(det(Sigma)))/2
        check_conv]]; end; %% end of while loop %%
phi0 = phi(1); %% end function %

Bartlet1.m function.

function elrt = Bartlet1(Gam,Lam,phi,T1,T2,A2,L,V3,n,m,nu,Dgs)
    p = length(Gam);
    ci = size(T1,2);
    k = length(m);
    df=round(sum(nu));
    Idf=speye(df);
    Ip = speye(p);
    Ic = speye(ci);
    Idf2=speye(df^2);
    permdf=commute(df,df);
    permp = commute(p,p);
    permp2 = commute(p^2,p^2);
    permp1 = commute(nu(1),nu(1));
    Ipp = speye(p^2);
    Ndf=Idf2+permdf;
    tNp = (Ipp+permp);
    Impp = Ipp-permp;
    Ik = speye(k);
    Dp = dup(p);
    Lam = sparse(Lam);
    Sigma = Gam*Lam*Gam';
```
Sig = Sigma;
Sigi = inv(Sigma);
for i1=1:2
    I{i1}=speye(nu(i1));
    Ip{i1}=speye(nu(i1)*p);
    for i2=i1:2
        if isempty(nu(i1)*nu(i2))==0
            I_{i1,i2}=commute(nu(i1),nu(i2));
            I_{i2,i1}=I_{i1,i2}';
        else
            I_{i1,i2}=[];
            I_{i2,i1}=[];
        end
    end
end
I11 = speye(nu(1)^2);
N1 = (I11+I_{1,1})/2;
nparam=round(sum(nu));
e1rt=nparam;
%%
%% Create Dg(1), Dg(2),Dg(3)  %%
%%
D1_g_m = sparse(Dgs*V3);
K4 = sparse(kron(Ip,vec(Ip)',Ip));
kdg1 = sparse(kron(D1_g_m,D1_g_m));
K = K4*kdg1;
D2_g_mm = sparse(A2*Dp'*K);
K1 = sparse(kron(I_{1,1},permp1));
K2 = sparse(kron(D2_g_mm,D1_g_m));
K3 = sparse(kK2*permp2*K2*K1);
K5 = sparse(kron(D1_g_m,permp*D2_g_mm));
K6 = sparse(kron(Ipp,permp));
clear K K1 K2 Dg0 Dg W1
D3_gmmm = -A2*Dp'*K4*(K6*K3+K5);
clear K3 K4 K5 K6
%%
%% Create D1(1),D1(2),D1(3)  %%
%%
D1_l_p = sparse(T1*diag(exp(T2*phi))*T2);
W2 = sparse(1,1);
W3 = sparse(1,1);
for i = 1 : ci
    ei = Ic(:,i);
    W2 = W2 + kron(ei*ei',ei*ei');
    W3 = W3 + kron3(ei*ei',ei*ei',ei');
end;
ket1 = sparse((kron(exp(T2*phi),T1'))');
D2_l_pp = sparse(ket1*W2*kron(T2,T2));
D3_l_ppp = sparse(ket1*W3*kron3(T2,T2,T2));
clear W2 W3

Sig=sparse(Sig);
Sigi=sparse(Sigi);
V2i=kron(Sigi,Sigi);
index=[1:2]';
jj=find(nu==0);
index(jj)=[];

%%
%% Use D1_{i} for first derivative of vec Sigma with respect to theta{i}'
%% Theta = (mu ' phi)' %
%% Use D2_{i,j} for second derivative of vec Sigma with respect to %
%% theta{i}' x theta{j}'
%% Use D3_{i,j,k} for third derivative of vec Sigma with respect to %
%% theta{i}' x theta{j}' x theta{j}'
%%
kkf = sparse(kron(Gam,Gam));
k1 = sparse(kron(Gam*Gam,Gam));
k1g = sparse(kron3(Gam,vec(Lam)',Gam));
k1ig = sparse(kron3(Gam,vec(lam)',Gam));
k1d = sparse(kron(lam*D1_l_p,D1_g_m));

D1_{1} = sparse(tN*p*kron(Gam*Gam,Gam)*D1_g_m);
D1_{2} = sparse(kkf*l*D1_l_p);
D2_{1,1} = tNp*((k1g*((-kdf)))+k1*D2_g_mm);
D2_{2,2} = k1f*l*D2_l_pp;
D2_{2,1} = tNp*kig*k1d;
D2_{2,2} = D2_{2,1}*I_{1,2};
k1d2 = kron(-D1_g_m,D2_g_mm);
k1d2 = kron(permp*D2_g_mm,D1_g_m);
D4 = k1d2+2*k1d2kron(I{1},N1);
D3_{1,1,1} = tNp*(k1g*D3_g_mmm+k1g*D4);
clear D4
D3_{2,2,2} = kkg*l*D3_l_pp;
D3_{2,2,1} = tNp*kkg*kron(l*D3_l_pp,D1_g_m);
D3_{2,1,2} = D3_{2,2,1}*kron(I{2},I_{1,2});
D3_{1,2,2} = D3_{2,1,2}*kron(I_{1,2},I_{1,2});
D3_{0,0} = kron3(lam,permp,lam)*kdf;
D3_{2,1,1} = tNp*((kkg*kron(l*D1_l_p,D2_g_mm)) +
(kron3(vec(lam*D1_l_p)',Gam,Gam)*kron(I{2},D3_{0,0}));
clear D3
D3_{1,2,1} = D3_{2,1,1}*kron(I{1,2},I{1});
D3_{1,1,2} = D3_{2,1,2}*kron(I{1},I_{1,2});
clear kgg kgl kglg kgig kld kd1d2 kpd2d1

%% Computation of F1 %%
F1=sparse(1,1);
F2=sparse(1,1);
F3=sparse(1,1);
for l=1:length(index)
    ll=index(l);
    F1=F1+D1_{ll}*make_E(ll,nu);
end;
Itheta=sparse(F1'*V2i*F1/2);
Ithetai=inv(Itheta);
Fisher_Info=Itheta;
clear D1_{11} D1_{22}

%% Computation of F2, F11, & F3 %%
nindex=length(index);
for i1=1:nindex
    l1=index(i1);
    E1=make_E(l1,nu);
    for i2=1:nindex
        l2=index(i2);
        E2=make_E(l2,nu);
        if isempty(D2_{11,l2}) == 0
            F2=F2+D2_{11,l2}*kron(E1,E2);
        end;
    end
end
F11=reshape(F2,p^2*nparam,nparam);
clear D2_{11,1} D2_{2,2} D2_{1,2} D2_{2,1}
for i1=1:nindex
    l1=index(i1);
    E1=make_E(l1,nu);
    for i2=1:nindex
        l2=index(i2);
        E2=make_E(l2,nu);
        E12=kron(E1,E2);
        for i3=1:nindex
            l3=index(i3);
            if isempty(D3_{ll,l2,l3}) == 0
                E3=make_E(l3,nu);
                E123=kron(E12,E3);
                F3=F3+D3_{ll,l2,l3}*E123;
end
end
end
clear D3_{1,1,1} D3_{2,2,2} D3_{2,2,1} D3_{2,1,1} D3_{1,2,2} D3_{2,1,2} D3_{1,2,1} D3_{1,1,2}
zeta=zeros(10,1);
Ithetai2=kron(Ithetai,Ithetai);
ithetai=vec(Ithetai);
k3_tmp4=kron3(Ip,vec(Sigma)',Ip);
V3=kron3(Sigi,vec(Sigi),Sigi);
F2dd=V2i*F2;
F1dd=V2i*F1;
F1ddd=V3*F1;
F1F1=kron(F1,F1);
F1Ithetai=F1*Ithetai;
tmp=F1dd'*F2;
K3s=2*F1ddd'*F1F1-reshape(tmp,df^2,df)'-tmp/2;
tmp=V3'*F1F1;
U1=reshape(F2dd,p^2*df,df)'- ... 
2*reshape(reshape(tmp,df^2,df)',df,df*p^2);
tmp1= kron(Ip,kron(vec(Sigi)',Ip)*kron(Ip,F1));
U2= -6*reshape(F1ddd',p^2*df,p^2)*F2 + ... 
+2*tmp1'*(tNp*permp)*tmp + ... 
reshape(reshape(V2i*F3,p^2*df^2,df)',p^2*df,df^2);
L1=U1*vec(F1Ithetai)/2;
L2=vec(U1*kron(Idf,F1Ithetai))/2;
tmp=kron3(Sigi,vec(Sigi)',Ip)*F1F1;
K4s=12*F1F1'*V3*F2 ...
-9*tmp'*permp*tmp ... 
-2*reshape(F1ddd'*F3,df^2,df^2) ...
-(3/2)*reshape(F2dd'*F2,df^2,df^2);
clear V3 F2dd F1F1 tmp tmp1

zeta(4)=L1' *Ithetai *L1+L2' *kron(permdf,Ithetai)*L2;
clear L1 L2
tmp=K3s*iithetai;
zeta(5)=vec(K3s*Ndf*Ithetai2*K3s')'*ithetai/4+tmp'*Ithetai*tmp/4;
clear tmp
zeta(10)=vec(Ndf*Ithetai2)'*vec(K4s)/12+iithetai' *K4s*iithetai/12;

zeta(1)=(1/2)*iithetai' *vec(U1*kron(Ithetai*F1dd',tNp)*k3_tmp4'*F1);
zeta(2)=(1/3)*iithetai' *vec(K3s*kron(F1Ithetai,F1Ithetai)'*F1dd);
zeta(3)=(1/4)*iithetai' *vec(U1*kron(Ithetai,tNp*kron(Sigma,Sigma))*U1');
zeta(6)=(1/2)*iithetai' *vec(K3s*Ndf*kron(Ithetai,F1Ithetai')*U1');
Ttmp=K3s' *Ithetai*U1;
zeta(7)=itheta1'*Tmp*vec(F1Ithetain)/2;
zeta(8)=(1/6)*vec(kron(F1Ithetain,Ithetain)*Ndf)'*vec(U2);
zeta(9)=itheta1'*U2'*vec(F1Ithetain')/6;
zeta=zeta/n;

elrt=df+sum(zeta);
%%% end function

schott_func.m function

function [Tkq,Tstar1,Tstar2,P_Tkq,P_Tstar1,P_Tstar2] = schott_func(A,S,n,p,q,k)
Ip = speye(p);
df = k*q;
%%% find Tkq using SVD od S %
%%% [Ghat,Lhat,Tmp] = svd(S);
l = diag(Lhat);
Y2 = Ghat(:,k+1:p);
T = 0;
for j = 1:k
    d{j} = [];
    for i = k+1:p
        d{j} = [d{j} ; 1(i)*l(i)/(l(i)-1(j))^2];
    end;
    D{j} = diag(d{j});
    Tj = trace(A'*Ghat(:,j)*Ghat(:,j)'*A*inv(A'*Y2*D{j}*Y2'*A));
    T = T+Tj;
end; % end for
Tkq = n*T;
R = Y2*Y2'*A;
[u,d1,v] = svd(R);
r = sum(diag(d1)>1e-20);
if r == q
    P_Tkq = 1 - chi2cdf(Tkq,df);
end;
if r < q P_Tkq = 0; end;
%%% find Bartlett adjustment factor %
%%% Ghat22 = Ghat(p-q+1:p,k+1:p);
c1 = 0;c2 = 0;c3 = 0;
for j = 1:k
    Sj = Ghat22*D{j}*Ghat22';
\[ Q(j) = \text{Ghat22} \cdot \text{inv(Sj)} \cdot \text{Ghat22}; \]
for \( u = k+1: p \)
\[
\begin{align*}
c11 &= q \cdot 1(u)/(1(u) - 1(j)); \\
c12 &= 2 \cdot Q(j \cdot u-k, u-k)^2 \cdot 1(u) \cdot 1(j)^3 \cdot 3 \cdot (2 \cdot 1(j) + 3 \cdot 1(u)) / (1(u) - 1(j))^6; \\
c13 &= Q(j \cdot u-k, u-k) \cdot 1(u) \cdot 1(j) \cdot 2 \cdot 1(u) \cdot 1(j) + 4 \cdot 1(u) \cdot 1(j) \cdot 2 \cdot 1(j)^3 / (1(u) - 1(j))^5; \\
c1 &= c1 + (c11 + c12 + c13); \\
\end{align*}
\]
for \( v = k+1: p \)
\[
\begin{align*}
c21 &= (Q(j \cdot v-k, v-k) + Q(j \cdot v-k, v-k) \cdot 1(u) \cdot 1(v) \cdot 1(j) \cdot 2 / ((1(u) - 1(j))^2) \cdot (1(v) - 1(j))^2); \\
deno &= ((1(u) - 1(j))^2) \cdot (1(v) - 1(j))^4); \\
c22 &= 2 \cdot Q(j \cdot u-k, u-k) \cdot 2 \cdot 1(u) \cdot 1(v) \cdot 1(j)^3 \cdot (1(u) \cdot 1(v) \cdot 1(j) - 3 \cdot 1(j)^3) \cdot (1(u) + 1(v)) \cdot (1(j) - 1(u)) + 1(u) \cdot (1(j) - 1(v)) + 1(j)^2) / deno; \\
c23 &= 2 \cdot Q(j \cdot u-k, u-k) \cdot 1(u) \cdot 1(v) \cdot 1(j)^3 \cdot 3 \cdot 1(u) \cdot 1(v) \cdot 1(j) \cdot 3 \cdot 3 \cdot 1(u) \cdot 1(v) \cdot 1(j) \cdot 3 \cdot 3 \cdot 1(u) \cdot 1(v) \cdot 1(j) / deno; \\
c2 &= c2 + (c21 + c22 - c23); \\
\end{align*}
\]
end; end; for \( i = 1:k \)
\[
\begin{align*}
\text{for } j = 1:k \\
\text{if } i \neq j \\
\text{for } u = k+1: p \\
\text{c31} &= Q(j \cdot u-k, u-k) \cdot 1(u) \cdot 1(j) \cdot 1(i) \cdot 1(u) \cdot 2 \cdot 1(j) \cdot 1(i) / ((1(u) - 1(j))^2 \cdot (1(u) - 1(i))^2); \\
c3 &= c3 + c31; \\
\end{align*}
\]
end; end; Ckq = 1/2* (k-1)+ inv(k*q) * (c1 - c2 - c3); Tstar1 = (1-Ckq/n)*Tkq; Tstar2 = Tkq/(1+Ckq/n);

\[
\begin{align*}
P_{\text{Tstar1}} &= 1-\text{chi2cdf}(\text{Tstar1}, df); \\
P_{\text{Tstar2}} &= 1-\text{chi2cdf}(\text{Tstar2}, df); \\
\text{end function} \% \%
\end{align*}
\]

\textbf{Cl function.}

function f = CI(c01,n,p,A1,A2,L, Gam ini, lhat, S, Dgs, T3, T4, C1, loglk)  
\[
p = \text{pvalue}(c01,n,p,A1,A2,L, Gam ini, lhat, S, Dgs, T3, T4, C1, loglk); \\
f = (p-0.05)^2;
pvalue function.

```matlab
function p = pvalue(c01,n,p,A1,A2,L,Gam_ini,lhat,S,Dgs,T3,T4,C1,loglk1)
    C2 = T3'*(C1-ones(p,1)*c01);
    [Gam0,Lam0,phi00,Xi,La,loglk,conv0] = solve_mle2(n,A1,A2,L,Gam_ini,lhat,S,
        Dgs,T3,T4,C2);
    LRT = 2*(loglk1-loglk(end,1));
    p = 1 - chi2cdf(LRT,1);
```

The matrix $C^*$ from Example 2.

```matlab
C = []; L = []; l = [];
Ip = speye(p);
linear = [1;2;3;4;5;6;7];
for i = 1:6
    l{i} = linear.^i;
    H = [ones(p,1) C];
    P = ppo(H);
    Ci = (Ip-P)*l{i};
    C = [C Ci];
end;
for i = 1:6
    C(:,i) = C(:,i)/sqrt(C(:,i)'*C(:,i));
end;
```