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

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.

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 Robert J. Boik (Signature) 7/22/02 Date

Approved for the Department of Mathematical Sciences

Kenneth L. Bowers Kenneth L. Bowers (Signature) 7/25/02 Date

Approved for the College of Graduate Studies

Bruce McLeod Bruce R. McLeod (Signature) 7-25-02 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 July, 22 2002

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 truely 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

LIST OF FIGURES

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 $\mathrm{diag}(a_1, a_2, \ldots, a_n)$ and the trace of matrix $\boldsymbol{A}$ is denoted by $\mathrm{tr}(\boldsymbol{A})$. The $p \times p$ identity matrix is denoted by $\boldsymbol{I}_p$.

Suppose that the vector of original variables $\boldsymbol{y} = (y_1\ y_2\ \cdots\ y_p)'$ has a positive definite covariance matrix $\boldsymbol{\Sigma}$. 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 $\boldsymbol{\Sigma}$ is symmetric and positive definite, all eigenvalues of $\boldsymbol{\Sigma}$ are real and positive. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$ be the ordered eigenvalues of $\boldsymbol{\Sigma}$ and let $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1\ \boldsymbol{\gamma}_2\ \cdots\ \boldsymbol{\gamma}_p)$ be a $p \times p$ orthogonal matrix such that

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}', \tag{1.1}$$

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

$$\boldsymbol{z} = \boldsymbol{\Gamma}'\boldsymbol{y}. \tag{1.2}$$

The covariance matrix of $\boldsymbol{z}$, $\boldsymbol{\Sigma}_z$, can be written as

$$\boldsymbol{\Sigma}_z = \boldsymbol{\Gamma}'\boldsymbol{\Sigma}\boldsymbol{\Gamma}.$$

Substituting $\boldsymbol{\Sigma}$ from (1.1) yields

$$\boldsymbol{\Sigma}_z = \boldsymbol{\Gamma}'(\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}')\boldsymbol{\Gamma} = \boldsymbol{\Lambda}.$$

Hence the components $\boldsymbol{z}_1 = \boldsymbol{\gamma}_1'\boldsymbol{y}$, $\boldsymbol{z}_2 = \boldsymbol{\gamma}_2'\boldsymbol{y}, \ldots, \boldsymbol{z}_p = \boldsymbol{\gamma}_p'\boldsymbol{y}$ are uncorrelated. The variance of z_i is λ_i. The maximum value of variance of $\boldsymbol{\gamma}_1'\boldsymbol{y}$ satisfying $\boldsymbol{\gamma}_1'\boldsymbol{\gamma}_1 = 1$ is equal to λ_1, the largest eigenvalue of $\boldsymbol{\Sigma}$. This maximum occurs when $\boldsymbol{\gamma}_1$ is an eigenvector of $\boldsymbol{\Sigma}$ corresponding to λ_1. That is,

$$\max_{\boldsymbol{l}'\boldsymbol{l}=1} \boldsymbol{l}'\boldsymbol{\Sigma}\boldsymbol{l} = \boldsymbol{\gamma}_1'\boldsymbol{\Sigma}\boldsymbol{\gamma}_1 = \lambda_1.$$

Then, the component z_1 has the largest variance λ_1, z_2 has the second largest variance λ_2, and so on. The total variance of the p principal components is equal to the total variance of the original variables so that

$$\sum_{i=1}^{p} \lambda_i = \mathrm{tr}(\boldsymbol{\Sigma}).$$

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

$$t = \frac{\lambda_j}{\mathrm{tr}(\boldsymbol{\Sigma})}$$

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.

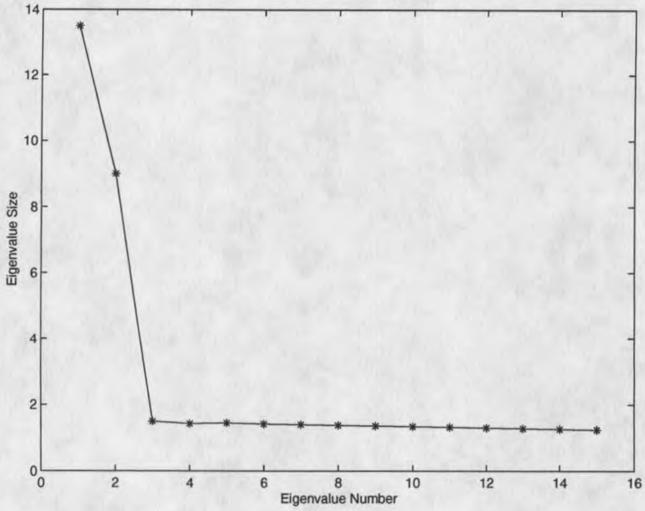
Reduction in the Number of Components

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 $\log(\boldsymbol{\lambda})$ rather than $\boldsymbol{\lambda}$. This plot is known as a log-eigenvalue diagram.

The second method is testing the hypothesis that the last $m = p - k$ eigenvalues of $\boldsymbol{\Sigma}$ are equal. Anderson (1963) derived the likelihood ratio test for $\mathrm{H}_{0k}\colon \lambda_{k+1} = \cdots = \lambda_p = \lambda$ against the alternative that at least two of the last m eigenvalues are different.

Figure 1. Ideal Scree Graph.



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 $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. If $\boldsymbol{y}$ can be written as $\boldsymbol{C}\boldsymbol{w} + \boldsymbol{\varepsilon}$, where $\boldsymbol{C}$ is a $p \times k$ matrix with rank k, $\mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}$, $\mathrm{var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_p$, and $\mathrm{var}(\boldsymbol{w}) = \boldsymbol{I}_k$, then $\mathrm{var}(\boldsymbol{y}) = \boldsymbol{\Sigma} = \boldsymbol{C}\boldsymbol{C}' + \sigma^2 \boldsymbol{I}_p$. Write $\boldsymbol{C}\boldsymbol{C}'$ in diagonal form as $\boldsymbol{C}\boldsymbol{C}' = \boldsymbol{\Gamma}_1\boldsymbol{\Lambda}_1\boldsymbol{\Gamma}_1'$ and let $\boldsymbol{\Gamma}_2$ be an orthogonal complement to $\boldsymbol{\Gamma}_1$. That is, $\boldsymbol{I}_p = \boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_1' + \boldsymbol{\Gamma}_2\boldsymbol{\Gamma}_2'$ and $(\boldsymbol{\Gamma}_1 \quad \boldsymbol{\Gamma}_2) \in \mathcal{O}(p)$, where $\mathcal{O}(p)$ denotes group of orthogonal $p \times p$ matrices. The covariance matrix of $\boldsymbol{y}$ can be rewritten as

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \boldsymbol{\Gamma}_1\boldsymbol{\Lambda}_1\boldsymbol{\Gamma}_1' + \sigma^2 \boldsymbol{I}_p \\
&= \boldsymbol{\Gamma}_1\boldsymbol{\Lambda}_1\boldsymbol{\Gamma}_1' + \sigma^2(\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_1' + \boldsymbol{\Gamma}_2\boldsymbol{\Gamma}_2') \\
&= \boldsymbol{\Gamma}_1(\boldsymbol{\Lambda}_1 + \sigma^2)\boldsymbol{\Gamma}_1' + \boldsymbol{\Gamma}_2\sigma^2\boldsymbol{\Gamma}_2'.
\end{aligned}
$$

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

$$\mathbf{\Lambda}_k=\left\{\prod_{i=k+1}^{p} l_i\Big/\left(\sum_{i=k+1}^{p} l_i/m\right)^m\right\}^{n/2}, \tag{1.3}$$

where l_i, $i=1,2,\ldots,p$, are eigenvalues of the sample covariance matrix, $\boldsymbol{S}$. If H_{0k} is true, then the limiting distribution of $-2\ln\mathbf{\Lambda}_k$ is χ^2_ν, 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_{0k} 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-α) percentile of the χ^2_ν distribution.

Lawley (1956) proposed a further improvement in Bartlett's multiplying factor. Lawley's test is to reject H_{0k} whenever

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

Acceptance of H_{0k} 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 $\text{H}_{0m}\colon \lambda_{k+i} = \alpha + \beta x_i,\ x_i = m - i,\ i = 1, 2, \ldots, m$. The likelihood ratio test statistic is

$$\boldsymbol{\Lambda}_m = \frac{\operatorname*{Sup}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}_0} \ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}_0)}{\operatorname*{Sup}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}} \ell(\boldsymbol{\mu}, \boldsymbol{\Sigma})},$$

where $\boldsymbol{\Sigma}_0 = \boldsymbol{\Gamma}_0 \boldsymbol{\Lambda}_0 \boldsymbol{\Gamma}_0'$ and $\boldsymbol{\Lambda}_0 = \text{diag}(\lambda_1, \ldots, \lambda_k, \alpha + \beta x_1, \alpha + \beta x_2, \ldots, \alpha + \beta x_m)$. For $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}$, $\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is maximized when $\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{y}}$. To maximize $\ell(\bar{\boldsymbol{y}}, \boldsymbol{\Sigma}_0)$, Bentler and Yuan showed that $\hat{\lambda}_i = \frac{n}{n+1} l_i$ 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 α and β were numerically obtained by solving the following equations:

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

The asymptotic distribution of $-2 \ln \boldsymbol{\Lambda}_m$ is χ^2_ν 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:

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Gamma}_i \boldsymbol{\Lambda}_i \boldsymbol{\Gamma}_i', \ i = 1, 2, \ldots, g,$$

where the eigenvector matrices $\{\boldsymbol{\Gamma}_i\}$ and eigenvalue matrices $\{\boldsymbol{\Lambda}_i\}$ 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, δ_k, and the last $m = p - k$ eigenvalues, δ_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 $\hat{\delta}_k = \dfrac{\sum_{i=1}^{k} l_i}{\sum_{i=1}^{p} l_i}$ and $\hat{\delta}_m = \dfrac{\sum_{i=k+1}^{p} l_i}{\sum_{i=1}^{p} l_i}$, where l_i, $i = 1, 2, \ldots, p$, are defined in (1.3).

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

$$\text{H}_{0m} \colon \delta_m = \delta_{m,0}, \tag{1.4}$$

where $\delta_{m,0}$ is a known constant. If δ_m is small, then little information is lost by discarding the corresponding principal components. Anderson showed that if $\lambda_k >$

λ_{k+1}, then $\hat{\delta}_m$ is asymptotically distributed as

$$\sqrt{n}(\hat{\delta}_m - \delta_m) \xrightarrow{L} N\left(0, \frac{2\delta_m^2 \sum_{i=1}^{k} \lambda_i^2 + 2(1-\delta_m)^2 \sum_{i=k+1}^{p} \lambda_i^2}{(\sum_{i=1}^{p} \lambda)^2}\right). \tag{1.5}$$

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

$$z^* = \frac{\hat{\delta}_m - \delta_{m,0}}{\sqrt{\hat{V}(\hat{\delta}_m)}}, \tag{1.6}$$

$z_{\alpha/2}$ is the upper $100\alpha/2$ percentage point of the standard normal distribution, and $\hat{V}(\hat{\delta}_m)$ is an estimator of the variance of $\hat{\delta}_m$ which can be computed by replacing λ_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, δ_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

$$\mathbf{\Omega}_q(\omega) = \{\boldsymbol{\lambda};\ \delta_q \geq \omega\},$$

where ω 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,

$$\min_k \left\{ \inf_{\boldsymbol{\lambda} \in \Omega_k(\omega)} P_\lambda(\hat{\delta}_k \geq c) \right\} \geq P^*. \tag{1.7}$$

Huang and Tseng showed that if $\omega \geq 0.5$, $k \leq \frac{p}{2}$, 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

$$\mathrm{H}_0 : \frac{\boldsymbol{C}_1'\boldsymbol{\lambda}}{\mathrm{tr}(\boldsymbol{\Sigma})} = \boldsymbol{c}_0, \tag{1.8}$$

where $\boldsymbol{C}_1$ is a matrix of known constants and $\boldsymbol{c}_0$ is a vector of known constants. This hypothesis is more flexible and general than the hypothesis in (1.4). If $\boldsymbol{C}_1$ is $(\boldsymbol{0}_k' \quad \boldsymbol{1}_{p-k}')'$ and $\boldsymbol{c}_0$ 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 $\frac{\boldsymbol{C}_1'\boldsymbol{\lambda}}{\mathrm{tr}(\boldsymbol{\Sigma})}$. 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 likelihood 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 $\boldsymbol{y}$ are defined as in (1.2). It follows that

$$\boldsymbol{y} = \boldsymbol{\Gamma}\mathbf{z}.$$

The coefficient for obtaining the i^{th} variable from the j^{th} principal component is γ_{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, λ_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 λ_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 λ_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 λ_0. Jolliffe (1972) discussed the appropriate λ_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

$$r_{xy} = \max r_{ij}, \tag{1.9}$$

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 $\mathbf{\Sigma}$ be partitioned as

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

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

$$\mathbf{\Sigma}_{22\cdot 1} = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}.$$

Let $\theta_i,\ i = 1, 2, \ldots, p-k$, be the eigenvalues of $\mathbf{\Sigma}_{22\cdot 1}$ and $\lambda_i,\ i = 1, 2, \ldots, p$, be the eigenvalues of $\mathbf{\Sigma}$. McCabe proposed the following four criteria:

1. $\min |\mathbf{\Sigma}_{22\cdot 1}| = \min \prod_{i=1}^{p-k} \theta_i$,

2. $\min \operatorname{tr}(\mathbf{\Sigma}_{22\cdot 1}) = \min \sum_{i=1}^{p-k} \theta_i$,

3. $\min ||\mathbf{\Sigma}_{22\cdot 1}||^2 = \min \sum_{i=1}^{p-k} \theta_i^2$,

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

where ρ_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

$$\hat{\rho}_{ij} = \hat{\gamma}_{ij}\left(\frac{l_j}{s_i^2}\right)^{1/2},$$

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

$$d = \frac{\|\boldsymbol{X}\hat{\gamma}_i - \boldsymbol{X}_k\hat{\gamma}_i^k\|}{\|\boldsymbol{X}\hat{\gamma}_i\|} = \left(1 - 2\hat{\gamma}_j^{k\prime}\hat{\gamma}_j^k + \frac{\hat{\gamma}_j^{k\prime}\boldsymbol{S}_k\hat{\gamma}_j^k}{l_j}\right)^{1/2},$$

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, $\boldsymbol{S}_k$ is the sample covariance of the k selected variables, and $\boldsymbol{X}_k$ is the sub-matrix of a $n \times p$ column-centered data matrix, $\boldsymbol{X}$, 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 $\mathrm{H}_0\colon \gamma_1 = \gamma_0$, where γ_1 is the eigenvector of the first principal component and γ_0 is a specified vector such that $\gamma_0'\gamma_0 = 1$. If the hypothesis is true, then

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

has the limiting Chi-square distribution with degree of freedom $p-1$, where l_1 is the first eigenvalue of the sample covariance matrix, $\boldsymbol{S}$. 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 $\boldsymbol{H}$ be a $p \times p$ matrix and $\boldsymbol{W}$ be a $p \times p$ positive definite symmetric matrix such that $\boldsymbol{WH}$ is symmetric. These conditions ensure that the eigenvalues and eigenvectors of $\boldsymbol{H}$ are real. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ be the ordered eigenvalues of $\boldsymbol{H}$ and let $\boldsymbol{A}$ 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 : The columns of $\boldsymbol{A}$ lie in the subspace generated by the set of eigenvectors of $\boldsymbol{H}$ associated with m eigenvalues, λ_i, $i \in w$, and

for $q \geq m$,

H_0^* : The eigenvectors of $\boldsymbol{H}$ associated with the m eigenvalues, λ_i, $i \in w$, lie in the subspace generated by the columns of $\boldsymbol{A}$.

Let $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_p$ be the eigenvectors of $\boldsymbol{H}$ and let $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p$ be the eigenvectors of $\boldsymbol{H}_n$, where

$$\sqrt{n}\,\mathrm{vec}(\boldsymbol{H}_n - \boldsymbol{H}) \xrightarrow{L} N(\mathbf{0}, \Xi).$$

Denote the $p \times m$ matrix whose columns are $\boldsymbol{\beta}_i$ for $i \in w$ by $\boldsymbol{\beta}_w$ and denote the $p \times m$ matrix whose columns are $\mathbf{b}_i$ for $i \in w$ by $\mathbf{B}_w$. Let $d_1 \geq d_2 \geq \cdots \geq d_p$ be the ordered eigenvalues of $\boldsymbol{H}_n$, $\boldsymbol{P}_0 = \boldsymbol{\beta}_w(\boldsymbol{\beta}_w'\boldsymbol{\beta}_w)^{-1}\boldsymbol{\beta}_w'$, and $\hat{\boldsymbol{P}}_0 = \sum_{i \in w} \hat{\boldsymbol{P}}_i = \mathbf{B}_w(\mathbf{B}_w'\mathbf{B}_w)^{-1}\mathbf{B}_w'$. Then the test statistic for H_0 is

$$T_n(\boldsymbol{A}) = n\left\{\mathrm{vec}\left[(\boldsymbol{I}_p - \hat{\boldsymbol{P}}_0)\boldsymbol{A}\right]\right\}'\left[\hat{\Xi}(\boldsymbol{A})\right]^{-}\mathrm{vec}\left[(\boldsymbol{I}_p - \hat{\boldsymbol{P}}_0)\boldsymbol{A}\right], \tag{1.10}$$

where

$$\hat{\Xi}(\boldsymbol{A}) = (\boldsymbol{A}' \otimes \boldsymbol{I}_p)\boldsymbol{C}_w' \hat{\Xi}\, \boldsymbol{C}_w(\boldsymbol{A} \otimes \boldsymbol{I}_p),$$

$$\boldsymbol{C}_w = \sum_{i \in w} \sum_{j \notin w} (d_i - d_j)^{-1} \hat{\boldsymbol{P}}_i \otimes \hat{\boldsymbol{P}}_j'.$$

Tyler showed that, under H_0, $T_n(\boldsymbol{A}) \xrightarrow{L} \chi^2_{(p-m)q}$. The test is to reject H_0 if $r = \mathrm{rank}(\hat{\boldsymbol{P}}_0\boldsymbol{A}) < q$ or if $r = q$ and $T_n(\boldsymbol{A}) > \chi^2_{1-\alpha,(p-m)q}$. The hypothesis H_0^* can be tested by using the following approach. Let $\boldsymbol{B}$ be a fixed $p \times (p-q)$ matrix with rank $p-q$ whose columns are orthogonal to $\boldsymbol{A}$, i.e. $\boldsymbol{A}'\boldsymbol{B} = \boldsymbol{0}$. The hypothesis H_0^* can be rephrased as

H_0 : The columns of $\boldsymbol{B}$ lie in the subspace generated by the set of eigenvectors of $\boldsymbol{H}'$ associated with $p-m$ eigenvalues, λ_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 $\boldsymbol{H}$. Let $\hat{\kappa}$ be an estimator of κ, the kurtosis parameter. An estimator of the covariance matrix of $\mathrm{vec}(\boldsymbol{H}_n)$ can be written as

$$\hat{\Xi} = (1+\hat{\kappa})(\boldsymbol{I}_{p^2} + \boldsymbol{I}_{(p,p)})(\boldsymbol{H}_n \otimes \boldsymbol{H}_n) + \hat{\kappa}\, \mathrm{vec}(\boldsymbol{H}_n)[\mathrm{vec}(\boldsymbol{H}_n)]'.$$

The test statistic in (1.10) can be simplified as

$$T_n(\boldsymbol{A}) = n(1+\hat{\kappa}) \sum_{j \notin w} d_j^{-1}\, \mathrm{tr}\left\{\boldsymbol{A}'\hat{\boldsymbol{P}}_i\boldsymbol{A}\left[\boldsymbol{A}'\boldsymbol{B}_w\boldsymbol{D}_j\boldsymbol{B}_w'\boldsymbol{A}\right]^{-1}\right\}, \tag{1.11}$$

where $\boldsymbol{D}_j$ is an $m \times m$ diagonal matrix with entries $d_i/(d_i - d_j)^2$, $j \notin w$. For $q = m = 1$, $T_n(\boldsymbol{A})$ 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 Σ_1 and Σ_2 be the covariance matrices of two independent p-vectors. Flury analyzed the eigenvectors of $\Sigma_1^{-1}\Sigma_2$. Let $\{\boldsymbol{\beta}_j\}_{j=1}^{p}$ be a set of eigenvectors of $\Sigma_1^{-1}\Sigma_2$ and partition $\{\boldsymbol{\beta}_j\}_{j=1}^{p}$ into the first $p-q$ and the last q rows as follows:

$$\{\boldsymbol{\beta}_j\}_{j=1}^{p} = \left\{ \begin{array}{c} \boldsymbol{\beta}_{j1} \\ \boldsymbol{\beta}_{j2} \end{array} \right\}_{j=1}^{p}.$$

The hypothesis of interest is

$$\mathrm{H}_0(p,q)\colon \boldsymbol{\beta}_{j2} = \mathbf{0} \quad \text{for all } j \in w. \tag{1.12}$$

This hypothesis can be formulated in the form H_0^* of Tyler (1981), by putting $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{I}_{p-q} & \mathbf{0}' \end{pmatrix}'$.

Then by using

$$\boldsymbol{B} = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{I}_q \end{pmatrix}, \tag{1.13}$$

the hypothesis can be rephrased as

$\mathrm{H}_0(p,q)$: The columns of $\boldsymbol{B}$ lie in the subspace generated by the eigenvector $\boldsymbol{\alpha}_i$, $(i \notin w)$ of $\Sigma_2\Sigma_1^{-1}$.

The test statistic in (1.10) can be applied. Let $\{\mathbf{b}_j\}_{j=1}^p$ be a set of eigenvectors of $\boldsymbol{S}_1^{-1}\boldsymbol{S}_2$ and partition $\{\mathbf{b}_j\}_{j=1}^p$ into the first $p-q$ and the last q rows as follows:

$$\{\mathbf{b}_j\}_{j=1}^p = \left\{ \begin{array}{c} \mathbf{b}_{j1} \\ \mathbf{b}_{j2} \end{array} \right\}_{j=1}^p .$$

The test statistic is

$$R(p,q) = n \sum_{j \in w} \mathbf{b}_{j2}' \left(\sum_{i \notin w} \frac{k_1 d_j^2 + k_2 d_i d_j}{(d_i - d_j)^2} \mathbf{b}_{i2}\mathbf{b}_{i2}' \right)^{-1} \mathbf{b}_{j2},$$

where $k_i = \lim_{n_i \to \infty} \frac{n}{n_i}$ for $i = 1, 2$ and $d_j, j = 1, 2, \ldots, p$, are the eigenvalues of $\boldsymbol{S}_2\boldsymbol{S}_1^{-1}$. Under $\mathrm{H}_0(p,q)$, $R(p,q) \xrightarrow{L} \chi^2_{mq}$.

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 $\boldsymbol{\Sigma}$ 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 $\mathrm{H}_0(k,q)\colon \gamma_{j2} = \mathbf{0}$ for all $j \in w,\ w = \{1, 2, \ldots, k\}$, which is equivalent to

$\mathrm{H}_0(k,q)$: The columns of $\boldsymbol{B}$ lie in the subspace generated by the eigenvectors $\gamma_i, (i \notin w)$ of $\boldsymbol{\Sigma}$, (1.14)

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

$$T_n(\boldsymbol{B}) = n \sum_{j \in w} \hat{\gamma}'_{j2} \left(\sum_{i \notin w} \frac{l_i l_j}{(l_i - l_j)^2} \hat{\gamma}_{i2} \hat{\gamma}'_{i2} \right)^{-1} \hat{\gamma}_{j2}, \tag{1.15}$$

where under $\mathrm{H}_0(k, q)$, $T_n(\boldsymbol{B}) \xrightarrow{L} \chi^2_{kq}$.

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 $\mathrm{H}_0(k, q)$ if $r = \mathrm{rank}(\boldsymbol{\Gamma\Gamma}'\boldsymbol{B}) \leq q$, or if $r = q$ and

$$T_{k,q} = n \sum_{j \notin w} l_j^{-1} \,\mathrm{tr}\left\{ \boldsymbol{B}'\hat{\gamma}_j \hat{\gamma}'_j \boldsymbol{B} [\boldsymbol{B}'\boldsymbol{\Gamma} \boldsymbol{D}_j \boldsymbol{\Gamma}'\boldsymbol{B}]^{-1} \right\} \geq \chi^2_{1-\alpha, kq}, \tag{1.16}$$

where $\boldsymbol{B}$ is defined in (1.13) and $\boldsymbol{D}_j$ 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 $T_n(\boldsymbol{B})$ 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

$$\mathrm{E}(T) = a \left\{ 1 + \frac{c}{n} + O(n^{-3/2}) \right\}, \tag{1.17}$$

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

$$T^* = \left(1 - \frac{c}{n}\right) T, \tag{1.18}$$

approaches a. It follows from $T_{k,q} \xrightarrow{L} \chi^2_{kq}$ 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 $\mathbf{\Sigma}$. Let $\mathbf{\Omega}$ 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 $\mathbf{\Sigma} \in \mathbf{\Omega}$, let $\mathbf{\Gamma} \in \mathcal{O}(p)$ such that $\mathbf{\Gamma}'\mathbf{\Sigma}\mathbf{\Gamma} = \mathbf{\Lambda}$, where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$ 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

$$\text{H}_0 \colon \mathbf{\Sigma} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}',$$

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

$$\boldsymbol{\mu} = \arg \min_{\boldsymbol{\nu} \in \boldsymbol{\lambda}(\mathbf{\Omega})} \|\boldsymbol{\nu} - \boldsymbol{\lambda}_m(\boldsymbol{S})\|^2,$$

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

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

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

$$t_k(\boldsymbol{S}) = \sum_{i=1}^{p} \ln(\frac{\mu_i}{l_i}).$$

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 $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$ be the diagonal matrix of the eigenvalues of $\mathbf{\Sigma}$ and $\mathbf{\Gamma}^*$ be a set of the corresponding eigenvectors. The subspace spanned by the columns of a matrix, $\boldsymbol{T}$, is denoted by $\mathcal{R}(\boldsymbol{T})$. Let $\boldsymbol{A}$ be a known $p \times q$ semi-orthogonal matrix with rank q, and $\boldsymbol{M}^*$ be a $p \times m$ semi-orthogonal matrix such that $\mathcal{R}(\boldsymbol{M}^*)$ spans the subspace generated by m specific columns of $\boldsymbol{I}_p$. The hypotheses of interest are

$$\text{H}_0 \quad : \quad \boldsymbol{A} \in \mathcal{R}(\mathbf{\Gamma}^* \boldsymbol{M}^*) \quad \text{for } q \leq m, \text{ and} \tag{1.19}$$

$$\text{H}_0^* \quad : \quad \mathbf{\Gamma}^* \boldsymbol{M}^* \in \mathcal{R}(\boldsymbol{A}) \quad \text{for } q > m. \tag{1.20}$$

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

$$\mathbf{\Gamma}^* \boldsymbol{M}^* \in \mathcal{R}(\boldsymbol{A}) \Longleftrightarrow \boldsymbol{A}^c \in \mathcal{R}(\mathbf{\Gamma}^* \boldsymbol{M}^{*^c}),$$

where $\boldsymbol{A}^c$ and $\boldsymbol{M}^{*^c}$ are orthogonal complement to $\boldsymbol{A}$ and $\boldsymbol{M}^*$, respectively. That is, $\boldsymbol{A}^c \boldsymbol{A}^{c\prime} = \boldsymbol{I}_p - \boldsymbol{A}\boldsymbol{A}'$ and $\boldsymbol{M}^{*^c} \boldsymbol{M}^{*^{c\prime}} = \boldsymbol{I}_p - \boldsymbol{M}^* \boldsymbol{M}^{*\prime}$. 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, $\boldsymbol{M}^*$ can always be equated to the first m columns of $\boldsymbol{I}_p$. This simplification is possible because the columns of $\mathbf{\Gamma}^*$ and the rows of $\boldsymbol{M}^*$ can be permuted to satisfy

$$\mathcal{R}(\boldsymbol{M}^*) = \mathcal{R}\begin{pmatrix} \boldsymbol{I}_m \\ \mathbf{0} \end{pmatrix}.$$

If $\mathcal{R}(\boldsymbol{M}^*) \neq \mathcal{R}\begin{pmatrix} \boldsymbol{I}_m \\ \boldsymbol{0} \end{pmatrix}$, then replace $\boldsymbol{\Gamma}^*\boldsymbol{M}^*$ by $\boldsymbol{\Gamma}\boldsymbol{M}$, where $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^*\boldsymbol{P}$ and $\boldsymbol{M} = \boldsymbol{P}'\boldsymbol{M}^*$, where $\boldsymbol{P}$ is a $p \times p$ permutation matrix. Choose $\boldsymbol{P}$ so that

$$\mathcal{R}(\boldsymbol{P}'\boldsymbol{M}^*) = \mathcal{R}\begin{pmatrix} \boldsymbol{I}_m \\ \boldsymbol{0} \end{pmatrix}.$$

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

$$\text{H}_0\colon \boldsymbol{A} \in \mathcal{R}(\boldsymbol{\Gamma}\boldsymbol{M}), \text{ where } \mathcal{R}(\boldsymbol{M}) = \mathcal{R}\begin{pmatrix} \boldsymbol{I}_m \\ \boldsymbol{0} \end{pmatrix}. \tag{1.21}$$

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 $\boldsymbol{\Gamma}$ be a matrix whose columns are the corresponding eigenvectors. Then, by equating $\boldsymbol{A}$ to $\boldsymbol{A} = (\boldsymbol{I}_q \quad \boldsymbol{0}')'$, 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 $\boldsymbol{y} = \boldsymbol{\Gamma z}$. Partition $\boldsymbol{y}$ as

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix},$$

where $\boldsymbol{y}_1$ is $(p-q) \times 1$ and $\boldsymbol{y}_2$ is $q \times 1$. Partition $\boldsymbol{z}$ as

$$\boldsymbol{z} = \begin{pmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \end{pmatrix},$$

where $\boldsymbol{z}_1$ is $k \times 1$ and $\boldsymbol{z}_2$ 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

$$\boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{\Gamma}_{11} & \boldsymbol{\Gamma}_{12} \\ \boldsymbol{0} & \boldsymbol{\Gamma}_{22} \end{pmatrix}. \qquad (1.22)$$

It follows from (1.2) that

$$\text{var}\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{z} \end{pmatrix} = \text{var}\begin{pmatrix} \boldsymbol{I}_p \\ \boldsymbol{\Gamma}' \end{pmatrix} \boldsymbol{y} = \begin{pmatrix} \boldsymbol{\Gamma\Lambda\Gamma}' & \boldsymbol{\Gamma\Lambda} \\ \boldsymbol{\Lambda\Gamma}' & \boldsymbol{\Lambda} \end{pmatrix},$$

where $\boldsymbol{\Gamma}$ is defined in (1.22). The conditional covariance of $\boldsymbol{y}_2$ given $\boldsymbol{z}_2$ can be written as

$$\text{var}(\boldsymbol{y}_2|\boldsymbol{z}_2) = \text{var}(\boldsymbol{y}_2) - \text{cov}(\boldsymbol{y}_2, \boldsymbol{z}_2)\{\text{var}(\boldsymbol{z}_2)\}^{-1}\text{cov}(\boldsymbol{z}_2, \boldsymbol{y}_2).$$

If the last q variables are redundant, then

$$\text{var}(\boldsymbol{y}_2|\boldsymbol{z}_2) = \boldsymbol{\Gamma}_{22}\boldsymbol{\Lambda}_2\boldsymbol{\Gamma}_{22}' - \boldsymbol{\Gamma}_{22}\boldsymbol{\Lambda}_2\boldsymbol{\Lambda}_2^{-1}\boldsymbol{\Lambda}_2\boldsymbol{\Gamma}_{22}' = \boldsymbol{0}.$$

That is, if $\boldsymbol{y}_2$ is redundant with respect to $\boldsymbol{z}_1$ then $\text{var}(\boldsymbol{y}_2|\boldsymbol{z}_2) = \boldsymbol{0}$.

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 $\boldsymbol{y}$ can be approximately reproduced as linear combination of $\boldsymbol{y}_1$. 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 $\boldsymbol{\Gamma}$ satisfies (1.22), then the conditional covariance of $\boldsymbol{y}_2$ given $\boldsymbol{y}_1$ can be written as

$$\boldsymbol{\Sigma}_{22\cdot 1} = \left(\boldsymbol{\Gamma}_{22}\boldsymbol{\Lambda}_2^{-1}\boldsymbol{\Gamma}'_{22}\right)^{-1},$$

which can have norm as large as $\left\|\left(\boldsymbol{\Gamma}_{22}\boldsymbol{\Lambda}_2^{-1}\boldsymbol{\Gamma}'_{22}\right)^{-1}\right\|^2 = \text{tr}(\boldsymbol{\Lambda}_2^2) = (p-k)\lambda_2^2$ if all eigenvalues in $\boldsymbol{\Lambda}_2$ are equal to λ_2. Conversely, $\|\boldsymbol{\Sigma}_{22\cdot 1}\|^2$ can be small even though redundancy is not satisfied. For example, suppose that $\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_1 + \boldsymbol{\varepsilon} \end{pmatrix}$, where $\text{var}(\boldsymbol{y}_1) = \boldsymbol{\Sigma}_{11}$, $\text{var}(\boldsymbol{\varepsilon}) = a\boldsymbol{\Sigma}_{11}$, and a is a scalar. The covariance matrix of $\boldsymbol{y}$ can be written as

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

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

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

The conditional covariance of $\boldsymbol{y}_2$ given $\boldsymbol{y}_1$ can be written as

$$\boldsymbol{\Sigma}_{22\cdot 1} = (1+a)\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{11} = a\boldsymbol{\Sigma}_{11}.$$

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

$$\boldsymbol{\Gamma} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes \boldsymbol{\Gamma}_{11}$$

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.

Component	Eigenvalue	Cumulative proportion of total variance
1	878.52	0.6092
2	196.10	0.7452
3	128.64	0.8344
4	103.43	0.9062
5	81.26	0.9625
6	37.85	0.9888
7	6.98	0.9936
8	5.71	0.9976
9	3.52	1.0000

Table 2. Eigenvectors of the First Two Principal Components.

Component	Eigenvector								
	High density			Middle tone			Toe portion		
	red	green	blue	red	green	blue	red	green	blue
1	0.305	0.653	0.483	0.261	0.324	0.271	0.002	0.006	0.014
2	-0.486	-0.151	0.588	-0.491	-0.038	0.373	0.057	0.054	0.088

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?

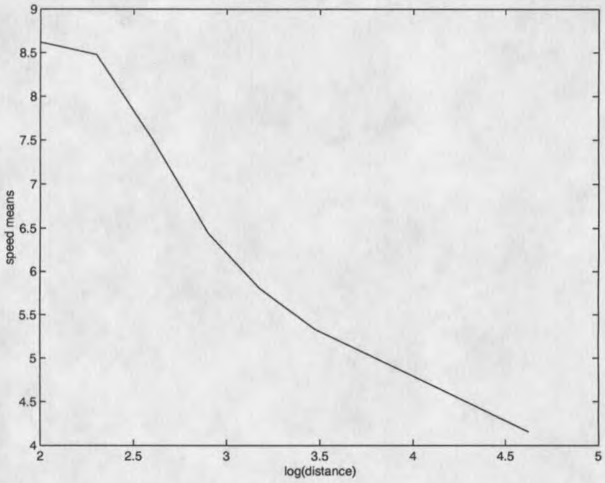
2. How many and which variables can be considered as redundant? Consider the first two principal components. These components account for 75% of the total sample variance. Table 2 reveals that first two components have small coefficients on the last three variables. The question is whether these three 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 $\boldsymbol{C}^*$ 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 $\boldsymbol{C}^*$ contains coefficients of the linear contrast. The second column of $\boldsymbol{C}^*$ contains coefficients of the quadratic contrast etc. The matrix $\boldsymbol{C}^*$ can be computed by using the Matlab program given in Appendix C.

The eigenvalues of $\boldsymbol{C}^{*\prime}\boldsymbol{S}\boldsymbol{C}^*$ 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%. Second, which polynomials contribute to the largest

Table 3. Eigenvalues of Covariance Matrix: Women's Track Example.

Component	Eigenvalue	Cumulative proportion of total variance
1	0.1246	0.5509
2	0.0586	0.8102
3	0.0240	0.9164
4	0.0091	0.9566
5	0.0061	0.9836
6	0.0037	1.0000

Table 4. Eigenvectors of the First Two Principal Components.

Component	Eigenvector					
	linear	quadratic	cubic	quartic	quintic	septic
1	-0.9784	0.0065	0.1738	-0.0764	-0.0783	-0.0229
2	0.0316	0.8985	0.0666	-0.3833	0.2005	0.0059

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 $\mathbf{\Sigma}$ 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:

$$\mathbf{\Sigma} = \mathbf{\Gamma \Lambda \Gamma}', \tag{2.1}$$

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

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

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\varphi} \end{pmatrix}, \quad \begin{pmatrix} \dim(\boldsymbol{\mu}) \\ \dim(\boldsymbol{\varphi}) \end{pmatrix} = \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix},$$

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 $\mathbf{\Sigma}$ by $\boldsymbol{\lambda}$. Then,

$$\begin{aligned}
\mathbf{\Lambda} &= \sum_{j=1}^{p}\sum_{i=1}^{p} \boldsymbol{e}_j^p \boldsymbol{e}_j^{p\prime} \mathbf{\Lambda} \boldsymbol{e}_i^p \boldsymbol{e}_i^{p\prime} \\
&= \sum_{j=1}^{p} \boldsymbol{e}_j^p \boldsymbol{e}_j^{p\prime} \mathbf{\Lambda} \boldsymbol{e}_j^p \boldsymbol{e}_j^{p\prime} \text{ because } \boldsymbol{e}_j^{p\prime} \mathbf{\Lambda} \boldsymbol{e}_i^p = 0 \text{ if } j \neq i, \\
&= \sum_{j=1}^{p} \boldsymbol{e}_j^p \lambda_j \boldsymbol{e}_j^{p\prime},
\end{aligned}$$

where $\boldsymbol{e}_j^p$ is the j^{th} column of $\boldsymbol{I}_p$.

Taking the vec of both sides yields

$$\begin{aligned}
\operatorname{vec}(\mathbf{\Lambda}) &= \sum_{j=1}^{p} (\boldsymbol{e}_j^p \otimes \boldsymbol{e}_j^p) \lambda_j \\
&= \sum_{j=1}^{p} (\boldsymbol{e}_j^p \otimes \boldsymbol{e}_j^p) \boldsymbol{e}_j^{p\prime} \boldsymbol{\lambda}.
\end{aligned}$$

Accordingly, the relationship between $\mathbf{\Lambda}$ and $\boldsymbol{\lambda}$ is

$$\operatorname{vec}(\mathbf{\Lambda}) = \boldsymbol{L}\boldsymbol{\lambda} \quad \text{and} \quad \boldsymbol{\lambda} = \boldsymbol{L}' \operatorname{vec}(\mathbf{\Lambda}), \tag{2.2}$$

where

$$\boldsymbol{L} = \sum_{j=1}^{p} (\boldsymbol{e}_j^p \otimes \boldsymbol{e}_j^p) \boldsymbol{e}_j^{p\prime} \text{ and } \boldsymbol{L}'\boldsymbol{L} = \boldsymbol{I}_p\,. \tag{2.3}$$

Two parameterizations of $\boldsymbol{\lambda}$ are used in two applications of the dimension reduction in principal component analysis. These two structures for $\boldsymbol{\lambda}$ are described in each of the applications below.

Reduction in the Number of Variables

It is assumed that $\boldsymbol{\lambda}$ is a differentiable function of $\boldsymbol{\varphi}$, where $\boldsymbol{\varphi}$ 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:

$$\boldsymbol{\lambda} = \boldsymbol{T}_1 \exp\{\odot(\boldsymbol{T}_2\boldsymbol{\varphi})\}, \tag{2.4}$$

where $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ 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 $\boldsymbol{u}$ is a $p \times 1$ vector, then $\exp(\odot\boldsymbol{u}) = (e^{u_1} \;\; e^{u_2} \cdots e^{u_p})'$. The following examples show how to construct $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$.

Suppose that $\boldsymbol{\lambda}$ contains ν_2 distinct values. Let $\boldsymbol{m} = (m_1 \, m_2 \;\cdots\; m_{\nu_2})'$ be the vector of eigenvalue multiplicities, where $\sum_{i=1}^{\nu_2} m_i = p$. For example, if $\boldsymbol{\lambda} = (10\ 10\ 5\ 1\ 1)'$ then $\boldsymbol{m} = (2\ 1\ 2)'$. The parameterization in (2.4) allows arbitrary eigenvalue multiplicities. Consider the $p \times 1$ vector of $\boldsymbol{\lambda}$ with multiplicity $\boldsymbol{m}$ and

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & & & \mathbf{0} \\ & \lambda_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \lambda_p \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_{m_1}\exp(\varphi_1) & & & \mathbf{0} \\ & \boldsymbol{I}_{m_2}\exp(\varphi_2) & & \\ & & \ddots & \\ \mathbf{0} & & & \boldsymbol{I}_{m_{\nu_2}}\exp(\varphi_{\nu_2}) \end{pmatrix}.$$

If the distinct eigenvalues are unordered, then $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ can be written as

$$\boldsymbol{T}_1 = \bigoplus_{i=1}^{\nu_2} \mathbf{1}_{m_i} \quad \text{and} \quad \boldsymbol{T}_2 = \boldsymbol{I}_{\nu_2}.$$

To order all the eigenvalues from small to large when $\boldsymbol{m} = \mathbf{1}_p$ and $p = 4 = \nu_2$; $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ can be written as

$$\boldsymbol{T}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \text{ and } \quad \boldsymbol{T}_2 = \boldsymbol{I}_{\nu_2}. \tag{2.5}$$

To order all the eigenvalues from small to large when $\boldsymbol{m} = (1\ 1\ 2)'$, $p = 4$, and $\nu_2 = 3$; $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ can be written as

$$\boldsymbol{T}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \text{ and } \quad \boldsymbol{T}_2 = \boldsymbol{I}_{\nu_2}.$$

To ensure that the first eigenvalue is the largest when $\boldsymbol{m} = \mathbf{1}_p$ and $p = 4 = \nu_2$; set $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ as

$$\boldsymbol{T}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix} \text{ and } \quad \boldsymbol{T}_2 = \boldsymbol{I}_{\nu_2}.$$

To order all the eigenvalues such that $\lambda_i = \exp\{\alpha + \beta(p - i)\}$, $i = 1, 2, \ldots, p$, where α and β are constants; $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ can be constructed as

$$\boldsymbol{T}_1 = \boldsymbol{I}_p \text{ and } \quad \boldsymbol{T}_2 = (\mathbf{1}_p \quad \boldsymbol{p}),$$

where $\boldsymbol{p}$ 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; $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ can be constructed as follows:

$$\boldsymbol{T}_1 = \begin{pmatrix} \mathbf{1}_7 & \mathbf{0}_7 & -\boldsymbol{I}_7 \\ 1 & 0 & \mathbf{0}_7' \\ 1 & 1 & \mathbf{0}_7' \end{pmatrix} \text{ and } \quad \boldsymbol{T}_2 = \boldsymbol{I}_9.$$

Expressions for the first, second, and third derivatives of $\boldsymbol{\lambda}$ in (2.4) with respect to $\boldsymbol{\varphi}'$ 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* $\boldsymbol{\lambda}$ *in (2.4) with respect to* $\boldsymbol{\varphi}'$ *can be written as follows:*

$$
\begin{aligned}
\boldsymbol{D}_{\lambda}^{(1)} &= \frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\varphi}'} = \boldsymbol{T}_1 \operatorname{diag}[\exp\{\odot(\boldsymbol{T}_2\boldsymbol{\varphi})\}]\boldsymbol{T}_2, \\
\boldsymbol{D}_{\lambda}^{(2)} &= \frac{\partial^2 \boldsymbol{\lambda}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'} = \{\exp(\odot \boldsymbol{T}_2\boldsymbol{\varphi}) \otimes \boldsymbol{T}_1'\}' \boldsymbol{W}^{(2)}(\boldsymbol{T}_2 \otimes \boldsymbol{T}_2), \\
\boldsymbol{D}_{\lambda}^{(3)} &= \frac{\partial^3 \boldsymbol{\lambda}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'} = \{\exp(\odot \boldsymbol{T}_2\boldsymbol{\varphi}) \otimes \boldsymbol{T}_1'\}' \boldsymbol{W}^{(3)}(\boldsymbol{T}_2 \otimes \boldsymbol{T}_2 \otimes \boldsymbol{T}_2),
\end{aligned}
$$

where

$$
\boldsymbol{W}^{(2)} = \sum_{i=1}^{q_2} (\boldsymbol{e}_i^{q_2}\boldsymbol{e}_i^{q_2\prime} \otimes \boldsymbol{e}_i^{q_2}\boldsymbol{e}_i^{q_2\prime}) \quad \text{and} \quad \boldsymbol{W}^{(3)} = \sum_{i=1}^{q_2} (\boldsymbol{e}_i^{q_2}\boldsymbol{e}_i^{q_2\prime} \otimes \boldsymbol{e}_i^{q_2}\boldsymbol{e}_i^{q_2\prime} \otimes \boldsymbol{e}_i^{q_2\prime}).
$$

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

$$
\boldsymbol{\lambda} = \boldsymbol{T}_1 \sum_{i=1}^{q_2} \boldsymbol{e}_i^{q_2} \exp\{\boldsymbol{e}_i^{q_2\prime}\boldsymbol{T}_2\boldsymbol{\varphi}\}.
$$

Taking the derivative of $\boldsymbol{\lambda}$ *with respect to* $\boldsymbol{\varphi}'$ *yields*

$$
\begin{aligned}
\boldsymbol{D}_{\lambda}^{(1)} &= \boldsymbol{T}_1 \sum_{i=1}^{q_2} \boldsymbol{e}_i^{q_2} \exp\{\boldsymbol{e}_i^{q_2\prime}\boldsymbol{T}_2\boldsymbol{\varphi}\}\boldsymbol{e}_i^{q_2\prime}\boldsymbol{T}_2 \qquad (2.6) \\
&= \boldsymbol{T}_1 \operatorname{diag}[\exp\{\odot(\boldsymbol{T}_2\boldsymbol{\varphi})\}]\boldsymbol{T}_2,
\end{aligned}
$$

where

$$
\operatorname{diag}[\exp\{\odot(\boldsymbol{T}_2\boldsymbol{\varphi})\}] = \sum_{i=1}^{q_2} \boldsymbol{e}_i^{q_2} \exp\{\boldsymbol{e}_i^{q_2\prime}\boldsymbol{T}_2\boldsymbol{\varphi}\}\boldsymbol{e}_i^{q_2\prime}.
$$

Taking the derivative of (2.6) yields

$$\begin{aligned}\boldsymbol{D}_\lambda^{(2)} &= \boldsymbol{T}_1 \sum_{i=1}^{q_2} \boldsymbol{e}_i^{q_2} \frac{\partial}{\partial \boldsymbol{\varphi}'} \left[\exp\{\boldsymbol{e}_i^{q_2\prime}\boldsymbol{T}_2\boldsymbol{\varphi}\}\boldsymbol{e}_i^{q_2\prime}\boldsymbol{T}_2\right] \\ &= \boldsymbol{T}_1 \sum_{i=1}^{q_2} \boldsymbol{e}_i^{q_2} \big\{ \exp\{\boldsymbol{e}_i^{q_2\prime}\boldsymbol{T}_2\boldsymbol{\varphi}\}\boldsymbol{e}_i^{q_2\prime}\boldsymbol{T}_2(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{e}_i^{q_2\prime}\boldsymbol{T}_2)\big\}.\end{aligned}$$

Expressing $\exp\{\boldsymbol{e}_i^{q_2\prime}\boldsymbol{T}_2\boldsymbol{\varphi}\}$ *as* $\boldsymbol{e}_i^{q_2\prime}\exp\{\odot(\boldsymbol{T}_2\boldsymbol{\varphi})\}$ *yields*

$$\begin{aligned}\boldsymbol{D}_\lambda^{(2)} &= \sum_{i=1}^{q_2} \boldsymbol{T}_1\boldsymbol{e}_i^{q_2}\boldsymbol{e}_i^{q_2\prime}\exp\{\odot(\boldsymbol{T}_2\boldsymbol{\varphi})\}(\boldsymbol{e}_i^{q_2\prime}\otimes\boldsymbol{e}_i^{q_2\prime})(\boldsymbol{T}_2\otimes\boldsymbol{T}_2) \\ &= \sum_{i=1}^{q_2} (\exp\{\odot(\boldsymbol{T}_2\boldsymbol{\varphi})\}'\boldsymbol{e}_i^{q_2}\otimes\boldsymbol{T}_1)\,\boldsymbol{e}_i^{q_2}(\boldsymbol{e}_i^{q_2\prime}\otimes\boldsymbol{e}_i^{q_2\prime})(\boldsymbol{T}_2\otimes\boldsymbol{T}_2) \\ &= \sum_{i=1}^{q_2} (\exp\{\odot(\boldsymbol{T}_2\boldsymbol{\varphi})\}'\otimes\boldsymbol{T}_1)\,(\boldsymbol{e}_i^{q_2}\otimes\boldsymbol{e}_i^{q_2})\,(\boldsymbol{e}_i^{q_2\prime}\otimes\boldsymbol{e}_i^{q_2\prime})(\boldsymbol{T}_2\otimes\boldsymbol{T}_2),\end{aligned}$$

where $(\boldsymbol{e}_i^{q_2}\otimes\boldsymbol{I}_{q_2})\boldsymbol{e}_i^{q_2} = (\boldsymbol{e}_i^{q_2}\otimes\boldsymbol{e}_i^{q_2})$.

Simplification yields

$$\begin{aligned}\boldsymbol{D}_\lambda^{(2)} &= (\exp\{\odot(\boldsymbol{T}_2\boldsymbol{\varphi})\}\otimes\boldsymbol{T}_1')'\sum_{i=1}^{q_2}(\boldsymbol{e}_i^{q_2}\boldsymbol{e}_i^{q_2\prime}\otimes\boldsymbol{e}_i^{q_2}\boldsymbol{e}_i^{q_2\prime})(\boldsymbol{T}_2\otimes\boldsymbol{T}_2) \\ &= (\exp\{\odot(\boldsymbol{T}_2\boldsymbol{\varphi})\}\otimes\boldsymbol{T}_1')'\boldsymbol{W}^{(2)}(\boldsymbol{T}_2\otimes\boldsymbol{T}_2).\end{aligned}$$

Similarly, the third derivative can be obtained as follows:

$$\begin{aligned}\boldsymbol{D}_\lambda^{(3)} &= \boldsymbol{T}_1 \sum_{i=1}^{q_2} \boldsymbol{e}_i^{q_2}\frac{\partial}{\partial\boldsymbol{\varphi}'}\big(\exp\{\boldsymbol{e}_i^{q_2\prime}\boldsymbol{T}_2\boldsymbol{\varphi}\}(\boldsymbol{e}_i^{q_2\prime}\boldsymbol{T}_2\otimes\boldsymbol{e}_i^{q_2\prime}\boldsymbol{T}_2)\big) \\ &= \boldsymbol{T}_1 \sum_{i=1}^{q_2} \boldsymbol{e}_i^{q_2}\big\{\exp\{\boldsymbol{e}_i^{q_2\prime}\boldsymbol{T}_2\boldsymbol{\varphi}\}\boldsymbol{e}_i^{q_2\prime}\boldsymbol{T}_2(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{e}_i^{q_2\prime}\boldsymbol{T}_2\otimes\boldsymbol{e}_i^{q_2\prime}\boldsymbol{T}_2)\big\} \\ &= \sum_{i=1}^{q_2} \boldsymbol{T}_1\boldsymbol{e}_i^{q_2}\boldsymbol{e}_i^{q_2\prime}\exp\{\odot(\boldsymbol{T}_2\boldsymbol{\varphi})\}(\boldsymbol{e}_i^{q_2\prime}\otimes\boldsymbol{e}_i^{q_2\prime}\otimes\boldsymbol{e}_i^{q_2\prime})(\boldsymbol{T}_2\otimes\boldsymbol{T}_2\otimes\boldsymbol{T}_2) \\ &= \sum_{i=1}^{q_2} (\exp\{\odot(\boldsymbol{T}_2\boldsymbol{\varphi})\}'\boldsymbol{e}_i^{q_2}\otimes\boldsymbol{T}_1)\,\boldsymbol{e}_i^{q_2}(\boldsymbol{e}_i^{q_2\prime}\otimes\boldsymbol{e}_i^{q_2\prime}\otimes\boldsymbol{e}_i^{q_2\prime})(\boldsymbol{T}_2\otimes\boldsymbol{T}_2\otimes\boldsymbol{T}_2) \\ &= \sum_{i=1}^{q_2} (\exp\{\odot(\boldsymbol{T}_2\boldsymbol{\varphi})\}'\otimes\boldsymbol{T}_1)\,(\boldsymbol{e}_i^{q_2}\otimes\boldsymbol{e}_i^{q_2})(\boldsymbol{e}_i^{q_2\prime}\otimes\boldsymbol{e}_i^{q_2\prime}\otimes\boldsymbol{e}_i^{q_2\prime})(\boldsymbol{T}_2\otimes\boldsymbol{T}_2\otimes\boldsymbol{T}_2).\end{aligned}$$

Finally,

$$\begin{aligned}\boldsymbol{D}_\lambda^{(3)} &= (\exp\{\odot(\boldsymbol{T}_2\boldsymbol{\varphi})\}\otimes\boldsymbol{T}_1')'\sum_{i=1}^{q_2}(\boldsymbol{e}_i^{q_2}\boldsymbol{e}_i^{q_2\prime}\otimes\boldsymbol{e}_i^{q_2}\boldsymbol{e}_i^{q_2\prime}\otimes\boldsymbol{e}_i^{q_2\prime})(\boldsymbol{T}_2\otimes\boldsymbol{T}_2\otimes\boldsymbol{T}_2)\\ &= (\exp\{\odot(\boldsymbol{T}_2\boldsymbol{\varphi})\}\otimes\boldsymbol{T}_1')'\boldsymbol{W}^{(3)}(\boldsymbol{T}_2\otimes\boldsymbol{T}_2\otimes\boldsymbol{T}_2).\end{aligned}$$

□

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:

$$\mathbf{1}_p'\boldsymbol{\lambda} = \varphi_0, \text{ where } \quad \varphi_0 = \text{tr}(\boldsymbol{\Sigma}), \text{ and} \tag{2.7}$$

$$\frac{\boldsymbol{C}_1'\boldsymbol{\lambda}}{\varphi_0} = \boldsymbol{c}_0, \tag{2.8}$$

where $\boldsymbol{C}_1$ is a matrix of known constants and $\boldsymbol{c}_0$ is a vector of known constants. The following parameterization is employed:

$$\boldsymbol{\lambda} = \frac{\varphi_0\boldsymbol{T}_3\exp\{\odot(\boldsymbol{T}_4\boldsymbol{\xi})\}}{\mathbf{1}_p'\boldsymbol{T}_3\exp\{\odot(\boldsymbol{T}_4\boldsymbol{\xi})\}}, \tag{2.9}$$

where $\boldsymbol{\xi} = \boldsymbol{\xi}(\boldsymbol{\varphi})$ is a differentiable function of $\boldsymbol{\varphi}$, $\boldsymbol{T}_3$ and $\boldsymbol{T}_4$ are full column-rank design matrices of known constants with dimensions $p \times q_3$ and $q_3 \times q_4$, respectively, $\dim(\boldsymbol{\varphi}) = \nu_2$, and $\nu_2 \leq q_4$. It is assumed that (2.8) is consistent. That is, $\exists\ \boldsymbol{\xi}$ such that $\frac{\boldsymbol{C}_1'\boldsymbol{\lambda}}{\varphi_0} = \boldsymbol{c}_0$.

The vector of eigenvalues in (2.9) depends on $\mathcal{R}(\boldsymbol{T}_4)$, the vector space generated by the columns of $\boldsymbol{T}_4$. If $\mathbf{1}_{q_3} \in \mathcal{R}(\boldsymbol{T}_4)$, then $\boldsymbol{T}_4$ in (2.9) can be reduced to a matrix that has one fewer column than the original matrix in the following manner. Let $\boldsymbol{P}_1 = \text{ppo}(\mathbf{1}_{q_3}) = \mathbf{1}_{q_3} q_3^{-1} \mathbf{1}'_{q_3}$ and write $\boldsymbol{T}_4\boldsymbol{\xi}$ as

$$\begin{aligned} \boldsymbol{T}_4\boldsymbol{\xi} &= (\boldsymbol{I}_{q_3} - \boldsymbol{P}_1)\boldsymbol{T}_4\boldsymbol{\xi} + \boldsymbol{P}_1\boldsymbol{T}_4\boldsymbol{\xi} \\ &= (\boldsymbol{I}_{q_3} - \boldsymbol{P}_1)\boldsymbol{T}_4\boldsymbol{\xi} + \mathbf{1}_{q_3} q_3^{-1} \mathbf{1}'_{q_3}\boldsymbol{T}_4\boldsymbol{\xi}. \end{aligned}$$

Substituting in (2.9) yields

$$\boldsymbol{\lambda} = \frac{\varphi_0 \boldsymbol{T}_3 \exp\{\odot((\boldsymbol{I}_{q_3} - \boldsymbol{P}_1)\boldsymbol{T}_4\boldsymbol{\xi})\}}{\mathbf{1}'_p \boldsymbol{T}_3 \exp\{\odot(\boldsymbol{I}_{q_3} - \boldsymbol{P}_1)\boldsymbol{T}_4\boldsymbol{\xi})\}}$$

because $\exp(q_3^{-1}\mathbf{1}'_{q_3}\boldsymbol{T}_4\boldsymbol{\xi})$ is a scalar and can be canceled out. If $\mathbf{1}_{q_3} \in \mathcal{R}(\boldsymbol{T}_4)$, then $\mathbf{1}_{q_3} = \boldsymbol{T}_4\boldsymbol{b}$ for some $\boldsymbol{b}$. Therefore, if $\mathbf{1}_{q_3} \in \mathcal{R}(\boldsymbol{T}_4)$, then

$$\begin{aligned} \text{rank}[(\boldsymbol{I}_{q_3} - \boldsymbol{P}_1)\boldsymbol{T}_4] &= \text{rank}\left(\boldsymbol{T}_4 - \mathbf{1}_{q_3} q_3^{-1} \mathbf{1}'_{q_3}\boldsymbol{T}_4\right) \\ &= \text{rank}\left(\boldsymbol{T}_4 - \boldsymbol{T}_4\boldsymbol{b} q_3^{-1} \mathbf{1}'_{q_3}\boldsymbol{T}_4\right) \\ &= \text{rank}[\boldsymbol{T}_4\left(\boldsymbol{I}_{q_4} - \boldsymbol{b} q_3^{-1} \mathbf{1}'_{q_3}\boldsymbol{T}_4\right)] \\ &= \text{rank}\left(\boldsymbol{I}_{q_4} - \boldsymbol{b} q_3^{-1} \mathbf{1}'_{q_3}\boldsymbol{T}_4\right) \quad \text{because } \boldsymbol{T}_4 \text{ has full column}-\text{rank} \\ &= \text{tr}(\boldsymbol{I}_{q_4}) - \text{tr}\left(\mathbf{1}_{q_3} q_3^{-1} \mathbf{1}'_{q_3}\right) \quad \text{because } \boldsymbol{b} q_3^{-1} \mathbf{1}'_{q_3}\boldsymbol{T}_4 \text{ is idempotent} \\ &= \text{rank}(\boldsymbol{T}_4) - 1. \end{aligned}$$

Factor $(\boldsymbol{I}_{q_3} - \boldsymbol{P}_1)\boldsymbol{T}_4$ as

$$(\boldsymbol{I}_{q_3} - \boldsymbol{P}_1)\boldsymbol{T}_4 = \boldsymbol{U}_p\boldsymbol{V}'_p,$$

where $\boldsymbol{U}_p$ and $\boldsymbol{V}_p$ each have full column-rank. Then

$$(\boldsymbol{I}_{q_3} - \boldsymbol{P}_1)\boldsymbol{T}_4\boldsymbol{\xi} = \boldsymbol{U}_p\boldsymbol{\xi}^*, \text{ where } \boldsymbol{\xi}^* = \boldsymbol{V}'_p\boldsymbol{\xi}.$$

Accordingly, $\boldsymbol{T}_4$ should be replaced by $\boldsymbol{U}_p$, a matrix whose columns form a basis set for $\mathcal{R}\left[\left(\boldsymbol{I}_{q_3} - \boldsymbol{1}_{q_3} q_3^{-1} \boldsymbol{1}'_{q_3}\right) \boldsymbol{T}_4\right]$. For convenience, the remainder of the thesis still uses symbols $\boldsymbol{T}_4$ and $\boldsymbol{\xi}$ instead of $\boldsymbol{U}_p$ and $\boldsymbol{\xi}^*$. For the example in (2.5), $\boldsymbol{T}_4 = \boldsymbol{I}_{\nu_2}$ should be replaced by

$$\boldsymbol{T}_4 = \begin{pmatrix} \boldsymbol{I}_{\nu_2 - 1} \\ -\boldsymbol{1}' \end{pmatrix},$$

or any other full column-rank matrix that has column space equal to $\mathcal{R}\begin{pmatrix} \boldsymbol{I}_{\nu_2 - 1} \\ -\boldsymbol{1}' \end{pmatrix}$. Consider the second constraint. It follows from (2.8) that

$$\frac{\boldsymbol{C}'_1 \boldsymbol{\lambda}}{\varphi_0} = \boldsymbol{c}_0 \iff \boldsymbol{C}'_2 \exp\{\odot(\boldsymbol{T}_4 \boldsymbol{\xi})\} = \boldsymbol{0}, \tag{2.10}$$

where $\boldsymbol{C}_2$ is a $q_3 \times r$ matrix with rank r whose columns are a basis set for $\mathcal{R}\left[\boldsymbol{T}'_3(\boldsymbol{C}_1 - \boldsymbol{1}_p \boldsymbol{c}'_0)\right]$. To show the equivalence in equation (2.10), substitute $\boldsymbol{\lambda}$ from (2.9) to obtain

$$\frac{\boldsymbol{C}'_1 \boldsymbol{T}_3 \exp\{\odot(\boldsymbol{T}_4 \boldsymbol{\xi})\}}{\boldsymbol{1}'_p \boldsymbol{T}_3 \exp\{\odot(\boldsymbol{T}_4 \boldsymbol{\xi})\}} = \boldsymbol{c}_0$$

$$\iff \boldsymbol{C}'_1 \boldsymbol{T}_3 \exp\{\odot(\boldsymbol{T}_4 \boldsymbol{\xi})\} = \boldsymbol{c}_0 \boldsymbol{1}'_p \boldsymbol{T}_3 \exp\{\odot(\boldsymbol{T}_4 \boldsymbol{\xi})\}$$

$$\iff \left(\boldsymbol{C}'_1 - \boldsymbol{c}_0 \boldsymbol{1}'_p\right) \boldsymbol{T}_3 \exp\{\odot(\boldsymbol{T}_4 \boldsymbol{\xi})\} = \boldsymbol{0}$$

$$\iff \boldsymbol{C}'_2 \exp\{\odot(\boldsymbol{T}_4 \boldsymbol{\xi})\} = \boldsymbol{0}.$$

Let $\boldsymbol{V} = (\boldsymbol{V}_1 \quad \boldsymbol{V}_2)$ be a $q_4 \times q_4$ nonsingular matrix of constants, where $\boldsymbol{V}_1$ has dimension $q_4 \times r$, $\boldsymbol{V}_2$ has dimension $q_4 \times \nu_2$, $\nu_2 = q_4 - r$, and $\boldsymbol{V}'_1 \boldsymbol{V}_2 = \boldsymbol{0}$. Then, $\boldsymbol{T}_4 \boldsymbol{\xi}$ can be written as

$$\begin{aligned} \boldsymbol{T}_4 \boldsymbol{\xi} &= \boldsymbol{T}_4 \boldsymbol{V} \boldsymbol{V}^{-1} \boldsymbol{\xi} \\ &= \boldsymbol{T}_4 \left[\boldsymbol{V}_1 (\boldsymbol{V}'_1 \boldsymbol{V}_1)^{-1} \boldsymbol{V}'_1 \boldsymbol{\xi} + \boldsymbol{V}_2 (\boldsymbol{V}'_2 \boldsymbol{V}_2)^{-1} \boldsymbol{V}'_2 \boldsymbol{\xi}\right] \end{aligned}$$

$$= \boldsymbol{T}_4(\boldsymbol{V}_1\boldsymbol{\tau} + \boldsymbol{V}_2\boldsymbol{\varphi}), \tag{2.11}$$

where $\boldsymbol{\tau} = (\boldsymbol{V}_1'\boldsymbol{V}_1)^{-1}\boldsymbol{V}_1'\boldsymbol{\xi}$, $\boldsymbol{\varphi} = (\boldsymbol{V}_2'\boldsymbol{V}_2)^{-1}\boldsymbol{V}_2'\boldsymbol{\xi}$, and $\boldsymbol{\tau}$ is an implicit function of $\boldsymbol{\varphi}$. Substituting $\boldsymbol{T}_4\boldsymbol{\xi}$ from (2.11) in the constraint (2.10) yields

$$\boldsymbol{C}_2' \exp\{\odot(\boldsymbol{T}_4(\boldsymbol{V}_1\boldsymbol{\tau} + \boldsymbol{V}_2\boldsymbol{\varphi}))\} = \boldsymbol{0}.$$

Taking the derivative of the above constraint with respect to $\boldsymbol{\tau}'$ yields

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\tau}'}\{\boldsymbol{C}_2' \exp\{\odot(\boldsymbol{T}_4(\boldsymbol{V}_1\boldsymbol{\tau} + \boldsymbol{V}_2\boldsymbol{\varphi}))\}\} &= \boldsymbol{C}_2' \operatorname{diag}(\exp\{\odot(\boldsymbol{T}_4\boldsymbol{\xi})\})\boldsymbol{T}_4\boldsymbol{V}_1 \\ &= \boldsymbol{C}_2'\boldsymbol{W}_1\boldsymbol{V}_1, \end{aligned} \tag{2.12}$$

where

$$\boldsymbol{W}_1 = \operatorname{diag}(\exp\{\odot(\boldsymbol{T}_4\boldsymbol{\xi})\})\boldsymbol{T}_4. \tag{2.13}$$

By the implicit function theorem (Taylor and Mann, 1983), if the matrix of derivatives in (2.12) is nonsingular, then $\boldsymbol{\tau}$ is an implicit function of $\boldsymbol{\varphi}$ implying that $\frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\varphi}'}$ exists. Theorem 2 chooses a matrix $\boldsymbol{V}$ such that $\boldsymbol{C}_2'\boldsymbol{W}_1\boldsymbol{V}_1$ in (2.12) is nonsingular. In the theorem, it is assumed that $\boldsymbol{C}_2'\boldsymbol{W}_1$ has full row-rank when evaluated in an open neighborhood of $\boldsymbol{\xi}^*$ and $\boldsymbol{\xi}^*$ satisfies $\boldsymbol{C}_2' \exp\{\odot(\boldsymbol{T}_4\boldsymbol{\xi}^*)\} = \boldsymbol{0}$.

Theorem 2. *Assuming that $\boldsymbol{C}_2'\boldsymbol{W}_1$ has full row-rank when evaluated in an open neighborhood of $\boldsymbol{\xi}^*$ and $\boldsymbol{\xi}^*$ satisfies $\boldsymbol{C}_2' \exp\{\odot(\boldsymbol{T}_4\boldsymbol{\xi}^*)\} = \boldsymbol{0}$. Write the singular value decomposition of $\boldsymbol{C}_2'\boldsymbol{W}_1$ as follows:*

$$\boldsymbol{C}_2'\boldsymbol{W}_1 = \boldsymbol{U}_\xi(\boldsymbol{D}_\xi \quad \boldsymbol{0})\begin{pmatrix}\boldsymbol{V}_1' \\ \boldsymbol{V}_2'\end{pmatrix} = \boldsymbol{U}_\xi\boldsymbol{D}_\xi\boldsymbol{V}_1', \tag{2.14}$$

where $\boldsymbol{D}_\xi$ is an $r \times r$ diagonal matrix of singular values, $\boldsymbol{U}_\xi \in \mathcal{O}(r)$, $\boldsymbol{V} = (\boldsymbol{V}_1 \quad \boldsymbol{V}_2) \in \mathcal{O}(q_4)$, and $\boldsymbol{W}_1$ is defined in (2.13). This choice of $\boldsymbol{V}$ ensures that the derivative matrix in (2.12) is nonsingular.

Proof. *Assuming that $\boldsymbol{C}_2'\boldsymbol{W}_1$ has full row-rank. It follows from (2.14) that*

$$\begin{aligned} \boldsymbol{C}_2'\boldsymbol{W}_1\boldsymbol{V}_1 &= \boldsymbol{U}_\xi\boldsymbol{D}_\xi\boldsymbol{V}_1'\boldsymbol{V}_1 = \boldsymbol{U}_\xi\boldsymbol{D}_\xi \\ \Longrightarrow (\boldsymbol{C}_2'\boldsymbol{W}_1\boldsymbol{V}_1)^{-1} &= \boldsymbol{D}_\xi^{-1}\boldsymbol{U}_\xi'. \end{aligned}$$

Therefore, $\boldsymbol{C}_2'\boldsymbol{W}_1\boldsymbol{V}_1$ is nonsingular for this choice of $\boldsymbol{V}$. □

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

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\varphi}'}\boldsymbol{C}_2' \exp\{\odot \boldsymbol{T}_4[\boldsymbol{V}_1\boldsymbol{\tau} + \boldsymbol{V}_2\boldsymbol{\varphi}]\} &= \boldsymbol{0} \\ \Longrightarrow \boldsymbol{C}_2'\boldsymbol{W}_1(\boldsymbol{V}_1\frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\varphi}'} + \boldsymbol{V}_2) &= \boldsymbol{0} \\ \Longrightarrow \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\varphi}'} = -(\boldsymbol{C}_2'\boldsymbol{W}_1\boldsymbol{V}_1)^{-1}\boldsymbol{C}_2'\boldsymbol{W}_1\boldsymbol{V}_2 &= \boldsymbol{0}, \end{aligned} \tag{2.15}$$

because $\boldsymbol{C}_2'\boldsymbol{W}_1 = \boldsymbol{U}_\xi\boldsymbol{D}_\xi\boldsymbol{V}_1'$ and $\boldsymbol{V}_1'\boldsymbol{V}_2 = \boldsymbol{0}$.

For fixed $\boldsymbol{V}$ from Theorem 2, the first three derivatives of $\boldsymbol{\lambda}$ in (2.9) with respect to $\boldsymbol{\varphi}'$ 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 $\boldsymbol{V}$ from Theorem 2, the first three derivatives of $\boldsymbol{\lambda}$ in (2.9) with respect to $\boldsymbol{\varphi}'$ can be written as follows:*

$$
\begin{aligned}
\boldsymbol{D}_{\lambda}^{(1)} &= \frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\varphi}'} = w\varphi_0 \boldsymbol{H}_{\boldsymbol{\lambda}} \boldsymbol{T}_3 \boldsymbol{W}_1 \boldsymbol{V}_2, \\
\boldsymbol{D}_{\lambda}^{(2)} &= \frac{\partial^2 \boldsymbol{\lambda}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'} = -2w \left(\boldsymbol{D}_{\lambda}^{(1)} \otimes \boldsymbol{1}_p' \boldsymbol{T}_3 \boldsymbol{W}_1 \boldsymbol{V}_2 \right) \boldsymbol{N}_{\nu_2} + w\varphi_0 \boldsymbol{H}_{\boldsymbol{\lambda}} \boldsymbol{T}_3 \boldsymbol{H}_{\boldsymbol{\xi}} \boldsymbol{W}_2 (\boldsymbol{V}_2 \otimes \boldsymbol{V}_2), \\
\boldsymbol{D}_{\lambda}^{(3)} &= \frac{\partial^3 \boldsymbol{\lambda}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'} = -2w\varphi_0 \boldsymbol{H}_{\boldsymbol{\lambda}} \boldsymbol{T}_3 \boldsymbol{H}_{\boldsymbol{\xi}} \boldsymbol{W}_2 \left[\boldsymbol{V}_1 \boldsymbol{D}_{\xi}^{-1} \boldsymbol{U}_{\xi}' \boldsymbol{C}_2' \boldsymbol{W}_2 (\boldsymbol{V}_2 \otimes \boldsymbol{V}_2) \otimes \boldsymbol{V}_2 \right] \boldsymbol{N}_{\nu_2}^* \\
&\quad -2w \left\{ \left(\boldsymbol{D}_{\lambda}^{(2)} \otimes \boldsymbol{1}_p' \boldsymbol{T}_3 \boldsymbol{W}_1 \boldsymbol{V}_2 \right) + \left[\boldsymbol{1}_p' \boldsymbol{T}_3 \boldsymbol{H}_{\boldsymbol{\xi}} \boldsymbol{W}_2 (\boldsymbol{V}_2 \otimes \boldsymbol{V}_2) \otimes \boldsymbol{D}_{\lambda}^{(1)} \right] \right\} \boldsymbol{N}_{\nu_2}^* \\
&\quad + w\varphi_0 \boldsymbol{H}_{\boldsymbol{\lambda}} \boldsymbol{T}_3 \boldsymbol{H}_{\boldsymbol{\xi}} \boldsymbol{W}_3 (\boldsymbol{V}_2 \otimes \boldsymbol{V}_2 \otimes \boldsymbol{V}_2), \quad \textit{where}
\end{aligned}
$$

$$
\begin{aligned}
w &= \left(\boldsymbol{1}_p' \boldsymbol{T}_3 \exp\{\odot(\boldsymbol{T}_4 \xi)\} \right)^{-1}, \qquad \boldsymbol{H}_{\boldsymbol{\lambda}} = \boldsymbol{I}_p - \boldsymbol{\lambda} {\varphi_0}^{-1} \boldsymbol{1}_p', \\
\boldsymbol{H}_{\boldsymbol{\xi}} &= \boldsymbol{I}_{q_3} - \boldsymbol{W}_1 \boldsymbol{V}_1 \boldsymbol{D}_{\xi}^{-1} \boldsymbol{U}_{\xi}' \boldsymbol{C}_2', \qquad \boldsymbol{N}_{\nu_2} = \frac{1}{2} (\boldsymbol{I}_{\nu_2^2} + \boldsymbol{I}_{(\nu_2,\nu_2)}), \\
\boldsymbol{N}_{\nu_2}^* &= \frac{1}{2} \boldsymbol{I}_{(\nu_2,\nu_2^2)} + (\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{N}_{\nu_2}), \\
\boldsymbol{W}_2 &= \sum_{i=1}^{q_3} \boldsymbol{e}_i^{q_3} \exp\{\boldsymbol{e}_i^{q_3\prime} \boldsymbol{T}_4 \boldsymbol{\xi}\} (\boldsymbol{e}_i^{q_3\prime} \boldsymbol{T}_4 \otimes \boldsymbol{e}_i^{q_3\prime} \boldsymbol{T}_4), \\
\boldsymbol{W}_3 &= \sum_{i=1}^{q_3} \boldsymbol{e}_i^{q_3} \exp\{\boldsymbol{e}_i^{q_3\prime} \boldsymbol{T}_4 \boldsymbol{\xi}\} (\boldsymbol{e}_i^{q_3\prime} \boldsymbol{T}_4 \otimes \boldsymbol{e}_i^{q_3\prime} \boldsymbol{T}_4 \otimes \boldsymbol{e}_i^{q_3\prime} \boldsymbol{T}_4),
\end{aligned}
$$

$\boldsymbol{e}_i^{q_3}$ is the i^{th} column of $\boldsymbol{I}_{q_3}$, and $\boldsymbol{I}_{(\cdot,\cdot)}$ is the commutation matrix.

Proof. *The first derivative of $\boldsymbol{\lambda}$ with respect to $\boldsymbol{\varphi}'$ can be written as*

$$
\begin{aligned}
\boldsymbol{D}_{\lambda}^{(1)} &= \frac{\partial}{\partial \boldsymbol{\varphi}'} \left\{ \varphi_0 \boldsymbol{T}_3 \exp\{\odot(\boldsymbol{T}_4 \boldsymbol{\xi})\} \left[\boldsymbol{1}_p' \boldsymbol{T}_3 \exp\{\odot(\boldsymbol{T}_4 \boldsymbol{\xi})\} \right]^{-1} \right\} \\
&= \varphi_0 \left\{ w \boldsymbol{T}_3 \boldsymbol{W}_1 \left(\boldsymbol{V}_1 \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\varphi}'} + \boldsymbol{V}_2 \right) - w \frac{\boldsymbol{\lambda}}{\varphi_0} \boldsymbol{1}_p' \boldsymbol{T}_3 \boldsymbol{W}_1 \left(\boldsymbol{V}_1 \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\varphi}'} + \boldsymbol{V}_2 \right) \right\},
\end{aligned}
$$

because $\frac{\partial \exp\{\odot(\boldsymbol{T}_4 \boldsymbol{\xi})\}}{\partial \boldsymbol{\varphi}'} = \boldsymbol{W}_1 \frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{\varphi}'}$, where $\boldsymbol{W}_1$ is defined in (2.13).

Therefore,

$$
\begin{aligned}
\boldsymbol{D}_{\lambda}^{(1)} &= \varphi_0 \left\{ w\boldsymbol{T}_3\,\boldsymbol{W}_1\,\boldsymbol{V}_2 - w\frac{\boldsymbol{\lambda}}{\varphi_0}\boldsymbol{1}_p'\boldsymbol{T}_3\,\boldsymbol{W}_1\,\boldsymbol{V}_2 \right\} \\
&= \varphi_0\, w \left(\boldsymbol{I}_p - \frac{\boldsymbol{\lambda}}{\varphi_0}\boldsymbol{1}_p' \right) \boldsymbol{T}_3\,\boldsymbol{W}_1\,\boldsymbol{V}_2 \\
&= \varphi_0\, w\, \boldsymbol{H}_{\boldsymbol{\lambda}}\,\boldsymbol{T}_3\,\boldsymbol{W}_1\,\boldsymbol{V}_2,
\end{aligned}
$$

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

$$
\begin{aligned}
\boldsymbol{D}_{\lambda}^{(2)} &= \frac{\partial}{\partial \boldsymbol{\varphi}'} \left\{ \varphi_0\, w\, \boldsymbol{H}_{\boldsymbol{\lambda}}\,\boldsymbol{T}_3\,\boldsymbol{W}_1 \left(\boldsymbol{V}_1 \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\varphi}'} + \boldsymbol{V}_2 \right) \right\} \\
&= \varphi_0 \left\{ \frac{\partial w}{\partial \boldsymbol{\varphi}'} (\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{H}_{\boldsymbol{\lambda}}\,\boldsymbol{T}_3\,\boldsymbol{W}_1\,\boldsymbol{V}_2) + w \frac{\partial\, \boldsymbol{H}_{\boldsymbol{\lambda}}}{\partial \boldsymbol{\varphi}'} (\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{T}_3\,\boldsymbol{W}_1\,\boldsymbol{V}_2) \right. \\
&\quad \left. + w\,\boldsymbol{H}_{\boldsymbol{\lambda}}\,\boldsymbol{T}_3 \frac{\partial\,\boldsymbol{W}_1}{\partial \boldsymbol{\varphi}'} (\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{V}_2) + w\,\boldsymbol{H}_{\boldsymbol{\lambda}}\,\boldsymbol{T}_3\,\boldsymbol{W}_1\,\boldsymbol{V}_1 \frac{\partial^2 \boldsymbol{\tau}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'} \right\}. \qquad (2.16)
\end{aligned}
$$

Applying standard procedures for differentiation yields the following results:

$$
\begin{aligned}
\frac{\partial w}{\partial \boldsymbol{\varphi}'} &= -w^2 \boldsymbol{1}_p'\boldsymbol{T}_3\,\boldsymbol{W}_1\,\boldsymbol{V}_2, & (2.17) \\
\frac{\partial\, \boldsymbol{H}_{\boldsymbol{\lambda}}}{\partial \boldsymbol{\varphi}'} &= -\frac{1}{\varphi_0} \left(\boldsymbol{D}_{\lambda}^{(1)} \otimes \boldsymbol{1}_p' \right), & (2.18) \\
\frac{\partial\, \boldsymbol{W}_1}{\partial \boldsymbol{\varphi}'} &= \boldsymbol{W}_2(\boldsymbol{V}_2 \otimes \boldsymbol{I}_{\nu_2}), \text{ and} & (2.19) \\
\frac{\partial^2 \boldsymbol{\tau}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'} &= -\left(\boldsymbol{D}_{\boldsymbol{\xi}}^{-1} \boldsymbol{U}_{\boldsymbol{\xi}}' \right) \boldsymbol{C}_2' \boldsymbol{W}_2 (\boldsymbol{V}_2 \otimes \boldsymbol{V}_2). & (2.20)
\end{aligned}
$$

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

$$
\begin{aligned}
\frac{\partial^2 \boldsymbol{\tau}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'} &= -(\boldsymbol{C}_2' \boldsymbol{W}_1 \boldsymbol{V}_1)^{-1} \boldsymbol{C}_2' \frac{\partial \boldsymbol{W}_1}{\partial \boldsymbol{\varphi}'} (\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{V}_2) \\
&= -\boldsymbol{D}_{\boldsymbol{\xi}}^{-1} \boldsymbol{U}_{\boldsymbol{\xi}}' \boldsymbol{C}_2' \frac{\partial \boldsymbol{W}_1}{\partial \boldsymbol{\varphi}'} (\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{V}_2), \qquad (2.21)
\end{aligned}
$$

because $V_1'V_1 = I_r$. *To compute* $\frac{\partial W_1}{\partial \varphi'}$, *express* W_1 *in term of summation as follows:*

$$W_1 = \sum_{i=1}^{q_3} e_i^{q_3} \exp\{e_i^{q_3\prime} T_4 \xi\} e_i^{q_3\prime} T_4.$$

Applying the product rule yields

$$\begin{aligned} \frac{\partial W_1}{\partial \varphi'} &= \sum_{i=1}^{q_3} \{e_i^{q_3} \exp\{e_i^{q_3\prime} T_4 \xi\}(e_i^{q_3\prime} T_4 \otimes e_i^{q_3\prime} T_4)\}(V_2 \otimes I_{\nu_2}) \\ &= W_2(V_2 \otimes I_{\nu_2}). \end{aligned} \tag{2.22}$$

Substituting (2.22) in (2.21) yields

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

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

$$\begin{aligned} D_\lambda^{(2)} &= -w\left(D_\lambda^{(1)} \otimes 1_p' T_3 W_1 V_2\right) I_{(\nu_2,\nu_2)} - w\left(D_\lambda^{(1)} \otimes 1_p' T_3 W_1 V_2\right) \\ &\quad + \varphi_0 w H_\lambda T_3 W_2 (V_2 \otimes V_2) - \varphi_0 w H_\lambda T_3 W_1 V_1 D_\xi^{-1} U_\xi' C_2' W_2 (V_2 \otimes V_2) \\ &= -2w\left(D_\lambda^{(1)} \otimes 1_p' T_3 W_1 V_2\right) N_{\nu_2} + \varphi_0 w H_\lambda T_3 H_\xi W_2 (V_2 \otimes V_2). \end{aligned}$$

To compute the third derivative of λ *with respect to* φ', *the following additional results are needed:*

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

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

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

$$(C_2' W_1 V_1)^{-1} = D_\xi^{-1} U_\xi'.$$

Then,

$$\begin{aligned}
\frac{\partial \boldsymbol{H}_\xi}{\partial \boldsymbol{\varphi}'} &= \frac{\partial}{\partial \boldsymbol{\varphi}'}\left\{\boldsymbol{I}_{q_3} - \boldsymbol{W}_1 \boldsymbol{V}_1 (\boldsymbol{C}_2' \boldsymbol{W}_1 \boldsymbol{V}_1)^{-1} \boldsymbol{C}_2'\right\} \\
&= -\frac{\partial \boldsymbol{W}_1}{\partial \boldsymbol{\varphi}'}\left(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{V}_1 \boldsymbol{D}_\xi^{-1} \boldsymbol{U}_\xi' \boldsymbol{C}_2'\right) - \boldsymbol{W}_1 \boldsymbol{V}_1 \frac{\partial (\boldsymbol{C}_2' \boldsymbol{W}_1 \boldsymbol{V}_1)^{-1}}{\partial \boldsymbol{\varphi}'}(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{C}_2') \\
&= -\boldsymbol{W}_2\left(\boldsymbol{V}_2 \otimes \boldsymbol{V}_1 \boldsymbol{D}_\xi^{-1} \boldsymbol{U}_\xi' \boldsymbol{C}_2'\right) + \boldsymbol{W}_1 \boldsymbol{V}_1 \boldsymbol{D}_\xi^{-1} \boldsymbol{U}_\xi' \boldsymbol{C}_2' \boldsymbol{W}_2\left(\boldsymbol{V}_2 \otimes \boldsymbol{V}_1 \boldsymbol{D}_\xi^{-1} \boldsymbol{U}_\xi' \boldsymbol{C}_2'\right) \\
&= -\left(\boldsymbol{I}_{q_3} - \boldsymbol{W}_1 \boldsymbol{V}_1 \boldsymbol{D}_\xi^{-1} \boldsymbol{U}_\xi' \boldsymbol{C}_2'\right) \boldsymbol{W}_2\left(\boldsymbol{V}_2 \otimes \boldsymbol{V}_1 \boldsymbol{D}_\xi^{-1} \boldsymbol{U}_\xi' \boldsymbol{C}_2'\right) \\
&= -\boldsymbol{H}_\xi \boldsymbol{W}_2\left(\boldsymbol{V}_2 \otimes \boldsymbol{V}_1 \boldsymbol{D}_\xi^{-1} \boldsymbol{U}_\xi' \boldsymbol{C}_2'\right),
\end{aligned}$$

by using the property

$$\frac{\partial \boldsymbol{X}^{-1}}{\partial \boldsymbol{\theta}'} = -\boldsymbol{X}^{-1}\frac{\partial \boldsymbol{X}}{\partial \boldsymbol{\theta}'}(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{X}^{-1}),$$

where $\boldsymbol{X}$ *is a nonsingular matrix. The third derivative of* $\boldsymbol{\lambda}$ *with respect to* $\boldsymbol{\varphi}'$ *can be written as*

$$\begin{aligned}
\boldsymbol{D}_\lambda^{(3)} &= \frac{\partial}{\partial \boldsymbol{\varphi}'}\Bigg\{-2w\left[\boldsymbol{D}_\lambda^{(1)} \otimes \boldsymbol{1}_p' \boldsymbol{T}_3 \boldsymbol{W}_1\left(\boldsymbol{V}_1 \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\varphi}'} + \boldsymbol{V}_2\right)\right] \boldsymbol{N}_{\nu_2} \\
&\quad + \varphi_0 w \boldsymbol{H}_{\boldsymbol{\lambda}} \boldsymbol{T}_3 \boldsymbol{H}_\xi \boldsymbol{W}_2\left[\left(\boldsymbol{V}_1 \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\varphi}'} + \boldsymbol{V}_2\right) \otimes \left(\boldsymbol{V}_1 \frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\varphi}'} + \boldsymbol{V}_2\right)\right]\Bigg\}. \qquad (2.25)
\end{aligned}$$

For convenience, the two parts on the right-hand-side of (2.25) are labeled as $\boldsymbol{D}_\lambda^{(3)}(a)$ *and* $\boldsymbol{D}_\lambda^{(3)}(b)$. *Simplifying* $\boldsymbol{D}_\lambda^{(3)}(a)$ *yields*

$$\begin{aligned}
\boldsymbol{D}_\lambda^{(3)}(a) &= -2\Bigg\{\frac{\partial w}{\partial \boldsymbol{\varphi}'}\left(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{D}_\lambda^{(1)} \otimes \boldsymbol{1}_p' \boldsymbol{T}_3 \boldsymbol{W}_1 \boldsymbol{V}_2\right) \\
&\quad + w\left(\mathbf{D}_{\lambda,\varphi}^{(2)} \otimes \boldsymbol{1}_p' \boldsymbol{T}_3 \boldsymbol{W}_1 \boldsymbol{V}_2\right) \\
&\quad + w\left(\boldsymbol{1}_p' \boldsymbol{T}_3 \frac{\partial \boldsymbol{W}_1}{\partial \boldsymbol{\varphi}'}(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{V}_2) \otimes \boldsymbol{D}_\lambda^{(1)}\right) \\
&\quad + w\left(\boldsymbol{1}_p' \boldsymbol{T}_3 \boldsymbol{W}_1 \boldsymbol{V}_1 \frac{\partial^2 \boldsymbol{\tau}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\varphi}'} \otimes \boldsymbol{D}_\lambda^{(1)}\right)\Bigg\}(\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{N}_{\nu_2}). \qquad (2.26)
\end{aligned}$$

Applying the results (2.17)-(2.20) and (2.23)-(2.24) gives

$$
\begin{aligned}
\boldsymbol{D}_\lambda^{(3)}(a) &= 2w\Big\{w\left(\mathbf{1}_p'\boldsymbol{T}_3\,\boldsymbol{W}_1\,\boldsymbol{V}_2\otimes\boldsymbol{D}_\lambda^{(1)}\otimes\mathbf{1}_p'\boldsymbol{T}_3\,\boldsymbol{W}_1\,\boldsymbol{V}_2\right)\\
&-\left(\boldsymbol{D}_\lambda^{(2)}\otimes\mathbf{1}_p'\boldsymbol{T}_3\,\boldsymbol{W}_1\,\boldsymbol{V}_2\right)\\
&-\left(\mathbf{1}_p'\boldsymbol{T}_3\,\boldsymbol{H}_\xi\,\boldsymbol{W}_2(\boldsymbol{V}_2\otimes\boldsymbol{V}_2)\otimes\boldsymbol{D}_\lambda^{(1)}\right)\Big\}(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{N}_{\nu_2}). \qquad (2.27)
\end{aligned}
$$

Simplifying $\boldsymbol{D}_\lambda^{(3)}(b)$ *yields*

$$
\begin{aligned}
\boldsymbol{D}_\lambda^{(3)}(b) &= \varphi_0\Big\{\frac{\partial w}{\partial\boldsymbol{\varphi}'}\left(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{H}_{\boldsymbol{\lambda}}\,\boldsymbol{T}_3\,\boldsymbol{H}_\xi\,\boldsymbol{W}_2(\boldsymbol{V}_2\otimes\boldsymbol{V}_2)\right)\\
&+w\frac{\partial\boldsymbol{H}_{\boldsymbol{\lambda}}}{\partial\boldsymbol{\varphi}'}(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{T}_3\,\boldsymbol{H}_\xi\,\boldsymbol{W}_2(\boldsymbol{V}_2\otimes\boldsymbol{V}_2))\\
&+w\,\boldsymbol{H}_{\boldsymbol{\lambda}}\,\boldsymbol{T}_3\frac{\partial\boldsymbol{H}_\xi}{\partial\boldsymbol{\varphi}'}(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{W}_2(\boldsymbol{V}_2\otimes\boldsymbol{V}_2))+w\,\boldsymbol{H}_{\boldsymbol{\lambda}}\,\boldsymbol{T}_3\,\boldsymbol{H}_\xi\,\frac{\partial\boldsymbol{W}_2}{\partial\boldsymbol{\varphi}'}(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{V}_2\otimes\boldsymbol{V}_2)\\
&+w\,\boldsymbol{H}_{\boldsymbol{\lambda}}\,\boldsymbol{T}_3\,\boldsymbol{H}_\xi\,\boldsymbol{W}_2\left(\boldsymbol{V}_1\frac{\partial^2\boldsymbol{\tau}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'}\otimes\boldsymbol{V}_2\right)\\
&+w\,\boldsymbol{H}_{\boldsymbol{\lambda}}\,\boldsymbol{T}_3\,\boldsymbol{H}_\xi\,\boldsymbol{W}_2\left(\boldsymbol{I}_{(q_4,q_4)}\left(\boldsymbol{V}_1\frac{\partial^2\boldsymbol{\tau}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'}\otimes\boldsymbol{V}_2\right)(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{I}_{(\nu_2,\nu_2)})\right)\Big\}\\
&= -\varphi_0\,w^2\left(\mathbf{1}_p'\boldsymbol{T}_3\,\boldsymbol{W}_1\,\boldsymbol{V}_2\otimes\boldsymbol{H}_{\boldsymbol{\lambda}}\,\boldsymbol{T}_3\,\boldsymbol{H}_\xi\,\boldsymbol{W}_2(\boldsymbol{V}_2\otimes\boldsymbol{V}_2)\right)\\
&-w\left(\boldsymbol{D}_\lambda^{(1)}\otimes\mathbf{1}_p'\boldsymbol{T}_3\,\boldsymbol{H}_\xi\,\boldsymbol{W}_2(\boldsymbol{V}_2\otimes\boldsymbol{V}_2)\right)\\
&-\varphi_0\,w\,\boldsymbol{H}_{\boldsymbol{\lambda}}\,\boldsymbol{T}_3\,\boldsymbol{H}_\xi\,\boldsymbol{W}_2\left(\boldsymbol{V}_2\otimes\boldsymbol{V}_1\,\boldsymbol{D}_\xi^{-1}\,\boldsymbol{U}_\xi'\,\boldsymbol{C}_2'\boldsymbol{W}_2(\boldsymbol{V}_2\otimes\boldsymbol{V}_2)\right)\\
&+\varphi_0\,w\,\boldsymbol{H}_{\boldsymbol{\lambda}}\,\boldsymbol{T}_3\,\boldsymbol{H}_\xi\,\boldsymbol{W}_3(\boldsymbol{V}_2\otimes\boldsymbol{V}_2\otimes\boldsymbol{V}_2)\\
&+\varphi_0\,w\,\boldsymbol{H}_{\boldsymbol{\lambda}}\,\boldsymbol{T}_3\,\boldsymbol{H}_\xi\,\boldsymbol{W}_2\left(\boldsymbol{V}_1\,\boldsymbol{D}_\xi^{-1}\,\boldsymbol{U}_\xi'\,\boldsymbol{C}_2'\boldsymbol{W}_2(\boldsymbol{V}_2\otimes\boldsymbol{V}_2)\otimes\boldsymbol{V}_2\right)\\
&+\varphi_0\,ww\,\boldsymbol{H}_{\boldsymbol{\lambda}}\,\boldsymbol{T}_3\,\boldsymbol{H}_\xi\,\boldsymbol{W}_2\left(\boldsymbol{V}_1\,\boldsymbol{D}_\xi^{-1}\,\boldsymbol{U}_\xi'\,\boldsymbol{C}_2'\boldsymbol{W}_2(\boldsymbol{V}_2\otimes\boldsymbol{V}_2)\otimes\boldsymbol{V}_2\right)\\
&\times(\boldsymbol{I}_{\nu_2}\otimes\boldsymbol{I}_{(\nu_2,\nu_2)}). \qquad (2.28)
\end{aligned}
$$

Additional simplifications yield

$$\begin{aligned}
\boldsymbol{D}_{\lambda}^{(3)}(b) &= -\varphi_0\, w^2 \left(\mathbf{1}_p' \boldsymbol{T}_3\, \boldsymbol{W}_1\, \boldsymbol{V}_2 \otimes \boldsymbol{H}_{\boldsymbol{\lambda}}\, \boldsymbol{T}_3\, \boldsymbol{H}_{\boldsymbol{\xi}}\, \boldsymbol{W}_2(\boldsymbol{V}_2 \otimes \boldsymbol{V}_2)\right) \\
&\quad -w \left(\boldsymbol{D}_{\lambda}^{(1)} \otimes \mathbf{1}_p' \boldsymbol{T}_3\, \boldsymbol{H}_{\boldsymbol{\xi}}\, \boldsymbol{W}_2(\boldsymbol{V}_2 \otimes \boldsymbol{V}_2)\right) \\
&\quad + \varphi_0\, w\, \boldsymbol{H}_{\boldsymbol{\lambda}}\, \boldsymbol{T}_3\, \boldsymbol{H}_{\boldsymbol{\xi}}\, \boldsymbol{W}_3(\boldsymbol{V}_2 \otimes \boldsymbol{V}_2 \otimes \boldsymbol{V}_2) \\
&\quad -2\, \varphi_0\, w\, \boldsymbol{H}_{\boldsymbol{\lambda}}\, \boldsymbol{T}_3\, \boldsymbol{H}_{\boldsymbol{\xi}}\, \boldsymbol{W}_2 \left(\boldsymbol{V}_1\, \boldsymbol{D}_{\boldsymbol{\xi}}^{-1}\, \boldsymbol{U}_{\boldsymbol{\xi}}'\, \boldsymbol{C}_2'(\boldsymbol{V}_2 \otimes \boldsymbol{V}_2) \otimes \boldsymbol{V}_2\right) \boldsymbol{N}_{\nu_2}^{*}, \quad (2.29)
\end{aligned}$$

where $\boldsymbol{N}_{\nu_2}^{*} = \frac{1}{2}\boldsymbol{I}_{(\nu_2,\nu_2^2)} + (\boldsymbol{I}_{\nu_2} \otimes \boldsymbol{N}_{\nu_2})$. *Combining (2.27) and (2.29) yields*

$$\begin{aligned}
\boldsymbol{D}_{\lambda}^{(3)} &= -2w\varphi_0 \boldsymbol{H}_{\boldsymbol{\lambda}} \boldsymbol{T}_3 \boldsymbol{H}_{\boldsymbol{\xi}} \boldsymbol{W}_2 \left[\boldsymbol{V}_1 \boldsymbol{D}_{\boldsymbol{\xi}}^{-1} \boldsymbol{U}_{\boldsymbol{\xi}}' \boldsymbol{C}_2' \boldsymbol{W}_2(\boldsymbol{V}_2 \otimes \boldsymbol{V}_2) \otimes \boldsymbol{V}_2\right] \boldsymbol{N}_{\nu_2}^{*} \\
&\quad -2w \left\{\left(\boldsymbol{D}_{\lambda}^{(1)} \otimes \mathbf{1}_p' \boldsymbol{T}_3 \boldsymbol{W}_1 \boldsymbol{V}_2\right) + \left[\mathbf{1}_p' \boldsymbol{T}_3 \boldsymbol{H}_{\boldsymbol{\xi}} \boldsymbol{W}_2(\boldsymbol{V}_2 \otimes \boldsymbol{V}_2) \otimes \boldsymbol{D}_{\lambda}^{(1)}\right]\right\} \boldsymbol{N}_{\nu_2}^{*} \\
&\quad + w\varphi_0 \boldsymbol{H}_{\boldsymbol{\lambda}} \boldsymbol{T}_3 \boldsymbol{H}_{\boldsymbol{\xi}} \boldsymbol{W}_3(\boldsymbol{V}_2 \otimes \boldsymbol{V}_2 \otimes \boldsymbol{V}_2).
\end{aligned}$$

□

Corollary 1. If $\boldsymbol{C}_2$ *has rank = 0, that is there are no additional constraints, then* $r = 0,\ \boldsymbol{V}_2 = \boldsymbol{I}_{q_4},\ \nu_2 = q_4,\ \boldsymbol{\varphi} = \boldsymbol{\xi}$ *and the derivatives simplify to*

$$\begin{aligned}
\boldsymbol{D}_{\lambda}^{(1)} &= w\varphi_0 \boldsymbol{H}_{\boldsymbol{\lambda}} \boldsymbol{T}_3 \boldsymbol{W}_1, \\
\boldsymbol{D}_{\lambda}^{(2)} &= -2w \left(\boldsymbol{D}_{\lambda}^{(1)} \otimes \mathbf{1}_p' \boldsymbol{T}_3 \boldsymbol{W}_1\right) \boldsymbol{N}_{\nu_2} + w\varphi_0 \boldsymbol{H}_{\boldsymbol{\lambda}} \boldsymbol{T}_3 \boldsymbol{W}_2, \\
\boldsymbol{D}_{\lambda}^{(3)} &= -2w \left\{\left(\boldsymbol{D}_{\lambda}^{(2)} \otimes \mathbf{1}_p' \boldsymbol{T}_3 \boldsymbol{W}_1\right) + \left[\mathbf{1}_p' \boldsymbol{T}_3 \boldsymbol{W}_2 \otimes \boldsymbol{D}_{\lambda}^{(1)}\right]\right\} \boldsymbol{N}_{\nu_2}^{*} \\
&\quad + w\varphi_0 \boldsymbol{H}_{\boldsymbol{\lambda}} \boldsymbol{T}_3 \boldsymbol{W}_3.
\end{aligned}$$

Since φ_0 defined in (2.7) is a parameter, the first three derivatives of $\boldsymbol{\lambda}$ in (2.9) with respect to φ_0 are useful for estimating parameters and for constructing the tests. The results are shown in Theorem 4.

Theorem 4. *The first derivative of* $\boldsymbol{\lambda}$ *in (2.9) with respect to* φ_0 *can be written as*

$$\boldsymbol{D}^{(1)}_{\lambda,\varphi_0} = \frac{\partial \boldsymbol{\lambda}}{\partial \varphi_0} = \frac{\boldsymbol{T}_3 \exp\{\odot(\boldsymbol{T}_4\boldsymbol{\xi})\}}{\mathbf{1}_p'\boldsymbol{T}_3 \exp\{\odot(\boldsymbol{T}_4\boldsymbol{\xi})\}} = \frac{\boldsymbol{\lambda}}{\varphi_0}$$

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 $\boldsymbol{\Gamma}$ be a $p \times p$ matrix whose columns consist of the eigenvectors of $\boldsymbol{\Sigma}$ and let $\boldsymbol{\Gamma}_0$ be an orthogonal matrix in a small neighborhood of $\boldsymbol{\Gamma}$. That is $\boldsymbol{\Gamma} \approx \boldsymbol{\Gamma}_0$. Boik's parameterization is based on the trivial equality $\boldsymbol{\Gamma} = \boldsymbol{I}_p\boldsymbol{\Gamma}$. Write the identity matrix as an orthogonal matrix, $\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0'$. Let $\boldsymbol{G}(\boldsymbol{\mu}^*) = \boldsymbol{\Gamma}_0'\boldsymbol{\Gamma}$, where $\boldsymbol{G}(\boldsymbol{\mu}^*) \in \mathcal{O}(p)$ and $\boldsymbol{\mu}^*$ is a vector of parameters. Then, $\boldsymbol{G}(\boldsymbol{\mu}^*) \approx \boldsymbol{I}_p$ and the parameterization of eigenvectors can be written as

$$\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_0\boldsymbol{G}(\boldsymbol{\mu}^*). \tag{2.30}$$

It is assumed, for now, that $\boldsymbol{\Gamma}_0$ is a known matrix. This assumption is not made in Chapter 3 when estimation is discussed. By orthogonality, $\boldsymbol{G}$ is subject to $p(p+1)/2$ constraints, $\boldsymbol{G}\boldsymbol{G}' = \boldsymbol{I}_p$. If each eigenvalue has multiplicity 1, then $\boldsymbol{G}$ is a function of $p(p-1)/2$ parameters denoted by $\boldsymbol{\mu}^*$. The $p(p-1)/2$ elements of $\boldsymbol{\mu}^*$ correspond to the $p(p-1)/2$ elements in the upper triangle of $\boldsymbol{G}$. Let $\boldsymbol{\eta}^*$ be an implicit function of $\boldsymbol{\mu}^*$, then $\boldsymbol{\eta}^*$ has $p(p+1)/2$ elements. Let $\boldsymbol{m} = (m_1\ m_2\ \cdots\ m_{\nu_2})'$ be the vector of eigenvalue multiplicities and denote the distinct eigenvalues by $(\varphi_1, \varphi_2, \ldots, \varphi_{\nu_2})$ where $\nu_2 = \dim(\boldsymbol{\varphi})$, $\nu_2 \leq p$, and $\sum_{i=1}^{\nu_2} m_i = p$. The columns of $\boldsymbol{G}$ can be partitioned with respect to the multiplicity vector $\boldsymbol{m}$ as follows:

$$\boldsymbol{G} = \begin{pmatrix} \underset{p\times m_1}{\boldsymbol{G}_1} & \underset{p\times m_2}{\boldsymbol{G}_2} & \cdots & \underset{p\times m_{\nu_2}}{\boldsymbol{G}_{\nu_2}} \end{pmatrix}.$$

The diagonal form of $\boldsymbol{\Sigma}$ can be written as

$$\begin{aligned} \boldsymbol{\Sigma} &= \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}' = \boldsymbol{\Gamma}_0\boldsymbol{G}\boldsymbol{\Lambda}\boldsymbol{G}'\boldsymbol{\Gamma}_0' = \sum_{i=1}^{\nu_2} \boldsymbol{\Gamma}_0\boldsymbol{G}_i\varphi_i\boldsymbol{I}_{m_i}\boldsymbol{G}_i'\boldsymbol{\Gamma}_0' \\ &= \boldsymbol{\Gamma}_0\left\{\sum_{i=1}^{\nu_2}\varphi_i\boldsymbol{G}_i\boldsymbol{G}_i'\right\}\boldsymbol{\Gamma}_0' = \boldsymbol{\Gamma}_0\left\{\sum_{i=1}^{\nu_2}\varphi_i\boldsymbol{G}_i\boldsymbol{Q}_i\boldsymbol{Q}_i'\boldsymbol{G}_i'\right\}\boldsymbol{\Gamma}_0' \end{aligned}$$

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

$$\boldsymbol{G} = \begin{pmatrix} \eta_1^* & \mu_1^* & \mu_2^* & \mu_3^* & \mu_4^* & \mu_8^* \\ \eta_2^* & \eta_7^* & 0 & 0 & \mu_5^* & \mu_9^* \\ \eta_3^* & \eta_8^* & \eta_{12}^* & 0 & \mu_6^* & \mu_{10}^* \\ \eta_4^* & \eta_9^* & \eta_{13}^* & \eta_{16}^* & \mu_7^* & \mu_{11}^* \\ \eta_5^* & \eta_{10}^* & \eta_{14}^* & \eta_{17}^* & \eta_{19}^* & 0 \\ \eta_6^* & \eta_{11}^* & \eta_{15}^* & \eta_{18}^* & \eta_{20}^* & \eta_{21}^* \end{pmatrix} = \begin{pmatrix} \underset{6\times 1}{\boldsymbol{G}_1} & \underset{6\times 3}{\boldsymbol{G}_2} & \underset{6\times 2}{\boldsymbol{G}_3} \end{pmatrix}.$$

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

$$\operatorname{vec}\boldsymbol{G} = \boldsymbol{A}_1\boldsymbol{\mu}^* + \boldsymbol{A}_2\boldsymbol{\eta}^*, \tag{2.31}$$

where $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are known indicator matrices with dimensions $p^2 \times ((p^2 - \boldsymbol{m}'\boldsymbol{m})/2)$ and $p^2 \times (p(p+1)/2)$, respectively. Because $\boldsymbol{A}_1'\boldsymbol{A}_1 = \boldsymbol{I}_{(p^2-\boldsymbol{m}'\boldsymbol{m})/2}$, $\boldsymbol{A}_2'\boldsymbol{A}_2 = \boldsymbol{I}_{p(p+1)/2}$, and $\boldsymbol{A}_1'\boldsymbol{A}_2 = \boldsymbol{0}$, $\operatorname{vec}\boldsymbol{G}$ satisfies $\boldsymbol{A}_1'\operatorname{vec}\boldsymbol{G} = \boldsymbol{\mu}^*$ and $\boldsymbol{A}_2'\operatorname{vec}\boldsymbol{G} = \boldsymbol{\eta}^*$.

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

$$\mathrm{H}_0 : \boldsymbol{A} \in \mathcal{R}(\boldsymbol{\Gamma M}),$$

where $\mathbf{A}$ is a known $p \times q$ semi-orthogonal matrix with rank q and $\boldsymbol{M}$ is a $p \times m$ semi-orthogonal matrix such that $\mathcal{R}(\boldsymbol{M}) = \mathcal{R}\begin{pmatrix} \boldsymbol{I}_m \\ \boldsymbol{0} \end{pmatrix}$. The restriction $\boldsymbol{A} \in \mathcal{R}(\boldsymbol{\Gamma M})$ further reduces the number of entries in $\boldsymbol{\mu}^*$. Theorem 5 describes this reduction. First, however, note that the parameterization in (2.30) is valid for any matrix $\boldsymbol{\Gamma}_0$ in a neighborhood of $\boldsymbol{\Gamma}$. It is most convenient to equate $\boldsymbol{\Gamma}_0$ to $\boldsymbol{\Gamma}$. In this way, the true value of $\boldsymbol{\mu}^*$ becomes zero. That is,

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

because $\boldsymbol{G}(\boldsymbol{0}) = \boldsymbol{I}_p$. In the remainder of this thesis, $\boldsymbol{\Gamma}_0$ will be taken to be $\boldsymbol{\Gamma}$.

Theorem 5. *Denote* $\boldsymbol{C} = (\boldsymbol{M}^{c\prime} \otimes \boldsymbol{A}'\boldsymbol{\Gamma})$*, where* $\boldsymbol{M}^c\boldsymbol{M}^{c\prime} = \boldsymbol{I}_p - \boldsymbol{M}\boldsymbol{M}'$ *and* $\boldsymbol{\Gamma} = \boldsymbol{\Gamma G}(\boldsymbol{\mu}^*)\big|_{\boldsymbol{\mu}^*=\boldsymbol{0}}$. *Then,*

$$\boldsymbol{A} \in \mathcal{R}(\boldsymbol{\Gamma M}) \Longleftrightarrow \boldsymbol{C}\operatorname{vec}\boldsymbol{G} = \boldsymbol{0}.$$

Proof. It will be first shown that $\boldsymbol{A} \in \mathcal{R}(\boldsymbol{\Gamma M}) \Longleftrightarrow \boldsymbol{A}'\boldsymbol{\Gamma M}^c = \boldsymbol{0}$. *The proof consists of the following two steps:*

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

Consider $\boldsymbol{A} = (\boldsymbol{\Gamma M M}'\boldsymbol{\Gamma}' + \boldsymbol{\Gamma M}^c\boldsymbol{M}^{c\prime}\boldsymbol{\Gamma}')\,\boldsymbol{A}$.

$$
\begin{aligned}
\boldsymbol{A} \in \mathcal{R}(\boldsymbol{\Gamma M}) &\Longrightarrow \mathbf{\Gamma M M'\Gamma' A = A} \\
&\Longrightarrow \boldsymbol{\Gamma M}^c\boldsymbol{M}^{c\prime}\boldsymbol{\Gamma}'\boldsymbol{A} = 0 \\
&\Longrightarrow \boldsymbol{A}'\boldsymbol{\Gamma M}^c = \boldsymbol{0}.
\end{aligned}
$$

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

$$
\begin{aligned}
\boldsymbol{A}'\boldsymbol{\Gamma M}^c = \boldsymbol{0} &\Longrightarrow \boldsymbol{A} = \boldsymbol{\Gamma M M}'\boldsymbol{\Gamma}'\boldsymbol{A} \\
&\Longrightarrow \boldsymbol{A} \in \mathcal{R}(\boldsymbol{\Gamma M}).
\end{aligned}
$$

Applying the property of the vec,

$$
\operatorname{vec}(\boldsymbol{XYZ}) = (\boldsymbol{Z}' \otimes \boldsymbol{X}) \operatorname{vec} \boldsymbol{Y}, \tag{2.32}
$$

where $\boldsymbol{X}$, $\boldsymbol{Y}$, *and* $\boldsymbol{Z}$ *are any matrices that are conformable for multiplication, yields*

$$
\begin{aligned}
\boldsymbol{A} \in \mathcal{R}(\boldsymbol{\Gamma M}) &\Longleftrightarrow \boldsymbol{A}'\boldsymbol{\Gamma M}^c = 0 \\
&\Longleftrightarrow \operatorname{vec}(\boldsymbol{A}'\boldsymbol{\Gamma M}^c) = 0 \\
&\Longleftrightarrow \operatorname{vec}(\boldsymbol{A}'\boldsymbol{\Gamma G M}^c) = 0 \\
&\Longleftrightarrow \boldsymbol{C} \operatorname{vec} \boldsymbol{G} = 0.
\end{aligned}
$$

□

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

Let $V^\star$ be a $(p^2 - m'm)/2 \times (p^2 - m'm)/2$ nonsingular matrix and $V^\star = (V_3 \quad V_4)$, 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, $\operatorname{vec} G$ can be rewritten as

$$\begin{aligned} \operatorname{vec} G &= A_1 \left(V^* V^{*^{-1}}\right) \mu^* + A_2 \eta^* \\ &= A_1 V_3 (V_3' V_3)^{-1} V_3' \mu^* + A_1 V_4 (V_4' V_4)^{-1} V_4' \mu^* + A_2 \eta^* \\ &= A_1 V_3 \mu + A_1 V_4 \eta_1 + A_2 \eta_2, \end{aligned} \tag{2.33}$$

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,

$$\operatorname{vec} G = A_1 V_3 \mu + [A_1 V_4 \quad A_2] \eta, \text{ where } \eta = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \tag{2.34}$$

and η is an implicit function of μ. Taking the derivative of the constraints with respect to η' yields

$$\frac{\partial}{\partial \eta'} \begin{pmatrix} D_p' \operatorname{vec} GG' \\ C \operatorname{vec} G \end{pmatrix} \Bigg|_{\mu=0} = \begin{bmatrix} 2D_p' A_1 V_4 & 2D_p' A_2 \\ C A_1 V_4 & C A_2 \end{bmatrix}, \tag{2.35}$$

where D_p is the duplication matrix (Magnus and Neudecker, 1999), D_p satisfies $\operatorname{vec} H = D_p \operatorname{vech} H$, 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 η is an implicit function of μ implying that $\frac{\partial \eta}{\partial \mu'}$ exists. To verify that (2.35) is nonsingular, the following results are useful:

$$D_p' A_2 = I_{p(p+1)/2} \text{ and } \tag{2.36}$$

$$A_2 D_p' A_1 = I_{(p,p)} A_1. \tag{2.37}$$

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

$$\text{rank}\begin{bmatrix} 2D_p' A_1 V_4 & 2D_p' A_2 \\ C A_1 V_4 & C A_2 \end{bmatrix} = \text{rank}\begin{bmatrix} 2D_p' A_2 & 2D_p' A_1 V_4 \\ C A_2 & C A_1 V_4 \end{bmatrix}.$$

Because the matrix $D_p' A_2$ is nonsingular with rank $\frac{p(p+1)}{2}$ by (2.36), the derivative matrix in (2.35) is nonsingular if and only if rank $\left(C A_1 V_4 - C A_2 D_p' A_1 V_4\right) = (p - m)q$ (Ouellette, 1981). That is,

$$C(I_{p^2} - I_{(p,p)}) A_1 V_4 \tag{2.38}$$

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 μ' for a special choice of $V^\star$ 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*

$$W_4 = U(D \quad 0)\begin{pmatrix} V_4' \\ V_3' \end{pmatrix} = U D V_4', \tag{2.39}$$

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* $\boldsymbol{W}_4 = \boldsymbol{C}(\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)})\boldsymbol{A}_1$ *has full row-rank. The quantity* $\boldsymbol{W}_4$ *has full row-rank if and only if*

$$\boldsymbol{t}'\boldsymbol{C}(\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)})\boldsymbol{A}_1 = \boldsymbol{0} \quad \Longrightarrow \quad \boldsymbol{t} = \boldsymbol{0}. \tag{2.40}$$

It follows from (1.21) that

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

where $\boldsymbol{R}$ *is any* $(p-m) \times (p-m)$ *orthogonal matrix.*

Therefore,

$$\boldsymbol{C} = (\boldsymbol{M}^{c\prime} \otimes \boldsymbol{A}'\boldsymbol{\Gamma}) \;=\; \Big[(\boldsymbol{0}_{(p-m\times m)} \quad \boldsymbol{R}') \otimes \boldsymbol{A}'\boldsymbol{\Gamma}\Big]. \tag{2.41}$$

Let $\boldsymbol{\Gamma} = (\boldsymbol{\Gamma}_1 \quad \boldsymbol{\Gamma}_2)$ *which have dimensions* $p \times m$ *and* $p \times (p-m)$*, respectively. It also follows from (1.21) that*

$$\begin{aligned} \boldsymbol{A} \in \mathcal{R}(\boldsymbol{\Gamma}\boldsymbol{M}) \Longrightarrow \boldsymbol{A} \in \mathcal{R}(\boldsymbol{\Gamma}_1) \quad &\Longrightarrow \quad \boldsymbol{A} = \boldsymbol{\Gamma}_1\boldsymbol{T} \quad \text{for some } \boldsymbol{T} \\ &\Longrightarrow \quad \boldsymbol{A}' = \boldsymbol{T}'\boldsymbol{\Gamma}_1' \\ &\Longrightarrow \quad \boldsymbol{A}'\boldsymbol{\Gamma} = (\boldsymbol{T}' \quad 0), \end{aligned}$$

because $\boldsymbol{\Gamma}_1$ *and* $\boldsymbol{\Gamma}_2$ *are orthogonal. Substituting in (2.41) yields*

$$\boldsymbol{C} = (\boldsymbol{M}^{c\prime} \otimes \boldsymbol{A}'\boldsymbol{\Gamma}) = \Big([\boldsymbol{0} \quad \boldsymbol{R}'] \otimes [\boldsymbol{T}' \quad 0]\Big). \tag{2.42}$$

Expressing $\boldsymbol{A}_1$ *as* $\sum_{i=1}^{p}\sum_{j=i}^{p}(\boldsymbol{e}_j^p \otimes \boldsymbol{e}_i^p)\boldsymbol{e}_f^{w\prime}$*, where* $j \geq i$ *and* $w = \frac{p^2 - \boldsymbol{m}'\boldsymbol{m}}{2}$*, the left-hand-side of (2.40) can be written as*

$$\sum_{i=1}^{p}\sum_{j=i}^{p}\boldsymbol{t}'\Big([\boldsymbol{0} \quad \boldsymbol{R}'] \otimes [\boldsymbol{T}' \quad 0]\Big)(\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)})(\boldsymbol{e}_j^p \otimes \boldsymbol{e}_i^p)\boldsymbol{e}_f^{w\prime} = \boldsymbol{0}.$$

The following properties of the vec are applied:

$$(\boldsymbol{X}' \otimes \boldsymbol{Y})\boldsymbol{a}_i = \text{vec}(\boldsymbol{Y}\,\text{dvec}(\boldsymbol{a}_i)\boldsymbol{X}) \quad \text{and} \tag{2.43}$$

$$\text{dvec}\left(\text{vec}(\boldsymbol{e}_i^p \boldsymbol{e}_j^{p\prime}), p, p\right) = \boldsymbol{e}_i^p \boldsymbol{e}_j^{p\prime}, \tag{2.44}$$

where $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *are any matrices.*

Accordingly,

$$\sum_{i=1}^{p}\sum_{j=i}^{p}\left\{\boldsymbol{t}'\,\text{vec}\left([\boldsymbol{T}' \quad \boldsymbol{0}](\boldsymbol{e}_i^p \boldsymbol{e}_j^{p\prime})\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{R}\end{pmatrix}\right)\boldsymbol{e}_f^{w\prime} - \boldsymbol{t}'\,\text{vec}\left([\boldsymbol{T}' \quad \boldsymbol{0}](\boldsymbol{e}_j^p \boldsymbol{e}_i^{p\prime})\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{R}\end{pmatrix}\right)\boldsymbol{e}_f^{w\prime}\right\} = \boldsymbol{0}.$$

Therefore,

$$\sum_{i=1}^{p}\sum_{j=i}^{p}\boldsymbol{t}'\,\text{vec}\left([\boldsymbol{T}' \quad \boldsymbol{0}](\boldsymbol{e}_i^p \boldsymbol{e}_j^{p\prime})\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{R}\end{pmatrix}\right) = \sum_{i=1}^{p}\sum_{j=i}^{p}\boldsymbol{t}'\,\text{vec}\left([\boldsymbol{T}' \quad \boldsymbol{0}](\boldsymbol{e}_j^p \boldsymbol{e}_i^{p\prime}) \times \begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{R}\end{pmatrix}\right).$$

Define $\boldsymbol{T}^*$ *so that* $\boldsymbol{t} = \text{vec}\,\boldsymbol{T}^{*\prime}$, *then*

$$\begin{aligned}
\text{tr}\left(\boldsymbol{T}^*[\boldsymbol{T}' \quad \boldsymbol{0}](\boldsymbol{e}_i^p \boldsymbol{e}_j^{p\prime})\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{R}\end{pmatrix}\right) &= \text{tr}\left(\boldsymbol{T}^*[\boldsymbol{T}' \quad \boldsymbol{0}](\boldsymbol{e}_j^p \boldsymbol{e}_i^{p\prime})\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{R}\end{pmatrix}\right)\\
\Longrightarrow \boldsymbol{e}_j^{p\prime}\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{R}\end{pmatrix}\boldsymbol{T}^*[\boldsymbol{T}' \quad \boldsymbol{0}]\boldsymbol{e}_i^p &= \boldsymbol{e}_i^{p\prime}\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{R}\end{pmatrix}\boldsymbol{T}^*[\boldsymbol{T}' \quad \boldsymbol{0}]\boldsymbol{e}_j^p,
\end{aligned}$$

which implies that $\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{R}\end{pmatrix}\boldsymbol{T}^*[\boldsymbol{T}' \quad \boldsymbol{0}]$ *is symmetric.*

Finally,

$$\begin{aligned}
\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{R}\end{pmatrix}\boldsymbol{T}^*[\boldsymbol{T}' \quad \boldsymbol{0}] &= \begin{pmatrix}\boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{R}\boldsymbol{T}^*\boldsymbol{T}' & \boldsymbol{0}\end{pmatrix} \quad \text{and}\\
\boldsymbol{R}\boldsymbol{T}^*\boldsymbol{T}' &= \boldsymbol{0}\\
\Longrightarrow \boldsymbol{T}^* &= \boldsymbol{0} \Longrightarrow \boldsymbol{t} = \boldsymbol{0},
\end{aligned}$$

because $\boldsymbol{R}$ is orthogonal and $\boldsymbol{T}$ has full column-rank. Hence, $\boldsymbol{W}_4$ has full row-rank and for fixed $\boldsymbol{V}^*$ in (2.39), $\boldsymbol{W}_4\boldsymbol{V}_4 = \boldsymbol{U}\boldsymbol{D}$ is nonsingular. □

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

$$\frac{\partial \boldsymbol{\eta}_1}{\partial \boldsymbol{\mu}'} = \boldsymbol{0}, \tag{2.45}$$

$$\frac{\partial \boldsymbol{\eta}_2}{\partial \boldsymbol{\mu}'} = -\boldsymbol{D}_p'\boldsymbol{A}_1\boldsymbol{V}_3. \tag{2.46}$$

To verify the above equations, take the derivative of the vec of the first constraint, $\boldsymbol{G}\boldsymbol{G}' = \boldsymbol{I}_p$, and use property (2.32) to obtain

$$\begin{aligned}\frac{\partial \operatorname{vec} \boldsymbol{G}\boldsymbol{G}'}{\partial \boldsymbol{\mu}'} &= (\boldsymbol{I}_{p^2} + \boldsymbol{I}_{(p,p)})(\boldsymbol{I}_p \otimes \boldsymbol{G}')\frac{\partial \operatorname{vec} \boldsymbol{G}}{\partial \boldsymbol{\mu}'} = \boldsymbol{0} \quad \text{and} \\ \left.\frac{\partial \operatorname{vec} \boldsymbol{G}\boldsymbol{G}'}{\partial \boldsymbol{\mu}'}\right|_{\mu=0} &= 2\,\boldsymbol{N}_p \left.\frac{\partial \operatorname{vec} \boldsymbol{G}}{\partial \boldsymbol{\mu}'}\right|_{\mu=0} = \boldsymbol{0},\end{aligned}$$

where

$$\boldsymbol{N}_p = \frac{1}{2}(\boldsymbol{I}_{p^2} + \boldsymbol{I}_{(p,p)}). \tag{2.47}$$

Equivalently,

$$\boldsymbol{D}_p' \left.\frac{\partial \operatorname{vec} \boldsymbol{G}}{\partial \boldsymbol{\mu}'}\right|_{\mu=0} = \boldsymbol{0}, \tag{2.48}$$

because $\boldsymbol{N}_p = \boldsymbol{D}_p(\boldsymbol{D}_p\boldsymbol{D}_p')^{-1}\boldsymbol{D}_p'$.

Simplification yields

$$\boldsymbol{D}_p'\boldsymbol{A}_1\boldsymbol{V}_3 + \boldsymbol{D}_p'\boldsymbol{A}_1\boldsymbol{V}_4\frac{\partial \boldsymbol{\eta}_1}{\partial \boldsymbol{\mu}'} + \boldsymbol{D}_p'\boldsymbol{A}_2\frac{\partial \boldsymbol{\eta}_2}{\partial \boldsymbol{\mu}'} = \boldsymbol{0}$$

$$\Longrightarrow \frac{\partial \boldsymbol{\eta}_2}{\partial \boldsymbol{\mu}'} = -[\boldsymbol{D}_p' \boldsymbol{A}_1 \boldsymbol{V}_3 + \boldsymbol{D}_p' \boldsymbol{A}_1 \boldsymbol{V}_4 \frac{\partial \boldsymbol{\eta}_1}{\partial \boldsymbol{\mu}'}], \tag{2.49}$$

because of the result in (2.36).

Taking the derivative of the second constraint, $\boldsymbol{C} \operatorname{vec} \boldsymbol{G} = \boldsymbol{0}$, yields

$$\begin{aligned} \boldsymbol{C} \frac{\partial \operatorname{vec} \boldsymbol{G}}{\partial \boldsymbol{\mu}'} &= \boldsymbol{0} \\ \Longrightarrow \boldsymbol{C}\boldsymbol{A}_1 \boldsymbol{V}_3 + \boldsymbol{C}\boldsymbol{A}_1 \boldsymbol{V}_4 \frac{\partial \boldsymbol{\eta}_1}{\partial \boldsymbol{\mu}'} + \boldsymbol{C}\boldsymbol{A}_2 \frac{\partial \boldsymbol{\eta}_2}{\partial \boldsymbol{\mu}'} &= \boldsymbol{0}. \end{aligned} \tag{2.50}$$

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

$$\begin{aligned} \boldsymbol{C}(\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)})\boldsymbol{A}_1 \boldsymbol{V}_4 \frac{\partial \boldsymbol{\eta}_1}{\partial \boldsymbol{\mu}'} &= -\boldsymbol{C}(\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)})\boldsymbol{A}_1 \boldsymbol{V}_3 \\ \Longrightarrow \boldsymbol{W}_4 \boldsymbol{V}_4 \frac{\partial \boldsymbol{\eta}_1}{\partial \boldsymbol{\mu}'} &= -\boldsymbol{W}_4 \boldsymbol{V}_3 \\ \Longrightarrow \frac{\partial \boldsymbol{\eta}_1}{\partial \boldsymbol{\mu}'} &= \boldsymbol{0}, \end{aligned}$$

because $\boldsymbol{W}_4 = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}_4'$ and $\boldsymbol{V}_4'\boldsymbol{V}_3 = \boldsymbol{0}$. Then, it follows from (2.49) that

$$\frac{\partial \boldsymbol{\eta}_2}{\partial \boldsymbol{\mu}'} = -\boldsymbol{D}_p' \boldsymbol{A}_1 \boldsymbol{V}_3.$$

Expressions for the first, second, and third derivatives of $\operatorname{vec} \boldsymbol{G}$ in (2.33) with respect to $\boldsymbol{\mu}'$ 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* $\operatorname{vec} \boldsymbol{G}$ *in (2.33) with respect to* $\boldsymbol{\mu}'$*, evaluated at* $\boldsymbol{\mu} = \mathbf{0}$*, can be written as follows:*

$$
\begin{aligned}
\boldsymbol{D}_G^{(1)} &= \left.\frac{\partial \operatorname{vec} \boldsymbol{G}}{\partial \boldsymbol{\mu}'}\right|_{\mu=0} = (\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)})\boldsymbol{A}_1\boldsymbol{V}_3, \\
\boldsymbol{D}_G^{(2)} &= \left.\frac{\partial^2 \operatorname{vec} \boldsymbol{G}}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}'}\right|_{\mu=0} = (\boldsymbol{I}_{p^2} - \boldsymbol{P})\boldsymbol{A}_2\boldsymbol{D}_p'(\boldsymbol{I}_p \otimes \operatorname{vec} \boldsymbol{I}_p \otimes \boldsymbol{I}_p)'\left(\boldsymbol{D}_G^{(1)} \otimes \boldsymbol{D}_G^{(1)}\right), \\
\boldsymbol{D}_G^{(3)} &= \left.\frac{\partial^3 \operatorname{vec} \boldsymbol{G}}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}'}\right|_{\mu=0} = (\boldsymbol{P} - \boldsymbol{I}_{p^2})\boldsymbol{A}_2\boldsymbol{D}_p'(\boldsymbol{I}_p \otimes \operatorname{vec} \boldsymbol{I}_p \otimes \boldsymbol{I}_p)' \\
&\quad \times\left[\left(\boldsymbol{D}_G^{(1)} \otimes \boldsymbol{I}_{(p,p)}\boldsymbol{D}_G^{(2)}\right) - \left(\boldsymbol{D}_G^{(2)} \otimes \boldsymbol{D}_G^{(1)}\right)\{\boldsymbol{I}_{\nu_1^3} + (\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{I}_{\nu_1,\nu_1})\}\right], \\
\textit{where } \boldsymbol{P} &= \boldsymbol{K}_1(\boldsymbol{C}\boldsymbol{K}_1)^{-1}\boldsymbol{C}, \quad \boldsymbol{K}_1 = (\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)})\boldsymbol{A}_1\boldsymbol{V}_4, \\
\boldsymbol{C}\boldsymbol{K}_1 &= \boldsymbol{W}_4\boldsymbol{V}_4 = \boldsymbol{U}\boldsymbol{D}, \textit{ and } \boldsymbol{C} \textit{ is defined in Theorem } 5. \qquad (2.51)
\end{aligned}
$$

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

$$
\begin{aligned}
\boldsymbol{D}_G^{(1)} &= \boldsymbol{A}_1\boldsymbol{V}_3 + \boldsymbol{A}_1\boldsymbol{V}_4\boldsymbol{\eta}_1^{(1)} + \boldsymbol{A}_2\boldsymbol{\eta}_2^{(1)} \\
&= \boldsymbol{A}_1\boldsymbol{V}_3 - \boldsymbol{A}_2\boldsymbol{D}_p'\boldsymbol{A}_1\boldsymbol{V}_3 \text{ by (2.45) and (2.46)} \\
&= (\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)})\boldsymbol{A}_1\boldsymbol{V}_3 \text{ by (2.37).}
\end{aligned}
$$

The second derivative of $\operatorname{vec} \boldsymbol{G}$ *can be written as*

$$
\boldsymbol{D}_G^{(2)} = \boldsymbol{A}_1\boldsymbol{V}_4\boldsymbol{\eta}_1^{(2)} + \boldsymbol{A}_2\boldsymbol{\eta}_2^{(2)}. \qquad (2.52)
$$

To compute $\boldsymbol{\eta}_1^{(2)}$ *and* $\boldsymbol{\eta}_2^{(2)}$*, first examine the second derivative of* $\operatorname{vec} \boldsymbol{G}'\boldsymbol{G}$*:*

$$
\begin{aligned}
\frac{\partial^2 \operatorname{vec} \boldsymbol{G}'\boldsymbol{G}}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}'} &= \frac{\partial}{\partial \boldsymbol{\mu}'}\left\{2\boldsymbol{N}_p(\boldsymbol{I}_p \otimes \boldsymbol{G}')\frac{\partial \operatorname{vec} \boldsymbol{G}}{\partial \boldsymbol{\mu}'}\right\} = \mathbf{0} \\
&= 2\,\boldsymbol{N}_p\left\{\frac{\partial(\boldsymbol{I}_p \otimes \boldsymbol{G}')}{\partial \boldsymbol{\mu}'}\left(\boldsymbol{I}_{\nu_2} \otimes \frac{\partial \operatorname{vec} \boldsymbol{G}}{\partial \boldsymbol{\mu}'}\right) + (\boldsymbol{I}_p \otimes \boldsymbol{G}')\frac{\partial^2 \operatorname{vec} \boldsymbol{G}}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}'}\right\} = \mathbf{0}.
\end{aligned}
$$

Using the properties,

$$(\boldsymbol{X} \otimes \boldsymbol{Y})\boldsymbol{Z} = \boldsymbol{I}_{(r,p)}\,(\boldsymbol{I}_r \otimes (\operatorname{vec}\boldsymbol{I}_s)' \otimes \boldsymbol{X})\,(\operatorname{vec}\boldsymbol{Y}' \otimes \boldsymbol{I}_{(q,s)}\boldsymbol{Z}),$$

where $\boldsymbol{X}$ *is* $p \times q$ *and* $\boldsymbol{Y}$ *is* $r \times s$, *and* $\boldsymbol{N}_p = \boldsymbol{N}_p\,\boldsymbol{I}_{(p,p)}$, *yields*

$$\begin{aligned}
\left.\frac{\partial^2 \operatorname{vec}\boldsymbol{G}'\boldsymbol{G}}{\partial\boldsymbol{\mu}' \otimes \partial\boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}} &= \boldsymbol{N}_p\left\{(\boldsymbol{I}_p \otimes (\operatorname{vec}\boldsymbol{I}_p)' \otimes \boldsymbol{I}_p)\left(\boldsymbol{D}_G^{(1)} \otimes \boldsymbol{I}_{(p,p)}\,\boldsymbol{D}_G^{(1)}\right) + \boldsymbol{D}_G^{(2)}\right\} = \boldsymbol{0} \\
&= \boldsymbol{D}_p'\left\{-(\boldsymbol{I}_p \otimes (\operatorname{vec}\boldsymbol{I}_p)' \otimes \boldsymbol{I}_p)\left(\boldsymbol{D}_G^{(1)} \otimes \boldsymbol{D}_G^{(1)}\right) + \boldsymbol{D}_G^{(2)}\right\} = \boldsymbol{0} \\
&= \boldsymbol{D}_p'\left\{-\boldsymbol{K} + \boldsymbol{D}_G^{(2)}\right\} = \boldsymbol{0}, \qquad (2.53)
\end{aligned}$$

where $\boldsymbol{K} = (\boldsymbol{I}_p \otimes \operatorname{vec}\boldsymbol{I}_p \otimes \boldsymbol{I}_p)'\left(\boldsymbol{D}_G^{(1)} \otimes \boldsymbol{D}_G^{(1)}\right)$ *and* $\boldsymbol{I}_{(p,p)}\,\boldsymbol{D}_G^{(1)} = -\,\boldsymbol{D}_G^{(1)}$.

Taking the second derivative of the second constraint, $\boldsymbol{C}\operatorname{vec}\boldsymbol{G} = \boldsymbol{0}$, *yields*

$$\begin{aligned}
\boldsymbol{C}\left.\frac{\partial^2 \operatorname{vec}\boldsymbol{G}}{\partial\boldsymbol{\mu}' \otimes \partial\boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}} &= \boldsymbol{0} \\
\boldsymbol{C}\,\boldsymbol{D}_G^{(2)} &= \boldsymbol{0}. \qquad (2.54)
\end{aligned}$$

Substituting $\boldsymbol{D}_G^{(2)}$ *from (2.52) in (2.53) yields*

$$\boldsymbol{D}_p'\left\{-\boldsymbol{K} + \boldsymbol{A}_1\boldsymbol{V}_4\boldsymbol{\eta}_1^{(2)} + \boldsymbol{A}_2\boldsymbol{\eta}_2^{(2)}\right\} = \boldsymbol{0},$$

then

$$\boldsymbol{\eta}_2^{(2)} = \boldsymbol{D}_p'\boldsymbol{K} - \boldsymbol{D}_p'\boldsymbol{A}_1\boldsymbol{V}_4\boldsymbol{\eta}_1^{(2)}, \qquad (2.55)$$

by using result (2.37). Substituting $\boldsymbol{D}_G^{(2)}$ *from (2.52) in (2.54) yields*

$$\boldsymbol{C}\left[\boldsymbol{A}_1\boldsymbol{V}_4\boldsymbol{\eta}_1^{(2)} + \boldsymbol{A}_2\boldsymbol{\eta}_2^{(2)}\right] = \boldsymbol{0}.$$

Substituting $\boldsymbol{\eta}_2^{(2)}$ *from (2.55) yields*

$$\begin{aligned}
\boldsymbol{C}(\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)})\boldsymbol{A}_1\boldsymbol{V}_4\boldsymbol{\eta}_1^{(2)} + \boldsymbol{C}\boldsymbol{A}_2\boldsymbol{D}_p'\boldsymbol{K} &= \boldsymbol{0} \\
\Longrightarrow \boldsymbol{W}_4\boldsymbol{V}_4\boldsymbol{\eta}_1^{(2)} + \boldsymbol{C}\boldsymbol{A}_2\boldsymbol{D}_p'\boldsymbol{K} &= \boldsymbol{0}.
\end{aligned}$$

Replacing $\boldsymbol{W}_4 = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}_4'$ *yields*

$$\boldsymbol{\eta}_1^{(2)} = -(\boldsymbol{U}\boldsymbol{D})^{-1}\boldsymbol{C}\boldsymbol{A}_2\boldsymbol{D}_p'\boldsymbol{K}.$$

Then, it follows from (2.55) that

$$\boldsymbol{\eta}_2^{(2)} = \boldsymbol{D}_p'\boldsymbol{K} + \boldsymbol{D}_p'\boldsymbol{A}_1\boldsymbol{V}_4(\boldsymbol{U}\boldsymbol{D})^{-1}\boldsymbol{C}\boldsymbol{A}_2\boldsymbol{D}_p'\boldsymbol{K}.$$

Consequently,

$$\begin{aligned}\boldsymbol{D}_G^{(2)} &= \boldsymbol{A}_2\boldsymbol{D}_p'\boldsymbol{K} - \boldsymbol{P}\boldsymbol{A}_2\boldsymbol{D}_p'\boldsymbol{K} \\ &= (\boldsymbol{I}_{p^2} - \boldsymbol{P})\boldsymbol{A}_2\boldsymbol{D}_p'(\boldsymbol{I}_p \otimes \operatorname{vec}\boldsymbol{I}_p \otimes \boldsymbol{I}_p)'\left(\boldsymbol{D}_G^{(1)} \otimes \boldsymbol{D}_G^{(1)}\right).\end{aligned}$$

Similarly, $\boldsymbol{D}_G^{(3)}$ *can be obtained. It follows from (2.33) that*

$$\boldsymbol{D}_G^{(3)} = \boldsymbol{A}_1\boldsymbol{V}_4\boldsymbol{\eta}_1^{(3)} + \boldsymbol{A}_2\boldsymbol{\eta}_2^{(3)}. \tag{2.56}$$

To compute $\boldsymbol{\eta}_1^{(3)}$ *and* $\boldsymbol{\eta}_2^{(3)}$*, the third derivative of* $\operatorname{vec}\boldsymbol{G}'\boldsymbol{G}$ *with respect to* $\boldsymbol{\mu}'$ *is obtained as follows:*

$$\begin{aligned}\left.\frac{\partial^3 \operatorname{vec}\boldsymbol{G}'\boldsymbol{G}}{\partial\boldsymbol{\mu}' \otimes \partial\boldsymbol{\mu}' \otimes \partial\boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}} &= \boldsymbol{N}_p\left\{(\boldsymbol{I}_p \otimes (\operatorname{vec}\boldsymbol{I}_p)' \otimes \boldsymbol{I}_p)(\boldsymbol{I}_{p^2} \otimes \boldsymbol{I}_{(p,p)})\left.\frac{\partial\left(\boldsymbol{D}_G^{(1)} \otimes \boldsymbol{D}_G^{(1)}\right)}{\partial\boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}}\right. \\ &\quad \left. + \left.\frac{\partial(\boldsymbol{I}_p \otimes \boldsymbol{G}')\,\boldsymbol{D}_G^{(2)}}{\partial\boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}}\right\} = \boldsymbol{0}.\end{aligned} \tag{2.57}$$

For convenience, the two parts on the right-hand-side of (2.57) are labeled as $\boldsymbol{G}_a$ *and* $\boldsymbol{G}_b$*. Using the Kronecker product rule,* $\boldsymbol{G}_a$ *can be simplified to*

$$\begin{aligned}\boldsymbol{G}_a &= -\boldsymbol{N}_p(\boldsymbol{I}_p \otimes (\operatorname{vec}\boldsymbol{I}_p)' \otimes \boldsymbol{I}_p)(\boldsymbol{I}_{p^2} \otimes \boldsymbol{I}_{(p,p)})\Big[\left(\boldsymbol{D}_G^{(2)} \otimes \boldsymbol{D}_G^{(1)}\right) \\ &\quad + \boldsymbol{I}_{(p^2,p^2)}\left(\boldsymbol{D}_G^{(2)} \otimes \boldsymbol{D}_G^{(1)}\right)(\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{I}_{(\nu_1,\nu_1)})\Big].\end{aligned}$$

Express $(\operatorname{vec}\boldsymbol{I}_p)'$ *as* $\sum_{i=1}^{p}(\boldsymbol{e}_i^{p\prime}\otimes\boldsymbol{e}_i^{p\prime})$ *and repeatedly use the permutation matrix:*

$$(\boldsymbol{A}\otimes\boldsymbol{B})=\boldsymbol{I}_{(p,r)}(\boldsymbol{B}\otimes\boldsymbol{A})\boldsymbol{I}_{(c,q)},\ \textit{where } \boldsymbol{A} \textit{ is } r\times c \textit{ and } \boldsymbol{B} \textit{ is } p\times q,$$

to simplify $\boldsymbol{G}_a$ *as follows:*

$$\boldsymbol{G}_a=-\boldsymbol{N}_p\left(\boldsymbol{I}_p\otimes(\operatorname{vec}\boldsymbol{I}_p)'\otimes\boldsymbol{I}_p\right)\left(\boldsymbol{D}_G^{(2)}\otimes\boldsymbol{D}_G^{(1)}\right)\left[\boldsymbol{I}_{\nu_1^3}+(\boldsymbol{I}_{\nu_1}\otimes\boldsymbol{I}_{(\nu_1,\nu_1)})\right].$$

Similarly, $\boldsymbol{G}_b$ *can be simplified to*

$$\boldsymbol{G}_b=\boldsymbol{N}_p\left\{\left(\boldsymbol{I}_p\otimes(\operatorname{vec}\boldsymbol{I}_p)'\otimes\boldsymbol{I}_p\right)\left(\boldsymbol{D}_G^{(1)}\otimes\boldsymbol{I}_{(p,p)}\boldsymbol{D}_G^{(2)}\right)+\boldsymbol{D}_G^{(3)}\right\}.$$

Combining $\boldsymbol{G}_a$ *and* $\boldsymbol{G}_b$ *yields*

$$\begin{aligned}\left.\frac{\partial^3\operatorname{vec}\boldsymbol{G}'\boldsymbol{G}}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}} &= \boldsymbol{N}_p\Big\{\left(\boldsymbol{I}_p\otimes(\operatorname{vec}\boldsymbol{I}_p)'\otimes\boldsymbol{I}_p\right)\Big[-\left(\boldsymbol{D}_G^{(2)}\otimes\boldsymbol{D}_G^{(1)}\right)\\ &\quad\left[\boldsymbol{I}_{\nu_1^3}+(\boldsymbol{I}_{\nu_1}\otimes\boldsymbol{I}_{(\nu_1,\nu_1)})\right]+\left(\boldsymbol{D}_G^{(1)}\otimes\boldsymbol{I}_{(p,p)}\boldsymbol{D}_G^{(2)}\right)\Big]+\boldsymbol{D}_G^{(3)}\Big\}\\ &= \boldsymbol{D}_p'\left\{\boldsymbol{K}^\star+\boldsymbol{D}_G^{(3)}\right\}=\boldsymbol{0}, \end{aligned}\tag{2.58}$$

where

$$\begin{aligned}\boldsymbol{K}^\star &= (\boldsymbol{I}_p\otimes\operatorname{vec}\boldsymbol{I}_p\otimes\boldsymbol{I}_p)'\Big\{\left(\boldsymbol{D}_G^{(1)}\otimes\boldsymbol{I}_{(p,p)}\boldsymbol{D}_G^{(2)}\right)\\ &\quad-\left(\boldsymbol{D}_G^{(2)}\otimes\boldsymbol{D}_G^{(1)}\right)\left[\boldsymbol{I}_{\nu_1^3}+(\boldsymbol{I}_{\nu_1}\otimes\boldsymbol{I}_{(\nu_1,\nu_1)})\right]\Big\}.\end{aligned}$$

It follows from (2.54) that

$$\boldsymbol{C}\,\boldsymbol{D}_G^{(3)}=\boldsymbol{0}.\tag{2.59}$$

Replacing $\boldsymbol{D}_G^{(3)}$ *from (2.56) in (2.58) yields*

$$\boldsymbol{D}_p'\left\{\boldsymbol{K}^\star+\boldsymbol{A}_1\boldsymbol{V}_4\boldsymbol{\eta}_1^{(3)}+\boldsymbol{A}_2\boldsymbol{\eta}_2^{(3)}\right\}=\boldsymbol{0}.$$

Therefore,

$$\boldsymbol{\eta}_2^{(3)} = -\boldsymbol{D}_p'\boldsymbol{K}^\star - \boldsymbol{D}_p'\boldsymbol{A}_1\boldsymbol{V}_4\boldsymbol{\eta}_1^{(3)}. \tag{2.60}$$

Substituting $\boldsymbol{D}_G^{(3)}$ *from (2.56) in (2.59) yields*

$$\boldsymbol{C}\left[\boldsymbol{A}_1\boldsymbol{V}_4\boldsymbol{\eta}_1^{(3)} + \boldsymbol{A}_2\boldsymbol{\eta}_2^{(3)}\right] = \mathbf{0}.$$

Substituting $\boldsymbol{\eta}_2^{(3)}$ *from (2.60) yields*

$$\begin{aligned} \boldsymbol{C}(\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)})\boldsymbol{A}_1\boldsymbol{V}_4\boldsymbol{\eta}_1^{(3)} - \boldsymbol{C}\boldsymbol{A}_2\boldsymbol{D}_p'\boldsymbol{K}^\star &= \mathbf{0} \\ \Longrightarrow \boldsymbol{W}_4\boldsymbol{V}_4\boldsymbol{\eta}_1^{(3)} - \boldsymbol{C}\boldsymbol{A}_2\boldsymbol{D}_p'\boldsymbol{K}^\star &= \mathbf{0}. \end{aligned}$$

Replacing $\boldsymbol{W}_4 = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}_4'$ *yields*

$$\boldsymbol{\eta}_1^{(3)} = (\boldsymbol{U}\boldsymbol{D})^{-1}\boldsymbol{C}\boldsymbol{A}_2\boldsymbol{D}_p'\boldsymbol{K}^\star.$$

It follows from (2.60) that

$$\boldsymbol{\eta}_2^{(3)} = -\boldsymbol{D}_p'\boldsymbol{K}^\star - \boldsymbol{D}_p'\boldsymbol{A}_1\boldsymbol{V}_4(\boldsymbol{U}\boldsymbol{D})^{-1}\boldsymbol{C}\boldsymbol{A}_2\boldsymbol{D}_p'\boldsymbol{K}^\star.$$

Finally, replacing $\boldsymbol{\eta}_1^{(3)}$ *and* $\boldsymbol{\eta}_2^{(3)}$ *into (2.56) yields*

$$\begin{aligned} \boldsymbol{D}_G^{(3)} &= (\boldsymbol{P} - \boldsymbol{I}_{p^2})\boldsymbol{A}_2\boldsymbol{D}_p'\boldsymbol{K}^\star \\ &= (\boldsymbol{P} - \boldsymbol{I}_{p^2})\boldsymbol{A}_2\boldsymbol{D}_p'(\boldsymbol{I}_p \otimes \operatorname{vec}\boldsymbol{I}_p \otimes \boldsymbol{I}_p)'\Big\{\left(\boldsymbol{D}_G^{(1)} \otimes \boldsymbol{I}_{(p,p)}\,\boldsymbol{D}_G^{(2)}\right) \\ &\quad - \left(\boldsymbol{D}_G^{(2)} \otimes \boldsymbol{D}_G^{(1)}\right)\left[\boldsymbol{I}_{\nu_1^3} + (\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{I}_{(\nu_1,\nu_1)})\right]\Big\}. \end{aligned}$$

□

Corollary 2. The first three derivatives of $\operatorname{vec}\boldsymbol{G}$ *in (2.33) with respect to* $\boldsymbol{\mu}'$*, evaluated at* $\boldsymbol{\mu} = \mathbf{0}$*, under constraint* $\boldsymbol{A} \in \mathcal{R}(\boldsymbol{\Gamma M})$*, where* $\mathcal{R}(\boldsymbol{M}) = \mathcal{R}\begin{pmatrix} \boldsymbol{I}_m \\ \mathbf{0} \end{pmatrix}$ *can be written as follows:*

$$\begin{aligned} \boldsymbol{D}_G^{(1)} &= \left.\frac{\partial \operatorname{vec}\boldsymbol{G}}{\partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = (\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)})\boldsymbol{A}_1\boldsymbol{V}_3, \\ \boldsymbol{D}_G^{(2)} &= \left.\frac{\partial^2 \operatorname{vec}\boldsymbol{G}}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \boldsymbol{A}_2\boldsymbol{D}_p' \left(\boldsymbol{I}_p \otimes \operatorname{vec}\boldsymbol{I}_p \otimes \boldsymbol{I}_p\right)' \left(\boldsymbol{D}_G^{(1)} \otimes \boldsymbol{D}_G^{(1)}\right), \end{aligned}$$

$$\begin{aligned} \boldsymbol{D}_G^{(3)} &= \left.\frac{\partial^3 \operatorname{vec}\boldsymbol{G}}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = -\boldsymbol{A}_2\boldsymbol{D}_p'(\boldsymbol{I}_p \otimes \operatorname{vec}\boldsymbol{I}_p \otimes \boldsymbol{I}_p)' \Big[\left(\boldsymbol{D}_G^{(1)} \otimes \boldsymbol{I}_{(p,p)}\boldsymbol{D}_G^{(2)}\right) \\ &\quad - \left(\boldsymbol{D}_G^{(2)} \otimes \boldsymbol{D}_G^{(1)}\right) \{\boldsymbol{I}_{\nu_1^3} + (\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{I}_{\nu_1,\nu_1})\} \Big]. \end{aligned}$$

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

$$\begin{aligned} \boldsymbol{\eta}_1^{(2)} &= \mathbf{0},\ \boldsymbol{\eta}_2^{(2)} = \boldsymbol{D}_p'\boldsymbol{K}, \\ \boldsymbol{\eta}_1^{(3)} &= \mathbf{0},\ \text{and } \boldsymbol{\eta}_2^{(3)} = -\boldsymbol{D}_p'\boldsymbol{K}^*. \end{aligned}$$

□

Under constraint $\boldsymbol{A} \in \mathcal{R}(\boldsymbol{\Gamma M})$, where $\mathcal{R}(\boldsymbol{M}) = \mathcal{R}\begin{pmatrix} \boldsymbol{I}_m \\ \mathbf{0} \end{pmatrix}$, the implicit parameter $\boldsymbol{\eta}$ is restricted that $\boldsymbol{\eta}_1 = \mathbf{0}$. This condition reduces form of the first three derivatives of $\operatorname{vec}\boldsymbol{G}$ in (2.33) with respect to $\boldsymbol{\mu}'$ as shown in Corollary 2. The proof is shown in the following Theorem.

Theorem 8. *For fixed $\boldsymbol{V}^*$ from Theorem 6 and under constraint $\boldsymbol{A} \in \mathcal{R}(\boldsymbol{\Gamma M})$ yielding $\boldsymbol{C}\,\mathrm{vec}\,\boldsymbol{G} = \boldsymbol{0}$, where $\mathcal{R}(\boldsymbol{M}) = \mathcal{R}\begin{pmatrix} \boldsymbol{I}_m \\ \boldsymbol{0} \end{pmatrix}$, then*

$$\boldsymbol{\eta}_1 = \boldsymbol{0},$$

where $\boldsymbol{\eta}_1$ has dimension $(p-m)q \times 1$.

Proof. *Note that*

$$\mathrm{vec}\,\boldsymbol{G} = \boldsymbol{A}_1\boldsymbol{V}_3\boldsymbol{\mu} + \boldsymbol{A}_1\boldsymbol{V}_4\boldsymbol{\eta}_1 + \boldsymbol{A}_2\boldsymbol{\eta}_2,$$

where $\boldsymbol{\mu} = \boldsymbol{V}_3'\boldsymbol{\mu}^*$, $\boldsymbol{\eta}_1 = \boldsymbol{V}_4'\boldsymbol{\mu}^*$, and $\boldsymbol{\eta}_2 = \boldsymbol{\eta}^*$.

It follows from (2.39) that

$$\begin{aligned} \boldsymbol{C}(\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)})\boldsymbol{A}_1 &= \boldsymbol{UDV}_4' \\ \Longrightarrow \boldsymbol{CA}_1 &= \boldsymbol{UDV}_4' + \boldsymbol{C}\,\boldsymbol{I}_{(p,p)}\,\boldsymbol{A}_1. \end{aligned} \tag{2.61}$$

Since,

$$\boldsymbol{C}\,\mathrm{vec}\,\boldsymbol{G} = \boldsymbol{CA}_1\boldsymbol{\mu}^* + \boldsymbol{CA}_2\boldsymbol{\eta}^* = \boldsymbol{0},$$

substituting (2.61) in yields

$$\boldsymbol{UDV}_4'\boldsymbol{\mu}^* + \boldsymbol{C}\,\boldsymbol{I}_{(p,p)}\,\boldsymbol{A}_1\boldsymbol{\mu}^* + \boldsymbol{CA}_2\boldsymbol{\eta}^* = \boldsymbol{0}.$$

Therefore,

$$\begin{aligned} \boldsymbol{UD}\boldsymbol{\eta}_1 &= -\boldsymbol{C}(\boldsymbol{I}_{(p,p)}\,\boldsymbol{A}_1\boldsymbol{\mu}^* + \boldsymbol{A}_2\boldsymbol{\eta}^*) \\ &= -\boldsymbol{CA}_2\left(\boldsymbol{D}_p'\boldsymbol{A}_1\boldsymbol{\mu}^* + \boldsymbol{\eta}^*\right). \end{aligned} \tag{2.62}$$

It will now be shown that $\boldsymbol{C}\boldsymbol{A}_2 = (\boldsymbol{M}^{c\prime} \otimes \boldsymbol{A}'\Gamma)\boldsymbol{A}_2 = \boldsymbol{0}$. *It follows from (2.42) that*

$$(\boldsymbol{M}^{c\prime} \otimes \boldsymbol{A}'\Gamma)\boldsymbol{A}_2 = \Big([0 \quad \boldsymbol{R}'] \otimes [\boldsymbol{T}' \quad 0]\Big)\boldsymbol{A}_2,$$

where $\boldsymbol{R}$ *is any* $(p-m)\times(p-m)$ *orthogonal matrix. Expressing* $\boldsymbol{A}_2$ *as* $\sum_{j=1}^{p}\sum_{i=j}^{p}(\boldsymbol{e}_j^p \otimes \boldsymbol{e}_i^p)\boldsymbol{e}_l^{w\prime}$, *where* $i \geq j$, $w = p(p+1)/2$, *and* $l = \frac{(j-1)(2p-j)}{2} + i$ *yields*

$$(\boldsymbol{M}^{c\prime} \otimes \boldsymbol{A}'\Gamma)\boldsymbol{A}_2 \;=\; \sum_{j=1}^{p}\sum_{i=j}^{p}\Big([0 \quad \boldsymbol{R}'] \otimes [\boldsymbol{T}' \quad 0]\Big)(\boldsymbol{e}_j^p \otimes \boldsymbol{e}_i^p)\boldsymbol{e}_l^{w\prime}.$$

Accordingly, using (2.43) and (2.44) yields

$$(\boldsymbol{M}^{c\prime} \otimes \boldsymbol{A}'\Gamma)\boldsymbol{A}_2 = \sum_{j=1}^{p}\sum_{i=j}^{p}\operatorname{vec}\left([\boldsymbol{T}' \quad 0](\boldsymbol{e}_i^p\boldsymbol{e}_j^{p\prime})\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{R}\end{pmatrix}\right)\boldsymbol{e}_l^{w\prime},$$

where $i \geq j$.

Choosing $j \leq i \leq m$ *yields*

$$\boldsymbol{e}_j^{p\prime}\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{R}\end{pmatrix} = 0.$$

Choosing $i \geq j \geq m$ *yields*

$$[\boldsymbol{T}' \quad 0]\boldsymbol{e}_i^p = \boldsymbol{0}.$$

Consequently,

$$(\boldsymbol{M}^{c\prime} \otimes \boldsymbol{A}'\Gamma)\boldsymbol{A}_2 = \boldsymbol{C}\boldsymbol{A}_2 = \boldsymbol{0}.$$

By replacing in (2.62), the result is

$$\begin{aligned}\boldsymbol{U}\boldsymbol{D}\boldsymbol{\eta}_1 \;&=\; \boldsymbol{0}\\ \Longrightarrow \boldsymbol{\eta}_1 \;&=\; \boldsymbol{0}.\end{aligned}$$

□

Solving for Implicit Parameters

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

$$\begin{aligned} \ell(\hat{\boldsymbol{\theta}}) &= \ell(\hat{\boldsymbol{\theta}}_i) + (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_i)' \left[\left. \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i} \right] \\ &\quad + \frac{1}{2}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_i)' \left[\left. \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \otimes \partial \boldsymbol{\theta}'} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i} \right] (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_i) + o(\|\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_i\|^2) \\ &\approx \ell(\hat{\boldsymbol{\theta}}_i) + (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_i)' \boldsymbol{g}_{\hat{\boldsymbol{\theta}}_i} + \frac{1}{2}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_i)' \boldsymbol{H}_{\hat{\boldsymbol{\theta}}_i} (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_i), \end{aligned} \tag{2.63}$$

where

$$\boldsymbol{g}_{\hat{\boldsymbol{\theta}}_i} = \left. \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i} \quad \text{and} \quad \boldsymbol{H}_{\hat{\boldsymbol{\theta}}_i} = \left. \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \otimes \partial \boldsymbol{\theta}'} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i}.$$

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

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_i - \boldsymbol{H}_{\hat{\boldsymbol{\theta}}_i}^{-1} \boldsymbol{g}_{\hat{\boldsymbol{\theta}}_i}. \tag{2.64}$$

Solving for $\boldsymbol{\tau}$

In reduction in the number of components, $\boldsymbol{T}_4\boldsymbol{\xi}$ can be written as (2.11), where $\boldsymbol{\tau}$ is an implicit function of $\boldsymbol{\varphi}$. For given $\boldsymbol{\varphi}$, to solve for $\boldsymbol{\tau}$ in (2.11) subject to constraint (2.10), $\boldsymbol{C}_2' \exp\{\odot(\boldsymbol{T}_4\boldsymbol{\xi})\} = \boldsymbol{0}$, the equation (2.64) is employed as follows:

$$\hat{\boldsymbol{\tau}} = \hat{\boldsymbol{\tau}}_i - \boldsymbol{H}_{\hat{\boldsymbol{\tau}}_i}^{-1} \boldsymbol{g}_{\hat{\boldsymbol{\tau}}_i}, \tag{2.65}$$

where

$$\begin{aligned} \boldsymbol{g}_{\hat{\boldsymbol{\tau}}_i} &= \boldsymbol{C}_2' \exp\{\odot(\boldsymbol{T}_4\boldsymbol{\xi}^{(i)})\} \quad \text{and} \\ \boldsymbol{H}_{\hat{\boldsymbol{\tau}}_i} &= \boldsymbol{C}_2' \operatorname{diag}\left(\exp\{\odot(\boldsymbol{T}_4\boldsymbol{\xi}^{(i)})\}\right) \boldsymbol{T}_4 \boldsymbol{V}_1^{(i)}. \end{aligned}$$

Solving for $\boldsymbol{\eta}$

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

$$\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\eta}}_i - \boldsymbol{H}_{\hat{\boldsymbol{\eta}}_i}^{-1} \boldsymbol{g}_{\hat{\boldsymbol{\eta}}_i},$$

where $\boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \end{pmatrix}$,

$$\begin{aligned} \boldsymbol{g}_{\hat{\boldsymbol{\eta}}_i} &= \begin{pmatrix} \boldsymbol{D}_p' \operatorname{vec}(\boldsymbol{G}^{(i)}\boldsymbol{G}^{(i)\prime} - \boldsymbol{I}_p) \\ \boldsymbol{C}\operatorname{vec}\boldsymbol{G}^{(i)} \end{pmatrix}, \quad \text{and} \\ \boldsymbol{H}_{\hat{\boldsymbol{\eta}}_i} &= \begin{pmatrix} 2\boldsymbol{D}_p'(\boldsymbol{G}^{(i)} \otimes \boldsymbol{I}_p)[\boldsymbol{A}_1\boldsymbol{V}_4 \quad \boldsymbol{A}_2] \\ \boldsymbol{C}[\boldsymbol{A}_1\boldsymbol{V}_4 \quad \boldsymbol{A}_2] \end{pmatrix}. \end{aligned}$$

In the special case described in Theorem 8, $\operatorname{vec}\boldsymbol{G}$ can be reduced to

$$\operatorname{vec}\boldsymbol{G} = \boldsymbol{A}_1\boldsymbol{V}_3\boldsymbol{\mu} + \boldsymbol{A}_2\boldsymbol{\eta}_2. \tag{2.66}$$

To solve for $\boldsymbol{\eta}_2$, $\boldsymbol{g}_{\hat{\boldsymbol{\eta}}_i}$ and $\boldsymbol{H}_{\hat{\boldsymbol{\eta}}_i}$ can be simplified to

$$\begin{aligned} \boldsymbol{g}_{\hat{\boldsymbol{\eta}}_i} &= \boldsymbol{D}_p' \operatorname{vec}(\boldsymbol{G}^{(i)}\boldsymbol{G}^{(i)\prime} - \boldsymbol{I}_p) \quad \text{and} \\ \boldsymbol{H}_{\hat{\boldsymbol{\eta}}_i} &= 2\boldsymbol{D}_p'(\boldsymbol{G}^{(i)} \otimes \boldsymbol{I}_p)\boldsymbol{A}_2. \end{aligned}$$

CHAPTER 3

ESTIMATING PARAMETERS AND CONSTRUCTING ASYMPTOTIC DISTRIBUTIONS

In this chapter, the likelihood function will be shown to depend on the unknown parameter Σ 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 Σ 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 Σ under the model in (2.1). To find the maximum likelihood estimator of Σ, 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 $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$, then $\boldsymbol{\tau}(\hat{\boldsymbol{\theta}})$ is the maximum likelihood estimator of $\boldsymbol{\tau}(\boldsymbol{\theta})$, where $\boldsymbol{\tau}(\boldsymbol{\theta})$ is a function of $\boldsymbol{\theta}$. Using this property, the MLE of Σ is given by

$$\hat{\Sigma} = \hat{\Gamma}\hat{\Lambda}\hat{\Gamma}',$$

where $\hat{\mathbf{\Gamma}}$ and $\hat{\mathbf{\Lambda}}$ are the MLEs of $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$, respectively. Specific goals of this chapter are to develop procedures for estimating $\mathbf{\Gamma}$, $\mathbf{\Lambda}$, and $\mathbf{\Sigma}$ and to construct asymptotic distributions of $\hat{\mathbf{\Gamma}}$ and $\hat{\boldsymbol{\lambda}}$.

Loglikelihood Function

Assume that the sample covariance matrix, $\boldsymbol{S}$, follows a Wishart distribution with covariance $\mathbf{\Sigma}$ and n degrees of freedom. The distribution of $\boldsymbol{S}$ is denoted by

$$n\boldsymbol{S} \sim W_p(n, \mathbf{\Sigma}).$$

The loglikelihood function can be written as

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

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

$$\begin{aligned}
\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \theta_j} &= -\mathbf{\Sigma}^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \theta_j}\mathbf{\Sigma}^{-1}, \\
\frac{\partial \ln|\mathbf{\Sigma}|}{\partial \theta_j} &= \operatorname{tr}\left(\mathbf{\Sigma}^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \theta_j}\right), \\
\operatorname{tr}(\boldsymbol{AB}) &= (\operatorname{vec}\boldsymbol{A}')'\operatorname{vec}\boldsymbol{B},
\end{aligned}$$

the first derivative of $\ell(\boldsymbol{\theta})$ with respect to θ_j can be written as

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_j} &= \frac{n}{2}\operatorname{tr}\left(\boldsymbol{S}\mathbf{\Sigma}^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \theta_j}\mathbf{\Sigma}^{-1}\right) - \frac{n}{2}\operatorname{tr}\left(\mathbf{\Sigma}\mathbf{\Sigma}^{-1}\frac{\partial \mathbf{\Sigma}}{\partial \theta_j}\mathbf{\Sigma}^{-1}\right) \\
&= \frac{n}{2}\left\{\operatorname{vec}(\boldsymbol{S} - \mathbf{\Sigma})\right\}'\left(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1}\right)\frac{\partial \operatorname{vec}\mathbf{\Sigma}}{\partial \theta_j}.
\end{aligned}$$

Accordingly, the first derivative of $\ell(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}'$ can be written as

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} &= \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_1} \quad \frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_2} \cdots \frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_\nu}\right) \\ &= \frac{n}{2}\{\operatorname{vec}(\boldsymbol{S}-\boldsymbol{\Sigma})\}'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\theta}'}.\end{aligned}$$

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

$$\left.\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \frac{n}{2}\boldsymbol{F}^{(1)\prime}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}(\boldsymbol{S}-\boldsymbol{\Sigma}), \tag{3.2}$$

where

$$\boldsymbol{F}^{(1)} = \left.\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = \left(\left.\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} \quad \left.\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'}\right|_{\boldsymbol{\mu}=\mathbf{0}}\right). \tag{3.3}$$

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

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{G}\boldsymbol{\Lambda}\boldsymbol{G}'\boldsymbol{\Gamma}'.$$

Using property (2.32) yields

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

Therefore,

$$\left.\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}'}\right|_{\boldsymbol{\mu}=\mathbf{0}} = (\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\boldsymbol{L}\,\boldsymbol{D}_{\lambda}^{(1)},$$

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

$$\begin{aligned}\operatorname{vec}\boldsymbol{\Sigma} &= (\boldsymbol{\Gamma}\boldsymbol{G}\boldsymbol{\Lambda}\otimes\boldsymbol{\Gamma})\operatorname{vec}\boldsymbol{G} \quad \text{or} \\ \operatorname{vec}\boldsymbol{\Sigma} &= (\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma}\boldsymbol{G}\boldsymbol{\Lambda})\operatorname{vec}\boldsymbol{G}' = \boldsymbol{I}_{(p,p)}(\boldsymbol{\Gamma}\boldsymbol{G}\boldsymbol{\Lambda}\otimes\boldsymbol{\Gamma})\operatorname{vec}\boldsymbol{G}.\end{aligned}$$

Therefore,

$$\begin{aligned} \frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\mu}'} &= (\boldsymbol{I}_{p^2} + \boldsymbol{I}_{(p,p)})(\boldsymbol{\Gamma} \boldsymbol{G} \boldsymbol{\Lambda} \otimes \boldsymbol{\Gamma}) \frac{\partial \operatorname{vec} \boldsymbol{G}}{\partial \boldsymbol{\mu}'} \quad \text{and} \\ \left. \frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\mu}'} \right|_{\boldsymbol{\mu}=\mathbf{0}} &= 2\boldsymbol{N}_p(\boldsymbol{\Gamma} \boldsymbol{\Lambda} \otimes \boldsymbol{\Gamma})\, \boldsymbol{D}_G^{(1)}, \end{aligned} \tag{3.4}$$

where $\boldsymbol{N}_p = (\boldsymbol{I}_{p^2} + \boldsymbol{I}_{(p,p)})/2$ and $\boldsymbol{D}_G^{(1)} = (\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)})\boldsymbol{A}_1 \boldsymbol{V}_3$.

Boik (1998) showed that the maximum likelihood estimate of $\boldsymbol{\Sigma}$, $\hat{\boldsymbol{\Sigma}}$, can be obtained by solving

$$\left. \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\mu}=\mathbf{0}} = \mathbf{0}. \tag{3.5}$$

There is no simple, closed form solution for $\hat{\boldsymbol{\Sigma}}$. Therefore, the solution, $\hat{\boldsymbol{\Sigma}}$, 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 $\ell(\boldsymbol{\theta})$ the loglikelihood function of $\boldsymbol{\theta}$ given an observed $\boldsymbol{Y}$. The MLE of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$, is the vector that maximizes $\ell(\boldsymbol{\theta})$. Let $\hat{\boldsymbol{\theta}}_0$ be an initial guess and $\hat{\boldsymbol{\theta}}_i$ be the estimate of $\boldsymbol{\theta}$ after the i^{th} iteration and let $\boldsymbol{I}_{\hat{\boldsymbol{\theta}}_i}$ be Fisher's information at the value $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_i$. That is,

$$\boldsymbol{I}_{\hat{\boldsymbol{\theta}}_i} = -\boldsymbol{E}\left[\left. \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \otimes \partial \boldsymbol{\theta}'} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i} \right]. \tag{3.6}$$

Expanding $\ell(\hat{\boldsymbol{\theta}})$ in a Taylor series around $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_i$ yields

$$\ell(\hat{\boldsymbol{\theta}}) \approx \ell(\hat{\boldsymbol{\theta}}_i) + (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_i)' \boldsymbol{g}_{\hat{\boldsymbol{\theta}}_i} - \frac{1}{2}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_i)' \boldsymbol{I}_{\hat{\boldsymbol{\theta}}_i} (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_i). \tag{3.7}$$

Set the derivative of $\ell(\hat{\boldsymbol{\theta}})$, expressed as (3.7), to zero and solve for $\hat{\boldsymbol{\theta}}$. The result is

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_i + \boldsymbol{I}_{\hat{\boldsymbol{\theta}}_i}^{-1} \boldsymbol{g}_{\hat{\boldsymbol{\theta}}_i}. \tag{3.8}$$

where $\boldsymbol{g}_{\hat{\boldsymbol{\theta}}_i}$ 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 $\boldsymbol{\lambda}$ in (2.4), Boik (2002a)'s algorithm is adopted as follows:

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

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

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

Step 3 Set $\hat{\boldsymbol{\mu}}_i = \boldsymbol{0}$ then $\hat{\boldsymbol{\theta}}_i = (\boldsymbol{0}' \quad \hat{\boldsymbol{\varphi}}_i')'$. Use one iteration of the Fisher scoring algorithm to update $\hat{\boldsymbol{\theta}}_i$, that is,

$$\hat{\boldsymbol{\theta}}_{i+1} = \hat{\boldsymbol{\theta}}_i + \boldsymbol{I}_{\hat{\boldsymbol{\theta}}_i}^{-1} \boldsymbol{g}_{\hat{\boldsymbol{\theta}}_i},$$

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

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

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

Step 6 Iterate steps 2-5 until convergence.

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

Step 4 By using the updated $\hat{\boldsymbol{\theta}}_{i+1}$ from step 3, solve for $\hat{\boldsymbol{\xi}}_{i+1} = \boldsymbol{V}_1\hat{\boldsymbol{\tau}}_{i+1} + \boldsymbol{V}_2\hat{\boldsymbol{\varphi}}_{i+1}$, where $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are given in Theorem 2. For given $\hat{\boldsymbol{\varphi}}_{i+1}$, $\hat{\boldsymbol{\tau}}_{i+1}$ is obtained from (2.65). Compute $\hat{\boldsymbol{\Lambda}}_{i+1} = \boldsymbol{\Lambda}(\hat{\boldsymbol{\xi}}_{i+1})$ and $\boldsymbol{G}(\hat{\boldsymbol{\mu}}_{i+1})$ 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 $\hat{\boldsymbol{\Gamma}}$ and $\hat{\boldsymbol{\Lambda}}$.

To find an initial guess for $\hat{\mathbf{\Gamma}}$ under the hypothesis (1.21), H_0: $\boldsymbol{A} \in \mathcal{R}(\mathbf{\Gamma}\boldsymbol{M})$, the singular value decomposition of the sample covariance matrix is employed as follows:

$$\boldsymbol{S} = \hat{\mathbf{\Gamma}}^* \hat{\mathbf{\Lambda}}^* \hat{\mathbf{\Gamma}}^{*\prime}.$$

Assume that the diagonal entries of $\hat{\mathbf{\Lambda}}^*$ are sorted from small to large and $\hat{\mathbf{\Gamma}}^*$ is a matrix of the corresponding eigenvectors. The quantity $\hat{\mathbf{\Gamma}}^*$ can be partitioned as $\begin{bmatrix}\hat{\mathbf{\Gamma}}_1^* & \hat{\mathbf{\Gamma}}_2^*\end{bmatrix}$, where $\hat{\mathbf{\Gamma}}_1^*$ and $\hat{\mathbf{\Gamma}}_2^*$ have dimensions $p \times m$ and $p \times (p - m)$, respectively. Denote the original initial guesses for $\hat{\mathbf{\Gamma}}$ and $\hat{\mathbf{\Lambda}}$ by $\hat{\mathbf{\Gamma}}_0$ and $\hat{\mathbf{\Lambda}}_0$, respectively. Let $\boldsymbol{A}^*$ be a $p \times m$ semi-orthogonal matrix, where $\boldsymbol{A}^* = (\boldsymbol{A} \quad \boldsymbol{A}^c)$ and $\boldsymbol{A}^c$ is orthogonal to $\boldsymbol{A}$. Let $\boldsymbol{Q}$ be a block diagonal $p \times p$ orthogonal matrix which, when applied to $\begin{bmatrix}\boldsymbol{A}^* & \boldsymbol{A}^{*c}\end{bmatrix}$, results in a matrix which is as similar as possible to $\hat{\mathbf{\Gamma}}^*$. To find $\hat{\mathbf{\Gamma}}_0$, choose $\boldsymbol{Q}$ such that

$$\hat{\mathbf{\Gamma}}_0 = \begin{bmatrix}\boldsymbol{A}^* & \boldsymbol{A}^{*c}\end{bmatrix} \boldsymbol{Q} \approx \hat{\mathbf{\Gamma}}^*, \tag{3.9}$$

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

$$\begin{bmatrix}\boldsymbol{A}^* & \boldsymbol{A}^{*c}\end{bmatrix} \begin{pmatrix} \boldsymbol{Q}_1 & 0 \\ 0 & \boldsymbol{Q}_2 \end{pmatrix} \approx \begin{pmatrix}\hat{\mathbf{\Gamma}}_1^* & \hat{\mathbf{\Gamma}}_2^*\end{pmatrix},$$

where $\boldsymbol{Q}_1 \in \mathcal{O}(m)$ and $\boldsymbol{Q}_2 \in \mathcal{O}(p - m)$. Accordingly,

$$\boldsymbol{A}^*\boldsymbol{Q}_1 \approx \hat{\mathbf{\Gamma}}_1^* \text{ and} \tag{3.10}$$

$$\boldsymbol{A}^{*c}\boldsymbol{Q}_2 \approx \hat{\mathbf{\Gamma}}_2^*. \tag{3.11}$$

To solve for the orthogonal matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$, the orthogonal rotation to congruence by Cliff (1966) is repeatedly applied. This method minimizes $\|\boldsymbol{A}^*\boldsymbol{Q}_1 - \hat{\mathbf{\Gamma}}_1^*\|^2$

with respect to $\boldsymbol{Q}_1$ for (3.10) and $\|\boldsymbol{A}^{*c}\boldsymbol{Q}_2 - \hat{\boldsymbol{\Gamma}}_2^*\|^2$ with respect to $\boldsymbol{Q}_2$ for (3.11) and maximizes tr($\hat{\boldsymbol{\Gamma}}_1^*\boldsymbol{Q}_1'\boldsymbol{A}^{*\prime}$) and tr($\hat{\boldsymbol{\Gamma}}_2^*\boldsymbol{Q}_2'\boldsymbol{A}^{*c'}$), respectively. The results are shown in the algorithm below.

It follows from (3.10) that

$$[\boldsymbol{A} \quad \boldsymbol{A}^c] = \hat{\boldsymbol{\Gamma}}_1^*\boldsymbol{Q}_1' = [\boldsymbol{B}_1 \quad \boldsymbol{B}_2], \tag{3.12}$$

where $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are the first q and the last $m-q$ columns of $\hat{\boldsymbol{\Gamma}}_1^*\boldsymbol{Q}_1'$, respectively. The quantity $\boldsymbol{A}^c$ must be orthogonal to $\boldsymbol{A}$. Accordingly, let $\boldsymbol{A}^c = (\boldsymbol{I}_p - \boldsymbol{A}\boldsymbol{A}')\boldsymbol{F}$ for some $\boldsymbol{F}$, where $\boldsymbol{F}$ is orthogonal by columns. Thus,

$$(\boldsymbol{I}_p - \boldsymbol{A}\boldsymbol{A}')\boldsymbol{F} = \boldsymbol{B}_2. \tag{3.13}$$

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

Step 1 Begin with any orthogonal matrix $\boldsymbol{Q}_1$.

Step 2 By using $\boldsymbol{Q}_1$ from step 1, compute $\boldsymbol{B}_2$ from (3.12). Solve for $\boldsymbol{F}$ (Cliff, 1966) in (3.13) as follows: write the singular value decomposition of $\boldsymbol{B}_2'(\boldsymbol{I}_p - \boldsymbol{A}\boldsymbol{A}')$ as

$$\begin{aligned} \boldsymbol{B}_2'(\boldsymbol{I}_p - \boldsymbol{A}\boldsymbol{A}') &= \boldsymbol{U}_1\boldsymbol{D}_1\boldsymbol{J}_1' \\ \Longrightarrow \boldsymbol{F} &= \boldsymbol{J}_1\boldsymbol{U}_1'. \end{aligned}$$

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

$$\begin{aligned} \hat{\boldsymbol{\Gamma}}_1^{*\prime}\boldsymbol{A}^* &= \boldsymbol{U}_2\boldsymbol{D}_2\boldsymbol{J}_2' \\ \Longrightarrow \boldsymbol{Q}_1 &= \boldsymbol{J}_2\boldsymbol{U}_2'. \end{aligned}$$

Step 4 Repeat steps 2-3 until $\boldsymbol{Q}_1$ converges.

Step 5 Solve for $\boldsymbol{Q}_2$ (Cliff, 1966) in (3.11) as follows: write the singular value decomposition of $\hat{\boldsymbol{\Gamma}}_2^{*\prime}\boldsymbol{A}^{*c}$ as

$$\begin{aligned}\hat{\boldsymbol{\Gamma}}_2^{*\prime}\boldsymbol{A}^{*c} &= \boldsymbol{U}_3\boldsymbol{D}_3\boldsymbol{J}_3' \\ \Longrightarrow \boldsymbol{Q}_2 &= \boldsymbol{J}_3\boldsymbol{U}_3'.\end{aligned}$$

Step 6 Obtain $\hat{\boldsymbol{\Gamma}}_0$ from (3.9) by using $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ from steps 4-5, respectively.

To find $\hat{\boldsymbol{\Lambda}}_0$ or $\hat{\boldsymbol{\varphi}}_0$ that satisfies model (2.4), let $\hat{\boldsymbol{\lambda}}^* = \text{diag}(\hat{\boldsymbol{\Lambda}}^*)$. Differentiating $\left\|\hat{\boldsymbol{\lambda}}^* - \boldsymbol{T}_1 \exp\{\odot\boldsymbol{T}_2\boldsymbol{\varphi}_0\}\right\|^2$ or minimizing $\left(\hat{\boldsymbol{\lambda}}^* - \boldsymbol{T}_1\exp\{\odot\boldsymbol{T}_2\boldsymbol{\varphi}_0\}\right)'\left(\hat{\boldsymbol{\lambda}}^* - \boldsymbol{T}_1\exp\{\odot\boldsymbol{T}_2\boldsymbol{\varphi}_0\}\right)$ with respect to $\exp\{\odot\boldsymbol{T}_2\boldsymbol{\varphi}_0\}$ yields

$$\frac{\partial\left(\hat{\boldsymbol{\lambda}}^* - \boldsymbol{T}_1\exp\{\odot\boldsymbol{T}_2\boldsymbol{\varphi}_0\}\right)'\left(\hat{\boldsymbol{\lambda}}^* - \boldsymbol{T}_1\exp\{\odot\boldsymbol{T}_2\boldsymbol{\varphi}_0\}\right)}{\partial\exp\{\odot\boldsymbol{T}_2\boldsymbol{\varphi}_0\}}.$$

Equating the derivative to zero yields

$$\begin{aligned}\boldsymbol{T}_1'\boldsymbol{T}_1\exp\{\odot\boldsymbol{T}_2\boldsymbol{\varphi}_0\} &= \boldsymbol{T}_1'\hat{\boldsymbol{\lambda}}^* \\ \Longrightarrow \exp\{\odot\boldsymbol{T}_2\boldsymbol{\varphi}_0\} &= (\boldsymbol{T}_1'\boldsymbol{T}_1)^{-1}\boldsymbol{T}_1'\hat{\boldsymbol{\lambda}}^*.\end{aligned}$$

If $(\boldsymbol{T}_1'\boldsymbol{T}_1)^{-1}\boldsymbol{T}_1'\hat{\boldsymbol{\lambda}}^*$ contains positive numbers only, then taking the log of both sides yields

$$\hat{\boldsymbol{\varphi}}_0 = (\boldsymbol{T}_2'\boldsymbol{T}_2)^{-1}\boldsymbol{T}_2'\ln\left\{(\boldsymbol{T}_1'\boldsymbol{T}_1)^{-1}\boldsymbol{T}_1'\hat{\boldsymbol{\lambda}}^*\right\}. \tag{3.14}$$

If $(\boldsymbol{T}_1'\boldsymbol{T}_1)^{-1}\boldsymbol{T}_1'\hat{\boldsymbol{\lambda}}^*$ contains non-positive numbers, a way to find $\hat{\boldsymbol{\varphi}}_0$ is to minimize $\left\|\hat{\boldsymbol{\lambda}}^* - \boldsymbol{T}_1\exp\{\odot\boldsymbol{T}_2\boldsymbol{\varphi}_0\}\right\|^2$ with respect to $\boldsymbol{\varphi}_0$ which can be done numerically by using the Newton-Raphson algorithm. To compute $\hat{\boldsymbol{\lambda}}_0$, φ in (2.4) is replaced by $\hat{\varphi}_0$.

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

Step 1 Set $\hat{\varphi}_0 = \text{tr}(\boldsymbol{S})$ and $\hat{\boldsymbol{\Lambda}}^* = \text{diag}(\hat{\boldsymbol{\lambda}}^*)$. Let $\boldsymbol{X} = [\mathbf{1}_{q_3} \quad \boldsymbol{T}_4]$ and if $(\boldsymbol{T}_3'\boldsymbol{T}_3)^{-1}\boldsymbol{T}_3'\hat{\boldsymbol{\lambda}}^*$ contains positive numbers only, define

$$\begin{pmatrix} \hat{\boldsymbol{\xi}}_0 \\ d \end{pmatrix} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' \ln\left[(\boldsymbol{T}_3'\boldsymbol{T}_3)^{-1}\boldsymbol{T}_3'\hat{\boldsymbol{\lambda}}^*\right],$$

where d is a scalar. If $(\boldsymbol{T}_3'\boldsymbol{T}_3)^{-1}\boldsymbol{T}_3'\hat{\boldsymbol{\lambda}}^*$ contains non-positive numbers, $\hat{\boldsymbol{\xi}}_0$ can be obtained numerically by minimizing $\left\|\hat{\boldsymbol{\lambda}}^* - \boldsymbol{\lambda}\right\|^2$ with respect to $\boldsymbol{\xi}_0$, where $\boldsymbol{\lambda}$ is defined in (2.9).

Step 2 Compute $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ from Theorem 2 and obtain $\hat{\varphi}_0 = (\boldsymbol{V}_2'\boldsymbol{V}_2)^{-1}\boldsymbol{V}_2'\hat{\boldsymbol{\xi}}_0$.

Step 3 Set $\hat{\boldsymbol{\xi}}_0 = \boldsymbol{V}_1\hat{\boldsymbol{\tau}}_0 + \boldsymbol{V}_2\hat{\varphi}_0$. For fixed $\hat{\varphi}_0$, solve for $\hat{\boldsymbol{\tau}}_0$ to satisfy the constraint (2.10) by using the Newton-Raphson algorithm in (2.65). Use $\hat{\boldsymbol{\tau}} = (\boldsymbol{V}_1'\boldsymbol{V}_1)^{-1}\boldsymbol{V}_1'\hat{\boldsymbol{\xi}}_0$ to be an initial guess.

Step 4 Obtain new $\hat{\boldsymbol{\xi}}_0$ from

$$\hat{\boldsymbol{\xi}}_0 = \boldsymbol{V}_1\hat{\boldsymbol{\tau}}_0 + \boldsymbol{V}_2\hat{\varphi}_0.$$

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

Asymptotic Distributions of Estimators

Normal Population

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

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

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

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

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

$$\begin{aligned} \bar{\boldsymbol{I}}_{\boldsymbol{\theta}} &= \frac{n}{4}\boldsymbol{F}^{(1)'}\left(\Sigma^{-1} \otimes \Sigma^{-1}\right) \mathrm{E}[\mathrm{vec}(\boldsymbol{S} - \Sigma)\{\mathrm{vec}(\boldsymbol{S} - \Sigma)\}'] \left(\Sigma^{-1} \otimes \Sigma^{-1}\right) \boldsymbol{F}^{(1)} \\ &= \frac{n}{4}\boldsymbol{F}^{(1)'}\left(\Sigma^{-1} \otimes \Sigma^{-1}\right) \mathrm{var}(\mathrm{vec}\,\boldsymbol{S})(\Sigma^{-1} \otimes \Sigma^{-1})\boldsymbol{F}^{(1)}, \end{aligned}$$

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

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

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

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

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

$$\bar{\boldsymbol{I}}_{\boldsymbol{\theta}} = \begin{pmatrix} \bar{\boldsymbol{I}}_{\theta,\mu\mu} & \mathbf{0} \\ \mathbf{0} & \bar{\boldsymbol{I}}_{\theta,\varphi\varphi} \end{pmatrix}.$$

Proof. *To show that* $\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}$ *is a block diagonal matrix,* $\bar{\boldsymbol{I}}_{\theta,\mu\varphi}$ *and* $\bar{\boldsymbol{I}}_{\theta,\varphi\mu}$ *must be zeros. It follows from (3.18) that*

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

Because

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

it follows that

$$\begin{aligned}
-\frac{1}{n}\,\mathrm{E}\left[\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\mu} \otimes \partial \boldsymbol{\varphi}'}\right]\Bigg|_{\boldsymbol{\mu}=\boldsymbol{0}} &= \frac{1}{2}\left(\boldsymbol{D}_G^{(1)\prime}(\boldsymbol{\Lambda}'\boldsymbol{\Gamma}'\otimes\boldsymbol{\Gamma}')\right)\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\boldsymbol{L}\,\boldsymbol{D}_\lambda^{(1)} \\
&= \frac{1}{2}\boldsymbol{D}_G^{(1)\prime}\left(\boldsymbol{I}_p\otimes\boldsymbol{\Lambda}^{-1}\right)\boldsymbol{L}\,\boldsymbol{D}_\lambda^{(1)}.
\end{aligned}$$

Expressing $\boldsymbol{L}$ *as in (2.3) gives*

$$\begin{aligned}
\bar{\boldsymbol{I}}_{\theta,\mu\varphi} &= \frac{1}{2}\boldsymbol{D}_G^{(1)\prime}\sum_{j=1}^{p}\left(\boldsymbol{e}_j^p\otimes\boldsymbol{e}_j^p\right)\boldsymbol{e}_j^{p\prime}\lambda_i^{-1}\,\boldsymbol{D}_\lambda^{(1)} \\
&= \frac{1}{2}\boldsymbol{V}_3\boldsymbol{A}_1\sum_{j=1}^{p}(\boldsymbol{I}_{p^2}-\boldsymbol{I}_{(p,p)})(\boldsymbol{e}_j^p\otimes\boldsymbol{e}_j^p)\boldsymbol{e}_j^{p\prime}\lambda_i^{-1}\,\boldsymbol{D}_\lambda^{(1)} \\
&= \boldsymbol{0},
\end{aligned}$$

because $\boldsymbol{I}_{(p,p)}(\boldsymbol{e}_j^p\otimes\boldsymbol{e}_j^p) = (\boldsymbol{e}_j^p\otimes\boldsymbol{e}_j^p)$.

Similarly, It can be shown that

$$\begin{aligned}
\bar{\boldsymbol{I}}_{\theta,\varphi\mu} &= -\frac{1}{n}\,\mathrm{E}\left[\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi} \otimes \partial \boldsymbol{\mu}'}\right] \\
&= \frac{1}{2}\boldsymbol{D}_\lambda^{(1)\prime}\boldsymbol{L}'\left(\boldsymbol{I}_p\otimes\boldsymbol{\Lambda}^{-1}\right)\boldsymbol{D}_G^{(1)} \\
&= \boldsymbol{0}.
\end{aligned}$$

Hence, $\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}$ *is a block diagonal matrix.* □

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

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

$$\sqrt{n}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}) \xrightarrow{L} N\left(\mathbf{0}, \bar{\boldsymbol{I}}_{\theta,\varphi\varphi}^{-1}\right), \tag{3.20}$$

and $\hat{\boldsymbol{\mu}} \perp\!\!\!\perp \hat{\boldsymbol{\varphi}}$.

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

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

where

$$\boldsymbol{D}(\boldsymbol{\theta}) = \left.\frac{\partial g(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}'}\right|_{\hat{\boldsymbol{\theta}}=\boldsymbol{\theta}}.$$

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

To find the asymptotic distributions of eigenvector estimators, consider that

$$\begin{aligned} \hat{\boldsymbol{\Gamma}} &= \boldsymbol{\Gamma}\hat{\boldsymbol{G}} \\ \Longrightarrow \operatorname{vec}\hat{\boldsymbol{\Gamma}} &= (\boldsymbol{I}_p \otimes \boldsymbol{\Gamma}) \operatorname{vec}\hat{\boldsymbol{G}}. \end{aligned}$$

The Taylor expansion of $\operatorname{vec}\hat{\mathbf{\Gamma}}$ around $\hat{\boldsymbol{\mu}} = \mathbf{0}$ can be expressed as

$$\begin{aligned}\operatorname{vec}\hat{\mathbf{\Gamma}} &= \operatorname{vec}\mathbf{\Gamma} + (\boldsymbol{I}_p \otimes \mathbf{\Gamma})\,\boldsymbol{D}_G^{(1)}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) + O_p\left(n^{-1}\right) \quad \text{and} \\ \sqrt{n}\operatorname{vec}(\hat{\mathbf{\Gamma}} - \mathbf{\Gamma}) &= \sqrt{n}(\boldsymbol{I}_p \otimes \mathbf{\Gamma})\,\boldsymbol{D}_G^{(1)}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) + O_p\left(n^{-1/2}\right).\end{aligned}$$

Using the delta method and (3.19) yields

$$\sqrt{n}\operatorname{vec}(\hat{\mathbf{\Gamma}} - \mathbf{\Gamma}) \xrightarrow{L} N\left[\mathbf{0}, (\boldsymbol{I}_p \otimes \mathbf{\Gamma})\,\boldsymbol{D}_G^{(1)}\,\bar{\boldsymbol{I}}_{\theta,\mu\mu}^{-1}\,\boldsymbol{D}_G^{(1)\prime}(\boldsymbol{I}_p \otimes \mathbf{\Gamma}')\right]. \tag{3.21}$$

If γ_i is identifiable, it follows from (3.21) that

$$\sqrt{n}(\hat{\gamma}_i - \gamma_i) \xrightarrow{L} N\left[\mathbf{0}, (\boldsymbol{I}_p \otimes \mathbf{e}_i'\mathbf{\Gamma})\,\boldsymbol{D}_G^{(1)}\,\bar{\boldsymbol{I}}_{\theta,\mu\mu}^{-1}\,\boldsymbol{D}_G^{(1)\prime}(\boldsymbol{I}_p \otimes \mathbf{\Gamma}'\mathbf{e}_i)\right]. \tag{3.22}$$

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

$$\hat{\boldsymbol{\lambda}} = \boldsymbol{L}'\operatorname{vec}\hat{\mathbf{\Lambda}}.$$

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

$$\begin{aligned}\hat{\boldsymbol{\lambda}} &= \boldsymbol{\lambda} + \boldsymbol{D}_\lambda^{(1)}(\hat{\varphi} - \varphi) + O_p(n^{-1}) \quad \text{and} \\ \sqrt{n}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) &= \sqrt{n}\,\boldsymbol{D}_\lambda^{(1)}(\hat{\varphi} - \varphi) + O_p(n^{-1/2}).\end{aligned}$$

Using the delta method and (3.20) yields

$$\sqrt{n}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) \xrightarrow{L} N\left[\mathbf{0}, \boldsymbol{D}_\lambda^{(1)}\,\bar{\boldsymbol{I}}_{\theta,\varphi\varphi}^{-1}\,\boldsymbol{D}_\lambda^{(1)\prime}\right]. \tag{3.23}$$

Non-normal Population

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

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

where $\boldsymbol{\Omega} = \lim_{n\to\infty} \frac{1}{4} \boldsymbol{F}^{(1)'} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{\Psi}_n \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}^{(1)}$ and $\boldsymbol{\Psi}_n = n\,\mathrm{var}(\mathrm{vec}\,\boldsymbol{S})$. Boik (1998) provided an unbiased estimator of $\boldsymbol{\Psi}_n$. Asymptotic distributions of $\hat{\boldsymbol{\Gamma}}$, $\hat{\gamma}_i$, and $\hat{\boldsymbol{\lambda}}$ can be obtained by substituting $\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1} \Omega \bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}$ for $\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}$ 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 Σ is rarely known and tests of hypotheses must be based on an appropriate estimate of Σ (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 χ^2 distribution.

Hypotheses

Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_p$ be the eigenvalues of Σ and $\boldsymbol{\Gamma} = (\gamma_1 \quad \gamma_2 \cdots \gamma_p)$ 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

$$\mathrm{H}_0: \frac{\boldsymbol{C}_1'\boldsymbol{\lambda}}{\mathrm{tr}(\boldsymbol{\Sigma})} = \boldsymbol{c}_0,$$

where $\boldsymbol{C}_1$ is a matrix of known constants and $\boldsymbol{c}_0$ 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, $\boldsymbol{C}_1$ is defined as $(\mathbf{1}_{p-k}' \quad \mathbf{0}_k')'$ and $\boldsymbol{c}_0$ 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

$$\mathrm{H}_0: \boldsymbol{A} \in \mathcal{R}(\boldsymbol{\Gamma M}), \text{ where } \mathcal{R}(\boldsymbol{M}) = \mathcal{R}\begin{pmatrix} \boldsymbol{I}_m \\ \mathbf{0} \end{pmatrix},$$

where $\boldsymbol{A}$ is a fixed $p \times q$ semi-orthogonal matrix with rank q and $\boldsymbol{\Gamma}$ is a matrix of eigenvectors. The alternative hypothesis is the general H_a which says H_0 is not true. Suppose $\boldsymbol{A}$ is equated to $\boldsymbol{A} = (\boldsymbol{I}_q \quad \mathbf{0}')'$. 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

$$\Lambda^* = \frac{L_0(\hat{\boldsymbol{\theta}}_0)}{L_a(\hat{\boldsymbol{\theta}}_a)}, \tag{4.1}$$

where $L_0(\hat{\boldsymbol{\theta}}_0)$ and $L_a(\hat{\boldsymbol{\theta}}_a)$ are the likelihood functions maximized under H_0 and H_a, and $\hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\theta}}_a$ 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

$$Q = -2 \ln \Lambda^* = 2\left\{\ell_a(\hat{\boldsymbol{\theta}}_a) - \ell_0(\hat{\boldsymbol{\theta}}_0)\right\} = D(\hat{\boldsymbol{\theta}}_a; \boldsymbol{\theta}) - D(\hat{\boldsymbol{\theta}}_0; \boldsymbol{\theta}), \tag{4.2}$$

where $D(\hat{\boldsymbol{\theta}}; \boldsymbol{\theta})$ is defined as $2\left\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\boldsymbol{\theta})\right\}$ and $\ell(\cdot) = \ln L(\cdot)$. Denote the dimensions of $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_a$ by ν_0 and ν_a, respectively. The test statistic Q asymptotically follows

$$Q \xrightarrow{L} \chi^2_{\nu_\triangle}, \tag{4.3}$$

where $\nu_\triangle = \nu_a - \nu_0$. 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 $D(\hat{\boldsymbol{\theta}}; \boldsymbol{\theta})$ in the following manner.

Denote the i^{th} derivative of the loglikelihood function, evaluated at $\boldsymbol{\mu} = \mathbf{0}$, by $\ell_i(\boldsymbol{\theta})$. That is,

$$\begin{aligned}
\ell_1(\boldsymbol{\theta}) &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \\
\ell_2(\boldsymbol{\theta}) &= \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}' \otimes \partial \boldsymbol{\theta}}, \\
\ell_3(\boldsymbol{\theta}) &= \frac{\partial^3 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}' \otimes \partial \boldsymbol{\theta}' \otimes \partial \boldsymbol{\theta}},
\end{aligned}$$

and so forth. Let

$$\boldsymbol{K}_i = n^{-1}\,\mathrm{E}(\ell_i(\boldsymbol{\theta})) \quad \text{and}$$

$$\boldsymbol{Z}_i = \sqrt{n}\left(\frac{\ell_i(\boldsymbol{\theta})}{n} - \boldsymbol{K}_i\right),$$

where $\mathrm{E}(\boldsymbol{Z}_i) = \mathbf{0}$. Therefore,

$$\ell_i(\boldsymbol{\theta}) = \sqrt{n}\boldsymbol{Z}_i + n\boldsymbol{K}_i. \tag{4.4}$$

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

$$\begin{aligned}
\mathbf{0} &= \boldsymbol{Z}_1 + \left(n^{-\frac{1}{2}}\boldsymbol{Z}_2 + \boldsymbol{K}_2\right)\boldsymbol{g} + \frac{1}{2}\left(n^{-1}\boldsymbol{Z}_3 + n^{-\frac{1}{2}}\boldsymbol{K}_3\right)(\boldsymbol{g}\otimes\boldsymbol{g}) \\
&\quad + \frac{1}{6}\left(n^{-\frac{3}{2}}\boldsymbol{Z}_4 + n^{-1}\boldsymbol{K}_4\right)(\boldsymbol{g}\otimes\boldsymbol{g}\otimes\boldsymbol{g}) + O_p\left(n^{-\frac{3}{2}}\right),
\end{aligned} \tag{4.5}$$

where $\boldsymbol{g} = \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ and $\boldsymbol{K}_1 = \mathbf{0}$. Let

$$\boldsymbol{g} = \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \hat{\boldsymbol{\delta}}_0 + n^{-\frac{1}{2}}\hat{\boldsymbol{\delta}}_1 + n^{-1}\hat{\boldsymbol{\delta}}_2 + O_p\left(n^{-\frac{3}{2}}\right). \tag{4.6}$$

To solve for $\hat{\boldsymbol{\delta}}_i$ in terms of $\boldsymbol{Z}_i$, replace $\boldsymbol{g}$ in (4.5) by (4.6) to obtain polynomial functions of n which have zero coefficients. Collecting each of coefficients and solving for $\hat{\boldsymbol{\delta}}_i$ yields the following results:

$$
\begin{aligned}
\hat{\boldsymbol{\delta}}_0 &= \bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\boldsymbol{Z}_1, \\
\hat{\boldsymbol{\delta}}_1 &= \bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\left\{\boldsymbol{Z}_2\hat{\boldsymbol{\delta}}_0 + \frac{1}{2}\boldsymbol{K}_3\left(\hat{\boldsymbol{\delta}}_0\otimes\hat{\boldsymbol{\delta}}_0\right)\right\}, \\
\hat{\boldsymbol{\delta}}_2 &= \bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\left\{\boldsymbol{Z}_2\hat{\boldsymbol{\delta}}_1 + \frac{1}{2}\boldsymbol{Z}_3\left(\hat{\boldsymbol{\delta}}_0\otimes\hat{\boldsymbol{\delta}}_0\right) + \boldsymbol{K}_3\left(\hat{\boldsymbol{\delta}}_0\otimes\hat{\boldsymbol{\delta}}_1\right) + \frac{1}{6}\boldsymbol{K}_4\left(\hat{\boldsymbol{\delta}}_0\otimes\hat{\boldsymbol{\delta}}_0\otimes\hat{\boldsymbol{\delta}}_0\right)\right\},
\end{aligned}
$$

where $\boldsymbol{K}_2 = -\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}$.

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

$$
\begin{aligned}
D(\hat{\boldsymbol{\theta}};\boldsymbol{\theta}) &= 2\boldsymbol{g}'\boldsymbol{Z}_1 + \boldsymbol{g}'\left(n^{-\frac{1}{2}}\boldsymbol{Z}_2 + \boldsymbol{K}_2\right)\boldsymbol{g} + \frac{1}{3}\boldsymbol{g}'\left(n^{-1}\boldsymbol{Z}_3 + n^{-\frac{1}{2}}\boldsymbol{K}_3\right)(\boldsymbol{g}\otimes\boldsymbol{g}) \\
&\quad + \frac{1}{12}\boldsymbol{g}'\left(n^{-\frac{3}{2}}\boldsymbol{Z}_4 + n^{-1}\boldsymbol{K}_4\right)(\boldsymbol{g}\otimes\boldsymbol{g}\otimes\boldsymbol{g}) + O_p\left(n^{-\frac{3}{2}}\right).
\end{aligned}
$$

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

$$
\begin{aligned}
D(\hat{\boldsymbol{\theta}};\boldsymbol{\theta}) &= \boldsymbol{Z}_1^{*\prime}\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}\boldsymbol{Z}_1^* + n^{-\frac{1}{2}}\left\{\boldsymbol{Z}_1^{*\prime}\boldsymbol{K}_3\left(\boldsymbol{Z}_1^*\otimes\boldsymbol{Z}_1^*\right)/3 + \boldsymbol{Z}_1^{*\prime}\boldsymbol{Z}_2\boldsymbol{Z}_1^*\right\} \\
&\quad + n^{-1}\Big[\left\{\boldsymbol{Z}_2\boldsymbol{Z}_1^* + \boldsymbol{K}_3(\boldsymbol{Z}_1^*\otimes\boldsymbol{Z}_1^*)/2\right\}'\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\left\{\boldsymbol{Z}_2\boldsymbol{Z}_1^* + \boldsymbol{K}_3(\boldsymbol{Z}_1^*\otimes\boldsymbol{Z}_1^*)/2\right\} \\
&\quad + \boldsymbol{Z}_1^{*\prime}\boldsymbol{K}_4\left(\boldsymbol{Z}_1^*\otimes\boldsymbol{Z}_1^*\otimes\boldsymbol{Z}_1^*\right)/12 + \boldsymbol{Z}_1^{*\prime}\boldsymbol{Z}_3\left(\boldsymbol{Z}_1^*\otimes\boldsymbol{Z}_1^*\right)/3\Big] + O_p\left(n^{-\frac{3}{2}}\right),
\end{aligned} \tag{4.7}
$$

where $\boldsymbol{Z}_1^* = \bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\boldsymbol{Z}_1$. Because it can be shown that

$$
\boldsymbol{Z}_1 \xrightarrow{L} N\left(\boldsymbol{0}, \bar{\boldsymbol{I}}_{\boldsymbol{\theta}}\right),
$$

the $O_p(1)$ term in (4.7), $\boldsymbol{Z}_1^{*\prime}\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}\boldsymbol{Z}_1^*$, is asymptotically distributed as χ^2_ν and $\mathrm{E}\left(\boldsymbol{Z}_1^{*\prime}\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}\boldsymbol{Z}_1^*\right) = \nu$. 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 $\boldsymbol{Z}_i$ and $\boldsymbol{K}_i$ for $\ell(\boldsymbol{\theta})$ 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

$$Q_c = \frac{Q\ \nu_\Delta}{\widehat{\mathrm{E}}(Q|\mathrm{H}_0)} \tag{4.8}$$

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

$$\widehat{\mathrm{E}}(Q|\mathrm{H}_0) = \widehat{\mathrm{E}}\left\{D(\hat{\boldsymbol{\theta}}_a;\boldsymbol{\theta}_a)|\mathrm{H}_0\right\} - \widehat{\mathrm{E}}\left\{D(\hat{\boldsymbol{\theta}}_0;\boldsymbol{\theta}_0)|\mathrm{H}_0\right\}. \tag{4.9}$$

The approximate expectations can be obtained by evaluation of the expectation of equation (4.7). To compute the expectation of $D(\hat{\boldsymbol{\theta}};\boldsymbol{\theta})$ in (4.7), the third and fourth order moments of $\mathrm{vec}(\boldsymbol{S}-\boldsymbol{\Sigma})$ 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 $\boldsymbol{S} = \frac{\sum_{i=1}^N \boldsymbol{u}_i\boldsymbol{u}_i'}{N}$, where $\boldsymbol{u}_i \overset{iid}{\sim} N(\boldsymbol{0},\boldsymbol{\Sigma})$, then $\boldsymbol{S} \sim W_p(n,\boldsymbol{\Sigma}/n)$. Note that $\boldsymbol{D}_p$ and $\boldsymbol{H}_p$ satisfy $\boldsymbol{D}_p\,\mathrm{vech}\,\boldsymbol{X} =$

$\operatorname{vec} \boldsymbol{X}$, $\boldsymbol{H}_p \operatorname{vec} \boldsymbol{X} = \operatorname{vech} \boldsymbol{X}$, and $\boldsymbol{D}_p \boldsymbol{H}_p = \boldsymbol{N}_p$, where $\boldsymbol{X}$ is any $p \times p$ symmetric matrix and $\boldsymbol{N}_p$ is defined in (2.47). The moment generating function of $\operatorname{vech}(\boldsymbol{S})$ can be written as

$$\begin{aligned} \boldsymbol{M}_{\operatorname{vech}(\boldsymbol{S})}(\boldsymbol{t}) &= \mathrm{E}\left[\exp(\boldsymbol{t}' \operatorname{vech}(\boldsymbol{S}))\right] \\ &= \mathrm{E}\left[\exp(\boldsymbol{t}' \boldsymbol{H}_p \operatorname{vec}(\boldsymbol{S}))\right] \\ &= \frac{|\boldsymbol{\Sigma}^*|^{n/2}}{|\boldsymbol{\Sigma}|^{n/2}}, \end{aligned}$$

where $\operatorname{vech} \boldsymbol{\Sigma}^{*^{-1}} = \operatorname{vech} \boldsymbol{\Sigma}^{-1} - \frac{2}{n} \left(\boldsymbol{D}_p' \boldsymbol{D}_p\right)^{-1} \boldsymbol{t}$. Then, the cumulant generating function of $\operatorname{vech}(\boldsymbol{S})$ is

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

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

$$\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}^{*^{-1}}}{\partial \boldsymbol{t}'} = -\frac{2}{n} \boldsymbol{H}_p' \quad \text{and} \tag{4.10}$$

$$\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}^*}{\partial \boldsymbol{t}'} = \frac{2}{n} \left(\boldsymbol{\Sigma}^* \otimes \boldsymbol{\Sigma}^*\right) \boldsymbol{H}_p'. \tag{4.11}$$

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

$$\begin{aligned} \boldsymbol{C}^{(1)}(\boldsymbol{t}) &= \frac{\partial \boldsymbol{C}^{(0)}(\boldsymbol{t})}{\partial \boldsymbol{t}} = \frac{n}{2} \frac{\partial \ln |\boldsymbol{\Sigma}^*|}{\partial \boldsymbol{t}} \\ \Longrightarrow \frac{\partial \boldsymbol{C}^{(0)}(\boldsymbol{t})}{\partial t_i} &= \frac{n}{2} \operatorname{tr}\left(\boldsymbol{\Sigma}^{*-1} \frac{\partial \boldsymbol{\Sigma}^*}{\partial t_i}\right) = -\frac{n}{2} \operatorname{tr}\left(\boldsymbol{\Sigma}^* \frac{\partial \boldsymbol{\Sigma}^{*^{-1}}}{\partial t_i}\right) \\ &= -\frac{n}{2} (\operatorname{vec} \boldsymbol{\Sigma}^*)' \frac{\partial \operatorname{vec} \boldsymbol{\Sigma}^{*^{-1}}}{\partial t_i} \\ \Longrightarrow \frac{\partial \boldsymbol{C}^{(0)}(\boldsymbol{t})}{\partial \boldsymbol{t}'} &= (\operatorname{vec} \boldsymbol{\Sigma}^*)' \boldsymbol{H}_p'. \end{aligned}$$

Therefore,

$$C^{(1)}(t) \;=\; H_p \operatorname{vec} \Sigma^* \text{ or } \operatorname{vech} \Sigma^*. \tag{4.12}$$

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

$$\begin{aligned} C^{(2)}(t) &= \frac{\partial^2 C^{(0)}(t)}{\partial t' \otimes \partial t} = H_p \frac{\partial \operatorname{vec} \Sigma^*}{\partial t'} \\ &= \frac{2}{n} H_p (\Sigma^* \otimes \Sigma^*) H_p'. \end{aligned} \tag{4.13}$$

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

$$\begin{aligned} C^{(3)}(t) &= \frac{\partial^3 C^{(0)}(t)}{\partial t' \otimes \partial t' \otimes \partial t} = \frac{2}{n} H_p \frac{\partial (\Sigma^* \otimes \Sigma^*) H_p'}{\partial t'} \\ &= \frac{2}{n} H_p \left\{ 2 \left(I_p \otimes (\operatorname{vec} I_p)' \otimes \Sigma^* \right) \left(\frac{\partial \operatorname{vec} \Sigma^*}{\partial t'} \otimes H_p' \right) \right\} \\ &= \frac{8}{n^2} H_p \left(\Sigma^* \otimes (\operatorname{vec} \Sigma^*)' \otimes \Sigma^* \right) (H_p \otimes H_p)', \end{aligned} \tag{4.14}$$

by $(\operatorname{vec} I_p)' (\Sigma^* \otimes I_p) = (\operatorname{vec} \Sigma^*)'$.

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

$$E(ABC \otimes D) = \{(\operatorname{vec} B)' \otimes E\} \left[(I_r \otimes \operatorname{vec} A) C \otimes D \right],$$

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

$$E(A \otimes BCD) = \{(\operatorname{vec} C)' \otimes E\} I_{(tp,qr)} \left[A \otimes (\operatorname{vec} B)' \otimes D \right],$$

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

$$
\begin{aligned}
\boldsymbol{C}^{(4)}(\boldsymbol{t}) &= \frac{\partial^4 \boldsymbol{C}^{(0)}(\boldsymbol{t})}{\partial \boldsymbol{t} \otimes \partial \boldsymbol{t}' \otimes \partial \boldsymbol{t}' \otimes \partial \boldsymbol{t}} \\
&= \frac{8}{n^2} \boldsymbol{H}_p \Bigg\{ \left(\frac{\partial (\operatorname{vec} \boldsymbol{\Sigma}^*)'}{\partial \boldsymbol{t}} \otimes \boldsymbol{H}_p \right) (\boldsymbol{I}_p \otimes \operatorname{vec} \boldsymbol{I}_p \otimes (\operatorname{vec} \boldsymbol{\Sigma}^*)' \otimes \boldsymbol{\Sigma}^*) \\
&\quad + \left(\frac{\partial (\operatorname{vec} \boldsymbol{\Sigma}^*)'}{\partial \boldsymbol{t}} \otimes \boldsymbol{H}_p \right) \boldsymbol{I}_{(p^2,p^2)} (\boldsymbol{\Sigma}^* \otimes (\operatorname{vec} \boldsymbol{\Sigma}^*)' \otimes \operatorname{vec} \boldsymbol{I}_p \otimes \boldsymbol{I}_p) \\
&\quad + \left(\boldsymbol{I}_{\frac{p(p+1)}{2}} \otimes \boldsymbol{H}_p \right) \left(\frac{\partial (\operatorname{vec} \boldsymbol{\Sigma}^*)'}{\partial \boldsymbol{t}} \otimes \boldsymbol{\Sigma}^* \otimes \boldsymbol{\Sigma}^* \right) \left(\boldsymbol{I}_{(p,p^2)} \otimes \boldsymbol{I}_p \right) \Bigg\} (\boldsymbol{H}_p \otimes \boldsymbol{H}_p)'.
\end{aligned}
$$

Substituting (4.11) and simplifying yields

$$
\begin{aligned}
\boldsymbol{C}^{(4)}(\boldsymbol{t}) &= \frac{32}{n^3} (\boldsymbol{H}_p \otimes \boldsymbol{H}_p) \boldsymbol{N}_{p^2} \{ \boldsymbol{\Sigma}^* \otimes (\operatorname{vec} \boldsymbol{\Sigma}^*)' \otimes \operatorname{vec} \boldsymbol{\Sigma}^* \otimes \boldsymbol{\Sigma}^* \} (\boldsymbol{H}_p \otimes \boldsymbol{H}_p)' \qquad (4.15) \\
&\quad + \frac{16}{n^3} (\boldsymbol{H}_p \otimes \boldsymbol{H}_p)(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(p,p)} \otimes \boldsymbol{I}_p)(\boldsymbol{\Sigma}^* \otimes \boldsymbol{\Sigma}^* \otimes \boldsymbol{\Sigma}^* \otimes \boldsymbol{\Sigma}^*)(\boldsymbol{H}_p \otimes \boldsymbol{H}_p)'.
\end{aligned}
$$

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

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

$$
\begin{aligned}
\boldsymbol{C}^{(1)}(\boldsymbol{0}) &= \mathrm{E}(\boldsymbol{w}) \quad \text{and} \\
\boldsymbol{C}^{(2)}(\boldsymbol{0}) &= \mathrm{E}\{(\boldsymbol{w} - \boldsymbol{\sigma})(\boldsymbol{w} - \boldsymbol{\sigma})'\}.
\end{aligned}
$$

Equivalently,

$$
\boldsymbol{C}^{(2)}(\boldsymbol{0}) = \boldsymbol{H}_p \, \mathrm{E}\left[\operatorname{vec}(\boldsymbol{S} - \boldsymbol{\Sigma})\{\operatorname{vec}(\boldsymbol{S} - \boldsymbol{\Sigma})\}'\right] \boldsymbol{H}_p', \tag{4.16}
$$

where $\boldsymbol{w} = \operatorname{vech} \boldsymbol{S} = \boldsymbol{H}_p \operatorname{vec} \boldsymbol{S}$. Pre-multiplying (4.13) by $\boldsymbol{D}_p$ and post-multiplying by $\boldsymbol{D}_p'$ and using (4.16) yield

$$
\mathrm{E}\{\operatorname{vec}(\boldsymbol{S} - \boldsymbol{\Sigma})\{\operatorname{vec}(\boldsymbol{S} - \boldsymbol{\Sigma})\}'\} = \frac{2}{n} \boldsymbol{N}_p (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) = \operatorname{var}(\operatorname{vec} \boldsymbol{S}), \tag{4.17}
$$

because $\boldsymbol{N}_p \operatorname{vec}(\boldsymbol{S}-\Sigma) = \operatorname{vec}(\boldsymbol{S}-\Sigma)$ and $\boldsymbol{N}_p(\Sigma \otimes \Sigma) = (\Sigma \otimes \Sigma)\,\boldsymbol{N}_p$. The third cumulant of $\boldsymbol{w}$ can be written as

$$\boldsymbol{C}^{(3)}(\boldsymbol{0}) = \mathrm{E}\left\{(\boldsymbol{w}-\boldsymbol{\sigma})' \otimes (\boldsymbol{w}-\boldsymbol{\sigma})(\boldsymbol{w}-\boldsymbol{\sigma})'\right\}.$$

Equivalently,

$$\boldsymbol{C}^{(3)}(\boldsymbol{0}) = \boldsymbol{H}_p \,\mathrm{E}\left[\{\operatorname{vec}(\boldsymbol{S}-\Sigma)\}' \otimes \operatorname{vec}(\boldsymbol{S}-\Sigma)\{\operatorname{vec}(\boldsymbol{S}-\Sigma)\}'\right](\boldsymbol{H}_p \otimes \boldsymbol{H}_p)', \quad (4.18)$$

where $\boldsymbol{w} = \operatorname{vech} \boldsymbol{S} = \boldsymbol{H}_p \operatorname{vec} \boldsymbol{S}$. Pre-multiplying (4.14) by $\boldsymbol{D}_p$ and post-multiplying by $(\boldsymbol{D}_p \otimes \boldsymbol{D}_p)'$ and using (4.18) yield

$$\begin{aligned}\mathrm{E}\left[\{\operatorname{vec}(\boldsymbol{S}-\Sigma)\}' \otimes \operatorname{vec}(\boldsymbol{S}-\Sigma)\{\operatorname{vec}(\boldsymbol{S}-\Sigma)\}'\right] &= \frac{8}{n^2}\boldsymbol{N}_p\{\Sigma \otimes (\operatorname{vec}\Sigma)' \otimes \Sigma\} \\ &\quad \times(\boldsymbol{N}_p \otimes \boldsymbol{N}_p). \qquad (4.19)\end{aligned}$$

Similarly, the fourth cumulant of $\operatorname{vech} \boldsymbol{S}$ can be written as

$$\begin{aligned}\boldsymbol{C}^{(4)}(\boldsymbol{0}) &= (\boldsymbol{H}_p \otimes \boldsymbol{H}_p)\Big[\mathrm{E}\left\{\operatorname{vec}(\boldsymbol{S}-\Sigma)\{\operatorname{vec}(\boldsymbol{S}-\Sigma)\}' \otimes \operatorname{vec}(\boldsymbol{S}-\Sigma)\{\operatorname{vec}(\boldsymbol{S}-\Sigma)\}'\right\} \\ &\quad -2\boldsymbol{N}_{p^2}\left[\mathrm{E}\{\operatorname{vec}(\boldsymbol{S}-\Sigma)\{\operatorname{vec}(\boldsymbol{S}-\Sigma)\}'\} \otimes \mathrm{E}\{\operatorname{vec}(\boldsymbol{S}-\Sigma)\{\operatorname{vec}(\boldsymbol{S}-\Sigma)\}'\}\right] \\ &\quad - \mathrm{E}\{\operatorname{vec}(\boldsymbol{S}-\Sigma) \otimes \operatorname{vec}(\boldsymbol{S}-\Sigma)\}\,\mathrm{E}\{\{\operatorname{vec}(\boldsymbol{S}-\Sigma)\}' \otimes \{\operatorname{vec}(\boldsymbol{S}-\Sigma)\}'\}\Big] \\ &\quad \times(\boldsymbol{H}_p \otimes \boldsymbol{H}_p)'. \qquad (4.20)\end{aligned}$$

Pre-multiplying (4.15) by $(\boldsymbol{D}_p \otimes \boldsymbol{D}_p)$ and post-multiplying by $(\boldsymbol{D}_p \otimes \boldsymbol{D}_p)'$ and using (4.20) yields

$$\mathrm{E}\left[\,\mathrm{vec}(\boldsymbol{S}-\Sigma)\{\mathrm{vec}(\boldsymbol{S}-\Sigma)\}'\otimes\mathrm{vec}(\boldsymbol{S}-\Sigma)\{\mathrm{vec}(\boldsymbol{S}-\Sigma)\}'\right]=$$

$$\frac{32}{n^3}\boldsymbol{N}_{p^2}(\boldsymbol{N}_p\otimes\boldsymbol{N}_p)\{\Sigma\otimes(\mathrm{vec}\,\Sigma)'\otimes\mathrm{vec}\,\Sigma\otimes\Sigma\}(\boldsymbol{N}_p\otimes\boldsymbol{N}_p)$$
$$+\frac{16}{n^3}(\boldsymbol{N}_p\otimes\boldsymbol{N}_p)(\boldsymbol{I}_p\otimes\boldsymbol{I}_{(p,p)}\otimes\boldsymbol{I}_p)(\Sigma\otimes\Sigma\otimes\Sigma\otimes\Sigma)(\boldsymbol{N}_p\otimes\boldsymbol{N}_p)$$
$$+2\boldsymbol{N}_{p^2}\left[\mathrm{E}\{\mathrm{vec}(\boldsymbol{S}-\Sigma)\{\mathrm{vec}(\boldsymbol{S}-\Sigma)\}'\}\otimes\mathrm{E}\{\mathrm{vec}(\boldsymbol{S}-\Sigma)\{\mathrm{vec}(\boldsymbol{S}-\Sigma)\}'\}\right]$$
$$+\mathrm{E}\{\mathrm{vec}(\boldsymbol{S}-\Sigma)\otimes\mathrm{vec}(\boldsymbol{S}-\Sigma)\}\,\mathrm{E}\{\{\mathrm{vec}(\boldsymbol{S}-\Sigma)\}'\otimes\{\mathrm{vec}(\boldsymbol{S}-\Sigma)\}'\}.\quad(4.21)$$

To compute the last part of (4.21), the second derivative of $\boldsymbol{C}^{(0)}(\boldsymbol{t})$ with respect to $\boldsymbol{t}$ evaluated at $\boldsymbol{t}=\boldsymbol{0}$ can be obtained as

$$\frac{\partial^2\boldsymbol{C}^{(0)}(\boldsymbol{t})}{\partial\boldsymbol{t}\otimes\partial\boldsymbol{t}}=\frac{2}{n}(\boldsymbol{H}_p\otimes\boldsymbol{H}_p)\,\mathrm{vec}\,(\Sigma\otimes\Sigma)\,.\quad(4.22)$$

Algebraically,

$$\frac{\partial^2\boldsymbol{C}^{(0)}(\boldsymbol{t})}{\partial\boldsymbol{t}\otimes\partial\boldsymbol{t}}=\mathrm{E}\{\mathrm{vec}(\boldsymbol{S}-\Sigma)\otimes\mathrm{vec}(\boldsymbol{S}-\Sigma)\}.\quad(4.23)$$

Pre-multiplying (4.22) by $(\boldsymbol{D}_p\otimes\boldsymbol{D}_p)$ and using (4.23) yields

$$\mathrm{E}\{\mathrm{vec}(\boldsymbol{S}-\Sigma)\otimes\mathrm{vec}(\boldsymbol{S}-\Sigma)\}=\frac{2}{n}(\boldsymbol{N}_p\otimes\boldsymbol{N}_p)\,\mathrm{vec}\,(\Sigma\otimes\Sigma)\,.\quad(4.24)$$

Similarly,

$$\mathrm{E}\,[\{\mathrm{vec}(\boldsymbol{S}-\Sigma)\}'\otimes\{\mathrm{vec}(\boldsymbol{S}-\Sigma)\}']=\frac{2}{n}\{\mathrm{vec}\,(\Sigma\otimes\Sigma)\}'(\boldsymbol{N}_p\otimes\boldsymbol{N}_p).\quad(4.25)$$

Substituting (4.24) and (4.25) in (4.21) yields

$$\mathrm{E}\left[\,\mathrm{vec}(\boldsymbol{S}-\Sigma)\{\mathrm{vec}(\boldsymbol{S}-\Sigma)\}'\otimes\mathrm{vec}(\boldsymbol{S}-\Sigma)\{\mathrm{vec}(\boldsymbol{S}-\Sigma)\}'\right]=$$

$$\frac{32}{n^3}\boldsymbol{N}_{p^2}(\boldsymbol{N}_p\otimes\boldsymbol{N}_p)\{\Sigma\otimes(\mathrm{vec}\,\Sigma)'\otimes\mathrm{vec}\,\Sigma\otimes\Sigma\}(\boldsymbol{N}_p\otimes\boldsymbol{N}_p)$$

$$+\frac{16}{n^3}(\boldsymbol{N}_p \otimes \boldsymbol{N}_p)(\boldsymbol{I}_p \otimes \boldsymbol{I}_{(p,p)} \otimes \boldsymbol{I}_p)(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})(\boldsymbol{N}_p \otimes \boldsymbol{N}_p)$$
$$+\frac{8}{n^2}\boldsymbol{N}_{p^2}(\boldsymbol{N}_p \otimes \boldsymbol{N}_p)(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})(\boldsymbol{N}_p \otimes \boldsymbol{N}_p)$$
$$+\frac{4}{n^2}(\boldsymbol{N}_p \otimes \boldsymbol{N}_p) \operatorname{vec}(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\{\operatorname{vec}(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\}'(\boldsymbol{N}_p \otimes \boldsymbol{N}_p). \tag{4.26}$$

The results of the third and fourth order moments of $\operatorname{vec}(\boldsymbol{S} - \boldsymbol{\Sigma})$ are used for computing the expectation of $D(\hat{\boldsymbol{\theta}}; \boldsymbol{\theta})$ in (4.7).

Expectation of LRT Statistic

Partition $\boldsymbol{\theta}$ as $\boldsymbol{\theta} = (\boldsymbol{\mu}' \quad \boldsymbol{\varphi}')'$ with dimension $(\nu_1 \quad \nu_2)'$, where $\nu = \nu_1 + \nu_2$. In particular, $k = 2$. Denote the derivatives of $\operatorname{vec} \boldsymbol{\Sigma}$ with respect to $\boldsymbol{\theta}$ as follows:

$$\begin{aligned}
\boldsymbol{F}^{(1)} &= \frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\theta}'} = \sum_{s=1}^{k} \frac{\partial \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\theta}'_s} \boldsymbol{E}_{s,\nu}; \\
\boldsymbol{F}^{(2)} &= \frac{\partial^2 \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\theta}' \otimes \partial \boldsymbol{\theta}'} = \sum_{s=1}^{k} \sum_{t=1}^{k} \frac{\partial^2 \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\theta}'_s \otimes \partial \boldsymbol{\theta}'_t} (\boldsymbol{E}_{s,\nu} \otimes \boldsymbol{E}_{t,\nu}); \\
\boldsymbol{F}^{(11)} &= \frac{\partial^2 \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\theta} \otimes \partial \boldsymbol{\theta}'} = \operatorname{dvec}\left(\boldsymbol{F}^{(2)}, p^2\nu, \nu\right); \\
\boldsymbol{F}^{(3)} &= \frac{\partial^3 \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\theta}' \otimes \partial \boldsymbol{\theta}' \otimes \partial \boldsymbol{\theta}'} = \sum_{s=1}^{k} \sum_{t=1}^{k} \sum_{u=1}^{k} \frac{\partial^3 \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\theta}'_s \otimes \partial \boldsymbol{\theta}'_t \otimes \partial \boldsymbol{\theta}'_u} (\boldsymbol{E}_{s,\nu} \otimes \boldsymbol{E}_{t,\nu} \otimes \boldsymbol{E}_{u,\nu}); \\
\boldsymbol{F}^{(111)} &= \operatorname{dvec}\left(\boldsymbol{F}^{(3)}, p^2\nu^2, \nu\right).
\end{aligned}$$

The matrix $\boldsymbol{E}_{i,\nu}$ with dimension $\nu_i \times \nu$ for $1 \le i \le k$ is defined as

$$\boldsymbol{E}_{i,\nu} = \begin{bmatrix} \boldsymbol{0}_{\nu_i \times a_i} & \boldsymbol{I}_{\nu_i} & \boldsymbol{0}_{\nu_i \times (\nu - a_i - \nu_i)} \end{bmatrix}.$$

where $a_i = \sum_{j=1}^{i} \nu_j - \nu_i$. For convenience, define $\ddot{\boldsymbol{F}}^{(1)}, \ddot{\boldsymbol{F}}^{(2)}, \dddot{\boldsymbol{F}}^{(1)}$, and $\dddot{\boldsymbol{F}}^{(2)}$ as

$$\begin{aligned}
\ddot{\boldsymbol{F}}^{(1)} &= \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}^{(1)}; \quad \ddot{\boldsymbol{F}}^{(2)} = \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}^{(2)}; \\
\dddot{\boldsymbol{F}}^{(1)} &= \left(\boldsymbol{\Sigma}^{-1} \otimes \operatorname{vec} \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}^{(1)}; \quad \text{and} \quad \dddot{\boldsymbol{F}}^{(2)} = \left(\boldsymbol{\Sigma}^{-1} \otimes \operatorname{vec} \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}^{(2)}.
\end{aligned}$$

Boik (2002a, in a supplement) derived the expectation of the right-hand-side of the equation (4.7) as follows:

$$\mathrm{E}\{D(\hat{\boldsymbol{\theta}};\boldsymbol{\theta})\} = \dim(\boldsymbol{\theta}) + \frac{1}{n}\sum_{i=1}^{10}\zeta_i + O(n^{-2}), \tag{4.27}$$

where

$$\begin{aligned}
\zeta_1 &= \operatorname{tr}\left\{\left(\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\,\ddot{\boldsymbol{F}}^{(1)\prime}\otimes \boldsymbol{N}_p\right)(\boldsymbol{I}_p\otimes \operatorname{vec}\boldsymbol{\Sigma}\otimes \boldsymbol{I}_p)\,\boldsymbol{F}^{(1)}\,\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\boldsymbol{U}^{(1)}\right\};\\
\boldsymbol{U}^{(1)} &= \boldsymbol{F}^{(11)\prime}\left(\boldsymbol{I}_\nu\otimes\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right) - 2\,\dddot{\boldsymbol{F}}^{(1)\prime}\left(\boldsymbol{F}^{(1)}\otimes\boldsymbol{I}_{p^2}\right);\\
\zeta_2 &= \frac{1}{3}\operatorname{tr}\left\{\left(\boldsymbol{F}^{(1)}\,\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\otimes\boldsymbol{F}^{(1)}\,\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\right)'\,\dddot{\boldsymbol{F}}^{(1)\prime}\,\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\boldsymbol{K}_3^*\right\};\\
\boldsymbol{K}_3^* &= 2\,\dddot{\boldsymbol{F}}^{(1)\prime}\left(\boldsymbol{F}^{(1)}\otimes\boldsymbol{F}^{(1)}\right) - \operatorname{dvec}\left(\ddot{\boldsymbol{F}}^{(1)\prime}\,\boldsymbol{F}^{(2)},\nu^2,\nu\right)' - \frac{1}{2}\,\ddot{\boldsymbol{F}}^{(1)\prime}\,\boldsymbol{F}^{(2)};\\
\zeta_3 &= \frac{1}{2}\operatorname{tr}\left[\left\{\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\otimes\boldsymbol{N}_p(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\right\}\boldsymbol{U}^{(1)\prime}\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\boldsymbol{U}^{(1)}\right];\\
\zeta_4 &= \boldsymbol{L}_1'\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\boldsymbol{L}_1 + \boldsymbol{L}_2'\left(\boldsymbol{I}_{(\nu,\nu)}\otimes\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\right)\boldsymbol{L}_2;\\
\boldsymbol{L}_1 &= \frac{1}{2}\boldsymbol{U}^{(1)}\operatorname{vec}\left(\boldsymbol{F}^{(1)}\,\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\right);\quad \boldsymbol{L}_2 = \frac{1}{2}\operatorname{vec}\left\{\boldsymbol{U}^{(1)}\left(\boldsymbol{I}_\nu\otimes\boldsymbol{F}^{(1)}\,\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\right)\right\};\\
\zeta_5 &= \frac{1}{2}\operatorname{tr}\left\{\boldsymbol{N}_\nu\left(\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\otimes\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\right)\boldsymbol{K}_3^{*\prime}\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\boldsymbol{K}_3^*\right\} + \frac{1}{4}\left(\operatorname{vec}\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\right)'\boldsymbol{K}_3^{*\prime}\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\boldsymbol{K}_3^*\operatorname{vec}\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1};\\
\zeta_6 &= \operatorname{tr}\left\{\boldsymbol{N}_\nu\left(\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\otimes\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\,\boldsymbol{F}^{(1)\prime}\right)\boldsymbol{U}^{(1)\prime}\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\boldsymbol{K}_3^*\right\};\\
\zeta_7 &= \frac{1}{2}\left(\operatorname{vec}\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\right)'\boldsymbol{K}_3^{*\prime}\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\boldsymbol{U}^{(1)}\operatorname{vec}\left(\boldsymbol{F}^{(1)}\,\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\right);\\
\zeta_8 &= \frac{1}{3}\operatorname{tr}\left\{\boldsymbol{N}_\nu\left(\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\,\boldsymbol{F}^{(1)\prime}\otimes\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\right)\widetilde{\mathrm{U}}^{(2)}\right\};\\
\widetilde{\mathrm{U}}^{(2)} &= -6\operatorname{dvec}\left(\dddot{\boldsymbol{F}}^{(2)\prime},p^2\nu,p^2\right)\boldsymbol{F}^{(2)}\\
&\quad+\operatorname{dvec}\left[\operatorname{dvec}\left\{\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}^{(3)},p^2\nu^2,\nu\right\}',p^2\nu,p^2\right]\\
&\quad+2\left(\boldsymbol{I}_{p^2}\otimes\boldsymbol{F}^{(1)}\right)'\left(\boldsymbol{I}_p\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{I}_p\right)\left(2\,\boldsymbol{N}_p+\boldsymbol{I}_{(p,p)}\right)\boldsymbol{L}_3;
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{L}_3 &= \left(\boldsymbol{\Sigma}^{-1} \otimes \operatorname{vec} \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_p\right)' \left(\boldsymbol{F}^{(1)} \otimes \boldsymbol{F}^{(1)}\right); \\
\zeta_9 &= \frac{1}{6} \left\{\operatorname{vec}\left(\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1} \boldsymbol{F}^{(1)\prime}\right)\right\}' \widetilde{\mathrm{U}}^{(2)} \left(\operatorname{vec} \bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\right); \\
\zeta_{10} &= \frac{1}{6} \operatorname{tr}\left\{\boldsymbol{N}_\nu \left(\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1} \otimes \bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\right) \boldsymbol{K}_4^*\right\} + \frac{1}{12} \left(\operatorname{vec} \bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}\right)' \boldsymbol{K}_4^* \operatorname{vec} \bar{\boldsymbol{I}}_{\boldsymbol{\theta}}^{-1}; \quad \text{and} \\
\boldsymbol{K}_4^* &= 12 \left(\boldsymbol{F}^{(1)} \otimes \boldsymbol{F}^{(1)}\right)' \dddot{\boldsymbol{F}}^{(2)} - 9 \boldsymbol{L}_3' \boldsymbol{I}_{(p,p)} \boldsymbol{L}_3 \\
&\quad - 2 \operatorname{dvec}\left(\ddot{\boldsymbol{F}}^{(1)\prime} \boldsymbol{F}^{(3)}, \nu^2, \nu^2\right) - \frac{3}{2} \boldsymbol{F}^{(2)\prime} \ddot{\boldsymbol{F}}^{(2)}.
\end{aligned}$$

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 $\widehat{\mathrm{E}}\left\{D(\hat{\boldsymbol{\theta}}_0; \boldsymbol{\theta}_0)|\mathrm{H}_0\right\}$ in (4.9). To obtain $\widehat{\mathrm{E}}\left\{D(\hat{\boldsymbol{\theta}}_a; \boldsymbol{\theta}_a)|\mathrm{H}_0\right\}$, Boik (2002a) showed that if $\boldsymbol{\Sigma}$ is unrestricted under H_a, then the asymptotic distribution of Q in (4.2) is $\chi^2_{\nu_\triangle}$, where $\nu_\triangle = \frac{p(p+1)}{2} - \nu_0$. Furthermore,

$$\widehat{\mathrm{E}}\left\{D(\hat{\boldsymbol{\theta}}_a; \boldsymbol{\theta}_a)|\mathrm{H}_0\right\} = \frac{p(p+1)}{2} + \frac{p\left(2p^2 + 3p - 1\right)}{12n} + O\left(n^{-2}\right). \tag{4.28}$$

To verify equation (4.28), it follows from (4.2) that

$$\begin{aligned}
\widehat{\mathrm{E}}\left\{D(\hat{\boldsymbol{\theta}}_a; \boldsymbol{\theta}_a)|\mathrm{H}_0\right\} &= 2\widehat{\mathrm{E}}\left\{\ell(\hat{\boldsymbol{\theta}}_a) - \ell(\boldsymbol{\theta}_a)\right\} \\
&= 2\widehat{\mathrm{E}}\left\{-\frac{n}{2} \ln|\boldsymbol{S}| + \frac{n}{2} \ln|\boldsymbol{\Sigma}|\right\} \\
&= 2\widehat{\mathrm{E}}\left\{-\frac{n}{2} \ln|n\boldsymbol{S}| + \frac{np}{2} \ln(n) + \frac{n}{2} \ln|\boldsymbol{\Sigma}|\right\}.
\end{aligned}$$

Seber (1984) showed that $|n\boldsymbol{S}| \sim |\boldsymbol{\Sigma}| \prod_{i=1}^{p} \chi^2_{n-i+1}$. Accordingly,

$$\widehat{\mathrm{E}}\left\{D(\hat{\boldsymbol{\theta}}_a; \boldsymbol{\theta}_a)|\mathrm{H}_0\right\} = np\ln(n) - n\sum_{i=1}^{p} \widehat{\mathrm{E}}\left(\ln \chi^2_{n-i+1}\right). \tag{4.29}$$

The expectation of $\ln(y)$, $\mathrm{E}\{\ln(y)\}$, assuming $y \sim \chi^2_n$, can be obtained by differentiating both sides of

$$\Gamma\left(\frac{n}{2}\right) = \int_0^\infty \frac{y^{\frac{n}{2}-1}\exp\{-\frac{y}{2}\}}{2^{\frac{n}{2}}} dy \tag{4.30}$$

with respect to n, where $\Gamma(\cdot)$ is the gamma function. The useful result concerning the digamma (ψ) 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

$$\psi\left(\frac{n}{2}\right) = \int_0^\infty \ln(y)\frac{y^{\frac{n}{2}-1}\exp\{-\frac{y}{2}\}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} dy - \ln(2).$$

Therefore,

$$\mathrm{E}\{\ln(y)\} = \psi\left(\frac{n}{2}\right) + \ln(2).$$

Expand the digamma function to obtain

$$\psi\left(\frac{n}{2}\right) + \ln(2) = \ln(n) - \frac{1}{n} - \frac{1}{3n^2} + \cdots \tag{4.31}$$

Applying result (4.31) to (4.29), simplification yields the result (4.28).

Matrix Expressions for $\boldsymbol{Z}_i$ and $\boldsymbol{K}_i$

To obtain matrix expressions for $\boldsymbol{Z}_i$ and $\boldsymbol{K}_i$, the second and third derivatives of $\operatorname{vec}\boldsymbol{\Sigma}$ with respect to $\boldsymbol{\theta}'$ evaluated at $\boldsymbol{\mu}=\mathbf{0}$ are needed. The following properties are repeatedly used:

$$(\boldsymbol{ABC}\otimes\boldsymbol{D})\boldsymbol{E} = [\boldsymbol{A}\otimes\{\operatorname{vec}(\boldsymbol{C}')\}'\otimes\boldsymbol{D}]\{\operatorname{vec}(\boldsymbol{B}')\otimes\boldsymbol{E}\} \text{ and} \quad (4.32)$$

$$(\boldsymbol{A}\otimes\boldsymbol{BCD})\boldsymbol{E} = \{\boldsymbol{A}\otimes\operatorname{vec}(\boldsymbol{D}')\otimes\boldsymbol{B}\}(\boldsymbol{E}\otimes\operatorname{vec}\boldsymbol{C}), \quad (4.33)$$

where $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, and $\boldsymbol{E}$ are any matrices that are conformable for multiplication. It follows from (3.4) that

$$\frac{\partial^2\operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'} = 2\boldsymbol{N}_p\left[(\boldsymbol{\Gamma G\Lambda}\otimes\boldsymbol{\Gamma})\frac{\partial^2\operatorname{vec}\boldsymbol{G}}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'} + \frac{\partial}{\partial\boldsymbol{\mu}'}\{\boldsymbol{\Gamma G\Lambda}\otimes\boldsymbol{\Gamma}\}\times\left(\boldsymbol{I}_{\nu_1}\otimes\frac{\partial\operatorname{vec}\boldsymbol{G}}{\partial\boldsymbol{\mu}'}\right)\right].$$

Applying (4.33) yields

$$\frac{\partial^2\operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'} = 2\boldsymbol{N}_p\left[(\boldsymbol{\Gamma G\Lambda}\otimes\boldsymbol{\Gamma})\frac{\partial^2\operatorname{vec}\boldsymbol{G}}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'} + \{\boldsymbol{\Gamma}\otimes(\operatorname{vec}\boldsymbol{\Lambda})'\otimes\boldsymbol{\Gamma}\}\left(\boldsymbol{I}_{(p,p)}\frac{\partial\operatorname{vec}\boldsymbol{G}}{\partial\boldsymbol{\mu}'}\otimes\frac{\partial\operatorname{vec}\boldsymbol{G}}{\partial\boldsymbol{\mu}'}\right)\right]. \quad (4.34)$$

Therefore, evaluated at $\boldsymbol{\mu}=\mathbf{0}$,

$$\left.\frac{\partial^2\operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'}\right|_{\mu=0} = 2\boldsymbol{N}_p\left[(\boldsymbol{\Gamma\Lambda}\otimes\boldsymbol{\Gamma})\,\boldsymbol{D}_G^{(2)} + \{\boldsymbol{\Gamma}\otimes(\operatorname{vec}\boldsymbol{\Lambda})'\otimes\boldsymbol{\Gamma}\}\left(\boldsymbol{I}_{(p,p)}\,\boldsymbol{D}_G^{(1)}\otimes\boldsymbol{D}_G^{(1)}\right)\right].$$

Also, it follows from (3.4) that

$$\begin{aligned}\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\mu}'} &= 2\,\boldsymbol{N}_p\frac{\partial}{\partial\boldsymbol{\varphi}'}\left\{(\boldsymbol{\Gamma G\Lambda}\otimes\boldsymbol{\Gamma})\frac{\partial\operatorname{vec}\boldsymbol{G}}{\partial\boldsymbol{\mu}'}\right\}\\ &= 2\,\boldsymbol{N}_p\frac{\partial}{\partial\boldsymbol{\varphi}'}\left\{(\boldsymbol{\Gamma}\otimes(\operatorname{vec}\boldsymbol{I}_p)'\otimes\boldsymbol{\Gamma})\left(\operatorname{vec}(\boldsymbol{G\Lambda})'\otimes\frac{\partial\operatorname{vec}\boldsymbol{G}}{\partial\boldsymbol{\mu}'}\right)\right\} \text{ by (4.32).}\end{aligned}$$

Therefore, evaluated at $\boldsymbol{\mu}=\boldsymbol{0}$,

$$\left.\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}} = 2\,\boldsymbol{N}_p\,(\boldsymbol{\Gamma}\otimes(\operatorname{vec}\boldsymbol{I}_p)'\otimes\boldsymbol{\Gamma})\left\{\boldsymbol{L}\,\boldsymbol{D}_\lambda^{(1)}\otimes\boldsymbol{D}_G^{(1)}\right\}. \tag{4.35}$$

The second derivative of $\operatorname{vec}\boldsymbol{\Sigma}$ with respect to $\boldsymbol{\varphi}'$ can be written as

$$\left.\frac{\partial^2 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}} = (\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\boldsymbol{L}\,\boldsymbol{D}_\lambda^{(2)}.$$

Similarly, the third derivative of $\operatorname{vec}\boldsymbol{\Sigma}$ with respect to $\boldsymbol{\varphi}'$ can be written as

$$\left.\frac{\partial^3 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}} = (\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\boldsymbol{L}\,\boldsymbol{D}_\lambda^{(3)}.$$

Next, it follows from (4.35) that

$$\left.\frac{\partial^3 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\mu}'}\right|_{\boldsymbol{\mu}=\boldsymbol{0}} = 2\,\boldsymbol{N}_p\,\{\boldsymbol{\Gamma}\otimes(\operatorname{vec}\boldsymbol{I}_p)'\otimes\boldsymbol{\Gamma}\}\left\{\boldsymbol{L}\,\boldsymbol{D}_\lambda^{(2)}\otimes\boldsymbol{D}_G^{(1)}\right\}.$$

It follows from (4.34) that

$$\begin{aligned}\frac{\partial^3 \operatorname{vec}\boldsymbol{\Sigma}}{\partial\boldsymbol{\varphi}'\otimes\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'} &= 2\,\boldsymbol{N}_p\left[\frac{\partial}{\partial\boldsymbol{\varphi}'}\left\{(\boldsymbol{\Gamma G\Lambda}\otimes\boldsymbol{\Gamma})\frac{\partial^2\operatorname{vec}\boldsymbol{G}}{\partial\boldsymbol{\mu}'\otimes\partial\boldsymbol{\mu}'}\right\}\right. \\ &\quad\left.+\frac{\partial}{\partial\boldsymbol{\varphi}'}\{\boldsymbol{\Gamma}\otimes(\operatorname{vec}\boldsymbol{\Lambda})'\otimes\boldsymbol{\Gamma}\}\left(\boldsymbol{I}_{(p,p)}\frac{\partial\operatorname{vec}\boldsymbol{G}}{\partial\boldsymbol{\mu}'}\otimes\frac{\partial\operatorname{vec}\boldsymbol{G}}{\partial\boldsymbol{\mu}'}\right)\right].\end{aligned} \tag{4.36}$$

For convenience, the two parts on the right-hand-side of (4.36) are labeled as $\boldsymbol{E}_a$ and $\boldsymbol{E}_b$. Using (4.33) and evaluating at $\boldsymbol{\mu}=\boldsymbol{0}$, $\boldsymbol{E}_a$ can be simplified to

$$\boldsymbol{E}_a = (\boldsymbol{\Gamma}\otimes(\operatorname{vec}\boldsymbol{I}_p)'\otimes\boldsymbol{\Gamma})\left\{\boldsymbol{L}\,\boldsymbol{D}_\lambda^{(1)}\otimes\boldsymbol{D}_G^{(2)}\right\}. \tag{4.37}$$

The quantity $\boldsymbol{E}_b$, evaluated at $\boldsymbol{\mu} = \mathbf{0}$, can be written as

$$\begin{aligned}
\boldsymbol{E}_b &= \frac{\partial}{\partial \boldsymbol{\varphi}'} \left\{ (\boldsymbol{\Gamma} \otimes (\operatorname{vec} \boldsymbol{\Lambda})' \otimes \boldsymbol{\Gamma}) \left(\boldsymbol{I}_{(p^2,p)} \boldsymbol{I}_{(p,p^2)} \otimes \boldsymbol{I}_p \right) \left(\boldsymbol{I}_{(p,p)} \boldsymbol{D}_G^{(1)} \otimes \boldsymbol{D}_G^{(1)} \right) \right\} \\
&= \frac{\partial}{\partial \boldsymbol{\varphi}'} \left\{ ((\operatorname{vec} \boldsymbol{\Lambda})' \otimes \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}) \left(\boldsymbol{I}_{(p,p^2)} \otimes \boldsymbol{I}_p \right) \left(\boldsymbol{I}_{(p,p)} \boldsymbol{D}_G^{(1)} \otimes \boldsymbol{D}_G^{(1)} \right) \right\} \\
&= \left(\frac{\partial (\operatorname{vec} \boldsymbol{\Lambda})'}{\partial \boldsymbol{\varphi}'} \otimes \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma} \right) \left[\boldsymbol{I}_{\nu_2} \otimes \left(\boldsymbol{I}_{(p,p^2)} \otimes \boldsymbol{I}_p \right) \left(\boldsymbol{I}_{(p,p)} \boldsymbol{D}_G^{(1)} \otimes \boldsymbol{D}_G^{(1)} \right) \right] \\
&= \left(\left\{ \operatorname{vec} \left(\boldsymbol{\Lambda} \boldsymbol{D}_\lambda^{(1)} \right) \right\}' \otimes \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma} \right) \left[\boldsymbol{I}_{\nu_2} \otimes (\boldsymbol{I}_p \otimes \boldsymbol{I}_{(p,p)} \otimes \boldsymbol{I}_p) \left(\boldsymbol{D}_G^{(1)} \otimes \boldsymbol{D}_G^{(1)} \right) \right] \quad (4.38)
\end{aligned}$$

Combining (4.37) and (4.38) yields

$$\begin{aligned}
\left. \frac{\partial^3 \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\varphi}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}'} \right|_{\boldsymbol{\mu}=\mathbf{0}} &= 2 \boldsymbol{N}_p \Bigg[\{ \boldsymbol{\Gamma} \otimes (\operatorname{vec} \boldsymbol{I}_p)' \otimes \boldsymbol{\Gamma} \} \left\{ \boldsymbol{L} \boldsymbol{D}_\lambda^{(1)} \otimes \boldsymbol{D}_G^{(2)} \right\} \\
&\quad + \left\{ \left(\operatorname{vec} \boldsymbol{L} \boldsymbol{D}_\lambda^{(1)} \right)' \otimes \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma} \right\} \\
&\quad \times \left[\boldsymbol{I}_{\nu_2} \otimes (\boldsymbol{I}_p \otimes \boldsymbol{I}_{(p,p)} \otimes \boldsymbol{I}_p) \left(\boldsymbol{D}_G^{(1)} \otimes \boldsymbol{D}_G^{(1)} \right) \right] \Bigg].
\end{aligned}$$

Next, it follows from (4.34) that

$$\begin{aligned}
\frac{\partial^3 \operatorname{vec} \boldsymbol{\Sigma}}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}'} &= 2 \boldsymbol{N}_p \Bigg[\frac{\partial}{\partial \boldsymbol{\mu}'} \left\{ (\boldsymbol{\Gamma} \boldsymbol{G} \boldsymbol{\Lambda} \otimes \boldsymbol{\Gamma}) \frac{\partial^2 \operatorname{vec} \boldsymbol{G}}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}'} \right\} \quad (4.39) \\
&\quad + \frac{\partial}{\partial \boldsymbol{\mu}'} \{ \boldsymbol{\Gamma} \otimes (\operatorname{vec} \boldsymbol{\Lambda})' \otimes \boldsymbol{\Gamma} \} \left(\boldsymbol{I}_{(p,p)} \frac{\partial \operatorname{vec} \boldsymbol{G}}{\partial \boldsymbol{\mu}'} \otimes \frac{\partial \operatorname{vec} \boldsymbol{G}}{\partial \boldsymbol{\mu}'} \right) \Bigg].
\end{aligned}$$

For convenience, the two parts on the right-hand-side of (4.39) are labeled as $\boldsymbol{E}_c$ and $\boldsymbol{E}_d$. Using the product rule and (4.33) and evaluating at $\boldsymbol{\mu} = \mathbf{0}$ yield

$$\boldsymbol{E}_c = \{ \boldsymbol{\Gamma} \otimes (\operatorname{vec} \boldsymbol{\Lambda})' \otimes \boldsymbol{\Gamma} \} \left(\boldsymbol{I}_{(p,p)} \boldsymbol{D}_G^{(1)} \otimes \boldsymbol{D}_G^{(2)} \right) + (\boldsymbol{\Gamma} \boldsymbol{\Lambda} \otimes \boldsymbol{\Gamma}) \boldsymbol{D}_G^{(3)}. \quad (4.40)$$

Using the Kronecker product rule, $\boldsymbol{E}_d$, evaluated at $\boldsymbol{\mu} = \mathbf{0}$, simplifies to

$$\begin{aligned}
\boldsymbol{E}_d &= 2 \boldsymbol{N}_p \{ \boldsymbol{\Gamma} \otimes (\operatorname{vec} \boldsymbol{\Lambda})' \otimes \boldsymbol{\Gamma} \} \Bigg[\left(\boldsymbol{I}_{(p,p)} \boldsymbol{D}_G^{(2)} \otimes \boldsymbol{D}_G^{(1)} \right) \\
&\quad + \boldsymbol{I}_{(p^2,p^2)} \left(\boldsymbol{D}_G^{(2)} \otimes \boldsymbol{I}_{(p,p)} \boldsymbol{D}_G^{(1)} \right) \left(\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{I}_{(\nu_1,\nu_1)} \right) \Bigg].
\end{aligned}$$

Using the following results repeatedly:

$$(\boldsymbol{A} \otimes \boldsymbol{B}) = \boldsymbol{I}_{(p,r)}(\boldsymbol{B} \otimes \boldsymbol{A})\boldsymbol{I}_{(c,q)},$$

where $\boldsymbol{A}$ is $r \times c$ and $\boldsymbol{B}$ is $p \times q$, yields

$$\begin{aligned} \boldsymbol{E}_d &= 2\,\boldsymbol{N}_p \left\{ (\boldsymbol{\Gamma} \otimes (\text{vec}\,\boldsymbol{\Lambda})' \otimes \boldsymbol{\Gamma}) \left(\boldsymbol{I}_{(p,p)}\,\boldsymbol{D}_G^{(2)} \otimes \boldsymbol{D}_G^{(1)} \right) \left(\boldsymbol{I}_{\nu_1^3} + (\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{I}_{(\nu_1,\nu_1)}) \right) \right\} \\ &= 4\,\boldsymbol{N}_p \left\{ (\boldsymbol{\Gamma} \otimes (\text{vec}\,\boldsymbol{\Lambda})' \otimes \boldsymbol{\Gamma}) \left(\boldsymbol{I}_{(p,p)}\,\boldsymbol{D}_G^{(2)} \otimes \boldsymbol{D}_G^{(1)} \right) (\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{N}_{\nu_1}) \right\}, \qquad (4.41) \end{aligned}$$

where $\boldsymbol{N}_{\nu_1} = \frac{1}{2}\left(\boldsymbol{I}_{\nu_1^2} + \boldsymbol{I}_{(\nu_1,\nu_1)} \right)$. Combining (4.40) and (4.41) yields

$$\begin{aligned} \left. \frac{\partial^3\,\text{vec}\,\boldsymbol{\Sigma}}{\partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}' \otimes \partial \boldsymbol{\mu}'} \right|_{\boldsymbol{\mu}=\boldsymbol{0}} &= 2\,\boldsymbol{N}_p \Big[(\boldsymbol{\Gamma}\boldsymbol{\Lambda} \otimes \boldsymbol{\Gamma})\,\boldsymbol{D}_G^{(3)} + \{\boldsymbol{\Gamma} \otimes (\text{vec}\,\boldsymbol{\Lambda})' \otimes \boldsymbol{\Gamma}\} \\ &\quad \times \left\{ \left(\boldsymbol{I}_{(p,p)}\,\boldsymbol{D}_G^{(1)} \otimes \boldsymbol{D}_G^{(2)} \right) + 2\left(\boldsymbol{I}_{(p,p)}\,\boldsymbol{D}_G^{(2)} \otimes \boldsymbol{D}_G^{(1)} \right) (\boldsymbol{I}_{\nu_1} \otimes \boldsymbol{N}_{\nu_1}) \right\} \Big]. \end{aligned}$$

The other order derivatives can be rearranged by post-multiplying by commutation matrices.

$$\begin{aligned} \frac{\partial^2\,\text{vec}\,\boldsymbol{\Sigma}}{\partial \boldsymbol{\theta}_t' \otimes \partial \boldsymbol{\theta}_s'} &= \frac{\partial^2\,\text{vec}\,\boldsymbol{\Sigma}}{\partial \boldsymbol{\theta}_s' \otimes \partial \boldsymbol{\theta}_t'} \boldsymbol{I}_{(\nu_t,\nu_s)} \\ \frac{\partial^3\,\text{vec}\,\boldsymbol{\Sigma}}{\partial \boldsymbol{\theta}_t' \otimes \partial \boldsymbol{\theta}_s' \otimes \partial \boldsymbol{\theta}_u'} &= \frac{\partial^3\,\text{vec}\,\boldsymbol{\Sigma}}{\partial \boldsymbol{\theta}_s' \otimes \partial \boldsymbol{\theta}_t' \otimes \partial \boldsymbol{\theta}_u'} (\boldsymbol{I}_{(\nu_t,\nu_s)} \otimes \boldsymbol{I}_{\nu_u}) \\ &= \frac{\partial^3\,\text{vec}\,\boldsymbol{\Sigma}}{\partial \boldsymbol{\theta}_t' \otimes \partial \boldsymbol{\theta}_u' \otimes \partial \boldsymbol{\theta}_s'} (\boldsymbol{I}_{\nu_t} \otimes \boldsymbol{I}_{(\nu_s,\nu_u)}) \end{aligned}$$

The following properties are useful for simplifying expressions.

$$\begin{aligned} \boldsymbol{F}^{(1)} &= \boldsymbol{I}_{(p,p)}\,\boldsymbol{F}^{(1)} \\ \boldsymbol{F}^{(2)} &= \boldsymbol{I}_{(p,p)}\,\boldsymbol{F}^{(2)} = \boldsymbol{F}^{(2)}\,\boldsymbol{I}_{(\nu,\nu)} \\ \boldsymbol{F}^{(11)} &= (\boldsymbol{I}_{\nu} \otimes \boldsymbol{I}_{(p,p)})\,\boldsymbol{F}^{(11)} \end{aligned}$$

$$\begin{aligned}
\boldsymbol{F}^{(3)} &= \boldsymbol{I}_{(p,p)}\,\boldsymbol{F}^{(3)} = \boldsymbol{F}^{(3)}\,\boldsymbol{I}_{(\nu,\nu^2)} = \boldsymbol{F}^{(3)}\,\boldsymbol{I}_{(\nu^2,\nu)} \\
&= \boldsymbol{F}^{(3)}(\boldsymbol{I}_\nu \otimes \boldsymbol{I}_{(\nu,\nu)}) = \boldsymbol{F}^{(3)}(\boldsymbol{I}_{(\nu,\nu)} \otimes \boldsymbol{I}_\nu) \\
\boldsymbol{F}^{(111)} &= (\boldsymbol{I}_{(p,p)} \otimes \boldsymbol{I}_{df^2})\,\boldsymbol{F}^{(111)} = (\boldsymbol{I}_{p^2} \otimes \boldsymbol{I}_{(\nu,\nu)})\,\boldsymbol{F}^{(111)} \\
\boldsymbol{F}^{(1)\prime}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\,\boldsymbol{F}^{(1)} &= 2\bar{\boldsymbol{I}}_{\boldsymbol{\theta}} \\
\ddot{\boldsymbol{F}}^{(1)\prime}(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\,\ddot{\boldsymbol{F}}^{(1)} &= 2\bar{\boldsymbol{I}}_{\boldsymbol{\theta}} \\
\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \operatorname{vec} \boldsymbol{\Sigma} &= \operatorname{vec} \boldsymbol{\Sigma}^{-1} \\
\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}^{-1}}{\partial \boldsymbol{\theta}'} = -\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}^{(1)} &= \ddot{\boldsymbol{F}}^{(1)}
\end{aligned}$$

It follows from (4.4) that $\boldsymbol{K}_1 = \boldsymbol{0}$. Then

$$\begin{aligned}
\boldsymbol{Z}_1 &= \frac{\sqrt{n}}{2}\,\boldsymbol{F}^{(1)\prime}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \operatorname{vec}(\boldsymbol{S} - \boldsymbol{\Sigma}) \\
&= \frac{\sqrt{n}}{2}\,\ddot{\boldsymbol{F}}^{(1)\prime} \operatorname{vec}(\boldsymbol{S} - \boldsymbol{\Sigma}).
\end{aligned}$$

The second derivatives of $\ell(\boldsymbol{\theta})$ can be written as

$$\begin{aligned}
\ell_2(\boldsymbol{\theta}) &= \frac{n}{2}\,\boldsymbol{F}^{(11)\prime}\left\{\boldsymbol{I}_\nu \otimes \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \operatorname{vec}(\boldsymbol{S} - \boldsymbol{\Sigma})\right\} \\
&\quad + n\,\boldsymbol{F}^{(1)\prime}\left\{\boldsymbol{I}_p \otimes (\operatorname{vec} \boldsymbol{I}_p)' \otimes \boldsymbol{\Sigma}^{-1}\right\} \left(\frac{\partial \operatorname{vec} \boldsymbol{\Sigma}^{-1}}{\partial \boldsymbol{\theta}'} \otimes \operatorname{vec}(\boldsymbol{S} - \boldsymbol{\Sigma})\right) \\
&\quad - \frac{n}{2}\,\boldsymbol{F}^{(1)\prime}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}^{(1)}.
\end{aligned}$$

Accordingly,

$$\begin{aligned}
\ell_2(\boldsymbol{\theta}) &= \frac{n}{2}\,\boldsymbol{F}^{(11)\prime}\left\{\boldsymbol{I}_\nu \otimes \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \operatorname{vec}(\boldsymbol{S} - \boldsymbol{\Sigma})\right\} \\
&\quad - n\,\boldsymbol{F}^{(1)\prime}\left(\boldsymbol{\Sigma}^{-1} \otimes \left(\operatorname{vec} \boldsymbol{\Sigma}^{-1}\right)' \otimes \boldsymbol{\Sigma}^{-1}\right) \left(\boldsymbol{F}^{(1)} \otimes \operatorname{vec}(\boldsymbol{S} - \boldsymbol{\Sigma})\right) \\
&\quad - \frac{n}{2}\,\boldsymbol{F}^{(1)\prime}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \boldsymbol{F}^{(1)}.
\end{aligned}$$

Taking the expectation of $\ell_2(\boldsymbol{\theta})$ and dividing by n yields

$$\boldsymbol{K}_2 = -\frac{1}{2}\,\boldsymbol{F}^{(1)\prime}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{F}^{(1)} = -\frac{1}{2}\,\ddot{\boldsymbol{F}}^{(1)\prime}\,\boldsymbol{F}^{(1)} = -\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}.$$

Therefore,

$$\boldsymbol{Z}_2 = \frac{\sqrt{n}}{2}\,\boldsymbol{F}^{(11)\prime}\left\{\boldsymbol{I}_\nu\otimes\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}(\boldsymbol{S}-\boldsymbol{\Sigma})\right\} - \sqrt{n}\,\dddot{\boldsymbol{F}}^{(1)\prime}\left\{\boldsymbol{F}^{(1)}\otimes\operatorname{vec}(\boldsymbol{S}-\boldsymbol{\Sigma})\right\}.$$

Boik (2002a, in supplement) gave expressions for $\boldsymbol{Z}_3$, $\boldsymbol{K}_3$, and $\boldsymbol{K}_4$ as follows:

$$\begin{aligned}
\boldsymbol{Z}_3 &= \sqrt{n}\left(\operatorname{vec}\ddot{\boldsymbol{F}}^{(1)}\otimes\ddot{\boldsymbol{F}}^{(1)}\right)'\left(\boldsymbol{I}_{p\,\nu}\otimes\boldsymbol{I}_{(p,p)}\otimes\boldsymbol{I}_p\right)\left\{\boldsymbol{I}_\nu\otimes\boldsymbol{F}^{(1)}\otimes\operatorname{vec}(\boldsymbol{S}-\boldsymbol{\Sigma})\right\}\\
&\quad +2\sqrt{n}\,\dddot{\boldsymbol{F}}^{(1)\prime}\left[\boldsymbol{F}^{(1)}\otimes\boldsymbol{N}_p(\boldsymbol{I}_p\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{I}_p)'\left\{\boldsymbol{F}^{(1)}\otimes\operatorname{vec}(\boldsymbol{S}-\boldsymbol{\Sigma})\right\}\right]\\
&\quad +\frac{\sqrt{n}}{2}\,\boldsymbol{F}^{(111)\prime}\left\{\boldsymbol{I}_{df^2}\otimes\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}(\boldsymbol{S}-\boldsymbol{\Sigma})\right\}\\
&\quad -2\sqrt{n}\,\boldsymbol{F}^{(11)\prime}\left(\boldsymbol{I}_\nu\otimes\boldsymbol{\Sigma}^{-1}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)'\left\{\left(\boldsymbol{I}_\nu\otimes\boldsymbol{F}^{(1)}\right)\boldsymbol{N}_\nu\otimes\operatorname{vec}(\boldsymbol{S}-\boldsymbol{\Sigma})\right\}\\
&\quad -\sqrt{n}\,\dddot{\boldsymbol{F}}^{(1)\prime}\left\{\boldsymbol{F}^{(2)}\otimes\operatorname{vec}(\boldsymbol{S}-\boldsymbol{\Sigma})\right\};\\
\boldsymbol{K}_3 &= 2\,\dddot{\boldsymbol{F}}^{(1)\prime}\left(\boldsymbol{F}^{(1)}\otimes\boldsymbol{F}^{(1)}\right)\boldsymbol{N}_\nu - \boldsymbol{F}^{(11)\prime}\left(\boldsymbol{I}_\nu\otimes\ddot{\boldsymbol{F}}^{(1)}\right)\boldsymbol{N}_\nu - \frac{1}{2}\,\ddot{\boldsymbol{F}}^{(1)\prime}\,\boldsymbol{F}^{(2)};\\
\boldsymbol{K}_4 &= \boldsymbol{F}^{(11)\prime}\left\{\boldsymbol{I}_\nu\otimes\left(\boldsymbol{\Sigma}^{-1}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)'\left(\boldsymbol{F}^{(1)}\otimes\boldsymbol{F}^{(1)}\right)\right\}\\
&\quad \times\left\{4(\boldsymbol{N}_\nu\otimes\boldsymbol{I}_\nu)+2\,\boldsymbol{I}_{(\nu^2,\nu)}\right\} + 4\,\dddot{\boldsymbol{F}}^{(1)\prime}\left(\boldsymbol{F}^{(2)}\otimes\boldsymbol{F}^{(1)}\right)\left\{(\boldsymbol{I}_\nu\otimes\boldsymbol{N}_\nu)+\frac{1}{2}\,\boldsymbol{I}_{(\nu,\nu^2)}\right\}\\
&\quad -\boldsymbol{F}^{(1)\prime}\left(\boldsymbol{\Sigma}^{-1}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\left(\boldsymbol{F}^{(1)}\otimes\boldsymbol{F}^{(1)}\otimes\boldsymbol{F}^{(1)}\right)\\
&\quad \times\left\{2\,\boldsymbol{I}_{(\nu,\nu^2)}+3\,\boldsymbol{I}_{(\nu^2,\nu)}+4(\boldsymbol{N}_\nu\otimes\boldsymbol{I}_\nu)\right\}\\
&\quad -\boldsymbol{F}^{(111)\prime}\left(\boldsymbol{I}_{\nu^2}\otimes\ddot{\boldsymbol{F}}^{(1)}\right)\left\{(\boldsymbol{I}_\nu\otimes\boldsymbol{N}_\nu)+\frac{1}{2}\,\boldsymbol{I}_{(\nu,\nu^2)}\right\}\\
&\quad -\boldsymbol{F}^{(11)\prime}\left(\boldsymbol{I}_\nu\otimes\ddot{\boldsymbol{F}}^{(2)}\right)\left\{(\boldsymbol{N}_\nu\otimes\boldsymbol{I}_\nu)+\frac{1}{2}\,\boldsymbol{I}_{(\nu^2,\nu)}\right\} - \frac{1}{2}\,\ddot{\boldsymbol{F}}^{(1)\prime}\,\boldsymbol{F}^{(3)};
\end{aligned}$$

where $\boldsymbol{N}_\nu = \frac{1}{2}(\boldsymbol{I}_{\nu^2}+\boldsymbol{I}_{(\nu,\nu)})$.

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 $T^{**} = \left(\frac{T_{k,q}}{1+\frac{c}{n}}\right)$. 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 $\lambda_1 = \cdots = \lambda_{p-k} = 1$.

<u>The Comparison Test</u>

Partition $\mathbf{\Gamma}$ as

$$\mathbf{\Gamma} = \begin{pmatrix} \underset{q\times m}{\mathbf{\Gamma}_{11}} & \underset{q\times(p-m)}{\mathbf{\Gamma}_{12}} \\ \underset{(m-q)\times m}{\mathbf{\Gamma}_{21}} & \underset{(m-q)\times(p-m)}{\mathbf{\Gamma}_{22}} \\ \underset{(p-m)\times m}{\mathbf{\Gamma}_{31}} & \underset{(p-m)\times(p-m)}{\mathbf{\Gamma}_{32}} \end{pmatrix}.$$

Consider the special case where $\mathbf{\Gamma}_{31} = \mathbf{0}$. In this section a test of hypothesis (1.21), where $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{I}_q \\ \mathbf{0} \end{pmatrix}$ and $\boldsymbol{M}$ is the first m columns of $\boldsymbol{I}_p$, will be developed. Although this test is not useful in practice (because $\mathbf{\Gamma}_{31} = \mathbf{0}$ 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 $\mathbf{\Gamma}_{31} = \mathbf{0}$, $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{I}_q \\ \mathbf{0} \end{pmatrix}$, and $\boldsymbol{M} = \begin{pmatrix} \boldsymbol{I}_m \\ \mathbf{0} \end{pmatrix}$. Then,

$$\boldsymbol{A} \in \mathcal{R}(\mathbf{\Gamma}\boldsymbol{M}) \Longleftrightarrow \mathbf{\Gamma}'\boldsymbol{A} \in \mathcal{R}(\boldsymbol{M})$$

$$\begin{aligned}
&\iff \begin{pmatrix} \boldsymbol{\Gamma}_{11}' & \boldsymbol{\Gamma}_{21}' & \boldsymbol{\Gamma}_{31}' \\ \boldsymbol{\Gamma}_{12}' & \boldsymbol{\Gamma}_{22}' & \boldsymbol{\Gamma}_{32}' \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_q \\ \boldsymbol{0} \end{pmatrix} \in \mathcal{R}\begin{pmatrix} \boldsymbol{I}_m \\ \boldsymbol{0} \end{pmatrix} \\
&\iff \begin{pmatrix} \boldsymbol{\Gamma}_{11}' \\ \boldsymbol{\Gamma}_{12}' \end{pmatrix} \in \mathcal{R}\begin{pmatrix} \boldsymbol{I}_m \\ \boldsymbol{0} \end{pmatrix} \\
&\iff \boldsymbol{\Gamma}_{12} = \boldsymbol{0}.
\end{aligned}$$

Let $\boldsymbol{Y}$ be an $n \times p$ matrix with distribution $\operatorname{vec}\boldsymbol{Y} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n)$. Then, $\boldsymbol{S} = \frac{\boldsymbol{Y}'\boldsymbol{Y}}{n} \sim W_p(n, \boldsymbol{\Sigma})$ and $\boldsymbol{\Sigma} = \boldsymbol{\Gamma\Lambda\Gamma}'$. Partition $\boldsymbol{Y}$ as $\begin{pmatrix} \underset{n\times q}{\boldsymbol{Y}_1} & \underset{n\times(m-q)}{\boldsymbol{Y}_2} & \underset{n\times(p-m)}{\boldsymbol{Y}_3} \end{pmatrix}$. If it is known that $\boldsymbol{\Gamma}_{31} = \boldsymbol{0}$, then

$$\begin{aligned}
\boldsymbol{A} \in \mathcal{R}(\boldsymbol{\Gamma M}) &\iff \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Gamma}_{11} & \boldsymbol{0} \\ \boldsymbol{\Gamma}_{21} & \boldsymbol{\Gamma}_{22} \\ \boldsymbol{0} & \boldsymbol{\Gamma}_{32} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Lambda}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\Gamma}_{11}' & \boldsymbol{\Gamma}_{21}' & \boldsymbol{0}' \\ \boldsymbol{0}' & \boldsymbol{\Gamma}_{22}' & \boldsymbol{\Gamma}_{32}' \end{pmatrix} \\
&\iff \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Gamma}_{11}\boldsymbol{\Lambda}_1\boldsymbol{\Gamma}_{11}' & \boldsymbol{\Gamma}_{11}\boldsymbol{\Lambda}_1\boldsymbol{\Gamma}_{21}' & \boldsymbol{0} \\ \boldsymbol{\Gamma}_{21}\boldsymbol{\Lambda}_1\boldsymbol{\Gamma}_{11}' & \boldsymbol{\Gamma}_{21}\boldsymbol{\Lambda}_1\boldsymbol{\Gamma}_{21}' + \boldsymbol{\Gamma}_{22}\boldsymbol{\Lambda}_2\boldsymbol{\Gamma}_{22}' & \boldsymbol{\Gamma}_{22}\boldsymbol{\Lambda}_2\boldsymbol{\Gamma}_{32}' \\ \boldsymbol{0} & \boldsymbol{\Gamma}_{32}\boldsymbol{\Lambda}_2\boldsymbol{\Gamma}_{22}' & \boldsymbol{\Gamma}_{32}\boldsymbol{\Lambda}_2\boldsymbol{\Gamma}_{32}' \end{pmatrix} \\
&\iff \boldsymbol{Y}_1 \perp\!\!\!\perp \boldsymbol{Y}_3.
\end{aligned}$$

Accordingly, a test of independence between $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_3$ is a test of $\mathrm{H}_0\colon \boldsymbol{A} \in \mathcal{R}(\boldsymbol{\Gamma M})$. if $\boldsymbol{\Gamma}_{31} = \boldsymbol{0}$, $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{I}_q \\ \boldsymbol{0} \end{pmatrix}$, and $\boldsymbol{M} = \begin{pmatrix} \boldsymbol{I}_m \\ \boldsymbol{0} \end{pmatrix}$. Partition $\boldsymbol{S}$ as

$$\boldsymbol{S} = \begin{pmatrix} \underset{q\times q}{\boldsymbol{S}_{11}} & \underset{q\times(m-q)}{\boldsymbol{S}_{12}} & \underset{q\times(p-m)}{\boldsymbol{S}_{13}} \\ \underset{(m-q)\times q}{\boldsymbol{S}_{21}} & \underset{(m-q)\times(m-q)}{\boldsymbol{S}_{22}} & \underset{(m-q)\times(p-m)}{\boldsymbol{S}_{23}} \\ \underset{(p-m)\times q}{\boldsymbol{S}_{31}} & \underset{(p-m)\times(m-q)}{\boldsymbol{S}_{32}} & \underset{(p-m)\times(p-m)}{\boldsymbol{S}_{33}} \end{pmatrix}.$$

The test statistic is

$$\Lambda_w = \frac{\begin{vmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{13} \\ \boldsymbol{S}_{31} & \boldsymbol{S}_{33} \end{vmatrix}}{|\boldsymbol{S}_{11}||\boldsymbol{S}_{33}|},$$

which is distributed as Wilk's $\boldsymbol{\Lambda}$, $\boldsymbol{\Lambda}_{q,k,n-k}$, where $k = p - m$. An approximate F-statistic (Rencher, 1995) is given by

$$F = \frac{(1 - \Lambda_w^{1/t})\, df_2}{\Lambda_w^{1/t}\, df_1}, \tag{5.1}$$

which has an approximate F-distribution with df_1 and df_2 degrees of freedom, where

$$\begin{aligned} df_1 &= qk, \\ df_2 &= wt - \frac{1}{2}qk + 1, \\ w &= n - \frac{1}{2}(q + k + 1), \\ t &= \sqrt{\frac{q^2k^2 - 4}{q^2 + k^2 - 5}}, \text{ and} \end{aligned}$$

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 $M = -s\,\log(\mathbf{\Lambda}_{q,k,n-k})$, where $s = (n-k) - \frac{1}{2}(q - k + 1)$. The error is of the order $O\left(N^{-6}\right)$. Box (1949) provided the coefficients to be used in term of $O\left(N^{-6}\right)$. The cdf of M can be written as

$$\begin{aligned} \mathrm{P}(M \leq m_0) &= P_{qk} + \frac{\gamma_2}{s^2}\left(P_{qk+4} - P_{qk}\right) \\ &+ \frac{1}{s^4}\left[\gamma_4\left(P_{qk+8} - P_{qk}\right) - \gamma_2^2\left(P_{qk+4} - P_{qk}\right)\right] \\ &+ \frac{1}{s^6}\left[\gamma_6\left(P_{qk+12} - P_{qk}\right) - \frac{\gamma_2^3}{2}\left(P_{qk+8} - P_{qk+4}\right)\right] + O\left(N^{-8}\right), \end{aligned}$$

where P_f is $\mathrm{P}\left(\chi_f^2 \leq m_0\right)$,

$$\begin{aligned} \gamma_2 &= \alpha_2 = \frac{qk}{48}\left(q^2 + k^2 - 5\right), \\ \gamma_4 &= \frac{\alpha_2^2}{2} + \frac{qk}{1920}\left\{3q^4 + 3k^4 + 10q^2k^2 - 50\left(q^2 + k^2\right) + 159\right\}, \text{ and} \\ \gamma_6 &= \frac{\alpha_2^3}{6} + \frac{qk}{16,128}\left\{\left(q^6 + k^6\right) - 105\left(q^4 + k^4\right) + 1,113\left(q^2 + k^2\right)\right. \\ &\left. + \left(21q^2 - 350 + 21k^2\right)q^2k^2 - 2,995\right\}. \end{aligned}$$

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 2×2 table of the number of observations can be written as follows.

		F		
		1	0	Total
L	1	n_{11}	n_{12}	$n_{1\cdot}$
	0	n_{21}	n_{22}	$n_{2\cdot}$
	Total	$n_{\cdot 1}$	$n_{\cdot 2}$	N

Let P_{ij} be the probability of an observation in the i^{th} row and the j^{th} column. Let α and α^* be the sizes of the F and L tests, respectively. Because the comparison test is exact (or nearly so), it is known that $\mathrm{E}(n_{\cdot 1}) = \alpha N$, whereas the expectation of $n_{1\cdot}$ is unknown. The likelihood function, conditional on $n_{\cdot 1}$, can be written as

$$L(P|n_{\cdot 1}) = \binom{n_{\cdot 1}}{n_{11}} \binom{n_{\cdot 2}}{n_{12}} \left(\frac{P_{11}}{\alpha}\right)^{n_{11}} \left(1 - \frac{P_{11}}{\alpha}\right)^{n_{\cdot 1} - n_{11}} \left(\frac{P_{22}}{1-\alpha}\right)^{n_{\cdot 2} - n_{12}} \left(1 - \frac{P_{22}}{1-\alpha}\right)^{n_{12}},$$

where $\alpha = P_{11} + P_{21}$ and $1 - \alpha = P_{12} + P_{22}$.

Then, the loglikelihood function, conditional on $n_{\cdot 1}$, is

$$\begin{aligned}\ell(P|n_{\cdot 1}) &= n_{11}\ln\left(\frac{P_{11}}{\alpha}\right) + (n_{\cdot 1}-n_{11})\ln\left(1-\frac{P_{11}}{\alpha}\right) + (n_{\cdot 2}-n_{12})\ln\left(\frac{P_{22}}{1-\alpha}\right)\\ &\quad + n_{12}\ln\left(1-\frac{P_{22}}{1-\alpha}\right) + \text{constants.}\end{aligned}$$

The MLE of P_{11}, $\widehat{P}_{11}$, can be obtained by solving

$$\frac{\partial \ell(P|n_{\cdot 1})}{\partial P_{11}} = 0.$$

Therefore,

$$\widehat{P}_{11} = \frac{n_{11}}{n_{\cdot 1}}\alpha.$$

Similarly, by solving $\frac{\partial \ell(P|n_{\cdot 1})}{\partial P_{12}} = 0$, the MLE of P_{12}, $\widehat{P}_{12}$ is

$$\widehat{P}_{12} = \frac{n_{12}}{n_{\cdot 2}}(1-\alpha).$$

Accordingly, an estimator of α^*, say $\widetilde{\alpha}$, is

$$\widetilde{\alpha} = \widehat{P}_{11} + \widehat{P}_{12} = \frac{n_{11}}{n_{\cdot 1}}\alpha + \frac{n_{12}}{n_{\cdot 2}}(1-\alpha).$$

The expectation and variance of $\widetilde{\alpha}$, conditional on $n_{\cdot 1}$, are

$$\begin{aligned}\mathrm{E}\left(\widetilde{\alpha}|n_{\cdot 1}\right) &= P_{11}+P_{12} = \alpha^* \quad \text{and}\\ \mathrm{var}\left(\widetilde{\alpha}|n_{\cdot 1}\right) &= \frac{\alpha}{n_{\cdot 1}}P_{11}\left(1-\frac{P_{11}}{\alpha}\right) + \frac{1-\alpha}{n_{\cdot 2}}P_{12}\left(1-\frac{P_{12}}{1-\alpha}\right).\end{aligned}$$

Accordingly,

$$\widehat{\mathrm{var}}\left(\widetilde{\alpha}\right) = \mathrm{E}\left[\mathrm{var}\left(\widetilde{\alpha}|n_{\cdot 1}\right)\right].$$

Because $n_{.1} \sim B(N, \alpha)$ and $n_{.2} \sim B(N, 1-\alpha)$, then using

$$\begin{aligned} \mathrm{E}\left(n_{.1}^{-1}\right) &\approx \frac{1}{\alpha N} \quad \text{and} \\ \mathrm{E}\left(n_{.2}^{-1}\right) &\approx \frac{1}{(1-\alpha)N} \end{aligned}$$

yields

$$\widehat{\mathrm{var}}\left(\widetilde{\alpha}\right) \approx \frac{\alpha^*\left(1-\alpha^*\right)}{N}\left(1-\varphi^2\right),$$

where $\alpha^* = P_{11} + P_{12}$ and

$$\varphi^2 = \frac{\left(n_{11}n_{22} - n_{12}n_{21}\right)^2}{n_{.1}n_{.2}n_{1.}n_{2.}}.$$

Let $\hat{\alpha} = \frac{n_{1.}}{N}$ be the usual estimator of α^*. Then,

$$\widehat{\mathrm{var}}\left(\hat{\alpha}\right) = \frac{\alpha^*(1-\alpha^*)}{N}.$$

The relative efficiency of $\widetilde{\alpha}$ to $\hat{\alpha}$ is given by

$$\frac{\widehat{\mathrm{var}}\left(\hat{\alpha}\right)}{\widehat{\mathrm{var}}\left(\widetilde{\alpha}\right)} = \frac{1}{(1-\varphi^2)}.$$

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:

$$\begin{pmatrix} \hat{\alpha} \\ \widetilde{\alpha} \\ N^* \end{pmatrix}$$

for a test with nominal $\alpha = 0.05$.

Table 5. Probabilities of Type I error when $(p, k, q) = (10, 2, 4)$ and $\lambda = (\mathbf{1}_8', 10, 25)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
20	0.0610	0.0440	0.0470	0.2690	0.0700	0.1170
	0.0625	0.0452		0.2713	0.0728	0.1198
	1263	1205		1154	2479	1592
30	0.0820	0.0540	0.0590	0.1780	0.0750	0.0970
	0.0763	0.0497		0.1701	0.0662	0.0884
	1420	1339		1407	4410	2402
50	0.0560	0.0480	0.0520	0.1100	0.0580	0.0620
	0.0546	0.0468		0.1081	0.0560	0.0600
	1895	1629		1797	9164	5877
75	0.0510	0.0430	0.0450	0.0860	0.0560	0.0590
	0.0546	0.0465		0.0908	0.0609	0.0639
	1829	2045		2003	4861	4024

Table 6. Probabilities of Type I error when $(p, k, q) = (10, 2, 4)$ and $\lambda = (\mathbf{1}_8', 5, 10)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
20	0.0680	0.0410	0.0470	0.3050	0.0470	0.1190
	0.0695	0.0421		0.3072	0.0491	0.1216
	1201	1198		1126	1896	1494
30	0.0750	0.0490	0.0590	0.2050	0.0610	0.0990
	0.0694	0.0441		0.1974	0.0535	0.0907
	1450	1539		1321	3096	2120
50	0.0570	0.0460	0.0520	0.1170	0.0640	0.0730
	0.0556	0.0447		0.1151	0.0621	0.0710
	1863	1966		1706	4340	3295
75	0.0560	0.0490	0.0510	0.0960	0.0620	0.0630
	0.0553	0.0483		0.0950	0.0611	0.0621
	1837	1881		2024	3944	3756

Table 7. Probabilities of Type I error when $(p, k, q) = (10, 2, 4)$ and $\lambda = (\mathbf{1}'_8, 2, 10)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
20	0.1520	0.0460	0.0550	0.5460	0.0320	0.1920
	0.1495	0.0441		0.5437	0.0311	0.1894
	1108	1199		1046	1056	1103
30	0.1200	0.0490	0.0580	0.4140	0.0370	0.1530
	0.1156	0.0458		0.4093	0.0351	0.1481
	1185	1238		1083	1092	1183
50	0.0910	0.0520	0.0520	0.2630	0.0550	0.1080
	0.0898	0.0511		0.2614	0.0541	0.1066
	1289	1257		1181	1236	1306
75	0.0670	0.0420	0.0450	0.1760	0.0580	0.0770
	0.0703	0.0447		0.1802	0.0610	0.0808
	1411	1446		1264	1378	1530

Table 8. Probabilities of Type I error when $(p, k, q) = (10, 2, 4)$ and $\lambda = (\mathbf{1}'_8, 1.5, 10)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
20	0.1690	0.0390	0.0470	0.6550	0.0400	0.2530
	0.1701	0.0395		0.6560	0.0401	0.2540
	1048	1031		1023	1001	1027
30	0.1560	0.0490	0.0440	0.5680	0.0620	0.2200
	0.1590	0.0513		0.5706	0.0630	0.2226
	1087	1146		1032	1022	1048
50	0.1350	0.0450	0.0470	0.4500	0.0750	0.1860
	0.1366	0.0457		0.4516	0.0758	0.1873
	1119	1065		1054	1052	1057
75	0.1160	0.0450	0.0440	0.3860	0.0740	0.1400
	0.1194	0.0474		0.3893	0.0762	0.1427
	1152	1190		1056	1094	1075

Table 9. Probabilities of Type I error when $(p,k,q) = (10,2,1)$ and $\lambda = (\mathbf{1}'_8, 10, 25)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
20	0.0660	0.0480	0.0470	0.0870	0.0600	0.0650
	0.0681	0.0498		0.0899	0.0630	0.0679
	1528	1537		2072	4398	3441
30	0.0640	0.0530	0.0480	0.0760	0.0540	0.0560
	0.0655	0.0545		0.0779	0.0560	0.0580
	1804	1953		2584	8568	6664
50	0.0530	0.0500	0.0510	0.0680	0.0560	0.0560
	0.0522	0.0492		0.0670	0.0550	0.0550
	2340	2369		3796	7614	7614
75	0.0520	0.0500	0.0500	0.0580	0.0520	0.0520
	0.0520	0.0500		0.0580	0.0520	0.0520
	2967	3241		6887	24700	24700

Table 10. Probabilities of Type I error when $(p,k,q) = (10,2,1)$ and $\lambda = (\mathbf{1}'_8, 5, 10)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
20	0.0590	0.0400	0.0530	0.0860	0.0580	0.0620
	0.0575	0.0388		0.0833	0.0555	0.0595
	1316	1258		2098	2785	2476
30	0.0650	0.0550	0.0570	0.0700	0.0540	0.0540
	0.0598	0.0502		0.0635	0.0480	0.0480
	1988	1947		3505	4310	4310
50	0.0560	0.0540	0.0490	0.0650	0.0580	0.0580
	0.0568	0.0548		0.0660	0.0590	0.0590
	2670	2617		3863	5024	5024
75	0.0560	0.0490	0.0520	0.0580	0.0560	0.0560
	0.0542	0.0474		0.0561	0.0541	0.0541
	3935	2806		6890	6704	6704

Table 11. Probabilities of Type I error when $(p, k, q) = (10, 2, 1)$ and $\lambda = (\mathbf{1}'_8, 2, 10)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
20	0.0970	0.0560	0.0530	0.1250	0.0730	0.0790
	0.0960	0.0552		0.1231	0.0717	0.0775
	1069	1080		1236	1168	1202
30	0.0890	0.0580	0.0470	0.1020	0.0660	0.0710
	0.0909	0.0594		0.1044	0.0675	0.0727
	1270	1226		1432	1234	1286
50	0.0790	0.0540	0.0560	0.0780	0.0710	0.0720
	0.0751	0.0508		0.0727	0.0665	0.0674
	1451	1408		2336	1849	1905
75	0.0570	0.0480	0.0510	0.0630	0.0560	0.0580
	0.0563	0.0473		0.0622	0.0553	0.0572
	2003	1919		2393	2043	2208

Table 12. Probabilities of Type I error when $(p, k, q) = (10, 2, 1)$ and $\lambda = (\mathbf{1}'_8, 1.5, 10)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
20	0.1210	0.0710	0.0520	0.1520	0.0850	0.0940
	0.1204	0.0705		0.1509	0.0844	0.0934
	1049	1040		1130	1050	1057
30	0.1220	0.0700	0.0530	0.1270	0.0880	0.1050
	0.1205	0.0686		0.1252	0.0870	0.1038
	1138	1199		1197	1071	1087
50	0.1080	0.0740	0.0490	0.1220	0.0920	0.089
	0.1086	0.0745		0.1226	0.0924	0.0894
	1224	1206		1203	1085	1101
75	0.0910	0.0640	0.0400	0.0920	0.0670	0.0660
	0.0971	0.0698		0.0989	0.0728	0.0716
	1207	1280		1275	1261	1238

Table 13. Probabilities of Type I error when $(p, k, q) = (10, 4, 2)$ and $\lambda = (\mathbf{1}_6', 10, 10, 10, 25)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
20	0.1350	0.0610	0.0480	0.2910	0.0700	0.1240
	0.1362	0.0620		0.2925	0.0718	0.1258
	1151	1228		1140	2397	1553
30	0.0870	0.0530	0.0430	0.1850	0.0550	0.0750
	0.0920	0.0570		0.1910	0.0614	0.0818
	1354	1374		1246	2960	2243
50	0.0710	0.0400	0.0390	0.1150	0.0500	0.0580
	0.0787	0.0460		0.1251	0.0606	0.0688
	1385	1409		1454	3703	2933
75	0.0640	0.0450	0.0400	0.0920	0.0490	0.0510
	0.0717	0.0521		0.1015	0.0586	0.0609
	1604	1812		1698	4286	4450

Table 14. Probabilities of Type I error when $(p, k, q) = (10, 4, 2)$ and $\lambda = (\mathbf{1}_6', 5, 5, 10, 10)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
20	0.1620	0.0630	0.0480	0.3380	0.0620	0.1400
	0.1631	0.0639		0.3394	0.0635	0.1418
	1124	1173		1109	1857	1418
30	0.0950	0.0490	0.0430	0.2130	0.0510	0.0810
	0.0999	0.0522		0.2188	0.0566	0.0870
	1309	1228		1199	2175	1700
50	0.0710	0.0460	0.0390	0.1290	0.0540	0.0670
	0.0784	0.0519		0.1390	0.0639	0.0774
	1346	1330		1377	2500	2148
75	0.0650	0.0490	0.0400	0.1000	0.0510	0.0530
	0.0727	0.0558		0.1094	0.0606	0.0626
	1588	1610		1600	3773	3398

Table 15. Probabilities of Type I error when $(p,k,q) = (10,4,2)$ and $\lambda = (\mathbf{1}_6', 2, 5, 10, 10)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
20	0.2140	0.0650	0.0440	0.4570	0.0820	0.1810
	0.2171	0.0670		0.4603	0.0842	0.1844
	1070	1084		1052	1080	1102
30	0.1340	0.0550	0.0460	0.3230	0.0630	0.1170
	0.1365	0.0568		0.3258	0.0648	0.1199
	1179	1199		1112	1184	1282
50	0.0820	0.0430	0.0390	0.1870	0.0580	0.0890
	0.0884	0.0478		0.1963	0.0644	0.0965
	1202	1208		1214	1300	1273
75	0.0850	0.0590	0.0500	0.1520	0.0700	0.0840
	0.0850	0.0590		0.1520	0.0700	0.0840
	1344	1520		1415	2018	1764

Table 16. Probabilities of Type I error when $(p,k,q) = (10,4,2)$ and $\lambda = (\mathbf{1}_6', 1.5, 5, 10, 10)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
20	0.2320	0.0650	0.0470	0.5070	0.0760	0.1980
	0.2336	0.0661		0.5085	0.0769	0.1994
	1078	1103		1045	1060	1064
30	0.1610	0.0640	0.0450	0.3940	0.0890	0.1570
	0.1642	0.0661		0.3972	0.0911	0.1599
	1153	1146		1078	1102	1122
50	0.1260	0.0590	0.0630	0.3280	0.1070	0.1420
	0.1164	0.0517		0.3192	0.1020	0.1366
	1425	1510		1151	1106	1095
75	0.0940	0.0500	0.0400	0.2190	0.0750	0.1070
	0.1011	0.0547		0.2269	0.0786	0.1124
	1293	1216		1161	1079	1132

Table 17. Probabilities of Type I error when $(p, k, q) = (10, 4, 5)$ and $\lambda = (\mathbf{1}_6', 10, 10, 10, 25)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
20	0.1560	0.0740	0.0800	0.7100	0.0840	0.3190
	0.1370	0.0593		0.7005	0.0590	0.2968
	1287	1347		1036	2968	1227
30	0.1120	0.0600	0.0650	0.4460	0.0890	0.1770
	0.1019	0.0518		0.4371	0.0746	0.1638
	1378	1469		1094	3202	1477
50	0.0750	0.0550	0.0500	0.2470	0.0660	0.0890
	0.0750	0.0550		0.2470	0.0660	0.0890
	1605	1905		1191	3918	2167
75	0.0660	0.0510	0.0430	0.1750	0.0570	0.0700
	0.0720	0.0568		0.1810	0.0639	0.0768
	1950	2346		1268	3896	2481

Table 18. Probabilities of Type I error when $(p, k, q) = (10, 4, 5)$ and $\lambda = (\mathbf{1}_6', 5, 5, 10, 10)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
20	0.1560	0.0460	0.0710	0.7980	0.0410	0.3500
	0.1433	0.0397		0.7934	0.0333	0.3356
	1224	1155		1019	1288	1157
30	0.1170	0.0560	0.0640	0.5180	0.0670	0.1880
	0.1078	0.0489		0.5108	0.0568	0.1759
	1337	1412		1067	2041	1419
50	0.0760	0.0500	0.0500	0.2890	0.0650	0.0980
	0.0760	0.0500		0.2890	0.0650	0.0980
	1591	1701		1148	2967	1939
75	0.0620	0.0480	0.0440	0.2070	0.0590	0.0730
	0.0670	0.0525		0.2120	0.0649	0.0788
	2034	2123		1214	3760	2406

Table 19. Probabilities of Type I error when $(p, k, q) = (10, 4, 5)$ and $\lambda = (\mathbf{1}_6', 2, 5, 10, 10)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
20	0.1800	0.0340	0.0790	0.9010	0.0460	0.4900
	0.1637	0.0271		0.8979	0.0387	0.4763
	1183	1143		1009	1118	1069
30	0.1340	0.0430	0.0640	0.7620	0.0730	0.3220
	0.1243	0.0385		0.7584	0.0680	0.3123
	1327	1176		1021	1126	1150
50	0.0930	0.0460	0.0500	0.4940	0.0740	0.1590
	0.0930	0.0460		0.4940	0.0740	0.1590
	1492	1413		1057	1430	1295
75	0.0690	0.0470	0.0460	0.3230	0.0780	0.1200
	0.0724	0.0500		0.3258	0.0810	0.1233
	1927	2215		1112	1547	1402

Table 20. Probabilities of Type I error when $(p, k, q) = (10, 4, 5)$ and $\lambda = (\mathbf{1}_6', 1.5, 5, 10, 10)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
20	0.2020	0.0250	0.0630	0.9450	0.0560	0.5740
	0.1936	0.0220		0.9442	0.0541	0.5694
	1182	1143		1003	1023	1031
30	0.1620	0.0260	0.0680	0.8490	0.0990	0.4540
	0.1515	0.0214		0.8461	0.0955	0.4454
	1189	1196		1013	1027	1061
50	0.1080	0.0220	0.0490	0.7180	0.1260	0.3170
	0.1087	0.0223		0.7183	0.1264	0.3175
	1338	1285		1020	1066	1057
75	0.1040	0.0420	0.0520	0.5620	0.1330	0.2540
	0.1022	0.0408		0.5611	0.1321	0.2527
	1763	1720		1044	1103	1121

Table 21. Probabilities of Type I error when $(p, k, q) = (15, 2, 1)$ and $\lambda = (\mathbf{1}'_{13}, 15, 35)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
25	0.0610	0.0470	0.0580	0.0650	0.0450	0.0470
	0.0606	0.0468		0.0647	0.0455	0.0465
	1001	1000		1001	1004	1004
35	0.0490	0.0390	0.0520	0.0590	0.0410	0.0420
	0.0489	0.0389		0.0580	0.0409	0.0419
	1002	1004		1001	1001	1001
50	0.0380	0.0310	0.0460	0.0500	0.0370	0.0370
	0.0381	0.0311		0.0503	0.0373	0.0373
	1001	1000		1006	1007	1007
75	0.0400	0.0370	0.0560	0.0410	0.0380	0.0380
	0.0396	0.0366		0.0405	0.0374	0.0374
	1007	1008		1010	1012	1012

Table 22. Probabilities of Type I error when $(p, k, q) = (15, 2, 1)$ and $\lambda = (\mathbf{1}'_{13}, 10, 15)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
25	0.0620	0.0460	0.0580	0.0610	0.0480	0.0490
	0.0616	0.0458		0.0606	0.0475	0.0485
	1001	1000		1001	1004	1003
35	0.0470	0.0380	0.0520	0.0580	0.0440	0.0450
	0.0469	0.0370		0.0579	0.0439	0.0449
	1003	1005		1001	1001	1001
50	0.0370	0.0310	0.0460	0.0480	0.0380	0.0380
	0.0461	0.0394		0.0618	0.0520	0.0520
	1001	1001		1004	1006	1006
75	0.0560	0.0480	0.0460	0.0680	0.0600	0.0600
	0.0563	0.0482		0.0684	0.0602	0.0602
	1005	1001		1005	1002	1002

Table 23. Probabilities of Type I error when $(p, k, q) = (15, 2, 1)$ and $\lambda = (\mathbf{1}'_{13}, 2, 15)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
25	0.1000	0.0740	0.0580	0.0830	0.0630	0.0650
	0.0995	0.0735		0.0822	0.0625	0.0644
	1002	1003		1006	1003	1005
35	0.0800	0.0570	0.0520	0.0720	0.0540	0.0570
	0.0798	0.0568		0.0719	0.0539	0.0569
	1009	1013		1000	1001	1003
50	0.0900	0.0650	0.0450	0.0670	0.0640	0.0660
	0.0901	0.0650		0.0668	0.0638	0.0658
	1000	1000		1001	1001	1001
75	0.0660	0.0550	0.0560	0.0470	0.0400	0.0410
	0.0655	0.0548		0.0466	0.0397	0.0407
	1005	1001		1004	1003	1003

Table 24. Probabilities of Type I error when $(p, k, q) = (15, 2, 1)$ and $\lambda = (\mathbf{1}'_{13}, 1.5, 15)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^*	T^{**}
25	0.1230	0.0850	0.0580	0.0890	0.0650	0.0710
	0.1221	0.0844		0.0881	0.0645	0.0701
	1005	1003		1007	1003	1009
35	0.1000	0.0620	0.0520	0.0910	0.0730	0.0790
	0.0998	0.0618		0.0909	0.0728	0.0788
	1007	1008		1002	1005	1004
50	0.1060	0.0780	0.0530	0.0860	0.0860	0.0820
	0.1059	0.0778		0.0859	0.0860	0.0820
	1001	1002		1001	1000	1000
75	0.0930	0.0640	0.0560	0.0740	0.0620	0.0700
	0.0923	0.0635		0.0736	0.0616	0.0694
	1007	1006		1004	1004	1007

Table 25. Probabilities of Type I error when $(p, k, q) = (15, 4, 2)$ and $\lambda = (\mathbf{1}'_{11}, 15, 15, 15, 35)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^{*}	T^{**}
25	0.1360	0.0700	0.1520	0.2080	0.0690	0.0930
	0.1247	0.0626		0.1919	0.0623	0.0820
	1013	1010		1020	1008	1018
35	0.1050	0.0640	0.0830	0.1440	0.0710	0.0860
	0.1005	0.0602		0.1379	0.0675	0.0809
	1015	1016		1021	1013	1024
50	0.0740	0.0440	0.0660	0.0113	0.0580	0.0620
	0.0722	0.0419		0.0110	0.0564	0.0602
	1012	1025		1009	1011	1013
75	0.0770	0.0630	0.0640	0.0910	0.0610	0.0620
	0.0744	0.0604		0.0882	0.0586	0.0597
	1024	1035		1030	1030	1029

Table 26. Probabilities of Type I error when $(p, k, q) = (15, 4, 2)$ and $\lambda = (\mathbf{1}'_{11}, 10, 10, 15, 15)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^{*}	T^{**}
25	0.1350	0.0710	0.1520	0.2100	0.0710	0.0940
	0.1243	0.0629		0.1933	0.0637	0.0839
	1012	1012		1021	1010	1014
35	0.1060	0.0680	0.0830	0.1530	0.0710	0.0820
	0.1016	0.0639		0.1468	0.0671	0.0776
	1014	1018		1021	1016	1018
50	0.0750	0.0440	0.0660	0.1120	0.0580	0.0630
	0.0729	0.0419		0.1100	0.0564	0.0612
	1015	1025		1009	1011	1013
75	0.0740	0.0620	0.0640	0.0880	0.0610	0.0610
	0.0714	0.0594		0.0853	0.0586	0.0586
	1031	1036		1027	1030	1030

Table 27. Probabilities of Type I error when $(p,k,q) = (15,4,2)$ and $\lambda = (\mathbf{1}'_{11}, 2, 10, 15, 15)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^{*}	T^{**}
25	0.1970	0.0990	0.1520	0.2950	0.0550	0.1110
	0.1843	0.0880		0.2830	0.0490	0.1014
	1012	1017		1008	1008	1011
35	0.1450	0.0880	0.0830	0.2320	0.0540	0.0820
	0.1407	0.0851		0.2273	0.0520	0.0784
	1010	1007		1008	1005	1011
50	0.1030	0.0590	0.0660	0.1650	0.0510	0.0740
	0.1009	0.0572		0.1619	0.0493	0.0724
	1012	1015		1071	1015	1008
75	0.0960	0.0600	0.0640	0.1240	0.0650	0.0690
	0.0937	0.0580		0.1209	0.0636	0.0672
	1019	1020		1026	1009	1015

Table 28. Probabilities of Type I error when $(p,k,q) = (15,4,2)$ and $\lambda = (\mathbf{1}'_{11}, 1.5, 10, 15, 15)'$.

n	LRT	$BART$	F	$T_{k,q}$	T^{*}	T^{**}
25	0.2290	0.1160	0.1520	0.3460	0.0650	0.1300
	0.2162	0.1054		0.3386	0.0617	0.1227
	1011	1013		1003	1002	1005
35	0.1790	0.0870	0.0830	0.2970	0.0660	0.1080
	0.1746	0.0854		0.2912	0.0653	0.1054
	1009	1002		1011	1000	1005
50	0.1230	0.0700	0.0660	0.2340	0.0760	0.0910
	0.1210	0.0683		0.2323	0.0752	0.0894
	1009	1010		1003	1002	1007
75	0.1240	0.0780	0.0640	0.1940	0.0750	0.0980
	0.1216	0.0761		0.1911	0.0743	0.0967
	1015	1015		1017	1002	1006

The following tables report power of the tests when $p = 10,\ k = 2, 4$, and $q = 1, 2$. Figues compare the power of the $BART$ and T^* tests both of which are controlled for Type I error.

Figure 3. Power of test when $\boldsymbol{\lambda} = (\mathbf{1}_8', 10, 25)',\ k = 2,\ q = 1$.

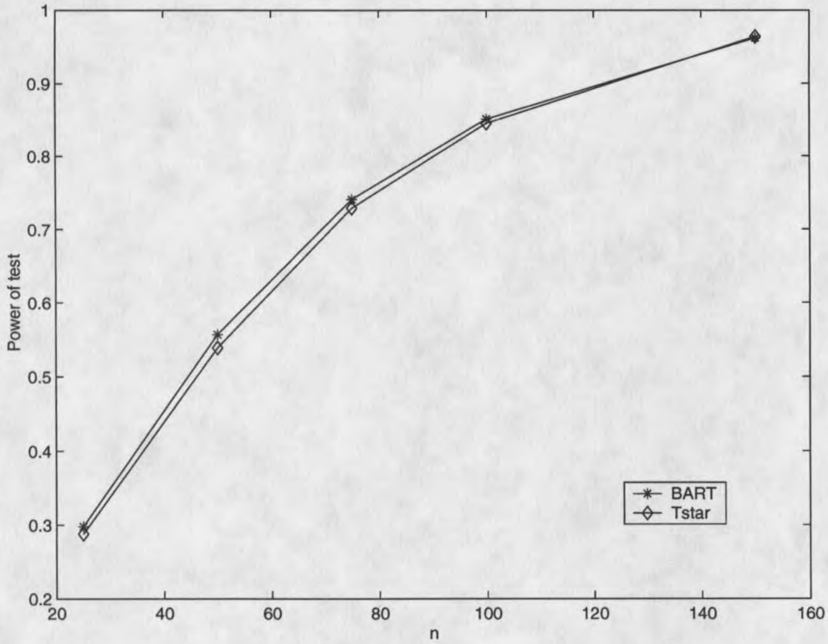


Table 29. Power of test when $\boldsymbol{\lambda} = (\mathbf{1}_8', 10, 25)',\ k = 2,\ q = 1$.

n	Power of test				
	LRT	$BART$	$T_{k,q}$	T^*	T^{**}
25	0.3330	0.2980	0.3550	0.2880	0.2960
50	0.5890	0.5570	0.5720	0.5390	0.5410
75	0.7480	0.7410	0.7540	0.7290	0.7300
100	0.8580	0.8510	0.8540	0.8450	0.8450
150	0.9620	0.9620	0.9640	0.9640	0.9640

Figure 4. Power of test when $\boldsymbol{\lambda} = (\mathbf{1}_8', 2, 10)'$, $k = 2$, $q = 1$.

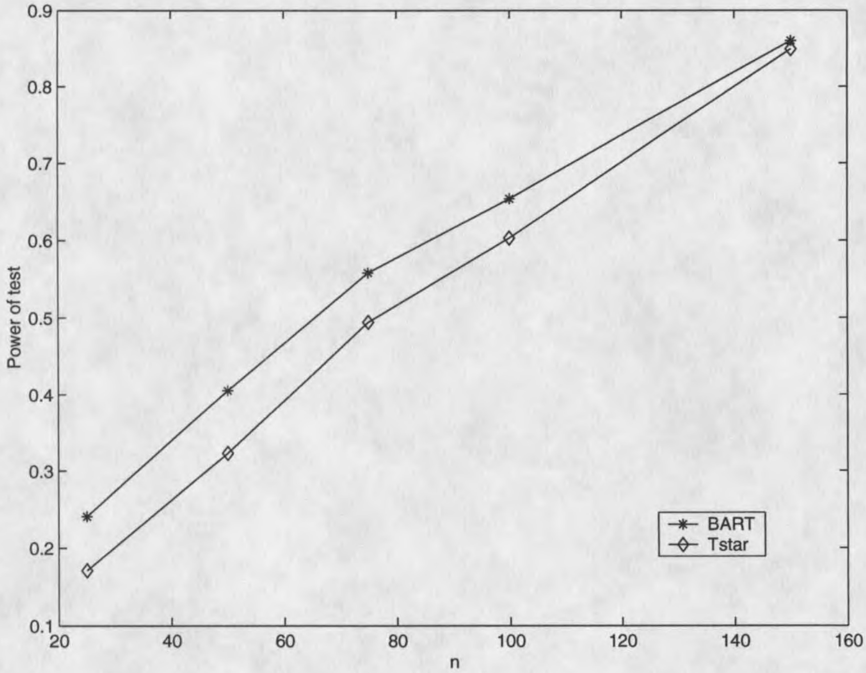


Table 30. Power of test when $\boldsymbol{\lambda} = (\mathbf{1}_8', 2, 10)'$, $k = 2$, $q = 1$.

n	Power of test				
	LRT	$BART$	$T_{k,q}$	T^*	T^{**}
25	0.3150	0.2410	0.3200	0.1710	0.2000
50	0.4580	0.4050	0.4620	0.3230	0.3640
75	0.5960	0.5580	0.5870	0.4940	0.5180
100	0.6930	0.6540	0.6610	0.6030	0.6240
150	0.8750	0.8600	0.8660	0.8490	0.8530

Figure 5. Power of test when $\boldsymbol{\lambda} = (\mathbf{1}_6', 10, 10, 10, 25)'$, $k = 4$, $q = 2$.

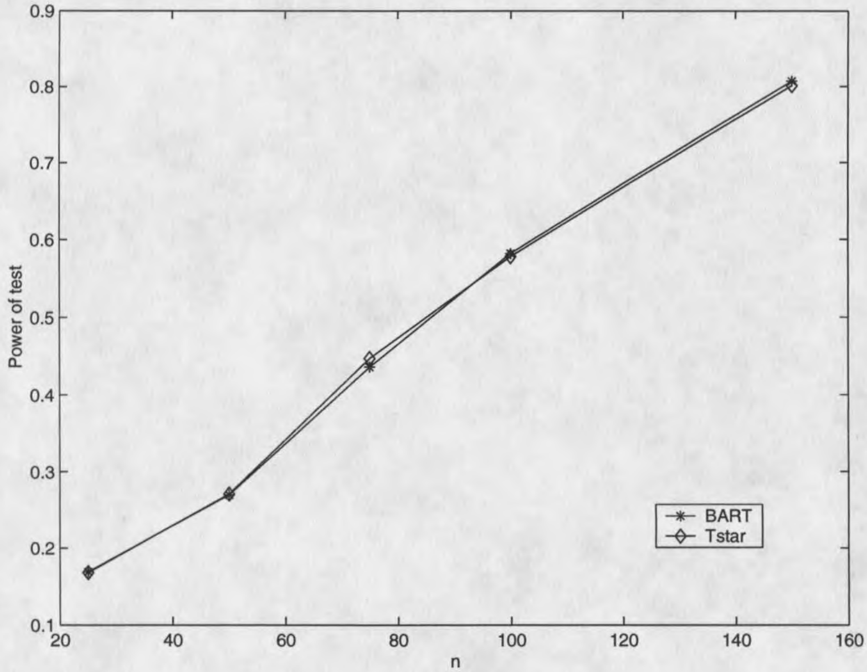


Table 31. Power of test when $\boldsymbol{\lambda} = (\mathbf{1}_6', 10, 10, 10, 25)'$, $k = 4$, $q = 2$.

n	Power of test				
	LRT	$BART$	$T_{k,q}$	T^*	T^{**}
25	0.2750	0.1690	0.4130	0.1680	0.2330
50	0.3420	0.2700	0.4150	0.2710	0.2940
75	0.4990	0.4360	0.5370	0.4470	0.4600
100	0.6130	0.5820	0.6360	0.5780	0.5800
150	0.8230	0.8070	0.8230	0.8010	0.8040

Figure 6. Power of test when $\boldsymbol{\lambda} = (\mathbf{1}_6', 2, 5, 10, 10)'$, $k = 4$, $q = 2$.

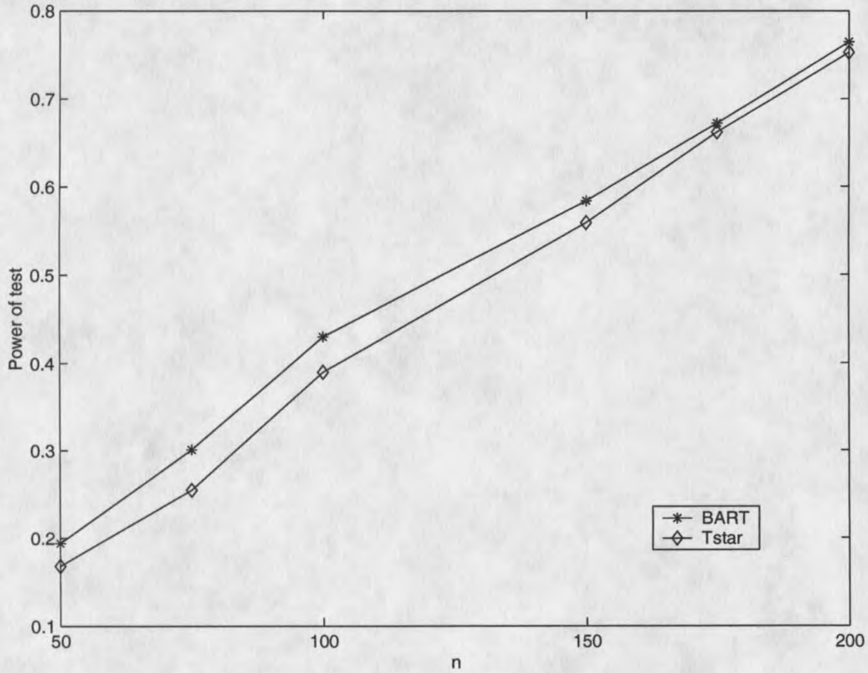


Table 32. Power of test when $\boldsymbol{\lambda} = (\mathbf{1}_6', 2, 5, 10, 10)'$, $k = 4$, $q = 2$.

n	Power of test				
	LRT	$BART$	$T_{k,q}$	T^*	T^{**}
50	0.2900	0.1940	0.4070	0.1680	0.2200
75	0.3770	0.3010	0.4570	0.2550	0.2870
100	0.4840	0.4300	0.5320	0.3900	0.4160
150	0.6240	0.5840	0.6510	0.5590	0.5710
175	0.7020	0.6720	0.7160	0.6620	0.6670
200	0.7890	0.7650	0.8000	0.7530	0.7550

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 $\frac{1}{2}p$. 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 $\frac{C_1'\boldsymbol{\lambda}}{\varphi_0}$ 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 $\frac{C_1'\boldsymbol{\lambda}}{\varphi_0}$ can be constructed by inverting the LRT and $BART$ of $\mathrm{H}_0\colon \frac{C_1'\boldsymbol{\lambda}}{\varphi_0} = c_0$ against $\mathrm{H}_a\colon \frac{C_1'\boldsymbol{\lambda}}{\varphi_0} \neq c_0$. The interval consists of all values c_0 for which $\mathrm{H}_0\colon \frac{C_1'\boldsymbol{\lambda}}{\varphi_0} = c_0$ cannot be rejected. That is, a $(1-\alpha)100\%$ CI for $\frac{C_1'\boldsymbol{\lambda}}{\varphi_0}$ can be obtained as

$$CI\left(\frac{C_1'\boldsymbol{\lambda}}{\varphi_0}\right) = \left\{c_0; Q \leq \chi^2_{1-\alpha,r}\right\},$$

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 $\mathrm{H}_0\colon \frac{C_1'\boldsymbol{\lambda}}{\varphi_0} = c_{01}$ and $\mathrm{H}_0\colon \frac{C_1'\boldsymbol{\lambda}}{\varphi_0} = c_{02}$ are each equal to α. 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, 1000 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 $\boldsymbol{C}_1$ was chosen so that $\frac{C_1'\boldsymbol{\lambda}}{\varphi_0}$ are $\frac{\sum_{i=1}^{m}\lambda_i}{\mathrm{tr}(\Sigma)}$. The nominal test size used was 0.05. Tables with the percentage coverage of confidence interval for $\frac{C_1'\boldsymbol{\lambda}}{\varphi_0}$ are shown below.

Table 33. Percentage coverage of CI when $(p, k) = (8, 2)$ and $\lambda = (\mathbf{1}_6', 10, 25)'$.

n	*LRI*	*BARI*	*ZI*
20	90.70	92.40	88.90
75	94.10	94.90	93.90

Table 34. Percentage coverage of CI when $(p, k) = (8, 2)$ and $\lambda = (\mathbf{1}_6', 2, 10)'$.

n	*LRI*	*BARI*	*ZI*
20	84.70	90.10	84.90
75	92.30	93.50	92.60

Table 35. Percentage coverage of CI when $(p, k) = (8, 4)$ and $\lambda = (\mathbf{1}_4', 10, 10, 10, 25)'$.

n	*LRI*	*BARI*	*ZI*
20	78.80	93.90	72.50
75	90.40	94.90	89.90

Table 36. Percentage coverage of CI when $(p, k) = (8, 4)$ and $\lambda = (\mathbf{1}_4', 2, 10, 10, 10)'$.

n	*LRI*	*BARI*	*ZI*
20	62.30	91.00	60.10
75	87.40	93.60	87.40

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 $\frac{C_1'\boldsymbol{\lambda}}{\varphi_0}$ 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 $\frac{\boldsymbol{C}_1'\boldsymbol{\lambda}}{\varphi_0}$ was computed, where $\boldsymbol{C}_1 = (0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 1)'$. 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 χ^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 $\frac{C_1'\lambda}{\varphi_0}$ was computed, where $C_1 = (0\ 0\ 0\ 0\ 1\ 1)'$. 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.

k	q	LRT	p-value	$BART$	p-value
1	1	0.5877	0.4433	0.6082	0.4355
1	2	2.5382	0.2811	2.5549	0.2787
1	3	8.7912	0.0322*	8.5538	0.0359*
1	4	9.7598	0.0447*	9.3310	0.0533
1	5	14.8112	0.0112*	13.9096	0.0162*
2	1	0.6040	0.7393	0.5991	0.7411
2	2	8.5104	0.0746	8.2538	0.0827
2	3	31.3926	<0.0001**	30.1616	<0.0001**

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 χ^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 $\frac{C_1'\lambda}{\varphi_0}$. Results show that the Bartlett-corrected intervals substantially improve the control over the percentage coverage of CI for $\frac{C_1'\lambda}{\varphi_0}$ 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 $\frac{C_1'\lambda}{\varphi_0}$. 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

[1] Anderson, T.W. (1963). Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics*, **10**, 203-224.

[2] Anderson, T.W. (1984). *Introduction to Multivariate Statistical Analysis*, Second edition. New York: Wiley.

[3] Bartlett, M.S. (1954). A note on the multiplying factors for various χ^2 approximations. *Journal of the Royal Statistical Society, Series B*, **16**, 296-298.

[4] Beale, E.M.L, Kendall, M.G. and Mann, D.W. (1967). The discarding of variables in multivariate analysis. *Biometrika*, **54**, 357-366.

[5] Bentler, P.M. and Yuan, K-H. (1996). Test of linear trend in eigenvalues of a covariance matrix. *The British Journal of Mathematical and Statistical Psychology*, **49**, 299-312.

[6] Boik, R.J. (1998). A local parameterization of orthogonal and semi-orthogonal matrices with applications. *J. Mult. Anal.*, **67**, 244-276.

[7] Boik, R.J. (2002a). Spectral models for covariance matrices. *Biometrika*, **89**, 159-182.

[8] Boik, R.J. (2002b). Principal components models for correlation matrices. in press.

[9] Box, G.E.P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika*, **36**, 317-346.

[10] Cattell, R.B. (1966). The scree test for the number of factors. *Journal of Multivariate Behav. Res.*, **1**, 245-276.

[11] Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika*, **31**, 33-42.

[12] Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics.* Chapman and Hall.

[13] Dawkins, B. (1989). Multivariate analysis of national track records. *The American Statistician*, **43**, 110-115.

[14] Dümbgen L. (1995). Likelihood ratio tests for principal components. *Journal of Multivariate Analysis*, **52**, 245-258.

[15] Flury B.N. (1986). An asymptotic test for redundancy of variables in the comparison of two covariance matrices. *Statistics and Probability Letters*, **4**, 123-126.

[16] Flury B.N. (1988). *Common Principal Components and Related Multivariate Models.* New York: Wiley.

[17] Fujikoshi, Y. (1985). Selection of variables in discriminant analysis and canonical correlation analysis. *Multivariate Analysis-VI, P.K. Krishnaiah, ed.*, Elsevier, 219-236.

[18] Fujikoshi, Y. (1989). Tests for redundancy of some variables in multivariate analysis. *Statistical data analysis and inference*, New York: Elsevier.

[19] Fujikoshi, Y. and Rao, C.R. (1991). Selection of covariance in the growth curve model. *Biometrika*, **78**, Issue 4, 779-785.

[20] Gleason, T.C. (1976). On redundancy in canonical analysis. *Psychological*, **83**, No.6, 1004-1006.

[21] Huang, D. and Tseng, S. (1992). A decision procedure for determining the number of components in principal component analysis. *Journal of Statistical Planning and Inference*, **30**,63-71.

[22] Jackson, J.E. and R.H. Morris (1957). An application of multivariate quality control to photographic processing. *Journal Amer. Statist. Assoc*, **52**,186-199.

[23] Jolliffe, I.T. (1972). Discarding variables in a principal component analysis, I: Artificial data. *Appl. Statist.*, **21**, 160-173.

[24] Jolliffe, I.T. (1973). Discarding variables in a principal component analysis, II: Real data. *Appl. Statist.*, **22**, 22-31.

[25] Jolliffe, I.T. (1986). *Principal Component Analysis.* New York: Springer.

[26] Jolliffe, I.T. and Cadima J. (1995). Loadings and correlations in the interpretation of principal components. *Journal of Applied Statistics*, **22**, No.2, 203-214.

[27] Kaiser, H.F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, **23**, 187-200.

[28] Lawley, D.N. (1956). Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika*, **43**, 128-136.

[29] Lazraq, A. (2001). Statistical inference concerning several redundancy indices. *Journal of Multivariate Analysis*, **79**, 71-88.

[30] Magnus, J.R. and Neudecker, H. (1979). The commutation matrix: some properties and applications. *Ann. Statist.*, **7**, 381-394.

[31] Magnus, J.R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Revised Ed. Chichester: Wiley.

[32] Mathworks, Inc. (2000). *Matlab Version 6.1.* Mathworks, Inc.

[33] McCabe, G.P. (1984). Principal variables. *Technometrics*, **26**, 137-144.

[34] Naik, D.N. and Khattree, R. (1996). Revisiting olympic track records: Some practical considerations in the principal component analysis. *The American Statistician*, **50**, 140-144.

[35] Ouellette, D.V. (1981). Schur complements and statistics. *Linear Algebra and its Applications*, **36**, 187-295.

[36] Rencher, A.C. (1995). *Methods of Multivariate Analysis.* New York: Wiley.

[37] Schott, J.R. (1991). Testing for the redundancy of variables in principal components analysis. *Statistics and Probability Letters*, **11**, 495-501.

[38] Seber, G.A.F. (1984). *Multivariate Observations.* New York: Wiley.

[39] Sen, P.K. and Singer, J.M. (1974). *Large Sample Methods in Statistics: An introduction with applications.* Chapman and Hall.

[40] Severini, T.A. (2000). *Likelihood Methods in Statistics.* Oxford University Press.

[41] Sugiyama, T. and Tong, H. (1976). On a statistic useful in dimensionality reduction in multivariate linear stochastic system. *Commun. Statist.*, **A5**, 711-721.

[42] Taylor, A.E. and Mann, W.R. (1983). *Advanced Calculus.* New York: Wiley.

[43] Tyler, D.E. (1981). Asymptotic inference for eigenvectors, *The Annals of Statistic*, **9**, 725-736.

[44] Van Den Wollenberg, A.L. (1977). Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika*, **42**, No.2, 207-219.

APPENDICES

APPENDIX A

Notation

Notation

$\mathrm{diag}(\boldsymbol{a})$	diagonal matrix having diagonal elements as a vector $\boldsymbol{a}$
det	determinant
tr	trace
exp	exponential function
dim	dimension
$\mathcal{O}(p)$	group of orthogonal $p \times p$ matrices
$\mathrm{vec}(\boldsymbol{A})$	stacking the columns of $\boldsymbol{A}$ into a single column vector
$\mathrm{dvec}(\boldsymbol{a}, p, p)$	converging a vector $\boldsymbol{a}$ to a $p \times p$ matrix
$\otimes$	direct or Kronecker product
$\oplus$	direct sum
$\odot$	the operation is to be performed element-wise
$\mathcal{R}(\boldsymbol{T})$	subspace spanned by the columns of a matrix $\boldsymbol{T}$
$\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	p-variate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$\mathrm{W}_p(n, \boldsymbol{\Sigma})$	$p \times p$ matrix variate Wishart distribution with n degrees of freedom and covariance matrix $\boldsymbol{\Sigma}$
$L(\cdot)$	likelihood function
$\ell(\cdot)$	loglikelihood function
$\lVert\boldsymbol{X}\rVert^2$	norm of $\boldsymbol{X}$ i.e $\sum_{i=1}^{p} X_i^2$
$\boldsymbol{L}$	the matrix $p^2 \times p$ of $\sum_{i=1}^{p} \left(\boldsymbol{e}_j^p \otimes \boldsymbol{e}_j^p\right) \boldsymbol{e}_j^{p\prime}$
$\boldsymbol{I}_p$	the $p \times p$ identity matrix
$\boldsymbol{e}_j^p$	the j^{th} column of $\boldsymbol{I}_p$
$\boldsymbol{I}_{(p,p)}$	the $p^2 \times p^2$ commutation matrix
$\boldsymbol{D}_G^{(i)}$	the i^{th} derivative of $\mathrm{vec}\boldsymbol{G}$ with respect to $\boldsymbol{\mu}'$ evaluated at $\boldsymbol{\mu} = \boldsymbol{0}$
$\boldsymbol{D}_\lambda^{(i)}$	the i^{th} derivative of $\boldsymbol{\lambda}$ with respect to $\boldsymbol{\varphi}'$
$\boldsymbol{D}_p$	the duplication matrix
$\boldsymbol{N}_p$	the matrix $\left(\boldsymbol{I}_{p^2} + \boldsymbol{I}_{(p,p)}\right)/2$
$\boldsymbol{I}_{\boldsymbol{\theta}}$	Fisher's information
$\bar{\boldsymbol{I}}_{\boldsymbol{\theta}}$	the average Fisher's information
$\psi(n)$	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, $\boldsymbol{I}_{(a,b)}$.

2. dup.m : Computes the duplication matrix, $\boldsymbol{D}_p$.

3. kron.m : Computes Kronecker products of three or more input matrices.

4. make_E.m : Constructs the elementary matrix, $\boldsymbol{E}_{i,\nu}$.

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 $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, $\boldsymbol{L}$, and $(\boldsymbol{I}_{p^2} - \boldsymbol{I}_{(p,p)})\boldsymbol{A}_1$.

9. solve_eta.m : Given parameter $\boldsymbol{\mu}$, solves for the implicit parameter $\boldsymbol{\eta}$ under model (2.31).

10. solve_eta1.m : Given parameter $\boldsymbol{\mu}$, solves for the implicit parameter $\boldsymbol{\eta}$ under model (2.33).

11. solve_eta2.m : Given parameter $\boldsymbol{\mu}$, solves for the implicit parameter $\boldsymbol{\eta}_2$ under model (2.66).

12. solve_tau.m : Given parameter $\boldsymbol{\varphi}$, solves for the implicit parameter $\boldsymbol{\tau}$ and computes $\boldsymbol{\xi}$ under model (2.11).

13. find_guess.m : Computes initial guesses for $\boldsymbol{\Gamma}_0$ and $\boldsymbol{\Lambda}_0$.

14. fisher_scoring1.m : Updates the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\varphi}$ under the parameterization of $\boldsymbol{\lambda}$ in (2.4).

15. fisher_scoring2.m : Updates the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\varphi}$ under the parameterization of $\boldsymbol{\lambda}$ in (2.9).

16. find_Vstar.m : Computes $\boldsymbol{V}^* = (\boldsymbol{V}_3 \;\; \boldsymbol{V}_4)$ ensured that the matrix of derivatives in (2.35) is nonsingular. This subprogram is used for estimating parameters.

17. find_Vstar_bart.m : Computes $\boldsymbol{V}^* = (\boldsymbol{V}_3 \;\; \boldsymbol{V}_4)$ 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 $\boldsymbol{\Gamma}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\Sigma}$ under model (2.31) and the parameterization of $\boldsymbol{\lambda}$ in (2.4).

19. solve_mle1.m : Computes the maximum likelihood estimates of $\boldsymbol{\Gamma}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\Sigma}$ under model (2.33) and the parameterization of $\boldsymbol{\lambda}$ in (2.4).

20. solve_mle2.m : Computes the maximum likelihood estimates of $\boldsymbol{\Gamma}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\Sigma}$ under model (2.31) and the parameterization of $\boldsymbol{\lambda}$ in (2.9).

21. Bartlett.m : Computes $\mathrm{E}\left(D(\hat{\boldsymbol{\theta}};\boldsymbol{\theta})\right)$ under model (2.31) and the parameterization of $\boldsymbol{\lambda}$ in (2.4).

22. Bartlett1.m : Computes $\mathrm{E}\left(D(\hat{\boldsymbol{\theta}};\boldsymbol{\theta})\right)$ under model (2.33) and the parameterization of $\boldsymbol{\lambda}$ in (2.4).

23. Bartlett2.m : Computes $\mathrm{E}\left(D(\hat{\boldsymbol{\theta}};\boldsymbol{\theta})\right)$ under model (2.31) and the parameterization of $\boldsymbol{\lambda}$ in (2.9).

24. Exp_Da.m : Computes $\mathrm{E}\left(D(\hat{\boldsymbol{\theta}};\boldsymbol{\theta})\right)$ 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 $\mathrm{H}_0\colon \frac{\boldsymbol{C}_1'\boldsymbol{\lambda}}{\varphi_0} = c_0$ is equal to α by minimizing (p-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 $\mathbf{\Gamma}$, $\mathbf{\Lambda}$, and $\mathbf{\Sigma}$, 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 :

<http://www.math.montana.edu/~rjboik/spectral/>.

solve_eta2.m function.

```
function [G,eta2] = solve_eta2(A,A1,A2,V3,M,mu,Gam)
%%
%% given a solution for mu, 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 columm %%
%%
 b = M'*[1:p]'; Mc = Ip ; Mc(:,b) = [];
%%
%% initial guess of eta2 %%
%%
 eta2 = vech(Ip);
 C = kron(Mc',A'*Gam);
 VG = (A1*V3*mu)+(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*mu)+(A2*eta2);
   G = dvec(VG,p,p);
 end    %% end while loop %%
```

solve_tau.m function.

```
function [Xi,tau] = solve_tau(phi,Xi,T4,V1,V2,C2)
%%
%% given a solution for phi, solve for the implicit parameter tau %%
%% initial guess of tau %%
%%
 tau = inv(V1'*V1)*V1'*Xi;
%%
%% find tau using Newton's method %%
%%
while norm(C2'*exp(T4*Xi),'fro') > 1e-8
   F = C2'*exp(T4*Xi);
   J = C2'*diag(exp(T4*Xi))*T4*V1;
   Jinv = inv(J);
   check = Jinv*F;
   tau = tau-check;
   Xi = V1*tau+V2*phi;
end %% end while loop %%
```

find_guess.m function.

```
function [Gam0,lam0] = find_guess(A,U,S,p,q,m,order)
%%
%% find initial guesses for Gamma and Lamda %%
%% using Orthogonal Rotation to Congruence %%
%% write S as its svd %%
%%
[Ghat,Lhat,Gh]= svd(S);
%%
%% obtain initial guess for lamda %%
%%
 lam0 = diag(Lhat);
%%
%% order eigenvalues from small to large and %%
%% also the corresponding eigenvectors %%
%%
if order == 2
  lam0 = lam0([p:-1:1]);
  Ghat = Ghat(:,[p:-1:1]);
  Lam0 = diag(lam0);
end;
%%
%% partition the eigenvector matrix %%
%%
 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 %%
```

fisher_scoring2.m function.

```
function [mu,newphi,check_conv] = fisher_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);  Ip2 = 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'*T3*exp(T4*Xi));
  lam = wi*phi(1)*T3*exp(T4*Xi);
  Lamda = dvec(L*lam,p,p);
  Sigma = Gam*Lamda*Gam';
  invsigma = inv(Sigma);
  F11 = 2*Np*kron(Gam*Lamda,Gam)*Dg1;
  Hp = sparse(Ip - (lam/phi(1))*onep');
  W1 = sparse(diag(exp(T4*Xi))*T4);
%%
%% D11 with respect to phi0 and phi %%
%%
  D11 = [wi*T3*exp(T4*Xi)  wi*phi(1)*Hp*T3*W1*V2];
  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);
%%
%% score function %%
%%
  Szeta = pre*vec(S-Sigma);
  delta = invihat*Szeta;
```

```
  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 %%
```

<u>find_Vstar.m function.</u>

```
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,lam,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'*invihat*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*Lamda*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 %%
```

```
%%
 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 %%
```

<u>Bartlett1.m function.</u>

```
function elrt = Bartlett1(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));
 Ipf{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));
elrt=nparam;
%%
%% Create Dg(1), Dg(2),Dg(3) %%
%%
D1_g_m = sparse(Dgs*V3);
K4 = sparse(kron3(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},permp1));
K2 = sparse(kron(D2_g_mm,D1_g_m));
K3 = sparse(K2+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_g_mmm = -A2*Dp'*K4*(K6*K3+K5);
clear K3 K4 K5 K6
%%
%% Create Dl(1),Dl(2),Dl(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}'%%
%%
kgg = sparse(kron(Gam,Gam));
kgl = sparse(kron(Gam*Lam,Gam));
kglg = sparse(kron3(Gam,vec(Lam)',Gam));
kgig = sparse(kron3(Gam,vec(Ip)',Gam));
kld = sparse(kron(L*D1_l_p,D1_g_m));

D1_{1} = sparse(tNp*kron(Gam*Lam,Gam)*D1_g_m);
D1_{2} = sparse(kgg*L*D1_l_p);
D2_{1,1} = tNp*((kglg*(-kdg1))+kgl*D2_g_mm);
D2_{2,2} = kgg*L*D2_l_pp;
D2_{2,1} = tNp*kgig*kld;
D2_{1,2} = D2_{2,1}*I_{1,2};
kd1d2 = kron(-D1_g_m,D2_g_mm);
kpd2d1 = kron(permp*D2_g_mm,D1_g_m);
DD_4 = kd1d2+2*kpd2d1*kron(I{1},N1);
D3_{1,1,1} = tNp*(kgl*D3_g_mmm+kglg*DD_4);
clear DD_4
D3_{2,2,2} = kgg*L*D3_l_ppp;
D3_{2,2,1} = tNp*kgig*kron(L*D2_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{2});
DD_3 = kron3(Ip,permp,Ip)*kdg1;
D3_{2,1,1} = tNp*((kgig*kron(L*D1_l_p,D2_g_mm))+
(kron3(vec(L*D1_l_p)',Gam,Gam)*kron(I{2},DD_3)));
clear DD_3
D3_{1,2,1}=D3_{2,1,1}*kron(I_{1,2},I{1});
D3_{1,1,2}=D3_{1,2,1}*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)
  l1=index(l);
  F1=F1+D1_{l1}*make_E(l1,nu);
end;
Itheta=sparse(F1'*V2i*F1/2);
Ithetai=inv(Itheta);
Fisher_Info=Itheta;
clear D1_{1} D1_{2}
%%
%% 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_{l1,l2}) ==0
      F2=F2+D2_{l1,l2}*kron(E1,E2);
    end;
  end
end
F11=reshape(F2,p^2*nparam,nparam);
clear D2_{1,1} 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_{l1,l2,l3}) == 0
        E3=make_E(l3,nu);
        E123=kron(E12,E3);
        F3=F3+D3_{l1,l2,l3}*E123;
```

```
      end
    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*p^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(F1dd'*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*ithetai;
zeta(5)=vec(K3s*Ndf*Ithetai2*K3s')'*ithetai/4+tmp'*Ithetai*tmp/4;
clear tmp
zeta(10)=vec(Ndf*Ithetai2)'*vec(K4s)/12+ithetai'*K4s*ithetai/12;

zeta(1)=(1/2)*ithetai'*vec(U1*kron(Ithetai*F1dd',tNp)*k3_tmp4'*F1);
zeta(2)=(1/3)*ithetai'*vec(K3s*kron(F1Ithetai,F1Ithetai)'*F1ddd);
zeta(3)=(1/4)*ithetai'*vec(U1*kron(Ithetai,tNp*kron(Sigma,Sigma))*U1');
zeta(6)=(1/2)*ithetai'*vec(K3s*Ndf*kron(Ithetai,F1Ithetai')*U1');
Tmp=K3s'*Ithetai*U1;
```

```
 zeta(7)=ithetai'*Tmp*vec(F1Ithetai)/2;
 zeta(8)=(1/6)*vec(kron(F1Ithetai,Ithetai)*Ndf)'*vec(U2);
 zeta(9)=ithetai'*U2'*vec(F1Ithetai')/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} ; l(j)*l(i)/(l(i)-l(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} = Ghat22'*inv(Sj)*Ghat22;
  for u = k+1:p
     c11 = q*l(u)/(l(u)-l(j));
     c12 = 2*Q{j}(u-k,u-k)^2*(l(u)^2*l(j)^3*(2*l(j)+3*l(u))/(l(u)-l(j))^6);
     c13 =Q{j}(u-k,u-k)*(l(u)*l(j)*(2*l(u)^3-7*l(u)^2*l(j)+4*l(u)*l(j)^2+
          l(j)^3)/(l(u)-l(j))^5);
       c1 = c1+(c11+c12+c13);
     for v = k+1:p
        if v < u
       c21 = (Q{j}(u-k,u-k)+Q{j}(v-k,v-k))*l(u)*l(v)*l(j)^2/((l(u)-l(j))^2
             *(l(v)-l(j))^2);
        deno = ((l(u)-l(j))^4*(l(v)-l(j))^4);
        c22 = 2*Q{j}(u-k,v-k)^2*l(u)*l(v)*l(j)^3*(l(u)*l(v)*l(j)-3*l(j)^3+
             (l(u)+l(v))*(l(v)*(l(j)-l(u))+l(u)*(l(j)-l(v))+l(j)^2))/deno;
        c23 = 2*Q{j}(u-k,u-k)*Q{j}(v-k,v-k)*l(u)*l(v)*l(j)^3*(l(j)^3-3*l(u)
             *l(v)*l(j)+l(u)*l(v)*(l(u)+l(v)))/deno;
         c2 = c2+ (c21+c22-c23);
        end;
     end;
  end;
end;
for i = 1:k
  for j = 1:k
     if i ~= j
        for u = k+1:p
           c31 = Q{j}(u-k,u-k)*l(u)*l(j)*l(i)*(l(u)^2-l(j)*l(i))/((l(u)
               -l(j))^2*(l(u)-l(i))^2*(l(j)-l(i)));
           c3 = c3 + c31;
        end;
     end;
  end;
end;
Ckq = 1/2*(k-1)+inv(k*q)*(c1 - c2 - c3);
Tstar1 = (1-Ckq/n)*Tkq;
Tstar2 = Tkq/(1+Ckq/n);

P_Tstar1 = 1-chi2cdf(Tstar1,df);
P_Tstar2 = 1-chi2cdf(Tstar2,df);
%% end function %%
```

CI function.

```
function f = CI(c01,n,p,A1,A2,L,Gam_ini,lhat,S,Dgs,T3,T4,C1,loglk)
  p = pvalue(c01,n,p,A1,A2,L,Gam_ini,lhat,S,Dgs,T3,T4,C1,loglk);
  f = (p-0.05)^2;
```

pvalue function.

```
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.

```
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;
```