PRINCIPAL COMPONENT MODELS APPLIED TO CONFIRMATORY FACTOR ANALYSIS

by

Yin Chang

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Statistics

MONTANA STATE UNIVERSITY
Bozeman, Montana

December, 2012

APPROVAL

of a dissertation submitted by

Yin Chang

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 Graduate School.

Dr. Robert J. Boik (Co-chair)

Dr. Steven Cherry (Co-chair)

Approved for the Department of Mathematical Sciences

Dr. Kenneth L. Bowers

Approved for The Graduate School

Dr. Ronald W. Larsen

STATEMENT OF PERMISSION TO USE

Yin Chang

December, 2012

ACKNOWLEDGEMENTS

I want to thank Dr. Robert Boik for providing tireless guidance and constant support through thick and thin, during my graduate study at Montana State University. I do not think I could have done this without his insight, and the very helpful daily emails we had during this project.

I thank my mom, Xiaoming Yang, my dad, Haoyi Chang, and my brother, Le Chang, for their support.

Thanks to the people with whom I talked about statistics and mathematics: Dr. Mark Greenwood, Dr. Lukas Geyer, Dr. John Borkowski, and more.

I would like to thank the Department of Mathematics of Montana State University, from which I received all the support needed these years.

TABLE OF CONTENTS

TABLE OF CONTENTS – CONTINUED

TABLE OF CONTENTS – CONTINUED

LIST OF TABLES

Table Page

LIST OF TABLES – CONTINUED

Table Page

LIST OF TABLES – CONTINUED

LIST OF TABLES – CONTINUED

ABSTRACT

Testing structural equation models, in practice, may not always go smoothly, and the solution in the output may be an improper solution. The term *improper solution* refers to several possible problems with model estimation. Improper solution involves boundary estimates or exterior estimates, such as Heywood cases. Improper solution can be found in the output even without an error message.

The dissertation achieves the following goals: develop a stable algorithm to generate proper estimates of parameters, and the stable algorithm would be robust to data set variability and converge with high probability; use statistical theory to construct confidence intervals for functions of parameters, under non-normality and equations were derived in this thesis for computing confidence intervals; and use statistical theory to construct hypotheses tests, such as the goodness-of-fit tests and model comparison tests to determine the number of factors, structure of $\mathbf{\Lambda}$ and structure of $\mathbf{\Phi}$, especially under non-normality.

Based on the large simulation results, it can be demonstrated that the inference procedures for the proposed model work well enough to be used in practice and that the proposed model has advantages over the conventional model, in terms of proportion of proper solutions; average success rates and coverage rates of upper one-sided nominal 95% confidence intervals, lower one-sided nominal 95% confidence intervals, and two-sided nominal 95% confidence intervals; and average of the ratios of widths of two-sided nominal 95% confidence intervals.

CHAPTER 1

INTRODUCTION

1.1. Introduction

British psychologist Charles Spearman is commonly credited with the initial development of factor analysis. He used this technique in his 1904 article [1] to determine whether a general intelligence unobserved variable underlies individual performance on tests. Even though Spearman's belief in a single unobserved variable gave way to solutions with several unobserved variables, the purpose of factor analysis has remained the same. The primary goal of factor analysis is to explain the covariances or correlations among observable variables $\mathbf{y}$ in terms of fewer unobservable variables called factors $\mathbf{f}$, where $\mathbf{y} = (y_1\ y_2\ \cdots\ y_p)'$, $\mathbf{f} = (f_1\ f_2\ \cdots\ f_q)'$, and $p > q$. Like the observable variables $\mathbf{y}$, the factors $\mathbf{f}$ vary from individual to individual; but unlike the observable variables $\mathbf{y}$, the factors $\mathbf{f}$ can not be observed. The factors are underlying latent variables that *generate* $\mathbf{y}$. Factor analysis also is considered to be a dimension reduction technique because q is less than p.

Example. The scores of a sample of 220 boys were collected on the 6 school subjects: 1. Gaelic, 2. English, 3. History, 4. Arithmetic, 5. Algebra, and 6. Geometry. The following correlation matrix, $\mathbf{R}_{lm}$, is taken from Table 6.1 in Lawley and Maxwell [2].

$$
\mathbf{R}_{lm} = \begin{array}{c} \begin{array}{cccccc} \textit{\textbf{Gaelic}} & \textit{\textbf{English}} & \textit{\textbf{History}} & \textit{\textbf{Arithmetic}} & \textit{\textbf{Algebra}} & \textit{\textbf{Geometry}} \end{array} \\ \left(\begin{array}{ccc|ccc} 1 & 0.439 & 0.410 & 0.288 & 0.329 & 0.248 \\ 0.439 & 1 & 0.351 & 0.354 & 0.320 & 0.329 \\ 0.410 & 0.351 & 1 & 0.164 & 0.190 & 0.181 \\ \hline 0.288 & 0.354 & 0.164 & 1 & 0.595 & 0.470 \\ 0.329 & 0.320 & 0.190 & 0.595 & 1 & 0.464 \\ 0.248 & 0.329 & 0.181 & 0.470 & 0.464 & 1 \end{array} \right) \end{array}. \tag{1}
$$

It seems that $\mathbf{R}_{lm}$ suggests two groups of variables: {Gaelic, English, History} and {Arithmetic, Algebra, Geometry} based on the correlation strength in $\mathbf{R}_{lm}$. Accordingly, it is expected that the correlations among the 6 variables can be explained well by 2 factors. This example is discussed again in § 1.3.2.

1.2. Factor Analysis Model

In factor analysis, the components of $\mathbf{y}$ are modeled as linear combinations of factors $\mathbf{f}$, plus an error term $\boldsymbol{\epsilon}$. The basic factor analysis model can be written as follows:

$$
\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f} + \boldsymbol{\epsilon}, \tag{2}
$$

where $\mathbf{y}$ is a $p \times 1$ random vector of responses for a single subject, $\boldsymbol{\mu}$ is a $p \times 1$ vector of fixed population means, $\boldsymbol{\Lambda}$ is a $p \times q$ matrix of fixed factor loadings, $\mathbf{f}$ is a $q \times 1$ random vector of latent factor scores for the same subject, and $\boldsymbol{\epsilon}$ is a $p \times 1$ vector of random deviations.

The assumptions made for factor analysis model (2) can be expressed concisely using vector and matrix notation:

$$\mathrm{E}(\mathbf{f}) = \mathbf{0}, \quad \mathrm{Var}(\mathbf{f}) = \mathbf{\Phi}, \quad \mathrm{E}(\boldsymbol{\epsilon}) = \mathbf{0}, \quad \mathrm{Var}(\boldsymbol{\epsilon}) = \mathbf{\Psi},$$
$$\mathbf{\Psi} = \mathrm{diag}(\boldsymbol{\psi}) = \begin{pmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{pmatrix}, \text{ and } \mathrm{Cov}(\mathbf{f}, \boldsymbol{\epsilon}) = \mathbf{0}, \tag{3}$$

where $\mathrm{diag}(\cdot)$ is defined in Table 56. Accordingly, the variance of $\mathbf{y}$ in factor analysis model (2) is

$$\mathbf{\Sigma} \stackrel{\text{def}}{=} \mathrm{Var}(\mathbf{y}) = \mathbf{\Lambda}\mathrm{Var}(\mathbf{f})\mathbf{\Lambda}' + \mathbf{\Psi} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi}. \tag{4}$$

Note that $\mathbf{\Phi}$ in (4) is a correlation matrix of $\mathbf{f}$. If $\mathbf{\Phi}$ is a covariance matrix, then $\mathbf{\Sigma}$ is not identified because identifiability of $\mathbf{\Sigma}$ depends on the structural equation model defined in (4). Specifically, let $\mathbf{A}$ be any nonsingular $q \times q$ matrix. Then $\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' = (\mathbf{\Lambda}\mathbf{A})\left(\mathbf{A}^{-1}\mathbf{\Phi}\mathbf{A}'^{-1}\right)(\mathbf{\Lambda}\mathbf{A})'$, and $\mathbf{A}^{-1}\mathbf{\Phi}\mathbf{A}'^{-1}$ is a covariance matrix whenever $\mathbf{\Phi}$ is a covariance matrix. Special cases for $\mathbf{A}$ include diagonal matrices, orthogonal matrices, and matrices that are proportional to $\mathbf{I}_q$. The latter two preserve structural zeros in $\mathbf{\Lambda}$ in confirmatory factor analysis, where confirmatory factor analysis is discussed in §1.4.

In the following sections, exploratory factor analysis and confirmatory factor analysis are discussed, which are two major approaches to factor analysis.

1.3. Exploratory Factor Analysis

Exploratory factor analysis (EFA) is used to uncover the underlying structure of covariances or correlations among a set of variables. The researcher's *a priori* assumption is that any variable could be associated with any factor. Therefore, EFA allows all loadings in $\mathbf{\Lambda}$ to be free to vary. In EFA, equation 4 can be written as $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Phi}^{\frac{1}{2}}\mathbf{\Phi}^{\frac{1}{2}}\mathbf{\Lambda}' + \mathbf{\Psi} = \mathbf{\Lambda}^{*}\mathbf{I}_q\mathbf{\Lambda}^{*\prime} + \mathbf{\Psi}$, where $\mathbf{\Lambda}^{*} = \mathbf{\Lambda}\mathbf{\Phi}^{\frac{1}{2}}$. Consequently, in EFA, without loss of generality, $\text{Var}(\mathbf{f}) = \mathbf{\Phi}$ can be taken to be an identity matrix $\mathbf{I}_q$, which implies uncorrelated factors. After letting $\mathbf{\Phi} = \mathbf{I}_q$, (4) can be simplified as

$$\text{Var}(\mathbf{y}) = \mathbf{\Sigma} = \mathbf{\Lambda}\text{Var}(\mathbf{f})\mathbf{\Lambda}' + \mathbf{\Psi} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}. \tag{5}$$

In § 1.3.3, it is shown that rotation can induce nonzero correlations among the factors. In such cases, $\mathbf{\Phi}$ no longer is an identity matrix $\mathbf{I}_q$. The covariance matrix $\mathbf{\Sigma}$ in (5) can be written as $\mathbf{\Sigma} = \mathbf{\Sigma}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of parameters. There are many ways to parameterize $\mathbf{\Sigma}$. For example, let $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\boldsymbol{\lambda}}' \ \ \boldsymbol{\theta}_{\psi}')'$, where $\boldsymbol{\theta}_{\boldsymbol{\lambda}} = \text{vec}\,\mathbf{\Lambda}$, $\boldsymbol{\theta}_{\psi} = \text{diag}(\mathbf{\Psi})$, and vec is as in Table 56. Accordingly, $\mathbf{\Lambda} = \mathbf{\Lambda}(\boldsymbol{\theta}_{\boldsymbol{\lambda}})$ and $\mathbf{\Psi} = \mathbf{\Psi}(\boldsymbol{\theta}_{\psi})$. However, if the structural equation model is not identifiable, then there is difficulty in doing inference. For example, if $\mathbf{\Sigma}(\boldsymbol{\theta}_1) = \mathbf{\Sigma}(\boldsymbol{\theta}_2)$ and $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, then it can not be determined whether the true value of the parameter is $\boldsymbol{\theta}_1$ or $\boldsymbol{\theta}_2$. Before estimation can be discussed, it is necessary to discuss parameter identification.

1.3.1. Identification

A family of distributions can be written as $\{f(\mathbf{y};\boldsymbol{\theta}),\ \forall\ \mathbf{y} \in \mathcal{Y} \text{ and } \boldsymbol{\theta} \in \mathbf{\Theta}\}$, where $f(\mathbf{y};\boldsymbol{\theta})$ is the pdf or pmf of $\mathbf{y}$, $\mathcal{Y}$ is the support set of $\mathbf{y}$ and $\mathbf{\Theta}$ is the parameter space. Note that $\mathcal{Y}$ is the set of values that a variable $\mathbf{y}$ can take for which $f(\mathbf{y};\boldsymbol{\theta}) > 0$ and $\mathbf{\Theta}$ is the set of values that a parameter can take. For instance, $\mathcal{Y}$ of the Bernoulli distribution is 0 and 1 and $\mathbf{\Theta}$ of the Bernoulli parameter is the interval $(0, 1)$.

Casella and Berger [3] defined *identifiability* as follows: a parameter $\boldsymbol{\theta}$ for a family of distributions $\{f(\mathbf{y};\boldsymbol{\theta}),\ \forall\ \mathbf{y} \in \mathcal{Y} \text{ and } \boldsymbol{\theta} \in \mathbf{\Theta}\}$ is *identifiable* if distinct values of $\boldsymbol{\theta}$ correspond to distinct pdfs or pmfs, that is,

$$\{f(\mathbf{y};\boldsymbol{\theta}_1) = f(\mathbf{y};\boldsymbol{\theta}_2),\ \ \forall\ \ \mathbf{y} \in \mathcal{Y}\ \ \text{and}\ \ \boldsymbol{\theta}_i \in \mathbf{\Theta}\ \ \text{for}\ \ i = 1, 2\} \implies \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2.$$

Casella and Berger's [3] definition for identifiability can not be applied here because neither pdfs nor pmfs are specified in (3).

Bollen [4] gave a definition of *identifiability* in terms of a general structural equation model $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. Bollen's definition can be rephrased in the following manner: a structural equation model is identified if

$$\{\boldsymbol{\Sigma}(\boldsymbol{\theta}_1) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_2) \ \text{ and } \ \boldsymbol{\theta}_i \in \boldsymbol{\Theta} \ \text{ for } \ i = 1, 2\} \quad \Longrightarrow \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2. \tag{6}$$

Furthermore, the parameter vector $\boldsymbol{\theta}$ in $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is said to be identified if and only if the structural equation model $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is identified.

Anderson and Rubin [5] provided a few theorems on necessary conditions and sufficient conditions for *identification* of $\boldsymbol{\Lambda}$. For instance, Theorem 5.1 on page 118 of Anderson and Rubin [5] stated that a sufficient condition for identification of $\boldsymbol{\Psi}$ and $\boldsymbol{\Lambda}$ up to multiplication on the right by an orthogonal matrix is that if any row of $\boldsymbol{\Lambda}$ is deleted, there remain two disjoint submatrices of rank q. Also, Theorem 5.6 on page 120 of Anderson and Rubin [5] stated that a necessary condition for identification of $\boldsymbol{\Lambda}$ is that each column of $\boldsymbol{\Lambda}\mathbf{A}$ has at least three nonzero elements for every nonsingular $\mathbf{A}$. According to the Lemma on page 160 of Vinograde [6], if $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ are $p \times q$, then $\boldsymbol{\Lambda}_1\boldsymbol{\Lambda}_1' = \boldsymbol{\Lambda}_2\boldsymbol{\Lambda}_2'$ if and only if $\boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_2\mathbf{Q}$, where $\mathbf{Q} \in \mathcal{O}(q)$ and $\mathcal{O}(\cdot)$ is defined in Table 57.

Theorem 1. [Adapted from Bartlett [7], 1950, Page 84]. *Assume that the $p \times q$ matrix $\boldsymbol{\Lambda}$ has rank q and that $\boldsymbol{\Lambda}$ is identified up to orthogonal rotation. Then, $\boldsymbol{\Lambda}$ can be parameterized as $\boldsymbol{\Lambda}(\boldsymbol{\theta}_{\boldsymbol{\lambda}})$, where $\dim(\boldsymbol{\theta}_{\boldsymbol{\lambda}}) = pq - q(q-1)/2$ and $\boldsymbol{\theta}_{\boldsymbol{\lambda}}$ is identified.*

All proofs of theorems and corollaries in this thesis can be found in Appendix B. In the remainder of this thesis, it is assumed that $\boldsymbol{\Lambda}$ is identified up to orthogonal rotation.

The parameter $\boldsymbol{\theta}_{\boldsymbol{\lambda}}$ of $\boldsymbol{\Lambda}$ is said to be identified in the EFA model if the following relation holds:

$$\{\boldsymbol{\Lambda}(\boldsymbol{\theta}_{\boldsymbol{\lambda}_1})\boldsymbol{\Lambda}(\boldsymbol{\theta}_{\boldsymbol{\lambda}_1})' = \boldsymbol{\Lambda}(\boldsymbol{\theta}_{\boldsymbol{\lambda}_2})\boldsymbol{\Lambda}(\boldsymbol{\theta}_{\boldsymbol{\lambda}_2})' \text{ and } \boldsymbol{\theta}_{\boldsymbol{\lambda}_i} \in \boldsymbol{\Theta}_{\boldsymbol{\lambda}} \text{ for } i = 1, 2\} \quad \Longrightarrow \quad \boldsymbol{\theta}_{\boldsymbol{\lambda}_1} = \boldsymbol{\theta}_{\boldsymbol{\lambda}_2},$$

where $\boldsymbol{\Theta}_{\boldsymbol{\lambda}}$ is the parameter space of $\boldsymbol{\theta}_{\boldsymbol{\lambda}}$.

Anderson and Rubin [5] wrote that one way to identify $\boldsymbol{\theta}_{\boldsymbol{\lambda}}$ is to choose an orthogonal matrix $\mathbf{Q}$ such that $\mathbf{Q}'\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}\mathbf{Q} = \boldsymbol{\Lambda}^{*\prime}\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}^*$ is diagonal, where $\mathbf{Q} \in \mathcal{O}_q$ and $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}\mathbf{Q}$. If the diagonal elements are distinct and arranged in descending order, then $\boldsymbol{\theta}_{\boldsymbol{\lambda}} \stackrel{\text{def}}{=} \text{vec}(\boldsymbol{\Lambda}^*)$ is identified. As a result, $\boldsymbol{\Lambda}^*$ needs to satisfy the constraint $\boldsymbol{\lambda}^{*\prime}_i\boldsymbol{\Psi}^{-1}\boldsymbol{\lambda}^*_j = 0$ for all $i \neq j$, where $\boldsymbol{\lambda}^*_i$ is the i^{th} column of $\boldsymbol{\Lambda}$. Hence, $\boldsymbol{\Lambda}^*$ is subject to $\binom{q}{2}$ constraints. This leaves $pq - \binom{q}{2} = pq - q(q-1)/2$ identified parameters in $\boldsymbol{\Lambda}^*$ because the dimension of $\boldsymbol{\theta}_{\boldsymbol{\lambda}} = \text{vec}(\boldsymbol{\Lambda}^*)$ is $pq \times 1$. This implies that $\boldsymbol{\Lambda}^*$ can be parameterized, at least locally, in terms of $pq - q(q-1)/2$ identified parameters.

1.3.2. Estimation

To estimate the $p \times p$ population covariance or correlation matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$, one can let $\hat{\boldsymbol{\theta}}$ represent an estimator of the parameter vector $\boldsymbol{\theta}$ in $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. The estimator $\hat{\boldsymbol{\theta}}$ is chosen to minimize a function that measures the discrepancy between the sample covariance matrix $\mathbf{S}$ and $\boldsymbol{\Sigma}$ with respect to $\boldsymbol{\theta}$. These functions are called discrepancy functions and are denoted by $F(\boldsymbol{\Sigma}, \mathbf{S})$. A discrepancy function is a numerical value that expresses how well $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ reproduces the observed data. The parameter estimator, $\hat{\boldsymbol{\theta}}$, for a given structural model, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, is generally selected to make a discrepancy function as small as possible. Browne [8] gave a detailed discussion on discrepancy functions. The following properties of discrepancy functions were adapted from Browne [8]:

(1) $F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S}) \geq 0$;

(2) $F(\mathbf{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S}) = 0$ if and only if $\mathbf{\Sigma}(\hat{\boldsymbol{\theta}}) = \mathbf{S}$;

(3) $F(\mathbf{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ is a continuous function of $\mathbf{S}$ and $\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})$.

There are two commonly used discrepancy functions:

$$\begin{aligned} &\text{(a) } F(\mathbf{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S}) = \operatorname{tr}\{[\mathbf{S} - \mathbf{\Sigma}(\hat{\boldsymbol{\theta}})]\mathbf{S}^{-1}\}^2 \quad \text{and} \\ &\text{(b) } F(\mathbf{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S}) = \operatorname{tr}[\mathbf{\Sigma}^{-1}(\hat{\boldsymbol{\theta}})\mathbf{S}] + \ln|\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})| - p - \ln|\mathbf{S}|, \end{aligned} \tag{7}$$

where $\operatorname{tr}(\cdot)$ is defined in Table 56. Discrepancy function (a) in (7) is called the Generalized Least Squares (GLS) discrepancy function. This yields a scale-free method. Discrepancy function (b) in (7) is called a maximum likelihood (ML) discrepancy function. This measure also is scale-free and, when $\mathbf{y}$ is multinormally distributed, leads to efficient estimators in large samples. Jöreskog and Goldberger [9] showed that discrepancy function (b) can be viewed as an approximation to discrepancy function (a). Browne [8] wrote that if $\boldsymbol{\theta}$ is identified and $\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})$ is continuous, then all estimators, $\hat{\boldsymbol{\theta}}$, obtained by minimizing the discrepancy functions $F(\mathbf{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ are consistent.

Consider the sample correlation matrix, $\mathbf{R}_{lm}$, given in (1), where $p = 6$. If there are $q = 2$ factors, then the factor analysis model can be written as follows:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \\ \mu_6 \end{pmatrix} + \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & \lambda_{52} \\ \lambda_{61} & \lambda_{62} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix}.$$

The following estimates for $\boldsymbol{\Lambda}$ and $\boldsymbol{\psi}$ without rotation are generated through *factanal* in R.

$$\widehat{\boldsymbol{\Lambda}}_{no} = \begin{pmatrix} 0.5533 & 0.4286 \\ 0.5682 & 0.2883 \\ 0.3922 & 0.4500 \\ 0.7404 & -0.2728 \\ 0.7239 & -0.2113 \\ 0.5954 & -0.1317 \end{pmatrix} \quad \text{and} \quad \widehat{\boldsymbol{\psi}}_{no} = \begin{pmatrix} 0.5102 \\ 0.5941 \\ 0.6437 \\ 0.3773 \\ 0.4314 \\ 0.6282 \end{pmatrix}, \tag{8}$$

where $\widehat{\boldsymbol{\Sigma}}_{no} = \widehat{\boldsymbol{\Lambda}}_{no}\widehat{\boldsymbol{\Lambda}}'_{no} + \widehat{\boldsymbol{\Psi}}_{no}$. The estimates in (8) are obtained by optimizing the log likelihood assuming multivariate normality of $\boldsymbol{\epsilon}$ in (2), which is equivalent to minimizing the discrepancy function (b) in (7).

Based on the above structure of $\widehat{\boldsymbol{\Lambda}}_{no}$, it is hard to identify the natural groupings of variables. In § 1.3.1, it is discussed that $\widehat{\boldsymbol{\Lambda}}$ can be replaced by $\widehat{\boldsymbol{\Lambda}}\mathbf{Q}$, where $\mathbf{Q} \in \mathcal{O}_q$. In order to find a factor loading structure that is easier to explain, rotation of the factor loading matrix can be used to achieve this goal.

1.3.3. Rotation

According to the rotation criteria discussed on page 372 to page 385 in Rummel [10], major substantive rotation criteria involve *simple structure*, *parsimony*, *factorial invariance*, *hypothetical structure*, *partialling* and *casual exploration*. In some sense, all those criteria are designed to make the structure of $\boldsymbol{\Lambda}$ as simple as possible, with most elements of $\boldsymbol{\Lambda}$ either *close to zero* or *far from zero*, and with as few as possible of the elements taking *intermediate values*.

Rotation of $\boldsymbol{\Lambda}$ to a simple structure typically involves optimizing a function of the rotated loadings over the family of orthogonal or oblique rotations. The function to be optimized is called the criterion and is denoted by $\mathcal{L}(\boldsymbol{\Lambda}^*)$. Detailed discussions of oblique rotation and orthogonal rotation were given by Browne [11], Browne and Cudeck [12], and Bernaards and Jennrich [13]. After rotation, factor analysis model (2) can be written in the following form:

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f} + \boldsymbol{\epsilon} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{T}\mathbf{T}^{-1}\mathbf{f} + \boldsymbol{\epsilon} = \boldsymbol{\mu} + \boldsymbol{\Lambda}^{*}\mathbf{f}^{*} + \boldsymbol{\epsilon}, \tag{9}$$

where $\mathbf{T}$ is a nonsingular rotation matrix, $\boldsymbol{\Lambda}^{*} = \boldsymbol{\Lambda}\mathbf{T}$ is the rotated loading matrix, $\mathbf{f}^{*} = \mathbf{T}^{-1}\mathbf{f}$ is the rotated vector of factor scores and $\text{Var}(\mathbf{f}^{*}) = \mathbf{T}^{-1}(\mathbf{T})^{-1\prime} = (\mathbf{T}'\mathbf{T})^{-1} = \boldsymbol{\Phi}$. Typically, the rotation matrix $\mathbf{T}$ is subjected to either of two constraints:

(a) $\mathbf{T}$ is a $q \times q$ nonsingular matrix and $(\mathbf{T}'\mathbf{T})^{-1}$ is a correlation matrix, or

(b) $\mathbf{T} \in \mathcal{O}_q$, where $\mathcal{O}_q$ is defined in Table 56.

The rotation in (a) is called oblique rotation whereas the rotation in (b) is called orthogonal rotation.

1.3.3.1. Oblique Rotation: Under oblique rotation, the variance of $\mathbf{y}$ in factor analysis model (9) is

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\mathbf{T}(\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'\boldsymbol{\Lambda}' + \boldsymbol{\Psi} = \boldsymbol{\Lambda}^{*}\boldsymbol{\Phi}\boldsymbol{\Lambda}^{*\prime} + \boldsymbol{\Psi}, \tag{10}$$

where $\boldsymbol{\Lambda}^{*} = \boldsymbol{\Lambda}\mathbf{T}$ and $\boldsymbol{\Phi} = (\mathbf{T}'\mathbf{T})^{-1}$ is a correlation matrix. Note that the factor correlation matrix, $\boldsymbol{\Phi} \neq \mathbf{I}_q$, which implies that the factor scores, $\mathbf{f}$, are correlated.

For oblique rotations, the quartic loss criterion in Carroll [14], the linear component loss criterion in Jennrich [15], and the modified component loss criterion in Jennrich [15] can be used. A matrix representation of the quartic loss criterion was given by Boik [16] as follows:

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\Lambda}^*) &= k_1[\mathrm{tr}(\boldsymbol{\Lambda}^{*\prime}\boldsymbol{\Lambda}^*)]^2 + k_2\sum_{i=1}^{p}(\mathbf{e}_i^{p\prime}\boldsymbol{\Lambda}^*\boldsymbol{\Lambda}^{*\prime}\mathbf{e}_i^p)^2 + k_3\sum_{j=1}^{q}(\mathbf{e}_j^{q\prime}\boldsymbol{\Lambda}^{*\prime}\boldsymbol{\Lambda}^*\mathbf{e}_j^q)^2 \\
&\quad + k_4\sum_{i=1}^{p}\sum_{j=1}^{q}\lambda^{*4}_{ij} = (\boldsymbol{\lambda}^*\otimes\boldsymbol{\lambda}^*)'\mathbf{W}(\boldsymbol{\lambda}^*\otimes\boldsymbol{\lambda}^*), \quad \text{where} \quad \boldsymbol{\lambda}^* = \mathrm{vec}\,\boldsymbol{\Lambda}^*, \\
\mathbf{W} &= \sum_{i=1}^{4}k_i\mathbf{W}_i, \quad \mathbf{W}_1 = \mathbf{I}_{q^2p^2}, \quad \mathbf{W}_2 = \mathbf{B}_{pq}[\mathrm{vec}(\mathbf{I}_q)\mathrm{vec}'(\mathbf{I}_q)\otimes\mathbf{L}_{22,p}]\mathbf{B}'_{pq}, \\
\mathbf{W}_3 &= \mathbf{B}_{pq}[\mathbf{L}_{22,q}\otimes\mathrm{vec}(\mathbf{I}_p)\mathrm{vec}'(\mathbf{I}_p)]\mathbf{B}'_{pq}, \quad \mathbf{W}_4 = \mathbf{L}_{22,qp}, \quad \mathbf{B}_{pq} = \mathbf{I}_q\otimes\mathbf{K}_{p,q}\otimes\mathbf{I}_p,
\end{aligned}
\tag{11}
$$

$\mathrm{tr}(\cdot)$, $\mathbf{K}_{p,q}$, $\mathbf{L}_{22,q}$ are defined in Table 56, and $k_1, ..., k_4$ are chosen to emphasize different aspects of the rotated loadings. Clarkson and Jennrich [17] listed the choices of k_1, k_2, k_3 and k_4 in (11) that lead to various oblique rotation criteria, such as quartimin criterion by setting $k_1 = 0$, $k_2 = 1$, $k_3 = 0$ and $k_4 = -1$ in (11). The modified component loss criterion can be written as $\mathcal{L}(\boldsymbol{\Lambda}^*) = \sum_{i=1}^{pq} h(|\lambda_i^*|)$, where λ_i^* is the ith component of $\boldsymbol{\lambda}^* = \mathrm{vec}\,\boldsymbol{\Lambda}^*$ and h is a component loss function. The component loss function h, for example, could be defined as $h(|\lambda|) = |\lambda|$.

Boik [18] showed that the optimally rotated factor loading matrix $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}\mathbf{T}$ necessarily satisfies

$$
\mathbf{D}^{(1)}_{\mathcal{L}(\boldsymbol{\Lambda}^*);\boldsymbol{\lambda}^{*\prime}}(\mathbf{I}_q\otimes\boldsymbol{\Lambda}^*)\,[\mathbf{I}_{q^2} - \mathbf{L}_{22,q}(\boldsymbol{\Phi}\otimes\mathbf{I}_q)]\,\mathbf{A}_4 = \mathbf{0}, \tag{12}
$$

where $\mathbf{A}_4 = \sum_{i=1}^{q}\sum_{j\neq i}^{q}\left(\mathbf{e}_j^q\otimes\mathbf{e}_i^q\right)\mathbf{e}_{f_{3,ij}}^{h_3\prime}$: $q^2\times q(q-1)$, $h_3 = q(q-1)$, $f_{3,ij} = (q-1)(j-1)+i-I(i>j)$ and $\mathbf{e}_i^q$ is defined in Table 56. Define $\boldsymbol{\theta}_{\boldsymbol{\lambda}^*}$ as $\boldsymbol{\theta}_{\boldsymbol{\lambda}^*} \stackrel{\text{def}}{=} \mathrm{vec}\,\boldsymbol{\Lambda}^*$ and define $\boldsymbol{\theta}_{\boldsymbol{\phi}}$ to be the vector that contains the $q(q-1)/2$ components in the upper triangle of $\boldsymbol{\Phi}$. If the optimizer of $\mathcal{L}(\boldsymbol{\Lambda}^*)$ is unique, then $\boldsymbol{\theta}_{\boldsymbol{\lambda}^*}$ is identified. There are $pq + q(q-1)/2 - q(q-1) = pq - q(q-1)/2$ identified parameters in $\boldsymbol{\Lambda}^*$ and $\boldsymbol{\Phi}$, because $\dim(\boldsymbol{\theta}_{\boldsymbol{\lambda}^*}) = pq$, $\dim(\boldsymbol{\theta}_{\boldsymbol{\phi}}) = q(q-1)/2$ and $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}\mathbf{T}$ is subject to the $q(q-1)$ constraints in (12).

Oblique rotations may turn the factor loadings $\mathbf{\Lambda}$ into a simple structure that are easily interpreted. However, the resulting factors are correlated. Oblique rotation is a transformation in which axes do not remain perpendicular. Technically, the term oblique rotation is a misnomer, because rotation implies an orthogonal transformation.

1.3.3.2. Orthogonal Rotation: If the matrix $\mathbf{T}$ in factor analysis model (9) is an orthogonal matrix, then it is called an orthogonal rotation. Specifically,

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{\Lambda f} + \boldsymbol{\epsilon} = \boldsymbol{\mu} + \mathbf{\Lambda T T' f} + \boldsymbol{\epsilon} = \boldsymbol{\mu} + \mathbf{\Lambda}^* \mathbf{f}^* + \boldsymbol{\epsilon} \tag{13}$$

where $\mathbf{T} \in \mathcal{O}_q$, $\mathbf{\Lambda}^* = \mathbf{\Lambda T}$, $\mathbf{f}^* = \mathbf{T}'\mathbf{f}$, and $\text{Var}(\mathbf{f}^*) = \text{Var}(\mathbf{T}'\mathbf{f}) = \mathbf{T}'\mathbf{T} = \mathbf{I}_q$. In orthogonal rotation, the factor scores, $\mathbf{f}$, are uncorrelated. Furthermore, orthogonal rotation avoids the issue of interpreting the matrix of factor intercorrelations because $\text{Var}(\mathbf{f}^*) = \mathbf{I}_q$.

For orthogonal rotations, Clarkson and Jennrich [17] listed the choices of k_1, k_2, k_3 and k_4 in quartic loss criteria (11) that lead to various orthogonal rotation criteria. An example of a component loss criterion is to set $k_1 = 0$, $k_2 = 0$, $k_3 = 0$ and $k_4 = 1$ in (11), which is called the quartimax criterion. The quartimax rotation maximizes a criterion of the form $\mathcal{L}(\mathbf{\Lambda}^*) = \sum_{i=1}^{p}\sum_{j=1}^{q} \lambda_{ij}^{*4}$, where λ_{ij}^* is the ijth component of $\mathbf{\Lambda}^*$.

Boik [18] showed that the optimal orthogonal rotation matrix $\mathbf{\Lambda}^* = \mathbf{\Lambda T}$ necessarily satisfies

$$\mathbf{D}^{(1)\prime}_{\mathcal{L}(\mathbf{\Lambda}^*);\boldsymbol{\lambda}^*}(\mathbf{I}_q \otimes \mathbf{\Lambda}^*) 2\mathbf{N}_q^{\perp}\mathbf{A}_2 = \mathbf{0}, \tag{14}$$

where $\mathbf{A}_2 = \sum_{i=1}^{q-1}\sum_{j=i+1}^{q}\left(\mathbf{e}_j^q \otimes \mathbf{e}_i^q\right)\mathbf{e}_{f_{2,ij}}^{h_2\prime}$: $q^2 \times [q(q-1)/2]$, $h_2 = q(q-1)/2$ and $f_{2,ij} = \frac{(j-1)(i-1)}{2} + i$. Define $\boldsymbol{\theta}_{\boldsymbol{\lambda}^*}$ as $\boldsymbol{\theta}_{\boldsymbol{\lambda}^*} \overset{\text{def}}{=} \operatorname{vec}\mathbf{\Lambda}^*$. If the optimizer of $\mathcal{L}(\mathbf{\Lambda}^*)$ is unique, then $\boldsymbol{\theta}_{\boldsymbol{\lambda}^*}$ is identified. There are $pq - q(q-1)/2$ identified parameters in $\mathbf{\Lambda}^*$ because $\dim(\boldsymbol{\theta}_{\boldsymbol{\lambda}^*}) = pq$ with the $q(q-1)/2$ constraints in (14).

Let

$$(l_{ij}) = \mathbf{\Lambda}, \quad \text{and} \quad h_i^2 = \sum_{j=1}^{q} l_{ij}^2, \quad \text{for} \quad i = 1, \ldots, p, \tag{15}$$

where h_i^2 is the communality of the i^{th} variable and is the proportion of the i^{th} variable's total variance that is accounted for by all the factors.

Thus far, $\mathbf{\Lambda}$ can be rotated obliquely as well as orthogonally. All the rotation criteria tend to give equal weight to variables having very low communalities and those having near-unity communalities. Kaiser [19] corrected for this tendency by dividing each factor loading within a given row of the factor structure by the square root of the communality of that variable.

1.3.3.3. Kaiser Normalization: Kaiser normalization can be applied to any rotation criteria. For orthogonal rotations, it consists of finding the optimal $q \times q$ orthogonal matrix $\mathbf{T}$ such that the rotated row-standardized loadings are simple, where the rotated row-standardized loadings are elements of

$$\mathbf{D}^{-1/2}\mathbf{\Lambda}\mathbf{T}, \tag{16}$$

where $\mathbf{D} = \text{Diag}[\text{diag}(\mathbf{\Lambda}\mathbf{\Lambda}')]$, $\text{Diag}(\cdot)$ and $\text{diag}(\cdot)$ are defined in Table 56. After the optimal matrix $\mathbf{T}$ is found, then the loadings are restandardized by premultiplying by $\mathbf{D}^{1/2}$.

For oblique rotation, $\mathbf{T}$ in (16) is a $q \times q$ nonsingular matrix instead of an orthogonal matrix, and $(\mathbf{T}'\mathbf{T})^{-1}$ is a factor correlation matrix given in (10), where $(\mathbf{T}'\mathbf{T})^{-1} \neq \mathbf{I}_q$. Kaiser normalization is not used when factors are not rotated because premultiplying $\mathbf{D}^{1/2}$ to $\mathbf{D}^{-1/2}\mathbf{\Lambda}$ in (16) does not affect the structure of $\mathbf{\Lambda}$.

Note that the diagonal elements of $\mathbf{D}$ in (16) are communalities, which are defined in (15). The communality, h_i^2, remains constant under orthogonal and oblique rotations for $i = 1, \ldots, p$.

Most statistical computer programs, such as, SAS, R and Matlab, provide options for several different rotation criteria, such as varimax, quartimax and quartimin, with options to choose Kaiser Normalization.

A simpler structure of the loading matrix $\widehat{\mathbf{\Lambda}}_{no}$ in (8) can be achieved by orthogonal rotations or oblique rotations. The following estimates are obtained in R by imposing the normal varimax criterion and the quartmin criterion with Kaiser normalization.

$$\widehat{\mathbf{\Lambda}}_{varimax} = \begin{pmatrix} 0.2347 & 0.6593 \\ 0.3229 & 0.5493 \\ 0.0875 & 0.5904 \\ 0.7706 & 0.1697 \\ 0.7235 & 0.2125 \\ 0.5724 & 0.2103 \end{pmatrix} \quad \text{and} \quad \widehat{\mathbf{\Lambda}}_{quartmin} = \begin{pmatrix} 0.0563 & 0.6692 \\ 0.1901 & 0.5179 \\ -0.0875 & 0.6373 \\ 0.8129 & -0.0483 \\ 0.7465 & 0.0146 \\ 0.5772 & 0.0590 \end{pmatrix}, \tag{17}$$

where $\widehat{\mathbf{\Lambda}}_{varimax}$ is the estimate for $\mathbf{\Lambda}$ under the varimax rotation and $\widehat{\mathbf{\Lambda}}_{quartmin}$ is the estimate for $\mathbf{\Lambda}$ under the quartmin rotation. Note that under the quartmin rotation, the factor correlation is no longer $\mathbf{I}_2$, instead the estimate for $\mathbf{\Phi}$ is as follows:

$$\widehat{\mathbf{\Phi}}_{quartmin} = \begin{pmatrix} 1.0000 & 0.5161 \\ 0.5161 & 1.0000 \end{pmatrix}. \tag{18}$$

Based on $\widehat{\mathbf{\Lambda}}_{quartmin}$ in (17), the first 3 variables, {Gaelic, English, History}, have large loadings on factor 2 and small loadings on factor 1, whereas the last 3 variables {Arithmetic, Algebra, Geometry} have large loadings on factor 1 and small loadings on factor 2. This result is consistent with the initial guess of two groups of variables based on the correlation strength in $\mathbf{R}_{lm}$ of (1).

Thus far we have concentrated on EFA. In EFA, a factor typically influences all observed variables and the number of factors is not necessarily determined before the analysis. Although rotation can be used to achieve a simpler structure for loadings

$\mathbf{\Lambda}$, there is no prior structure for $\mathbf{\Lambda}$. In contrast to EFA, confirmatory factor analysis (CFA) seeks to determine if the number of factors and the loadings conform to what is expected on the basis of theory or experience.

1.4. Confirmatory Factor Analysis

The researcher's *a priori* assumption is that each factor is associated with a specified subset of variables. Therefore, there is a prior structure for $\mathbf{\Lambda}$. Specifically, the number of factors is set by the analyst, whether a factor influences an observed variable is specified, and some factor loadings on observed variables are fixed to zero or some other constant, such as one. More constraints than needed for identification are imposed. In CFA, rotation is not necessary because factor loading matrix $\mathbf{\Lambda}$ already has a simple structure and $\mathbf{\Lambda}$ already is identified.

Under CFA, in factor analysis model (2), factor scores have distribution $\mathbf{f} \sim (\mathbf{0}, \mathbf{\Phi})$, where $\mathbf{\Phi}$ is the correlation matrix of the common factors. In CFA, the covariance of $\mathbf{y}$ in (2) can be written as

$$\mathrm{Var}(\mathbf{y}) = \mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi}, \tag{19}$$

where there is a prior structure for $\mathbf{\Lambda}$.

In short, CFA requires a detailed and identified initial model. In later sections of this dissertation, results and inferences are made based on CFA.

1.5. Solution Issues

Classify parameter estimates $\hat{\boldsymbol{\theta}}$ as follows:

(a) a proper or interior estimate: $\hat{\boldsymbol{\theta}} \in \mathbf{\Theta}$, where $\mathbf{\Theta}$ is an open set,

(b) a boundary estimate: $\hat{\boldsymbol{\theta}} \in \overline{\mathbf{\Theta}}$, but $\hat{\boldsymbol{\theta}} \notin \mathbf{\Theta}$, where $\overline{\mathbf{\Theta}}$ is a closure set, and

(c) an exterior estimate: $\hat{\boldsymbol{\theta}} \notin \overline{\boldsymbol{\Theta}}$.

Both (b) and (c) are called improper solutions.

For instance, $\boldsymbol{\Theta}$ for the Bernoulli parameter is $(0, 1)$. If $\hat{\boldsymbol{\theta}}$ of the Bernoulli parameter is 0 or 1, then it is called a boundary estimate defined in (b). If $\hat{\boldsymbol{\theta}} \notin [0, 1]$, then it is called an exterior estimate defined in (c).

Negative unique variance estimates are called Heywood cases [20]. A Heywood case is improper because it yields an exterior estimate. However, not every improper estimate is a Heywood case. For example, correlation estimates greater than one in absolute value are exterior estimates, but they are not Heywood cases. Any non-positive eigenvalue estimates of a covariance or correlation matrix are exterior estimates, but not Heywood cases.

Improper solutions have many causes. In the following section, the possible conceptual and empirical reasons for improper solutions are reviewed.

1.5.1. Causes of Improper Solutions

(1) Improper solutions may be caused by sampling fluctuations in combination with a true parameter value close to the boundary of proper region (Bollen [4] and Driel [21]). Rindskopf's [22] examples show how small factor loadings, factor correlations near zero, and factor correlations near one can lead to improper solutions.

(2) Inappropriateness or misspecification of a model can cause improper solutions (Bollen [4] and Driel [21]). Model specification depends on the researcher's substantive knowledge and is specific to each problem (Bollen [4]). In some cases, there does not exist any factor analysis model that fits $\boldsymbol{\Sigma}$ (Driel [21]).

(3) If a model is not identified (unidentified model), then it can cause an improper solution (Driel [21]). Anderson and Gerbing [23] suggested three or more variables per factor to decrease the chances of improper solutions. This suggestion agrees with Theorem 5.6 of Anderson and Rubin [5] for parameter identification.

(4) Outliers or influential observations can lead to distorted measures of association among the observed variables, which in turn affects the parameter estimators (Bollen [4]). It is wise to check for outliers before using a sample covariance matrix, particularly when the number of observations is small. Of course, the cause of the outliers should be sought. In one example given by Bollen [24], improper estimates were eliminated after outliers were dropped.

(5) Boomsma [25] and Anderson and Gerbing [23] found that small sample size is associated with the occurrence of improper solutions. Anderson and Gerbing [23] recommended a sample size of 150 or more to decrease the chances of improper solutions.

To illustrate the problem of improper solutions, 1000 random samples of sizes $N = 100, 200,$ and 500 were generated from a multivariate normal distribution, that is, $\mathbf{y}_i \overset{\text{iid}}{\sim} MVN(\mathbf{0}, \mathbf{\Sigma})$, for $i = 1, \ldots, N$, where $\mathbf{y}_i$ is a $p \times 1$ random vector. Let $\mathbf{\Phi} = \mathbf{\Gamma}\mathbf{\Delta}\mathbf{\Gamma}'$, where $\mathbf{\Phi}$ is the factor correlation matrix, $\mathbf{\Gamma}$ is the matrix of eigenvectors of $\mathbf{\Phi}$, and $\mathbf{\Delta}$ are the diagonal matrix of eigenvalues of $\mathbf{\Phi}$. Recall that $\mathbf{\Sigma}$ can be written as $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi}$, where $\mathbf{\Phi} = \mathbf{\Gamma}\mathbf{\Delta}\mathbf{\Gamma}'$.

Design $\mathbf{\Sigma}$ in the following manner:

$$\mathbf{\Lambda} = 0.7 * (\mathbf{I}_q \otimes \mathbf{1}_3); \quad \boldsymbol{\delta} = \begin{pmatrix} \delta_1 & [(q - \delta_1)/(q-1)] \times \mathbf{1}_{q-1} \end{pmatrix}'; \quad \mathbf{\Delta} = \text{Diag}(\boldsymbol{\delta});$$

$$\boldsymbol{\gamma}_1 = \mathbf{1}_q(1/\sqrt{q}); \quad \boldsymbol{\gamma}_2 \in \mathcal{N}(\boldsymbol{\gamma}_1'); \quad \mathbf{\Gamma} = (\boldsymbol{\gamma}_1 \; \boldsymbol{\gamma}_2) \quad \boldsymbol{\psi} = 0.51 * \mathbf{1}_{3q}; \text{ and } q = 4;$$

where δ_1 is the largest eigenvalue of $\mathbf{\Phi}$ and is given the value of 2, 3, 3.5, and 3.9 in the simulation. In Table 1 and Table 2, the percentages of improper solutions out of 1000 random samples were recorded in EFA and CFA at $N = 100, 200,$ and 500. Also, in Table 3, the percentage of improper solutions were recorded in CFA by using a model to set the smallest 3 eigenvalues of $\mathbf{\Phi}$ to be homogeneous. A list of possible models for eigenvalues of $\mathbf{\Phi}$ can be found in Table 4 of Chapter 2.

In this example, when the largest eigenvalue δ_1 increases, the other smallest eigenvalues are getting close to 0, that is, the true eigenvalues are close to the boundary. Consequently, the percentage of improper solutions increases. Also, note that when sample size increases, the percentage of improper solutions decreases. Compare Table 2 to Table 3, the percentage of improper solutions in Table 3 is far less than the percentage of improper solutions in Table 2. Especially, the percentage of improper solutions for $\mathbf{\Phi}$ is essentially zero in Table 3. This result suggests that fitting structure to $\mathbf{\Delta}$ has reduced improper solutions and eliminated the improper solutions for $\mathbf{\Phi}$. In Chapter 2, there is a detailed explanation on parameterization of eigenvalues of $\mathbf{\Phi}$.

1.5.2. Existing Strategies

To distinguish among the first three reasons for improper solutions listed in § 1.5.1, sample fluctuation, model misspecification and unidentified model, Driel [21] used a confidence region of unique variance based on the estimated standard error for the exterior estimate. Driel [21] listed three types of confidence regions of unique variance as follows:

(1) positive values were within the confidence region constructed around the exterior estimate. In this case, the improper solution was interpreted as being caused by sample fluctuation,

Table 1: Percentage of Improper Solutions in EFA

δ_1	N	Heywood Cases	N	Heywood Cases	N	Heywood Cases
2	100	0.5%	200	0%	500	0%
3	100	4.1%	200	0%	500	0%
3.5	100	9.8%	200	2.8%	500	0%
3.9	100	16.7%	200	7.7%	500	3.2%

Table 2: Percentage of Improper Solutions in CFA

δ_1	N	Heywood Cases	Heywood Cases and Improper Solutions for $\mathbf{\Phi}$	Improper Solutions for $\mathbf{\Phi}$
2	100	0.5%	0%	0%
	200	0%	0%	0%
	500	0%	0%	0%
3	100	0%	0%	0.5%
	200	0%	0%	0%
	500	0%	0%	0%
3.5	100	0%	0%	16%
	200	0%	0%	1.4%
	500	0%	0%	0%
3.9	100	0.1%	0.1%	88.3%
	200	0%	0%	79.5%
	500	0%	0%	55.4%

Table 3: Percentage of Improper Solutions in CFA with model fitting for $\mathbf{\Delta}$

δ_1	N	Heywood Cases	Heywood Cases and Improper Solutions for $\mathbf{\Phi}$	Improper Solutions for $\mathbf{\Phi}$
2	100	0.1%	0%	0%
	200	0%	0%	0%
	500	0%	0%	0%
3	100	0%	0%	0%
	200	0%	0%	0%
	500	0%	0%	0%
3.5	100	0%	0%	0%
	200	0%	0%	0%
	500	0%	0%	0%
3.9	100	0%	0%	19.9%
	200	0%	0%	9.1%
	500	0%	0%	1%

(2) the confidence region constructed around the exterior estimate did not include any positive values. In this case, there may not exist any factor analysis model that fits the data,

(3) the confidence region was much larger in one direction than in the others. In this case, the cause of the improper solution may result from an unidentified model.

Bollen [24] described a multidimensional outlier screening device, explained the role of outliers in generating improper solutions and recommended that analysts screen data for outliers before estimating a factor analysis model.

When improper solutions occur and the model is not believed to be misspecified, one can use equality or inequality constraints to force the estimate to be proper. Lawley and Maxwell [2] suggested that when a Heywood case is encountered, the corresponding unique variance should be set to zero or some very small value, and the parameters of the model can be estimated again. Lee [26] pointed out that fixing nonpositive variance estimates at zero is not perfect, especially when more than one Heywood case is encountered. Lee [26] also provided an artificial example to illustrate that estimates produced by using inequality constraints are better than estimates produced by fixing nonpositive estimates at zeros. To prevent Heywood cases, Bentler and Weeks [27] fixed unique variances at one and estimated the linear coefficients. To be more specific, consider the following factor analysis model:

$$\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f} + \mathbf{D}_{\boldsymbol{\alpha}}\boldsymbol{\epsilon}, \tag{20}$$

where $\mathbf{D}_{\boldsymbol{\alpha}} = \mathrm{Diag}(\alpha_1, \alpha_2, \alpha_3, \cdots, \alpha_p)$ and $\mathrm{Var}(\boldsymbol{\epsilon}) = \boldsymbol{\Psi}$. Conventionally, $\mathbf{D}_{\boldsymbol{\alpha}}$ is defined to be $\mathbf{I}_p$ and $\boldsymbol{\Psi}$ is estimated. In contrast to the above usual practice, Bentler and Weeks let $\boldsymbol{\Psi} = \mathbf{I}_p$ and $\mathbf{D}_{\boldsymbol{\alpha}}$ is estimated. Accordingly, $\mathrm{Var}(\mathbf{D}_{\boldsymbol{\alpha}}\boldsymbol{\epsilon}) = \mathbf{D}_{\boldsymbol{\alpha}}\mathbf{I}_p\mathbf{D}_{\boldsymbol{\alpha}} =$

$\text{Diag}(\alpha_1^2, \alpha_2^2, \alpha_3^2, \cdots, \alpha_p^2)$. As the estimate of each unique variance is equal to the square of the estimated unique factor loading, it is constrained to be nonnegative. Rindskopf [28] used Bentler and Weeks approach in a variety of other situations besides Heywood cases, such as constraining parameters to be greater than a specified value or imposing ordered inequalities.

Thus far, existing strategies for avoiding improper solutions have been heavily concentrated on the Heywood case, which refers to a negative estimate of unique variance $\mathbf{\Psi}$. However, in the structural equation model (4): $\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Psi}$, an improper solution also could be an exterior estimate of factor correlation $\mathbf{\Phi}$ (e.g. a negative estimate of an eigenvalue or a correlation estimate out of range). To ensure an interior estimate of unique variance $\mathbf{\Psi}$ does not guarantee all the other estimates of the components of $\mathbf{\Sigma}$ are proper.

Bentler and Jamshidian [29] discussed the issue that factor correlation estimates of $\mathbf{\Phi}$ can be improper. Bentler and Jamshidian [29] obtained estimates of $\mathbf{\Lambda}$, $\mathbf{\Phi}$ and $\mathbf{\Psi}$ by minimizing the maximum likelihood discrepancy function under constraints on eigenvalues of $\mathbf{\Phi}$. The maximum likelihood discrepancy function Bentler and Jamshidian [29] used is $F(\mathbf{\Sigma}, \mathbf{S}) = \text{trace}(\mathbf{\Sigma}^{-1}\mathbf{S}) - \ln|\mathbf{\Sigma}^{-1}\mathbf{S}| - p$, which is the discrepancy function (b) in (7). The constraints Bentler and Jamshidian [29] used were to set all the eigenvalues of $\mathbf{\Phi}$ to be non-negative. From the examples given in Bentler and Jamshidian [29], it can be concluded that Bentler and Jamshidian [29]'s approach allows boundary estimates. Specifically, one or more eigenvalue estimates can be 0. If one or more eigenvalue estimates of $\mathbf{\Phi}$ are 0, then factor correlation matrix $\mathbf{\Phi}$ is no longer full rank, therefore, one of the factors is linearly dependent on the others, which can cause interpretation problems. In Bentler and Jamshidian [29], the numerical solution was provided; but no standard errors for the estimators were given. Also, none of asymptotic theory for the discrepancy function was provided, therefore, no tests, such as goodness of fit tests, can be performed.

After surveying existing strageties for avoiding improper solutions, it is worthwhile to investigate how existing programs handle improper solutions. The following discussion is based on the most current available versions of the two program packages R-2.15.1 [30] and SAS(STAT 9.2). These two packages are examined with respect to estimates of $\mathbf{\Phi}$ and $\mathbf{\Psi}$.

1.5.3. Factor Analysis Programs in R and SAS

First, the *sem* package [31] in R can be used to estimate EFA and CFA parameters. Specifically, the EFA model can be specified using the *factanal( )* function with various orthogonal and oblique rotation options. The argument *lower* in *factanal( )* sets the lower bound for estimates in $\boldsymbol{\psi}$ to be 0.005 by default, where $\boldsymbol{\psi}$ is given in (3). The *GPFoblq( )* function in the *GPArotation* package can be used to obtain the rotated loadings matrix and the covariance matrix of the rotated factors based on an initial loadings matrix and a given rotation criteria. The CFA model can be specified using the *specifyModel()* function. But no options in *specifyModel()* are provided so that estimates for $\mathbf{\Phi}$ or $\mathbf{\Psi}$ are proper.

Secondly, PROC CALIS or PROC FACTOR in SAS can be used to fit the EFA model, but only PROC CALIS can be used to fit the CFA model. In PROC FACTOR, the option HEYWOOD|HEY sets any communality greater than 1 to 1, where communality is defined in (15). When a communality equals 1, the corresponding uniqueness is 0. This means that HEYWOOD|HEY in PROC FACTOR allows improper estimate 0, although it prevents Heywood cases. In PROC CALIS, either the LINEQS statement or the FACTOR statement can be used to estimate EFA or CFA parameters. No option is mentioned in the LINEQS statement regarding any constraint for estimates of $\mathbf{\Phi}$ or $\mathbf{\Psi}$. The option HEYWOOD|HEY under the FACTOR statement in PROC CALIS constrains the unique variance estimates to be nonnegative, which prevents Heywood cases, but allows improper estimates 0. In summary, neither PROC FACTOR nor PROC CALIS provides options to constrain estimates for $\mathbf{\Phi}$ and $\mathbf{\Psi}$ to be proper. MacCallum [32] provided a discussion on SAS, BMDP, and SPSS, with respect to how, and whether, those packages treat improper estimates of $\mathbf{\Phi}$ and Heywood cases. Nevertheless, SAS should be used with caution for factor analysis because SAS allows improper estimates of $\mathbf{\Phi}$ and $\mathbf{\Psi}$.

In Chapter 5 of this thesis, examples are given to demonstrate proper estimates generated by R and SAS, and also examples are given to demonstrate improper estimates generated by R and SAS.

In closing, none of the existing strategies or computer packages discussed here completely eliminate improper solutions in factor analysis, which leads to the purpose of this dissertation.

1.6. Dissertation Outline

The goal of the present dissertation is to provide a solution such that all the estimates of the components of $\mathbf{\Sigma}$ are proper, which includes proper estimates of eigenvalues of $\mathbf{\Phi}$. The following includes the tasks of the present dissertation:

(1.) Develop statistical theory to perform hypotheses tests, such as goodness of fit and model comparison tests to determine the number of factors, structure of $\mathbf{\Lambda}$ and structure of $\mathbf{\Phi}$, especially under non-normality of responses.

(2.) Develop statistical theory to compute confidence regions for functions of parameters, under non-normality of responses.

(3.) Develop a stable algorithm to estimate parameters, which would be robust to data set variability and converge with high probability.

(4.) Apply to real data sets and compare to conventional methods.

CHAPTER 2

MODEL AND PARAMETERIZATION

2.1. Model and Parameterization

In this chapter, principal component models are applied to confirmatory factor analysis. Parameterizations of factor loadings and unique variances are presented. A parameterization of eigenvalues and eigenvectors of the correlation matrix among factors also is given.

2.2. Embed Principal Component Structure into CFA

In Chapter 1, the basic factor analysis model, $\mathbf{y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f} + \boldsymbol{\epsilon}$, was given in (2). In this model, the $p \times p$ covariance matrix of $\mathbf{y}$ in (4), is $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$. To embed the principal component structure into CFA, one can write the $q \times q$ factor correlation matrix $\boldsymbol{\Phi}$ in diagonal form: $\boldsymbol{\Phi} = \boldsymbol{\Gamma}\boldsymbol{\Delta}\boldsymbol{\Gamma}'$, where $\boldsymbol{\Delta} = \mathrm{Diag}(\delta_1, \cdots, \delta_q)$ is a matrix of eigenvalues $\boldsymbol{\Phi}$ and $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_q)$ is a matrix of eigenvectors of $\boldsymbol{\Phi}$. Then, the $p \times p$ covariance matrix $\boldsymbol{\Sigma}$ can be written as follows:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\Delta}\boldsymbol{\Gamma}'\boldsymbol{\Lambda}' + \boldsymbol{\Psi}. \tag{21}$$

Let $\boldsymbol{\theta}_{\boldsymbol{\lambda}}$, $\boldsymbol{\theta}_{\boldsymbol{\delta}}$, $\boldsymbol{\theta}_{\gamma}$ and $\boldsymbol{\theta}_{\psi}$ be vectors of unknown parameters. The quantities, $\boldsymbol{\Lambda}$, $\boldsymbol{\Delta}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Psi}$ are parameterized as functions of $\boldsymbol{\theta}_{\boldsymbol{\lambda}}$, $\boldsymbol{\theta}_{\boldsymbol{\delta}}$, $\boldsymbol{\theta}_{\gamma}$ and $\boldsymbol{\theta}_{\psi}$ such that $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\boldsymbol{\theta}_{\boldsymbol{\lambda}})$, $\boldsymbol{\Delta} = \boldsymbol{\Delta}(\boldsymbol{\theta}_{\boldsymbol{\delta}})$, $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}(\boldsymbol{\theta}_{\boldsymbol{\delta}}, \boldsymbol{\theta}_{\gamma})$ and $\boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{\theta}_{\psi})$. The parameters are arranged in a vector,

$\boldsymbol{\theta}$:

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_{\boldsymbol{\lambda}} \\ \boldsymbol{\theta}_{\boldsymbol{\delta}} \\ \boldsymbol{\theta}_{\gamma} \\ \boldsymbol{\theta}_{\psi} \end{pmatrix}, \text{ where } \begin{pmatrix} \dim(\boldsymbol{\theta}_{\boldsymbol{\lambda}}) \\ \dim(\boldsymbol{\theta}_{\boldsymbol{\delta}}) \\ \dim(\boldsymbol{\theta}_{\gamma}) \\ \dim(\boldsymbol{\theta}_{\psi}) \end{pmatrix} = \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \end{pmatrix}$$

and $\dim(\cdot)$ stands for the dimension of vector. Define ν as $\nu = \nu_1 + \nu_2 + \nu_3 + \nu_4$. For notational convenience, rewrite $\boldsymbol{\theta}$ as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1' \ \ \boldsymbol{\theta}_2' \ \ \boldsymbol{\theta}_3' \ \ \boldsymbol{\theta}_4')' = (\boldsymbol{\theta}_\lambda' \ \ \boldsymbol{\theta}_\delta' \ \ \boldsymbol{\theta}_\gamma' \ \ \boldsymbol{\theta}_\psi')'$, where $\boldsymbol{\theta}_i$ has dimension $\nu_i \times 1$.

An example for derivative notations of this thesis is given next. Let $\mathbf{W}$ be a matrix function of $\boldsymbol{\theta}$. Derivatives of $\mathbf{W}$ with respect to $\boldsymbol{\theta}$ are denoted as

$$\begin{gathered} \mathbf{D}^{(1)}_{\mathbf{W};\boldsymbol{\theta}} \stackrel{\text{def}}{=} \frac{\partial \mathbf{W}}{\partial \boldsymbol{\theta}}, \quad \mathbf{D}^{(2)}_{\mathbf{W};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2} \stackrel{\text{def}}{=} \frac{\partial^2 \mathbf{W}}{\partial \boldsymbol{\theta}_1 \otimes \partial \boldsymbol{\theta}_2}, \\ \mathbf{D}^{(3)}_{\mathbf{W};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\boldsymbol{\theta}_3} \stackrel{\text{def}}{=} \frac{\partial^3 \mathbf{W}}{\partial \boldsymbol{\theta}_1 \otimes \partial \boldsymbol{\theta}_2 \otimes \partial \boldsymbol{\theta}_3}, \quad \text{and so forth.} \end{gathered} \tag{22}$$

2.3. Parameterization of CFA Model

2.3.1. Parameterization of Factor Loadings $\boldsymbol{\Lambda}$

The structure of factor loadings $\boldsymbol{\Lambda}$ can be divided into two parts, one part is known and the other is unknown. In CFA, the known part of $\boldsymbol{\Lambda}$ comes from prior knowledge. Let $\boldsymbol{\lambda} = \text{vec}(\boldsymbol{\Lambda})$ with $\dim(\boldsymbol{\lambda}) = pq \times 1$. The structure of $\boldsymbol{\lambda}$ can be expressed as

$$\boldsymbol{\lambda} = \mathbf{W}_1 \mathbf{L}_1 + \mathbf{W}_2 \boldsymbol{\theta}_{\boldsymbol{\lambda}}, \tag{23}$$

where $\mathbf{W}_1$ and $\mathbf{W}_2$ are known design matrices of full column-rank, $\mathbf{L}_1$ is a known vector and $\boldsymbol{\theta}_{\boldsymbol{\lambda}}$ is an unknown vector with $\dim(\boldsymbol{\theta}_{\boldsymbol{\lambda}}) = \nu_1$. Typically, the components of $\mathbf{W}_1$ and $\mathbf{W}_2$ are 0's and 1's. Without loss of generality, the matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ can be assumed to satisfy $\mathbf{W}_2'\mathbf{W}_1 = \mathbf{0}$. A proof is as follows:

$$\boldsymbol{\lambda} \ = \ \mathbf{W}_1 \mathbf{L}_1 + \mathbf{W}_2 \boldsymbol{\theta}_{\boldsymbol{\lambda}}$$

$$
\begin{aligned}
&= \mathbf{H}_2\mathbf{W}_1\mathbf{L}_1 + (\mathbf{I}_{pq} - \mathbf{H}_2)\mathbf{W}_1\mathbf{L}_1 + \mathbf{W}_2\boldsymbol{\theta}_{\boldsymbol{\lambda}} \\
&\quad \text{because } \mathbf{W}_1 = \mathbf{H}_2\mathbf{W}_1 + (\mathbf{I}_{pq} - \mathbf{H}_2)\mathbf{W}_1, \text{ where } \mathbf{H}_2 = \text{ppo}(\mathbf{W}_2) \\
&= (\mathbf{I}_{pq} - \mathbf{H}_2)\mathbf{W}_1\mathbf{L}_1 + \mathbf{W}_2\boldsymbol{\theta}^*_{\boldsymbol{\lambda}} \\
&= \mathbf{W}^*\mathbf{L}_1 + \mathbf{W}_2\boldsymbol{\theta}^*_{\boldsymbol{\lambda}},
\end{aligned}
$$

where $\mathbf{W}^* = (\mathbf{I}_{pq} - \mathbf{H}_2)\mathbf{W}_1$, $\boldsymbol{\theta}^*_{\boldsymbol{\lambda}} = \boldsymbol{\theta}_{\boldsymbol{\lambda}} + (\mathbf{W}_2'\mathbf{W}_2)^{-1}\mathbf{W}_2'\mathbf{W}_1\mathbf{L}_1$, $\mathbf{W}_2'\mathbf{W}^* = \mathbf{0}$ and ppo$(\cdot)$ is defined in Table 56.

It follows from (23) that the parameter vector $\boldsymbol{\theta}_{\boldsymbol{\lambda}}$ can be expressed as $\boldsymbol{\theta}_{\boldsymbol{\lambda}} = (\mathbf{W}_2'\mathbf{W}_2)^{-1}\mathbf{W}_2'\boldsymbol{\lambda}$.

For example, suppose that the factor loading matrix $\boldsymbol{\Lambda}$ is

$$
\boldsymbol{\Lambda} = \begin{pmatrix} 1 & 0 \\ 0.73 & 0.22 \\ 0.81 & 0.38 \\ 0 & 1 \\ 0.27 & 0.90 \\ 0.13 & 0.77 \end{pmatrix}.
$$

Assume that the values at coordinates $(1,1)$, $(1,2)$,$(4,1)$ and $(4,2)$ in $\boldsymbol{\Lambda}$ are known and that the remaining eight values in $\boldsymbol{\Lambda}$ are unknown. The matrices $\mathbf{W}_1$, $\mathbf{L}_1$ and

$\mathbf{W}_2$ in (23) are the following:

$$\mathbf{W}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \mathbf{L}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \mathbf{W}_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \text{ and } \boldsymbol{\theta}_{\boldsymbol{\lambda}} = \begin{pmatrix} 0.73 \\ 0.81 \\ 0.27 \\ 0.13 \\ 0.22 \\ 0.38 \\ 0.90 \\ 0.77 \end{pmatrix}.$$

Note that the columns of $\mathbf{W}_1$ that correspond to 0 in $\mathbf{L}_1$ can be dropped. Also, column permutations of $\mathbf{W}_1$ and $\mathbf{W}_2$ and row permutations of $\mathbf{L}_1$ and $\boldsymbol{\theta}_{\boldsymbol{\lambda}}$ are allowed because $\mathbf{W}_1\mathbf{L}_1 = \mathbf{W}_1\mathbf{P}_1'\mathbf{P}_1\mathbf{L}_1$ and $\mathbf{W}_2\boldsymbol{\theta}_{\boldsymbol{\lambda}} = \mathbf{W}_2\mathbf{P}_2'\mathbf{P}_2\boldsymbol{\theta}_{\boldsymbol{\lambda}}$, where $\mathbf{P}_1$ and $\mathbf{P}_2$ are permutation matrices.

The first three derivatives of $\boldsymbol{\lambda}$ with respect to $\boldsymbol{\theta}_{\boldsymbol{\lambda}}$ are

$$\mathbf{D}^{(1)}_{\boldsymbol{\lambda};\boldsymbol{\theta}'_\lambda} = \mathbf{W}_2, \quad \mathbf{D}^{(2)}_{\boldsymbol{\lambda};\boldsymbol{\theta}'_\lambda,\boldsymbol{\theta}'_\lambda} = \mathbf{0}_{pq\times\nu_1^2}, \quad \text{and} \quad \mathbf{D}^{(3)}_{\boldsymbol{\lambda};\boldsymbol{\theta}'_\lambda,\boldsymbol{\theta}'_\lambda,\boldsymbol{\theta}'_\lambda} = \mathbf{0}_{pq\times\nu_1^3}. \tag{24}$$

2.3.2. Parameterization of Eigenvalues of $\mathbf{\Phi}$

Arrange the eigenvalues of the factor correlation matrix $\mathbf{\Phi}$ into a $q \times 1$ vector $\boldsymbol{\delta} = \operatorname{diag}(\mathbf{\Delta})$, that is, $\mathbf{\Delta} = \operatorname{Diag}(\boldsymbol{\delta})$, where $\operatorname{diag}(\cdot)$ and $\operatorname{Diag}(\cdot)$ are defined in Table 56. Note that

$$\operatorname{vec}(\mathbf{\Delta}) = \operatorname{vec}\left(\sum_{j=1}^{q} \mathbf{e}_j^q \delta_j \mathbf{e}_j^{q\prime}\right) = \sum_{j=1}^{q} \left(\mathbf{e}_j^q \otimes \mathbf{e}_j^q\right) \delta_j = \sum_{j=1}^{q} \left(\mathbf{e}_j^q \otimes \mathbf{e}_j^q\right) \mathbf{e}_j^{q\prime} \boldsymbol{\delta} = \mathbf{L}_{21,q}\boldsymbol{\delta},$$

where $\mathbf{e}_j^q$ and $\mathbf{L}_{21,q}$ are defined in Table 56. It follows from $\mathbf{L}_{21,q}'\mathbf{L}_{21,q} = \mathbf{I}_q$ that $\boldsymbol{\delta} = \mathbf{L}_{21,q}' \operatorname{vec}(\boldsymbol{\Delta})$.

It is assumed that the q-vector of eigenvalues, $\boldsymbol{\delta}$, is a differentiable function of $\boldsymbol{\theta}_{\boldsymbol{\delta}}$, where $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ is a $\nu_2 \times 1$ vector of unknown parameters. Furthermore, $\boldsymbol{\delta} = \boldsymbol{\delta}(\boldsymbol{\theta}_{\boldsymbol{\delta}})$ must satisfy $\mathbf{1}_q'\boldsymbol{\delta} = q$ because $\operatorname{tr}(\boldsymbol{\Phi}) = q$.

Table 4 provides a list of possible structures for eigenvalue parameterizations. Structure 1a to structure 3b were proposed by Boik [33]. From structure 1b to 3b, if there are no constraints imposed on $\boldsymbol{\delta}$, then $\boldsymbol{\xi}_\delta$ is equal to $\boldsymbol{\theta}_{\boldsymbol{\delta}}$. Otherwise, the parameter $\boldsymbol{\xi}_\delta$ is an implicit function of $\boldsymbol{\theta}_{\boldsymbol{\delta}}$, which can be shown by the implicit function theorem and also was described by Boik [34]. In general, the relationship between $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ and $\boldsymbol{\xi}_\delta$ depends on the constraints imposed.

For structure 1a, structure 1b and structure 4, $\mathbf{T}_2$ is a known full column-rank design matrix with dimension $q \times q_2$ and $\odot$ is defined in Table 56. For structure 2a to structure 3b, $\mathbf{T}_1$ and $\mathbf{T}_2$ are known full column-rank design matrices with dimension $q \times q_1$ and $q_1 \times q_2$, respectively. In Table 4, each of $\mathbf{C}_1 : q \times r_c$ and $\mathbf{c}_0 : r_c \times 1$ is a full column-rank matrix of known constants and stated dimensions.

Structure 4 is written as follows:

$$\boldsymbol{\delta} = \mathbf{T}_2\boldsymbol{\xi}_{\boldsymbol{\delta}}^{\odot 2} = \mathbf{T}_2(\boldsymbol{\xi}_\delta \odot \boldsymbol{\xi}_\delta) = \mathbf{T}_2 \begin{pmatrix} \boldsymbol{\xi}_{\delta 1}^{\,2} \\ \vdots \\ \boldsymbol{\xi}_{\delta q_2}^{\,2} \end{pmatrix}, \tag{25}$$

where the design matrix $\mathbf{T}_2$ has no negative entries. It follows that all entries in $\boldsymbol{\delta}$ are non-negative. Furthermore, $\boldsymbol{\Phi}$ is non-negative definite. Structure 4 in Table 4 is motivated by Bentler and Weeks' [27] idea on how to parameterize nonnegative unique variances in $\boldsymbol{\Psi}$. For the remainder of this section, all the discussions are based on structure 4 in Table 4. Structure 4 is used in §5.2.2.1, which is used to generate nonnegative eigenvalue estimates for $\boldsymbol{\Phi}$.

Table 4: Eigenvalue Structures for Correlation Matrices

Structure for $\boldsymbol{\delta}$	Optional Constraints	Requirements on $\mathbf{T}_1$, $\mathbf{T}_2$ & $\mathbf{C}_1$
1a. $\mathbf{T}_2\boldsymbol{\xi}_\delta$	$\mathbf{C}_1'\boldsymbol{\delta} = \mathbf{c}_0$	$\mathbf{1}_q'\mathbf{T}_2 \neq \mathbf{0}$ $\exists \mathbf{h} \ni \mathbf{h}'\mathbf{C}_1' = \mathbf{1}_q'$ and $\mathbf{h}'\mathbf{c}_0 = q$ $\text{rank}(\mathbf{C}_1'\mathbf{T}_2) = r_c$
1b. $\mathbf{1}_q + \mathbf{T}_2\boldsymbol{\xi}_\delta$	$\mathbf{C}_1'\boldsymbol{\delta} = \mathbf{c}_0$	$\mathbf{1}_q'\mathbf{T}_2 = \mathbf{0}$ $\mathbf{C}_1'\mathbf{1}_q = \mathbf{0}$ $\text{rank}(\mathbf{C}_1'\mathbf{T}_2) = r_c$
2a. $\dfrac{q\mathbf{T}_1 \exp_\odot\{\mathbf{T}_2\boldsymbol{\xi}_\delta\}}{\mathbf{1}_q'\mathbf{T}_1 \exp_\odot\{\mathbf{T}_2\boldsymbol{\xi}_\delta\}}$	$\mathbf{C}_1'\boldsymbol{\delta} = \mathbf{c}_0$	$\mathbf{1}_q'\mathbf{T}_1 \neq \mathbf{0}$ $\mathbf{1}_{q_1}'\mathbf{T}_2 = \mathbf{0}$ $\text{rank}(\mathbf{C}'\mathbf{T}_1) = r_c$ $\mathbf{C} = \mathbf{C}_1 - \mathbf{1}_q q^{-1}\mathbf{c}_0'$
2b. $\dfrac{q\mathbf{T}_1 \exp_\odot\{\mathbf{T}_2\boldsymbol{\xi}_\delta\}}{\mathbf{1}_q'\mathbf{T}_1 \exp_\odot\{\mathbf{T}_2\boldsymbol{\xi}_\delta\}}$	$\mathbf{C}_1' \ln_\odot(\boldsymbol{\delta}) = \mathbf{c}_0$	$\mathbf{1}_q'\mathbf{T}_1 \neq \mathbf{0}$ $\mathbf{1}_{q_1}'\mathbf{T}_2 = \mathbf{0}$ $\text{rank}(\mathbf{C}_1'\mathbf{T}_1) = r_c$
2c. $\dfrac{q\mathbf{T}_1 \exp_\odot\{\mathbf{T}_2\boldsymbol{\xi}_\delta\}}{\mathbf{1}_q'\mathbf{T}_1 \exp_\odot\{\mathbf{T}_2\boldsymbol{\xi}_\delta\}}$	$\mathbf{C}_1' \ln_\odot\left(\dfrac{\boldsymbol{\delta}}{\lvert\boldsymbol{\Phi}\rvert^{\frac{1}{q}}}\right) = \mathbf{c}_0$	$\mathbf{1}_q'\mathbf{T}_1 \neq \mathbf{0}$ $\mathbf{1}_{q_1}'\mathbf{T}_2 = \mathbf{0}$ $\text{rank}(\mathbf{C}_1'\mathbf{T}_1) = r_c$
3a. $\dfrac{q \exp_\odot\{\mathbf{T}_1 \exp_\odot[\mathbf{T}_2\boldsymbol{\xi}_\delta]\}}{\mathbf{1}_q' \exp_\odot\{\mathbf{T}_1 \exp_\odot[\mathbf{T}_2\boldsymbol{\xi}_\delta]\}}$	$\mathbf{C}_1' \ln_\odot(\boldsymbol{\delta}) = \mathbf{c}_0$	$\mathbf{1}_q'\mathbf{T}_1 = \mathbf{0}$ $\text{rank}(\mathbf{C}_1'\mathbf{T}_1) = r_c$
3b. $\dfrac{q \exp_\odot\{\mathbf{T}_1 \exp_\odot[\mathbf{T}_2\boldsymbol{\xi}_\delta]\}}{\mathbf{1}_q' \exp_\odot\{\mathbf{T}_1 \exp_\odot[\mathbf{T}_2\boldsymbol{\xi}_\delta]\}}$	$\mathbf{C}_1' \ln_\odot\left(\dfrac{\boldsymbol{\delta}}{\lvert\boldsymbol{\Phi}\rvert^{\frac{1}{q}}}\right) = \mathbf{c}_0$	$\mathbf{1}_q'\mathbf{T}_1 = \mathbf{0}$ $\text{rank}(\mathbf{C}_1'\mathbf{T}_1) = r_c$
4. $\mathbf{T}_2\boldsymbol{\xi}_\delta^{\odot 2}$	$\mathbf{C}_1'\boldsymbol{\delta} = \mathbf{c}_0$	$\mathbf{T}_2$ has non-negative entries $\exists \mathbf{h} \ni \mathbf{h}'\mathbf{C}_1' = \mathbf{1}_q'$ and $\mathbf{h}'\mathbf{c}_0 = q$ $\text{rank}(\mathbf{C}_1'\mathbf{T}_2) = r_c$

Let $\mathbf{m} = (m_1 \ m_2 \ \ldots \ m_k)'$ be the vector of eigenvalue multiplicities, where $\sum_{i=1}^{k} m_i = q$. The multiplicity of a eigenvalue is the number of times of this eigenvalues appears in $\boldsymbol{\delta}$. For example, if $\boldsymbol{\delta} = (1.5 \ 1.5 \ 1 \ 0.5 \ 0.5)'$, then $\mathbf{m} = (2 \ 1 \ 2)'$. The parameterization in (25) allows arbitrary eigenvalue multiplicities.

The design matrix $\mathbf{T}_2$ in structure 4 can be used to model linear relationships among $\delta_1, \ldots, \delta_q$. One can manipulate the structure of $\mathbf{T}_2$ to obtain structures as follows:

(1) If the distinct eigenvalues are unordered, then $\mathbf{T}_2$ can be written as

$$\mathbf{T}_2 = \bigoplus_{i=1}^{k} \mathbf{1}_{m_i}, \text{ where } \sum_{i=i}^{k} m_i = q.$$

(2) If all eigenvalues are ordered from large to small with $q = 3$ and $\mathbf{m} = \mathbf{1}_3$, then $\mathbf{T}_2$ can be written as

$$\mathbf{T}_2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

(3) If all eigenvalues are ordered from large to small with $q = 4$ and $\mathbf{m} = \begin{pmatrix} 2 & 1 & 1 \end{pmatrix}'$, then $\mathbf{T}_2$ can be written as

$$\mathbf{T}_2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

From structure 1a to structure 3b, Boik [33] provided expressions for first, second, and third derivatives of $\boldsymbol{\delta}$ with respect to $\boldsymbol{\theta}_{\boldsymbol{\delta}}$. First, second and third derivatives of structure 4 are given in Theorem 2, which can be used for point estimation and statistical inferences, such as interval estimation in Chapter 4.

Theorem 2. [Adapted from Boik [35], Supplement, 2009, Theorem 13, Page 15]. *Let $\boldsymbol{\delta}$ be parameterized as structure 4 in Table 4. Define $\mathbf{D}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}$ and $\mathbf{W}_{\boldsymbol{\delta}}$ as*

$$\mathbf{D}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}} \stackrel{\text{def}}{=} \operatorname{Diag}(\boldsymbol{\xi}_{\delta}) \text{ and } \mathbf{W}_{\boldsymbol{\delta}} \stackrel{\text{def}}{=} \mathbf{C}_1'\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}. \tag{26}$$

Assume that $\mathbf{C}_1$ has been chosen such that the $r_c \times q_2$ matrix $\mathbf{W}_{\boldsymbol{\delta}}$ has full row-rank. Write the singular value decomposition of $\mathbf{W}_{\boldsymbol{\delta}}$ as follows:

$$\begin{aligned}
&\mathbf{W}_{\boldsymbol{\delta}} = \mathbf{U}_{\boldsymbol{\delta}}\mathbf{D}_{\boldsymbol{\delta}}\mathbf{V}_{\boldsymbol{\delta}}', \text{ where}\\
&\mathbf{U}_{\boldsymbol{\delta}} \in \mathcal{O}(r_c), \quad \mathbf{V}_{\boldsymbol{\delta}} = \begin{pmatrix} \mathbf{V}_{\boldsymbol{\delta},1} & \mathbf{V}_{\boldsymbol{\delta},2} \end{pmatrix} \in \mathcal{O}(q_2),\\
&\mathbf{V}_{\boldsymbol{\delta},1} \in \mathcal{O}(q_2, r_c), \quad \mathbf{D}_{\boldsymbol{\delta}} = \begin{pmatrix} \mathbf{D}_{\boldsymbol{\delta},1} & \mathbf{0}_{r_c\times(q_2-r_c)} \end{pmatrix}, \text{ and } \mathbf{D}_{\boldsymbol{\delta},1} \in \mathcal{D}^+(r_c).
\end{aligned} \tag{27}$$

Then,

a. *the parameter $\boldsymbol{\xi}_{\delta}$ can be written as $\boldsymbol{\xi}_{\delta} = \mathbf{V}_1\boldsymbol{\eta}_{\boldsymbol{\delta}} + \mathbf{V}_2\boldsymbol{\theta}_{\boldsymbol{\delta}}$, where $\boldsymbol{\eta}_{\boldsymbol{\delta}} = \mathbf{V}_1'\boldsymbol{\xi}_{\delta}$, $\boldsymbol{\theta}_{\boldsymbol{\delta}} = \mathbf{V}_2'\boldsymbol{\xi}_{\delta}$, $\mathbf{V}_1 = \mathbf{V}_{\boldsymbol{\delta},1}$, and $\mathbf{V}_2 = \mathbf{V}_{\boldsymbol{\delta},2}$;*

b. *the parameters $\boldsymbol{\eta}_{\boldsymbol{\delta}}$ and $\boldsymbol{\xi}_{\delta}$ are implicit functions of $\boldsymbol{\theta}_{\boldsymbol{\delta}}$. Therefore, $\partial\boldsymbol{\eta}_{\boldsymbol{\delta}}/\partial\boldsymbol{\theta}_{\boldsymbol{\delta}}'$ and $\partial\boldsymbol{\xi}_{\delta}/\partial\boldsymbol{\theta}_{\boldsymbol{\delta}}'$ exist;*

c. [Original result]. *the first three derivatives of $\boldsymbol{\delta}$ with respect to $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ can be written as follows:*

$$\begin{gathered}
\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta'} = 2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}\mathbf{D}^{(1)}_{\boldsymbol{\xi}_{\boldsymbol{\delta}};\boldsymbol{\theta}_\delta'} = 2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}\mathbf{V}_2,\\
\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} = 2\mathbf{T}_2\left(\mathbf{I}_{q_2} - \mathbf{P}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}\right)\left(\mathbf{V}_2' * \mathbf{V}_2'\right)', \text{ and}\\
\mathbf{D}^{(3)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} = 2\mathbf{T}_2\left(\mathbf{I}_{q_2} - \mathbf{P}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}\right)\left(\mathbf{D}^{(2)\prime}_{\boldsymbol{\xi}_{\boldsymbol{\delta}};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi}_{\boldsymbol{\delta}};\boldsymbol{\theta}_\delta'}\right)'\mathbf{J}_{\nu_2}, \text{ where}\\
\mathbf{D}^{(1)}_{\boldsymbol{\xi}_{\boldsymbol{\delta}};\boldsymbol{\theta}_\delta'} = \mathbf{V}_2, \quad \mathbf{D}^{(2)}_{\boldsymbol{\xi}_{\boldsymbol{\delta}};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} = -\mathbf{D}^{-1}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}\mathbf{P}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\xi}_{\boldsymbol{\delta}};\boldsymbol{\theta}_\delta'} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi}_{\boldsymbol{\delta}};\boldsymbol{\theta}_\delta'}\right)', \quad \mathbf{P}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}} = \mathbf{D}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}\mathbf{W}_{\boldsymbol{\delta}}^+\mathbf{C}_1'\mathbf{T}_2,
\end{gathered}$$

$*$ is the Khatri-Rao column-wise product, $\mathbf{J}_{\nu_2}$ and $\mathbf{N}_{\nu_2}$ are defined in Table 56 and $\mathcal{D}^+(r_c)$ is defined in Table 57; and

d. [Original result]. $\mathbf{P}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}$ *is a projection operator.*

The notation used in Theorem 2 is consistent for the remainder of the thesis, unless otherwise specified. The following Theorem 3, Theorem 4, Theorem 5 and Corollary 5.1 provide relevant properties of $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_{\delta}}$ and $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ and assume that $\mathbf{W}_{\boldsymbol{\delta}}$ has full row-rank.

Theorem 3. [Original result]. *Let $\boldsymbol{\delta}$ be parameterized as structure 4 in Table 4. Suppose that the $q \times q_2$ design matrix $\mathbf{T}_2$ has full column-rank. Then $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_{\delta}}$ has full column-rank if and only if $\boldsymbol{\xi}_{\boldsymbol{\delta},i} \neq 0$ for $i = 1, 2, \ldots, q_2$.*

Theorem 4. [Original result]. *Let $\boldsymbol{\delta}$ be parameterized as structure 4 in Table 4. If $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_{\delta}}$ does not have full column-rank, then $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ is not identified.*

Theorem 5 below provides discussions on the dimension of $\boldsymbol{\theta}_{\boldsymbol{\delta}}$.

Theorem 5. [Original result]. *Let $\boldsymbol{\delta}$ be parameterized as structure 4 in Table 4. If $\mathbf{T_2}$ has full column-rank and $\boldsymbol{\xi}_{\boldsymbol{\delta},i} \neq 0$ for $i = 1, 2, \ldots, q_2$, then*

a. The dimension of $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ is $\nu_2 = q_2 - r_c$

b. A special case of part a: if $\mathbf{T}_2 = \bigoplus_{i=1}^{k} \mathbf{1}_{m_i}$ and $\sum_{i=1}^{k} m_i = q$, then $\nu_2 = k - r_c$.

Corollary 5.1. [Original result]. *Let $\boldsymbol{\delta}$ be parameterized as structure 4 in Table 4. If $\mathbf{T_2} = \mathbf{I}_q$, no entry in $\boldsymbol{\xi}_\delta$ is $\mathbf{0}$, $\mathbf{C_1} = \mathbf{1}_q$ and $\mathbf{c}_0 = q$, then $\mathbf{V}_1 = \boldsymbol{\xi}_\delta / \sqrt{\boldsymbol{\xi}_\delta{}' \boldsymbol{\xi}_\delta}$, $\mathcal{R}(\mathbf{V}_2) = \mathcal{N}(\boldsymbol{\xi}_\delta{}')$, $\boldsymbol{\theta}_{\boldsymbol{\delta}} = \mathbf{0}$ and $\nu_2 = q - 1$.*

For structure 4 in (25), if there are one or more values in $\boldsymbol{\xi}_\delta$ that are $\mathbf{0}$, then the dimension of $\boldsymbol{\xi}_\delta$ needs to be modified so that $\boldsymbol{\xi}_{\boldsymbol{\delta},i} \neq 0$ for $i = 1, 2, \ldots, q_2$ and the structure of $\mathbf{T}_2$ needs to be modified accordingly. The reason for these modifications is as follows: if $\boldsymbol{\xi}_{\boldsymbol{\delta},i} \neq 0$ for $i = 1, 2, \ldots, q_2$ is not satisfied, then $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_{\delta}}$ does not have full column-rank by Theorem 3. Furthermore, if $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_{\delta}}$ does not have full column-rank, then $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ is not identified by Theorem 4. Therefore, if there are one or more values in $\boldsymbol{\xi}_\delta$ that are $\mathbf{0}$, then the corresponding column of $\mathbf{T}_2$ should be removed and $\boldsymbol{\xi}_\delta$ should be modified accordingly. This is illustrated by the following example.

Example. Suppose that

$$\mathbf{T}_2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \text{ and } \boldsymbol{\xi}_\delta = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}, \text{ then } \boldsymbol{\delta} = \mathbf{T_2}\boldsymbol{\xi}_\delta{}^{\odot 2} = \begin{pmatrix} \xi_1^2 + \xi_2^2 + \xi_3^2 \\ \xi_1^2 + \xi_2^2 \\ \xi_1^2 \end{pmatrix}. \tag{28}$$

If $\xi_3 = 0$ in (28), then

$$\boldsymbol{\delta} = \mathbf{T_2}\boldsymbol{\xi}_\delta{}^{\odot 2} = \mathbf{T}_\mathbf{2}^*\boldsymbol{\xi}_{\boldsymbol{\delta}}^{*\odot 2} = \begin{pmatrix} \xi_1^2 + \xi_2^2 \\ \xi_1^2 + \xi_2^2 \\ \xi_1^2 \end{pmatrix}, \text{ where } \mathbf{T}_2^* = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix} \text{ and } \boldsymbol{\xi}_{\boldsymbol{\delta}}^* = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}. \tag{29}$$

For the above example, $\mathbf{T}_2$ and $\boldsymbol{\xi}_\delta$ are modified to be $\mathbf{T}_2^*$ and $\boldsymbol{\xi}_{\boldsymbol{\delta}}^*$, respectively. Given there is only one constraint, $\mathbf{1}_3'\boldsymbol{\delta} = 3$, imposed for $\boldsymbol{\delta} = \mathbf{T_2}\boldsymbol{\xi}_\delta{}^{\odot 2}$, the dimension $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ is $\nu_2 = 1$ by Theorem 5.

2.3.3. Parameterization of Eigenvectors of $\boldsymbol{\Phi}$

The parameterization of eigenvectors of $\boldsymbol{\Phi}$ is adapted from the eigenvector parameterization in Boik [34]. Boik's eigenvector parameterization [34] is based on the trivial equality $\boldsymbol{\Gamma} = \mathbf{I}_q\boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma}$ is a matrix of eigenvectors of $\boldsymbol{\Phi}$. Let $\boldsymbol{\Gamma}_0$ be a $q \times q$ orthogonal matrix in a small neighborhood of $\boldsymbol{\Gamma}$, which means $\boldsymbol{\Gamma} \approx \boldsymbol{\Gamma}_0$. The identity matrix $\mathbf{I}_q$ can be written as $\mathbf{I}_q = \boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0'$. Then $\boldsymbol{\Gamma} = \mathbf{I}_q\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0'\boldsymbol{\Gamma}$. Let $\mathbf{G} = \boldsymbol{\Gamma}_0'\boldsymbol{\Gamma}$ and note that $\mathbf{G} \in \mathcal{O}(q)$. It follows that $\mathbf{G} \approx \mathbf{I}_q$ because $\boldsymbol{\Gamma} \approx \boldsymbol{\Gamma}_0$. To develop a parameterization for $\boldsymbol{\Gamma}$, the matrix $\boldsymbol{\Gamma}_0$ is treated as known whereas $\mathbf{G}$ is treated as unknown and is parameterized locally near $\mathbf{G} = \mathbf{I}_q$. The above approach employs the implicit function-based parameterization approach in Boik [18]. The assumption that the matrix $\mathbf{\Gamma}_0$ is treated as known is not made in Chapter 3 when parameter estimation is discussed.

By orthogonality, $\mathbf{G}$ is subject to $q(q+1)/2$ constraints. If each eigenvalue has multiplicity 1, then $\mathbf{G}$ can be parameterized using at most $q(q-1)/2$ parameters. These $q(q-1)/2$ parameters correspond to the $q(q-1)/2$ elements in the upper triangle of $\mathbf{G}$. To be consistent with the previous notation, let $\mathbf{m} = (m_1 \ m_2 \ \cdots \ m_k)'$ be the vector of eigenvalue multiplicities, where $\sum_{i=1}^{k} m_i = q$. The columns of $\mathbf{G}$ can be partitioned as follows:

$$\mathbf{G} = \begin{pmatrix} \underset{q\times m_1}{\mathbf{G}_1} & \underset{q\times m_2}{\mathbf{G}_2} & \cdots & \underset{q\times m_k}{\mathbf{G}_k} \end{pmatrix}. \tag{30}$$

Accordingly, the diagonal form of $\mathbf{\Phi}$ can be rewritten as follows:

$$\begin{aligned} \mathbf{\Phi} = \mathbf{\Gamma\Delta\Gamma}' = \mathbf{\Gamma}_0\mathbf{G\Delta G}'\mathbf{\Gamma}_0' = \sum_{i=1}^{k} \mathbf{\Gamma}_0\mathbf{G}_i\delta_i\mathbf{I}_{m_i}\mathbf{G}_i'\mathbf{\Gamma}_0' \\ = \mathbf{\Gamma}_0\left\{\sum_{i=1}^{k}\delta_i\mathbf{G}_i\mathbf{G}_i'\right\}\mathbf{\Gamma}_0' = \mathbf{\Gamma}_0\left\{\sum_{i=1}^{k}\delta_i\mathbf{G}_i\mathbf{Q}_i\mathbf{Q}_i'\mathbf{G}_i'\right\}\mathbf{\Gamma}_0', \end{aligned} \tag{31}$$

where $\mathbf{Q}_i \in \mathcal{O}(m_i)$.

Note that the correlation matrix, $\mathbf{\Phi}$, depends on $\mathbf{G}_j$ solely through $\mathbf{G}_j\mathbf{G}_j' = \mathbf{G}_j\mathbf{Q}_j\mathbf{Q}_j'\mathbf{G}_j'$. If any eigenvalue multiplicity exceeds one, say $m_j \geq 2$, then the columns of $\mathbf{G}_j$ can be rotated to annihilate the $m_j(m_j-1)/2$ elements just above the main diagonal of $\mathbf{G}_j$. Specifically, the orthogonal matrix $\mathbf{Q}_j$ that annihilates entries in $\mathbf{G}_j$ can be computed using the QR decomposition:

$$\mathbf{G}_j' = \mathbf{Q}_j\mathbf{R}_j, \tag{32}$$

where $\mathbf{Q}_j$ is orthogonal and $\mathbf{R}_j$ is upper triangular. Based on (32), it can be concluded that

$$\mathbf{G}_j\mathbf{Q}_j = \mathbf{R}_j', \tag{33}$$

where $\mathbf{R}_j'$ is lower triangular. Thus, the $m_j(m_j-1)/2$ elements in the upper right-hand corner of $\mathbf{G}_j$ have been annihilated. Accordingly, the $m_j(m_j-1)/2$ elements in $\mathbf{G}_j\mathbf{Q}_j$ above the main diagonal of $\mathbf{G}$ can be arranged to be zeros through permutation. Furthermore, the number of remaining parameters in $\mathbf{G}$ is

$$\nu_3^* = q(q-1)/2 - \sum_{i=1}^{k} m_i(m_i-1)/2 = \frac{1}{2}(q^2 - \mathbf{m}'\mathbf{m}). \tag{34}$$

Let $\boldsymbol{\theta}_{\gamma}^*$ be a parameter vector of length ν_3^*, whose components correspond to the non-annihilated elements that are above the main diagonal of $\mathbf{G}$. Let $\boldsymbol{\eta}^*$ be an implicit function of $\boldsymbol{\theta}_{\gamma}^*$, whose components correspond to the $q(q+1)/2$ elements that are on or below the main diagonal of $\mathbf{G}$. For example, if $q=4$ and the vector of multiplicities is $\mathbf{m} = (1 \;\; 3)'$, then $\mathbf{G}$ can be written as follows:

$$\mathbf{G} = \left(\begin{array}{c|ccc} \eta_1^* & \theta_{\gamma 1}^* & \theta_{\gamma 2}^* & \theta_{\gamma 3}^* \\ \eta_2^* & \eta_7^* & 0 & 0 \\ \eta_3^* & \eta_8^* & \eta_{12}^* & 0 \\ \eta_4^* & \eta_9^* & \eta_{13}^* & \eta_{16}^* \end{array}\right) = \left(\begin{array}{c|c} \underset{4\times 1}{\mathbf{G}_1} & \underset{4\times 3}{\mathbf{G}_2} \end{array}\right). \tag{35}$$

In Boik [34], $\operatorname{vec}\mathbf{G}$ was written as

$$\begin{aligned} \mathbf{g} &\stackrel{\text{def}}{=} \operatorname{vec}\mathbf{G} = \mathbf{A}_1\boldsymbol{\eta}_{\gamma}^* + \mathbf{A}_2\boldsymbol{\theta}_{\gamma}^*, \text{ where } \dim(\mathbf{A}_2) = q^2 \times \nu_3^*, \\ \mathbf{A}_1 &= \sum_{j=1}^{q}\sum_{i=j}^{q}(\mathbf{e}_j^q \otimes \mathbf{e}_i^q)\mathbf{e}_{f_{ij}}^{h\prime}, \quad h = \frac{q(q+1)}{2}, \quad f_{ij} = \frac{(2q-j)(j-1)}{2} + i, \end{aligned} \tag{36}$$

and $\mathbf{A}_2$ is a known indicator matrix. These matrices satisfy $\mathbf{A}_i'\mathbf{A}_i = \mathbf{I}$ for $i = 1,2$; $\mathbf{A}_1'\mathbf{g} = \boldsymbol{\eta}_{\gamma}^*$; $\mathbf{A}_2'\mathbf{g} = \boldsymbol{\theta}_{\gamma}^*$; and $\mathbf{A}_1'\mathbf{A}_2 = \mathbf{0}$. Also, note that $\dim(\mathbf{A}_1) = q^2 \times q(q+1)/2$. In addition to $\mathbf{G} \in \mathcal{O}(q)$, $\mathbf{G}$ must satisfy

$$\mathbf{e}_i^{q\prime}\boldsymbol{\Phi}\mathbf{e}_i^q = 1, \text{ for } i = 1, \ldots, q, \tag{37}$$

where $\boldsymbol{\Phi} = \boldsymbol{\Gamma}_0\mathbf{G}\boldsymbol{\Delta}\mathbf{G}'\boldsymbol{\Gamma}_0'$ and $\boldsymbol{\Delta} = \boldsymbol{\Delta}(\boldsymbol{\delta})$. Equation 37 can be reduced as follows:

$$\mathbf{e}_i^{q\prime}\boldsymbol{\Phi}\mathbf{e}_i^q = 1, \text{ for } i = 1, \ldots, q-1, \tag{38}$$

because $\boldsymbol{\delta}$ is parameterized to satisfy $\mathbf{1}_q'\boldsymbol{\delta} = q$. Let $\boldsymbol{\Phi} = \{\phi_{ij}\}$ for $i \in \{1, 2, \ldots, q\}$ and $j \in \{1, 2, \ldots, q\}$. Besides orthogonality, Boik [34] expressed these additional constraints as $\mathbf{1}_q'\boldsymbol{\delta} = q$ and

$$\mathbf{C}_3' \operatorname{vec}(\boldsymbol{\Phi}) = \begin{pmatrix} \phi_{11} - \phi_{qq} \\ \phi_{22} - \phi_{qq} \\ \vdots \\ \phi_{(q-1)(q-1)} - \phi_{qq} \end{pmatrix} = \mathbf{0}, \tag{39}$$

where $\mathbf{C}_3 = \mathbf{L}_{21,q}\mathbf{K}$, $\mathbf{K} = (\mathbf{I}_{q-1} \quad -\mathbf{1}_{q-1})'$, and $\mathbf{L}_{21,q}$ is defined in Table 56. Accordingly, imposing (39) reduces the number of identified parameters in $\mathbf{G}$ by $q-1$.

Similar to the approach described in Boik [34], the expression of $\operatorname{vec}\mathbf{G}$ can be written in the following manner: let $\mathbf{V} \in \mathcal{O}(\nu_3^*)$ such that $\mathbf{V} = (\mathbf{V}_3 \;\; \mathbf{V}_4)$, $\dim(\mathbf{V}_3) = \nu_3^* \times (q-1)$, $\dim(\mathbf{V}_4) = \nu_3^* \times \nu_3$, $\nu_3 = \nu_3^* - (q-1)$ and $\mathbf{V}_3'\mathbf{V}_4 = \mathbf{0}$. The equality $\boldsymbol{\theta}_\gamma^* = \mathbf{V}\mathbf{V}'\boldsymbol{\theta}_\gamma^* = (\mathbf{V}_3\mathbf{V}_3' + \mathbf{V}_4\mathbf{V}_4')\,\boldsymbol{\theta}_\gamma^*$ implies that

$$\mathbf{g} = \operatorname{vec}\mathbf{G} = (\mathbf{A}_1 \;\; \mathbf{A}_2\mathbf{V}_3)\,\boldsymbol{\eta}_\gamma + \mathbf{A}_2\mathbf{V}_4\boldsymbol{\theta}_\gamma = \mathbf{A}_1\boldsymbol{\eta}_{\gamma,1} + \mathbf{A}_2\mathbf{V}_3\boldsymbol{\eta}_{\gamma,2} + \mathbf{A}_2\mathbf{V}_4\boldsymbol{\theta}_\gamma, \tag{40}$$

where $\mathbf{g}$ is defined in (36), $\boldsymbol{\eta}_\gamma = (\boldsymbol{\eta}_{\gamma,1}' \;\; \boldsymbol{\eta}_{\gamma,2}')'$, $\boldsymbol{\eta}_{\gamma,1} = \boldsymbol{\eta}_\gamma^*$, $\boldsymbol{\eta}_{\gamma,2} = \mathbf{V}_3'\boldsymbol{\theta}_\gamma{}^*$, and $\boldsymbol{\theta}_\gamma = \mathbf{V}_4'\boldsymbol{\theta}_\gamma^*$. Thus far, there are two constraints imposed on $(\boldsymbol{\theta}_{\boldsymbol{\delta}}, \boldsymbol{\theta}_\gamma)$, $\mathbf{G}\mathbf{G}' = \mathbf{I}_q$ and $\mathbf{C}_3'\operatorname{vec}(\boldsymbol{\Phi}) = \mathbf{0}$. Theorem 6 is needed for deriving Theorem 9 results.

Theorem 6. [Boik [35], Supplement, 2009, Theorem 1, Page 60]. $\mathbf{G}\mathbf{G}' = \mathbf{I}_q \Leftrightarrow \mathbf{D}_q' \operatorname{vec}(\mathbf{G}\mathbf{G}' - \mathbf{I}_q) = \mathbf{0}$.

By Theorem 6, the two constraints imposed on $(\boldsymbol{\theta}_{\boldsymbol{\delta}}, \boldsymbol{\theta}_\gamma)$ can be written as follows:

$$\mathbf{h}(\boldsymbol{\eta}_\gamma; \boldsymbol{\theta}_{\boldsymbol{\delta}}, \boldsymbol{\theta}_\gamma) = \begin{pmatrix} \mathbf{D}_q' \operatorname{vec}(\mathbf{G}\mathbf{G}' - \mathbf{I}_q) \\ \mathbf{C}_3' \operatorname{vec}(\boldsymbol{\Phi}) \end{pmatrix} = \mathbf{0}. \tag{41}$$

The following results, which were provided by Boik [18], are useful for later computations.

$$\text{(a) } \mathbf{D}_q'\mathbf{A}_1 = \mathbf{I}_{\frac{q(q+1)}{2}}, \quad \text{(b) } \mathbf{A}_1\mathbf{D}_q'\mathbf{A}_2 = \mathbf{K}_{q,q}\mathbf{A}_2, \qquad \text{and (c) } \mathbf{A}_1\mathbf{D}_q' \operatorname{vec}\mathbf{I}_{q^2} = \operatorname{vec}\mathbf{I}_{q^2}, \tag{42}$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ are defined in (36). Theorem 7 and Theorem 8 are constructed for the use of later theorems. Theorem 7 is given first because Theorem 8 needs the result from Theorem 7.

Theorem 7. [Boik [35], Supplement, 2009, Theorem 1, Page 61]. *Define* $\mathbf{W}_\gamma$ *as*

$$\mathbf{W}_\gamma \stackrel{\text{def}}{=} \mathbf{C}_3'\left(\boldsymbol{\Gamma}\boldsymbol{\Delta} \otimes \boldsymbol{\Gamma}\right) 2\mathbf{N}_q^{\perp}\mathbf{A}_2, \tag{43}$$

where $\mathbf{C}_3'$ *is defined in (39),* $\mathbf{A}_2$ *is defined in (36) and* $\mathbf{N}_p^{\perp}$ *is defined in Table 56. Assume the* $r_k \times \nu_3^*$ *matrix* $\mathbf{W}_\gamma$ *has full row-rank, where* $r_k = q - 1$ *and* ν_3^* *is defined in (34). Write the singular value decomposition of* $\mathbf{W}_\gamma$ *as follows:*

$$\begin{aligned}
&\mathbf{W}_\gamma = \mathbf{U}_\gamma\mathbf{D}_\gamma\mathbf{V}_\gamma', \ \textit{where } \mathbf{U}_\gamma \in \mathcal{O}(r_k) \\
&\mathbf{V}_\gamma = \begin{pmatrix} \mathbf{V}_{\gamma,1} & \mathbf{V}_{\gamma,2} \end{pmatrix} \in \mathcal{O}(\nu_3^*), \quad \mathbf{V}_{\gamma,1} \in \mathcal{O}(\nu_3^*, r_k), \\
&\mathbf{D}_\gamma = \begin{pmatrix} \mathbf{D}_{\gamma,1} & \mathbf{0}_{r_k \times (\nu_3^* - r_k)} \end{pmatrix}, \ \textit{and } \mathbf{D}_{\gamma,1} \in \mathcal{D}^+(r_k).
\end{aligned} \tag{44}$$

Choose $\mathbf{V}_3$ *to be any matrix whose columns form a basis for* $\mathcal{R}(\mathbf{W}_\gamma')$*, then the matrix* $\partial\,\mathbf{h}(\boldsymbol{\eta}_\gamma; \boldsymbol{\theta}_\delta, \boldsymbol{\theta}_\gamma)/\partial\,\boldsymbol{\eta}_\gamma'\Big|_{\mathbf{G}=\mathbf{I}}$ *is nonsingular and* $\boldsymbol{\eta}_\gamma$ *is an implicit function of* $(\boldsymbol{\theta}_\delta, \boldsymbol{\theta}_\gamma)$ *in a neighborhood of* $\mathbf{G} = \mathbf{I}$.

For the remainder of the thesis, assume that $\mathbf{W}_\gamma$ has full row-rank and $\mathbf{V}_3$ form a basis for $\mathcal{R}(\mathbf{W}_\gamma')$. Theorem 8 is needed for later theorems, such as Theorem 9 and Theorem 10.

Theorem 8. [Boik [34], 2003, §2.3, Page 683]. $\mathbf{G} = \mathbf{I}_q \Longleftrightarrow \boldsymbol{\theta}_\gamma = \mathbf{0}$.

First, second and third derivatives of $\operatorname{vec}\boldsymbol{\Sigma}$ with respect to $\boldsymbol{\theta}$ are needed when estimating parameters and conducting statistical inferences. First, an expression for the first three derivatives of $\operatorname{vec}\mathbf{G}$ in (40) with respect to $\boldsymbol{\theta}_{\boldsymbol{\gamma}}$, evaluated at $\boldsymbol{\theta}_{\boldsymbol{\gamma}} = \mathbf{0}$ and $\boldsymbol{\Gamma}_0 = \boldsymbol{\Gamma}$, must be found, which are given in Theorem 9.

Theorem 9. [Boik [33], Supplement, 2010, Theorem 29, Page 84]. *The first three derivatives of* $\operatorname{vec}\mathbf{G}$ *in* (40) *with respect to* $\boldsymbol{\theta}_{\boldsymbol{\gamma}}$*, evaluated at* $\boldsymbol{\theta}_{\boldsymbol{\gamma}} = \mathbf{0}$ *and* $\boldsymbol{\Gamma}_0 = \boldsymbol{\Gamma}$*, can be written as follows:*

$$\begin{aligned}
\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\gamma}} &= 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{V}_4, \\
\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_{\gamma},\boldsymbol{\theta}'_{\gamma}} &= \left[(\mathbf{I}_{q^2} - \mathbf{P}_{\boldsymbol{\gamma}})\,\mathbf{A}_1\mathbf{D}'_q\mathbf{I}'_{q,3} + 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}^{+}_{\boldsymbol{\gamma}}\mathbf{C}'_3(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\right] \\
&\times \left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\gamma}}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\gamma}}\right), \ \textit{and} \\
\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}'_{\gamma},\boldsymbol{\theta}'_{\gamma},\boldsymbol{\theta}'_{\gamma}} &= -\left\{2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}^{+}_{\boldsymbol{\gamma}}\mathbf{C}'_3(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q] + (\mathbf{I}_{q^2} - \mathbf{P}_{\boldsymbol{\gamma}})\,\mathbf{A}_1\mathbf{D}'_q\mathbf{I}'_{q,3}\right\} \\
&\times \left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_{\gamma},\boldsymbol{\theta}'_{\gamma}}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\gamma}}\right)\mathbf{J}_{21,\nu_3}
\end{aligned}$$

where $\mathbf{P}_{\boldsymbol{\gamma}} = 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}^{+}_{\boldsymbol{\gamma}}\mathbf{C}'_3\,(\boldsymbol{\Gamma}\boldsymbol{\Delta}\otimes\boldsymbol{\Gamma})$, $\mathbf{D}_q$, $\mathbf{I}_{q,3}$, $\mathbf{J}_{21,\nu_3}$, $\mathbf{N}_q^{\perp}$ *are defined in Table 56 and* $\mathbf{W}^{+}_{\boldsymbol{\gamma}}$ *is the Moore-Penrose inverse of* $\mathbf{W}_{\boldsymbol{\gamma}}$*.*

Theorem 10. [Adapted from Boik [33], Supplement, 2010, Theorem 29, Page 84]. *Assume that* $\mathbf{W}_{\boldsymbol{\gamma}}$ *has full row-rank and* $\mathbf{V}_3 = \mathbf{V}_{\boldsymbol{\gamma},1}$*. The derivatives of* $\operatorname{vec}\mathbf{G}$ *in* (40) *with respect to* $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ *and* $\boldsymbol{\theta}_{\boldsymbol{\gamma}}$*, evaluated at* $\boldsymbol{\theta}_{\boldsymbol{\gamma}} = \mathbf{0}$ *and* $\boldsymbol{\Gamma}_0 = \boldsymbol{\Gamma}$*, can be written as follows:*

$$\begin{aligned}
\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta}} &= -\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}^{+}_{\boldsymbol{\gamma}}\mathbf{C}'_3(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_{\delta}}, \\
\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta},\boldsymbol{\theta}'_{\delta}} &= (\mathbf{I}_{q^2} - \mathbf{P}_{\boldsymbol{\gamma}})\,\mathbf{A}_1\mathbf{D}'_q\mathbf{I}'_{q,3}\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta}}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta}}\right) \\
&+ 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}^{+}_{\boldsymbol{\gamma}}\mathbf{C}'_3[\boldsymbol{\Gamma}\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\boldsymbol{\Gamma}]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta}}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta}}\right) \\
&- \mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}^{+}_{\boldsymbol{\gamma}}\mathbf{C}'_3(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\mathbf{L}_{21,q}\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_{\delta},\boldsymbol{\theta}'_{\delta}} \\
&- 4\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}^{+}_{\boldsymbol{\gamma}}\mathbf{C}'_3[\boldsymbol{\Gamma}\otimes\operatorname{vec}'(\mathbf{I}_q)\otimes\boldsymbol{\Gamma}](\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_{\delta}}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta}})\mathbf{N}_{\nu_2}, \\
\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta},\boldsymbol{\theta}'_{\gamma}} &= (\mathbf{I}_{q^2} - \mathbf{P}_{\boldsymbol{\gamma}})\,\mathbf{A}_1\mathbf{D}'_q\mathbf{I}'_{q,3}\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta}}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\gamma}}\right)
\end{aligned}$$

$$
\begin{aligned}
& + \quad 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes \mathrm{vec}'(\boldsymbol{\Delta})\otimes\boldsymbol{\Gamma}]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\\
& - \quad 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes \mathrm{vec}'(\mathbf{I}_q)\otimes\boldsymbol{\Gamma}](\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}),\\
\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} &= -\left\{(\mathbf{I}_{q^2}-\mathbf{P}_{\boldsymbol{\gamma}})\,\mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}' + 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes \mathrm{vec}'(\boldsymbol{\Delta})\otimes\boldsymbol{\Gamma}]\right\}\\
& \times \quad \left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\delta'}\right)\mathbf{J}_{21,\nu_2}\\
& - \quad 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes \mathrm{vec}'(\mathbf{I}_q)\otimes\boldsymbol{\Gamma}]\\
& \times \quad \left[\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'}\otimes\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta'}\right)+\left(\mathbf{L}_{21,q}\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\delta'}\right)\right]\mathbf{J}_{21,\nu_2}\\
& + \quad 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes \mathrm{vec}'(\mathbf{I}_q)\otimes\mathrm{vec}'(\mathbf{I}_q)\otimes\boldsymbol{\Gamma}]\\
& \times \quad \left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\delta'}\otimes\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\delta'}\right)[(\mathbf{K}_{\nu_2,\nu_2}\otimes\mathbf{I}_{\nu_2})+(\mathbf{I}_{\nu_2}\otimes 2\mathbf{N}_{\nu_2})]\\
& - \quad \mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\mathbf{L}_{21,q}\mathbf{D}^{(3)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'},\\
\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\gamma'} &= -\left\{(\mathbf{I}_{q^2}-\mathbf{P}_{\boldsymbol{\gamma}})\,\mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}' + 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes \mathrm{vec}'(\boldsymbol{\Delta})\otimes\boldsymbol{\Gamma}]\right\}\\
& \times \quad \left[\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\delta'}\right)\mathbf{J}_{21,\nu_2\nu_3}+\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\right]\\
& - \quad 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes \mathrm{vec}'(\mathbf{I}_q)\otimes\boldsymbol{\Gamma}]\\
& \times \quad \left[\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta'}\right)\mathbf{J}_{21,\nu_2\nu_3}+\left(\mathbf{L}_{21,q}\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\right]\\
& + \quad 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes \mathrm{vec}'(\mathbf{I}_q)\otimes\mathrm{vec}'(\mathbf{I}_q)\otimes\boldsymbol{\Gamma}]\\
& \times \quad \left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\delta'}\otimes\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\mathbf{J}^{*}_{21,\nu_2\nu_3}\\
& \quad\quad \textit{and}\\
\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} &= -\left\{(\mathbf{I}_{q^2}-\mathbf{P}_{\boldsymbol{\gamma}})\,\mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}' + 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes \mathrm{vec}'(\boldsymbol{\Delta})\otimes\boldsymbol{\Gamma}]\right\}\\
& \times \quad \left[\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\left(\mathbf{I}_{\nu_3^2\nu_2}+\mathbf{I}_{\nu_2}\otimes\mathbf{K}_{\nu_3,\nu_3}\right)-\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\right)\right]\\
& - \quad 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes \mathrm{vec}'(\mathbf{I}_q)\otimes\boldsymbol{\Gamma}](\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'})\\
& + \quad 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes \mathrm{vec}'(\mathbf{I}_q)\otimes\mathrm{vec}'(\mathbf{I}_q)\otimes\boldsymbol{\Gamma}]\\
& \times \quad \left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)(\mathbf{K}_{\nu_3,\nu_2}\otimes\mathbf{I}_{\nu_3})\,,
\end{aligned}
$$

where $\mathbf{P}_\gamma = 2\mathbf{N}_q^\perp \mathbf{A}_2 \mathbf{W}_\gamma^+ \mathbf{C}_3' (\boldsymbol{\Gamma\Delta} \otimes \boldsymbol{\Gamma})$, $\mathbf{I}_{q,3}$, $\mathbf{J}_{21,\nu_3}$, $\mathbf{N}_q^\perp$ *are defined in Table 56,* $\mathbf{W}_\gamma^+$ *is the Moore-Penrose inverse of* $\mathbf{W}_\gamma$, $\mathbf{J}_{21,\nu_2\nu_3} = \mathbf{K}_{\nu_2\nu_3,\nu_2} + \mathbf{I}_{\nu_2} \otimes \mathbf{K}_{\nu_3,\nu_2}$ *and* $\mathbf{J}^*_{21,\nu_2\nu_3} = \mathbf{I}_{\nu_2^2\nu_3} + \mathbf{K}_{\nu_2,\nu_2} \otimes \mathbf{I}_{\nu_3}$.

Theorem 10 is based on the results from Theorem 29 on page 84 in the supplement of Boik [33]. Certain substitutions are made for the use of this thesis. The proof is similar to that of Theorem 9 and is omitted. The derivative expressions in Theorem 10 were checked numerically.

2.3.4. Parameterization of Unique Variances $\boldsymbol{\Psi}$

Define the diagonal entries of the unique variances matrix $\boldsymbol{\Psi}$ by $\boldsymbol{\psi}$, that is, $\boldsymbol{\psi} = \text{diag}(\boldsymbol{\Psi})$.

Table 5 provides a list of possible structures for $\boldsymbol{\psi}$. Structure 1a to structure 4b were proposed by Boik [35]. From structure 1a to structure 4b, Boik [35] provided expressions for first, second, and third derivatives of $\boldsymbol{\psi}$ with respect to $\boldsymbol{\theta}_\psi'$ and discussed how to obtain initial guesses for $\boldsymbol{\xi}_\psi$, $\boldsymbol{\theta}_\psi$ and $\boldsymbol{\psi}$. From structure 1b to 3b, if there are no constraints imposed, then $\boldsymbol{\xi}_\psi$ is equal to $\boldsymbol{\theta}_\psi$. Otherwise, the parameter $\boldsymbol{\xi}_\psi$ is an implicit function of $\boldsymbol{\theta}_\psi$, which can be shown by the implicit function theorem and also was described by Boik [34]. In general, the relationship between $\boldsymbol{\theta}_\psi$ and $\boldsymbol{\xi}_\psi$ depends on the constraints imposed.

For structure 1a, structure 1b and structure 5, $\mathbf{T}_4$ is a known full column-rank design matrix with dimension $p \times p_2$ and $\odot$ is defined in Table 56. For structure 2a to structure 3b, $\mathbf{T}_3$ and $\mathbf{T}_4$ are known full column-rank design matrices with dimension $p \times p_1$ and $p_1 \times p_2$, respectively. The structure of structure 5 in Table 5 is motivated by Bentler and Weeks' [27] idea on how to parameterize nonnegative unique variances in $\boldsymbol{\Psi}$.

Table 5: Eigenvalue Structures for Unique Variances

Structure for $\boldsymbol{\psi}$	Optional Constraints	Requirements on $\mathbf{T}_3$, $\mathbf{T}_4$ & $\mathbf{C}_1$
1a. $\mathbf{T}_4\boldsymbol{\xi}_\psi$	$\mathbf{C}_1'\boldsymbol{\psi} = \mathbf{c}_0$	$\mathbf{1}_p \notin \mathcal{N}(\mathbf{T}_4')$ $\operatorname{rank}(\mathbf{C}_1'\mathbf{T}_4) = r_p$
1b. $\mathbf{T}_4\boldsymbol{\xi}_\psi$	$\dfrac{\mathbf{C}_1'\boldsymbol{\psi}}{\operatorname{tr}\boldsymbol{\Psi}} = \mathbf{c}_0$	$\mathbf{1}_p \notin \mathcal{N}(\mathbf{T}_4')$ $\operatorname{rank}(\mathbf{C}'\mathbf{T}_4) = r_p$ $\mathbf{C} = \mathbf{C}_1 - \mathbf{1}_p\mathbf{c}_0'$
2a. $\mathbf{T}_3 \exp_\odot\{\mathbf{T}_4\boldsymbol{\xi}_\psi\}$	$\mathbf{C}_1'\boldsymbol{\psi} = \mathbf{c}_0$	$\mathbf{1}_p \notin \mathcal{N}(\mathbf{T}_3')$ $\operatorname{rank}(\mathbf{C}_1'\mathbf{T}_3) = r_p$
2b. $\mathbf{T}_3 \exp_\odot\{\mathbf{T}_4\boldsymbol{\xi}_\psi\}$	$\mathbf{C}_1' \ln_\odot(\boldsymbol{\psi}) = \mathbf{c}_0$	$\mathbf{1}_p \notin \mathcal{N}(\mathbf{T}_3')$ $\operatorname{rank}(\mathbf{C}_1'\mathbf{T}_3) = r_p$
3. $\boldsymbol{\theta}_{\psi,1}\left(\dfrac{\mathbf{T}_3\exp_\odot\{\mathbf{T}_2\boldsymbol{\xi}_\psi\}}{\mathbf{1}_p'\mathbf{T}_3\exp_\odot\{\mathbf{T}_4\boldsymbol{\xi}_\psi\}}\right)$	$\dfrac{\mathbf{C}_1'\boldsymbol{\psi}}{\operatorname{tr}\boldsymbol{\Psi}} = \mathbf{c}_0$	$\mathbf{1}_p \notin \mathcal{N}(\mathbf{T}_3')$ $\mathbf{1}_{p_1} \in \mathcal{N}(\mathbf{T}_4')$ $\operatorname{rank}(\mathbf{C}'\mathbf{T}_3) = r_p$ $\mathbf{C} = \mathbf{C}_1 - \mathbf{1}_p\mathbf{c}_0'$
4a. $\boldsymbol{\xi}_{\psi,1}\exp_\odot\left\{\mathbf{T}_3\exp_\odot\left[\mathbf{T}_2\boldsymbol{\xi}_{\psi,2}\right]\right\}$	$\mathbf{C}_1' \ln_\odot(\boldsymbol{\psi}) = \mathbf{c}_0$	$\mathbf{1}_p \in \mathcal{N}(\mathbf{T}_3')$ $\operatorname{rank}(\mathbf{C}_1'\mathbf{T}_3) = r_p$
4b. $\boldsymbol{\theta}_{\psi,1}\exp_\odot\left\{\mathbf{T}_3\exp_\odot\left(\mathbf{T}_4\boldsymbol{\xi}_\psi\right)\right\}$	$\mathbf{C}_1' \ln_\odot\left(\dfrac{\boldsymbol{\psi}}{\lvert\boldsymbol{\Psi}\rvert^{\frac{1}{p}}}\right) = \mathbf{c}_0$	$\mathbf{1}_p \in \mathcal{N}(\mathbf{T}_3')$ $\operatorname{rank}(\mathbf{C}_1'\mathbf{T}_3) = r_p$
5. $\mathbf{T}_4\boldsymbol{\xi}_\psi^{\odot 2}$	$\mathbf{C}_1'\boldsymbol{\psi} = \mathbf{c}_0$	All entries in $\mathbf{T}_4$ are non-negative $\operatorname{rank}(\mathbf{C}_1'\mathbf{T}_4) = r_p$

Structure 5 is examined in details because this structure is new. The structure is written as follows:

$$\boldsymbol{\psi} = \mathbf{T}_4 \boldsymbol{\xi}_{\psi}^{\odot 2} = \mathbf{T}_4(\boldsymbol{\xi}_{\psi} \odot \boldsymbol{\xi}_{\psi}) = \mathbf{T}_4 \begin{pmatrix} {\boldsymbol{\xi}_{\psi}}_1^2 \\ \vdots \\ {\boldsymbol{\xi}_{\psi}}_{p_2}^2 \end{pmatrix}, \tag{45}$$

where the design matrix $\mathbf{T}_4$ has no negative entries. It follows that all the eigenvalues in $\boldsymbol{\psi}$ are non-negative. Accordingly, the factor correlation matrix $\boldsymbol{\Psi}$ is guaranteed to be non-negative definite. The design matrix $\mathbf{T}_4$ in (45) is used to model linear relationships within $\boldsymbol{\psi}$ and manipulate multiplicities of $\boldsymbol{\psi}$. In § 2.3.2, examples can be found regarding the structures of $\mathbf{T}_2$ in structure 4 and multiplicities of $\boldsymbol{\phi}$. Examples regarding the structure of $\mathbf{T}_4$ and multiplicities of $\boldsymbol{\psi}$ are omitted because of the similarity between the parameterization of eigenvalues of $\boldsymbol{\Phi}$ and parameterization of eigenvalues of $\boldsymbol{\Psi}$.

First, second, and third derivatives of structure 5 are used for point estimation and statistical inferences, such as interval estimation. Expressions for the derivatives for structure 5 are given in Theorem 11. The structure of Theorem 11 is adapted from Theorem 13 of Boik [35].

Theorem 11. [Original result]. *Define* $\mathbf{D}_{\xi_\psi}$ *and* $\mathbf{W}_\psi$ *as*

$$\mathbf{D}_{\xi_\psi} \overset{\text{def}}{=} \text{Diag}(\boldsymbol{\xi}_\psi) \text{ and } \mathbf{W}_\psi \overset{\text{def}}{=} \mathbf{C}_1' \mathbf{T}_4 \mathbf{D}_{\xi_\psi}. \tag{46}$$

Assume that no entry in $\boldsymbol{\xi}_\psi$ *is* $\mathbf{0}$ *and* $\mathbf{C}_1$ *has been chosen such that the* $r_p \times p_2$ *matrix* $\mathbf{W}_\psi$ *has full row-rank. Write the singular value decomposition of* $\mathbf{W}_\psi$ *as follows:*

$$\begin{aligned} &\mathbf{W}_\psi = \mathbf{U}_\psi \mathbf{D}_\psi \mathbf{V}_\psi', \text{ where} \\ &\mathbf{U}_\psi \in \mathcal{O}(r_p), \quad \mathbf{V}_\psi = \begin{pmatrix} \mathbf{V}_{\psi,1} & \mathbf{V}_{\psi,2} \end{pmatrix} \in \mathcal{O}(p_2), \\ &\mathbf{V}_{\psi,1} \in \mathcal{O}(p_2, r_p), \quad \mathbf{D}_\psi = \begin{pmatrix} \mathbf{D}_{\psi,1} & \mathbf{0}_{r_p \times (p_2 - r_p)} \end{pmatrix}, \text{ and } \mathbf{D}_{\psi,1} \in \mathcal{D}^+(r_p). \end{aligned} \tag{47}$$

Then,

a. *the parameter* $\boldsymbol{\xi}_{\psi}$ *can be written as* $\boldsymbol{\xi}_{\psi} = \mathbf{V}_1\boldsymbol{\eta}_{\psi} + \mathbf{V}_2\boldsymbol{\theta}_{\psi}$*, where* $\boldsymbol{\eta}_{\psi} = \mathbf{V}_1'\boldsymbol{\xi}_{\psi}$*,* $\boldsymbol{\theta}_{\psi} = \mathbf{V}_2'\boldsymbol{\xi}_{\psi}$*,* $\mathbf{V}_1 = \mathbf{V}_{\psi,1}$*, and* $\mathbf{V}_2 = \mathbf{V}_{\psi,2}$*;*

b. *the parameters* $\boldsymbol{\eta}_{\psi}$ *and* $\boldsymbol{\xi}_{\psi}$ *are implicit functions of* $\boldsymbol{\theta}_{\psi}$*. Therefore,* $\partial\boldsymbol{\eta}_{\psi}/\partial\boldsymbol{\theta}_{\psi}'$ *and* $\partial\boldsymbol{\xi}_{\psi}/\partial\boldsymbol{\theta}_{\psi}'$ *exist;*

c. *the first three derivatives of* $\boldsymbol{\psi}$ *with respect to* $\boldsymbol{\theta}_{\psi}$ *can be written as follows:*

$$\mathbf{D}^{(1)}_{\psi;\boldsymbol{\theta}_{\psi}'} = 2\mathbf{T}_4\mathbf{D}_{\boldsymbol{\xi}_{\psi}}\mathbf{D}^{(1)}_{\boldsymbol{\xi}_{\psi};\boldsymbol{\theta}_{\psi}'} = 2\mathbf{T}_4\mathbf{D}_{\boldsymbol{\xi}_{\psi}}\mathbf{V}_2,$$

$$\mathbf{D}^{(2)}_{\psi;\boldsymbol{\theta}_{\psi}',\boldsymbol{\theta}_{\psi}'} = 2\mathbf{T}_4\left(\mathbf{I}_{p_2} - \mathbf{P}_{\boldsymbol{\xi}_{\psi}}\right)(\mathbf{V}_2' * \mathbf{V}_2')', \text{ and}$$

$$\mathbf{D}^{(3)}_{\psi;\boldsymbol{\theta}_{\psi}',\boldsymbol{\theta}_{\psi}',\boldsymbol{\theta}_{\psi}'} = 2\mathbf{T}_4\left(\mathbf{I}_{p_2} - \mathbf{P}_{\boldsymbol{\xi}_{\psi}}\right)\left(\mathbf{D}^{(2)\prime}_{\boldsymbol{\xi}_{\psi};\boldsymbol{\theta}_{\psi}',\boldsymbol{\theta}_{\psi}'} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi}_{\psi};\boldsymbol{\theta}_{\psi}'}\right)'\mathbf{J}_{\nu_4}, \text{ where}$$

$$\mathbf{D}^{(1)}_{\boldsymbol{\xi}_{\psi};\boldsymbol{\theta}_{\psi}'} = \mathbf{V}_2, \quad \mathbf{D}^{(2)}_{\boldsymbol{\xi}_{\psi};\boldsymbol{\theta}_{\psi}',\boldsymbol{\theta}_{\psi}'} = -\mathbf{D}^{-1}_{\boldsymbol{\xi}_{\psi}}\mathbf{P}_{\boldsymbol{\xi}_{\psi}}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\xi}_{\psi};\boldsymbol{\theta}_{\psi}'} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi}_{\psi};\boldsymbol{\theta}_{\psi}'}\right)', \quad \mathbf{P}_{\boldsymbol{\xi}_{\psi}} = \mathbf{D}_{\boldsymbol{\xi}_{\psi}}\mathbf{W}_{\psi}^{+}\mathbf{C}_1'\mathbf{T}_4,$$

$*$ *is the Khatri-Rao column-wise product,* $\mathbf{J}_{\nu_4}$ *and* $\mathbf{N}_{\nu_4}$ *are defined in Table 56 and* $\mathcal{D}^{+}(r_p)$ *is defined in Table 57; and*

d. $\mathbf{P}_{\boldsymbol{\xi}_{\psi}}$ *is a projection operator.*

The proof is similar to that of Theorem 4 and is ommited. The derivative expressions in Theorem 11 were checked numerically. Relevant properties of $\mathbf{D}^{(1)}_{\psi;\boldsymbol{\theta}_{\psi}'}$ and $\boldsymbol{\theta}_{\psi}$ are given in Theorem 12, Theorem 13 and Theorem 14. Theorem 14 is given below regarding the dimension of $\boldsymbol{\theta}_{\psi}$.

Theorem 12. [Original result]. *Suppose that the* $p \times p_2$ *design matrix* $\mathbf{T}_4$ *has full column-rank. Then* $\mathbf{D}^{(1)}_{\psi;\boldsymbol{\theta}_{\psi}'}$ *has full column-rank if and only if* $\boldsymbol{\xi}_{\psi,i} \neq 0$ *for* $i = 1, 2, \ldots, p_2$.

Theorem 13. [Original result]. *If* $\mathbf{D}^{(1)}_{\psi;\boldsymbol{\theta}_{\psi}'}$ *does not have full column-rank, then* $\boldsymbol{\theta}_{\psi}$ *is not identified.*

Theorem 14. [Original result]. *If* $\mathbf{T_4}$ *has full column-rank and* $\boldsymbol{\xi}_{\psi,i} \neq 0$ *for* $i = 1, 2, \ldots, p_2$*, then*

a. The dimension of $\boldsymbol{\theta}_{\psi}$ *is* $\nu_4 = p_2 - r_p$

b. A special case of part a: if $\mathbf{T}_4 = \bigoplus_{i=1}^{k} \mathbf{1}_{m_i}$ *and* $\sum_{i=1}^{k} m_i = p$*, then* $\nu_4 = k - r_p$*.*

Theorem 11, Theorem 12, Theorem 13 and Theorem 14 are quite similar to Theorem 2, Theorem 3, Theorem 4 and Theorem 5, respectively. Therefore, the proofs of Theorem 11, Theorem 12, Theorem 13 and Theorem 14 are omitted. The notation used in this section is consistent for the remainder of the thesis, unless otherwise specified.

For structure 5 in (45), if there are one or more values in $\boldsymbol{\xi}_{\psi}$ that are $\mathbf{0}$, then the dimension of $\boldsymbol{\xi}_{\psi}$ needs to be reduced so that $\boldsymbol{\xi}_{\psi,i} \neq 0$ for $i = 1, 2, \ldots, p_2$ and the structure of $\mathbf{T}_4$ needs to be reconstructed. One can refer to Theorem 12 and Theorem 13 for reasons.

2.3.5. Derivatives of vec $\boldsymbol{\Sigma}$ With Respect To $\boldsymbol{\theta}$

Let $\boldsymbol{\sigma} = \text{vec}\,\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ is the $p \times p$ covariance matrix defined in (21). Recall that $\boldsymbol{\theta}$ is defined as be $\boldsymbol{\theta} = (\boldsymbol{\theta}'_{\lambda}\;\; \boldsymbol{\theta}'_{\delta}\;\; \boldsymbol{\theta}'_{\gamma}\;\; \boldsymbol{\theta}'_{\psi})'$ and $\boldsymbol{\nu}$ is defined as $\boldsymbol{\nu} = (\nu_1\;\; \nu_2\;\; \nu_3\;\; \nu_4)'$. Derivatives of $\boldsymbol{\sigma}$ with respect to $\boldsymbol{\theta}$ are needed when estimating parameters and conducting statistical inferences. Accordingly, Theorem 15 is needed.

Theorem 15. [Original result]. *First, second and third derivatives of* $\boldsymbol{\sigma}$ *with respect to* $\boldsymbol{\theta_\lambda}$*,* $\boldsymbol{\theta_\delta}$*,* $\boldsymbol{\theta_\gamma}$ *and* $\boldsymbol{\theta}_{\psi}$*, evaluated at* $\mathbf{G} = \mathbf{I}_q$*, are listed as follows:*

$$
\begin{aligned}
\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_{\lambda}} &= 2\mathbf{N}_p(\boldsymbol{\Lambda}\boldsymbol{\Phi} \otimes \mathbf{I}_p)\mathbf{W}_2, \\
\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_{\delta}} &= 2\mathbf{N}_p(\boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\Delta} \otimes \boldsymbol{\Lambda}\boldsymbol{\Gamma})\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta}} + (\boldsymbol{\Lambda}\boldsymbol{\Gamma} \otimes \boldsymbol{\Lambda}\boldsymbol{\Gamma})\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_{\delta}}, \\
\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_{\gamma}} &= 2\mathbf{N}_p(\boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\Delta} \otimes \boldsymbol{\Lambda}\boldsymbol{\Gamma})\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\gamma}},
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi} &= \mathbf{L}_{21,p}\mathbf{D}^{(1)}_{\boldsymbol{\psi};\boldsymbol{\theta}'_\psi}, \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\lambda,\boldsymbol{\theta}'_\lambda} &= 2\mathbf{N}_p[\mathbf{I}_p \otimes \operatorname{vec}'(\boldsymbol{\Phi}) \otimes \mathbf{I}_p]\,[\mathbf{K}_{p,q}\mathbf{W}_2 \otimes \mathbf{W}_2]\,, \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\lambda} &= 4\mathbf{N}_p[\boldsymbol{\Lambda} \otimes \operatorname{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_p]\mathbf{N}_q(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}) \\
&\times \left\{\left[(\boldsymbol{\Delta} \otimes \mathbf{I}_q)\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta} + 1/2\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\right] \otimes \mathbf{W}_2\right\}, \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\lambda} &= 4\mathbf{N}_p[\boldsymbol{\Lambda} \otimes \operatorname{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_p]\left[\mathbf{N}_q(\boldsymbol{\Gamma}\boldsymbol{\Delta} \otimes \boldsymbol{\Gamma})\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} \otimes \mathbf{W}_2\right], \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\lambda} &= \mathbf{0}_{p^2\times\nu_1\nu_4}, \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} &= 2\mathbf{N}_p[\boldsymbol{\Lambda}\boldsymbol{\Gamma} \otimes \operatorname{vec}'(\mathbf{I}_q) \otimes \boldsymbol{\Lambda}\boldsymbol{\Gamma}](\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta})2\mathbf{N}_{\nu_2} \\
&+ 2\mathbf{N}_p(\boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\Delta} \otimes \boldsymbol{\Lambda}\boldsymbol{\Gamma})\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} \\
&- 2\mathbf{N}_p[\boldsymbol{\Lambda}\boldsymbol{\Gamma} \otimes \operatorname{vec}'(\boldsymbol{\Delta}) \otimes \boldsymbol{\Lambda}\boldsymbol{\Gamma}]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\right) \\
&+ (\boldsymbol{\Lambda}\boldsymbol{\Gamma} \otimes \boldsymbol{\Lambda}\boldsymbol{\Gamma})\mathbf{L}_{21,q}\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta}, \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\delta} &= 2\mathbf{N}_p(\boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\Delta} \otimes \boldsymbol{\Lambda}\boldsymbol{\Gamma})\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\delta} \\
&- 2\mathbf{N}_p[\boldsymbol{\Lambda}\boldsymbol{\Gamma} \otimes \operatorname{vec}'(\boldsymbol{\Delta}) \otimes \boldsymbol{\Lambda}\boldsymbol{\Gamma}]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\right) \\
&- 2\mathbf{N}_p[\boldsymbol{\Lambda}\boldsymbol{\Gamma} \otimes \operatorname{vec}'(\mathbf{I}_q) \otimes \boldsymbol{\Lambda}\boldsymbol{\Gamma}]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} \otimes \mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\right), \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\delta} &= \mathbf{0}_{p^2\times\nu_2\nu_4}, \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} &= 2\mathbf{N}_p(\boldsymbol{\Lambda}\boldsymbol{\Gamma}\boldsymbol{\Delta} \otimes \boldsymbol{\Lambda}\boldsymbol{\Gamma})\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} \\
&- 2\mathbf{N}_p[\boldsymbol{\Lambda}\boldsymbol{\Gamma} \otimes \operatorname{vec}'(\boldsymbol{\Delta}) \otimes \boldsymbol{\Lambda}\boldsymbol{\Gamma}]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\right), \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\gamma} &= \mathbf{0}_{p^2\times\nu_3\nu_4}, \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi} &= \mathbf{L}_{21,p}\mathbf{D}^{(2)}_{\boldsymbol{\psi};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi}, \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\lambda,\boldsymbol{\theta}'_\lambda,\boldsymbol{\theta}'_\lambda} &= \mathbf{0}_{p^2\times\nu_1^3}, \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\lambda,\boldsymbol{\theta}'_\lambda} &= 2\mathbf{N}_p[\mathbf{I}_p \otimes \operatorname{vec}'(\mathbf{I}_q) \otimes \operatorname{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_p](\mathbf{K}_{pq,q^2} \otimes \mathbf{I}_{pq}) \\
&\times \left\{\left[2\mathbf{N}_q(\boldsymbol{\Gamma}\boldsymbol{\Delta} \otimes \boldsymbol{\Gamma})\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta} + (\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\right] \otimes \mathbf{I}_{p^2q^2}\right\} \\
&\times (\mathbf{I}_{\nu_2} \otimes \mathbf{K}_{p,q}\mathbf{W}_2 \otimes \mathbf{W}_2)\,, \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\lambda,\boldsymbol{\theta}'_\lambda} &= 4\mathbf{N}_p[\mathbf{I}_p \otimes \operatorname{vec}'(\mathbf{I}_q) \otimes \operatorname{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_p](\mathbf{K}_{pq,q^2} \otimes \mathbf{I}_{pq})
\end{aligned}
$$

$$
\begin{aligned}
& \times \quad \left[\mathbf{N}_q(\boldsymbol{\Gamma\Delta}\otimes\boldsymbol{\Gamma})\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\otimes\mathbf{I}_{p^2q^2}\right](\mathbf{I}_{\nu_3}\otimes\mathbf{K}_{p,q}\mathbf{W}_2\otimes\mathbf{W}_2)\,,\\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\lambda,\boldsymbol{\theta}'_\lambda} &= \mathbf{0}_{p^2\times\nu_1^2\nu_4},\\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\lambda} &= 4\mathbf{N}_p[\boldsymbol{\Lambda}\otimes\mathrm{vec}'(\mathbf{I}_q)\otimes\mathbf{I}_p]\,[\mathbf{N}_q(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\mathbf{A}\otimes\mathbf{W}_2]\,,\ \textit{where}\\
\mathbf{A} &= -[\mathbf{I}_q\otimes\mathrm{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\right)\\
&+ [\mathbf{I}_q\otimes\mathrm{vec}'(\mathbf{I}_q)\otimes\mathbf{I}_q]\left(\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\right)\\
&+ (\boldsymbol{\Delta}\otimes\mathbf{I}_q)\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta}\\
&- \mathbf{N}_q[\mathbf{I}_q\otimes\mathrm{vec}'(\mathbf{I}_q)\otimes\mathbf{I}_q]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\otimes\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\right)+1/2\mathbf{L}_{21,q}\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta},\\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\lambda} &= 4\,(\mathbf{K}_{\nu_3,\nu_2}\otimes\mathbf{I}_{\nu_1p^2})\,\mathbf{N}_p[\boldsymbol{\Lambda}\otimes\mathrm{vec}'(\mathbf{I}_q)\otimes\mathbf{I}_p]\,[\mathbf{N}_q(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\mathbf{B}\otimes\mathbf{W}_2]\,,\ \textit{where}\\
\mathbf{B} &= (\boldsymbol{\Delta}\otimes\mathbf{I}_q)\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\gamma}-[\mathbf{I}_q\otimes\mathrm{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\right)\\
&+ [\mathbf{I}_q\otimes\mathrm{vec}'(\mathbf{I}_q)\otimes\mathbf{I}_q]\left(\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\right),\\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\lambda} &= \mathbf{0}_{\nu_1\nu_2\nu_4},\\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\lambda} &= 4\mathbf{N}_p[\boldsymbol{\Lambda}\otimes\mathrm{vec}'(\mathbf{I}_q)\otimes\mathbf{I}_p]\left\{\left[\mathbf{N}_q(\boldsymbol{\Gamma\Delta}\otimes\boldsymbol{\Gamma})\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma}\right]\otimes\mathbf{W}_2\right\}\\
&- 4\mathbf{N}_p[\boldsymbol{\Lambda}\otimes\mathrm{vec}'(\mathbf{I}_q)\otimes\mathbf{I}_p]\\
&\times \left\{\left[\mathbf{N}_q[\boldsymbol{\Gamma}\otimes\mathrm{vec}'(\boldsymbol{\Delta})\otimes\boldsymbol{\Gamma}]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\right)\right]\otimes\mathbf{W}_2\right\},\\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\lambda} &= \mathbf{0}_{\nu_1\nu_3\nu_4},\\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\lambda} &= \mathbf{0}_{\nu_1\nu_4^2},\\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} &= 2\mathbf{N}_p(\boldsymbol{\Lambda\Gamma}\otimes\boldsymbol{\Lambda\Gamma})\mathbf{C},\ \textit{where}\\
\mathbf{C} &= -[\mathbf{I}_q\otimes\mathrm{vec}'(\mathbf{I}_q)\otimes\mathrm{vec}'(\mathbf{I}_q)\otimes\mathbf{I}_q]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\otimes\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\right)\\
&\times [(\mathbf{K}_{\nu_2,\nu_2}\otimes\mathbf{I}_{\nu_2})+(\mathbf{I}_{\nu_2}\otimes 2\mathbf{N}_{\nu_2})]\\
&+ [\mathbf{I}_q\otimes\mathrm{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\right)\mathbf{J}_{21,\nu_2}\\
&+ [\mathbf{I}_q\otimes\mathrm{vec}'(\mathbf{I}_q)\otimes\mathbf{I}_q]\\
&\times \left[\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta}\otimes\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\right)+\left(\mathbf{L}_{21,q}\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\right)\right]\mathbf{J}_{21,\nu_2}\\
&+ (\boldsymbol{\Delta}\otimes\mathbf{I}_q)\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta}+1/2\mathbf{L}_{21,q}\mathbf{D}^{(3)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta},
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\gamma} &= 2\mathbf{N}_p(\boldsymbol{\Lambda\Gamma}\otimes\boldsymbol{\Lambda\Gamma})\mathbf{E},\ \textit{where}\\
\mathbf{E} &= [\mathbf{I}_q\otimes \operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\gamma}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\right)(\mathbf{K}_{\nu_2\nu_3,\nu_2}+\mathbf{I}_{\nu_2}\otimes\mathbf{K}_{\nu_3,\nu_2})\\
&+ [\mathbf{I}_q\otimes \operatorname{vec}'(\mathbf{I}_q)\otimes\mathbf{I}_q]\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\gamma}\otimes\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\right)\\
&\times (\mathbf{K}_{\nu_2\nu_3,\nu_2}+\mathbf{I}_{\nu_2}\otimes\mathbf{K}_{\nu_3,\nu_2})\\
&+ (\boldsymbol{\Delta}\otimes\mathbf{I}_q)\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\gamma} - [\mathbf{I}_q\otimes \operatorname{vec}'(\mathbf{I}_q)\otimes \operatorname{vec}'(\mathbf{I}_q)\otimes\mathbf{I}_q]\\
&\times \left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\otimes\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\right)\left[\mathbf{I}_{\nu_2^2\nu_3}+(\mathbf{K}_{\nu_2,\nu_2}\otimes\mathbf{I}_{\nu_3})\right]\\
&+ [\mathbf{I}_q\otimes \operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\right)\\
&+ [\mathbf{I}_q\otimes \operatorname{vec}'(\mathbf{I}_q)\otimes\mathbf{I}_q](\mathbf{L}_{21,q}\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}),\\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} &= \mathbf{0}_{\nu_2^2\nu_4},\\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} &= 2\mathbf{N}_p(\boldsymbol{\Lambda\Gamma}\otimes\boldsymbol{\Lambda\Gamma})\mathbf{F},\ \textit{where}\\
\mathbf{F} &= [\mathbf{I}_q\otimes \operatorname{vec}'(\mathbf{I}_q)\otimes\mathbf{I}_q](\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\otimes\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma})\\
&- [\mathbf{I}_q\otimes \operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q](\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\otimes\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma})+(\boldsymbol{\Delta}\otimes\mathbf{I}_q)\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma}\\
&- [\mathbf{I}_q\otimes \operatorname{vec}'(\mathbf{I}_q)\otimes \operatorname{vec}'(\mathbf{I}_q)\otimes\mathbf{I}_q]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\otimes\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\right)\\
&\times (\mathbf{K}_{\nu_3,\nu_2}\otimes\mathbf{I}_{\nu_3})\\
&+ [\mathbf{I}_q\otimes \operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q](\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\gamma}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma})\\
&\times \left[\mathbf{I}_{\nu_2\nu_3^2}+(\mathbf{I}_{\nu_2}\otimes\mathbf{K}_{\nu_3,\nu_3})\right],\\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\delta} &= \mathbf{0}_{\nu_2\nu_3\nu_4},\\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\delta} &= \mathbf{0}_{\nu_2\nu_4^2},\\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} &= 2\mathbf{N}_p(\boldsymbol{\Lambda\Gamma}\otimes\boldsymbol{\Lambda\Gamma})\mathbf{D},\ \textit{where}\\
\mathbf{D} &= [\mathbf{I}_q\otimes \operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\right)\mathbf{J}_{21,\nu_3}\\
&+ (\boldsymbol{\Delta}\otimes\mathbf{I}_q)\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma},\\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} &= \mathbf{0}_{\nu_3^2\nu_4},\\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\gamma} &= \mathbf{0}_{\nu_3\nu_4^2},
\end{aligned}
$$

$$\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi} = \mathbf{L}_{21,p}\mathbf{D}^{(3)}_{\boldsymbol{\psi};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi},$$

where the derivatives of $\boldsymbol{\delta}$, $\mathbf{g}$ and $\boldsymbol{\psi}$ with respective to $\boldsymbol{\theta}_{\boldsymbol{\delta}}$, $\boldsymbol{\theta}_\gamma$, $\boldsymbol{\theta}_\psi$ respectively were given in Theorem 2, Theorem 10, Theorem 9 and Theorem 11 and $\mathbf{0}_{a\times b}$ represents a matrix of zeros with dimension $a \times b$.

Derivatives of $\boldsymbol{\sigma}$ with respect to $\boldsymbol{\theta}$ can be re-assembled from derivatives with respect to $\{\boldsymbol{\theta}_j\}_{j=1}^4$ by using the elementary matrix, $\mathbf{E}_{i,\boldsymbol{\nu}}$, which is defined in Table 56.

The first derivative of $\boldsymbol{\sigma}$ with respect to $\boldsymbol{\theta}$ is

$$\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \stackrel{\text{def}}{=} \frac{\partial \text{vec}\boldsymbol{\Sigma}}{\partial\,\boldsymbol{\theta}'} = \begin{pmatrix} \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\lambda} & \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\delta} & \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\gamma} & \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi} \end{pmatrix} = \sum_{i=1}^{4} \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_i}\mathbf{E}'_{i,\boldsymbol{\nu}}. \tag{48}$$

The second derivative of $\boldsymbol{\sigma}$ with respect to $\boldsymbol{\theta}$ is

$$\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'} \stackrel{\text{def}}{=} \frac{\partial^2 \text{vec}\boldsymbol{\Sigma}}{\partial\,\boldsymbol{\theta}' \otimes \partial\,\boldsymbol{\theta}'} = \sum_{i=1}^{4}\sum_{j=1}^{4} \mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_i,\boldsymbol{\theta}'_j} \left(\mathbf{E}_{i,\boldsymbol{\nu}} \otimes \mathbf{E}_{j,\boldsymbol{\nu}}\right)'. \tag{49}$$

The third derivative of $\boldsymbol{\sigma}$ with respect to $\boldsymbol{\theta}$ is

$$\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}'} \stackrel{\text{def}}{=} \frac{\partial^3 \text{vec}\boldsymbol{\Sigma}}{\partial\,\boldsymbol{\theta}' \otimes \partial\,\boldsymbol{\theta}' \otimes \partial\,\boldsymbol{\theta}'} = \sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{4} \mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_i,\boldsymbol{\theta}'_j,\boldsymbol{\theta}'_k} \left(\mathbf{E}_{i,\boldsymbol{\nu}} \otimes \mathbf{E}_{j,\boldsymbol{\nu}} \otimes \mathbf{E}_{k,\boldsymbol{\nu}}\right)'. \tag{50}$$

2.4. Alternative Parameterization of CFA Model

In earlier sections, $\boldsymbol{\Sigma}$ is written as $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$, where $\boldsymbol{\Phi}$ is a $q \times q$ factor correlation matrix. An alternative parameterization of $\boldsymbol{\Sigma}$ is introduced next in order for use in Chapter 3. Define $\boldsymbol{\Phi}_*$ as a $q \times q$ factor covariance matrix. Define $\boldsymbol{\Lambda}_*$ as a $p \times q$ factor loading matrix in this alternative parameterization of $\boldsymbol{\Sigma}$. That is, $\boldsymbol{\Sigma}$ can be rewritten as

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}_*\boldsymbol{\Phi}_*\boldsymbol{\Lambda}'_* + \boldsymbol{\Psi}. \tag{51}$$

2.4.1. Parameterization of Factor Loadings $\mathbf{\Lambda}_*$

Recall that the structure of $\mathbf{\Lambda}$ in (23) can be divided into known and unknown parts. That is,

$$\operatorname{vec}(\mathbf{\Lambda}) = \mathbf{W}_1\mathbf{L}_1 + \mathbf{W}_2\boldsymbol{\theta}_{\boldsymbol{\lambda}}.$$

Similarly, the structure for $\operatorname{vec}\mathbf{\Lambda}_*$ can be expressed as

$$\operatorname{vec}\mathbf{\Lambda}_* = \mathbf{W}_{1*}\mathbf{L}_{1*} + \mathbf{W}_{2*}\boldsymbol{\theta}_{\boldsymbol{\lambda}*}, \tag{52}$$

where $\mathbf{W}_{1*}$ and $\mathbf{W}_{2*}$ are known design matrices of full column-rank, $\mathbf{L}_{1*}$ is a known vector and $\boldsymbol{\theta}_{\boldsymbol{\lambda}*}$ is an unknown vector with $\dim(\boldsymbol{\theta}_{\boldsymbol{\lambda}}) = \nu_{*1}$. Typically, the components of $\mathbf{W}_{1*}$ and $\mathbf{W}_{2*}$ are 0's and 1's. Without loss of generality, the matrices $\mathbf{W}_{1*}$ and $\mathbf{W}_{2*}$ can be assumed to satisfy $\mathbf{W}_{2*}'\mathbf{W}_{1*} = \mathbf{0}$ and $\mathbf{W}_{2*}'\mathbf{W}_{2*} = \mathbf{I}$. A proof similar to the proof of $\mathbf{W}_{2*}'\mathbf{W}_{1*} = \mathbf{0}$ has been shown in § 2.3.1.

The structures of $\mathbf{W}_{1*}$ and $\mathbf{W}_{2*}$ in (52) are different from the structures of $\mathbf{W}_1$ and $\mathbf{W}_2$ in (23). Specifically, any entries in $\mathbf{\Lambda}$ of (23) corresponding to diagonal values of $\mathbf{\Phi}$ are considered unknown because diagonal entries of $\mathbf{\Phi}$ are fixed at 1. However, any entries in $\mathbf{\Lambda}_*$ of (52) corresponding to diagonal values of $\mathbf{\Phi}_*$ can be considered known or unknown because diagonal entries of $\mathbf{\Phi}_*$ are not fixed. This is illustrated by the following example.

Example. Consider the following structures for $\mathbf{\Lambda}_*$, $\mathbf{\Lambda}$, $\mathbf{\Phi}_*$ and $\mathbf{\Phi}$ that satisfy $\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' = \mathbf{\Lambda}_*\mathbf{\Phi}_*\mathbf{\Lambda}_*'$:

$$\mathbf{\Lambda}_* = \begin{pmatrix} 1 & 0 \\ \boldsymbol{\theta}_{\boldsymbol{\lambda}*1} & 0 \\ \boldsymbol{\theta}_{\boldsymbol{\lambda}*2} & 0 \\ 0 & 1 \\ 0 & \boldsymbol{\theta}_{\boldsymbol{\lambda}*3} \\ 0 & \boldsymbol{\theta}_{\boldsymbol{\lambda}*4} \end{pmatrix}, \quad \mathbf{\Phi}_* = \begin{pmatrix} \boldsymbol{\theta}_{\phi*1} & \boldsymbol{\theta}_{\phi*3} \\ \boldsymbol{\theta}_{\phi*3} & \boldsymbol{\theta}_{\phi*2} \end{pmatrix}, \quad \mathbf{\Lambda} = \mathbf{\Lambda}_*[\operatorname{Diag}(\mathbf{\Phi}_*)^{\odot(0.5)}],$$

$$\text{and} \quad \mathbf{\Phi} = [\mathrm{Diag}(\mathbf{\Phi}_*)^{\odot(-0.5)}]\mathbf{\Phi}_*[\mathrm{Diag}(\mathbf{\Phi}_*)^{\odot(-0.5)}].$$

where $\boldsymbol{\theta}_{\boldsymbol{\lambda}*1}$, $\boldsymbol{\theta}_{\boldsymbol{\lambda}*2}$, $\boldsymbol{\theta}_{\boldsymbol{\lambda}*3}$, $\boldsymbol{\theta}_{\boldsymbol{\lambda}*4}$, $\boldsymbol{\theta}_{\boldsymbol{\phi}*1}$, $\boldsymbol{\theta}_{\boldsymbol{\phi}*2}$ and $\boldsymbol{\theta}_{\boldsymbol{\phi}*3}$ are unknown parameters, $\mathbf{\Phi}_*$ is a covariance matrix, $\mathbf{\Phi}$ is a correlation matrix, $\odot$ and $\mathrm{Diag}(\mathbf{\Phi}_*)$ are defined in Table 56.

In $\mathbf{\Lambda}_*$, 1's and 0's are considered known and the other entries are considered unknown. Accordingly, $\mathbf{W}_{1*}$, $\mathbf{L}_{1*}$, and $\mathbf{W}_{2*}$ in $\mathrm{vec}\,\mathbf{\Lambda}_* = \mathbf{W}_{1*}\mathbf{L}_{1*} + \mathbf{W}_{2*}\boldsymbol{\theta}_{\boldsymbol{\lambda}*}$ can be written as

$$\mathbf{W}_{1*} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}, \mathbf{L}_{1*} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{W}_{2*} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \text{ and } \boldsymbol{\theta}_{\boldsymbol{\lambda}*} = \begin{pmatrix} \boldsymbol{\theta}_{\boldsymbol{\lambda}*1} \\ \boldsymbol{\theta}_{\boldsymbol{\lambda}*2} \\ \boldsymbol{\theta}_{\boldsymbol{\lambda}*3} \\ \boldsymbol{\theta}_{\boldsymbol{\lambda}*4} \end{pmatrix}. \quad (53)$$

The structure of $\mathbf{\Lambda}$ can be written as follows:

$$\mathbf{\Lambda} = \begin{pmatrix} \boldsymbol{\theta}_{\boldsymbol{\lambda}1} & 0 \\ \boldsymbol{\theta}_{\boldsymbol{\lambda}2} & 0 \\ \boldsymbol{\theta}_{\boldsymbol{\lambda}3} & 0 \\ 0 & \boldsymbol{\theta}_{\boldsymbol{\lambda}4} \\ 0 & \boldsymbol{\theta}_{\boldsymbol{\lambda}5} \\ 0 & \boldsymbol{\theta}_{\boldsymbol{\lambda}6} \end{pmatrix},$$

where $\boldsymbol{\theta}_{\boldsymbol{\lambda}1}$, $\boldsymbol{\theta}_{\boldsymbol{\lambda}2}$, $\boldsymbol{\theta}_{\boldsymbol{\lambda}3}$, $\boldsymbol{\theta}_{\boldsymbol{\lambda}4}$, $\boldsymbol{\theta}_{\boldsymbol{\lambda}5}$ and $\boldsymbol{\theta}_{\boldsymbol{\lambda}6}$ are unknown because $\mathbf{\Lambda} = \mathbf{\Lambda}_*[\mathrm{Diag}(\mathbf{\Phi}_*)^{\odot(0.5)}]$ and $\boldsymbol{\theta}_{\boldsymbol{\lambda}*1}$, $\boldsymbol{\theta}_{\boldsymbol{\lambda}*2}$, $\boldsymbol{\theta}_{\boldsymbol{\lambda}*3}$, $\boldsymbol{\theta}_{\boldsymbol{\lambda}*4}$, $\boldsymbol{\theta}_{\boldsymbol{\phi}*1}$ and $\boldsymbol{\theta}_{\boldsymbol{\phi}*2}$ are unknown parameters.

Accordingly, $\mathbf{W}_1$, $\mathbf{L}_1$, and $\mathbf{W}_2$ in $\mathrm{vec}\,\mathbf{\Lambda} = \mathbf{W}_1\mathbf{L}_1 + \mathbf{W}_2\boldsymbol{\theta}_{\boldsymbol{\lambda}}$ can be written as

$$\mathbf{W}_1 \text{ is empty, } \mathbf{L}_1 \text{ is empty, } \mathbf{W}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \text{ and } \boldsymbol{\theta}_{\boldsymbol{\lambda}} = \begin{pmatrix} \boldsymbol{\theta}_{\boldsymbol{\lambda}1} \\ \boldsymbol{\theta}_{\boldsymbol{\lambda}2} \\ \boldsymbol{\theta}_{\boldsymbol{\lambda}3} \\ \boldsymbol{\theta}_{\boldsymbol{\lambda}4} \\ \boldsymbol{\theta}_{\boldsymbol{\lambda}5} \\ \boldsymbol{\theta}_{\boldsymbol{\lambda}6} \end{pmatrix}. \quad (54)$$

There are more parameters in $\boldsymbol{\theta}_{\boldsymbol{\lambda}}$ than in $\boldsymbol{\theta}_{\boldsymbol{\lambda}*}$ because $\mathbf{\Lambda}$ needs to satisfy $\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' = \mathbf{\Lambda}_*\mathbf{\Phi}_*\mathbf{\Lambda}_*'$ and diagonal entries of $\mathbf{\Phi}$ are fixed at 1.

For instance, if

$$\boldsymbol{\theta}_{\boldsymbol{\lambda}*} = \begin{pmatrix} 0.57 \\ 0.22 \\ 0.90 \\ 0.98 \end{pmatrix}, \boldsymbol{\Phi}_* = \begin{pmatrix} 0.72 & 0.35 \\ 0.35 & 2.32 \end{pmatrix}, \text{then } \boldsymbol{\Phi} = \begin{pmatrix} 1.0000 & 0.2708 \\ 0.2708 & 1.0000 \end{pmatrix} \text{ and } \boldsymbol{\theta}_{\boldsymbol{\lambda}} = \begin{pmatrix} 0.8485 \\ 0.4837 \\ 0.1867 \\ 1.5232 \\ 1.3708 \\ 1.4927 \end{pmatrix}.$$

The parameterization of eigenvalues of $\boldsymbol{\Phi}_*$ is the same as the parameterization of eigenvalues of $\boldsymbol{\Phi}$ in § 2.3.2. Define $\boldsymbol{\Gamma}_*$ as a matrix of eigenvectors of $\boldsymbol{\Phi}_*$. One can write the $q \times q$ factor covariance matrix $\boldsymbol{\Phi}_*$ in diagonal form: $\boldsymbol{\Phi}_* = \boldsymbol{\Gamma}_* \boldsymbol{\Delta} \boldsymbol{\Gamma}'_*$. The parameterization of eigenvectors of $\boldsymbol{\Phi}_*$ is discussed in the following section.

2.4.2. Parameterization of Eigenvectors of $\boldsymbol{\Phi}_*$

The parameterization of eigenvectors of $\boldsymbol{\Phi}_*$ is similar to the parameterization of eigenvectors of $\boldsymbol{\Phi}$. The structure of the following discussion is adapted from Boik [35]. Boik's eigenvector parameterization [35] is based on the trivial equality $\boldsymbol{\Gamma}_* = \mathbf{I}_q \boldsymbol{\Gamma}_*$. Let $\boldsymbol{\Gamma}_{*0}$ be a $q \times q$ orthogonal matrix in a small neighborhood of $\boldsymbol{\Gamma}_*$. This implies that $\boldsymbol{\Gamma}_* \approx \boldsymbol{\Gamma}_{*0}$. The identity matrix $\mathbf{I}_q$ can be written as $\mathbf{I}_q = \boldsymbol{\Gamma}_{*0} \boldsymbol{\Gamma}'_{*0}$ and it follows that $\boldsymbol{\Gamma}_* = \mathbf{I}_q \boldsymbol{\Gamma}_* = \boldsymbol{\Gamma}_{*0} \boldsymbol{\Gamma}'_{*0} \boldsymbol{\Gamma}_*$. Let $\mathbf{G}_* = \boldsymbol{\Gamma}'_{*0} \boldsymbol{\Gamma}_*$ and note that $\mathbf{G}_* \in \mathcal{O}(q)$. It further follows that $\mathbf{G}_* \approx \mathbf{I}_q$ because $\boldsymbol{\Gamma}_* \approx \boldsymbol{\Gamma}_{*0}$. To develop a parameterization for $\boldsymbol{\Gamma}_*$, the matrix $\boldsymbol{\Gamma}_{*0}$ is treated as known whereas $\mathbf{G}_*$ is treated as unknown and is parameterized locally near $\mathbf{G}_* = \mathbf{I}_q$. The above approach employs the implicit function-based parameterization approach in Boik [18].

In Boik [35], $\text{vec}\,\mathbf{G}_*$ was written as

$$\mathbf{g}_* \stackrel{\text{def}}{=} \text{vec}\,\mathbf{G}_* = \mathbf{A}_1\boldsymbol{\eta}_{\boldsymbol{\gamma}*} + \mathbf{A}_2\boldsymbol{\theta}_{\boldsymbol{\gamma}*}, \text{ where } \dim(\boldsymbol{\theta}_{\boldsymbol{\gamma}*}) = \nu_3^*, \quad \dim(\mathbf{A}_2) = q^2 \times \nu_3^*,$$
$$\mathbf{A}_1 = \sum_{j=1}^{q}\sum_{i=j}^{q}(\mathbf{e}_j^q \otimes \mathbf{e}_i^q)\mathbf{e}_{f_{ij}}^{h\prime}, \quad h = \frac{q(q+1)}{2}, \quad f_{ij} = \frac{(2q-j)(j-1)}{2} + i, \tag{55}$$

and $\mathbf{A}_2$ is a known indicator matrix. These matrices satisfy $\mathbf{A}_i'\mathbf{A}_i = \mathbf{I}$ for $i = 1, 2$; $\mathbf{A}_1'\mathbf{g}_* = \boldsymbol{\eta}_{\boldsymbol{\gamma}*}$; $\mathbf{A}_2'\mathbf{g}_* = \boldsymbol{\theta}_{\boldsymbol{\gamma}*}$; and $\mathbf{A}_1'\mathbf{A}_2 = \mathbf{0}$. Recall that in § 2.3.3, $\mathbf{G}$ in (30) needs to satisfy not only $\mathbf{G} \in \mathcal{O}(q)$, but also (37):

$$\mathbf{e}_i^{q\prime}\boldsymbol{\Phi}\mathbf{e}_i^q = 1, \text{ for } i = 1, \ldots, q,$$

where $\boldsymbol{\Phi} = \boldsymbol{\Gamma}_0\mathbf{G}\boldsymbol{\Delta}\mathbf{G}'\boldsymbol{\Gamma}_0'$ in (31) is a correlation matrix. Unlike $\mathbf{G}$, $\mathbf{G}_*$ only needs to satisfy $\mathbf{G}_* \in \mathcal{O}(q)$. Furthermore, $\mathbf{g}_*$ is only a function of $\boldsymbol{\theta}_{\boldsymbol{\gamma}*}$, not a function of $\boldsymbol{\theta}_{\boldsymbol{\delta}}$. Accordingly, $\dim(\boldsymbol{\theta}_{\boldsymbol{\gamma}*}) \neq \dim(\boldsymbol{\theta}_{\boldsymbol{\gamma}})$ because $\dim(\boldsymbol{\theta}_{\boldsymbol{\gamma}*}) = \nu_3^* = \frac{1}{2}(q^2 - \mathbf{m}'\mathbf{m})$, given in (34), and $\dim(\boldsymbol{\theta}_{\boldsymbol{\gamma}}) = \nu_3 = \nu_3^* - (q-1)$.

Derivatives of $\boldsymbol{\sigma}$ with respect to $\boldsymbol{\theta}_{\boldsymbol{\gamma}*}$ are needed when estimating parameters and conducting statistical inferences. Accordingly, Theorem 16 is needed.

Theorem 16. [Boik [35], 2010, Theorem 4, Page 13]. *The first three derivatives of* $\mathbf{g}_*$ *in* (55) *with respect to* $\boldsymbol{\theta}_{\boldsymbol{\gamma}*}$*, evaluated at* $\boldsymbol{\theta}_{\boldsymbol{\gamma}*} = \mathbf{0}$*, can be written as follows:*

$$\begin{aligned}
\mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}} &= 2\mathbf{N}_q^{\perp}\mathbf{A}_2, \\
\mathbf{D}^{(2)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\gamma*}} &= \mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}} \otimes \mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}}\right), \ \textit{and} \\
\mathbf{D}^{(3)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\gamma*}} &= -\mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\gamma*}} \otimes \mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}}\right)\mathbf{J}_{21,\nu_3^*}
\end{aligned}$$

where $\mathbf{I}_{q,3}$*,* $\mathbf{D}_q$ *and* $\mathbf{J}_{21,\nu_3}$ *are defined in Table 56.*

For notational convenience, define $\boldsymbol{\theta}_*$ as

$$\boldsymbol{\theta}_* = (\boldsymbol{\theta}'_{\boldsymbol{\lambda}*} \ \ \boldsymbol{\theta}'_{\delta} \ \ \boldsymbol{\theta}'_{\gamma*} \ \ \boldsymbol{\theta}'_{\psi})' = (\boldsymbol{\theta}'_{\lambda*} \ \ \boldsymbol{\theta}'_{\delta*} \ \ \boldsymbol{\theta}'_{\gamma*} \ \ \boldsymbol{\theta}'_{\psi*})', \tag{56}$$

where $\boldsymbol{\theta}_{\delta*} = \boldsymbol{\theta}_{\delta}$ and $\boldsymbol{\theta}_{\psi*} = \boldsymbol{\theta}_{\psi}$. Rewrite $\boldsymbol{\theta}_*$ as $\boldsymbol{\theta}_* = (\boldsymbol{\theta}'_{*1} \quad \boldsymbol{\theta}'_{*2} \quad \boldsymbol{\theta}'_{*3} \quad \boldsymbol{\theta}'_{*4})' = (\boldsymbol{\theta}'_{\lambda*} \quad \boldsymbol{\theta}'_{\delta*} \quad \boldsymbol{\theta}'_{\gamma*} \quad \boldsymbol{\theta}'_{\psi*})'$, where $\boldsymbol{\theta}_{*i}$ has dimension $\nu_{*i} \times 1$. Define ν_* as

$$\nu_* = \nu_{*1} + \nu_{*2} + \nu_{*3} + \nu_{*4}. \tag{57}$$

The first derivative of $\boldsymbol{\sigma}$ with respect to $\boldsymbol{\theta}$ is

$$\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_*} \stackrel{\text{def}}{=} \frac{\partial \text{vec}\boldsymbol{\Sigma}}{\partial\, \boldsymbol{\theta}'_*} = \begin{pmatrix} \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_{\lambda*}} & \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_{\delta*}} & \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_{\gamma*}} & \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_{\psi*}} \end{pmatrix} = \sum_{i=1}^{4} \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_{*i}} \mathbf{E}'_{i,\boldsymbol{\nu}_*}. \tag{58}$$

The second derivative of $\boldsymbol{\sigma}$ with respect to $\boldsymbol{\theta}$ is

$$\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_*,\boldsymbol{\theta}'_*} \stackrel{\text{def}}{=} \frac{\partial^2 \text{vec}\boldsymbol{\Sigma}}{\partial\, \boldsymbol{\theta}'_* \otimes \partial\, \boldsymbol{\theta}'_*} = \sum_{i=1}^{4}\sum_{j=1}^{4} \mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_{*i},\boldsymbol{\theta}'_{*j}} \left(\mathbf{E}_{i,\boldsymbol{\nu}_*} \otimes \mathbf{E}_{j,\boldsymbol{\nu}_*}\right)'. \tag{59}$$

The third derivative of $\boldsymbol{\sigma}$ with respect to $\boldsymbol{\theta}$ is

$$\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_*,\boldsymbol{\theta}'_*,\boldsymbol{\theta}'_*} \stackrel{\text{def}}{=} \frac{\partial^3 \text{vec}\boldsymbol{\Sigma}}{\partial\, \boldsymbol{\theta}'_* \otimes \partial\, \boldsymbol{\theta}'_* \otimes \partial\, \boldsymbol{\theta}'_*} = \sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{4} \mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_{*i},\boldsymbol{\theta}'_{*j},\boldsymbol{\theta}'_{*k}} \left(\mathbf{E}_{i,\boldsymbol{\nu}_*} \otimes \mathbf{E}_{j,\boldsymbol{\nu}_*} \otimes \mathbf{E}_{k,\boldsymbol{\nu}_*}\right)'. \tag{60}$$

The expressions for $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_{*i}}$, $\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_{*i},\boldsymbol{\theta}'_{*j}}$ and $\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_{*i},\boldsymbol{\theta}'_{*j},\boldsymbol{\theta}'_{*k}}$ are the same with $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_i}$, $\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_i,\boldsymbol{\theta}'_j}$ and $\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_i,\boldsymbol{\theta}'_j,\boldsymbol{\theta}'_k}$ in Theorem 15, except that $\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}$, $\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma}$ and $\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma}$ in Theorem 15 are replaced by $\mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}}$, $\mathbf{D}^{(2)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\gamma*}}$ and $\mathbf{D}^{(3)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\gamma*}}$ in Theorem 16. Note that all the derivatives of $\mathbf{g}_*$ with respect to $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ are $\mathbf{0}$ because $\mathbf{g}_*$ is not a function of $\boldsymbol{\theta}_{\boldsymbol{\delta}}$.

CHAPTER 3

PARAMETER ESTIMATION

3.1. Parameter Estimation

In this chapter, the method of Lagrange multipliers is applied to obtain a constrained minimizer of a loss function $F(\mathbf{\Sigma}, \mathbf{S})$. Boik [33] gave details on the Lagrange multipliers approach. The loss function used in this chapter is the discrepancy function (b) in (7) given in §1.3.2. Derivations of this loss function with respect to $\boldsymbol{\theta} = (\boldsymbol{\theta}'_\lambda \quad \boldsymbol{\theta}'_\delta \quad \boldsymbol{\theta}'_\gamma \quad \boldsymbol{\theta}'_\psi)'$ are obtained in this Chapter. The interest of this chapter is to estimate $\mathbf{\Sigma}$, where $\mathbf{\Sigma} = \mathbf{\Lambda\Gamma\Delta\Gamma'\Lambda'} + \mathbf{\Psi}$ was given in (21). Specifically, the goal of this chapter is to find a minimizer of the loss function, $\widehat{\mathbf{\Sigma}}$, under the constraints that estimators, $\widehat{\mathbf{\Lambda}}$, $\widehat{\mathbf{\Delta}}$, $\widehat{\mathbf{\Gamma}}$ and $\widehat{\mathbf{\Psi}}$, are proper and eigenvalues of $\widehat{\mathbf{\Delta}}$, namely $\widehat{\boldsymbol{\delta}}_i$ for $i = 1, \ldots, q$, are nonnegative.

3.2. Loss Function

Generate a random sample $\mathbf{Y}$ of size $\mathbf{N}$

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}'_1 \\ \mathbf{y}'_2 \\ \vdots \\ \mathbf{y}'_N \end{pmatrix}, \tag{61}$$

where $\mathbf{y}_i$ is a $p \times 1$ random vector of observed responses for a single subject for $i = 1, 2, \ldots, N$.

Recall that the factor analysis model (3) in § 1.2 for $\mathbf{y}_i$ is

$$\mathbf{y}_i = \boldsymbol{\mu}_i + \mathbf{\Lambda}\mathbf{f}_i + \boldsymbol{\epsilon}_i,$$

where

$$\mathrm{E}(\mathbf{f}_i) = \mathbf{0}, \quad \mathrm{Var}(\mathbf{f}_i) = \mathbf{\Phi}, \quad \mathrm{E}(\boldsymbol{\epsilon}_i) = \mathbf{0}, \quad \mathrm{Var}(\boldsymbol{\epsilon}_i) = \mathbf{\Psi},$$

$$\mathbf{\Psi} = \mathrm{diag}\,\boldsymbol{\psi} = \begin{pmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{pmatrix}, \text{ and } \mathrm{Cov}(\mathbf{f}_i, \boldsymbol{\epsilon}_i) = \mathbf{0}.$$

Note that the factor loading matrix $\mathbf{\Lambda}$ does not vary across all N observations.

The factor analytic model for $\mathbf{Y}$ is

$$\mathbf{Y} = \mathbf{XB} + \mathbf{F\Lambda}' + \mathbf{E}, \tag{62}$$

where $\mathbf{X}$ is an $\mathbf{N} \times d$ known matrix of constants, $\mathrm{rank}(\mathbf{X}) = r_x$, $\mathbf{B}$ is a $d \times p$ matrix of fixed regression coefficients, $\mathbf{F}$ is an $\mathbf{N} \times q$ matrix of random latent factor scores whose i^{th} row is $\mathbf{f}_i'$, $\mathbf{\Lambda}$ is a $p \times q$ matrix of fixed factor loadings, and $\mathbf{E}$ is an $\mathbf{N} \times p$ matrix of random errors whose i^{th} row is $\boldsymbol{\epsilon}_i'$. Assume that $\mathbf{y}_i$ in $\mathbf{Y}$, for $i = 1, 2, \ldots, N$, are independently distributed with common positive definite covariance matrix $\mathbf{\Sigma}$, where $\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Psi}$. Also, assumed that $r_x = O(1)$. These assumptions hold for the remainder of the thesis, unless otherwise specified.

Further, $\mathrm{E}[\mathbf{F}] = \mathbf{0}_{N\times q}$, $\mathrm{E}[\mathbf{E}] = \mathbf{0}_{N\times p}$, $\mathrm{E}[\mathbf{Y}] = \mathbf{XB}$, $\mathrm{Var}(\mathrm{vec}\,\mathbf{F}) = \mathbf{\Phi} \otimes \mathbf{I}_N$, $\mathrm{Var}(\mathrm{vec}\,\mathbf{E}) = \mathbf{\Psi} \otimes \mathbf{I}_N$, $\mathrm{Var}(\mathrm{vec}\,\mathbf{Y}) = \mathbf{\Sigma} \otimes \mathbf{I}_N$, $\mathrm{Cov}(\mathrm{vec}\,\mathbf{F}, \mathrm{vec}\,\mathbf{E}) = \mathbf{0}$, and

$$\boldsymbol{\mu}_i = \mathrm{E}[\mathbf{y}_i] = \mathrm{E}[\mathbf{Y}'\mathbf{e}_i^N] = (\mathbf{XB})'\mathbf{e}_i^N, \tag{63}$$

where $\mathbf{e}_i^N$ is the i^{th} column of $\mathbf{I}_N$ and is defined in Table 56.

Let $\mathbf{S}$ be the $p \times p$ sample covariance matrix. If $\mathbf{y}_i \overset{\mathrm{iid}}{\sim} \mathrm{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ for $i = 1, 2, \ldots, N$, then $\mathbf{A} \overset{\mathrm{def}}{=} (N - r_x)\mathbf{S} \sim \mathbf{W}_p(N - r_x, \mathbf{\Sigma})$. Accordingly, the probability density function of $\mathbf{A}$ is

$$f_{\mathbf{A}}(\mathbf{A}; \mathbf{\Sigma}) = \frac{|\mathbf{A}|^{\frac{n-p-1}{2}} e^{-\frac{1}{2}\mathrm{tr}(\mathbf{A\Sigma}^{-1})}}{|\mathbf{\Sigma}|^{\frac{n}{2}} 2^{\frac{np}{2}} \Gamma_p(\frac{n}{2})}, \tag{64}$$

where $n = N - r_x$ and $\Gamma_p\left(\frac{n}{2}\right) = \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^{p} \Gamma\left[\frac{n}{2} - \frac{(i-1)}{2}\right]$.

Accordingly, the log likelihood function can be written in terms of $\mathbf{S}$ as follows:

$$l(\boldsymbol{\theta}; \mathbf{S}) = -\frac{n}{2}\,\mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \frac{n}{2}\ln|\boldsymbol{\Sigma}| + Constants, \tag{65}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}'_\lambda \;\; \boldsymbol{\theta}'_\delta \;\; \boldsymbol{\theta}'_\gamma \;\; \boldsymbol{\theta}'_\psi)'$. The maximum likelihood (ML) discrepancy function,

$$F(\boldsymbol{\Sigma}, \mathbf{S}) = \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) + \ln|\boldsymbol{\Sigma}| - p - \ln|\mathbf{S}|, \tag{66}$$

is the discrepancy function (b) in (7) given in § 1.3.2. It is apparent that maximizing $l(\boldsymbol{\theta}; \mathbf{S})$ in (65) is equivalent to minimizing $F(\boldsymbol{\Sigma}, \mathbf{S})$ in (66).

Define $\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}$ as follows:

$$\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}} \;\; = \;\; \left.\frac{\partial F(\boldsymbol{\Sigma}, \mathbf{S})}{\partial\,\boldsymbol{\theta}}\right|_{\mathbf{G}=\mathbf{I}_q}.$$

Note that $\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}} = \mathbf{0}$ can and will be used as an estimating function whether or not $\mathbf{y}_i \overset{\text{iid}}{\sim} \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is satisfied. If $\mathbf{y}_i \overset{\text{iid}}{\sim} \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is satisfied, then the maximum likelihood estimator is a solution to $\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}} = \mathbf{0}$. In general, a solution to $\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}} = \mathbf{0}$ is an estimator, but not a maximum likelihood estimator. Lemma 1 and Lemma 2 provide useful properties for simplifying expressions.

Lemma 1. *Let* $\mathbf{A}$ *be a* $r \times s$ *matrix and* $\mathbf{B}$ *be a* $s \times r$ *matrix. Some useful results are the following.*

(a.) [Searle [36], 1982, Theorem 2, Page 333] $\mathrm{tr}(\mathbf{AB}) = \mathrm{vec}'(\mathbf{A}')\,\mathrm{vec}(\mathbf{B})$.

(b.) [Searle [36], 1982, equation 39, Page 337] $\partial \ln(|\mathbf{A}|)/\partial y = \mathrm{tr}\left[\mathbf{A}^{-1}\left(\partial\mathbf{A}/\partial y\right)\right]$ *for symmetric and non-symmetric* $\mathbf{A}$*, where* $\mathbf{A}$ *is a* $r \times r$ *matrix.*

(c.) [A Useful Result] $\partial\,\mathrm{vec}\boldsymbol{\Sigma}^{-1}/\partial\,\boldsymbol{\theta}' = -\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\partial\,\mathrm{vec}\boldsymbol{\Sigma}/\partial\,\boldsymbol{\theta}'$.

(d.) [A Useful Result] $\partial\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)/\partial\,\boldsymbol{\theta}' = -\left[\boldsymbol{\Sigma}^{-1} \otimes \mathrm{vec}'(\boldsymbol{\Sigma}^{-1}) \otimes \boldsymbol{\Sigma}^{-1}\right]$
$\times \left(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathbf{I}_{p^2}\right) - \mathbf{K}_{p,p}\left[\boldsymbol{\Sigma}^{-1} \otimes \mathrm{vec}'(\boldsymbol{\Sigma}^{-1}) \otimes \boldsymbol{\Sigma}^{-1}\right]\left(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathbf{K}_{p,p}\right)$.

(e.) [A Useful Result] $\partial\, \text{vec}\boldsymbol{\Sigma}^{-1}/\partial\, \boldsymbol{\theta} = -\left(\mathbf{I}_\nu \otimes \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \partial\, \text{vec}\boldsymbol{\Sigma}/\partial\, \boldsymbol{\theta}$.

(f.) [A Useful Result] $\mathbf{L}_{21,p}\mathbf{L}'_{21,p} = \mathbf{L}_{22,p}$ *and* $\mathbf{L}'_{21,p}\mathbf{L}_{21,p} = \mathbf{I}_p$, *where* $\mathbf{L}_{21,p}$ *is as Table 56.*

Lemma 2. [Brewer [37], 1978, Table I and II on Page 773 and Table IV on page 776 and Supplement of Boik [34], 2003, §6.5.2].

1. $\frac{\partial(\mathbf{X}\otimes\mathbf{Y})}{\partial\mathbf{Z}} = \left(\frac{\partial\mathbf{X}}{\partial\mathbf{Z}} \otimes \mathbf{Y}\right) + \left(\mathbf{I}_s \otimes \mathbf{K}_{m,q}\right)\left(\frac{\partial\mathbf{Y}}{\partial\mathbf{Z}} \otimes \mathbf{X}\right)\left(\mathbf{I}_t \otimes \mathbf{K}_{r,n}\right)$,
   where $\mathbf{X}$ *is* $m \times n$, $\mathbf{Y}$ *is* $q \times r$ *and* $\mathbf{Z}$ *is* $s \times t$;

2. $\frac{\partial(\mathbf{X}\mathbf{Y})}{\partial\mathbf{Z}} = \left(\frac{\partial\mathbf{X}}{\partial\mathbf{Z}}\right)\left(\mathbf{I}_t \otimes \mathbf{Y}\right) + \left(\mathbf{I}_s \otimes \mathbf{X}\right)\left(\frac{\partial\mathbf{Y}}{\partial\mathbf{Z}}\right)$,
   where $\mathbf{X}$ *is* $p \times n$, $\mathbf{Y}$ *is* $n \times q$ *and* $\mathbf{Z}$ *is* $s \times t$;

3. $\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\, \text{vec}\, \mathbf{B}$;

4. $(\mathbf{ABC} \otimes \mathbf{D})\, \mathbf{E} = \left[\mathbf{A} \otimes (\text{vec}\, \mathbf{C}')' \otimes \mathbf{D}\right](\text{vec}\, \mathbf{B}' \otimes \mathbf{E})$;

5. $(\mathbf{A} \otimes \mathbf{BCD})\, \mathbf{E} = \left[\mathbf{A} \otimes (\text{vec}\, \mathbf{D})' \otimes \mathbf{B}\right](\mathbf{E} \otimes \text{vec}\, \mathbf{C})$

6. $\mathbf{A} \otimes \mathbf{B} = \mathbf{K}_{r,p}(\mathbf{B} \otimes \mathbf{A})\mathbf{K}_{q,c}$, *where* $\mathbf{A}$ *is* $r \times c$ *and* $\mathbf{B}$ *is* $p \times q$;

7. $\mathbf{K}_{r,c}\mathbf{K}_{c,r} = \mathbf{I}_{rc}$;

8. $\mathbf{K}_{1,c} = \mathbf{K}_{c,1} = \mathbf{I}_c$;

First, second and third derivatives of $F(\boldsymbol{\Sigma}, \mathbf{S})$ in (66) with respect to $\boldsymbol{\theta}$ are are given in Theorem 17. These derivatives are needed for point estimation and statistical inferences.

Theorem 17. [Adapted from Boik [34], 2003, Page 689, & Supplement, Page 26].

(a.) *The first derivative of* $F(\boldsymbol{\Sigma},\mathbf{S})$ *in (66) with respect to* $\boldsymbol{\theta}$*, evaluated at* $\mathbf{G}=\mathbf{I}_q$*, is as follows:*

$$\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}} \stackrel{\text{def}}{=} \left.\frac{\partial F(\boldsymbol{\Sigma},\mathbf{S})}{\partial\,\boldsymbol{\theta}}\right|_{\mathbf{G}=\mathbf{I}_q} = -\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma}),$$

where $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}$ *is given in (48).*

(b.) *The second derivative of* $F(\boldsymbol{\Sigma},\mathbf{S})$ *in (66) with respect to* $\boldsymbol{\theta}$ *and* $\boldsymbol{\theta}'$*, evaluated at* $\mathbf{G}=\mathbf{I}_q$*, is as follows:*

$$\begin{aligned}\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'} &\stackrel{\text{def}}{=} \left.\frac{\partial^2 F(\boldsymbol{\Sigma},\mathbf{S})}{\partial\,\boldsymbol{\theta}\otimes\partial\,\boldsymbol{\theta}'}\right|_{\mathbf{G}=\mathbf{I}_q}\\ &= \mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}+2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right]\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\\ &- \mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}\left[\mathbf{I}_\nu\otimes\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right],\end{aligned}$$

where $\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}=\operatorname{dvec}\left(\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'},p^2\nu,\nu\right)$ *and* $\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'}$ *is given in (49).*

(c.) *The third derivative of* $F(\boldsymbol{\Sigma},\mathbf{S})$ *in (66) with respect to* $\boldsymbol{\theta}'$*,* $\boldsymbol{\theta}'$ *and* $\boldsymbol{\theta}$*, evaluated at* $\mathbf{G}=\mathbf{I}_q$*, is as follows:*

$$\begin{aligned}\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}} &\stackrel{\text{def}}{=} \left.\frac{\partial^3 F(\boldsymbol{\Sigma},\mathbf{S})}{\partial\,\boldsymbol{\theta}'\otimes\partial\,\boldsymbol{\theta}'\otimes\partial\,\boldsymbol{\theta}}\right|_{\mathbf{G}=\mathbf{I}_q}\\ &= -4\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]'\left(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right)\mathbf{N}_\nu\\ &+ 2\mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}\left\{\mathbf{I}_\nu\otimes\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]\right\}\mathbf{N}_\nu\\ &+ \left[\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]'\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'}\\ &- 2\left\{\operatorname{vec}\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]\otimes\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]\right\}'\\ &\times \left(\mathbf{I}_{p\nu}\otimes\mathbf{K}_{p,p}\otimes\mathbf{I}_p\right)\left[\mathbf{I}_\nu\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right]\\ &- 4\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]'\\ &\times \left\{\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\mathbf{N}_p\left[\mathbf{I}_p\otimes(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right]\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right\}\\ &- \mathbf{D}^{(3)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta},\boldsymbol{\theta}'}\left[\mathbf{I}_{\nu^2}\otimes\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right]\end{aligned}$$

$$
\begin{aligned}
&+ \quad 4\mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}\left(\mathbf{I}_\nu \otimes \boldsymbol{\Sigma}^{-1} \otimes \operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)' \\
&\times \quad \left[\left(\mathbf{I}_\nu \otimes \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right)\mathbf{N}_\nu \otimes \operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
&+ \quad 2\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]'\left[\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'} \otimes \operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right],
\end{aligned}
$$

where $\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'} = \operatorname{dvec}\left(\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'}, p^2\nu, \nu\right)$, $\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta},\boldsymbol{\theta}'} = \operatorname{dvec}\left(\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}'}, p^2\nu^2, \nu\right)$, $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}$, $\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'}$, *and* $\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}'}$ *are given in (50).*

Also, define $\mathbf{D}^{(3)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}',\boldsymbol{\theta}'}$ *as*

$$
\mathbf{D}^{(3)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}',\boldsymbol{\theta}'} \stackrel{\text{def}}{=} \left.\frac{\partial^3 F(\boldsymbol{\Sigma},\mathbf{S})}{\partial\,\boldsymbol{\theta} \otimes \partial\,\boldsymbol{\theta}' \otimes \partial\,\boldsymbol{\theta}'}\right|_{\mathbf{G}=\mathbf{I}_q},
$$

and $\mathbf{D}^{(3)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}',\boldsymbol{\theta}'} = \mathbf{D}^{(3)}_{F;\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}}$.

Four algorithms are discussed in the following sections. Each of these algorithms can be used to obtain an estimate of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ that minimizes $F(\boldsymbol{\Sigma},\mathbf{S})$ in (66). The first algorithm is called a modified Fisher Scoring algorithm based on a correlation parameterization of $\boldsymbol{\Phi}$. The second algorithm is called a modified Fisher Scoring algorithm based on a covariance parameterization of $\boldsymbol{\Phi}_*$. The third is the Lagrange multiplier algorithm based on a parameterization of $\boldsymbol{\Phi}_*$. Recall that $\boldsymbol{\Phi}$ is defined as a a $q \times q$ factor correlation matrix and $\boldsymbol{\Phi}_*$ is defined as a a $q \times q$ factor covariance matrix in Chapter 2. The last algorithm is called the 4-step procedure, which uses the second and the third algorithms. For notational convenience, $l(\boldsymbol{\theta};\mathbf{S})$ in (65) is written as $l(\boldsymbol{\theta})$ and $F(\boldsymbol{\Sigma},\mathbf{S})$ in (66) is written as $F(\boldsymbol{\theta})$ in the remainder of this thesis.

Initial guesses for $\boldsymbol{\Lambda}_*$, $\boldsymbol{\Phi}_*$ $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Delta}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_*$ are needed for the above iterative procedures. Therefore, a discussion on how to obtain these initial guesses are presented first.

3.3. Initial Guesses

Initial guesses for $\mathbf{\Phi}_*$, $\mathbf{\Lambda}_*$, $\mathbf{\Phi}$ and $\mathbf{\Lambda}$ are discussed by order because the latter one(s) are based on the former one(s).

3.3.1. An Initial Guess for $\mathbf{\Phi}_*$

Theorem 18 below supports the claim that the measured variables in $\mathbf{y}$ can be ordered such that the first q variables are influenced by one and only one factor.

Theorem 18. [A Well-Known Result]. *Rewrite* $\mathbf{\Lambda}_*$ *as* $\mathbf{\Lambda}_* = (\mathbf{\Lambda}_1^{*\prime} \quad \mathbf{\Lambda}^{*\prime})'$, *where* $\mathbf{\Lambda}_1^*$ *is a* $q \times q$ *nonsingular matrix. Without loss of generality,* $\mathbf{\Lambda}_1^*$ *in* $\mathbf{\Lambda}_* = (\mathbf{\Lambda}_1^{*\prime} \quad \mathbf{\Lambda}^{*\prime})'$ *can be replaced by a* $q \times q$ *identity matrix* $\mathbf{I}_q$ *so that* $\mathbf{\Lambda}_* = (\mathbf{I}_q \quad \mathbf{\Lambda}_2^{*\prime})'$, *where* $\mathbf{\Lambda}_2^*$ *is* $p_1 \times q$ *with* $p_1 = p - q$.

In confirmatory factor analysis, there is a prior structure for $\mathbf{\Lambda}_*$. Consequently, there is a prior structure for $\mathbf{\Lambda}_2^*$ by Theorem 18. Therefore, the measured variables in $\mathbf{y}$ can be ordered such that the first q variables are unifactorial; that is, they are influenced by one and only one factor. Then, $\mathbf{y}$ can be specialized to

$$\mathbf{y} = \boldsymbol{\mu} + \begin{pmatrix} \mathbf{I}_q \\ \mathbf{\Lambda}_2^* \end{pmatrix} \mathbf{f}^* + \boldsymbol{\epsilon}. \tag{67}$$

Furthermore, define $\mathbf{x}$ and $\mathbf{z}$ as follows:

$$\begin{aligned} \mathbf{x} &\stackrel{\text{def}}{=} \boldsymbol{\mu}_{\boldsymbol{x}} + \mathbf{f}^* + \boldsymbol{\epsilon}_{\boldsymbol{x}}, \quad \text{and} \\ \mathbf{z} &\stackrel{\text{def}}{=} \boldsymbol{\mu}_{\boldsymbol{z}} + \mathbf{\Lambda}_2^* \mathbf{f}^* + \boldsymbol{\epsilon}_{\boldsymbol{z}}, \end{aligned} \tag{68}$$

where $\mathbf{x}$ is $q \times 1$, $\mathbf{z}$ is $p_1 \times 1$, $\boldsymbol{\mu}' = [\boldsymbol{\mu}_{\boldsymbol{x}}' \ \boldsymbol{\mu}_{\boldsymbol{z}}']$ and $\boldsymbol{\epsilon}' = [\boldsymbol{\epsilon}_{\boldsymbol{x}}' \ \boldsymbol{\epsilon}_{\boldsymbol{z}}']$.

It follows that

$$\mathbf{y} = \begin{pmatrix} \mathbf{x} \\ \mathbf{z} \end{pmatrix} = \boldsymbol{\mu} + \begin{pmatrix} \mathbf{I}_q \\ \mathbf{\Lambda}_2^* \end{pmatrix} \mathbf{f}^* + \boldsymbol{\epsilon}. \tag{69}$$

The structure of (69) is adapted from Bentler [38]. Partition $\mathbf{\Psi}$ as

$$\mathbf{\Psi} = \begin{pmatrix} \mathbf{\Psi}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Psi}_2 \end{pmatrix}, \tag{70}$$

where $\mathbf{\Psi}_1$ is a $q \times q$ diagonal matrix and $\mathbf{\Psi}_2$ is a $p_1 \times p_1$ diagonal matrix.

In Bentler [38], $\mathbf{\Sigma}$ was written as

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{\boldsymbol{xx}} & \mathbf{\Sigma}_{\boldsymbol{xz}} \\ \mathbf{\Sigma}_{\boldsymbol{zx}} & \mathbf{\Sigma}_{\boldsymbol{zz}} \end{pmatrix}, \tag{71}$$

where $\mathbf{\Sigma}_{\boldsymbol{xx}} = \mathbf{\Phi}_* + \mathbf{\Psi}_1$, $\mathbf{\Sigma}_{\boldsymbol{zz}} = \mathbf{\Lambda}_2^* \mathbf{\Phi}_* \mathbf{\Lambda}_2^{*\prime} + \mathbf{\Psi}_2$, $\mathbf{\Sigma}_{\boldsymbol{zx}} = \mathbf{\Lambda}_2^* \mathbf{\Phi}_*$, and $\mathbf{\Phi}_* = \mathrm{Var}(\mathbf{f}^*)$. It follows that

$$\mathbf{\Sigma}_{\boldsymbol{zz}} = \mathbf{\Sigma}_{\boldsymbol{zx}} {\mathbf{\Phi}_*}^{-1} \mathbf{\Sigma}_{\boldsymbol{zx}}' + \mathbf{\Psi}_2, \tag{72}$$

which is from Bentler [38].

Let $\mathbf{S}$ be a $p \times p$ sample covariance matrix obtained from a sample of size N. Partition $\mathbf{S}$ as

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{\boldsymbol{xx}} & \mathbf{S}_{\boldsymbol{xz}} \\ \mathbf{S}_{\boldsymbol{zx}} & \mathbf{S}_{\boldsymbol{zz}}. \end{pmatrix}. \tag{73}$$

Accordingly, equation 72 implies

$$\mathbf{S}_{\boldsymbol{zz}} \approx \mathbf{S}_{\boldsymbol{zx}} {\mathbf{\Phi}_*}^{-1} \mathbf{S}_{\boldsymbol{zx}}' + \mathbf{\Psi}_2. \tag{74}$$

Theorem 19. [An Application of a Well-Known Result]. *Define* $\boldsymbol{\theta}_{\phi*}$ *as* $\boldsymbol{\theta}_{\phi*} = \mathrm{vech}({\mathbf{\Phi}_*}^{-1})$ *and* $\boldsymbol{\theta}_{\psi_2}$ *as* $\boldsymbol{\theta}_{\psi_2} \stackrel{\mathrm{def}}{=} \mathrm{diag}\,(\mathbf{\Psi}_2)$*, where* $\mathrm{vech}({\mathbf{\Phi}_*}^{-1})$ *and* $\mathrm{diag}\,(\mathbf{\Psi}_2)$ *are as in Table 56. Assume that* $\mathbf{S}_{\boldsymbol{zx}}$ *and* $\mathbf{S}_{\boldsymbol{zz}}$ *are matrices of constants.*

(a.) The approximate generalized least squares (GLS) estimator of $\boldsymbol{\theta}_{\phi*}$ *is*

$$\widehat{\boldsymbol{\theta}}_{\phi*} = \left(\mathbf{X}_{1.2}' \widehat{\mathbf{\Omega}}^{-1} \mathbf{X}_{1.2}\right)^{-1} \mathbf{X}_{1.2}' \widehat{\mathbf{\Omega}}^{-1} \,\mathrm{vec}\, \mathbf{S}_{\boldsymbol{zz}}, \tag{75}$$

where $\mathbf{X}_{1.2} = \left[\mathbf{I}_{p_1^2} - \mathbf{L}_{21,p_1} \left(\mathbf{L}_{21,p_1}' \widehat{\mathbf{\Omega}}^{-1} \mathbf{L}_{21,p_1}\right)^{-1} \mathbf{L}_{21,p_1}' \widehat{\mathbf{\Omega}}^{-1}\right] (\mathbf{S}_{\boldsymbol{zx}} \otimes \mathbf{S}_{\boldsymbol{zx}})\, \mathbf{D}_q$*,* $\widehat{\mathbf{\Omega}} = \mathbf{S}_{\boldsymbol{zz}} \otimes \mathbf{S}_{\boldsymbol{zz}}$*,* $\mathbf{D}_q$ *and* $\mathbf{L}_{21,p_1}$ *are defined in Table 56.*

(b.) The approximate ordinary least squares (OLS) estimator of $\boldsymbol{\theta}_{\phi *}$ *is*

$$\widehat{\boldsymbol{\theta}}_{\phi *} = \left(\mathbf{X}_{1.2}'\mathbf{X}_{1.2}\right)^{-1} \mathbf{X}_{1.2}' \operatorname{vec} \mathbf{S}_{\boldsymbol{zz}}, \tag{76}$$

where $\mathbf{X}_{1.2} = \left(\mathbf{I}_{p_1^2} - \mathbf{L}_{22,p_1}\right)(\mathbf{S}_{\boldsymbol{zx}} \otimes \mathbf{S}_{\boldsymbol{zx}})\, \mathbf{D}_q$.

Let $\widehat{\boldsymbol{\Phi}}_*$ be the initial guess for $\boldsymbol{\Phi}_*$. From Theorem 19, $\widehat{\boldsymbol{\Phi}}_*$ can be obtained from

$$\widehat{\boldsymbol{\Phi}}_* = (\operatorname{dvech} \widehat{\boldsymbol{\theta}}_{\phi *})^{-1}. \tag{77}$$

The eigenvalues of $\widehat{\boldsymbol{\Phi}}_*$ are checked because the covariance matrix $\widehat{\boldsymbol{\Phi}}_*$ is restricted to be positive definite. If one or more eigenvalues of $\widehat{\boldsymbol{\Phi}}_*$ are nonpositive, the following steps can be used.

Step 1. Let x_{min} be the minimum of the eigenvalues of $\widehat{\boldsymbol{\Phi}}_*$ and let x_{max} be the maximum of 0 and $-2x_{min}$.

Step 2. Generate a diagonal matrix, $\mathbf{A}$, whose diagonals are x_{max}.

Step 3. Replace $\widehat{\boldsymbol{\Phi}}_*$ by $\widehat{\boldsymbol{\Phi}}_* + \mathbf{A}$. The modified matrix, $\widehat{\boldsymbol{\Phi}}_* + \mathbf{A}$, is considered as the initial estimate for $\boldsymbol{\Phi}_*$.

3.3.2. An Initial Guess for $\boldsymbol{\Lambda}_*$

Recall $\boldsymbol{\Lambda}_*$ is defined as a $p \times q$ factor loading matrix based on $\widehat{\boldsymbol{\Phi}}_*$. Denote an initial guess for $\boldsymbol{\Lambda}_*$ by $\widehat{\boldsymbol{\Lambda}}_*$. By Theorem 18, the structure of $\boldsymbol{\Lambda}_* = (\mathbf{I}_q \quad \boldsymbol{\Lambda}_2^{*\prime})'$ contains two parts, one known and the other unknown. The structure for $\operatorname{vec} \boldsymbol{\Lambda}_2^*$ can be expressed as

$$\operatorname{vec} \boldsymbol{\Lambda}_2^* = \mathbf{W}_1^* \mathbf{L}_1^* + \mathbf{W}_2^* \boldsymbol{\theta}_{\boldsymbol{\lambda} *}, \tag{78}$$

where $\mathbf{W}_1^*$ and $\mathbf{W}_2^*$ are known matrices of 0's and 1's, $\mathbf{L}_1^*$ is a known vector, and $\boldsymbol{\theta}_{\boldsymbol{\lambda} *}$ is unknown.

Example. Consider the following hypothetical loading matrix $\boldsymbol{\Lambda}_*$:

$$\boldsymbol{\Lambda}_* = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0.57 & 0 \\ 0.22 & 0 \\ 0 & 0.90 \\ 0 & 0.98 \end{pmatrix}, \text{ where } q = 2 \text{ and } \boldsymbol{\Lambda}_2^* = \begin{pmatrix} 0.57 & 0 \\ 0.22 & 0 \\ 0 & 0.90 \\ 0 & 0.98 \end{pmatrix}. \tag{79}$$

In $\boldsymbol{\Lambda}_2^*$, 0's are considered known and the other entries of $\boldsymbol{\Lambda}_2^*$ are considered unknown. Accordingly, $\mathbf{W}_1^*$, $\mathbf{L}_1^*$, $\mathbf{W}_2^*$ and $\boldsymbol{\theta}_{\boldsymbol{\lambda}*}$ can be written as

$$\mathbf{W}_1^* \text{ is empty}, \quad \mathbf{L}_1^* \text{ is empty}, \quad \mathbf{W}_2^* = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \text{ and } \boldsymbol{\theta}_{\boldsymbol{\lambda}*} = \begin{pmatrix} \boldsymbol{\theta}_{\boldsymbol{\lambda}*1} \\ \boldsymbol{\theta}_{\boldsymbol{\lambda}*2} \\ \boldsymbol{\theta}_{\boldsymbol{\lambda}*3} \\ \boldsymbol{\theta}_{\boldsymbol{\lambda}*4} \end{pmatrix} = \begin{pmatrix} 0.57 \\ 0.22 \\ 0.90 \\ 0.98 \end{pmatrix}. \tag{80}$$

It follows from $\boldsymbol{\Sigma}_{\boldsymbol{zx}} = \boldsymbol{\Lambda}_2^* \boldsymbol{\Phi}_*$ that

$$\mathbf{S}_{\boldsymbol{zx}} \approx \boldsymbol{\Lambda}_2^* \boldsymbol{\Phi}_* \Longrightarrow \operatorname{vec} \mathbf{S}_{\boldsymbol{zx}} \approx (\boldsymbol{\Phi}_* \otimes \mathbf{I}_p) \operatorname{vec}(\boldsymbol{\Lambda}_2^*). \tag{81}$$

Combining (78) and (81) gives

$$\begin{aligned} &\operatorname{vec} \mathbf{S}_{\boldsymbol{zx}} \approx \left(\widehat{\boldsymbol{\Phi}}_* \otimes \mathbf{I}_p\right) \mathbf{W}_1^* \mathbf{L}_1^* + \left(\widehat{\boldsymbol{\Phi}}_* \otimes \mathbf{I}_p\right) \mathbf{W}_2^* \boldsymbol{\theta}_{\boldsymbol{\lambda}*} \Longrightarrow \\ &\operatorname{vec} \mathbf{S}_{\boldsymbol{zx}} - \left(\widehat{\boldsymbol{\Phi}}_* \otimes \mathbf{I}_p\right) \mathbf{W}_1^* \mathbf{L}_1^* \approx \left(\widehat{\boldsymbol{\Phi}}_* \otimes \mathbf{I}_p\right) \mathbf{W}_2^* \boldsymbol{\theta}_{\boldsymbol{\lambda}*}, \end{aligned} \tag{82}$$

where $\widehat{\boldsymbol{\Phi}}_*$ is the initial guess for $\boldsymbol{\Phi}_*$.

Now consider (82) as a linear regression model, the approximate ordinary least squares estimator of $\boldsymbol{\theta}_{\boldsymbol{\lambda}*}$ is

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{\lambda}*} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}, \tag{83}$$

where $\mathbf{X} = \left(\widehat{\boldsymbol{\Phi}}_* \otimes \mathbf{I}_p\right) \mathbf{W}_2^*$ and $\mathbf{y} = \text{vec}\, \mathbf{S}_{\boldsymbol{zx}} - \left(\widehat{\boldsymbol{\Phi}}_* \otimes \mathbf{I}_p\right) \mathbf{W}_1^* \mathbf{L}_1^*$.

Recall the structure for vec $\boldsymbol{\Lambda}_*$ is expressed as vec $\boldsymbol{\Lambda}_* = \mathbf{W}_{1*}\mathbf{L}_{1*} + \mathbf{W}_{2*}\boldsymbol{\theta}_{\boldsymbol{\lambda}*}$ in (52), where $\mathbf{W}_{1*}$ and $\mathbf{W}_{2*}$ are known matrices of 0's and 1's with the constraints that $\mathbf{W}_{2*}'\mathbf{W}_{1*} = \mathbf{0}$ and $\mathbf{W}_{2*}'\mathbf{W}_{2*} = \mathbf{I}$, $\mathbf{L}_{1*}$ is a known vector, and $\boldsymbol{\theta}_{\boldsymbol{\lambda}*}$ is unknown. It follows that the estimator of $\boldsymbol{\Lambda}_*$ can be obtained by

$$\widehat{\boldsymbol{\Lambda}}_* = \text{dvec}\left(\mathbf{W}_{1*}\mathbf{L}_{1*} + \mathbf{W}_{2*}\widehat{\boldsymbol{\theta}}_{\boldsymbol{\lambda}*}\right), \tag{84}$$

where dvec is as in Table 56. Note that the structures of $\mathbf{W}_{1*}$ and $\mathbf{W}_{2*}$ in this section are based on a factor covariance matrix $\boldsymbol{\Phi}_*$. If $\boldsymbol{\Phi}$ is restricted to be a factor correlation matrix, then the structures of $\mathbf{W}_{1*}$ and $\mathbf{W}_{2*}$ are modified accordingly. Details on this modification were discussed in Chapter 2.

3.3.3. An Initial Guess for $\boldsymbol{\Phi}$

Denote the initial guess for a factor correlation matrix $\boldsymbol{\Phi}$ by $\widehat{\boldsymbol{\Phi}}$. Define $\mathbf{D}_{\widehat{\phi}}$ as $\mathbf{D}_{\widehat{\phi}} = \text{Diag}\left(\left[\text{diag}(\widehat{\boldsymbol{\Phi}}_*)\right]^{\odot 1/2}\right)$. It follows that

$$\widehat{\boldsymbol{\Phi}} = \mathbf{D}_{\widehat{\phi}}^{-1}\widehat{\boldsymbol{\Phi}}_*\mathbf{D}_{\widehat{\phi}}^{-1}, \tag{85}$$

where the diagonal entries of $\widehat{\boldsymbol{\Phi}}$ are 1's and $\widehat{\boldsymbol{\Phi}}_*$ was given in (77).

3.3.4. An Initial Guess for $\boldsymbol{\Lambda}$

Denote the initial guess for a factor loading matrix $\boldsymbol{\Lambda}$ by $\widehat{\boldsymbol{\Lambda}}$ and the initial guess for $\boldsymbol{\theta}_{\boldsymbol{\lambda}}$ by $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\lambda}}$. It is trivial that

$$\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\Phi}}\widehat{\boldsymbol{\Lambda}}' = \widehat{\boldsymbol{\Lambda}}_*\widehat{\boldsymbol{\Phi}}_*\widehat{\boldsymbol{\Lambda}}_*'. \tag{86}$$

If follows from (85) and (86)

$$\widehat{\boldsymbol{\Lambda}} = \widehat{\boldsymbol{\Lambda}}_*\mathbf{D}_{\widehat{\phi}}. \tag{87}$$

where $\widehat{\mathbf{\Lambda}}_*$ was given in (84).

Note that $\widehat{\mathbf{\Lambda}}$ is calculated based on a factor correlation matrix $\mathbf{\Phi}$. It follows from $\mathbf{W}_2'\mathbf{W}_1 = \mathbf{0}$ and $\mathbf{W}_2'\mathbf{W}_2 = \mathbf{I}$ that

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{\lambda}} = (\mathbf{W}_2'\mathbf{W}_2)^{-1}\mathbf{W}_2' \operatorname{vec} \widehat{\mathbf{\Lambda}} \tag{88}$$

where $\operatorname{vec} \widehat{\mathbf{\Lambda}}$ was given in (87).

3.3.5. An Initial Guess for $\mathbf{\Psi}$

Denote the initial guess for $\mathbf{\Psi}$ by $\widehat{\mathbf{\Psi}}$ and $\widehat{\mathbf{\Psi}}$ is computed by

$$\widehat{\mathbf{\Psi}} = \mathbf{S} - \widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Phi}}\widehat{\mathbf{\Lambda}}'. \tag{89}$$

Initial guesses for $\mathbf{\Delta}$, $\mathbf{\Gamma}$ and $\mathbf{\Gamma}_*$ are discussed next in § 3.3.6, § 3.3.7 and § 3.3.8.

3.3.6. An Initial Guess for $\mathbf{\Delta}$

The supplement of Boik [33] gave initial guesses for $\boldsymbol{\theta}_{\boldsymbol{\delta}}$, $\boldsymbol{\delta}$ and $\mathbf{\Delta}$ under all structures in Table 4 except structure 4. The initial guesses for $\boldsymbol{\theta}_{\boldsymbol{\delta}}$, $\boldsymbol{\delta}$ and $\mathbf{\Delta}$ under structure 4 are provided in this section. The following discussion is composed of three subsections, (1), (2) and (3). The initial estimate for $\boldsymbol{\xi}_\delta$ is given in subsection (1). The initial guesses of $\boldsymbol{\theta}_{\boldsymbol{\delta}}$, $\boldsymbol{\delta}$ and $\mathbf{\Delta}$ are given in subsection (2). The discussion on solving for the implicit parameter $\boldsymbol{\eta}_{\boldsymbol{\delta}}(\boldsymbol{\theta}_{\boldsymbol{\delta}})$ is given in subsection (3).

(1) The initial estimate for $\boldsymbol{\xi}_\delta$ is obtained in this subsection. Denote the q-vector of eigenvalues of $\widehat{\mathbf{\Phi}}$ by $\boldsymbol{\ell}$, where $\widehat{\mathbf{\Phi}}$ is an initial guess for $\mathbf{\Phi}$. The approach of obtaining $\widehat{\mathbf{\Phi}}$ is discussed in § 3.3.3.

Define $\boldsymbol{\xi}_{\boldsymbol{\delta}t}$ as $\boldsymbol{\xi}_{\boldsymbol{\delta}t} \stackrel{\text{def}}{=} \boldsymbol{\xi}_\delta \odot \boldsymbol{\xi}_\delta$. The structure of structure 4 in (25) can be rewritten as follows:

$$\boldsymbol{\delta}\left(\boldsymbol{\xi}_{\boldsymbol{\delta}t}\right) = \mathbf{T}_2\boldsymbol{\xi}_{\boldsymbol{\delta}t}, \tag{90}$$

where $\mathbf{T}_2$ is a known full column-rank design matrix with dimension $q \times q_2$ and has non-negative entries. The ordinary least squares estimator, say $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}t}$, that minimizes $SSE(\boldsymbol{\xi}_{\boldsymbol{\delta}t}) = [\boldsymbol{\ell} - \mathbf{T}_2\boldsymbol{\xi}_{\boldsymbol{\delta}t}]' [\boldsymbol{\ell} - \mathbf{T}_2\boldsymbol{\xi}_{\boldsymbol{\delta}t}]$, is

$$\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}t} = (\mathbf{T}_2'\mathbf{T}_2)^{-1}\mathbf{T}_2'\boldsymbol{\ell}. \tag{91}$$

The value of $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}t}$ is used to obtain a starting value for $\boldsymbol{\xi}_\delta$. Denote this starting value for $\boldsymbol{\xi}_\delta$ by $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta},0}$ and $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta},0}$ is used in the iteration procedure discussed next. First, the following modification for $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}t}$ in (91) needs to be executed so that all elements of $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}t}$ are positive:

$$\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}t\_adj} = \frac{q}{k}(\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}t} + \mathbf{1}_{q_2}b), \tag{92}$$

where $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}t\_adj}$ is the adjusted value of $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}t}$, $k = \mathbf{1}_q'\mathbf{T}_2\left(\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}t} + \mathbf{1}_{q_2}b\right)$, and $b = \max\left(0, -1.25\min(\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}t})\right)$. The multiplier in b, -1.25, can be adjusted and the value of k will change accordingly. Accordingly, $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta},0}$ is computed as follows:

$$\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta},0} = \widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}t\_adj}^{\odot 0.5}. \tag{93}$$

The Gauss-Newton algorithm is employed to minimize the non-linear least squares loss function

$$SSE(\boldsymbol{\xi}_\delta) = [\boldsymbol{\ell} - \boldsymbol{\delta}(\boldsymbol{\xi}_\delta)]' [\boldsymbol{\ell} - \boldsymbol{\delta}(\boldsymbol{\xi}_\delta)] \tag{94}$$

with respect to $\boldsymbol{\xi}_\delta$. Note that the constraint $\mathbf{C}_1'\boldsymbol{\delta} = \mathbf{C}_1'\mathbf{T}_2\boldsymbol{\xi}_{\boldsymbol{\delta}}^{\odot 2} = \mathbf{c}_0$ is ignored in (94).

Taylor expansion is used to linearly approximate $\boldsymbol{\delta}(\boldsymbol{\xi}_\delta)$. That is, $\boldsymbol{\delta}(\boldsymbol{\xi}_\delta)$ can be written as follows:

$$\boldsymbol{\delta}(\boldsymbol{\xi}_\delta) \approx \boldsymbol{\delta}\left(\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}}\right) + \frac{\partial}{\partial\,\boldsymbol{\xi}_{\boldsymbol{\delta}}'}\boldsymbol{\delta}(\boldsymbol{\xi}_\delta)\Big|_{\boldsymbol{\xi}_{\boldsymbol{\delta}}=\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}}}\left(\boldsymbol{\xi}_\delta - \widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}}\right), \tag{95}$$

where $\partial\, \boldsymbol{\delta}(\boldsymbol{\xi}_\delta)/\partial\, \boldsymbol{\xi}_{\boldsymbol{\delta}}'\Big|_{\boldsymbol{\xi}_{\boldsymbol{\delta}}=\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}}} = 2\mathbf{T}_2\, \mathrm{Diag}\left(\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}}\right)$. Replace $\boldsymbol{\delta}\left(\boldsymbol{\xi}_\delta\right)$ in (94) by the right-hand-side of (95) and (94) can be rewritten as follows:

$$\left[\boldsymbol{\ell} - \boldsymbol{\delta}\left(\widehat{\boldsymbol{\xi}}_\delta\right) - \frac{\partial}{\partial\, \boldsymbol{\xi}_{\boldsymbol{\delta}}'}\boldsymbol{\delta}(\boldsymbol{\xi}_\delta)\Big|_{\boldsymbol{\xi}_{\boldsymbol{\delta}}=\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}}}\left(\boldsymbol{\xi}_\delta - \widehat{\boldsymbol{\xi}}_\delta\right)\right]' \left[\boldsymbol{\ell} - \boldsymbol{\delta}\left(\widehat{\boldsymbol{\xi}}_\delta\right) - \frac{\partial}{\partial\, \boldsymbol{\xi}_{\boldsymbol{\delta}}'}\boldsymbol{\delta}(\boldsymbol{\xi}_\delta)\Big|_{\boldsymbol{\xi}_{\boldsymbol{\delta}}=\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}}}\left(\boldsymbol{\xi}_\delta - \widehat{\boldsymbol{\xi}}_\delta\right)\right],$$

It follows that

$$\boldsymbol{\xi}_\delta - \widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}} \approx \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left[\boldsymbol{\ell} - \boldsymbol{\delta}\left(\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}}\right)\right],$$

where $\mathbf{X} = \partial\, \boldsymbol{\delta}(\boldsymbol{\xi}_\delta)/\partial\, \boldsymbol{\xi}_{\boldsymbol{\delta}}'\Big|_{\boldsymbol{\xi}_{\boldsymbol{\delta}}=\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}}} = 2\mathbf{T}_2\, \mathrm{Diag}\left(\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}}\right)$.

Accordingly, at $(i+1)^{\text{th}}$ iteration, the estimate for $\boldsymbol{\xi}_\delta$ is updated by

$$\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta},i+1} = \widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta},i} + \alpha_i\left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\mathbf{X}_i'(\boldsymbol{\ell} - \widehat{\boldsymbol{\delta}}_i), \tag{96}$$

where $\widehat{\boldsymbol{\delta}}_i = \boldsymbol{\delta}\left(\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta},i}\right)$, $\mathbf{X}_i = 2\mathbf{T}_2\, \mathrm{Diag}\left(\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta},i}\right)$ and $\alpha_i \in (0,1]$. The value of $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta},0}$ in (93) is used as a starting value for (96). The value of α_i is chosen to ensure that $SSE(\boldsymbol{\xi}_\delta)$ in (94) decreases at each iteration. The updating procedure is repeated until convergence. Accordingly, an initial estimate of $\boldsymbol{\xi}_\delta$ is obtained and denote this initial estimate by $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta},est}$.

(2) In this subsection, the Gauss-Newton algorithm is employed with the constraint $\mathbf{C}_1'\mathbf{T}_2\boldsymbol{\xi}_{\boldsymbol{\delta}}^{\odot 2} = \mathbf{c}_0$. The initial guess for $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ is obtained as a minimizer of

$$SSE(\boldsymbol{\theta}_{\boldsymbol{\delta}}) = \left[\boldsymbol{\ell} - \boldsymbol{\delta}(\boldsymbol{\theta}_{\boldsymbol{\delta}})\right]'\left[\boldsymbol{\ell} - \boldsymbol{\delta}(\boldsymbol{\theta}_{\boldsymbol{\delta}})\right] \tag{97}$$

with respect to $\boldsymbol{\theta}_{\boldsymbol{\delta}}$. Note that $\boldsymbol{\xi}_\delta = \mathbf{V}_1\boldsymbol{\eta}_{\boldsymbol{\delta}} + \mathbf{V}_2\boldsymbol{\theta}_{\boldsymbol{\delta}}$ and $\boldsymbol{\eta}_{\boldsymbol{\delta}}$ satisfies

$$\mathbf{C}_1'\boldsymbol{\delta} = \mathbf{C}_1'\mathbf{T}_2\left[(\mathbf{V}_1\boldsymbol{\eta}_{\boldsymbol{\delta}} + \mathbf{V}_2\boldsymbol{\theta}_\delta) \odot (\mathbf{V}_1\boldsymbol{\eta}_{\boldsymbol{\delta}} + \mathbf{V}_2\boldsymbol{\theta}_\delta)\right] = \mathbf{c}_0$$

for fixed $\boldsymbol{\theta}_{\boldsymbol{\delta}}$, where the matrices $\mathbf{V}_1$ and $\mathbf{V}_2$ are given in Theorem 2.

The procedure of linearizing the non-linear least squares loss function (97) is similar to the procedure used for estimating $\boldsymbol{\xi}_\delta$. Approximate the non-linear least squares loss function, (97), as follows:

$$\left[\boldsymbol{\ell}-\boldsymbol{\delta}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}\right)-\frac{\partial}{\partial\,\boldsymbol{\theta}_{\boldsymbol{\delta}}'}\boldsymbol{\delta}(\boldsymbol{\theta}_{\boldsymbol{\delta}})\Big|_{\boldsymbol{\theta}_\delta=\widehat{\boldsymbol{\theta}}_\delta}\left(\boldsymbol{\theta}_{\boldsymbol{\delta}}-\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}\right)\right]'\left[\boldsymbol{\ell}-\boldsymbol{\delta}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}\right)-\frac{\partial}{\partial\,\boldsymbol{\theta}_{\boldsymbol{\delta}}'}\boldsymbol{\delta}(\boldsymbol{\theta}_{\boldsymbol{\delta}})\Big|_{\boldsymbol{\theta}_\delta=\widehat{\boldsymbol{\theta}}_\delta}\left(\boldsymbol{\theta}_{\boldsymbol{\delta}}-\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}\right)\right],$$

which yields the normal equations in the algorithm.

At the $(i+1)^{\text{th}}$ iteration, the estimate of $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ is updated as

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta},i+1}=\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta},i}+\alpha_i\left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\mathbf{X}_i'(\boldsymbol{\ell}-\widehat{\boldsymbol{\delta}}_i),\ \text{where } \widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta},0}=\mathbf{V}_2'\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta},est},\tag{98}$$

$\widehat{\boldsymbol{\delta}}_i=\boldsymbol{\delta}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta},i}\right)$, $\mathbf{X}_i=\partial\,\boldsymbol{\delta}(\boldsymbol{\theta}_{\boldsymbol{\delta}})/\partial\,\boldsymbol{\theta}_{\boldsymbol{\delta}}'\Big|_{\boldsymbol{\theta}_\delta=\widehat{\boldsymbol{\theta}}_{\delta,i}}=2\mathbf{T}_2\,\mathrm{Diag}\left(\mathbf{V}_{1,i}\widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},i}+\mathbf{V}_{2,i}\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta},i}\right)\mathbf{V}_{2,i}$ and $\alpha_i\in(0,1]$. The value of $\widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},i}$ is computed based on $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta},i}$, which is discussed in the next subsection. The matrices $\mathbf{V}_{1,i}$ and $\mathbf{V}_{2,i}$ are the updated matrices $\mathbf{V}_1$ and $\mathbf{V}_2$ computed based on $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta},(i-1)}$ through Theorem 2, where $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta},(i-1)}=\mathbf{V}_{1,(i-1)}\widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},(i-1)}+\mathbf{V}_{2,(i-1)}\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta},(i-1)}$ is substituted for $\boldsymbol{\xi}_\delta$ in Theorem 2. The value of α_i is chosen to ensure that $SSE(\boldsymbol{\theta}_{\boldsymbol{\delta}})$ in (97) decreases at each iteration. An initial guess for $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ is obtained after convergence and denote this initial guess by $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta},ini}$.

Furthermore, denote $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta},i}$ at $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta},ini}$ by $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta},ini}$. Accordingly, the initial guess for $\boldsymbol{\delta}$, say $\widehat{\boldsymbol{\delta}}_{ini}$, is computed by

$$\widehat{\boldsymbol{\delta}}_{ini}=\mathbf{T}_2\left(\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta},ini}\odot\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta},ini}\right).\tag{99}$$

Denote the initial guess for $\boldsymbol{\Delta}$ by $\widehat{\boldsymbol{\Delta}}$ and $\widehat{\boldsymbol{\Delta}}$ is as follows:

$$\widehat{\boldsymbol{\Delta}}=\mathrm{Diag}(\widehat{\boldsymbol{\delta}}_{ini}),\tag{100}$$

where $\widehat{\boldsymbol{\delta}}_{ini}$ is given in (99).

The following subsection addresses how to obtain $\widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},i}$ for given $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta},i}$, $\mathbf{V}_{1,i}$ and $\mathbf{V}_{2,i}$.

(3) The constraint,

$$\mathbf{C}(\boldsymbol{\eta}_{\boldsymbol{\delta}}) \stackrel{\text{def}}{=} \mathbf{C}_1'\mathbf{T}_2\left[(\mathbf{V}_1\boldsymbol{\eta}_{\boldsymbol{\delta}} + \mathbf{V}_2\boldsymbol{\theta}_{\boldsymbol{\delta}}) \odot (\mathbf{V}_1\boldsymbol{\eta}_{\boldsymbol{\delta}} + \mathbf{V}_2\boldsymbol{\theta}_{\boldsymbol{\delta}})\right] - \mathbf{c}_0 = \mathbf{0}, \tag{101}$$

is considered to be a function of $\boldsymbol{\eta}_{\boldsymbol{\delta}}$, given a fixed vector $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ and fixed matrices $\mathbf{V}_1 \in \mathcal{O}(q_2, r_c)$ and $\mathbf{V}_2 \in \mathcal{O}(q_2, q_2 - r_c)$. The detailed computation for $\mathbf{V}_1$ and $\mathbf{V}_2$ are given in Theorem 2.

Equation 101 is expanded in a Taylor series around the i^{th} iteration guess for $\boldsymbol{\eta}_{\boldsymbol{\delta}}$, $\widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},i}$, that is,

$$\mathbf{0} = \mathbf{C}(\boldsymbol{\eta}_{\boldsymbol{\delta}}) \approx \mathbf{C}(\widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},i}) + \frac{\partial}{\partial\, \boldsymbol{\eta}_{\boldsymbol{\delta}}'}\mathbf{C}(\boldsymbol{\eta}_{\boldsymbol{\delta}})\Big|_{\boldsymbol{\eta}_{\boldsymbol{\delta}}=\widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},j}} (\boldsymbol{\eta}_{\boldsymbol{\delta}} - \widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},i}). \tag{102}$$

It follows that

$$\boldsymbol{\eta}_{\boldsymbol{\delta}} \approx \widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},i} - \left[\frac{\partial}{\partial\, \boldsymbol{\eta}_{\boldsymbol{\delta}}'}\mathbf{C}(\boldsymbol{\eta}_{\boldsymbol{\delta}})\Big|_{\boldsymbol{\eta}_{\boldsymbol{\delta}}=\widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},j}}\right]^{-1} \mathbf{C}(\widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},i}). \tag{103}$$

The initial guess for $\widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},0}$ is to set $\widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},0} = \mathbf{V}_1'\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}}$, where $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta}}$ is the current estimate of $\boldsymbol{\xi}_{\boldsymbol{\delta}}$.

A modified Newton-Raphson update for $\widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},i+1}$ is

$$\widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},i+1} = \widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},i} - \alpha_i \left[\frac{\partial}{\partial\, \boldsymbol{\eta}_{\boldsymbol{\delta}}'}\mathbf{C}(\boldsymbol{\eta}_{\boldsymbol{\delta}})\Big|_{\boldsymbol{\eta}_{\boldsymbol{\delta}}=\widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},j}}\right]^{-1} \mathbf{C}(\widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},i}), \tag{104}$$

where $\alpha_i \in (0, 1]$, $\partial\, \mathbf{C}(\boldsymbol{\eta}_{\boldsymbol{\delta}})/\partial\, \boldsymbol{\eta}_{\boldsymbol{\delta}}'\Big|_{\boldsymbol{\eta}_{\boldsymbol{\delta}}=\widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},j}} = 2\mathbf{T}_2\,\mathrm{Diag}(\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta},i})\mathbf{V}_1$ and $\widehat{\boldsymbol{\xi}}_{\boldsymbol{\delta},i} = \mathbf{V}_1\widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta},i} + \mathbf{V}_2\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}$.

The value of α_i is chosen so that $\sqrt{\mathbf{C}'(\widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta}})\mathbf{C}(\widehat{\boldsymbol{\eta}}_{\boldsymbol{\delta}})}$ decreases at each iteration.

The following discussion on how to obtain an initial guess for $\boldsymbol{\Gamma}$ is adapted from the supplement of Boik [33].

3.3.7. An Initial Guess for $\mathbf{\Gamma}$

Denote the orthogonal matrix of eigenvectors that corresponds to the ordered eigenvalues of $\widehat{\mathbf{\Phi}}$ by $\widehat{\mathbf{\Gamma}}_0$, where $\widehat{\mathbf{\Phi}}$ is the initial guess for $\mathbf{\Phi}$. The discussion on how to obtain $\widehat{\mathbf{\Phi}}$ can be found in (85) of § 3.3.3. Define $\widehat{\mathbf{\Phi}}_0$ as

$$\widehat{\mathbf{\Phi}}_0 = \widehat{\mathbf{\Gamma}}_0 \widehat{\mathbf{\Delta}} \widehat{\mathbf{\Gamma}}'_0, \tag{105}$$

where $\widehat{\mathbf{\Delta}}$ is the initial guess for $\mathbf{\Delta}$ and is given in (100). Note that $\widehat{\mathbf{\Delta}}$ is not the eigenvalues of $\widehat{\mathbf{\Phi}}$ and $\widehat{\mathbf{\Phi}}_0$ is not necessarily a correlation matrix. Therefore, $\widehat{\mathbf{\Gamma}}_0$ is not a suitable initial guess for $\mathbf{\Gamma}$.

The following discussion on the modified AS 213 algorithm is used to obtain an initial guess for $\mathbf{\Gamma}$ and this discussion is adapted from Lin and Bendel [39]. The algorithm AS 213 successively generates $(q-1)$ orthogonal matrix $\mathbf{P}_i$ so that

$$\widehat{\mathbf{\Phi}}_{q-1} = \mathbf{P}_{q-1}\mathbf{P}_{q-2}\ldots\mathbf{P}_2\mathbf{P}_1\widehat{\mathbf{\Phi}}_0\mathbf{P}'_1\mathbf{P}'_2\ldots\mathbf{P}'_{q-2}\mathbf{P}'_{q-1} \tag{106}$$

is a correlation matrix having the same eigenvalues as $\widehat{\mathbf{\Phi}}_0$.

The general form of $\mathbf{P}_i$ presented in Lin and Bendel [39] is as follows:

$$\mathbf{P}_i = \begin{pmatrix} \mathbf{I}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & c & 0 & s & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_2 & \mathbf{0} & \mathbf{0} \\ 0 & -s & 0 & c & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_3 \end{pmatrix}, \tag{107}$$

where $\mathbf{I}_j$ are identity submatrices for $j = 1, 2, 3$, $c = \cos(\theta)$ and $s = \sin(\theta)$. Each $\mathbf{P}_i$ in (106) is an elementary rotational matrix of the form:

$$\mathbf{P}_i = \begin{pmatrix} c & 0 & s \\ \mathbf{0} & \mathbf{I}_1 & \mathbf{0} \\ -s & 0 & c \end{pmatrix}, \tag{108}$$

where $\mathbf{I}_1$ is an identity submatrix, $c = \cos(\theta)$ and $s = \sin(\theta)$.

Define $\widehat{\mathbf{\Phi}}_1$ as $\widehat{\mathbf{\Phi}}_1 = \mathbf{P}_1\widehat{\mathbf{\Phi}}_0\mathbf{P}_1'$ and define $\widehat{\mathbf{\Phi}}_i$ as $\widehat{\mathbf{\Phi}}_i = \mathbf{P}_i\widehat{\mathbf{\Phi}}_{i-1}\mathbf{P}_i'$ for $i = 1, 2, \ldots, q-1$. Let α and γ be the smallest and largest diagonal elements of $\widehat{\mathbf{\Phi}}_{i-1}$ and choose $\mathbf{P}_i$ such that the locations of the diagonal elements c match the locations of α and γ. Let β be the element of $\widehat{\mathbf{\Phi}}_{i-1}$ corresponding to the element of s in (108). The rotation angle θ in c and s of (108) is chosen so that the smallest diagonal element of $\widehat{\mathbf{\Phi}}_i$ is set equal to one, that is,

$$\alpha\cos^2(\theta) + \gamma\sin^2(\theta) + \beta\sin(2\theta) = 1. \tag{109}$$

Equation 109 is equivalent to

$$1 - \frac{1}{2}(\alpha+\gamma) - \frac{1}{2}(\alpha-\gamma)\cos(2\theta) = \beta\sin(2\theta). \tag{110}$$

To solve equation 110, a quadratic equation in the variable $\cos(2\theta)$ is obtained when both sides of equation 110 are squared. Two solutions to (109) are

$$|\theta_1| = \left|\frac{\cos^{-1}\left(\frac{-B+\sqrt{B^2-4AC}}{2A}\right)}{2}\right| \quad \text{and} \quad |\theta_2| = \left|\frac{\cos^{-1}\left(\frac{-B-\sqrt{B^2-4AC}}{2A}\right)}{2}\right| \tag{111}$$

where $A = (\alpha-\gamma)^2/4+\beta^2/2$, $B = -2[1-(\alpha+\gamma)/2](\alpha-\gamma)/2$, $C = [1-(\alpha+\gamma)/2]^2-\beta^2$ and the signs of θ_1 and θ_2 are indeterminate. The choice between θ_1 and θ_2 is discussed in the following modified AS 213 algorithm.

Step 1 At the k^{th} iteration, denote the smallest and largest diagonal components of $\widehat{\mathbf{\Phi}}_k$ by α_k and γ_k, where $\widehat{\mathbf{\Phi}}_k = \widehat{\mathbf{\Gamma}}_k\widehat{\mathbf{\Delta}}\widehat{\mathbf{\Gamma}}_k'$ and $\widehat{\mathbf{\Gamma}}_k = \mathbf{P}_k\widehat{\mathbf{\Gamma}}_{k-1}$.

Step 2 Denote the value of β, θ_1 and θ_2 in (111) at the k^{th} iteration by β_k, $\theta_{1,k}$ and $\theta_{2,k}$, respectively. The sign of $\theta_{1,k}$ is determined based on if (109) is not satisfied, and so is the sign of $\theta_{2,k}$.

Step 3 At the k^{th} iteration, let $\mathbf{P}_{k,1}$ be the value of $\mathbf{P}_k$ in (108) computed based on $\theta_{1,k}$. Let $\mathbf{P}_{k,2}$ be the value of $\mathbf{P}_k$ in (108) computed based on $\theta_{2,k}$. Let $\widehat{\mathbf{\Gamma}}_{k+1,1}$ be $\widehat{\mathbf{\Gamma}}_{k+1,1} = \mathbf{P}_{k+1,1}\widehat{\mathbf{\Gamma}}_k$ and let $\widehat{\mathbf{\Gamma}}_{k+1,2}$ be $\widehat{\mathbf{\Gamma}}_{k+1,2} = \mathbf{P}_{k+1,2}\widehat{\mathbf{\Gamma}}_k$.

Step 4 Define SSE_1 as $SSE_1 = \operatorname{vec}'\left(\widehat{\mathbf{\Phi}}_{k+1,1} - \widehat{\mathbf{\Phi}}\right)\mathbf{V}_{\mathbf{\Phi}} \operatorname{vec}\left(\widehat{\mathbf{\Phi}}_{k+1,1} - \widehat{\mathbf{\Phi}}\right)$ and define SSE_2 as $SSE_2 = \operatorname{vec}'\left(\widehat{\mathbf{\Phi}}_{k+1,2} - \widehat{\mathbf{\Phi}}\right)\mathbf{V}_{\mathbf{\Phi}} \operatorname{vec}\left(\widehat{\mathbf{\Phi}}_{k+1,2} - \widehat{\mathbf{\Phi}}\right)$, where $\widehat{\mathbf{\Phi}}_{k+1,1} = \widehat{\mathbf{\Gamma}}_{k+1,1}\widehat{\mathbf{\Delta}}\widehat{\mathbf{\Gamma}}'_{k+1,1}$, $\widehat{\mathbf{\Phi}}_{k+1,2} = \widehat{\mathbf{\Gamma}}_{k+1,2}\widehat{\mathbf{\Delta}}\widehat{\mathbf{\Gamma}}'_{k+1,2}$, and $\mathbf{V}_{\mathbf{\Phi}}$ is a positive semi-definite matrix.

Step 5 Choose $\widehat{\mathbf{\Gamma}}_{k+1,1}$ over $\widehat{\mathbf{\Gamma}}_{k+1,2}$ if $SSE_1 < SSE_2$.

Step 6 Iterate Steps 1-5 for $k = 1, \ldots, q-1$.

In Step 4, the parameters can be estimated by ordinary least squares. That is, one choice for $\mathbf{V}_{\mathbf{\Phi}}$ can be chosen as an identity matrix. Another choice for $\mathbf{V}_{\mathbf{\Phi}}$ is

$$\mathbf{V}_{\mathbf{\Phi}} = \mathbf{E}'_q\left(\mathbf{E}_q\widehat{\mathbf{\Omega}}_{\mathbf{\Phi}}\mathbf{E}'_q\right)^{-1}\mathbf{E}_q, \tag{112}$$

where $\mathbf{E}_q = \sum_{s=1}^{q-1}\sum_{t=s+1}^{q}\mathbf{e}_u^{q(q-1)/2}\left(\mathbf{e}_s^q \otimes \mathbf{e}_t^q\right)'\mathbf{N}_q$, $u = (2q-1-s)s/2 + t - q$, $\mathbf{N}_q$ is defined in Table 56, $\widehat{\mathbf{\Omega}}_{\mathbf{\Phi}} = \widehat{\mathbf{U}}(\widehat{\mathbf{\Phi}} \otimes \widehat{\mathbf{\Phi}})\widehat{\mathbf{U}}'$, and $\widehat{\mathbf{U}} = \mathbf{I}_{q^2} - (\mathbf{I}_q \otimes \widehat{\mathbf{\Phi}})\mathbf{L}_{22,q}$. The $q(q-1)/2 \times q^2$ matrix $\mathbf{E}_q$ in (112) is constructed so that $\mathbf{E}_q \operatorname{vec}\mathbf{\Phi}$ is a vector that contains the $q(q-1)/2$ factor correlations below the main diagonal of $\mathbf{\Phi}$.

The choice for $\mathbf{E}_q\widehat{\mathbf{\Omega}}_{\mathbf{\Phi}}\mathbf{E}'_q$ in $\mathbf{V}_{\mathbf{\Phi}}$ of (112) is based on the asymptotic variance of $\operatorname{vec}\mathbf{R}$, where $\mathbf{R}$ is the sample correlation matrix. By equation 158 in the supplement of Boik [33], it showed that if the multivariate normality is satisfied, then

$$\mathbf{E}_q\mathbf{\Omega}_{\mathbf{\Phi}}\mathbf{E}'_q \stackrel{\text{def}}{=} \lim_{n\to\infty} \operatorname{Var}\left[\sqrt{n}\mathbf{E}_q \operatorname{vec}(\mathbf{R})\right] = 2\mathbf{E}_q\mathbf{U}(\mathbf{\Phi} \otimes \mathbf{\Phi})\mathbf{U}'\mathbf{E}'_q, \tag{113}$$

where $\mathbf{U} = \mathbf{I}_{q^2} - (\mathbf{I}_q \otimes \mathbf{\Phi})\mathbf{L}_{22,q}$.

Accordingly, $\widehat{\mathbf{\Omega}}_{\mathbf{\Phi}}$ in (112) is obtained through replacing $\mathbf{\Phi}$ in (113) by $\widehat{\mathbf{\Phi}}$. Although the choice for $\mathbf{V}_{\mathbf{\Phi}}$ in (112) is not based on the asymptotic variance of $\operatorname{vec}(\widehat{\mathbf{\Phi}})$, it is a sensible choice because the goal is merely to compute an initial estimate of $\mathbf{\Gamma}$.

3.3.8. An Initial Guess for $\mathbf{\Gamma}_*$

Denote the orthogonal matrix of eigenvectors that corresponds to the ordered eigenvalues of $\widehat{\mathbf{\Phi}}_*$ by $\widehat{\mathbf{\Gamma}}_*$, where $\widehat{\mathbf{\Phi}}_*$ is the initial guess for $\mathbf{\Phi}_*$. The discussion on how to obtain $\widehat{\mathbf{\Phi}}_*$ can be found in (77) of § 3.3.1. An initial guess for $\mathbf{\Gamma}_*$ is $\widehat{\mathbf{\Gamma}}_*$.

The following sections present four algorithms as mentioned before. They are a modified Fisher Scoring algorithm based on a correlation parameterization of $\mathbf{\Phi}$, a modified Fisher Scoring algorithm based on a covariance parameterization of $\mathbf{\Phi}_*$, the Lagrange multiplier algorithm based on a parameterization of $\mathbf{\Phi}_*$ and the 4-step procedure.

3.4. A Correlation Parameterization-Based Algorithm

Before a modified Fisher scoring algorithm based on a correlation parameterization of $\mathbf{\Phi}$ is discussed, a description of the Newton-Raphson algorithm is presented.

3.4.1. Newton-Raphson Algorithm

Denote the estimate of $\boldsymbol{\theta}$ after the i^{th} iteration by $\widehat{\boldsymbol{\theta}}_i$. Denote the minimizer of $F(\boldsymbol{\theta})$ by $\hat{\boldsymbol{\theta}}$, which is the maximizer of $l(\boldsymbol{\theta})$. Expand $F(\hat{\boldsymbol{\theta}})$ in a Taylor series around $\hat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_i$. That is,

$$F(\hat{\boldsymbol{\theta}}) = F(\widehat{\boldsymbol{\theta}}_i) + (\hat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_i)'\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i} + \frac{1}{2}(\hat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_i)'\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i}(\hat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_i) + o\left(\|\hat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_i\|^2\right), \quad (114)$$

where the expressions of $\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}$ and $\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}$ are given in Theorem 17. The equation 114 can be written as

$$F(\hat{\boldsymbol{\theta}}) \approx F(\widehat{\boldsymbol{\theta}}_i) + (\hat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_i)'\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i} + \frac{1}{2}(\hat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_i)'\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i}(\hat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_i). \quad (115)$$

Define $\mathbf{g}_{\hat{\boldsymbol{\theta}}_i}$ as $\mathbf{g}_{\hat{\boldsymbol{\theta}}_i} = -\frac{n}{2}\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i}$ and define $\mathbf{H}_{\hat{\boldsymbol{\theta}}_i}$ as $\mathbf{H}_{\hat{\boldsymbol{\theta}}_i} = -\frac{n}{2}\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i}$, which is called the Hessian matrix. To find the vector $\hat{\boldsymbol{\theta}}$ that minimizes $F(\boldsymbol{\theta})$, ignore the remainder and take the derivative of (115) with respect to $\hat{\boldsymbol{\theta}}$ and set it to zero. Specifically,

$$\frac{\partial F(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}} = \mathbf{0} \quad \Longrightarrow \quad \mathbf{g}_{\hat{\boldsymbol{\theta}}_i} + \mathbf{H}_{\hat{\boldsymbol{\theta}}_i}(\hat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_i) \approx \mathbf{0}.$$

It follows that

$$\hat{\boldsymbol{\theta}} \approx \widehat{\boldsymbol{\theta}}_i - \mathbf{H}^{-1}_{\hat{\boldsymbol{\theta}}_i}\mathbf{g}_{\hat{\boldsymbol{\theta}}_i}. \tag{116}$$

The left-hand-side of (116) becomes the new guess and the procedure is repeated. That is,

$$\hat{\boldsymbol{\theta}}_{i+1} = \widehat{\boldsymbol{\theta}}_i - \mathbf{H}^{-1}_{\hat{\boldsymbol{\theta}}_i}\mathbf{g}_{\hat{\boldsymbol{\theta}}_i}. \tag{117}$$

Iteration continues until convergence. Examine remainder in (114), $o\left(\|\hat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}_i\|^2\right)$, if the original guess is far from the minimizer of $F(\boldsymbol{\theta})$, then the Taylor series expansion is not accurate. Accordingly, the algorithm may not converge.

3.4.2. Fisher Scoring Algorithm

The Newton-Raphson algorithm does not require that $F(\boldsymbol{\theta})$ be a log likelihood function. $F(\boldsymbol{\theta})$ could be a function that contains no random variables. The Fisher scoring algorithm is applicable specifically if $F(\boldsymbol{\theta})$ is a log likelihood function. Define $\mathbf{I}_{\boldsymbol{\theta}}$ as

$$\mathbf{I}_{\boldsymbol{\theta}} = -E\left[\mathbf{H}_{\boldsymbol{\theta}}\right], \tag{118}$$

which is called Fisher's information matrix. Define $\mathbf{I}_{\hat{\boldsymbol{\theta}}_i}$ as $\mathbf{I}_{\hat{\boldsymbol{\theta}}_i} = \mathbf{I}_{\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_i}$. The Fisher scoring algorithm (named after R.A. Fisher) replaces $\mathbf{H}_{\hat{\boldsymbol{\theta}}_i}$ by $-\mathbf{I}_{\hat{\boldsymbol{\theta}}_i}$.

Based on (117), the corresponding Fisher scoring procedure can be summarized as

$$\hat{\boldsymbol{\theta}}_{i+1} = \widehat{\boldsymbol{\theta}}_i + \mathbf{I}^{-1}_{\hat{\boldsymbol{\theta}}_i}\mathbf{g}_{\hat{\boldsymbol{\theta}}_i}. \tag{119}$$

After plugging the expressions of $\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}$ and $\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}$ into (119), it follows that

$$\hat{\boldsymbol{\theta}}_{i+1} = \widehat{\boldsymbol{\theta}}_i + \left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_i}\left(\widehat{\boldsymbol{\Sigma}}_i \otimes \widehat{\boldsymbol{\Sigma}}_i\right)^{-1}\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_i}\right]^{-1}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_i}\left(\widehat{\boldsymbol{\Sigma}}_i \otimes \widehat{\boldsymbol{\Sigma}}_i\right)^{-1}\operatorname{vec}\left(\mathbf{S} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{i}}\right), \quad (120)$$

where $\widehat{\boldsymbol{\Sigma}}_i = \boldsymbol{\Sigma}\left(\widehat{\boldsymbol{\theta}}_i\right)$.

Both the Newton-Raphson algorithm and the Fisher scoring algorithm described above need to be modified to work with the implicit parameterization used in this thesis. A modified Fisher scoring algorithm is discussed in the following section. The modified Fisher scoring algorithm (MFSA) is based on the modified Newton algorithm in Boik [18] and the modified Fisher scoring algorithm in Boik [34].

3.4.3. MFSA Based On a Parameterization of $\boldsymbol{\Phi}$

The modified Fisher scoring iterative procedure consists of six steps:

Step 1 Compute initial guesses for $\widehat{\boldsymbol{\Lambda}}$, $\widehat{\boldsymbol{\Delta}}$, $\widehat{\boldsymbol{\Gamma}}$ and $\widehat{\boldsymbol{\Psi}}$. Denote these guesses by $\widehat{\boldsymbol{\Lambda}}_0 = \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\lambda},0})$, $\widehat{\boldsymbol{\Delta}}_0 = \boldsymbol{\Delta}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta},0})$, $\widehat{\boldsymbol{\Gamma}}_0 = \boldsymbol{\Gamma}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta},0}, \widehat{\boldsymbol{\theta}}_{\gamma,0})$ and $\widehat{\boldsymbol{\Psi}}_0 = \boldsymbol{\Psi}(\widehat{\boldsymbol{\theta}}_{\psi,0})$. An algorithm to obtain initial guesses was given in § 3.3.

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

$$\widehat{\boldsymbol{\Sigma}}_i = \widehat{\boldsymbol{\Lambda}}_i\widehat{\boldsymbol{\Gamma}}_i\widehat{\boldsymbol{\Delta}}_i\widehat{\boldsymbol{\Gamma}}'_i\widehat{\boldsymbol{\Lambda}}'_i + \widehat{\boldsymbol{\Psi}}_i.$$

Step 3 Set $\widehat{\boldsymbol{\theta}}_{\gamma,i} = \mathbf{0}$ and set $\hat{\boldsymbol{\theta}}_i = \begin{pmatrix}\widehat{\boldsymbol{\theta}}'_{\boldsymbol{\lambda},i} & \widehat{\boldsymbol{\theta}}'_{\boldsymbol{\delta},i} & \mathbf{0}' & \widehat{\boldsymbol{\theta}}'_{\psi,i}\end{pmatrix}'$. Use (120) to update $\hat{\boldsymbol{\theta}}_i$ as follows:

$$\hat{\boldsymbol{\theta}}_{i+1} = \widehat{\boldsymbol{\theta}}_i + \alpha\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_i}\left(\widehat{\boldsymbol{\Sigma}}_i \otimes \widehat{\boldsymbol{\Sigma}}_i\right)^{-1}\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_i}\right]^{-1}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_i}\left(\widehat{\boldsymbol{\Sigma}}_i \otimes \widehat{\boldsymbol{\Sigma}}_i\right)^{-1}\operatorname{vec}\left(\mathbf{S} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{i}}\right),$$

where $\alpha \in (0, 1]$ and $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_i}$ was given in (48). This is not the traditional Fisher scoring algorithm because $\widehat{\boldsymbol{\theta}}_{\gamma,i}$ is not updated not from iteration to iteration. Instead, $\widehat{\boldsymbol{\theta}}_{\gamma,i}$ is set to zero and updated from zero in each iteration.

Step 4 Use (40) in Chapter 2 to compute $\mathbf{G}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta},i+1}, \widehat{\boldsymbol{\theta}}_{\boldsymbol{\gamma},i+1}\right)$ based on the updated $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta},i+1}$ and $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\gamma},i+1}$ from the previous step.

Step 5 Compute $\widehat{\boldsymbol{\Lambda}}_{i+1} = \boldsymbol{\Lambda}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\lambda},i+1}\right)$, $\widehat{\boldsymbol{\Delta}}_{i+1} = \boldsymbol{\Delta}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta},i+1}\right)$, $\widehat{\boldsymbol{\Gamma}}_{i+1} = \widehat{\boldsymbol{\Gamma}}_i\mathbf{G}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta},i+1}, \widehat{\boldsymbol{\theta}}_{\boldsymbol{\gamma},i+1}\right)$ and $\widehat{\boldsymbol{\Psi}}_{i+1} = \widehat{\boldsymbol{\Psi}}\left(\widehat{\boldsymbol{\theta}}_{\psi,i+1}\right)$ based on the updated $\hat{\boldsymbol{\theta}}_{i+1}$ from Step 3.

Step 6 Iterate Steps 2-5 until convergence.

Note that the modified Fisher scoring algorithm discussed in this section is based on a correlation parameterization of $\boldsymbol{\Phi}$. The correlation parameterization of eigenvalues and eigenvectors of $\boldsymbol{\Phi}$ was given in Chapter 2. Two covariance parameterization-based algorithms are presented next. One is called the modified Fisher scoring algorithm based on a covariance parameterization of $\boldsymbol{\Phi}_*$ and the other is called the Lagrange algorithm based on a Parameterization of $\boldsymbol{\Phi}_*$. The relevant discussion on the parameterization of eigenvectors of $\boldsymbol{\Phi}_*$ was provided in Chapter 2.

3.5. Covariance Parameterization-Based Algorithms

3.5.1. MFSA Based On a Parameterization of $\boldsymbol{\Phi}_*$

Recall that the quantities, $\boldsymbol{\Lambda}_*$ and $\boldsymbol{\Phi}_*$ are parameterized as functions of $\boldsymbol{\theta}_{\boldsymbol{\lambda}*}$ and $\boldsymbol{\theta}_{\boldsymbol{\phi}*}$, respectively. That is, $\boldsymbol{\Lambda}_* = \boldsymbol{\Lambda}_*(\boldsymbol{\theta}_{\boldsymbol{\lambda}*})$ and $\boldsymbol{\Phi}_* = \boldsymbol{\Phi}_*(\boldsymbol{\theta}_{\boldsymbol{\phi}*})$. Theorem 20 below provides expressions that are useful for the modified Fisher scoring algorithm.

Theorem 20. [Original result].

$$\begin{aligned}
\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_{\lambda*}} &= 2\mathbf{N}_p(\boldsymbol{\Lambda}_*\boldsymbol{\Phi}_* \otimes \mathbf{I}_p)\mathbf{W}_2, \\
\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_{\phi*}} &= (\boldsymbol{\Lambda}_* \otimes \boldsymbol{\Lambda}_*)\mathbf{D}_q, \\
\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_{\psi*}} &= \mathbf{L}_{21,p}\mathbf{D}^{(1)}_{\boldsymbol{\psi};\boldsymbol{\theta}'_{\psi*}}.
\end{aligned}$$

For notational convenience, write a $\boldsymbol{\theta}^*$ as $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^{*\prime} \;\; \boldsymbol{\theta}_2^{*\prime} \;\; \boldsymbol{\theta}_3^{*\prime})' = (\boldsymbol{\theta}_{\lambda*}' \;\; \boldsymbol{\theta}_{\phi*}' \;\; \boldsymbol{\theta}_{\psi*}')'$, where $\boldsymbol{\theta}_i^*$ has dimension $\nu_i^* \times 1$. Derivatives of $\boldsymbol{\sigma}$ with respect to $\boldsymbol{\theta}^*$ can be re-assembled from derivatives with respect to $\{\boldsymbol{\theta}_j^*\}_{j=1}^3$ by using the elementary matrix, $\mathbf{E}_{i,\boldsymbol{\nu}^*}$, which is defined in Table 56.

The first derivative of $\boldsymbol{\sigma}$ with respect to $\boldsymbol{\theta}^*$ is

$$\mathbf{D}_{\boldsymbol{\sigma};\boldsymbol{\theta}^{*\prime}}^{(1)} \stackrel{\text{def}}{=} \frac{\partial \text{vec}\boldsymbol{\Sigma}}{\partial\, \boldsymbol{\theta}^{*\prime}} = \begin{pmatrix} \mathbf{D}_{\boldsymbol{\sigma};\boldsymbol{\theta}_{\lambda*}'}^{(1)} & \mathbf{D}_{\boldsymbol{\sigma};\boldsymbol{\theta}_{\phi*}'}^{(1)} & \mathbf{D}_{\boldsymbol{\sigma};\boldsymbol{\theta}_{\psi*}'}^{(1)} \end{pmatrix} = \sum_{i=1}^{3} \mathbf{D}_{\boldsymbol{\sigma};\boldsymbol{\theta}^{*\prime}_i}^{(1)} \mathbf{E}_{i,\boldsymbol{\nu}^*}'. \tag{121}$$

The modified Fisher scoring algorithm discussed in this section is used to obtain estimates for $\boldsymbol{\Lambda}_*$, $\boldsymbol{\Phi}_*$ and $\boldsymbol{\Psi}$ that minimizes $F(\boldsymbol{\Sigma}, \mathbf{S})$ in (66). This modified Fisher scoring iterative procedure consists of the following six steps:

Step 1 Compute initial guesses for $\widehat{\boldsymbol{\Lambda}}_*$, $\widehat{\boldsymbol{\Phi}}_*$ and $\widehat{\boldsymbol{\Psi}}$. Denote these guesses by $\widehat{\boldsymbol{\Lambda}}_{*0} = \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\lambda}*,0})$, $\widehat{\boldsymbol{\Phi}}_{*0} = \boldsymbol{\Phi}_*(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\psi}*,0})$ and $\widehat{\boldsymbol{\Psi}}_0 = \boldsymbol{\Psi}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\psi},0})$. An algorithm to obtain initial guesses was given in § 3.3.

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

$$\widehat{\boldsymbol{\Sigma}}_i = \widehat{\boldsymbol{\Lambda}}_{*i}\widehat{\boldsymbol{\Phi}}_{*i}\widehat{\boldsymbol{\Lambda}}_{*i}' + \widehat{\boldsymbol{\Psi}}_i.$$

Step 3 Set $\hat{\boldsymbol{\theta}}_i^* = \begin{pmatrix} \widehat{\boldsymbol{\theta}}_{\boldsymbol{\lambda}*,i}' & \widehat{\boldsymbol{\theta}}_{\boldsymbol{\phi}*,i}' & \widehat{\boldsymbol{\theta}}_{\boldsymbol{\psi},i}' \end{pmatrix}'$. Update $\hat{\boldsymbol{\theta}}_i^*$ as follows:

$$\hat{\boldsymbol{\theta}}_{i+1}^* = \hat{\boldsymbol{\theta}}_i^* + \alpha \left[\mathbf{D}_{\boldsymbol{\sigma};\boldsymbol{\theta}_i^{*\prime}}^{(1)\prime} \left(\widehat{\boldsymbol{\Sigma}}_i \otimes \widehat{\boldsymbol{\Sigma}}_i \right)^{-1} \mathbf{D}_{\boldsymbol{\sigma};\boldsymbol{\theta}_i^{*\prime}}^{(1)} \right]^{-1} \mathbf{D}_{\boldsymbol{\sigma};\boldsymbol{\theta}_i^{*\prime}}^{(1)\prime} \left(\widehat{\boldsymbol{\Sigma}}_i \otimes \widehat{\boldsymbol{\Sigma}}_i \right)^{-1} \text{vec} \left(\mathbf{S} - \widehat{\boldsymbol{\Sigma}}_{\mathbf{i}} \right),$$

where $\alpha \in (0, 1]$ and $\mathbf{D}_{\boldsymbol{\sigma};\boldsymbol{\theta}_i^{*\prime}}^{(1)}$ is given in (121).

Step 4 Compute $\widehat{\boldsymbol{\Lambda}}_{*i+1} = \boldsymbol{\Lambda}_* \left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\lambda}*,i+1} \right)$, $\widehat{\boldsymbol{\Phi}}_{*i+1} = \boldsymbol{\Phi}_*(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\phi}*,i+1})$ and $\widehat{\boldsymbol{\Psi}}_{i+1} = \widehat{\boldsymbol{\Psi}} \left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\psi},i+1} \right)$ based on the updated $\hat{\boldsymbol{\theta}}_{i+1}$ from Step 3.

Step 5 Iterate Steps 2-4 until convergence.

Note that this modified Fisher scoring algorithm is based on the covariance parameterization of $\mathbf{\Phi}_*$. The relevant covariance parameterization of eigenvalues and eigenvectors of $\mathbf{\Phi}_*$ was given in Chapter 2.

The Lagrange multipliers algorithm (LMA) provides an alternative method to solve constrained optimization problems. It can deal with both equality and inequality constraints.

3.5.2. LMA Based On a Parameterization of $\mathbf{\Phi}_*$

3.5.2.1. An Introduction on LMA: The Lagrange multipliers method for nonlinear optimization problems with equality constraints can be formulated as

$$\begin{aligned} &\text{minimize} \quad L(\boldsymbol{\theta}) \\ &\text{subject to} \quad \mathbf{z}(\boldsymbol{\theta}) = \mathbf{0}, \end{aligned} \tag{122}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_n)'$ is the variable vector, $L : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{z} : \mathbb{R}^n \to \mathbb{R}^m$ are given functions, $\mathbf{z}(\boldsymbol{\theta}) = (z_1(\boldsymbol{\theta}), z_2(\boldsymbol{\theta}), \cdots, z_m(\boldsymbol{\theta}))'$ is the constraints function vector and $m \leq n$.

The Lagrangian function $Q : \mathbb{R}^{n+m} \to \mathbb{R}$ is defined as

$$Q(\boldsymbol{\theta}, \boldsymbol{\zeta}) = L(\boldsymbol{\theta}) + \boldsymbol{\zeta}'\mathbf{z}, \tag{123}$$

where $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \cdots, \zeta_m)'$ is called an m-vector of Lagrange multipliers. Equation (122) and equation (123) are from Dimitri [40].

The solution to (122) satisfies

$$\frac{\partial Q}{\partial \boldsymbol{\omega}} = \mathbf{0}, \qquad \text{where } \boldsymbol{\omega} = (\boldsymbol{\theta}' \quad \boldsymbol{\zeta}')'. \tag{124}$$

Equation 124 gives a necessary condition for $\boldsymbol{\theta}$ to be a solution to (122).

<u>3.5.2.2. Application of Lagrange Algorithm</u>: The application of the Lagrange algorithm is adapted from Boik [33]. Define the $(q-1)$-dimensional constraints function $\mathbf{z}(\boldsymbol{\theta}_*)$ as

$$\mathbf{z}(\boldsymbol{\theta}_*) = \mathbf{C}'\mathbf{L}'_{21,q} \operatorname{vec} \boldsymbol{\Phi} = \mathbf{C}' \operatorname{diag}(\boldsymbol{\Phi}), \tag{125}$$

where $\boldsymbol{\theta}_* = (\boldsymbol{\theta}'_{\lambda*} \;\; \boldsymbol{\theta}'_{\delta*} \;\; \boldsymbol{\theta}'_{\gamma*} \;\; \boldsymbol{\theta}'_{\psi*})'$ with $\mathbf{C}$ a full column-rank matrix that satisfies $\mathcal{R}(\mathbf{C}) = \mathcal{N}(\mathbf{1}'_q)$. Note that $\mathbf{C}'\mathbf{1}_q = \mathbf{0}$ because $\mathcal{R}(\mathbf{C}) \subseteq \mathcal{N}(\mathbf{1}'_q)$. The equality constrained optimization problem can be written as follows:

$$\begin{aligned} &\text{minimize} \quad F(\boldsymbol{\theta}_*) \\ &\text{subject to} \quad \mathbf{z}(\boldsymbol{\theta}_*) = \mathbf{0}, \end{aligned} \tag{126}$$

where $F(\boldsymbol{\theta}_*) = \operatorname{trace}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) + \ln|\boldsymbol{\Sigma}|$, as defined in (66).

The constrained minimizer of $F(\boldsymbol{\theta}_*)$ is a solution to

$$\frac{\partial Q}{\partial \boldsymbol{\omega}} = \mathbf{0} \text{ for } \boldsymbol{\theta}_* \text{ and } \boldsymbol{\zeta}, \text{ where } \boldsymbol{\omega} = \begin{pmatrix}\boldsymbol{\theta}'_* & \boldsymbol{\zeta}'\end{pmatrix}', \quad Q = Q(\boldsymbol{\omega}) = F(\boldsymbol{\theta}_*) - \boldsymbol{\zeta}'\mathbf{z}(\boldsymbol{\theta}_*) \tag{127}$$

and $\boldsymbol{\zeta}$ is a $(q-1)$-vector of Lagrange multipliers.

Theorem 16 in Chapter 2 is used for covariance parameterization-based algorithms, such as the Lagrange multipliers algorithm based on $\boldsymbol{\Phi}_*$.

Derivatives of the $(q-1)$-dimensional constraints function $\mathbf{z}(\boldsymbol{\theta}_*)$ in (125) are presented next for the use of the Lagrange algorithm. Theorem 21 uses the results from Theorem 16 for derivatives of $\mathbf{z}(\boldsymbol{\theta}_*)$.

Theorem 21. [Original result]. *First and second derivatives of the $(q-1)$-dimensional constraints function $\mathbf{z}(\boldsymbol{\theta}_*)$ in (125) with respect to $\boldsymbol{\theta}_{\boldsymbol{\lambda}*}$, $\boldsymbol{\theta}_{\boldsymbol{\delta}*}$, $\boldsymbol{\theta}_{\gamma*}$ and $\boldsymbol{\theta}_{\boldsymbol{\psi}*}$, evaluated at $\mathbf{G}_* = \mathbf{I}_q$, are listed as follows:*

$$\begin{aligned} \mathbf{D}^{(1)}_{\mathbf{z};\boldsymbol{\theta}'_{\lambda*}} &= \mathbf{0}_{q-1\times\nu_{*1}}, \\ \mathbf{D}^{(1)}_{\mathbf{z};\boldsymbol{\theta}'_{\delta*}} &= \mathbf{C}'\mathbf{L}'_{21,q}(\boldsymbol{\Gamma}\mathbf{G}\otimes\boldsymbol{\Gamma}\mathbf{G})\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta} = \mathbf{C}'\mathbf{L}'_{21,q}(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}, \end{aligned}$$

$$
\begin{aligned}
\mathbf{D}^{(1)}_{\mathbf{z};\boldsymbol{\theta}'_{\gamma*}} &= \mathbf{C}'\mathbf{L}'_{21,q}2\mathbf{N}_q\left(\boldsymbol{\Gamma}\mathbf{G}\boldsymbol{\Delta}\otimes\boldsymbol{\Gamma}\right)\mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}} = 2\mathbf{C}'\mathbf{L}'_{21,q}(\boldsymbol{\Gamma}\boldsymbol{\Delta}\otimes\boldsymbol{\Gamma})\mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}},\\
\mathbf{D}^{(1)}_{\mathbf{z};\boldsymbol{\theta}'_{\psi*}} &= \mathbf{0}_{q-1\times\nu_{*4}},\\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\lambda*},\boldsymbol{\theta}'_{\lambda*}} &= \mathbf{0}_{q-1\times\nu_{*1}\nu_{*1}},\\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\delta*},\boldsymbol{\theta}'_{\lambda*}} &= \mathbf{0}_{q-1\times\nu_{*1}\nu_{*2}},\\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\lambda*}} &= \mathbf{0}_{q-1\times\nu_{*1}\nu_{*3}},\\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\psi*},\boldsymbol{\theta}'_{\lambda*}} &= \mathbf{0}_{q-1\times\nu_{*1}\nu_{*4}},\\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\delta*},\boldsymbol{\theta}'_{\delta*}} &= \mathbf{C}'\mathbf{L}'_{21,q}(\boldsymbol{\Gamma}\mathbf{G}\otimes\boldsymbol{\Gamma}\mathbf{G})\mathbf{L}_{21,q}\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_{\delta},\boldsymbol{\theta}'_{\delta}} = \mathbf{C}'\mathbf{L}'_{21,q}(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\mathbf{L}_{21,q}\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_{\delta},\boldsymbol{\theta}'_{\delta}},\\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\delta*}} &= -2\mathbf{C}'\mathbf{L}'_{21,q}[\boldsymbol{\Gamma}\otimes\operatorname{vec}'(\mathbf{I}_q)\otimes\boldsymbol{\Gamma}]\left(\mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}}\otimes\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_{\delta}}\right),\\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\psi*},\boldsymbol{\theta}'_{\delta*}} &= \mathbf{0}_{q-1\times\nu_{*2}\nu_{*4}},\\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\gamma*}} &= 2\mathbf{C}'\mathbf{L}'_{21,q}(\boldsymbol{\Gamma}\boldsymbol{\Delta}\otimes\boldsymbol{\Gamma})\mathbf{D}^{(2)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\gamma*}}\\
&- 2\mathbf{C}'\mathbf{L}'_{21,q}[\boldsymbol{\Gamma}\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\boldsymbol{\Gamma}]\left(\mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}}\otimes\mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}}\right),\\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\psi*},\boldsymbol{\theta}'_{\gamma*}} &= \mathbf{0}_{q-1\times\nu_{*3}\nu_{*4}},\ \textit{and}\\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\psi*},\boldsymbol{\theta}'_{\psi*}} &= \mathbf{0}_{q-1\times\nu_{*4}\nu_{*4}},
\end{aligned}
\tag{128}
$$

where $\mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}}$ *and* $\mathbf{D}^{(2)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\gamma*}}$ *were given in Theorem 16.*

Derivatives of $\mathbf{z}(\boldsymbol{\theta}_*)$ with respect to $\boldsymbol{\theta}_*$ can be re-assembled from derivatives with respect to $\{\boldsymbol{\theta}_{*j}\}_{j=1}^4$ by using the elementary matrix, $\mathbf{E}_{i,\boldsymbol{\nu}_*}$, which is defined in Table 56.

The first derivative of $\mathbf{z}(\boldsymbol{\theta}_*)$ with respect to $\boldsymbol{\theta}_*$ is

$$
\mathbf{D}^{(1)}_{\mathbf{z};\boldsymbol{\theta}'_*} \overset{\text{def}}{=} \frac{\partial\mathbf{z}(\boldsymbol{\theta}_*)}{\partial\,\boldsymbol{\theta}'_*} = \begin{pmatrix}\mathbf{D}^{(1)}_{\mathbf{z};\boldsymbol{\theta}'_{\lambda*}} & \mathbf{D}^{(1)}_{\mathbf{z};\boldsymbol{\theta}'_{\delta*}} & \mathbf{D}^{(1)}_{\mathbf{z};\boldsymbol{\theta}'_{\gamma*}} & \mathbf{D}^{(1)}_{\mathbf{z};\boldsymbol{\theta}'_{\psi*}}\end{pmatrix} = \sum_{i=1}^{4}\mathbf{D}^{(1)}_{\mathbf{z};\boldsymbol{\theta}'_{*i}}\mathbf{E}'_{i,\boldsymbol{\nu}_*}. \tag{129}
$$

The second derivative of $\mathbf{z}(\boldsymbol{\theta}_*)$ with respect to $\boldsymbol{\theta}_*$ is

$$
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_*,\boldsymbol{\theta}'_*} \overset{\text{def}}{=} \frac{\partial^2\mathbf{z}(\boldsymbol{\theta}_*)}{\partial\,\boldsymbol{\theta}'_*\otimes\partial\,\boldsymbol{\theta}'_*} = \sum_{i=1}^{4}\sum_{j=1}^{4}\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{*i},\boldsymbol{\theta}'_{*j}}\left(\mathbf{E}_{i,\boldsymbol{\nu}_*}\otimes\mathbf{E}_{j,\boldsymbol{\nu}_*}\right)'. \tag{130}
$$

<u>3.5.2.3. LMA Based On a Parameterization of $\boldsymbol{\Phi}_*$</u>: For notational convenience, define $\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}_*}$, $\mathbf{D}^{(2)}_{F;\boldsymbol{\theta}_*,\boldsymbol{\theta}'_*}$ and $\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}'_*,\boldsymbol{\theta}'_*,\boldsymbol{\theta}_*}$ as

$$\begin{aligned} \mathbf{D}^{(1)}_{F;\boldsymbol{\theta}_*} &\stackrel{\text{def}}{=} \frac{\partial F(\boldsymbol{\Sigma},\mathbf{S})}{\partial\,\boldsymbol{\theta}_*}\Big|_{\mathbf{G}=\mathbf{I}_q}, \quad \mathbf{D}^{(2)}_{F;\boldsymbol{\theta}_*,\boldsymbol{\theta}'_*} \stackrel{\text{def}}{=} \frac{\partial^2 F(\boldsymbol{\Sigma},\mathbf{S})}{\partial\,\boldsymbol{\theta}_*\otimes\partial\,\boldsymbol{\theta}'_*}\Big|_{\mathbf{G}=\mathbf{I}_q}, \\ \text{and} \quad \mathbf{D}^{(3)}_{F;\boldsymbol{\theta}'_*,\boldsymbol{\theta}'_*,\boldsymbol{\theta}_*} &\stackrel{\text{def}}{=} \frac{\partial^3 F(\boldsymbol{\Sigma},\mathbf{S})}{\partial\,\boldsymbol{\theta}'_*\otimes\partial\,\boldsymbol{\theta}'_*\otimes\partial\,\boldsymbol{\theta}_*}\Big|_{\mathbf{G}=\mathbf{I}_q}, \end{aligned} \tag{131}$$

where $F(\boldsymbol{\Sigma},\mathbf{S})$ is given in (66). The expressions for $\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}_*}$, $\mathbf{D}^{(2)}_{F;\boldsymbol{\theta}_*,\boldsymbol{\theta}'_*}$ and $\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}'_*,\boldsymbol{\theta}'_*,\boldsymbol{\theta}_*}$ are the same with the expressions provided in Theorem 17, except that $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}$, $\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'}$ and $\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}'}$ are replaced by $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_*}$, $\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_*,\boldsymbol{\theta}'_*}$ and $\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_*,\boldsymbol{\theta}'_*,\boldsymbol{\theta}'_*}$. The expressions for $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_*}$, $\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_*,\boldsymbol{\theta}'_*}$ and $\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_*,\boldsymbol{\theta}'_*,\boldsymbol{\theta}'_*}$ were given § 2.4.2.

The following structure on Lagrange algorithm is adapted from details on Lagrange algorithm in the supplement of Boik [33]. Denote the vector of parameter estimates after i^{th} iteration of the Lagrange algorithm by $\widehat{\boldsymbol{\omega}}_i = \begin{pmatrix}\widehat{\boldsymbol{\theta}}'_{*i} & \widehat{\boldsymbol{\zeta}}'_i\end{pmatrix}'$. Define $\mathbf{D}^{(1)}_{Q;\widehat{\boldsymbol{\omega}}_i}$ and $\mathbf{D}^{(2)}_{Q;\widehat{\boldsymbol{\omega}}_i,\widehat{\boldsymbol{\omega}}'_i}$ as follows:

$$\begin{aligned} \mathbf{D}^{(1)}_{Q;\widehat{\boldsymbol{\omega}}_i} &\stackrel{\text{def}}{=} \begin{pmatrix} \mathbf{D}^{(1)}_{F;\hat{\boldsymbol{\theta}}_{*i}} - \mathbf{D}^{(1)\prime}_{\mathbf{z};\hat{\boldsymbol{\theta}}'_{*i}}\widehat{\boldsymbol{\zeta}}_i \\ \mathbf{z}(\widehat{\boldsymbol{\theta}}_*) \end{pmatrix} \quad \text{and} \\ \mathbf{D}^{(2)}_{Q;\widehat{\boldsymbol{\omega}}_i,\widehat{\boldsymbol{\omega}}'_i} &\stackrel{\text{def}}{=} \begin{pmatrix} \widehat{\mathbf{A}}_{i,11} & \widehat{\mathbf{A}}_{i,12} \\ \widehat{\mathbf{A}}_{i,21} & \widehat{\mathbf{A}}_{i,22} \end{pmatrix} = \begin{pmatrix} \mathbf{D}^{(2)}_{F;\hat{\boldsymbol{\theta}}_{*i},\hat{\boldsymbol{\theta}}'_{*i}} - (\mathbf{I}_\nu\otimes\widehat{\boldsymbol{\zeta}}'_i)\mathbf{D}^{(2)}_{\mathbf{z};\hat{\boldsymbol{\theta}}_{*i},\hat{\boldsymbol{\theta}}'_{*i}} & -\mathbf{D}^{(1)\prime}_{\mathbf{z};\hat{\boldsymbol{\theta}}'_{*i}} \\ -\mathbf{D}^{(1)}_{\mathbf{z};\hat{\boldsymbol{\theta}}'_{*i}} & \mathbf{0}_{(q-1)\times(q-1)} \end{pmatrix}, \end{aligned} \tag{132}$$

where $\mathbf{D}^{(1)}_{F;\hat{\boldsymbol{\theta}}_{*i}} = \mathbf{D}^{(1)}_{F;\boldsymbol{\theta}_*}\Big|_{\boldsymbol{\theta}_*=\hat{\boldsymbol{\theta}}_{*i}}$, $\mathbf{D}^{(2)}_{F;\hat{\boldsymbol{\theta}}_{*i},\hat{\boldsymbol{\theta}}'_{*i}} = \mathbf{D}^{(2)}_{F;\boldsymbol{\theta}_*,\boldsymbol{\theta}'_*}\Big|_{\boldsymbol{\theta}_*=\hat{\boldsymbol{\theta}}_{*i}}$ and $\mathbf{D}^{(1)}_{\mathbf{z};\hat{\boldsymbol{\theta}}'_{*i}} = \mathbf{D}^{(1)}_{\mathbf{z};\boldsymbol{\theta}'_*}\Big|_{\boldsymbol{\theta}_*=\hat{\boldsymbol{\theta}}_{*i}}$.

The Newton update of $\widehat{\boldsymbol{\omega}}_i$ at iteration $i+1$ is

$$\widehat{\boldsymbol{\omega}}_{i+1} = \widehat{\boldsymbol{\omega}}_i - \left[\mathbf{D}^{(2)}_{Q;\widehat{\boldsymbol{\omega}}_i,\widehat{\boldsymbol{\omega}}'_i}\right]^{-1}\mathbf{D}^{(1)}_{Q;\widehat{\boldsymbol{\omega}}_i}, \tag{133}$$

where $\mathbf{D}^{(1)}_{Q;\widehat{\boldsymbol{\omega}}_i}$ and $\mathbf{D}^{(2)}_{Q;\widehat{\boldsymbol{\omega}}_i,\widehat{\boldsymbol{\omega}}'_i}$ is defined in (132). Furthermore, the modified Newton update is

$$\widehat{\boldsymbol{\omega}}_{i+1} = \widehat{\boldsymbol{\omega}}_i - \alpha\left[\mathbf{D}^{(2)}_{Q;\widehat{\boldsymbol{\omega}}_i,\widehat{\boldsymbol{\omega}}'_i}\right]^{-1}\mathbf{D}^{(1)}_{Q;\widehat{\boldsymbol{\omega}}_i}. \tag{134}$$

The following theorem provides an expression for the modified Newton update of $\widehat{\boldsymbol{\omega}}_i$, which is used in the Lagrange algorithm.

Theorem 22. [Boik [33], 2011, Details on Lagrange Algorithm, Theorem 1, Page 3]. *Assuming that a solution to the Lagrange equations in (127) exists in which the Lagrange multipliers are finite and the required inverses exist, one solution (the Moore-Penrose solution) to (134) yields the following modified Newton update,*

$$\widehat{\boldsymbol{\omega}}_{i+1} = \begin{pmatrix} \hat{\boldsymbol{\theta}}_{*(i+1)} \\ \widehat{\boldsymbol{\zeta}}_{(i+1)} \end{pmatrix} = \begin{pmatrix} \widehat{\boldsymbol{\theta}}_{*i} - \alpha \mathbf{D}^{(1)+}_{\mathbf{z};\hat{\boldsymbol{\theta}}'_{*i}} \mathbf{z}(\widehat{\boldsymbol{\theta}}_{*i}) \\ \widehat{\boldsymbol{\zeta}}_i \left(1-\alpha\right) \end{pmatrix}$$

$$+\alpha \begin{pmatrix} -\widehat{\mathbf{F}}_i^{\perp} \left(\widehat{\mathbf{F}}_i^{\perp\prime} \widehat{\mathbf{A}}_{i,11} \widehat{\mathbf{F}}_i^{\perp}\right)^{-1} \widehat{\mathbf{F}}_i^{\perp\prime} \\ \mathbf{D}^{(1)+\prime}_{\mathbf{z};\hat{\boldsymbol{\theta}}'_{*i}} \left[\mathbf{I}_{\nu*} - \widehat{\mathbf{A}}_{i,11} \widehat{\mathbf{F}}_i^{\perp} \left(\widehat{\mathbf{F}}_i^{\perp\prime} \widehat{\mathbf{A}}_{i,11} \widehat{\mathbf{F}}_i^{\perp}\right)^{-1} \widehat{\mathbf{F}}_i^{\perp\prime}\right] \end{pmatrix} \left(\mathbf{D}^{(1)}_{F;\hat{\boldsymbol{\theta}}_{*i}} - \widehat{\mathbf{A}}_{i,11} \mathbf{D}^{(1)+}_{\mathbf{z};\hat{\boldsymbol{\theta}}'_{*i}} \mathbf{z}(\widehat{\boldsymbol{\theta}}_{*i})\right),$$

where $\widehat{\mathbf{U}}_i \widehat{\mathbf{D}}_i \widehat{\mathbf{F}}_i'$ *is the full-rank SVD* $\left(\mathbf{D}^{(1)}_{\mathbf{z};\hat{\boldsymbol{\theta}}'_{*i}}\right)$, $\widehat{\mathbf{U}}_i \in \mathcal{O}(q-1)$, $\widehat{\mathbf{D}}_i \in \mathcal{D}^+(q-1)$, $\widehat{\mathbf{F}}_i \in \mathcal{O}(\nu^*, (q-1))$, $q-1 = rank\left(\mathbf{D}^{(1)}_{\mathbf{z};\hat{\boldsymbol{\theta}}'_{*i}}\right)$, ν^* *is defined in (57),* $\mathbf{D}^{(1)+}_{\mathbf{z};\hat{\boldsymbol{\theta}}'_{*i}}$ *and* $\widehat{\mathbf{F}}_i^{\perp}$ *are defined in Table 56, and* $\mathcal{O}(\nu^*, (q-1))$ *is defined in Table 57.*

A proof of Theorem 22 is omitted because it is provided on page 3 in details on Lagrange algorithm of Boik [33]. The Lagrange algorithm iterative procedure consists of six steps:

Step 1 Compute initial guesses for $\widehat{\boldsymbol{\Lambda}}_*$, $\widehat{\boldsymbol{\Delta}}_*$, $\widehat{\boldsymbol{\Gamma}}_*$ and $\widehat{\boldsymbol{\Psi}}_*$. Denote these guesses by $\widehat{\boldsymbol{\Lambda}}_0 = \boldsymbol{\Lambda}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\lambda}*,0})$, $\widehat{\boldsymbol{\Delta}}_0 = \boldsymbol{\Delta}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta}*,0})$, $\widehat{\boldsymbol{\Gamma}}_0 = \boldsymbol{\Gamma}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\gamma}*,0})$ and $\widehat{\boldsymbol{\Psi}}_0 = \boldsymbol{\Psi}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\psi}*,0})$. An algorithm to obtain initial guesses is given in § 3.3.6.

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

$$\widehat{\boldsymbol{\Sigma}}_i = \widehat{\boldsymbol{\Lambda}}_i \widehat{\boldsymbol{\Gamma}}_i \widehat{\boldsymbol{\Delta}}_i \widehat{\boldsymbol{\Gamma}}_i' \widehat{\boldsymbol{\Lambda}}_i' + \widehat{\boldsymbol{\Psi}}_i.$$

Step 3 Set $\hat{\boldsymbol{\theta}}_{\gamma*,i} = \mathbf{0}$ and set $\hat{\boldsymbol{\theta}}_{*i} = \begin{pmatrix}\boldsymbol{\theta}'_{\lambda*} & \boldsymbol{\theta}'_{\delta*} & \mathbf{0}' & \boldsymbol{\theta}'_{\psi*}\end{pmatrix}'$. Use the result in Theorem 22 to update $\hat{\boldsymbol{\theta}}_{*i}$. Note that $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\gamma}*,i}$ is set to zero and updated from zero in each iteration.

Step 4 Use equation 55 in § 2.4.2 to compute $\mathbf{G}\left(\widehat{\boldsymbol{\theta}}_{\gamma*,i+1}\right)$ based on the updated $\widehat{\boldsymbol{\theta}}_{\gamma*,i+1}$ from the previous step.

Step 5 Compute $\widehat{\boldsymbol{\Lambda}}_{i+1} = \boldsymbol{\Lambda}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\lambda}*,i+1}\right)$, $\widehat{\boldsymbol{\Delta}}_{i+1} = \boldsymbol{\Delta}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta}*,i+1}\right)$, $\widehat{\boldsymbol{\Gamma}}_{i+1} = \widehat{\boldsymbol{\Gamma}}_i\mathbf{G}\left(\widehat{\boldsymbol{\theta}}_{\gamma*,i+1}\right)$ and $\widehat{\boldsymbol{\Psi}}_{i+1} = \widehat{\boldsymbol{\Psi}}\left(\widehat{\boldsymbol{\theta}}_{\psi*,i+1}\right)$ based on the updated $\hat{\boldsymbol{\theta}}_{*(i+1)}$ from Step 3.

Step 6 Iterate Steps 2-5 until convergence.

In each iteration of the Lagrange algorithm, the $(q-1)$-dimensional constraints function $\mathbf{z}(\boldsymbol{\theta}_*) = \mathbf{0}$ is not necessarily satisfied. Therefore, the updated estimate for $\boldsymbol{\Phi}_*$ in each iteration is not necessarily a correlation matrix. However, the estimate for $\boldsymbol{\Phi}_*$ satisfies properties of a correlation matrix when convergence occurs.

3.6. The 4-Step Procedure

The purpose of the 4-step procedure is to obtain optimal proper estimates for $\boldsymbol{\Lambda}$, $\boldsymbol{\Delta}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ that minimize $F(\boldsymbol{\Sigma}, \mathbf{S})$ in (66) with no program failure. The 4-Step Procedure consists of the following 4 steps:

Step 1. Compute initial guesses $\widehat{\boldsymbol{\Lambda}}_*$, $\widehat{\boldsymbol{\Phi}}_*$ and $\widehat{\boldsymbol{\Psi}}$. The strategy on how to obtain those initial guesses can be found in § 3.3.2, § 3.3.1 and § 3.3.5.

Step 2. Apply § 3.5.1: Modified Fisher Scoring Algorithm Based On a Covariance Parameterization of $\boldsymbol{\Phi}_*$ using $\widehat{\boldsymbol{\Lambda}}_*$, $\widehat{\boldsymbol{\Phi}}_*$ and $\widehat{\boldsymbol{\Psi}}$.

Step 3. Use estimates for $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ obtained through Step 2 to compute initial guesses $\widehat{\boldsymbol{\Lambda}}$, $\widehat{\boldsymbol{\Delta}}$, $\widehat{\boldsymbol{\Gamma}}$ and $\widehat{\boldsymbol{\Psi}}$. The details on how to obtain those initial guesses can be found in § 3.3.4, § 3.3.6, § 3.3.7 and § 3.3.5.

Step 4. Apply § 3.5.2: Lagrange Multipliers Algorithm Based On a Parameterization of $\boldsymbol{\Phi}_*$ using $\widehat{\boldsymbol{\Lambda}}$, $\widehat{\boldsymbol{\Delta}}$, $\widehat{\boldsymbol{\Gamma}}$ and $\widehat{\boldsymbol{\Psi}}$. Accordingly, the optimal estimates for $\boldsymbol{\Lambda}$, $\boldsymbol{\Delta}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ that minimizes $F(\boldsymbol{\Sigma}, \mathbf{S})$ in (66) can be obtained.

In Step 2, the estimate for $\boldsymbol{\Phi}_*$, $\widehat{\boldsymbol{\Phi}}_*$, is non-negative definite, but not necessarily positive definite. This means that one of the eigenvalues of $\widehat{\boldsymbol{\Phi}}_*$ can be a boundary estimate, where a boundary estimate is defined in § 1.5. Therefore, it is necessary to proceed to Step 3 and Step 4, where $\boldsymbol{\Delta}$ is parameterized in order to avoid improper eigenvalue estimates. Parameterizations of $\boldsymbol{\Delta}$ are provided in § 2.3.2. When Step 4 is finished, all the estimates for $\boldsymbol{\Lambda}$, $\boldsymbol{\Delta}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Psi}$ are proper. Accordingly, the estimate for the factor correlation matrix $\boldsymbol{\Phi} = \boldsymbol{\Gamma}\boldsymbol{\Delta}\boldsymbol{\Gamma}'$ is proper as well.

CHAPTER 4

ASYMPTOTIC DISTRIBUTIONS OF ESTIMATORS

In this chapter, asymptotic distributions of estimators under the assumption of finite fourth-order moments are given. As a special case, simplifications are given under normality. First-order accurate confidence intervals and hypothesis tests are conducted.

4.1. Theoretical Background

In order to establish asymptotic distributions of estimators, related theorems are established first.

Theorem 23. [Slutsky's Theorem]. *Let $\mathbf{t}_n$ be a random $p \times 1$ vector that satisfies $\mathbf{t}_n \xrightarrow{\text{dist}} \mathbf{t}$ as $n \to \infty$, where $\mathbf{t}$ is a random $p \times 1$ vector. Suppose that $\mathbf{a}_n$ is a random $k \times 1$ vector and that $\mathbf{B}_n$ is a random $k \times p$ matrix that satisfy $\mathbf{a}_n \xrightarrow{\text{prob}} \mathbf{a}$ and $\mathbf{B}_n \xrightarrow{\text{prob}} \mathbf{B}$, where $\mathbf{a}$ is a $k \times 1$ vector of constants and $\mathbf{B}$ is a $k \times p$ matrix of constants. Then*

1. $\mathbf{a}_n + \mathbf{B}_n \mathbf{t}_n \xrightarrow{\text{dist}} \mathbf{a} + \mathbf{B}\mathbf{t}$

2. $\mathbf{a}_n + \mathbf{B}_n^{-1} \mathbf{t}_n \xrightarrow{\text{dist}} \mathbf{a} + \mathbf{B}^{-1}\mathbf{t}$, *if $k = p$ and $\mathbf{B}^{-1}$ exists.*

A proof of Theorem 23 is given on page 127 of Sen and Singer [41]. The following delta method can be used to obtain asymptotic distributions for differentiable functions of asymptotically normal statistics.

Theorem 24. [A Well-Known Result].

(1.) If g is a continuous function and $\mathbf{t}_n \xrightarrow{\text{dist}} \mathbf{t}$, then $g(\mathbf{t}_n) \xrightarrow{\text{dist}} g(\mathbf{t})$.

(2.) If g is a continuous function and $\mathbf{t}_n \xrightarrow{\text{prob}} \mathbf{t}$, then $g(\mathbf{t}_n) \xrightarrow{\text{prob}} g(\mathbf{t})$.

A proof of Theorem 24 is given on page 124 of Rao [42].

Theorem 25. [Delta Method]. *Let* $\mathbf{t}_n$ *be a random* $p \times 1$ *vector with asymptotic distribution* $\sqrt{n}(\mathbf{t}_n - \boldsymbol{\theta}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \boldsymbol{\Omega}_{\mathbf{t}})$. *Suppose that* $\mathbf{f}(\mathbf{t}_n)$ *is a vector-valued differentiable function. Then*

$$\sqrt{n}\left[\mathbf{f}(\mathbf{t}_n) - \mathbf{f}(\boldsymbol{\theta})\right] \xrightarrow{\text{dist}} \mathbf{N}\left[\mathbf{0}, \mathbf{D}(\boldsymbol{\theta})\boldsymbol{\Omega}_{\mathbf{t}}\mathbf{D}(\boldsymbol{\theta})'\right],$$

where $\mathbf{D}(\boldsymbol{\theta}) = \left.\frac{\partial\, \mathbf{f}(\mathbf{t}_n)}{\partial\, \mathbf{t}_n'}\right|_{\mathbf{t}_n = \boldsymbol{\theta}}$.

A proof of Theorem 25 is given on page 136 of Sen and Singer [41].

4.2. Asymptotic Distributions of Estimators

Recall in § 3.2, the model for $\mathbf{Y}$, (62), is

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{F}\boldsymbol{\Lambda}' + \mathbf{E},$$

where $\mathbf{X}$ is an $N \times d$ known matrix of constants, $\text{rank}(\mathbf{X}) = r_x$, $\mathbf{B}$ is a $d \times p$ matrix of fixed regression coefficients, $\mathbf{F}$ is an $N \times q$ matrix of random latent factor scores whose i^{th} row is $\mathbf{f}_i'$, and $\mathbf{E}$ is an $N \times p$ matrix of random errors whose i^{th} row is $\boldsymbol{\epsilon}_i'$. It is assumed that $\mathbf{y}_i$ in $\mathbf{Y}$, for $i = 1, 2, \ldots, N$, are independent, and $r_x = O(1)$.

Theorem 26. [A Well-Known Result]. *Let* $\mathbf{H}_{\mathbf{x}}$ *be the projection operator that projects onto* $\mathcal{R}(\mathbf{X})$ *along* $\mathcal{N}(\mathbf{X}')$, *that is,* $\mathbf{H}_{\mathbf{x}} = \text{ppo}(\mathbf{X}) = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'$, *where* $\text{ppo}(\cdot)$ *is defined in Table 56. Then,* $\mathbf{H}_{\mathbf{x}}\mathbf{X} = \mathbf{X}$, $\mathbf{H}_{\mathbf{x}} = \mathbf{H}_x'$ *and* $\text{rank}(\mathbf{H}_{\mathbf{x}}) = \text{tr}(\mathbf{H}_{\mathbf{x}})$.

A proof of Theorem 26 is given on page 386 and page 433 of Meyer [43] and page 370 of Schott [44]. By Theorem 26, it is readily shown that $\text{rank}(\mathbf{H}_{\mathbf{x}}) = \text{rank}(\mathbf{X})$ because $\text{rank}(\mathbf{X}) = \text{rank}(\mathbf{H}_{\mathbf{x}}\mathbf{X}) \leq \text{rank}(\mathbf{H}_{\mathbf{x}}) \leq \text{rank}(\mathbf{X})$.

Theorem 27. [Asymptotic Distribution of $\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})$][Muirhead [45], 1982, Theorem 1.2.17, Page 19]. *The model* $\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{F}\boldsymbol{\Lambda}' + \mathbf{E}$ *can be rewritten as*

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{Z}, \tag{135}$$

where $\mathbf{Z} = \mathbf{F}\mathbf{\Lambda}' + \mathbf{E}$ *is an* $N \times p$ *matrix of random errors with* $E[\mathbf{Z}] = \mathbf{0}_{N\times p}$ *and* $Var(\operatorname{vec}\mathbf{Z}) = \mathbf{\Sigma}\otimes\mathbf{I}_N$. *Denote the* i^{th} *row of* $\mathbf{Z}$ *by* $\mathbf{z}_i'$ *and assume that* $\mathbf{z}_i \overset{\text{iid}}{\sim} (\mathbf{0}, \mathbf{\Sigma})$ *with finite fourth moments, where* $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi}$.

The asymptotic distribution of $\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})$ *is*

$$\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \mathbf{\Omega}), \tag{136}$$

where $\mathbf{s} = \operatorname{vec}\mathbf{S}$, $\boldsymbol{\sigma} = \operatorname{vec}\mathbf{\Sigma}$, $\mathbf{\Omega} = \mathrm{E}\left[\mathbf{z}_i\mathbf{z}_i' \otimes \mathbf{z}_i\mathbf{z}_i'\right] - \boldsymbol{\sigma}\boldsymbol{\sigma}'$, *and* $\mathbf{S}$ *is the* $p \times p$ *sample covariance matrix.*

This assumption made for $\mathbf{z}_i$ in Theorem 27 holds for the remainder of the thesis, unless otherwise specified. Denote the diagonals of $\mathbf{I}_N - \mathbf{H}_{\mathbf{x}}$ by $\{c_{ii}\}_{i=1}^{N}$, $c_1 \overset{\text{def}}{=} \sum_{i=1}^{N} c_{ii}^2$, and denote the entries in $\mathbf{H}_{\mathbf{x}}$ by $\{h_{ij}\}_{i=1}^{N}{}_{j=1}^{N}$. Examine h_{ii} as follows:

$$\begin{aligned}
h_{ii} &= \mathbf{e}_i^{N\prime}\mathbf{H}_{\mathbf{x}}\mathbf{e}_i^N = \mathbf{e}_i^{N\prime}\mathbf{H}_{\mathbf{x}}'\mathbf{H}_{\mathbf{x}}\mathbf{e}_i^N \quad \text{because} \quad \mathbf{H}_{\mathbf{x}} = \mathbf{H}_{\mathbf{x}}'\mathbf{H}_{\mathbf{x}} \\
&= \mathbf{h}_i'\mathbf{h}_i \quad \text{where} \quad \mathbf{h}_i = \mathbf{H}_{\mathbf{x}}\mathbf{e}_i^N \\
&= \sum_{j=1}^{N} h_{ji}^2 = h_{ii}^2 + \sum_{j\neq i}^{N} h_{ji}^2.
\end{aligned} \tag{137}$$

Further, (137) implies that

$$h_{ii}^2 \leq h_{ii}, \quad \sum_{i=1}^{N} h_{ii}^2 \leq \sum_{i=1}^{N} h_{ii}, \quad \text{and} \quad \sum_{i=1}^{N} h_{ii}^2 = O(1), \tag{138}$$

because $\sum_{i=1}^{N} h_{ii} = \operatorname{tr}(\mathbf{H}_{\mathbf{x}})$ and $\operatorname{tr}(\mathbf{H}_{\mathbf{x}}) = \operatorname{rank}(\mathbf{H}_{\mathbf{x}}) = r_x = O(1)$ by Theorem 26.

Examine c_1 as follows:

$$\begin{aligned}
c_1 &= \sum_{i=1}^{N} c_{ii}^2 = \sum_{i=1}^{N} \left[\mathbf{e}_i^{N\prime}\left(\mathbf{I}_N - \mathbf{H}_{\mathbf{x}}\right)\mathbf{e}_i^N\right]^2 \\
&= \sum_{i=1}^{N} \left(1 - \mathbf{e}_i^{N\prime}\mathbf{H}_{\mathbf{x}}\mathbf{e}_i^N\right)^2 = \sum_{i=1}^{N} \left[1 - 2\mathbf{e}_i^{N\prime}\mathbf{H}_{\mathbf{x}}\mathbf{e}_i^N + \left(\mathbf{e}_i^{N\prime}\mathbf{H}_{\mathbf{x}}\mathbf{e}_i^N\right)^2\right] \\
&= N - 2\operatorname{tr}(\mathbf{H}_{\mathbf{x}}) + \sum_{i=1}^{N} h_{ii}^2 = N - 2r_x + \sum_{i=1}^{N} h_{ii}^2 \\
&= N + O(1) + O(1) = n + O(1),
\end{aligned} \tag{139}$$

because $\operatorname{tr}(\mathbf{H_x}) = \operatorname{rank}(\mathbf{H_x}) = r_x$, $N = n + r_x$, $r_x = O(1)$ and (138).

It can be concluded from (139) that

$$\lim_{n\to\infty}\left(\frac{c_1}{n}\right) = \lim_{n\to\infty}\left[\frac{n+O(1)}{n}\right] = 1. \tag{140}$$

From equation 12 of Boik [46], the variance of $\sqrt{n}(\operatorname{vec}\mathbf{S} - \operatorname{vec}\boldsymbol{\Sigma})$ is

$$\boldsymbol{\Omega}_n \stackrel{\text{def}}{=} \operatorname{Var}\left[\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})\right] = \frac{c_1}{n}\left[\operatorname{E}\left(\mathbf{z}_i\mathbf{z}_i' \otimes \mathbf{z}_i\mathbf{z}_i'\right) - \boldsymbol{\sigma}\boldsymbol{\sigma}'\right] + \left(1-\frac{c_1}{n}\right)2\mathbf{N}_p(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}), \tag{141}$$

where $\mathbf{s} = \operatorname{vec}\mathbf{S}$ and $\boldsymbol{\sigma} = \operatorname{vec}\boldsymbol{\Sigma}$.

It can be shown that $\boldsymbol{\Omega}$ in Theorem 27 can be obtained through $\boldsymbol{\Omega}_n$ as follows:

$$\begin{aligned}
\lim_{n\to\infty}\boldsymbol{\Omega}_n &= \lim_{n\to\infty}\left\{\frac{c_1}{n}\left[\operatorname{E}\left(\mathbf{z}_i\mathbf{z}_i' \otimes \mathbf{z}_i\mathbf{z}_i'\right) - \boldsymbol{\sigma}\boldsymbol{\sigma}'\right] + \left(1-\frac{c_1}{n}\right)2\mathbf{N}_p(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\right\} \\
&= \left[\operatorname{E}\left(\mathbf{z}_i\mathbf{z}_i' \otimes \mathbf{z}_i\mathbf{z}_i'\right) - \boldsymbol{\sigma}\boldsymbol{\sigma}'\right]\lim_{n\to\infty}\left(\frac{c_1}{n}\right) + 2\mathbf{N}_p(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\lim_{n\to\infty}\left(1-\frac{c_1}{n}\right) \\
&= \boldsymbol{\Omega} \quad \text{because of (140).}
\end{aligned}$$

Let $\widehat{\boldsymbol{\Omega}}_n$ be a consistent estimator of $\boldsymbol{\Omega}_n$. That is, $\widehat{\boldsymbol{\Omega}}_n$ and $\boldsymbol{\Omega}_n$ are asymptotically equal in distribution, or, $\widehat{\boldsymbol{\Omega}}_n - \boldsymbol{\Omega}_n = o_p(1)$. Equation 10 in Boik, Panishkan, and Hyde [35] provides an expression for $\widehat{\boldsymbol{\Omega}}_n$, that is,

$$\widehat{\boldsymbol{\Omega}}_n = \frac{N}{n}\left[\frac{1}{n}\sum_{i=1}^{N}\left(\hat{\mathbf{z}}_i\hat{\mathbf{z}}_i' \otimes \hat{\mathbf{z}}_i\hat{\mathbf{z}}_i'\right)\right] - \mathbf{s}\mathbf{s}' + (\mathbf{s}-\widehat{\boldsymbol{\sigma}})(\mathbf{s}-\widehat{\boldsymbol{\sigma}})', \tag{142}$$

where $\hat{\mathbf{z}}_i = \mathbf{Y}'(\mathbf{I}_N - \mathbf{H_x})\mathbf{e}_i^N$, $\mathbf{Y}$ is given in (62), $\mathbf{e}_i^N$ is defined in Table 56, $\mathbf{H_x} = \operatorname{ppo}(\mathbf{X})$, $\widehat{\boldsymbol{\sigma}} = \operatorname{vec}\widehat{\boldsymbol{\Sigma}}$, and $\widehat{\boldsymbol{\Sigma}}$ minimizes $F(\boldsymbol{\Sigma},\mathbf{S})$ in (66).

Corollary 27.1. [Asymptotic Distribution of $\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})$ Under Normality][Magnus and Neudecker [47], 1979, Corollary 4.2, Page 394 and Muirhead [45], 1982, Corollary 1.2.18, Page 19]. *Assume that* $\mathbf{z}_i \stackrel{\text{iid}}{\sim} \mathbf{N}(\mathbf{0},\boldsymbol{\Sigma})$, *where* $\mathbf{z}_i'$ *is the* i^{th} *row of* $\mathbf{Z}$ *in (135). The asymptotic distribution of* $\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})$ *is*

$$\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0},\boldsymbol{\Omega}). \tag{143}$$

where $\mathbf{s} = \operatorname{vec}\mathbf{S}$, $\boldsymbol{\sigma} = \operatorname{vec}\boldsymbol{\Sigma}$ *and* $\boldsymbol{\Omega} = 2\mathbf{N}_p\left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}\right)$.

Theorem 28. [Asymptotic Distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$][Browne [48], 1974, Proposition 6, Page 13 and Browne [49], 1984, Proposition 2, Page 67]. *Let $\hat{\boldsymbol{\theta}}$ be the minimum discrepancy estimator of $\boldsymbol{\theta}$ in $F(\boldsymbol{\Sigma}(\boldsymbol{\theta}),\mathbf{S})$, where $F(\boldsymbol{\Sigma}(\boldsymbol{\theta}),\mathbf{S})$ is given in (66). Assume that $\hat{\boldsymbol{\theta}}$ is a consistent root of $F(\boldsymbol{\Sigma}(\boldsymbol{\theta}),\mathbf{S})$ and $\operatorname{rank}(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'})=\nu$, where $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}$ is given in (48). The asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ is*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0},\boldsymbol{\Omega}_{\boldsymbol{\theta}}),$$

where $\boldsymbol{\Omega}_{\boldsymbol{\theta}} = \mathbf{A}'_{\boldsymbol{\theta}}\boldsymbol{\Omega}\mathbf{A}_{\boldsymbol{\theta}}$, $\mathbf{A}_{\boldsymbol{\theta}} = (\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\,\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\,\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]^{-1}$, *and* $\boldsymbol{\Omega} = \mathrm{E}\left[\mathbf{z}_i\mathbf{z}_i'\otimes\mathbf{z}_i\mathbf{z}_i'\right]-\boldsymbol{\sigma}\boldsymbol{\sigma}'$ *is given in Theorem 27.*

The assumptions made in Theorem 28 hold for the remainder of the thesis, unless otherwise specified. Define $\boldsymbol{\Omega}_{\boldsymbol{\theta},n}$ as

$$\boldsymbol{\Omega}_{\boldsymbol{\theta},n} \stackrel{\text{def}}{=} \mathbf{A}'_{\boldsymbol{\theta}}\boldsymbol{\Omega}_n\mathbf{A}_{\boldsymbol{\theta}}, \tag{144}$$

where $\boldsymbol{\Omega}_n$ is defined in (141), and $\mathbf{A}_{\boldsymbol{\theta}}$ is defined in Theorem 28. Note that $\boldsymbol{\Omega}_{\boldsymbol{\theta}}$ in Theorem 28 is $\boldsymbol{\Omega}_{\boldsymbol{\theta}} = \lim_{n\to\infty}\boldsymbol{\Omega}_{\boldsymbol{\theta},n}$.

Corollary 28.1. [Asymptotic Distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ Under Normality][Browne [49], 1984, Proposition 5, Page 76]. *Assume that $\mathbf{z}_i \stackrel{\text{iid}}{\sim} \mathbf{N}(\mathbf{0},\boldsymbol{\Sigma})$, where $\mathbf{z}_i'$ is the i^{th} row of $\mathbf{Z}$ in (135). Then the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ is*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0},\boldsymbol{\Omega}_{\boldsymbol{\theta}}),$$

where $\boldsymbol{\Omega}_{\boldsymbol{\theta}} = 2\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\,\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]^{-1}$.

Recall that in Chapter 2, $\boldsymbol{\lambda}$, $\boldsymbol{\delta}$, and $\boldsymbol{\psi}$ are defined as $\boldsymbol{\lambda}=\operatorname{vec}(\boldsymbol{\Lambda})$, $\boldsymbol{\delta}=\operatorname{diag}(\boldsymbol{\Delta})$ and $\boldsymbol{\psi}=\operatorname{diag}(\boldsymbol{\Psi})$, respectively. Let $\widehat{\boldsymbol{\lambda}}$, $\widehat{\boldsymbol{\delta}}$, $\widehat{\boldsymbol{\Gamma}}$ and $\widehat{\boldsymbol{\psi}}$ be estimators of $\boldsymbol{\lambda}$, $\boldsymbol{\delta}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\psi}$, respectively, that minimize $F(\boldsymbol{\Sigma}(\boldsymbol{\theta}),\mathbf{S})$, where $F(\boldsymbol{\Sigma}(\boldsymbol{\theta}),\mathbf{S})$ is given in (66). Accordingly, asymptotic distributions of $\sqrt{n}(\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda})$, $\sqrt{n}(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta})$, $\sqrt{n}\operatorname{vec}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})$ and $\sqrt{n}(\widehat{\boldsymbol{\psi}}-\boldsymbol{\psi})$ are presented next.

Theorem 29. [Asymptotic Distribution of $\sqrt{n}(\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda})$][Original result]. *The asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda})$ is*

$$\sqrt{n}(\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \mathbf{D}^{(1)}_{\boldsymbol{\lambda};\boldsymbol{\theta}'_\lambda}\mathbf{E}'_{1,\nu}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\mathbf{E}_{1,\nu}\mathbf{D}^{(1)'}_{\boldsymbol{\lambda};\boldsymbol{\theta}'_\lambda}),$$

where $\mathbf{D}^{(1)}_{\boldsymbol{\lambda};\boldsymbol{\theta}'_\lambda}$ is given in (24), $\mathbf{E}_{1,\nu_1}$ is the 1^{st} submatrix of
$\mathbf{I}_\nu = (\mathbf{E}_{1,\nu_1}\ \ \mathbf{E}_{2,\nu_2}\ \ \mathbf{E}_{3,\nu_3}\ \ \mathbf{E}_{4,\nu_4})$, $\dim(\mathbf{E}_{j,\nu_j}) = \nu\times\nu_j$ *and $\boldsymbol{\Omega}_{\boldsymbol{\theta}}$ is given in Theorem 28.*

Theorem 30. [Asymptotic Distribution of $\sqrt{n}(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta})$][Original result]. *The asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta})$ is*

$$\sqrt{n}(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\mathbf{E}'_{2,\nu}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\mathbf{E}_{2,\nu}\mathbf{D}^{(1)'}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}),$$

where $\mathbf{E}_{2,\nu_2}$ is the 2^{nd} submatrix of $\mathbf{I}_\nu = (\mathbf{E}_{1,\nu_1}\ \ \mathbf{E}_{2,\nu_2}\ \ \mathbf{E}_{3,\nu_3}\ \ \mathbf{E}_{4,\nu_4})$, $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}$ *is given in Theorem 2,* $\dim(\mathbf{E}_{j,\nu_j}) = \nu\times\nu_j$ *and $\boldsymbol{\Omega}_{\boldsymbol{\theta}}$ is given in Theorem 28.*

Theorem 31. [Asymptotic Distribution of $\sqrt{n}\,\mathrm{vec}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})$][Original result]. *The asymptotic distribution of $\sqrt{n}\,\mathrm{vec}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})$ is*

$$\sqrt{n}\,\mathrm{vec}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \mathbf{A}_{\boldsymbol{\theta}_{\boldsymbol{\delta\gamma}}}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\mathbf{A}'_{\boldsymbol{\theta}_{\boldsymbol{\delta\gamma}}}),$$

where $\mathbf{A}_{\boldsymbol{\theta}_{\boldsymbol{\delta\gamma}}} = (\mathbf{I}_q\otimes\boldsymbol{\Gamma})\left(\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\mathbf{E}'_{2,\nu_2}+\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\mathbf{E}'_{3,\nu_3}\right)$, $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}$ *is given in Theorem 2,* $\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}$ *is given in Theorem 9, $\mathbf{E}_{3,\nu_3}$ is the 3^{rd} submatrix of* $\mathbf{I}_\nu = (\mathbf{E}_{1,\nu_1}\ \ \mathbf{E}_{2,\nu_2}\ \ \mathbf{E}_{3,\nu_3}\ \ \mathbf{E}_{4,\nu_4})$, $\dim(\mathbf{E}_{j,\nu_j}) = \nu\times\nu_j$, *and $\boldsymbol{\Omega}_{\boldsymbol{\theta}}$ is given in Theorem 28.*

Theorem 32. [Asymptotic Distribution of $\sqrt{n}(\widehat{\boldsymbol{\psi}}-\boldsymbol{\psi})$][Original result]. *The asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\psi}}-\boldsymbol{\psi})$ is*

$$\sqrt{n}(\widehat{\boldsymbol{\psi}}-\boldsymbol{\psi}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \mathbf{D}^{(1)}_{\boldsymbol{\psi};\boldsymbol{\theta}'_\psi}\mathbf{E}'_{4,\nu}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\mathbf{E}_{4,\nu}\mathbf{D}^{(1)'}_{\boldsymbol{\psi};\boldsymbol{\theta}'_\psi}),$$

where $\mathbf{E}_{4,\nu_1}$ is the 4^{th} submatrix of $\mathbf{I}_\nu = (\mathbf{E}_{1,\nu_1}\ \ \mathbf{E}_{2,\nu_2}\ \ \mathbf{E}_{3,\nu_3}\ \ \mathbf{E}_{4,\nu_4})$, $\mathbf{D}^{(1)}_{\boldsymbol{\psi};\boldsymbol{\theta}'_\psi}$ *is given in Theorem 11,* $\dim(\mathbf{E}_{j,\nu_j}) = \nu\times\nu_j$ *and $\boldsymbol{\Omega}_{\boldsymbol{\theta}}$ is given in Theorem 28.*

4.3. Confidence Intervals for Functions of Parameters

Define $\boldsymbol{\tau}$, $\hat{\boldsymbol{\tau}}$, $\mathbf{D}^{(1)}_{\boldsymbol{\tau};\boldsymbol{\theta}'}$, $\mathbf{D}^{(1)}_{\boldsymbol{\tau};\hat{\boldsymbol{\theta}}'}$, $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}'}$, $\mathbf{A}_{\hat{\boldsymbol{\theta}}}$ and $\widehat{\boldsymbol{\Omega}}_{\boldsymbol{\theta},n}$ as

$$\begin{aligned}
&\boldsymbol{\tau} \stackrel{\text{def}}{=} \boldsymbol{\tau}(\boldsymbol{\theta}), \quad \hat{\boldsymbol{\tau}} \stackrel{\text{def}}{=} \boldsymbol{\tau}(\hat{\boldsymbol{\theta}}), \quad \mathbf{D}^{(1)}_{\boldsymbol{\tau};\boldsymbol{\theta}'} \stackrel{\text{def}}{=} \frac{\partial \boldsymbol{\tau}}{\partial\, \boldsymbol{\theta}'}, \quad \mathbf{D}^{(1)}_{\boldsymbol{\tau};\hat{\boldsymbol{\theta}}'} \stackrel{\text{def}}{=} \left.\frac{\partial \boldsymbol{\tau}}{\partial\, \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \\
&\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}'} \stackrel{\text{def}}{=} \left.\frac{\partial \boldsymbol{\sigma}}{\partial\, \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad \mathbf{A}_{\hat{\boldsymbol{\theta}}} \stackrel{\text{def}}{=} \left(\widehat{\boldsymbol{\Sigma}}^{-1} \otimes \widehat{\boldsymbol{\Sigma}}^{-1}\right) \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}'} \left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}'} \left(\widehat{\boldsymbol{\Sigma}}^{-1} \otimes \widehat{\boldsymbol{\Sigma}}^{-1}\right) \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}'}\right]^{-1},
\end{aligned} \tag{145}$$

and $\widehat{\boldsymbol{\Omega}}_{\boldsymbol{\theta},n} \stackrel{\text{def}}{=} \mathbf{A}'_{\hat{\boldsymbol{\theta}}} \widehat{\boldsymbol{\Omega}}_n \mathbf{A}_{\hat{\boldsymbol{\theta}}}$, where $\boldsymbol{\tau}(\boldsymbol{\theta})$ is a $b-$vector of functions of parameters, $\hat{\boldsymbol{\theta}}$ is the minimum discrepancy estimator and a consistent root of $F(\boldsymbol{\Sigma}(\boldsymbol{\theta}), \mathbf{S})$, $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Omega}}_n$ are given in (142). Assume that derivatives of $\boldsymbol{\tau}(\boldsymbol{\theta})$ up to order 3 exist and are continuous in an open neighborhood of $\boldsymbol{\theta}$.

Suppose that $\tau = \tau(\boldsymbol{\theta})$ in (145) is a scalar-valued function. Define T_τ as

$$T_\tau \stackrel{\text{def}}{=} \sqrt{n}(\hat{\tau} - \tau)/\hat{\sigma}_{\boldsymbol{\tau}}, \tag{146}$$

where $\hat{\sigma}^2_{\boldsymbol{\tau}} = \mathbf{D}^{(1)}_{\boldsymbol{\tau};\hat{\boldsymbol{\theta}}'} \widehat{\boldsymbol{\Omega}}_{\boldsymbol{\theta},n} \mathbf{D}^{(1)\prime}_{\boldsymbol{\tau};\hat{\boldsymbol{\theta}}'}$ and $\widehat{\boldsymbol{\Omega}}_{\boldsymbol{\theta},n}$ is defined in (145). According to Theorem 6 on page 127 of Boik, Panishkan, and Hyde [35], it can be concluded that

$$\begin{aligned}
&T_\tau \xrightarrow{\text{dist}} \mathbf{N}(0,1) \quad \text{as} \quad N \to \infty, \quad \text{where} \quad T_\tau \quad \text{is defined in} \quad (146), \\
&F_{T_\tau}(t) \stackrel{\text{def}}{=} P(T_\tau \le t) = \Phi(t) + O(n^{-1/2}), \\
&f_{T_\tau}(t) \stackrel{\text{def}}{=} \frac{d}{dt} F_{T_\tau}(t) = P(T_\tau \le t) = \phi(t) + O(n^{-1/2}),
\end{aligned} \tag{147}$$

where $\Phi(t)$ and $\phi(t)$ are the standard normal cumulative distribution and probability density functions. That is,

$$P\left[\frac{\sqrt{n}(\hat{\tau} - \tau)}{\hat{\sigma}_{\boldsymbol{\tau}}} \le z_\alpha\right] = \alpha + O(n^{-1/2}), \tag{148}$$

where z_α and $z_{1-\alpha}$ are the $100(\alpha)$ and the $100(1-\alpha)$ percentiles of the standard normal distribution, respectively. It can be derived from (148) that

$$\begin{aligned}
&\text{(i)} \quad P\left[\frac{\sqrt{n}(\hat{\tau}-\tau)}{\hat{\sigma}_{\boldsymbol{\tau}}} \geq z_\alpha\right] = P\left[\tau \leq (\hat{\tau} - z_\alpha \hat{\sigma}_{\boldsymbol{\tau}}/\sqrt{n})\right] = 1-\alpha+O(n^{-1/2}), \quad \text{and} \\
&\text{(ii)} \quad P\left[\frac{\sqrt{n}(\hat{\tau}-\tau)}{\hat{\sigma}_{\boldsymbol{\tau}}} \leq z_{1-\alpha}\right] = P\left[\tau \geq (\hat{\tau} - z_{1-\alpha} \hat{\sigma}_{\boldsymbol{\tau}}/\sqrt{n})\right] = 1-\alpha+O(n^{-1/2}).
\end{aligned} \tag{149}$$

Further,

$$\begin{aligned}
&P\left[(\hat{\tau} - z_{1-\alpha/2}\hat{\sigma}_{\boldsymbol{\tau}}/\sqrt{n}) \leq \tau \leq (\hat{\tau} - z_{\alpha/2}\hat{\sigma}_{\boldsymbol{\tau}}/\sqrt{n})\right] \\
&= P\left[\tau \leq (\hat{\tau} - z_{\alpha/2}\hat{\sigma}_{\boldsymbol{\tau}}/\sqrt{n})\right] - P\left[\tau \leq (\hat{\tau} - z_{1-\alpha/2}\hat{\sigma}_{\boldsymbol{\tau}}/\sqrt{n})\right] \\
&= \left[1-\alpha/2+O(n^{-1/2})\right] - \left[\alpha/2+O(n^{-1/2})\right] \quad \text{because of} \quad (149(i)) \\
&= 1-\alpha+O(n^{-1}).
\end{aligned} \tag{150}$$

Based on (149) and (150), the $100(1-\alpha)\%$ confidence intervals for τ are as follows:

$$\begin{aligned}
&\text{(a)} \quad (-\infty, \quad \hat{\tau} - z_\alpha\hat{\sigma}_{\boldsymbol{\tau}}/\sqrt{n}), \quad \text{(b)} \quad (\hat{\tau} - z_{1-\alpha}\hat{\sigma}_{\boldsymbol{\tau}}/\sqrt{n}, \quad \infty), \quad \text{and} \\
&\text{(c)} \quad (\hat{\tau} - z_{1-\alpha/2}\hat{\sigma}_{\boldsymbol{\tau}}/\sqrt{n}, \quad \hat{\tau} - z_{\alpha/2}\hat{\sigma}_{\boldsymbol{\tau}}/\sqrt{n}).
\end{aligned} \tag{151}$$

A definition for the m^{th} order accurate confidence interval (L, U) for τ is as follows:

$$P(L < \tau < U) = 1-\alpha+O(n^{-m/2}), \tag{152}$$

where the nominal probability is $1-\alpha$. Accordingly, the confidence intervals in (a) and (b) of (151) are first-order accurate and the one in (c) of (151) is second-order accurate.

4.4. Discrepancy Test

Suppose a test of $\mathbf{H}_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ against $\mathbf{H}_a : \boldsymbol{\theta} \in \boldsymbol{\Theta}_a$ is desired, where $\boldsymbol{\Theta}_0 \cap \boldsymbol{\Theta}_a = \emptyset$. Define $\boldsymbol{\Theta}$, ν_0, ν_a, $\hat{\boldsymbol{\theta}}_0$, $\hat{\boldsymbol{\theta}}_a$, $\hat{\boldsymbol{\Sigma}}_0$, $\hat{\boldsymbol{\Sigma}}_a$, and X^2 as follows:

$$\begin{aligned}
&\boldsymbol{\Theta} = \boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_a, \quad \nu_0 = \dim(\boldsymbol{\Theta}_0), \quad \nu_a = \dim(\boldsymbol{\Theta}_a), \\
&\hat{\boldsymbol{\theta}}_0 = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}_0}{\operatorname{argmin}}\ F(\boldsymbol{\Sigma}, \mathbf{S}), \quad \hat{\boldsymbol{\theta}}_a = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmin}}\ F(\boldsymbol{\Sigma}, \mathbf{S}), \quad \hat{\boldsymbol{\Sigma}}_0 = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_0), \quad \hat{\boldsymbol{\Sigma}}_a = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_a), \\
&\text{and} \quad X^2 = n\left[F(\hat{\boldsymbol{\Sigma}}_0, \mathbf{S}) - F(\hat{\boldsymbol{\Sigma}}_a, \mathbf{S})\right],
\end{aligned} \tag{153}$$

where $F(\boldsymbol{\Sigma}, \mathbf{S}) = \operatorname{trace}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) + \ln|\boldsymbol{\Sigma}| - p - \ln|\mathbf{S}|$ is the discrepancy function defined in (66), $F(\hat{\boldsymbol{\Sigma}}_0, \mathbf{S})$ and $F(\hat{\boldsymbol{\Sigma}}_a, \mathbf{S})$ are defined as $F(\boldsymbol{\Sigma}, \mathbf{S})$ in which $\hat{\boldsymbol{\Sigma}}_0$ and $\hat{\boldsymbol{\Sigma}}_a$ is substituted for $\boldsymbol{\Sigma}$.

4.4.1. Asymptotic Distribution: Normal Population

If $\mathbf{y}_i$ in (61) satisfies $\mathbf{y}_i \overset{\text{iid}}{\sim} \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for $i = 1, 2, \ldots, N$, then $(N - r_x)\mathbf{S} \sim \mathbf{W}_p(N - r_x, \boldsymbol{\Sigma})$ and the probability density function of $(N - r_x)\mathbf{S}$ was given in (64). Accordingly, the log likelihood function can be written as follows:

$$l(\boldsymbol{\Sigma}; \mathbf{S}) = -\frac{n}{2}\operatorname{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \frac{n}{2}\ln|\boldsymbol{\Sigma}| + Constants. \tag{154}$$

The generalized likelihood ratio (LR) test statistic is

$$-2\left[l(\hat{\boldsymbol{\Sigma}}_0; \mathbf{S}) - l(\hat{\boldsymbol{\Sigma}}_a; \mathbf{S})\right] = n\left[F(\hat{\boldsymbol{\Sigma}}_0, \mathbf{S}) - F(\hat{\boldsymbol{\Sigma}}_a, \mathbf{S})\right] = X^2, \tag{155}$$

where $l(\hat{\boldsymbol{\Sigma}}_0; \mathbf{S})$ and $l(\hat{\boldsymbol{\Sigma}}_a; \mathbf{S})$ are defined as $l(\boldsymbol{\Sigma}; \mathbf{S})$ in (154) in which $\hat{\boldsymbol{\Sigma}}_0$ and $\hat{\boldsymbol{\Sigma}}_a$ is substituted for $\boldsymbol{\Sigma}$. Wilks' Theorem in [50] established that under some fairly general regularity conditions,

$$X^2 \xrightarrow{\text{dist}} \chi^2_{\nu_a - \nu_0} \quad \text{as} \quad n \to \infty, \tag{156}$$

provided that $\mathbf{H}_0$ is true. Further, for a size α test, $\mathbf{H}_0$ is rejected if $X^2 > \chi^2_{1-\alpha, \nu_a - \nu_0}$.

4.4.2. Asymptotic Distribution: Non-normal Population

Recall that in §4.2, the model $\mathbf{Y} = \mathbf{XB} + \mathbf{F}\boldsymbol{\Lambda}' + \mathbf{E}$ is rewritten as $\mathbf{Y} = \mathbf{XB} + \mathbf{Z}$. In this section, it is assumed that the i^{th} row of $\mathbf{Z}$, $\mathbf{z}_i \overset{\text{iid}}{\sim} (\mathbf{0}, \boldsymbol{\Sigma})$ with finite fourth moments, and $\text{rank}(\mathbf{X}) = r_x = O(1)$. The sample covariance matrix $\mathbf{S}$ is no longer assumed to follow a Wishart distribution. The following Theorem presents the null distribution of $\mathbf{X}^2$ when the assumption for normality of $\mathbf{y}_i$ is not made.

Theorem 33. [Adapted from Boik, Panishkan, and Hyde [35], 2010, Page 131]. *Consider a test of* $\mathbf{H}_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ *against* $\mathbf{H}_a : \boldsymbol{\theta} \in \boldsymbol{\Theta}_a$*, where* $\boldsymbol{\Theta}_0 \cap \boldsymbol{\Theta}_a = \emptyset$*. Suppose that under* $\mathbf{H}_0$*, constraints are placed on* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*, where* $\boldsymbol{\Theta} = \boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_a$*. If* $\mathbf{H}_0$ *is true, then, by construction, there exists a value* $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}_0$ *and a value* $\boldsymbol{\theta}_a \in \boldsymbol{\Theta}$ *so that* $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_a) = \boldsymbol{\Sigma}$ *and* $\boldsymbol{\theta}_a = g(\boldsymbol{\theta}_0)$*. Assume that*

(a.) $\hat{\boldsymbol{\theta}}_0$ *and* $\hat{\boldsymbol{\theta}}_a$ *are consistent roots of* $F(\boldsymbol{\Sigma}(\boldsymbol{\theta}), \mathbf{S})$*, where* $\hat{\boldsymbol{\theta}}_0$ *and* $\hat{\boldsymbol{\theta}}_a$ *are defined in (153);*

(b.) $\text{rank}(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'}) = \nu_0$ *and* $\text{rank}(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}) = \nu_a$*, where* $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'} = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$*,* $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'} = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_a}$*, and* $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}$ *is given in (48); and*

(c.) the p^* *distinct elements in* $\mathbf{z}_i\mathbf{z}_i'$ *have a positive definite covariance matrix. That is,* $\text{Var}[\text{vech}(\mathbf{z}_i\mathbf{z}_i')]$ *is positive definite, where* $\text{vech}(\mathbf{z}_i\mathbf{z}_i')$ *is defined in Table 56.*

Define p^**,* $\mathbf{H}_p$*,* $\mathbf{A}_p$*,* $\mathbf{P}_0$*, and* $\mathbf{P}_a$ *as follows:*

$$\begin{aligned}
&p^* = p(p+1)/2, \quad \mathbf{H}_p = (\mathbf{D}_p'\mathbf{D}_p)^{-1}\mathbf{D}_p', \quad \mathbf{A}_p = \mathbf{D}_p'\boldsymbol{\Omega}\mathbf{D}_p, \\
&\mathbf{P}_0 = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'}\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'}\right]^{-1}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right), \text{ and} \\
&\mathbf{P}_a = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\right]^{-1}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right),
\end{aligned} \tag{157}$$

where $\mathbf{D}_p$ *is a* $p^2 \times p^*$ *duplication matrix defined in Table 56, and* $\boldsymbol{\Omega} = \mathrm{E}\left[\mathbf{z}_i\mathbf{z}_i' \otimes \mathbf{z}_i\mathbf{z}_i'\right] - \boldsymbol{\sigma}\boldsymbol{\sigma}'$ *is given in Theorem 27.*

Then,

(1.) *Define* ν_d *as* $\nu_d = \nu_a - \nu_0$*, where* $\nu_0 = \dim(\mathbf{\Theta}_0)$ *and* $\nu_a = \dim(\mathbf{\Theta})$*.* $\mathbf{P}_a - \mathbf{P}_0$ *is a projection operator and* $\operatorname{rank}(\mathbf{P}_a - \mathbf{P}_0) = \nu_d$*.*

(2.) $\mathbf{A}_p$ *is positive definite and can be diagonalized as* $\mathbf{A}_p = \mathbf{U}_p\mathbf{V}_p\mathbf{U}_p'$*, where* $\mathbf{U}_p \in \mathcal{O}(p^*)$*,* $\mathbf{V}_p = \operatorname{Diag}(v_1, v_2, \ldots, v_{p^*})$*, and* $v_i > 0$ *for* $i = 1, 2, \ldots, p^*$*. Define* $\mathbf{V}_p^{1/2}$ *as* $\mathbf{V}_p^{1/2} = \operatorname{Diag}(\sqrt{v_1}, \sqrt{v_2}, \ldots, \sqrt{v_{p^*}})$ *and* $\mathbf{A}_p^{1/2}$ *as* $\mathbf{A}_p^{1/2} = \mathbf{U}_p\mathbf{V}_p^{1/2}\mathbf{U}_p'$*. Then,* $\mathbf{A}_p^{1/2}$ *is positive definite.*

(3.) *Define* $\mathbf{A}$ *as* $\mathbf{A} = \mathbf{A}_p^{1/2}\mathbf{H}_p\left[\frac{1}{2}\left(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1}\right)(\mathbf{P}_a - \mathbf{P}_0)\right]\mathbf{H}_p'\mathbf{A}_p^{1/2}$ *and define* $\mathbf{J}$ *as* $\mathbf{J} = \frac{1}{2}\mathbf{\Omega}\left(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1}\right)(\mathbf{P}_a - \mathbf{P}_0)$*. The nonzero eigenvalues of* $\mathbf{A}$ *is the same as the nonzero eigenvalues of* $\mathbf{J}$ *and* $\mathbf{A}$ *can be diagonalized as* $\mathbf{A} = \mathbf{UVU}'$*, where* $\mathbf{U}$ *is a* $p^* \times \nu_d$ *semi-orthogonal matrix,* $\{v_i\}_{i=1}^{\nu_d}$ *are the nonzero eigenvalues of* $\mathbf{A}$*, and* $\mathbf{V} = \operatorname{Diag}(v_1, \ldots, v_{\nu_d})$*.*

(4.) *Given* $\mathbf{H}_0$ *is true,* $\mathbf{X}^2 = n\left[F(\hat{\mathbf{\Sigma}}_0, \mathbf{S}) - F(\hat{\mathbf{\Sigma}}_a, \mathbf{S})\right]$ *can be written as*

$$\mathbf{X}^2 = \frac{1}{2}\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})'\left(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1}\right)(\mathbf{P}_a - \mathbf{P}_0)\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma}) + O_p(n^{-\frac{1}{2}}).$$

(5.) *Given* $\mathbf{H}_0$ *is true,* $\mathbf{X}^2 \xrightarrow{\text{dist}} \sum_{i=1}^{\nu_d} v_i w_i^2$*, where* $w_i^2 \overset{\text{iid}}{\sim} \chi_1^2$ *for* $i = 1, \ldots, \nu_d$*.*

All the assumptions made in Theorem 33 hold for the remainder of the thesis, unless otherwise specified.

Corollary 33.1. [Adapted from Boik, Panishkan, and Hyde [35], 2010, Page 131]. *Consider a goodness-of-fit test of* $\mathbf{H}_0 : \boldsymbol{\theta} \in \mathbf{\Theta}_0$*. Assume that the* $p \times p$ *sample covariance matrix* $\mathbf{S}$ *is invertible. The test statistic for the goodness-of-fit test is*

$$X^2 = nF(\hat{\mathbf{\Sigma}}_0, \mathbf{S}),$$

where $F(\mathbf{\Sigma}, \mathbf{S})$ *is defined in (66), and* $\hat{\mathbf{\Sigma}}_0$ *is defined in (153). Reject* $\mathbf{H}_0$*, for large values of* $\mathbf{X}^2$*. Furthermore, the asymptotic null distribution of* X^2 *is* $\mathbf{X}^2 \xrightarrow{\text{dist}} \sum_{i=1}^{\nu_d} v_i w_i^2$*,* *where* $\mathbf{X}^2 = \frac{1}{2}\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)(\mathbf{I}_{p^2}-\mathbf{P}_0)\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) + O_p(n^{-\frac{1}{2}})$*,* $\nu_d = p(p+1)/2-\nu_0$*, and* $w_i^2 \overset{\text{iid}}{\sim} \chi_1^2$ *for* $i=1,\ldots,\nu_d$*.*

Corollary 33.2. [Adapted from Boik, Panishkan, and Hyde [35], 2010, Page 131]. *Provided that* $v_i \approx v$ *for all* i*, the Satterthwaite approximation [51] to the null distribution of* X^2 *is*

$$\lim_{n\to\infty} cX^2 \dot{\sim} \chi_f^2,$$

where $c = \operatorname{tr}(\mathbf{J})/\operatorname{tr}(\mathbf{J}^2)$*,* $\mathbf{X}^2 = n[F(\hat{\boldsymbol{\Sigma}}_0,\mathbf{S}) - F(\hat{\boldsymbol{\Sigma}}_a,\mathbf{S})]$*,* $f = [\operatorname{tr}(\mathbf{J})]^2/\operatorname{tr}(\mathbf{J}^2)$*, and* $\mathbf{J} = \frac{1}{2}\boldsymbol{\Omega}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)(\mathbf{P}_a-\mathbf{P}_0)$*. Reject* $\mathbf{H}_0$ *for large values of* cX^2*.*

4.5. Browne's Residual-Based Test Statistic

Boik, Panishkan, and Hyde [35] presented Browne's [49] residual-based statistic for the test of $\mathbf{H}_0 : \boldsymbol{\theta}\in\boldsymbol{\Theta}_0$ against $\mathbf{H}_a : \boldsymbol{\theta}\in\boldsymbol{\Theta}_a$, where $\boldsymbol{\Theta}_0\cap\boldsymbol{\Theta}_a=\emptyset$. The purpose of this section is to provide Browne's residual-based statistic and to prove its asymptotic null distribution is $\chi^2_{\nu_d}$, where $\nu_d = \nu_a - \nu_0$ is defined in Theorem 33. The notations used in the following theorems are consistent with the notations used in earlier sections.

Theorem 34. [Adapted from Boik, Panishkan, and Hyde [35], 2010, Page 131]. *Consider a test of* $\mathbf{H}_0 : \boldsymbol{\theta}\in\boldsymbol{\Theta}_0$ *against* $\mathbf{H}_a : \boldsymbol{\theta}\in\boldsymbol{\Theta}_a$*, where* $\boldsymbol{\Theta}_0\cap\boldsymbol{\Theta}_a=\emptyset$*. Suppose that under* $\mathbf{H}_0$*, constraints are placed on* $\boldsymbol{\theta}\in\boldsymbol{\Theta}$*, where* $\boldsymbol{\Theta}=\boldsymbol{\Theta}_0\cup\boldsymbol{\Theta}_a$*. If* $\mathbf{H}_0$ *is true, then, by construction, there exists a value* $\boldsymbol{\theta}_0\in\boldsymbol{\Theta}_0$ *and a value* $\boldsymbol{\theta}_a\in\boldsymbol{\Theta}$ *so that* $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)=\boldsymbol{\Sigma}(\boldsymbol{\theta}_a)=\boldsymbol{\Sigma}$ *and* $\boldsymbol{\theta}_a = g(\boldsymbol{\theta}_0)$*.*

Define $\mathbf{P}_{B,0}$, $\mathbf{P}_{B,a}$, $\widehat{\mathbf{P}}_{B,0}$, $\widehat{\mathbf{P}}_{B,a}$, *and* $SSE(\hat{\boldsymbol{\theta}}_0)$ *as follows:*

$$
\begin{aligned}
\mathbf{P}_{B,0} &= \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_0}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_0}\boldsymbol{\Omega}^{+}_{n}\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_0}\right)^{-1}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_0}\boldsymbol{\Omega}^{+}_{n},\\
\mathbf{P}_{B,a} &= \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}\boldsymbol{\Omega}^{+}_{n}\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}\right)^{-1}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}\boldsymbol{\Omega}^{+}_{n},\\
\widehat{\mathbf{P}}_{B,0} &= \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}'_0}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}'_0}\widehat{\boldsymbol{\Omega}}^{+}_{n}\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}'_0}\right)^{-1}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}'_0}\widehat{\boldsymbol{\Omega}}^{+}_{n},\\
\widehat{\mathbf{P}}_{B,a} &= \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}'_a}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}'_a}\widehat{\boldsymbol{\Omega}}^{+}_{n}\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}'_a}\right)^{-1}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}'_a}\widehat{\boldsymbol{\Omega}}^{+}_{n},\\
SSE(\hat{\boldsymbol{\theta}}_0) &= n(\mathbf{s}-\widehat{\boldsymbol{\sigma}}_0)'\widehat{\boldsymbol{\Omega}}^{+}_{n}(\mathbf{I}_{p^2}-\widehat{\mathbf{P}}_{B,0})(\mathbf{s}-\widehat{\boldsymbol{\sigma}}_0), \quad \textit{and}\\
SSE(\hat{\boldsymbol{\theta}}_a) &= n(\mathbf{s}-\widehat{\boldsymbol{\sigma}}_a)'\widehat{\boldsymbol{\Omega}}^{+}_{n}(\mathbf{I}_{p^2}-\widehat{\mathbf{P}}_{B,a})(\mathbf{s}-\widehat{\boldsymbol{\sigma}}_a),
\end{aligned}
\tag{158}
$$

where $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_0} = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$, $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a} = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_a}$, $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}'_0} = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0}$, $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}'_a} = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_a}$, $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}$ *is given in (48),* $\boldsymbol{\Omega}^{+}_{n}$ *and* $\widehat{\boldsymbol{\Omega}}^{+}_{n}$ *are the Moore-Penrose generalized inverses of* $\boldsymbol{\Omega}_n$ *and* $\widehat{\boldsymbol{\Omega}}_n$*, respectively,* $\boldsymbol{\Omega}_n$ *is defined in (141),* $\widehat{\boldsymbol{\Omega}}_n$ *is defined in (142),* $\widehat{\boldsymbol{\sigma}}_0 = \operatorname{vec}\hat{\boldsymbol{\Sigma}}_0$, $\widehat{\boldsymbol{\sigma}}_a = \operatorname{vec}\hat{\boldsymbol{\Sigma}}_a$, *and* $\hat{\boldsymbol{\Sigma}}_0$ *and* $\hat{\boldsymbol{\Sigma}}_a$ *are defined in (153). The Moore-Penrose generalized inverses of* $\boldsymbol{\Omega}_n$ *and* $\widehat{\boldsymbol{\Omega}}_n$ *can be written as* $\mathbf{D}_p(\mathbf{D}'_p\boldsymbol{\Omega}_n\mathbf{D}_p)^{-1}\mathbf{D}'_p$ *and* $\mathbf{D}_p(\mathbf{D}'_p\widehat{\boldsymbol{\Omega}}_n\mathbf{D}_p)^{-1}\mathbf{D}'_p$*, respectively.*

Browne's residual-based statistic is

$$
SSE_B = SSE(\hat{\boldsymbol{\theta}}_0) - SSE(\hat{\boldsymbol{\theta}}_a), \tag{159}
$$

where $SSE(\hat{\boldsymbol{\theta}}_0)$ *and* $SSE(\hat{\boldsymbol{\theta}}_a)$ *are defined in (158). Reject* $\mathbf{H}_0$*, for large values of* SSE_B*. Given* $\mathbf{H}_0$ *is true,* $SSE_B \xrightarrow{\text{dist}} \chi^2_{\nu_d}$*, where* $\nu_d = \nu_a - \nu_0$*,* ν_0 *and* ν_a *are defined in (153).*

4.6. Wald Test

Consider a test of $\mathbf{H}_0 : \boldsymbol{\tau}(\boldsymbol{\theta}) = \boldsymbol{\tau}_0$ against $\mathbf{H}_a : \boldsymbol{\tau}(\boldsymbol{\theta}) \neq \boldsymbol{\tau}_0$, where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and $\boldsymbol{\tau}(\boldsymbol{\theta})$ is a $b-$vector of functions of parameters defined in (145). According to Theorem 5 in Boik, Panishkan, and Hyde [35], based on Theorem 28, it can be concluded that

$$
\sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\tau}}), \tag{160}
$$

where $\boldsymbol{\Omega}_{\boldsymbol{\tau}} = \mathbf{D}^{(1)}_{\boldsymbol{\tau};\boldsymbol{\theta}'}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\mathbf{D}^{(1)\prime}_{\boldsymbol{\tau};\boldsymbol{\theta}'}$, $\mathbf{D}^{(1)}_{\boldsymbol{\tau};\boldsymbol{\theta}'}$ is defined in (145), and $\boldsymbol{\Omega}_{\boldsymbol{\theta}}$ is defined in Theorem 28.

Define $\boldsymbol{\Omega}_{\boldsymbol{\tau},n}$ as $\boldsymbol{\Omega}_{\boldsymbol{\tau},n} = \mathbf{D}^{(1)}_{\boldsymbol{\tau};\boldsymbol{\theta}'}\boldsymbol{\Omega}_{\boldsymbol{\theta},n}\mathbf{D}^{(1)\prime}_{\boldsymbol{\tau};\boldsymbol{\theta}'}$, where $\boldsymbol{\Omega}_{\boldsymbol{\theta},n}$ is defined in (144). Note that $\boldsymbol{\Omega}_{\boldsymbol{\tau}} = \lim_{n\to\infty}\boldsymbol{\Omega}_{\boldsymbol{\tau},n}$.

Assume that $\boldsymbol{\Omega}^{-1}_{\boldsymbol{\tau}}$ exists, define Q_W as

$$Q_W = n(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0)'\widehat{\boldsymbol{\Omega}}^{-1}_{\boldsymbol{\tau},n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0), \tag{161}$$

where $\widehat{\boldsymbol{\Omega}}^{-1}_{\boldsymbol{\tau},n} = \mathbf{D}^{(1)}_{\boldsymbol{\tau};\hat{\boldsymbol{\theta}}'}\widehat{\boldsymbol{\Omega}}_{\boldsymbol{\theta},n}\mathbf{D}^{(1)\prime}_{\boldsymbol{\tau};\hat{\boldsymbol{\theta}}'}$, $\mathbf{D}^{(1)}_{\boldsymbol{\tau};\hat{\boldsymbol{\theta}}'}$ and $\widehat{\boldsymbol{\Omega}}_{\boldsymbol{\theta},n}$ are defined in (145).

By (160), Theorem 1.3.6 part (1) on page 10 of Christensen [52], and Slutsky's Theorem 23, it can be concluded that

$$Q_W \xrightarrow{\text{dist}} \chi^2_b, \tag{162}$$

because $\operatorname{tr}\left(\boldsymbol{\Omega}^{-1}_{\boldsymbol{\tau},n}\boldsymbol{\Omega}_{\boldsymbol{\tau},n}\right) = \operatorname{tr}(\mathbf{I}_b) = b$ and $\widehat{\boldsymbol{\Omega}}^{-1}_{\boldsymbol{\tau},n}$ is a consistent estimator of $\boldsymbol{\Omega}^{-1}_{\boldsymbol{\tau},n}$. The result in (162) is from Boik, Panishkan, and Hyde [35].

CHAPTER 5

ILLUSTRATION

To demonstrate the 4-step procedure from this thesis, four different data sets are used. These data sets are from Levy [53], Ostrom [54], Bentler and Jamshidian [29] as well as Byrne and Shavelson [55]. For each data set, a comparison table is provided with respect to the results generated under the same constraint fit by three software routines, SAS, R and the 4-step procedure. For notational convenience, denote the 4-step procedure by 4SP in the remainder of this thesis.

5.1. Levy's Data

The data from Levy [53] corresponds to an investigation into the perceived quality of life of United States residents. The 2164 participants were asked to rate their satisfaction with respect to 15 variables. These 15 variables are city to live in (v_1), neighborhood (v_2), housing (v_3), life in USA (v_4), education (v_5), useful education (v_6), job satisfaction (v_7), spending spare time (v_8), health (v_9), standard of living (v_{10}), saving and investments (v_{11}), friends (v_{12}), marriage (v_{13}), family life (v_{14}) and life in general (v_{15}). From these ratings, a correlation matrix has been calculated. Based on the result of the goodness-of-fit test in Srivastava [56] on page 441, three factors seem adequate for this data set. Overall, for Levy's data, $N = 2164$, $p = 15$, and $q = 3$.

An identical model, fit by SAS, R and 4SP, is described by listing all of the constraints imposed on $\mathbf{\Lambda}$, $\mathbf{\Phi}$ ($\mathbf{\Gamma}$ and $\mathbf{\Delta}$), and $\mathbf{\Psi}$. This model used for Levy's data is called Levy Model I.

5.1.1. Levy Model I

5.1.1.1. Constraints Imposed On $\mathbf{\Lambda}$: The structure of $\mathbf{\Lambda}$ is chosen in the following manner: factor 1 has loadings only on v_1, v_2, v_3, v_4; factor 2 has loadings only on v_5, v_6, v_7, v_8, v_9; and factor 3 has loadings only on v_{10}, v_{11}, v_{12}, v_{13}, v_{14}, v_{15}. Therefore, $\mathbf{\Lambda}$ has the form

$$\mathbf{\Lambda} = \bigoplus_{i=1}^{3} \mathbf{\Lambda}_i, \tag{163}$$

where $\bigoplus$ is defined in Table 56, $\mathbf{\Lambda}_1 = (\lambda_1\ \lambda_2\ \lambda_3\ \lambda_4)'$, $\mathbf{\Lambda}_2 = (\lambda_5\ \lambda_6\ \lambda_7\ \lambda_8\ \lambda_9)'$, and $\mathbf{\Lambda}_3 = (\lambda_{10}\ \lambda_{11}\ \lambda_{12}\ \lambda_{13}\ \lambda_{14}\ \lambda_{15})'$.

5.1.1.2. Constraints Imposed On $\mathbf{\Phi}$: Structure 1a in Table 4 is chosen for the eigenvalues of $\mathbf{\Phi}$, where structure 1a is $\mathbf{T}_2\boldsymbol{\xi}_\delta$ with $\mathbf{1}_3'\boldsymbol{\delta} = 3$. The matrix $\mathbf{T}_2$ for structure 1a can be any 3×3 nonsingular matrix. For example,

$$\mathbf{T}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{or} \quad \mathbf{T}_2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}. \tag{164}$$

According to (37) in Chapter 2, $\mathbf{\Phi}$ satisfies $\mathbf{e}_i^{q\prime}\mathbf{\Phi}\mathbf{e}_i^q = 1$, for $i = 1, 2, 3$, equivalently, $\operatorname{diag}(\mathbf{\Phi}) = \mathbf{1}_3$. The corresponding $\widehat{\boldsymbol{\delta}}$ is provided in Table 6.

5.1.1.3. Constraints Imposed On $\mathbf{\Psi}$: The diagonal entries of $\mathbf{\Psi}$ are unconstrained. In 4SP, structure 1a in Table 5 is chosen for $\boldsymbol{\psi}$, where structure 1a is $\boldsymbol{\psi} = \mathbf{T}_4\boldsymbol{\xi}_\psi$. The matrix $\mathbf{T}_4$ is chosen to be an identity matrix with dimension $p = 15$, allowing unrestricted entries in $\boldsymbol{\psi}$.

5.1.1.4. Fit by SAS, R and 4SP: Denote the estimated value of $F(\mathbf{\Sigma}, \mathbf{S})$ by $F(\mathbf{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ in the remainder of this thesis, where $F(\mathbf{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S}) = \operatorname{tr}(\mathbf{S}[\mathbf{\Sigma}(\hat{\boldsymbol{\theta}})]^{-1}) +$

Table 6: Comparisons of the values of $\widehat{\boldsymbol{\delta}}$ and $F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ under Levy Model I

Levy Model I Fit By	SAS	R	4SP
$\widehat{\boldsymbol{\delta}}$	$\begin{pmatrix} 2.3686 \\ 0.5453 \\ 0.0862 \end{pmatrix}$	$\begin{pmatrix} 2.3686 \\ 0.5453 \\ 0.0862 \end{pmatrix}$	$\begin{pmatrix} 2.3686 \\ 0.5453 \\ 0.0862 \end{pmatrix}$
$F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$	0.7210	0.7210	0.7210

$\ln|\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})| - p - \ln|\mathbf{S}|$. Table 6 is comparing the values of $\widehat{\boldsymbol{\delta}}$ and $F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ for Levy's data. The estimates for $\boldsymbol{\Lambda}$, $\boldsymbol{\delta}$ and $\boldsymbol{\Psi}$ generated by SAS, R and 4SP are identical up to the fourth digit. The $F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ produced by SAS, R and 4SP are identical up to the fourth digit.

To generate the identical estimates in 4SP as shown in Table 6, the following structures for $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ can also be used. In Table 4, structure 4 can be chosen for the eigenvalues of $\boldsymbol{\Phi}$, where structure 4 is $\mathbf{T}_2\boldsymbol{\xi}_{\boldsymbol{\delta}}^{\odot 2}$ with $\mathbf{1}_3'\boldsymbol{\delta} = 3$. The matrix $\mathbf{T}_2$ for structure 4 can be any 3×3 nonsingular matrix with non-negative entries. For example,

$$\mathbf{T}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{or} \quad \mathbf{T}_2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}. \tag{165}$$

Structure 5 in Table 5 can be chosen for $\boldsymbol{\psi}$, where structure 5 is $\boldsymbol{\psi} = \mathbf{T}_4\boldsymbol{\xi}_{\boldsymbol{\psi}}^{\odot 2}$.

For illustration purposes, a diagram for Levy Model I, based on the graphviz graph visualisation software (Gansner [57]) and the output from the *sem* package [31] in R is given in Appendix C, where v_i is defined in Section 5.1 for $i = 1, \ldots, 15$ and lam_i in the diagram of Appendix C corresponds to λ_i given in (163) for $i = 1, \ldots, 15$.

5.1.2. Other Analysis and Model Fits in 4SP

To illustrate the flexibility of 4SP, two different models, Levy Model II and Levy Model III, are used. In those two models, the constraints imposed on $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ remain the same as those of Levy Model I.

5.1.2.1. Constraints Imposed On $\boldsymbol{\delta}$ in Levy Model II: In Levy Model II, all entries in $\boldsymbol{\delta}$ are constrained to be positive and decreasing. Structure 2a in Table 4 is used to parameterize $\boldsymbol{\delta}$, and is given as

$$\boldsymbol{\delta} = \frac{3\mathbf{T}_1 \exp_{\odot}\{\mathbf{T}_2\boldsymbol{\xi}_\delta\}}{\mathbf{1}_3'\mathbf{T}_1 \exp_{\odot}\{\mathbf{T}_2\boldsymbol{\xi}_\delta\}}.$$

The matrices $\mathbf{T}_1$ and $\mathbf{T}_2$ corresponding to structure 2a are

$$\mathbf{T}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{T}_2 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}. \tag{166}$$

It follows from structure 2a in Table 4 that $\mathbf{1}_3'\boldsymbol{\delta} = 3$. The corresponding values of $\widehat{\boldsymbol{\delta}}$ and $F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ are provided in Table 7.

5.1.2.2. Constraints Imposed On $\boldsymbol{\delta}$ in Levy Model III: In Levy Model III, the constraints that are imposed on $\boldsymbol{\delta}$ are motivated by the structure of $\widehat{\boldsymbol{\delta}}$ from Table 6. Structure 2a from Table 4 is used to parameterize $\boldsymbol{\delta}$ so that all the entries in $\boldsymbol{\delta}$ are positive and exponentially decreasing. The matrices $\mathbf{T}_1$ and $\mathbf{T}_2$ are

$$\mathbf{T}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and } \mathbf{T}_2 = \begin{pmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}. \tag{167}$$

The corresponding values of $\widehat{\boldsymbol{\delta}}$ and $F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ are provided in Table 7.

Table 7: Comparisons of the values of $\widehat{\boldsymbol{\delta}}$ and $F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ for 3 Models in 4SP

	Levy Model I	Levy Model II	Levy Model III
constraints for $\boldsymbol{\delta}$	$\mathbf{1}_3'\boldsymbol{\delta} = 3$	$\mathbf{1}_3'\boldsymbol{\delta} = 3$ and $\boldsymbol{\delta}$ is positive and decreasing	$\mathbf{1}_3'\boldsymbol{\delta} = 3$ and $\boldsymbol{\delta}$ is positive and exponentially decreasing
$\widehat{\boldsymbol{\delta}}$	$\begin{pmatrix} 2.3686 \\ 0.5453 \\ 0.0862 \end{pmatrix}$	$\begin{pmatrix} 2.3686 \\ 0.5453 \\ 0.0862 \end{pmatrix}$	$\begin{pmatrix} 2.3717 \\ 0.5160 \\ 0.1123 \end{pmatrix}$
$F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$	0.7210	0.7210	0.7228

Table 7 is comparing the values of $\widehat{\boldsymbol{\delta}}$ and $F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ under Levy Model I, II and III using 4SP. 4SP is the only program used to fit Levy Model II and III because no existing commands in SAS or R can be used to impose the constraints on $\boldsymbol{\delta}$ in Levy Model II or III. Levy Model I produces $\widehat{\boldsymbol{\Lambda}}$, $\widehat{\boldsymbol{\Delta}}$, $\widehat{\boldsymbol{\Gamma}}$, $\widehat{\boldsymbol{\Psi}}$ and $F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ that are identical as those of Levy Model II. Levy Model III has a larger discrepancy function value, because the exponentially decreasing model is not as good a fit as the other two models.

The following Theorem justifies the reasoning of why Levy Model I produces a $F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ that is identical with that of Levy Model II.

Theorem 35. [Original result]. *Let* $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\boldsymbol{\theta}_{\boldsymbol{\lambda}})$, $\boldsymbol{\Phi} = \boldsymbol{\Phi}(\boldsymbol{\theta}_{\boldsymbol{\delta}}, \boldsymbol{\theta}_{\gamma})$ *and* $\boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{\theta}_{\psi})$, *where* $\dim(\boldsymbol{\Lambda}) = p \times q$, $\dim(\boldsymbol{\Phi}) = q \times q$ *and* $\dim(\boldsymbol{\Psi}) = p \times p$. *Let* $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$ *represent a* $p \times p$ *covariance matrix, where* $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\lambda}' \quad \boldsymbol{\theta}_{\delta}' \quad \boldsymbol{\theta}_{\gamma}' \quad \boldsymbol{\theta}_{\psi}')'$ *and* $\dim(\boldsymbol{\theta}) = \nu$. *Write* $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ *as* $\boldsymbol{\Sigma}(\boldsymbol{\theta}) : \mathbb{R}^{\nu} \to \mathbb{R}^{p\times p}$. *Define* $\boldsymbol{\Theta}_1$ *as* $\boldsymbol{\Theta}_1 = \left\{\boldsymbol{\theta}\middle|\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi} \quad \text{and} \quad \operatorname{diag}(\boldsymbol{\Phi}(\boldsymbol{\theta}_{\boldsymbol{\delta}}, \boldsymbol{\theta}_{\gamma})) = \mathbf{1}_q\right\}$ *and* $\boldsymbol{\Theta}_2$ *as* $\boldsymbol{\Theta}_2 = \left\{\boldsymbol{\theta}\middle|\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi} \quad \text{and} \quad \boldsymbol{\Phi}(\boldsymbol{\theta}_{\boldsymbol{\delta}}, \boldsymbol{\theta}_{\gamma}) \text{ is a correlation matrix}\right\}$. *Define* $\boldsymbol{\Sigma}_1$ *as* $\boldsymbol{\Sigma}_1 : \boldsymbol{\Theta}_1 \to \mathbb{R}^{p\times p}$ *and* $\boldsymbol{\Sigma}_2$ *as* $\boldsymbol{\Sigma}_2 : \boldsymbol{\Theta}_2 \to \mathbb{R}^{p\times p}$. *Let* $\hat{\boldsymbol{\theta}}_1 = (\hat{\boldsymbol{\theta}}_{\lambda 1}' \quad \hat{\boldsymbol{\theta}}_{\delta 1}' \quad \hat{\boldsymbol{\theta}}_{\gamma 1}' \quad \hat{\boldsymbol{\theta}}_{\psi 1}')'$ *be the minimizer of* $F(\boldsymbol{\Sigma}_{\mathbf{1}}(\boldsymbol{\theta}), \mathbf{S})$ *and* $\hat{\boldsymbol{\theta}}_2 = (\hat{\boldsymbol{\theta}}_{\lambda 2}' \quad \hat{\boldsymbol{\theta}}_{\delta 2}' \quad \hat{\boldsymbol{\theta}}_{\gamma 2}' \quad \hat{\boldsymbol{\theta}}_{\psi 2}')'$ *be the minimizer of* $F(\boldsymbol{\Sigma}_{\mathbf{2}}(\boldsymbol{\theta}), \mathbf{S})$. *If* $\boldsymbol{\Phi}(\hat{\boldsymbol{\theta}}_{\delta 1}, \hat{\boldsymbol{\theta}}_{\gamma 1})$ *is a correlation matrix, then* $F(\boldsymbol{\Sigma}_{\mathbf{1}}(\hat{\boldsymbol{\theta}}_{\mathbf{1}}), \mathbf{S}) = F(\boldsymbol{\Sigma}_{\mathbf{2}}(\hat{\boldsymbol{\theta}}_{\mathbf{2}}), \mathbf{S})$.

The following sections provide examples for when the estimate of $\boldsymbol{\delta}$ is improper. The relevant models to avoid improper solutions are also discussed.

5.2. Ostrom's Data

Ostrom [54] used the multitrait-multimethod matrix procedure of Campbell and Fiske [58] to compute a 12×12 correlation matrix. There were 189 subjects who participated in this study. There were three traits: the affective (t_1), the behavioral (t_2) and the cognitive component (t_3). And there were four methods: Thurstone scale/equal-appearing intervals (m_1), Likert scale/summated ratings (m_2), Guttman scale/scalogram analysis (m_3) and self-rating (m_4). Accordingly, there were 12 variables listed with their sequential trait (t_i) and method (m_j) designation along with 4 method factors. Ostrom's data [54] can be studied by a CFA model, with $N = 189$, $p = 12$, and $q = 4$.

The identical model fit by SAS, R and 4SP is described by listing all of the constraints imposed on $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$ ($\boldsymbol{\Gamma}$ and $\boldsymbol{\Delta}$), and $\boldsymbol{\Psi}$. Denote this identical model for Ostrom's data by Ostrom Model I.

5.2.1. Ostrom Model I

5.2.1.1. Constraints Imposed On $\boldsymbol{\Lambda}$: The structure of $\boldsymbol{\Lambda}$ is chosen in the following manner: factor 1 has loadings only on t_1m_1, t_2m_1, t_3m_1; factor 2 has loadings only on t_1m_2, t_2m_2, t_3m_2; factor 3 has loadings only on t_1m_3, t_2m_3, t_3m_3; and factor 4 has loadings only on t_1m_4, t_2m_4, t_3m_4. Thus,

$$\boldsymbol{\Lambda} = \bigoplus_{i=1}^{4} \boldsymbol{\Lambda}_i, \tag{168}$$

where $\bigoplus$ is defined in Table 56, $\mathbf{\Lambda}_1 = (\lambda_1\ \lambda_2\ \lambda_3)'$, $\mathbf{\Lambda}_2 = (\lambda_4\ \lambda_5\ \lambda_6)'$, $\mathbf{\Lambda}_3 = (\lambda_7\ \lambda_8\ \lambda_9)'$, and $\mathbf{\Lambda}_4 = (\lambda_{10}\ \lambda_{11}\ \lambda_{12})'$.

5.2.1.2. Constraints Imposed On $\mathbf{\Phi}$: The diagonal entries of $\mathbf{\Phi}$ are restricted to be 1's. That is, $\operatorname{diag}(\mathbf{\Phi}) = \mathbf{1}_4$. To constrain $\operatorname{diag}(\mathbf{\Phi}) = \mathbf{1}_4$ in 4SP, structure 1a in Table 4 is used to parameterize $\boldsymbol{\delta}$, where structure 1a is $\boldsymbol{\delta} = \mathbf{T}_2\boldsymbol{\xi}_\delta$ with $\mathbf{1}_4'\boldsymbol{\delta} = 4$. The matrix $\mathbf{T}_2$ can be given as

$$\mathbf{T}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{169}$$

The corresponding $\widehat{\boldsymbol{\delta}}$ is provided in Table 8.

5.2.1.3. Constraints Imposed On $\mathbf{\Psi}$: The diagonal entries of $\mathbf{\Psi}$ are unconstrained. In 4SP, structure 1a in Table 5 is chosen to parameterize $\boldsymbol{\psi}$, where structure 1a is $\boldsymbol{\psi} = \mathbf{T}_4\boldsymbol{\xi}_\psi$. The matrix $\mathbf{T}_4$ is chosen to be an identity matrix with dimension $p = 12$; therefore, all entries in $\boldsymbol{\psi}$ are unrestricted.

5.2.1.4. Fit by SAS, R and 4SP: Table 8 is comparing the values of $\widehat{\boldsymbol{\delta}}$ and $F(\mathbf{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ for Ostrom's data. It shows that SAS, R and 4SP generate identical estimates for $\mathbf{\Lambda}$, $\mathbf{\Phi}$, $\mathbf{\Psi}$ and $F(\mathbf{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ under Ostrom Model I. One estimated eigenvalue, $\widehat{\boldsymbol{\delta}}$, is negative ($-0.0142$), which violates a property of a correlation matrix.

Table 8: Comparisons of the values of $\widehat{\boldsymbol{\delta}}$ and $F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}),\mathbf{S})$ under Ostrom Model I Among SAS, R and 4SP

Ostrom Model I Fit By	SAS	R	4SP
$\widehat{\boldsymbol{\delta}}$	$\begin{pmatrix} 3.8304 \\ 0.1527 \\ 0.0311 \\ -0.0142 \end{pmatrix}$	$\begin{pmatrix} 3.8304 \\ 0.1527 \\ 0.0311 \\ -0.0142 \end{pmatrix}$	$\begin{pmatrix} 3.8304 \\ 0.1527 \\ 0.0311 \\ -0.0142 \end{pmatrix}$
$F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}),\mathbf{S})$	0.3982	0.3982	0.3982

Furthermore, the estimate for $\boldsymbol{\Phi}$ is

$$\widehat{\boldsymbol{\Phi}} = \begin{pmatrix} 1.0000 & 0.9699 & 1.0072 & 0.8896 \\ 0.9699 & 1.0000 & 0.9911 & 0.8741 \\ 1.0072 & 0.9911 & 1.0000 & 0.9248 \\ 0.8896 & 0.8741 & 0.9248 & 1.0000 \end{pmatrix}. \tag{170}$$

A correlation of 1.0072 in $\widehat{\boldsymbol{\Phi}}$ is improper and further analysis is conducted to deal with this issue.

5.2.2. Further Analysis and Model Fits

There are two different models, Ostrom Model II and Ostrom Model III, discussed in the following sections. In both models, the constraints imposed on $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ remain the same as those of Ostrom Model I.

5.2.2.1. Constraints Imposed On $\boldsymbol{\delta}$ in Ostrom Model II: Bentler and Jamshidian [29] suggested to constrain all entries in $\boldsymbol{\delta}$ to be nonnegative, that is, $\delta_i \geq 0$ for $i = 1, \ldots, 4$, where $\boldsymbol{\delta} = (\delta_1\ \delta_2\ \delta_3\ \delta_4)'$. In order to reproduce Bentler's constrained solutions in 4SP, structure 4 in Table 4 is used to parameterize $\boldsymbol{\delta}$, where structure 4 is $\boldsymbol{\delta} = \mathbf{T}_2 \boldsymbol{\xi}_{\boldsymbol{\delta}}^{\odot 2}$ with $\mathbf{1}_4' \boldsymbol{\delta} = 4$. The matrix $\mathbf{T}_2$ corresponding to this structure is

$$\mathbf{T}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{171}$$

The corresponding values of $\widehat{\boldsymbol{\Phi}}$, $\widehat{\boldsymbol{\delta}}$ and $F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ are provided in Table 9.

5.2.2.2. Constraints Imposed On $\boldsymbol{\delta}$ in Ostrom Model III: A new structure for $\boldsymbol{\delta}$ is motivated by the structure of $\widehat{\boldsymbol{\delta}}$ in Table 8. Structure 2a from Table 4 is used to parameterize $\boldsymbol{\delta}$ so that the largest eigenvalue, δ_1, is unrestricted and $\delta_2/\delta_3 = \delta_3/\delta_4$. Hence, the structure of $\mathbf{T}_2$ is

$$\mathbf{T}_2 = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 2 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}. \tag{172}$$

The corresponding values of $\widehat{\boldsymbol{\Phi}}$, $\widehat{\boldsymbol{\delta}}$ and $F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ are provided in Table 9. For convenience, denote this new model by Ostrom Model III.

Table 9 is comparing the values of $\widehat{\boldsymbol{\Phi}}$, $\widehat{\boldsymbol{\delta}}$ and $F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ under Ostrom Model II and III. In Ostrom Model II, one eigenvalue in $\widehat{\boldsymbol{\delta}}$ is 0.0000; therefore, $\widehat{\boldsymbol{\Phi}}$ has rank 3 and one of the factors is linearly dependent on the others. A discrepancy value of 0.4003 produced by Ostrom Model III in 4SP is slightly larger than the 0.3992 produced by Ostrom Model II. However, unlike Ostrom Model II, $\widehat{\boldsymbol{\delta}}$ and $\widehat{\boldsymbol{\Phi}}$ produced by Ostrom Model III in 4SP are both proper.

Table 9: Comparisons of $\widehat{\boldsymbol{\Phi}}$, $\widehat{\boldsymbol{\delta}}$ and $F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ for Ostrom Model II and III

	Ostrom Model II	Ostrom Model III
$\widehat{\boldsymbol{\Phi}}$	$\begin{pmatrix} 1.0000 & 0.9691 & 0.9949 & 0.8887 \\ 0.9691 & 1.0000 & 0.9832 & 0.8741 \\ 0.9949 & 0.9832 & 1.0000 & 0.9176 \\ 0.8887 & 0.8741 & 0.9176 & 1.0000 \end{pmatrix}$	$\begin{pmatrix} 1.0000 & 0.9711 & 0.9905 & 0.8891 \\ 0.9711 & 1.0000 & 0.9809 & 0.8740 \\ 0.9905 & 0.9809 & 1.0000 & 0.9147 \\ 0.8891 & 0.8740 & 0.9147 & 1.0000 \end{pmatrix}$
$\widehat{\boldsymbol{\delta}}$	$\begin{pmatrix} 3.8156 \\ 0.1529 \\ 0.0315 \\ 0.0000 \end{pmatrix}$	$\begin{pmatrix} 3.8119 \\ 0.1533 \\ 0.0292 \\ 0.0056 \end{pmatrix}$
$F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$	0.3992	0.4003

5.3. Bentler's Example

Bentler and Jamshidian [29] provided a nine variable (v_1,...,v_9) and three factor model. The hypothetical $\boldsymbol{\Lambda}_*$, $\boldsymbol{\Phi}_*$, $\boldsymbol{\delta}$ and $\boldsymbol{\Psi}$ in Table 10 are also from Bentler and Jamshidian [29]. Recall that $\boldsymbol{\Phi}_*$ is defined as a covariance matrix and $\boldsymbol{\Lambda}_*$ is defined as a factor loading matrix in the alternative parameterization of $\boldsymbol{\Sigma}$ of Chapter 2. The hypothetical $\boldsymbol{\Phi}_*$ in Table 10 is an improper factor covariance matrix because one factor variance in $\boldsymbol{\Phi}_*$ (-1.46) and one eigenvalue of $\boldsymbol{\Phi}_*$ (-1.88) are negative.

Accordingly, the hypothetical covariance matrix is computed by $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}_*\boldsymbol{\Phi}_*\boldsymbol{\Lambda}_*' + \boldsymbol{\Psi}$, which is

Table 10: A Table of $\mathbf{\Lambda}_*$, $\mathbf{\Phi}_*$, $\boldsymbol{\delta}$ and $\mathbf{\Psi}$ for Bentler's Example

$\mathbf{\Lambda}_*$	$\begin{pmatrix} 1.00 & 0.00 & 0.00 \\ 0.57 & 0.00 & 0.00 \\ 0.22 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 \\ 0.00 & 0.90 & 0.00 \\ 0.00 & 0.98 & 0.00 \\ 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 1.06 \\ 0.00 & 0.00 & 0.22 \end{pmatrix}$
$\mathbf{\Phi}_*$	$\begin{pmatrix} -1.46 & 0.88 & 0.96 \\ 0.88 & 1.48 & 0.94 \\ 0.96 & 0.94 & 1.00 \end{pmatrix}$
$\boldsymbol{\delta}$	$\begin{pmatrix} 2.62 & 0.29 & -1.88 \end{pmatrix}'$
$\mathbf{\Psi}$	$\mathrm{Diag}(6.43 \times \mathbf{1}_9)$

$$\mathbf{\Sigma} = \begin{pmatrix} 4.9700 & -0.8322 & -0.3212 & 0.8800 & 0.7920 & 0.8624 & 0.9600 & 1.0176 & 0.2112 \\ -0.8322 & 5.9556 & -0.1831 & 0.5016 & 0.4514 & 0.4916 & 0.5472 & 0.5800 & 0.1204 \\ -0.3212 & -0.1831 & 6.3593 & 0.1936 & 0.1742 & 0.1897 & 0.2112 & 0.2239 & 0.0465 \\ 0.8800 & 0.5016 & 0.1936 & 7.9100 & 1.3320 & 1.4504 & 0.9400 & 0.9964 & 0.2068 \\ 0.7920 & 0.4514 & 0.1742 & 1.3320 & 7.6288 & 1.3054 & 0.8460 & 0.8968 & 0.1861 \\ 0.8624 & 0.4916 & 0.1897 & 1.4504 & 1.3054 & 7.8514 & 0.9212 & 0.9765 & 0.2027 \\ 0.9600 & 0.5472 & 0.2112 & 0.9400 & 0.8460 & 0.9212 & 7.4300 & 1.0600 & 0.2200 \\ 1.0176 & 0.5800 & 0.2239 & 0.9964 & 0.8968 & 0.9765 & 1.0600 & 7.5536 & 0.2332 \\ 0.2112 & 0.1204 & 0.0465 & 0.2068 & 0.1861 & 0.2027 & 0.2200 & 0.2332 & 6.4784 \end{pmatrix}. \tag{173}$$

The vector of eigenvalues of $\mathbf{\Sigma}$ is

$$\begin{pmatrix} 12.7315 & 3.7105 & 7.1152 & 6.4300 & 6.4300 & 6.4300 & 6.4300 & 6.4300 & 6.4300 \end{pmatrix}'.$$

Therefore, $\mathbf{\Sigma}$ is a covariance matrix because $\mathbf{\Sigma}$ is symmetric and its eigenvalues are non-negative. In the following section, the hypothetical $\mathbf{\Lambda}_*$, $\mathbf{\Phi}_*$ and $\mathbf{\Psi}$ in Table 10 are reproduced given the covariance matrix $\mathbf{\Sigma}$. Denote the model used to recover the exact $\mathbf{\Lambda}_*$, $\mathbf{\Phi}_*$ and $\mathbf{\Psi}$ in Table 10 by Bentler Model I.

5.3.1. Bentler Model I

5.3.1.1. Constraints Imposed On $\mathbf{\Lambda}_*$: The structure of $\mathbf{\Lambda}_*$ is chosen in the following manner:

$$\mathbf{\Lambda}_* = \bigoplus_{i=1}^{3} \mathbf{\Lambda}_i, \tag{174}$$

where $\bigoplus$ is defined in Table 56, $\mathbf{\Lambda}_1 = (1\ \lambda_1\ \lambda_2)'$, $\mathbf{\Lambda}_2 = (1\ \lambda_3\ \lambda_4)'$, and $\mathbf{\Lambda}_3 = (1\ \lambda_5\ \lambda_6)'$. Note that factor 1, factor 2 and factor 3 have fixed loadings on v_1, v_4 and v_7 at 1, respectively.

5.3.1.2. Constraints Imposed On $\mathbf{\Phi}_*$: $\mathbf{\Phi}_*$ is constrained to be symmetric.

5.3.1.3. Constraints Imposed On $\mathbf{\Psi}$: The diagonal entries of $\mathbf{\Psi}$ are constrained to be identical. Structure 1a in Table 5 is chosen for $\boldsymbol{\psi}$, where structure 1a is $\boldsymbol{\psi} = \mathbf{T}_4\boldsymbol{\xi}_{\psi}$. The matrix $\mathbf{T}_4$ is chosen to be a vector of $\mathbf{1}_9$.

The modified Fisher Scoring algorithm based on the covariance parameterization of $\mathbf{\Phi}_*$ is applied to generate the estimates for $\mathbf{\Lambda}_*$, $\mathbf{\Phi}_*$, $\boldsymbol{\delta}$ and $\mathbf{\Psi}$. The generated $\widehat{\mathbf{\Lambda}}$, $\widehat{\mathbf{\Phi}}$, $\widehat{\boldsymbol{\delta}}$ and $\widehat{\mathbf{\Psi}}$ are the same with the hypothetical $\mathbf{\Lambda}_*$, $\mathbf{\Phi}_*$, $\boldsymbol{\delta}$ and $\mathbf{\Psi}$ in Table 10, respectively. Under Bentler Model I, $F(\mathbf{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ is 0.

Marsh [59] suggested analyzing a covariance matrix with factor variances fixed and denote this process by the fixed factor variance parameterization. Specifically, this process was done by fixing the factor variance of each factor to be 1 and estimating all the factor loadings. Bentler and Jamshidian [29] provided a solution based on Marsh's suggestion. Denote the corresponding model by Bentler Model II.

5.3.2. Bentler Model II

5.3.2.1. Constraints Imposed On $\mathbf{\Lambda}$: The structure of $\mathbf{\Lambda}$ is chosen in the following manner. Factor 1 has loadings only on v_1, v_2, v_3; factor 2 has loadings only on v_4, v_5, v_6; and factor 3 has loadings only on v_7, v_8, v_9. Thus,

$$\mathbf{\Lambda} = \bigoplus_{i=1}^{3} \mathbf{\Lambda}_i, \tag{175}$$

where $\mathbf{\Lambda}_1 = (\lambda_1\ \lambda_2\ \lambda_3)'$, $\mathbf{\Lambda}_2 = (\lambda_4\ \lambda_5\ \lambda_6)'$, and $\mathbf{\Lambda}_3 = (\lambda_7\ \lambda_8\ \lambda_9)'$.

5.3.2.2. Constraints Imposed On $\mathbf{\Phi}$: The diagonal entries of $\mathbf{\Phi}$ are restricted to be all 1's and $\mathbf{\Phi}$ is constrained to be symmetric.

5.3.2.3. Constraints Imposed On $\mathbf{\Psi}$: The diagonal entries of $\mathbf{\Psi}$ are unconstrained.

Relevant estimates of $\widehat{\mathbf{\Lambda}}$, $\widehat{\mathbf{\Phi}}$, $\widehat{\boldsymbol{\delta}}$ are provided in Bentler and Jamshidian [29] and are also given in Table 11.

For notational convenience, denote the Lagrange multipliers algorithm based on a parameterization of $\mathbf{\Phi}_*$ by LMA for the remainder of the thesis. In the following section, Bentler Model II is fit by LMA. The purpose is to compare the solution generated in LMA to the corresponding solution in Bentler and Jamshidian [29].

5.3.2.4. Bentler Model II in LMA: In LMA, structure 1a in Table 4 is used to parameterize $\boldsymbol{\delta}$, where structure 1a is $\boldsymbol{\delta} = \mathbf{T}_2\boldsymbol{\xi}_{\boldsymbol{\delta}}$ with $\mathbf{1}_3'\boldsymbol{\delta} = 3$. The matrix $\mathbf{T}_2$ corresponding to this structure is

$$\mathbf{T}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{176}$$

In LMA, structure 1a in Table 5 is chosen to parameterize $\boldsymbol{\psi}$, where structure 1a is $\boldsymbol{\psi} = \mathbf{T}_4\boldsymbol{\xi}_\psi$. The matrix $\mathbf{T}_4$ is chosen to be an identity matrix with dimension $p = 9$.

The values of $\widehat{\boldsymbol{\Lambda}}$, $\widehat{\boldsymbol{\Phi}}$, $\widehat{\boldsymbol{\delta}}$ and $F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ generated in LMA under Bentler Model II are given in Table 11. Those values are similar to the corresponding values provided in Bentler and Jamshidian [29]. However, $\widehat{\boldsymbol{\Phi}}$ is an improper correlation or covariance matrix in either solution. The discrepancy value of 0.0336 produced in LMA is smaller than the discrepancy value of 0.0342 based on the solution in Bentler and Jamshidian [29].

Table 11: Comparisons of $\widehat{\boldsymbol{\Lambda}}$, $\widehat{\boldsymbol{\Phi}}$, $\widehat{\boldsymbol{\delta}}$ and $F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ for Bentler Model II

	Bentler Model II	
	In LMA	In Bentler and Jamshidian [29]
$\widehat{\boldsymbol{\Lambda}}$	$\begin{pmatrix} 0.3148 & 0.00 & 0.00 \\ 0.2103 & 0.00 & 0.00 \\ 0.0859 & 0.00 & 0.00 \\ 0.00 & 1.2261 & 0.00 \\ 0.00 & 1.1047 & 0.00 \\ 0.00 & 1.2019 & 0.00 \\ 0.00 & 0.00 & 1.0183 \\ 0.00 & 0.00 & 1.0790 \\ 0.00 & 0.00 & 0.2245 \end{pmatrix}$	$\begin{pmatrix} 0.34 & 0.00 & 0.00 \\ 0.22 & 0.00 & 0.00 \\ 0.09 & 0.00 & 0.00 \\ 0.00 & 1.24 & 0.00 \\ 0.00 & 1.12 & 0.00 \\ 0.00 & 1.22 & 0.00 \\ 0.00 & 0.00 & 1.05 \\ 0.00 & 0.00 & 1.12 \\ 0.00 & 0.00 & 0.22 \end{pmatrix}$
$\widehat{\boldsymbol{\Phi}}$	$\begin{pmatrix} 1.0000 & 2.5783 & 3.2946 \\ 2.5783 & 1.0000 & 0.8264 \\ 3.2946 & 0.8264 & 1.0000 \end{pmatrix}$	$\begin{pmatrix} 1.00 & 2.34 & 2.99 \\ 2.34 & 1.00 & 0.75 \\ 2.99 & 0.75 & 1.00 \end{pmatrix}$
$\widehat{\boldsymbol{\delta}}$	$\begin{pmatrix} 5.6078 \\ 0.1998 \\ -2.8076 \end{pmatrix}$	$\begin{pmatrix} 5.18 \\ 0.27 \\ -2.45 \end{pmatrix}$
$F(\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$	0.0336	0.0342

5.3.3. Other Analysis and Model Fits in 4SP

There are two different models, Bentler Model III and Bentler Model IV, discussed in the following sections. Each model is fit by 4SP in an attempt to produce a proper estimate for $\mathbf{\Phi}$.

5.3.4. Bentler Model III

5.3.4.1. Constraints Imposed On $\mathbf{\Lambda}$: The constraints imposed on $\mathbf{\Lambda}$ remain the same as those of Bentler Model II. The corresponding values of $\widehat{\mathbf{\Lambda}}$ are provided in Table 12.

5.3.4.2. Constraints Imposed On $\mathbf{\Phi}$: In Bentler Model III, the constraints that are imposed on $\boldsymbol{\delta}$ are motivated by the structure of $\widehat{\boldsymbol{\delta}}$ as seen in Table 11. Structure 2a from Table 4 is used to parameterize $\boldsymbol{\delta}$, so that all the entries in $\boldsymbol{\delta}$ are positive and exponentially decreasing. The matrices $\mathbf{T}_1$ and $\mathbf{T}_2$ are

$$\mathbf{T}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and } \mathbf{T}_2 = \begin{pmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}. \tag{177}$$

The corresponding values of $\widehat{\mathbf{\Phi}}$ and $\widehat{\boldsymbol{\delta}}$ are provided in Table 12.

5.3.4.3. Constraints Imposed On $\mathbf{\Psi}$: The constraints imposed on $\mathbf{\Psi}$ remain the same as those of Bentler Model I. That is, the diagonal entries of $\mathbf{\Psi}$ are constrained to be identical. The corresponding values of $\widehat{\mathbf{\Psi}}$ are provided in Table 12.

Table 12: Comparisons of $\widehat{\mathbf{\Lambda}}$, $\widehat{\mathbf{\Phi}}$, $\widehat{\boldsymbol{\delta}}$, $\widehat{\mathbf{\Psi}}$ and $F(\mathbf{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$ for Bentler Model III and IV

	Bentler Model III	Bentler Model IV
$\widehat{\mathbf{\Lambda}}$	$\begin{pmatrix} 0.6081 & 0 & 0 \\ 0.3466 & 0 & 0 \\ 0.1338 & 0 & 0 \\ 0 & 1.2748 & 0 \\ 0 & 1.1473 & 0 \\ 0 & 1.2493 & 0 \\ 0 & 0 & 1.1029 \\ 0 & 0 & 1.1690 \\ 0 & 0 & 0.2426 \end{pmatrix}$	$\begin{pmatrix} 0.5947 & 0 & 0 \\ 0.3390 & 0 & 0 \\ 0.1308 & 0 & 0 \\ 0 & 1.2409 & 0 \\ 0 & 1.1168 & 0 \\ 0 & 1.2160 & 0 \\ 0 & 0 & 1.0248 \\ 0 & 0 & 1.0863 \\ 0 & 0 & 0.2255 \end{pmatrix}$
$\widehat{\mathbf{\Phi}}$	$\begin{pmatrix} 1.0000 & 0.8934 & 0.9191 \\ 0.8934 & 1.0000 & 0.7223 \\ 0.9191 & 0.7223 & 1.0000 \end{pmatrix}$	$\begin{pmatrix} 1.0000 & 0.8848 & 0.8848 \\ 0.8848 & 1.0000 & 0.8848 \\ 0.8848 & 0.8848 & 1.0000 \end{pmatrix}$
$\widehat{\boldsymbol{\delta}}$	$\begin{pmatrix} 2.6927 \\ 0.2785 \\ 0.0288 \end{pmatrix}$	$\begin{pmatrix} 2.7696 \\ 0.1152 \\ 0.1152 \end{pmatrix}$
$\widehat{\mathbf{\Psi}}$	$\text{Diag}(6.0546 \times \mathbf{1}_9)$	$\text{Diag}(6.1229 \times \mathbf{1}_9)$
$F(\mathbf{\Sigma}(\hat{\boldsymbol{\theta}}), \mathbf{S})$	0.1151	0.1243

5.3.5. Bentler Model IV

5.3.5.1. Constraints Imposed On $\mathbf{\Lambda}$: The constraints imposed on $\mathbf{\Lambda}$ remain the same as those of Bentler Model III.

5.3.5.2. Constraints Imposed On $\mathbf{\Phi}$: Structure 1a as illustrated in Table 4 is used to parameterize $\boldsymbol{\delta}$, so that $\mathbf{m} = [1\ 2]'$, where $\mathbf{m}$ is the vector of multiplicities of $\boldsymbol{\delta}$. The matrix $\mathbf{T}_2$ is

$$\mathbf{T}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}. \tag{178}$$

5.3.5.3. Constraints Imposed On $\mathbf{\Psi}$: The constraints imposed on $\mathbf{\Psi}$ remain the same as those of Bentler Model III.

Bentler Model IV is fit by 4SP. The estimates for $\mathbf{\Lambda}$, $\mathbf{\Phi}$, $\boldsymbol{\delta}$, $\mathbf{\Psi}$ and $F(\mathbf{\Sigma}, \mathbf{S})$ are provided in Table 12. In SAS and R, in order to parameterize $\boldsymbol{\delta}$ with $\mathbf{m} = [1\ 2]'$, the following theorem is utilized.

Theorem 36. [A Well-Known Result]. *Let* $\mathbf{A}$ *be a* $q \times q$ *correlation matrix. If* $\mathbf{A}$ *is compound symmetric, then the multiplicity of an eigenvalue of* $\mathbf{A}$ *is* 1 *or* $q - 1$.

In SAS and R, $\mathbf{\Phi}$ is parameterized as a compound symmetric matrix. This is equivalent to parameterizing $\boldsymbol{\delta}$ with $\mathbf{m} = [1\ 2]'$ by Theorem 36. The estimates generated by SAS and R are identical to those generated by 4SP in Table 12.

CHAPTER 6

SIMULATION

The purpose of the simulation is to examine the properties of the proposed inference methods discussed in Chapter 4 when sample size is finite. In this chapter, both conventional models (CM) for confirmatory factor analysis used in R [30] and proposed models (PM) introduced in this thesis are employed. The following simulation results are provided:

(1). proportions of proper estimates for $\mathbf{\Phi}$ and $\mathbf{\Psi}$ under conventional models and proposed models, respectively;

(2). average coverage rates of upper one-sided nominal 95% confidence intervals, lower one-sided nominal 95% confidence intervals, and two-sided nominal 95% confidence intervals, regarding

   (a) nonzero elements in $\mathbf{\Lambda}$, distinct non-unit elements $\mathbf{\Phi}$, and distinct diagonal elements in $\mathbf{\Psi}$, respectively, under conventional models; and

   (b) nonzero elements in $\mathbf{\Lambda}$, diagonal elements in $\mathbf{\Delta}$, identified elements in $\mathbf{\Gamma}$, and distinct diagonal elements in $\mathbf{\Psi}$, respectively, under proposed models;

(3). average of all the ratios of widths of two-sided nominal 95% confidence intervals, between proposed models and conventional models, regarding (a) nonzero elements in $\mathbf{\Lambda}$, and (b) distinct diagonal elements in $\mathbf{\Psi}$, repectively;

(4). empirical test sizes of goodness-of-fit tests for conventional models and proposed models;

(5). empirical test sizes of model comparison tests, comparing proposed models with conventional models.

Both (4) and (5) above are performed using Browne's Residual-Based Test in Theorem 34 and the Satterthwaite approximation in Corollary 33.2.

Further discussions are given on the advantages and disadvantages of employing proposed models, compared to conventional models, in terms of (1), (2), and (3) listed above, are given. Analysis on the closeness between $1-\alpha$ and average coverage rates in (2); and the closeness between α and empirical test sizes in (4) and in (5) are provided.

6.1. Data Generation

Recall that the factor analysis model (3) in § 1.2 for $\mathbf{y}_i$ is

$$\mathbf{y}_i = \boldsymbol{\mu}_i + \boldsymbol{\Lambda}\mathbf{f}_i + \boldsymbol{\epsilon}_i,$$

where

$$\mathrm{E}(\mathbf{f}_i) = \mathbf{0}, \quad \mathrm{Var}(\mathbf{f}_i) = \boldsymbol{\Phi}, \quad \mathrm{E}(\boldsymbol{\epsilon}_i) = \mathbf{0}, \quad \mathrm{Var}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Psi},$$
$$\boldsymbol{\Psi} = \operatorname{diag}\boldsymbol{\psi} = \begin{pmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{pmatrix}, \text{ and } \mathrm{Cov}(\mathbf{f}_i, \boldsymbol{\epsilon}_i) = \mathbf{0}.$$

Accordingly, to generate a random sample $\mathbf{Y}$ of size $\mathbf{N}$ is to have

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}_1' \\ \mathbf{y}_2' \\ \vdots \\ \mathbf{y}_N' \end{pmatrix},$$

where $\mathbf{y}_i$ is a $p \times 1$ random vector of observed responses for a single subject. One thousand trials were conducted for $N \in \{50, 300\}$.

Population parameter values in $\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Psi} = \mathbf{\Lambda\Gamma\Delta\Gamma}'\mathbf{\Lambda}' + \mathbf{\Psi}$ for the simulation study are listed as follows, where $p = 9$ and $q = 3$.

1. $\mathbf{\Lambda} = 0.5 * (\mathbf{I}_q \otimes \mathbf{1}_3)$;

2. (a) $\boldsymbol{\delta}_1 = (300/111 \;\; 30/111 \;\; 3/111)'$, $\mathbf{\Delta}_1 = \text{Diag}(\boldsymbol{\delta}_1)$,

$$\mathbf{\Gamma}_1 = \begin{pmatrix} 0.5641 & -0.7071 & -0.4264 \\ 0.5641 & 0.7071 & -0.4264 \\ 0.6030 & 0 & 0.7977 \end{pmatrix},$$

$$\mathbf{\Phi_1} = \mathbf{\Gamma}_1\mathbf{\Delta}_1\mathbf{\Gamma}_1' = \begin{pmatrix} 1.0000 & 0.7297 & 0.9101 \\ 0.7297 & 1.0000 & 0.9101 \\ 0.9101 & 0.9101 & 1.0000 \end{pmatrix};$$

(b) $\boldsymbol{\delta}_2 = (2.86 \;\; 0.07 \;\; 0.07)'$, $\mathbf{\Delta}_2 = \text{Diag}(\boldsymbol{\delta}_2)$,

$$\mathbf{\Gamma}_2 = \begin{pmatrix} 0.5774 & -0.4082 & -0.7071 \\ 0.5774 & 0.8165 & 0 \\ 0.5774 & -0.4082 & 0.7071 \end{pmatrix},$$

$$\mathbf{\Phi_2} = \mathbf{\Gamma}_2\mathbf{\Delta}_2\mathbf{\Gamma}_2' = \begin{pmatrix} 1.00 & 0.93 & 0.93 \\ 0.93 & 1.00 & 0.93 \\ 0.93 & 0.93 & 1.00 \end{pmatrix}; \text{ and}$$

3. $\mathbf{\Psi}_1 = \text{Diag}(\boldsymbol{\psi}_1) = \text{Diag}(0.0324)$, and $\mathbf{\Psi}_2 = \text{Diag}(\boldsymbol{\psi}_2) = \text{Diag}(0.9025)$.

In 2(a), $\boldsymbol{\delta}_1$ follows structure 2a from Table 4 with

$$\mathbf{T}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and } \mathbf{T}_2 = \begin{pmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}, \tag{179}$$

Table 13: Population Model I, II, III and IV

Population Model I	$\mathbf{\Sigma}_1 = \mathbf{\Lambda\Phi}_1\mathbf{\Lambda}' + \mathbf{\Psi}_1 = \mathbf{\Lambda\Gamma}_1\mathbf{\Delta}_1\mathbf{\Gamma}_1'\mathbf{\Lambda}' + \mathbf{\Psi}_1$
Population Model II	$\mathbf{\Sigma}_2 = \mathbf{\Lambda\Phi}_1\mathbf{\Lambda}' + \mathbf{\Psi}_2 = \mathbf{\Lambda\Gamma}_1\mathbf{\Delta}_1\mathbf{\Gamma}_1'\mathbf{\Lambda}' + \mathbf{\Psi}_2$
Population Model III	$\mathbf{\Sigma}_3 = \mathbf{\Lambda\Phi}_2\mathbf{\Lambda}' + \mathbf{\Psi}_1 = \mathbf{\Lambda\Gamma}_2\mathbf{\Delta}_2\mathbf{\Gamma}_2'\mathbf{\Lambda}' + \mathbf{\Psi}_1$
Population Model IV	$\mathbf{\Sigma}_4 = \mathbf{\Lambda\Phi}_2\mathbf{\Lambda}' + \mathbf{\Psi}_2 = \mathbf{\Lambda\Gamma}_2\mathbf{\Delta}_2\mathbf{\Gamma}_2'\mathbf{\Lambda}' + \mathbf{\Psi}_2$

so that all the entries in $\boldsymbol{\delta}_1$ are positive and exponentially decreasing. In 2(b), $\boldsymbol{\delta}_2$ follows structure 4 from Table 4 with $\mathbf{m} = [1\ 2]'$ and

$$\mathbf{T}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}. \tag{180}$$

In 3, $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$ both follow structure 5 in Table 5 with $\mathbf{T}_4 = \mathbf{1}_9$ so that all diagonal entries in $\mathbf{\Psi}_1$ or $\mathbf{\Psi}_2$ are positive and homogeneous. The values of $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ are chosen so that the communality at $\boldsymbol{\psi}_1 = 0.0324$ is 0.8853 and the communality at $\boldsymbol{\psi}_2 = 0.9025$ is 0.2169, where the communality is defined in (15).

According to the different parameter values listed above, population models for $\mathbf{\Sigma}$, $\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Psi} = \mathbf{\Lambda\Gamma\Delta\Gamma}'\mathbf{\Lambda}' + \mathbf{\Psi}$, can be classified into four categories as in Table 13.

The random vectors $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ were sampled from the five distributions in Table 14 with assigned $\mathbf{\Phi}$ and $\mathbf{\Psi}$ for $\text{Var}(\mathbf{f}_i)$ and $\text{Var}(\boldsymbol{\epsilon}_i)$, respectively.

To generate independent $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$, or uncorrelated but not independent $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$, use

$$\begin{pmatrix} \mathbf{f}_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} = \mathbf{Q}\mathbf{z}_i, \tag{181}$$

where $\mathbf{Q}$ is an orthogonal matrix, $\mathbf{z}_i$ are iid and the entries of $\mathbf{z}_i$ are either $\text{N}(0,1)$, $\text{Unif}(-\sqrt{3},\sqrt{3})$, or centered and scaled χ^2_2. To create independent vectors, $\mathbf{Q}$ in

Table 14: Five Distributions

Distribution	Description
Distribution 1	$\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are independent; and are sampled from multivariate normal distribution;
Distribution 2	$\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are independent; and were sampled from uniform distribution;
Distribution 3	$\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are independent; and were sampled from χ^2 distribution;
Distribution 4	$\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent; and were sampled from uniform distribution;
Distribution 5	$\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent; and were sampled from χ^2 distribution.

(181) is set to be $\mathbf{Q} = \mathbf{I}_{q+p}$; for non-independence, the columns of $\mathbf{Q}$ are normalized coefficients of orthogonal polynomials.

The following section gives a table of fitted models, both conventional models and proposed models, to Population Model I, II, III, and IV, respectively.

6.2. Model Fitting

The detailed parameterizations for $\boldsymbol{\Lambda}$, $\boldsymbol{\Delta}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\Psi}$ in proposed models are provided in Chapter 2. In conventional models, the parameterization for $\boldsymbol{\Lambda}_*$ is provided in Section 2.4.1 and the parameterization for $\boldsymbol{\Psi}$ is provided in Section 2.3.4.

The parameterization for $\boldsymbol{\Phi}$ in the conventional models is

$$\text{vec}(\boldsymbol{\Phi}) = \text{vec}(\mathbf{I}_q) + 2\mathbf{N}_q\mathbf{L}'\boldsymbol{\theta}_\phi, \tag{182}$$

where $\mathbf{L}$ is the $q(q-1)/2 \times q^2$ strictly lower triangular elimination matrix (§6.5 in Magnus [60]). Thus, $\dim(\boldsymbol{\theta}_\phi) = 3$. If $\boldsymbol{\Phi}$ is compound symmetric, then in conventional models $\boldsymbol{\Phi}$ is parameterized as

$$\text{vec}(\boldsymbol{\Phi}) = \text{vec}(\mathbf{I}_q) + 2\mathbf{N}_q\mathbf{L}'\mathbf{1}_{q(q-1)/2}\boldsymbol{\theta}_\phi. \tag{183}$$

Accordingly, $\dim(\boldsymbol{\theta}_\phi) = 1$. Denote the following models as follows:

Model a. the saturated model; the saturated model parameterized as $\operatorname{vec}(\mathbf{\Sigma}) = \mathbf{D}_p\boldsymbol{\theta}_{\boldsymbol{\sigma}}$, where $\boldsymbol{\theta}_{\boldsymbol{\sigma}} = (\mathbf{D}_p{}'\mathbf{D}_p)^{-1}\mathbf{D}_p{}' \operatorname{vec}(\mathbf{\Sigma}) = \operatorname{vech}(\mathbf{\Sigma})$ and $\mathbf{D}_p$ is defined in Table 56. Corollary 33.1 includes more details about the saturated model.

Model b. conventional model; symmetric structure for $\mathbf{\Phi}$ parameterized as in (182), and homogeneous structure for $\boldsymbol{\psi}$, which is parameterized as structure 1a in Table 5 with $\mathbf{T}_4 = \mathbf{1}_9$;

Model c. conventional model; symmetric structure for $\mathbf{\Phi}$ parameterized as in (182), and heterogeneous structure for $\boldsymbol{\psi}$, which is parameterized as structure 1a in Table 5 with $\mathbf{T}_4 = \mathbf{I}_9$;

Model d. proposed model; structure 2a in Table 4 for $\boldsymbol{\delta}$ with

$$\mathbf{T}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{T}_2 = \begin{pmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 0 \end{pmatrix},$$

and structure 5 in Table 5 for $\boldsymbol{\psi}$ with $\mathbf{T}_4 = \mathbf{1}_9$;

Model e. proposed model; structure 2a in Table 4 for $\boldsymbol{\delta}$ with

$$\mathbf{T}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{T}_2 = \begin{pmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 0 \end{pmatrix},$$

and structure 5 in Table 5 for $\boldsymbol{\psi}$ with $\mathbf{T}_4 = \mathbf{I}_9$;

Model f. conventional model; compound symmetric structure for $\mathbf{\Phi}$ parameterized as in (183), and homogeneous structure for $\boldsymbol{\psi}$, which is parameterized as structure 1a in Table 5 with $\mathbf{T}_4 = \mathbf{1}_9$;

Table 15: Model Fitting for Population Model I, II, III and IV

	Conventional Models	**Proposed Models**
Population Model I	Model b and Model c	Model d and Model e
Population Model II	Model b and Model c	Model d and Model e
Population Model III	Model b, Model c, Model f and Model g	Model h and Model i
Population Model IV	Model b, Model c, Model f and Model g	Model h and Model i

Model g. conventional model; compound symmetric structure for $\mathbf{\Phi}$ parameterized as in (183), and heterogeneous structure for $\boldsymbol{\psi}$, which is parameterized as structure 1a in Table 5 with $\mathbf{T}_4 = \mathbf{I}_9$;

Model h. proposed model; structure 4 from Table 4 for $\boldsymbol{\delta}$ with $\mathbf{T}_1 = [\ \ ]$, $\mathbf{m} = [1\ 2]'$ and

$$\mathbf{T}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix},$$

and structure 5 in Table 5 for $\boldsymbol{\psi}$ with $\mathbf{T}_4 = \mathbf{1}_9$; and

Model i. proposed model; structure 4 from Table 4 for $\boldsymbol{\delta}$ with $\mathbf{T}_1 = [\ \ ]$, $\mathbf{m} = [1\ 2]'$ and

$$\mathbf{T}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix},$$

and structure 5 in Table 5 for $\boldsymbol{\psi}$ with $\mathbf{T}_4 = \mathbf{I}_9$.

Table 15 provides fitted model, both conventional models and proposed models, to Population Model I, II, III, and IV. The notations are consistent with the notations listed above.

For notational convenience, fitted models to Population Model I, Model b, Model c, Model d and Model e, are denoted by Ib, Ic, Id, and Ie. Likewise, fitted models to Population Model II, Model b, Model c, Model d and Model e, are denoted by IIb, IIc, IId, and IIe; fitted models to Population Model III, Model b, Model c, Model f, Model g, Model h and Model i, are denoted by IIIb, IIIc, IIIf, IIIg, IIIh and IIIi; and fitted models to Population Model IV, Model b, Model c, Model f, Model g, Model h and Model i, are denoted by IVb, IVc, IVf, IVg, IVh and IVi.

6.3. Simulation Results

Simulation results that include: (1) proportions of proper estimates; (2) average coverage rates; (3) average mean ratios of the width of two-sided CI; (4) empirical test sizes of goodness-of-fit tests; and (5) empirical test sizes of model comparison tests, are discussed according the five distributions in Table 14 at $N = 300$.

6.3.1. Proportion of Proper Solutions

Table 16, Table 17, Table 18, Table 19, and Table 20 display the proportions of proper solutions by fitting conventional models, Ib, Ic, IIb, and IIc, and by fitting proposed models, Id, Ie, IId, and IIe. Under Population Model I, both conventional models and proposed models generate 100% proper solutions; under Population Model II, proposed models generate much higher proportions of proper solutions, compared to conventional models. Note that improper solutions in the R package sem can be due to a failure to converge (Fox, [31]). In this thesis, if the *sem* package [31] in R provides a solution, then the solution is classified as a proper or an improper solution. The solution is counted as a failure to converge only when R stops and no solution is provided.

Table 16: Proportions of Proper Solutions, Improper Solutions and Failure to Converge under Population Model I and II at $N = 300$ Distribution 1: Multivariate Normal

	Population Model I				**Population Model II**			
	CM		**PM**		**CM**		**PM**	
	Ib	**Ic**	**Id**	**Ie**	**IIb**	**IIc**	**IId**	**IIe**
Proper Solutions	1.0000	1.0000	1.0000	1.0000	0.4210	0.4500	0.9060	0.8980
Improper $\widehat{\boldsymbol{\Psi}}$	0.0000	0.0000	0.0000	0.0000	0.0940	0.0000	0.0000	0.0000
Improper $\widehat{\boldsymbol{\Phi}}$	0.0000	0.0000	0.0000	0.0000	0.4850	0.5500	0.0000	0.0000
Improper $\widehat{\boldsymbol{\Psi}}$ **and** $\widehat{\boldsymbol{\Phi}}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Failure to Converge	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0940	0.1020

Table 17: Proportions of Proper Solutions, Improper Solutions and Failure to Converge under Population Model I and II at $N = 300$ Distribution 2: Uniform

	Population Model I				**Population Model II**			
	CM		**PM**		**CM**		**PM**	
	Ib	**Ic**	**Id**	**Ie**	**IIb**	**IIc**	**IId**	**IIe**
Proper Solutions	1.0000	1.0000	1.0000	1.0000	0.4380	0.4870	0.9060	0.9090
Improper $\widehat{\boldsymbol{\Psi}}$	0.0000	0.0000	0.0000	0.0000	0.0740	0.0000	0.0000	0.0000
Improper $\widehat{\boldsymbol{\Phi}}$	0.0000	0.0000	0.0000	0.0000	0.4870	0.5130	0.0000	0.0000
Improper $\widehat{\boldsymbol{\Psi}}$ **and** $\widehat{\boldsymbol{\Phi}}$	0.0000	0.0000	0.0000	0.0000	0.0010	0.0000	0.0000	0.0000
Failure to Converge	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0940	0.0910

Table 18: Proportions of Proper Solutions, Improper Solutions and Failure to Converge under Population Model I and II at $N = 300$ Distribution 3: χ^2

	Population Model I				**Population Model II**			
	CM		**PM**		**CM**		**PM**	
	Ib	**Ic**	**Id**	**Ie**	**IIb**	**IIc**	**IId**	**IIe**
Proper Solutions	1.0000	1.0000	1.0000	1.0000	0.4690	0.4640	0.9290	0.8790
Improper $\widehat{\boldsymbol{\Psi}}$	0.0000	0.0000	0.0000	0.0000	0.1130	0.0000	0.0000	0.0000
Improper $\widehat{\boldsymbol{\Phi}}$	0.0000	0.0000	0.0000	0.0000	0.4140	0.5360	0.0000	0.0000
Improper $\widehat{\boldsymbol{\Psi}}$ **and** $\widehat{\boldsymbol{\Phi}}$	0.0000	0.0000	0.0000	0.0000	0.0040	0.0000	0.0000	0.0000
Failure to Converge	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0710	0.1210

Table 19: Proportions of Proper Solutions, Improper Solutions and Failure to Converge under Population Model I and II at $N = 300$ Distribution 4: Uniform, $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent

	Population Model I				Population Model II			
	CM		PM		CM		PM	
	Ib	Ic	Id	Ie	IIb	IIc	IId	IIe
Proper Solutions	1.0000	1.0000	1.0000	1.0000	0.4450	0.4910	0.9310	0.9260
Improper $\widehat{\boldsymbol{\Psi}}$	0.0000	0.0000	0.0000	0.0000	0.1030	0.0000	0.0000	0.0000
Improper $\widehat{\boldsymbol{\Phi}}$	0.0000	0.0000	0.0000	0.0000	0.4510	0.5090	0.0000	0.0000
Improper $\widehat{\boldsymbol{\Psi}}$ and $\widehat{\boldsymbol{\Phi}}$	0.0000	0.0000	0.0000	0.0000	0.0010	0.0000	0.0000	0.0000
Failure to Converge	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0690	0.0740

Table 20: Proportions of Proper Solutions, Improper Solutions and Failure to Converge under Population Model I and II at $N = 300$ Distribution 5: χ^2, $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent

	Population Model I				Population Model II			
	CM		PM		CM		PM	
	Ib	Ic	Id	Ie	IIb	IIc	IId	IIe
Proper Solutions	1.0000	1.0000	1.0000	1.0000	0.3760	0.3980	0.8920	0.8730
Improper $\widehat{\boldsymbol{\Psi}}$	0.0000	0.0000	0.0000	0.0000	0.1130	0.0000	0.0000	0.0000
Improper $\widehat{\boldsymbol{\Phi}}$	0.0000	0.0000	0.0000	0.0000	0.5100	0.6020	0.0000	0.0000
Improper $\widehat{\boldsymbol{\Psi}}$ and $\widehat{\boldsymbol{\Phi}}$	0.0000	0.0000	0.0000	0.0000	0.0010	0.0000	0.0000	0.0000
Failure to Converge	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1080	0.1270

6.3.2. Success Rate and Coverage Rate

Success rate and coverage rate are defined as follows:

$$\text{success rate} = \frac{\text{the number of CIs that capture the true parameter}}{\text{the total number of trials}},$$

$$\text{and} \qquad (184)$$

$$\text{coverage rate} = \frac{\text{the number of CIs that capture the true parameter}}{\text{the total number of trials that generate proper solutions}},$$

where a confidence interval is computed only if a proper solution is attained.

Under Distribution 1: Multivariate Normal, Table 21, Table 22, Table 23 and Table 24 contain average success rate and average coverage rate of nominal 95% upper one-sided confidence interval, lower one-sided confidence interval and two-sided confidence intervals.

Under Distribution 2: Uniform, Table 25, Table 26, Table 27 and Table 28 contain average success rate and average coverage rate of nominal 95% upper one-sided confidence interval, lower one-sided confidence interval and two-sided confidence intervals.

Under Distribution 3: χ^2, Table 29, Table 30, Table 31 and Table 32 contain average success rate and average coverage rate of nominal 95% upper one-sided confidence interval, lower one-sided confidence interval and two-sided confidence intervals.

Under Distribution 4: Uniform, $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent, Table 33, Table 34, Table 35 and Table 36 contain average success rate and average coverage rate of nominal 95% upper one-sided confidence interval, lower one-sided confidence interval and two-sided confidence intervals.

Under Distribution 5: χ^2, $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent, Table 37, Table 38, Table 39 and Table 40 contain average success rate and average coverage rate of nominal 95% upper one-sided confidence interval, lower one-sided confidence interval and two-sided confidence intervals.

Average success rate is averaging all the success rates for nonzero elements in $\boldsymbol{\lambda}$, distinct non-unit elements in $\boldsymbol{\phi}$, all the elements in $\boldsymbol{\delta}$, identified elements in $\boldsymbol{\gamma}$, and distinct elements in $\boldsymbol{\psi}$, respectively. Similarly, average coverage rate is averaging all coverage rates for nonzero elements in $\boldsymbol{\lambda}$, distinct non-unit elements in $\boldsymbol{\phi}$, all the elements in $\boldsymbol{\delta}$, identified elements in $\boldsymbol{\gamma}$, and distinct elements in $\boldsymbol{\psi}$, respectively.

Overall, under Population Model I, there is not much difference between conventional model and proposed models, in terms of success rates and coverage rates. Under Population Model II, proposed models generate much higher success rate than conventional models, overall. However, under Population Model II, there is not much difference in coverage rate between proposed models and conventional models.

6.3.3. Average of the Ratios of Widths of Two-Sided CI

The ratio for nonzero elements in $\boldsymbol{\lambda}$ and distinct elements in $\boldsymbol{\psi}$ is computed as follows:

$$\text{ratio of widths of two-sided CI} = \frac{\text{widths of two-sided CI from PM}}{\text{widths of two-sided CI from CM}}. \tag{185}$$

Accordingly, average of the ratios of widths of two-sided confidence intervals is averaging all ratios computed for nonzero elements in $\boldsymbol{\lambda}$ and distinct elements in $\boldsymbol{\psi}$, respectively. Note that all the two-sided CI is of nominal 95%.

Based on Table 41, Table 42, Table 43, Table 44, and Table 45, under Population Model I and II, proposed models generate slightly narrower confidence intervals than conventional models on average, for elements in $\boldsymbol{\lambda}$ and $\boldsymbol{\psi}$. The only exception is under Population Model II, Model c (conventional model), generates slightly narrower confidence intervals than Model e (proposed model).

6.3.4. Empirical Test Sizes for the Goodness-of-fit Test

The empirical test size (ETS) is the proportion of simulation rejections when the null hypothesis, $\mathbf{H}_0$, is assumed to be true.

6.3.4.1. Two Ways to Compute ETS: There are two way to compute empirical test size (ETS), which are listed as follows:

$$\begin{aligned} ETS1 &= \frac{\text{the number of rejections}}{\text{the total number of proper solutions}}, \\ &\text{and} \\ ETS2 &= \frac{\text{the number of rejections, improper and failure solutions}}{\text{the total number of trials}}. \end{aligned} \tag{186}$$

Table 46, Table 47, Table 48, Table 49, and Table 50 provide ETS1 and ETS2 for the goodness-of-fit test. The Satterthwaite's approximation procedure generates empirical test sizes that are closer to the nominal test size of $\alpha = 0.05$ compared to Browne's test procedure, overall. However, under Model IIb and Model IId, Browne's test procedure generates ETS1 closer to $\alpha = 0.05$ than the Satterthwaite's approximation procedure. Under Population Model I, there is not much difference between ETS1 and ETS2. But under Population Model II, ETS2 is much higher than $\alpha = 0.05$, especially those under conventional models, because of the high proportion of improper solutions from the conventional models.

In Table 46, Table 47, and Table 48, the titles of those tables are shortened in terms of the specific names of Distribution 1, Distribution 2, and Distribution 3. As defined earlier in Table 14, Distribution 1 is a multivariate normal distribution with $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are independent; Distribution 2 is a uniform distribution with $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are independent; and Distribution 3 is a χ^2 with $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are independent.

Table 21: Average of Success Rates under Population Model I at $N = 300$ Distribution 1: Multivariate Normal

		upper one-sided CI	**lower one-sided CI**	**two-sided CI**
CM	**Model b ($\boldsymbol{\lambda}$)**	0.9319	0.9553	0.9393
	Model b ($\boldsymbol{\phi}$)	0.9657	0.9303	0.9467
	Model b ($\boldsymbol{\psi}$)	0.9400	0.9620	0.9470
	Model c ($\boldsymbol{\lambda}$)	0.9319	0.9556	0.9410
	Model c ($\boldsymbol{\phi}$)	0.9640	0.9307	0.9463
	Model c ($\boldsymbol{\psi}$)	0.9226	0.9678	0.9408
PM	**Model d ($\boldsymbol{\lambda}$)**	0.9334	0.9546	0.9418
	Model d ($\boldsymbol{\delta}$)	0.9453	0.9573	0.9497
	Model d ($\boldsymbol{\gamma}$)	0.9520	0.9527	0.9489
	Model d ($\boldsymbol{\psi}$)	0.9400	0.9590	0.9510
	Model e ($\boldsymbol{\lambda}$)	0.9330	0.9539	0.9423
	Model e ($\boldsymbol{\delta}$)	0.9430	0.9577	0.9500
	Model e ($\boldsymbol{\gamma}$)	0.9507	0.9510	0.9462
	Model e ($\boldsymbol{\psi}$)	0.9217	0.9670	0.9424

Table 22: Average of Success Rates under Population Model II at $N = 300$ Distribution 1: Multivariate Normal

		upper one-sided CI	lower one-sided CI	two-sided CI
CM	**Model b (λ)**	0.4048	0.3922	0.3976
	Model b (ϕ)	0.3870	0.4147	0.4013
	Model b (ψ)	0.3720	0.4170	0.3930
	Model c (λ)	0.4309	0.4212	0.4240
	Model c (ϕ)	0.4087	0.4460	0.4250
	Model c (ψ)	0.4104	0.4377	0.4196
PM	**Model d (λ)**	0.8676	0.8538	0.8602
	Model d (δ)	0.8743	0.8717	0.8757
	Model d (γ)	0.8368	0.8423	0.8133
	Model d (ψ)	0.8140	0.8880	0.8470
	Model e (λ)	0.8584	0.8488	0.8514
	Model e (δ)	0.8657	0.8630	0.8620
	Model e (γ)	0.8250	0.8306	0.7970
	Model e (ψ)	0.8286	0.8700	0.8442

Table 23: Average of Coverage Rates under Population Model I at $N = 300$ Distribution 1: Multivariate Normal

		upper one-sided CI	lower one-sided CI	two-sided CI
CM	**Model b (λ)**	0.9319	0.9553	0.9393
	Model b (ϕ)	0.9657	0.9303	0.9467
	Model b (ψ)	0.9400	0.9620	0.9470
	Model c (λ)	0.9319	0.9556	0.9410
	Model c (ϕ)	0.9640	0.9307	0.9463
	Model c (ψ)	0.9226	0.9678	0.9408
PM	**Model d (λ)**	0.9334	0.9546	0.9418
	Model d (δ)	0.9453	0.9573	0.9497
	Model d (γ)	0.9520	0.9527	0.9489
	Model d (ψ)	0.9400	0.9590	0.9510
	Model e (λ)	0.9330	0.9539	0.9423
	Model e (δ)	0.9430	0.9577	0.9500
	Model e (γ)	0.9507	0.9510	0.9462
	Model e (ψ)	0.9217	0.9670	0.9424

Table 24: Average of Coverage Rates under Population Model II at $N = 300$ Distribution 1: Multivariate Normal

		upper one-sided CI	**lower one-sided CI**	**two-sided CI**
CM	**Model b (λ)**	0.9615	0.9316	0.9443
	Model b (ϕ)	0.9192	0.9850	0.9533
	Model b (ψ)	0.8836	0.9905	0.9335
	Model c (λ)	0.9575	0.9360	0.9422
	Model c (ϕ)	0.9081	0.9911	0.9444
	Model c (ψ)	0.9121	0.9726	0.9323
PM	**Model d (λ)**	0.9576	0.9424	0.9495
	Model d (δ)	0.9650	0.9621	0.9665
	Model d (γ)	0.9236	0.9297	0.8977
	Model d (ψ)	0.8985	0.9801	0.9349
	Model e (λ)	0.9560	0.9452	0.9482
	Model e (δ)	0.9640	0.9610	0.9599
	Model e (γ)	0.9187	0.9249	0.8875
	Model e (ψ)	0.9227	0.9688	0.9401

Table 25: Average of Success Rates under Population Model I at $N = 300$ Distribution 2: Uniform

		upper one-sided CI	**lower one-sided CI**	**two-sided CI**
CM	**Model b (λ)**	0.9466	0.9596	0.9542
	Model b (ϕ)	0.9657	0.9227	0.9410
	Model b (ψ)	0.9360	0.9580	0.9480
	Model c (λ)	0.9478	0.9583	0.9532
	Model c (ϕ)	0.9643	0.9227	0.9393
	Model c (ψ)	0.9374	0.9624	0.9467
PM	**Model d (λ)**	0.9480	0.9568	0.9533
	Model d (δ)	0.9393	0.9560	0.9487
	Model d (γ)	0.9430	0.9423	0.9391
	Model d (ψ)	0.9360	0.9620	0.9480
	Model e (λ)	0.9482	0.9573	0.9538
	Model e (δ)	0.9380	0.9543	0.9450
	Model e (γ)	0.9423	0.9429	0.9366
	Model e (ψ)	0.9383	0.9621	0.9471

Table 26: Average of Success Rates under Population Model II at $N = 300$ Distribution 2: Uniform

		upper one-sided CI	**lower one-sided CI**	**two-sided CI**
CM	**Model b (λ)**	0.4238	0.4080	0.4149
	Model b (ϕ)	0.3993	0.4353	0.4167
	Model b (ψ)	0.3830	0.4370	0.4130
	Model c (λ)	0.4684	0.4550	0.4592
	Model c (ϕ)	0.4347	0.4837	0.4590
	Model c (ψ)	0.4496	0.4723	0.4607
PM	**Model d (λ)**	0.8711	0.8516	0.8583
	Model d (δ)	0.8730	0.8733	0.8750
	Model d (γ)	0.8310	0.8372	0.8019
	Model d (ψ)	0.8190	0.8950	0.8570
	Model e (λ)	0.8679	0.8580	0.8606
	Model e (δ)	0.8757	0.8707	0.8720
	Model e (γ)	0.8272	0.8347	0.7947
	Model e (ψ)	0.8486	0.8781	0.8644

Table 27: Average of Coverage Rates under Population Model I at $N = 300$ Distribution 2: Uniform

		upper one-sided CI	**lower one-sided CI**	**two-sided CI**
CM	**Model b (λ)**	0.9466	0.9596	0.9542
	Model b (ϕ)	0.9657	0.9227	0.9410
	Model b (ψ)	0.9360	0.9580	0.9480
	Model c (λ)	0.9478	0.9583	0.9532
	Model c (ϕ)	0.9643	0.9227	0.9393
	Model c (ψ)	0.9374	0.9624	0.9467
PM	**Model d (λ)**	0.9480	0.9568	0.9533
	Model d (δ)	0.9393	0.9560	0.9487
	Model d (γ)	0.9430	0.9423	0.9391
	Model d (ψ)	0.9360	0.9620	0.9480
	Model e (λ)	0.9482	0.9573	0.9538
	Model e (δ)	0.9380	0.9543	0.9450
	Model e (γ)	0.9423	0.9429	0.9366
	Model e (ψ)	0.9383	0.9621	0.9471

Table 28: Average of Coverage Rates under Population Model II at $N = 300$ Distribution 2: Uniform

		upper one-sided CI	**lower one-sided CI**	**two-sided CI**
CM	**Model b (λ)**	0.9675	0.9315	0.9472
	Model b (ϕ)	0.9117	0.9939	0.9513
	Model b (ψ)	0.8744	0.9977	0.9429
	Model c (λ)	0.9619	0.9343	0.9430
	Model c (ϕ)	0.8925	0.9932	0.9425
	Model c (ψ)	0.9231	0.9699	0.9459
PM	**Model d (λ)**	0.9615	0.9399	0.9474
	Model d (δ)	0.9636	0.9639	0.9658
	Model d (γ)	0.9172	0.9241	0.8851
	Model d (ψ)	0.9040	0.9879	0.9459
	Model e (λ)	0.9548	0.9439	0.9467
	Model e (δ)	0.9633	0.9578	0.9593
	Model e (γ)	0.9100	0.9182	0.8742
	Model e (ψ)	0.9335	0.9660	0.9510

Table 29: Average of Success Rates under Population Model I at $N = 300$ Distribution 3: χ^2

		upper one-sided CI	**lower one-sided CI**	**two-sided CI**
CM	**Model b (λ)**	0.9028	0.9652	0.9243
	Model b (ϕ)	0.9597	0.9180	0.9350
	Model b (ψ)	0.9300	0.9620	0.9550
	Model c (λ)	0.9033	0.9664	0.9232
	Model c (ϕ)	0.9597	0.9160	0.9327
	Model c (ψ)	0.8873	0.9780	0.9131
PM	**Model d (λ)**	0.9000	0.9637	0.9193
	Model d (δ)	0.9253	0.9447	0.9257
	Model d (γ)	0.9446	0.9411	0.9341
	Model d (ψ)	0.9330	0.9650	0.9550
	Model e (λ)	0.9014	0.9643	0.9204
	Model e (δ)	0.9257	0.9410	0.9270
	Model e (γ)	0.9441	0.9416	0.9361
	Model e (ψ)	0.8868	0.9783	0.9129

Table 30: Average of Success Rates under Population Model II at $N = 300$ Distribution 3: χ^2

		upper one-sided CI	**lower one-sided CI**	**two-sided CI**
CM	**Model b (λ)**	0.4376	0.4438	0.4387
	Model b (ϕ)	0.4127	0.4653	0.4350
	Model b (ψ)	0.4070	0.4650	0.4220
	Model c (λ)	0.4406	0.4367	0.4369
	Model c (ϕ)	0.4193	0.4600	0.4373
	Model c (ψ)	0.3970	0.4580	0.4137
PM	**Model d (λ)**	0.8689	0.8906	0.8780
	Model d (δ)	0.8903	0.8783	0.8810
	Model d (γ)	0.8396	0.8458	0.8000
	Model d (ψ)	0.8070	0.9150	0.8520
	Model e (λ)	0.8369	0.8349	0.8343
	Model e (δ)	0.8450	0.8410	0.8403
	Model e (γ)	0.8003	0.8072	0.7653
	Model e (ψ)	0.7636	0.8658	0.7936

Table 31: Average of Coverage Rates under Population Model I at $N = 300$ Distribution 3: χ^2

		upper one-sided CI	**lower one-sided CI**	**two-sided CI**
CM	**Model b (λ)**	0.9028	0.9652	0.9243
	Model b (ϕ)	0.9597	0.9180	0.9350
	Model b (ψ)	0.9300	0.9620	0.9550
	Model c (λ)	0.9033	0.9664	0.9232
	Model c (ϕ)	0.9597	0.9160	0.9327
	Model c (ψ)	0.8873	0.9780	0.9131
PM	**Model d (λ)**	0.9000	0.9637	0.9193
	Model d (δ)	0.9253	0.9447	0.9257
	Model d (γ)	0.9446	0.9411	0.9341
	Model d (ψ)	0.9330	0.9650	0.9550
	Model e (λ)	0.9014	0.9643	0.9204
	Model e (δ)	0.9257	0.9410	0.9270
	Model e (γ)	0.9441	0.9416	0.9361
	Model e (ψ)	0.8868	0.9783	0.9129

Table 32: Average of Coverage Rates under Population Model II at $N = 300$ Distribution 3: χ^2

		upper one-sided CI	**lower one-sided CI**	**two-sided CI**
CM	**Model b** ($\boldsymbol{\lambda}$)	0.9330	0.9462	0.9353
	Model b ($\boldsymbol{\phi}$)	0.8799	0.9922	0.9275
	Model b ($\boldsymbol{\psi}$)	0.8678	0.9915	0.8998
	Model c ($\boldsymbol{\lambda}$)	0.9495	0.9411	0.9416
	Model c ($\boldsymbol{\phi}$)	0.9037	0.9914	0.9425
	Model c ($\boldsymbol{\psi}$)	0.8556	0.9871	0.8915
PM	**Model d** ($\boldsymbol{\lambda}$)	0.9353	0.9586	0.9451
	Model d ($\boldsymbol{\delta}$)	0.9584	0.9455	0.9483
	Model d ($\boldsymbol{\gamma}$)	0.9037	0.9104	0.8611
	Model d ($\boldsymbol{\psi}$)	0.8687	0.9849	0.9171
	Model e ($\boldsymbol{\lambda}$)	0.9521	0.9498	0.9492
	Model e ($\boldsymbol{\delta}$)	0.9613	0.9568	0.9560
	Model e ($\boldsymbol{\gamma}$)	0.9105	0.9183	0.8707
	Model e ($\boldsymbol{\psi}$)	0.8687	0.9850	0.9028

Table 33: Average of Success Rates under Population Model I at $N = 300$ Distribution 4: Uniform, $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent

		upper one-sided CI	**lower one-sided CI**	**two-sided CI**
CM	**Model b** ($\boldsymbol{\lambda}$)	0.9424	0.9576	0.9477
	Model b ($\boldsymbol{\phi}$)	0.9610	0.9293	0.9380
	Model b ($\boldsymbol{\psi}$)	0.9420	0.9610	0.9570
	Model c ($\boldsymbol{\lambda}$)	0.9432	0.9571	0.9479
	Model c ($\boldsymbol{\phi}$)	0.9603	0.9317	0.9373
	Model c ($\boldsymbol{\psi}$)	0.9316	0.9656	0.9444
PM	**Model d** ($\boldsymbol{\lambda}$)	0.9000	0.9637	0.9193
	Model d ($\boldsymbol{\delta}$)	0.9407	0.9560	0.9453
	Model d ($\boldsymbol{\gamma}$)	0.9489	0.9511	0.9499
	Model d ($\boldsymbol{\psi}$)	0.9450	0.9630	0.9530
	Model e ($\boldsymbol{\lambda}$)	0.9431	0.9572	0.9480
	Model e ($\boldsymbol{\delta}$)	0.9387	0.9523	0.9427
	Model e ($\boldsymbol{\gamma}$)	0.9472	0.9492	0.9478
	Model e ($\boldsymbol{\psi}$)	0.9317	0.9649	0.9451

Table 34: Average of Success Rates under Population Model II at $N = 300$ Distribution 4: Uniform, $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent

		upper one-sided CI	**lower one-sided CI**	**two-sided CI**
CM	**Model b ($\boldsymbol{\lambda}$)**	0.4237	0.4168	0.4218
	Model b ($\boldsymbol{\phi}$)	0.4080	0.4393	0.4250
	Model b ($\boldsymbol{\psi}$)	0.3870	0.4430	0.4090
	Model c ($\boldsymbol{\lambda}$)	0.4689	0.4619	0.4663
	Model c ($\boldsymbol{\phi}$)	0.4473	0.4860	0.4647
	Model c ($\boldsymbol{\psi}$)	0.4513	0.4769	0.4649
PM	**Model d ($\boldsymbol{\lambda}$)**	0.8870	0.8823	0.8869
	Model d ($\boldsymbol{\delta}$)	0.9027	0.9013	0.9027
	Model d ($\boldsymbol{\gamma}$)	0.8558	0.8661	0.8309
	Model d ($\boldsymbol{\psi}$)	0.8230	0.9130	0.8690
	Model e ($\boldsymbol{\lambda}$)	0.8839	0.8810	0.8820
	Model e ($\boldsymbol{\delta}$)	0.8983	0.8920	0.8963
	Model e ($\boldsymbol{\gamma}$)	0.8473	0.8553	0.8166
	Model e ($\boldsymbol{\psi}$)	0.8600	0.8992	0.8780

Table 35: Average of Coverage Rates under Population Model I at $N = 300$ Distribution 4: Uniform, $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent

		upper one-sided CI	**lower one-sided CI**	**two-sided CI**
CM	**Model b ($\boldsymbol{\lambda}$)**	0.9424	0.9576	0.9477
	Model b ($\boldsymbol{\phi}$)	0.9610	0.9293	0.9380
	Model b ($\boldsymbol{\psi}$)	0.9420	0.9610	0.9570
	Model c ($\boldsymbol{\lambda}$)	0.9432	0.9571	0.9479
	Model c ($\boldsymbol{\phi}$)	0.9603	0.9317	0.9373
	Model c ($\boldsymbol{\psi}$)	0.9316	0.9656	0.9444
PM	**Model d ($\boldsymbol{\lambda}$)**	0.9448	0.9576	0.9478
	Model d ($\boldsymbol{\delta}$)	0.9407	0.9560	0.9453
	Model d ($\boldsymbol{\gamma}$)	0.9489	0.9511	0.9499
	Model d ($\boldsymbol{\psi}$)	0.9450	0.9630	0.9530
	Model e ($\boldsymbol{\lambda}$)	0.9431	0.9572	0.9480
	Model e ($\boldsymbol{\delta}$)	0.9387	0.9523	0.9427
	Model e ($\boldsymbol{\gamma}$)	0.9472	0.9492	0.9478
	Model e ($\boldsymbol{\psi}$)	0.9317	0.9649	0.9451

Table 36: Average of Coverage Rates under Population Model II at $N = 300$ Distribution 4: Uniform, $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent

		upper one-sided CI	**lower one-sided CI**	**two-sided CI**
CM	**Model b ($\boldsymbol{\lambda}$)**	0.9521	0.9366	0.9478
	Model b ($\boldsymbol{\phi}$)	0.9169	0.9873	0.9551
	Model b ($\boldsymbol{\psi}$)	0.8697	0.9955	0.9191
	Model c ($\boldsymbol{\lambda}$)	0.9550	0.9407	0.9498
	Model c ($\boldsymbol{\phi}$)	0.9111	0.9898	0.9464
	Model c ($\boldsymbol{\psi}$)	0.9192	0.9713	0.9468
PM	**Model d ($\boldsymbol{\lambda}$)**	0.9527	0.9477	0.9526
	Model d ($\boldsymbol{\delta}$)	0.9696	0.9681	0.9696
	Model d ($\boldsymbol{\gamma}$)	0.9192	0.9303	0.8925
	Model d ($\boldsymbol{\psi}$)	0.8840	0.9807	0.9334
	Model e ($\boldsymbol{\lambda}$)	0.9545	0.9514	0.9525
	Model e ($\boldsymbol{\delta}$)	0.9701	0.9633	0.9680
	Model e ($\boldsymbol{\gamma}$)	0.9150	0.9237	0.8818
	Model e ($\boldsymbol{\psi}$)	0.9287	0.9711	0.9482

Table 37: Average of Success Rates under Population Model I at $N = 300$ Distribution 5: χ^2, $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent

		upper one-sided CI	**lower one-sided CI**	**two-sided CI**
CM	**Model b ($\boldsymbol{\lambda}$)**	0.9097	0.9688	0.9263
	Model b ($\boldsymbol{\phi}$)	0.9547	0.9157	0.9283
	Model b ($\boldsymbol{\psi}$)	0.9350	0.9690	0.9510
	Model c ($\boldsymbol{\lambda}$)	0.9108	0.9680	0.9260
	Model c ($\boldsymbol{\phi}$)	0.9523	0.9180	0.9250
	Model c ($\boldsymbol{\psi}$)	0.9099	0.9731	0.9324
PM	**Model d ($\boldsymbol{\lambda}$)**	0.9142	0.9660	0.9258
	Model d ($\boldsymbol{\delta}$)	0.9280	0.9473	0.9307
	Model d ($\boldsymbol{\gamma}$)	0.9471	0.9449	0.9460
	Model d ($\boldsymbol{\psi}$)	0.9310	0.9670	0.9480
	Model e ($\boldsymbol{\lambda}$)	0.9147	0.9659	0.9281
	Model e ($\boldsymbol{\delta}$)	0.9260	0.9457	0.9297
	Model e ($\boldsymbol{\gamma}$)	0.9460	0.9429	0.9459
	Model e ($\boldsymbol{\psi}$)	0.9086	0.9727	0.9322

Table 38: Average of Success Rates under Population Model II at $N = 300$ Distribution 5: χ^2, $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent

		upper one-sided CI	**lower one-sided CI**	**two-sided CI**
CM	**Model b ($\boldsymbol{\lambda}$)**	0.3596	0.3519	0.3560
	Model b ($\boldsymbol{\phi}$)	0.3533	0.3720	0.3627
	Model b ($\boldsymbol{\psi}$)	0.3240	0.3740	0.3450
	Model c ($\boldsymbol{\lambda}$)	0.3817	0.3713	0.3754
	Model c ($\boldsymbol{\phi}$)	0.3630	0.3933	0.3787
	Model c ($\boldsymbol{\psi}$)	0.3517	0.3882	0.3667
PM	**Model d ($\boldsymbol{\lambda}$)**	0.8464	0.8483	0.8473
	Model d ($\boldsymbol{\delta}$)	0.8737	0.8723	0.8740
	Model d ($\boldsymbol{\gamma}$)	0.8193	0.8248	0.7910
	Model d ($\boldsymbol{\psi}$)	0.7920	0.8780	0.8260
	Model e ($\boldsymbol{\lambda}$)	0.8307	0.8281	0.8243
	Model e ($\boldsymbol{\delta}$)	0.8497	0.8447	0.8447
	Model e ($\boldsymbol{\gamma}$)	0.7970	0.8038	0.7672
	Model e ($\boldsymbol{\psi}$)	0.7872	0.8527	0.8119

Table 39: Average of Coverage Rates under Population Model I at $N = 300$ Distribution 5: χ^2, $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent

		upper one-sided CI	**lower one-sided CI**	**two-sided CI**
CM	**Model b ($\boldsymbol{\lambda}$)**	0.9097	0.9688	0.9263
	Model b ($\boldsymbol{\phi}$)	0.9547	0.9157	0.9283
	Model b ($\boldsymbol{\psi}$)	0.9350	0.9690	0.9510
	Model c ($\boldsymbol{\lambda}$)	0.9108	0.9680	0.9260
	Model c ($\boldsymbol{\phi}$)	0.9523	0.9180	0.9250
	Model c ($\boldsymbol{\psi}$)	0.9099	0.9731	0.9324
PM	**Model d ($\boldsymbol{\lambda}$)**	0.9142	0.9660	0.9258
	Model d ($\boldsymbol{\delta}$)	0.9280	0.9473	0.9307
	Model d ($\boldsymbol{\gamma}$)	0.9471	0.9449	0.9460
	Model d ($\boldsymbol{\psi}$)	0.9310	0.9670	0.9480
	Model e ($\boldsymbol{\lambda}$)	0.9147	0.9659	0.9281
	Model e ($\boldsymbol{\delta}$)	0.9260	0.9457	0.9297
	Model e ($\boldsymbol{\gamma}$)	0.9460	0.9429	0.9459
	Model e ($\boldsymbol{\psi}$)	0.9086	0.9727	0.9322

Table 40: Average of Coverage Rates under Population Model II at $N = 300$ Distribution 5: χ^2, $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent

		upper one-sided CI	**lower one-sided CI**	**two-sided CI**
CM	**Model b ($\boldsymbol{\lambda}$)**	0.9563	0.9359	0.9468
	Model b ($\boldsymbol{\phi}$)	0.9397	0.9894	0.9645
	Model b ($\boldsymbol{\psi}$)	0.8617	0.9947	0.9176
	Model c ($\boldsymbol{\lambda}$)	0.9590	0.9330	0.9433
	Model c ($\boldsymbol{\phi}$)	0.9121	0.9883	0.9514
	Model c ($\boldsymbol{\psi}$)	0.8836	0.9754	0.9213
PM	**Model d ($\boldsymbol{\lambda}$)**	0.9489	0.9510	0.9499
	Model d ($\boldsymbol{\delta}$)	0.9794	0.9780	0.9798
	Model d ($\boldsymbol{\gamma}$)	0.9185	0.9246	0.8868
	Model d ($\boldsymbol{\psi}$)	0.8879	0.9843	0.9260
	Model e ($\boldsymbol{\lambda}$)	0.9515	0.9486	0.9443
	Model e ($\boldsymbol{\delta}$)	0.9733	0.9675	0.9675
	Model e ($\boldsymbol{\gamma}$)	0.9129	0.9207	0.8788
	Model e ($\boldsymbol{\psi}$)	0.9017	0.9767	0.9300

Table 41: Average of the Ratios of Widths of Two-Sided CI at $N = 300$ Distribution 1: Multivariate Normal

	$\boldsymbol{\lambda}$	$\boldsymbol{\psi}$
Id VS Ib	0.9856	0.9880
IId VS IIb	0.9927	0.9616
Ie VS Ic	0.9858	0.9982
IIe VS IIc	1.0091	1.0004

Table 42: Average of the Ratios of Widths of Two-Sided CI at $N = 300$ Distribution 2: Uniform

	$\boldsymbol{\lambda}$	$\boldsymbol{\psi}$
Id VS Ib	0.9817	0.9802
IId VS IIb	0.9933	0.9329
Ie VS Ic	0.9821	0.9969
IIe VS IIc	1.0094	1.0003

Table 43: Average of the Ratios of Widths of Two-Sided CI at $N = 300$ Distribution 3: χ^2

	$\boldsymbol{\lambda}$	$\boldsymbol{\psi}$
Id VS Ib	0.9918	0.9960
IId VS IIb	0.9902	0.9895
Ie VS Ic	0.9918	0.9995
IIe VS IIc	1.0194	1.0026

Table 44: Average of the Ratios of Widths of Two-Sided CI at $N = 300$ Distribution 4: Uniform, $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent

	$\boldsymbol{\lambda}$	$\boldsymbol{\psi}$
Id VS Ib	0.9935	0.9895
IId VS IIb	0.9935	0.9442
Ie VS Ic	0.9937	0.9982
IIe VS IIc	1.0104	1.0004

Table 45: Average of the Ratios of Widths of Two-Sided CI at $N = 300$ Distribution 5: χ^2, $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent

	$\boldsymbol{\lambda}$	$\boldsymbol{\psi}$
Id VS Ib	0.9716	0.9858
IId VS IIb	0.9903	0.9775
Ie VS Ic	0.9713	0.9985
IIe VS IIc	1.0081	1.0003

6.3.5. ETS1 and ETS2 for Model Comparison Test

Empirical test sizes for model comparison tests were computed in two ways for Browne's test and the Satterthwaite's approximation version. Browne(2) and Satterthwaite(2) are computed based on the algorithm discussed in §5.3 in Boik [34], whereas, Browne(1) and Satterthwaite(1) are computed based the output generated under conventional models and proposed models, without further calculation, by assuming $\mathbf{H}_0$ is true.

Table 51, Table 52, Table 53, Table 54, and Table 55 provide ETS1 and ETS2 for Model Comparison Test. Under Population Model I, Browne's test procedure and Satterthewaite approximation procedure both generate ETS1 close to $\alpha = 0.05$, especially, for the comparison between Model Ie and Model Ic, ETS1 are very close to each other and are close to $\alpha = 0.05$. Under Population Model II, ETS1 are much lower than $\alpha = 0.05$. Under Population Model I, there is not much difference between ETS1 and ETS2. But under Population Model II, ETS2 is much higher than $\alpha = 0.05$, especially those under conventional models, because of the high proportion of improper solutions from the conventional models.

In Table 51, Table 52, and Table 53, the titles of those tables are shortened in terms of the specific names of Distribution 1, Distribution 2, and Distribution 3. As defined earlier in Table 14, Distribution 1 is a multivariate normal distribution with $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are independent; Distribution 2 is a uniform distribution with $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are independent; and Distribution 3 is a χ^2 with $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are independent.

Table 46: ETS1 and ETS2 for the Goodness-of-fit Test at $N = 300$ Distribution 1

ETS1	**Population Model I**				**Population Model II**			
	CM		**PM**		**CM**		**PM**	
	Ib	**Ic**	**Id**	**Ie**	**IIb**	**IIc**	**IId**	**IIe**
Browne	0.0740	0.0650	0.0850	0.0650	0.0428	0.0511	0.0596	0.0423
Satterthwaite	0.0500	0.0510	0.0520	0.0550	0.0333	0.0533	0.0375	0.0445
ETS2	**Population Model I**				**Population Model II**			
	CM		**PM**		**CM**		**PM**	
	Ib	**Ic**	**Id**	**Ie**	**IIb**	**IIc**	**IId**	**IIe**
Browne	0.0740	0.0650	0.0850	0.0650	0.5970	0.5730	0.1480	0.1400
Satterthwaite	0.0500	0.0510	0.0520	0.0550	0.5930	0.5740	0.1280	0.1420

Table 47: ETS1 and ETS2 for the Goodness-of-fit Test at $N = 300$ Distribution 2

ETS1	**Population Model I**				**Population Model II**			
	CM		**PM**		**CM**		**PM**	
	Ib	**Ic**	**Id**	**Ie**	**IIb**	**IIc**	**IId**	**IIe**
Browne	0.0720	0.0690	0.0730	0.0650	0.0731	0.0986	0.0740	0.0726
Satterthwaite	0.0400	0.0500	0.0400	0.0470	0.0548	0.0575	0.0497	0.0561
ETS2	**Population Model I**				**Population Model II**			
	CM		**PM**		**CM**		**PM**	
	Ib	**Ic**	**Id**	**Ie**	**IIb**	**IIc**	**IId**	**IIe**
Browne	0.0720	0.0690	0.0730	0.0650	0.5940	0.5610	0.1610	0.1570
Satterthwaite	0.0400	0.0500	0.0400	0.0470	0.5860	0.5410	0.1390	0.1420

Table 48: ETS1 and ETS2 for the Goodness-of-fit Test at $N = 300$ Distribution 3

ETS1	**Population Model I**				**Population Model II**			
	CM		**PM**		**CM**		**PM**	
	Ib	**Ic**	**Id**	**Ie**	**IIb**	**IIc**	**IId**	**IIe**
Browne	0.0630	0.0400	0.0640	0.0400	0.0725	0.0474	0.0872	0.0410
Satterthwaite	0.0200	0.0360	0.0210	0.0350	0.0277	0.0172	0.0344	0.0353
ETS2	**Population Model I**				**Population Model II**			
	CM		**PM**		**CM**		**PM**	
	Ib	**Ic**	**Id**	**Ie**	**IIb**	**IIc**	**IId**	**IIe**
Browne	0.0630	0.0400	0.0640	0.0400	0.5650	0.5580	0.1520	0.1570
Satterthwaite	0.0200	0.0360	0.0210	0.0350	0.5440	0.5440	0.1030	0.1520

Table 49: ETS1 and ETS2 for the Goodness-of-fit Test at $N = 300$ Distribution 4: Uniform, $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent

ETS1	**Population Model I**				**Population Model II**			
	CM		**PM**		**CM**		**PM**	
	Ib	**Ic**	**Id**	**Ie**	**IIb**	**IIc**	**IId**	**IIe**
Browne	0.0680	0.0570	0.0720	0.0570	0.0629	0.0428	0.0655	0.0443
Satterthwaite	0.0330	0.0420	0.0360	0.0400	0.0337	0.0387	0.0440	0.0443
ETS2	**Population Model I**				**Population Model II**			
	CM		**PM**		**CM**		**PM**	
	Ib	**Ic**	**Id**	**Ie**	**IIb**	**IIc**	**IId**	**IIe**
Browne	0.0680	0.0570	0.0720	0.0570	0.5830	0.5300	0.1300	0.1150
Satterthwaite	0.0330	0.0420	0.0360	0.0400	0.5700	0.5280	0.1100	0.1150

Table 50: ETS1 and ETS2 for the Goodness-of-fit Test at $N = 300$ Distribution 5: χ^2, $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent

ETS1	**Population Model I**				**Population Model II**			
	CM		**PM**		**CM**		**PM**	
	Ib	**Ic**	**Id**	**Ie**	**IIb**	**IIc**	**IId**	**IIe**
Browne	0.0840	0.0710	0.0890	0.0710	0.0771	0.0477	0.0673	0.0458
Satterthwaite	0.0230	0.0330	0.0190	0.0270	0.0213	0.0302	0.0235	0.0275
ETS2	**Population Model I**				**Population Model II**			
	CM		**PM**		**CM**		**PM**	
	Ib	**Ic**	**Id**	**Ie**	**IIb**	**IIc**	**IId**	**IIe**
Browne	0.0840	0.0710	0.0890	0.0710	0.6530	0.6210	0.1680	0.1670
Satterthwaite	0.0230	0.0330	0.0190	0.0270	0.6320	0.6140	0.1290	0.1510

Table 51: ETS1 and ETS2 for Model Comparison Test at $N = 300$ Distribution 1

ETS1	**Browne(1)**	**Satterthwaite(1)**	**Browne(2)**	**Satterthwaite(2)**
Id VS Ib	0.0590	0.0540	0.0610	0.0490
Ie VS Ic	0.0530	0.0540	0.0530	0.0540
IId VS IIb	0.0077	0.0026	0.0077	0.0026
IIe VS IIc	0.0144	0.0024	0.0192	0.0048
ETS2	**Browne(1)**	**Satterthwaite(1)**	**Browne(2)**	**Satterthwaite(2)**
Id VS Ib	0.0590	0.0540	0.0610	0.0490
Ie VS Ic	0.0530	0.0540	0.0530	0.0540
IId VS IIb	0.6120	0.6100	0.6120	0.6100
IIe VS IIc	0.5890	0.5840	0.5910	0.5850

Table 52: ETS1 and ETS2 for Model Comparison Test at $N = 300$ Distribution 2

ETS1	**Browne(1)**	**Satterthwaite(1)**	**Browne(2)**	**Satterthwaite(2)**
Id VS Ib	0.0520	0.0440	0.0520	0.0410
Ie VS Ic	0.0430	0.0360	0.0430	0.0350
IId VS IIb	0.0120	0.0000	0.0120	0.0000
IIe VS IIc	0.0198	0.0066	0.0176	0.0066
ETS2	**Browne(1)**	**Satterthwaite(1)**	**Browne(2)**	**Satterthwaite(2)**
Id VS Ib	0.0520	0.0440	0.0520	0.0410
Ie VS Ic	0.0430	0.0360	0.0430	0.0350
IId VS IIb	0.5890	0.5840	0.5890	0.5840
IIe VS IIc	0.5550	0.5490	0.5540	0.5490

Table 53: ETS1 and ETS2 for Model Comparison Test at $N = 300$ Distribution 3

ETS1	**Browne(1)**	**Satterthwaite(1)**	**Browne(2)**	**Satterthwaite(2)**
Id VS Ib	0.0540	0.0540	0.0530	0.0500
Ie VS Ic	0.0590	0.0490	0.0610	0.0530
IId VS IIb	0.0605	0.0135	0.0583	0.0112
IIe VS IIc	0.0313	0.0241	0.0337	0.0265
ETS2	**Browne(1)**	**Satterthwaite(1)**	**Browne(2)**	**Satterthwaite(2)**
Id VS Ib	0.0540	0.0540	0.0530	0.0500
Ie VS Ic	0.0590	0.0490	0.0610	0.0530
IId VS IIb	0.5810	0.5600	0.5800	0.5590
IIe VS IIc	0.5980	0.5950	0.5990	0.5960

Table 54: ETS1 and ETS2 for Model Comparison Test at $N = 300$ Distribution 4: Uniform, $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent

ETS1	**Browne(1)**	**Satterthwaite(1)**	**Browne(2)**	**Satterthwaite(2)**
Id VS Ib	0.0660	0.0610	0.0660	0.0580
Ie VS Ic	0.0640	0.0580	0.0660	0.0580
IId VS IIb	0.0143	0.0024	0.0143	0.0024
IIe VS IIc	0.0129	0.0043	0.0151	0.0086
ETS2	**Browne(1)**	**Satterthwaite(1)**	**Browne(2)**	**Satterthwaite(2)**
Id VS Ib	0.0540	0.0540	0.0530	0.0500
Ie VS Ic	0.0590	0.0490	0.0610	0.0530
IId VS IIb	0.5810	0.5600	0.5800	0.5590
IIe VS IIc	0.5980	0.5950	0.5990	0.5960

Table 55: ETS1 and ETS2 for Model Comparison Test at $N = 300$ Distribution 5: χ^2, $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are uncorrelated but not independent

ETS1	**Browne(1)**	**Satterthwaite(1)**	**Browne(2)**	**Satterthwaite(2)**
Id VS Ib	0.0760	0.0620	0.0780	0.0550
Ie VS Ic	0.0690	0.0590	0.0680	0.0600
IId VS IIb	0.0309	0.0000	0.0309	0.0000
IIe VS IIc	0.0279	0.0084	0.0306	0.0084
ETS2	**Browne(1)**	**Satterthwaite(1)**	**Browne(2)**	**Satterthwaite(2)**
Id VS Ib	0.0760	0.0620	0.0780	0.0550
Ie VS Ic	0.0690	0.0590	0.0680	0.0600
IId VS IIb	0.6550	0.6440	0.6550	0.6440
IIe VS IIc	0.6510	0.6440	0.6520	0.6440

6.4. Conclusion and Future Work

Based on the large simulation results, it has demonstrated that the inference procedures for the proposed model work well enough to be used in practice and that the proposed model has advantages over the conventional model, in terms of proportion of proper solutions; average success rates and average coverage rates of upper one-sided nominal 95% confidence intervals, lower one-sided nominal 95% confidence intervals, and two-sided nominal 95% confidence intervals; and average of the ratios of widths of two-sided nominal 95% confidence intervals.

Future work includes the algorithm used in this thesis becomes more robust to the data variability so that the convergence rate for the proposed models is higher than the current convergence rate. It is also likely to write a similar program as the one used in this thesis in R so that the programs used in R become robust as well.

Also, future work includes constructing models and inference procedures for fitting confirmatory factor analysis models to correlation matrices.

REFERENCES CITED

[1] C. Spearman. General intelligence, objectively determined and measured. *American Journal of Psychology*, 15:201–293, 1904.

[2] D. N. Lawley, A. E. Maxwell. *Factor Analysis as a Statistical Method*. edition second, 1971.

[3] G. Casella, R. L. Berger. *Statistical Inference*. Duxbury Press, 1990.

[4] K. A. Bollen. *Structural equations with latent variables*. Wiley-Interscience, 1989.

[5] T. W. Anderson, H. Rubin. Statistical inference in factor analysis. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 5:111–150, 1956.

[6] B. Vinograde. Canonical positive definite matrices under internal linear transformations. *Proceedings of the American Mathematical Society*, 1:159–161, 1950.

[7] M. S. Bartlett. Tests of significance in factor analysis. *The British Journal of Psychology*, 3:77–85, 1950.

[8] M. W. Browne. Covariance structures. *Topics in Applied Multivariate Analysis*, 1982.

[9] K. G. Jöreskog, A. S. Goldberger. Factor analysis by generalized least squares. *Psychometrika*, 37(3):243–260, 1972.

[10] R. J. Rummel. *Applied Factor Analysis*. Northwestern University Press, 1970.

[11] M. W. Browne. An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36:111–150, 2001.

[12] M. W. Browne, R. Cudeck, K. Tateneni, G. Mels. Cefa: Comprehensive exploratory factor analysis (version 3.02) [computer software manual]. 2008.

[13] C. A. Bernaards, R. I. Jennrich. Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational and Psychological Measurement*, 65:676–696, 2005.

[14] J. B. Carroll. An analytical solution for approximating simple structure in factor analysis. *Psychometrika*, 18:23–38, 1953.

[15] R. I. Jennrich. Rotation to simple loadings using component loss functions: The oblique case. *Psychometrika*, 71:173–191, 2006.

[16] R. J. Boik. An implicit function approach to constrained optimization with applications to asymptotic expansions. *Journal of Multivariate Analysis*, 2007. `doi:http://10.1016/j.jmva.2007.01.005`.

[17] D. Clarkson, R. Jennrich. Quartic rotation criteria and algorithms. *Psychometrika*, 53(2):251–259, June 1988.

[18] R. J. Boik. Newton algorithms for analytic rotation: an implicit function approach. *Psychometrica*, 73(2):231–259, 2008.

[19] H. F. Kaiser. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23:187–200, 1958.

[20] H. B. Heywood. On finite sequences of real numbers. *Proceedings of the Royal Society of London*, 134:486–501, 1931.

[21] O. P. Van Driel. On various causes of improper solutions in maximum likelihood factor analysis. *Psychometrika*, 43(2):225–243, 1978.

[22] D. Rindskopf. Structural equation models: Empirical identification, heywood cases, and related problems. *Sociological Methods Research*, 13:109–119, 1984.

[23] J. Anderson, D.W. Gerbing. The effects of sampling error on convergence, improper solutions and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49:155–173, 1984.

[24] K. A. Bollen. Outliers and improper solutions: A confirmatory factor analysis example. *Sociological Methods and Research*, 15:375–384, 1987.

[25] A. Boomsma. The robustness of lisrel against small sample sizes in factor analysis models. *Systems under Indirect Observation: Causality, Structure, Prediction*, pages 149–173, 1982.

[26] S. Y. Lee. Estimation of covariance structure models with parameters subject to functional restraints. *Psychometrika*, 45(3):309–324, 1980.

[27] P. M. Bentler, D. G. Weeks. Linear structural equations with latent variables. *Psychometrika*, 45:289–308, 1980.

[28] D. Rindskopf. Parameterizing inequality constraints on unique variances in linear structural models. *Psychometrika*, 48:73–83, 1983.

[29] P. M. Bentler, M. Jamshidian. Gramian matrices in covariance structure models. *Applied Psychological Measurement*, 18(1):79–94, 1994.

[30] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012.

[31] John Fox, Zhenghua Nie, Jarrett Byrnes. *sem: Structural Equation Models*, 2012. R package version 3.0-0.

[32] R. MacCallum. A comparison of factor analysis programs in spss, bmdp, and sas. *Psychometrika*, 48(2):223–231, June 1983.

[33] R. J. Boik. Model-based principal components of correlation matrices (accepted for publication). *Journal of Multivariate Analysis*, 2013.

[34] R. J. Boik. Principal component models for correlation matrices. *Biometrika*, 90(3):679–701, 2003.

[35] R. J. Boik, K. Panishkan, S. K. Hyde. Model-based principal components of covariance matrices. *British Journal of Mathematical and Statistical Psychology*, 63(1), 2010.

[36] S. R. Searle. *Matrix algebra useful for statistics.* John Wiley & Sons., 1982.

[37] J. W. Brewer. Kronecker products and matrix calculus in system theory. *IEEE Transactions on Circuits and Systems*, 25:772–781, 1978.

[38] P. M. Bentler. Confirmatory factor analysis via noniterative estimation: A fast, inexpensive method. *Journal of Marketing Research*, 19(4):417–424, 1982.

[39] S. P. Lin, R. B. Bendel. Algorithm as 213: Generation of population correlation matrices with specified eigenvalues. *Applied Statistics*, 34:193–198, 1985.

[40] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods.* Athena Scientific, edition 1, 1996.

[41] P. K. Sen, J. M. Singer. *Large sample methods in statistics: an introduction with applications.* Chapman & Hall Ltd., 1993.

[42] J. N. K. Rao. *Linear statistical inference and its applications.* Wiley, 1973.

[43] C. D. Meyer. *Matrix analysis and applied linear algebra.* Society for Industrial and Applied Mathematics, 2000.

[44] J. R. Schott. *Matrix Analysis for Statistics.* John Wiley & Sons., 1997.

[45] R. J. Muirhead. *Aspects of multivariate statistical theory.* Wiley, 1982.

[46] R. J. Boik. Second-order accurate inference on eigenvalues of covariance and correlation matrices. *Journal of Multivariate Analysis*, 96(1):136–171, 2005.

[47] J. R. Magnus, H. Neudecker. The commutation matrix: Some properties and applications. 7:381–394, 1979.

[48] M. W. Browne. Generalized least squares estimators in the analysis of covariance structures. *South African Statistical Journal*, 8:1–24, 1974.

[49] M. W. Browne. Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37:62–83, 1984.

[50] S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, 9:60–62, 1938.

[51] F. E. Satterthwaite. An approximate distribution of estimates of variance components. *Biometrika*, 33:110–114, 1946.

[52] Ronald Christensen. *Plane Answers to Complex Questions: The Theory of Linear Models*, Volume second edition. Springer, 2011.

[53] S. Levy. Use of the mapping sentence for coordinating theory and research: a cross-cultural example. *Quality and Quantity: European-American Journal of Methodology*, 10:117–125, 1976.

[54] T. M. Ostrom. The relationship between the affective, behavioral, and cognitive components of attitude. *Journal of Experimental Social Psychology*, 5:12–30, 1969.

[55] B. M. Byrne, R. J. Shavelson. On the structure of adolescent self-concept. *J. of Educ. Psychology*, 78:474–481, 1986.

[56] M. S. Srivastava. *Methods of multivariate statistics*. Wiley series in probability and statistics. Wiley, 2002.

[57] Emden R. Gansner, Stephen C. North. An open graph visualization system and its applications to software engineering. *Softw. Pract. Exper.*, 30(11), September 2000.

[58] D. T. Campbell, D. W. Fiske. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56:81–105, March 1959.

[59] H. W. Marsh. Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement*, 13:335–361, 1989.

[60] J. R. Magnus. *Linear Structures*. Griffin, 1988.

[61] J. J. Callahan. *Advanced Calculus: A Geometric View*. Springer, 2010.

[62] Ronald R. Hocking. *Methods and applications of linear models : regression and the analysis of variance*. Wiley, New York, 1996.

[63] J. Shao. *Mathematical Statistics*. Springer, 1999.

APPENDICES

APPENDIX A

NOTATION

Table 56: Matrix Functions

Function	Definition
$\mathrm{tr}(\mathbf{A})$	trace of the square matrix $\mathbf{A}$
$\mathrm{diag}(\mathbf{A})$	vector of diagonal components of the square matrix $\mathbf{A}$
$\mathrm{Diag}(\mathbf{v})$	diagonal matrix whose diagonals are the components of $\mathbf{v}$
$\dim(\mathbf{M})$	dimension of a vector, matrix, or vector space
$\mathrm{vec}(\mathbf{M})$	vector obtained by stacking the columns of $\mathbf{M}$
$\mathrm{dvec}(\mathbf{M}; a, b)$	$a \times b$ matrix $\mathbf{U}$ that satisfies $\mathrm{vec}(\mathbf{U}) = \mathrm{vec}(\mathbf{M})$, where $\dim(\mathrm{vec}\,\mathbf{M}) = ab$
$\mathrm{vech}(\mathbf{A})$	$\mathbf{A} = \mathbf{A}'$, the vech of $\mathbf{A}$ is the vector obtained by stacking the distinct elements in the columns of $\mathbf{A}$ (i.e. that are on and below the main diagonal) into a single vector.
$\mathrm{dvech}(\mathbf{a})$	symmetric matrix $\mathbf{A}$ that satisfies $\mathrm{vech}(\mathbf{A}) = \mathbf{a}$
$(\mathbf{M})^{\otimes r}$	r^{th} order Kronecker product of $\mathbf{M}$ with itself; e.g., $(\mathbf{M})^{\otimes 3} = \mathbf{M} \otimes \mathbf{M} \otimes \mathbf{M}$
$\otimes$	direct or Kronecker product
$\bigoplus$	direct sum, e.g., $\bigoplus_{i=1}^{k} \mathbf{A}_i = \mathrm{Diag}(\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k)$
$\odot$	elementwise operator
$\mathbf{1}_p$	$p \times 1$ column of ones
$\mathbf{e}_i^p$	elementary vector, i^{th} column of $\mathbf{I}_p$
$\mathbf{E}_{j,\mathbf{a}}$	elementary matrix, j^{th} submatrix of $\mathbf{I}_p = (\mathbf{E}_{1,\mathbf{a}} \;\; \mathbf{E}_{2,\mathbf{a}} \;\; \cdots \;\; \mathbf{E}_{c,\mathbf{a}})$, $\dim(\mathbf{E}_{j,\mathbf{a}}) = p \times a_j$, the components of $\mathbf{a} = (a_1 \;\; a_2 \;\; \cdots \;\; a_c)'$ are positive integers, and $p = \mathbf{1}_c'\mathbf{a}$
$\mathbf{I}_{p,3}$	$\mathbf{I}_p \otimes \mathrm{vec}(\mathbf{I}_p) \otimes \mathbf{I}_p$
$\mathbf{L}_{qr,p}$	$\sum_{i=1}^{p} (\mathbf{e}_i^p)^{\otimes q} \left(\mathbf{e}_i^{p\prime}\right)^{\otimes r}$
$\mathbf{D}_p$	$p^2 \times p(p+1)/2$ duplication matrix
$\mathbf{K}_{a,b}$	$ab \times ab$ commutation matrix
$\mathbf{N}_p$	$\left(\mathbf{I}_{p^2} + \mathbf{K}_{p,p}\right)/2$
$\mathbf{N}_p^{\perp}$	$\left(\mathbf{I}_{p^2} - \mathbf{K}_{p,p}\right)/2$
$\mathbf{J}_{21,p}$	$\mathbf{K}_{p^2,p} + (\mathbf{I}_p \otimes 2\mathbf{N}_p)$
$\mathrm{SVD}(\mathbf{B})$	singular value decomposition: $\mathrm{SVD}(\mathbf{B}) = \mathbf{UDV}'$
$\mathbf{B}^{+}$	Moore-Penrose generalized inverse of $\mathbf{B}$
$\mathrm{ppo}(\mathbf{X})$	a projection operator that projects onto $\mathcal{R}(\mathbf{X})$ along $\mathcal{N}(\mathbf{X}')$, $\mathrm{ppo}(\mathbf{X}) = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'$

Table 57: Matrix/Vector Spaces and Basis Sets

Set/Operator	Definition
$\mathcal{O}(p)$	set of all $p \times p$ orthogonal matrices
$\mathcal{O}(\mathbf{m})$	set of all $p \times p$ orthogonal matrices with structure $\text{BDiag}\left(\{\mathbf{Q}_i\}_{i=1}^{m_0}\right)$, where $\mathbf{Q}_i \in \mathcal{O}(m_i)$, $\sum_{i=1}^{m_0} m_i = p$ and m_i is a positive integer
$\mathcal{O}(p, q)$	set of all $p \times q$ semiorthogonal matrices, where $p > q$.
$\mathcal{D}^+(p)$	set of all $p \times p$ diagonal positive definite matrices
$\mathcal{R}(\mathbf{B})$	vector space generated by the columns of the matrix $\mathbf{B}$
$\mathcal{N}(\mathbf{B})$	null space (kernel) of the matrix $\mathbf{B}$
$\mathbf{F}^{\perp}$	any $a \times (a - r_f)$ matrix whose columns are a basis for $\mathcal{N}(\mathbf{F}')$, where $\dim(\mathbf{F}) = a \times b$ and $\text{rank}(\mathbf{F}) = r_f$
$\mathring{N}(\mathbf{B})$	open neighborhood of the matrix $\mathbf{B}$

APPENDIX B

PROOFS OF THEOREMS 1-35 AND LEMMA 1

B.1. Proof of Theorem 1

Theorem 1. [Adapted from Bartlett [7], 1950, Page 84]. *Assume that the* $p \times q$ *matrix* $\mathbf{\Lambda}$ *has rank* q *and that* $\mathbf{\Lambda}$ *is identified up to orthogonal rotation. Then,* $\mathbf{\Lambda}$ *can be parameterized as* $\mathbf{\Lambda}(\boldsymbol{\theta}_{\boldsymbol{\lambda}})$*, where* $\dim(\boldsymbol{\theta}_{\boldsymbol{\lambda}}) = pq - \frac{q(q-1)}{2}$ *and* $\boldsymbol{\theta}_{\boldsymbol{\lambda}}$ *is identified.*

Proof: [Proof With Advisor's Help]. *If* $\operatorname{rank}(\mathbf{\Lambda}) = q$*, then there exist* q *rows of* $\mathbf{\Lambda}$ *that are linearly independent. Without loss of generality, assume that the first* q *rows of* $\mathbf{\Lambda}$ *are linearly independent. If this is not the case, then the rows of* $\mathbf{y}$, $\boldsymbol{\mu}$, $\mathbf{\Lambda}$ *and* $\boldsymbol{\epsilon}$ *can be permuted so that the first* q *rows are linearly independent.*

Partition $\mathbf{\Lambda}$ *as* $\mathbf{\Lambda} = (\mathbf{\Lambda}_1' \;\; \mathbf{\Lambda}_2')'$*, where* $\mathbf{\Lambda}_1$ *is a* $q \times q$ *nonsingular matrix. Let* $\mathbf{QR} = \mathbf{\Lambda}_1'$ *be the QR factorization of* $\mathbf{\Lambda}_1'$*, where* $\mathbf{Q} \in \mathcal{O}_q$*,* $\mathbf{R}$ *is an upper triangular matrix (also called right triangular matrix), the diagonal components of* $\mathbf{R}$ *are positive and* $\mathcal{O}_q$ *is defined in Table 57. This factorization is unique because* $\mathbf{\Lambda}_1$ *has full rank.*

Write $\mathbf{\Lambda}$ *as* $\mathbf{\Lambda} = \begin{pmatrix} \mathbf{R}'\mathbf{Q}' \\ \mathbf{\Lambda}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{R}' \\ \mathbf{\Lambda}_2\mathbf{Q} \end{pmatrix} \mathbf{Q}' = \begin{pmatrix} \mathbf{R}' \\ \mathbf{B} \end{pmatrix} \mathbf{Q}'$*, where* $\mathbf{B} = \mathbf{\Lambda}_2\mathbf{Q}$.

Define $\boldsymbol{\theta}_{\boldsymbol{\lambda}}$ *as* $\boldsymbol{\theta}_{\boldsymbol{\lambda}} \stackrel{\text{def}}{=} \begin{pmatrix} \nu^*(\mathbf{R}) \\ \operatorname{vec}(\mathbf{B}) \end{pmatrix}$*, where* $\nu^*(\mathbf{R})$ *contains the* $\frac{q(q+1)}{2}$ *components in the upper triangle of* $\mathbf{R}$ *and* $\dim[\operatorname{vec}(\mathbf{B})] = (p-q)q$*. The dimension of* $\boldsymbol{\theta}_{\boldsymbol{\lambda}}$ *is* $pq - \frac{q(q-1)}{2}$ *because* $\dim(\boldsymbol{\theta}_{\boldsymbol{\lambda}}) = \dim(\nu^*(\mathbf{R})) + \dim(\operatorname{vec}(\mathbf{B})) = \frac{q(q+1)}{2} + (p-q)q = pq - \frac{q(q-1)}{2}$*. The first part of the proof is now finished because* $\dim(\boldsymbol{\theta}_{\boldsymbol{\lambda}}) = pq - \frac{q(q-1)}{2}$.

The second part of the proof is to show that $\mathbf{\Lambda}(\boldsymbol{\theta}_{\boldsymbol{\lambda}_1})\mathbf{\Lambda}(\boldsymbol{\theta}_{\boldsymbol{\lambda}_1})' = \mathbf{\Lambda}(\boldsymbol{\theta}_{\boldsymbol{\lambda}_2})\mathbf{\Lambda}(\boldsymbol{\theta}_{\boldsymbol{\lambda}_2})' \Longrightarrow \boldsymbol{\theta}_{\boldsymbol{\lambda}_1} = \boldsymbol{\theta}_{\boldsymbol{\lambda}_2}$.

Write $\mathbf{\Lambda\Lambda}'$ *as* $\mathbf{\Lambda\Lambda}' = \begin{pmatrix} \mathbf{R}' \\ \mathbf{B} \end{pmatrix} \mathbf{Q}'\mathbf{Q}\,(\mathbf{R} \;\; \mathbf{B}') = \begin{pmatrix} \mathbf{R}' \\ \mathbf{B} \end{pmatrix} (\mathbf{R} \;\; \mathbf{B}') = \begin{pmatrix} \mathbf{R}'\mathbf{R} & \mathbf{R}'\mathbf{B}' \\ \mathbf{BR} & \mathbf{BB}' \end{pmatrix}$.

If $\boldsymbol{\Lambda}(\boldsymbol{\theta}_{\boldsymbol{\lambda}_1})\boldsymbol{\Lambda}(\boldsymbol{\theta}_{\boldsymbol{\lambda}_1})' = \boldsymbol{\Lambda}(\boldsymbol{\theta}_{\boldsymbol{\lambda}_2})\boldsymbol{\Lambda}(\boldsymbol{\theta}_{\boldsymbol{\lambda}_2})'$, *then* $\begin{pmatrix} \mathbf{R}_1'\mathbf{R}_1 & \mathbf{R}_1'\mathbf{B}_1' \\ \mathbf{B}_1\mathbf{R}_1 & \mathbf{B}_1\mathbf{B}_1' \end{pmatrix} = \begin{pmatrix} \mathbf{R}_2'\mathbf{R}_2 & \mathbf{R}_2'\mathbf{B}_2' \\ \mathbf{B}_2\mathbf{R}_2 & \mathbf{B}_2\mathbf{B}_2' \end{pmatrix}$. *Note that* $\mathbf{R}_1'\mathbf{R}_1$ *and* $\mathbf{R}_2'\mathbf{R}_2$ *are each the unique Cholesky factorization of the positive definite matrix in the upper left hand corner of* $\boldsymbol{\Lambda}\boldsymbol{\Lambda}'$. *It follows that* $\mathbf{R}_1 = \mathbf{R}_2$. *It is true that* $\mathbf{B}_1 = \mathbf{B}_2$ *because* $\mathbf{B}_1\mathbf{R}_1 = \mathbf{B}_2\mathbf{R}_1$ *and* $\mathbf{R}_1$ *is nonsingular. It can be concluded that* $\boldsymbol{\theta}_{\boldsymbol{\lambda}_1} = \boldsymbol{\theta}_{\boldsymbol{\lambda}_2}$ *and* $\boldsymbol{\theta}_{\boldsymbol{\lambda}}$ *is identified.* □

B.2. Proof of Theorem 2

Theorem 2. [Adapted from Boik [35], Supplement, 2009, Theorem 13, Page 15]. *Let* $\boldsymbol{\delta}$ *be parameterized as model 4 in Table 4. Define* $\mathbf{D}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}$ *and* $\mathbf{W}_{\boldsymbol{\delta}}$ *as*

$$\mathbf{D}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}} \stackrel{\text{def}}{=} \operatorname{Diag}(\boldsymbol{\xi}_{\delta}) \text{ and } \mathbf{W}_{\boldsymbol{\delta}} \stackrel{\text{def}}{=} \mathbf{C}_1'\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}.$$

Assume that $\mathbf{C}_1$ *has been chosen such that the* $r_c \times q_2$ *matrix* $\mathbf{W}_{\boldsymbol{\delta}}$ *has full row-rank. Write the singular value decomposition of* $\mathbf{W}_{\boldsymbol{\delta}}$ *as follows:*

$$\begin{aligned}
&\mathbf{W}_{\boldsymbol{\delta}} = \mathbf{U}_{\boldsymbol{\delta}}\mathbf{D}_{\boldsymbol{\delta}}\mathbf{V}_{\boldsymbol{\delta}}', \text{ where} \\
&\mathbf{U}_{\boldsymbol{\delta}} \in \mathcal{O}(r_c), \quad \mathbf{V}_{\boldsymbol{\delta}} = \begin{pmatrix} \mathbf{V}_{\boldsymbol{\delta},1} & \mathbf{V}_{\boldsymbol{\delta},2} \end{pmatrix} \in \mathcal{O}(q_2), \\
&\mathbf{V}_{\boldsymbol{\delta},1} \in \mathcal{O}(q_2, r_c), \quad \mathbf{D}_{\boldsymbol{\delta}} = \begin{pmatrix} \mathbf{D}_{\boldsymbol{\delta},1} & \mathbf{0}_{r_c \times (q_2 - r_c)} \end{pmatrix}, \text{ and } \mathbf{D}_{\boldsymbol{\delta},1} \in \mathcal{D}^+(r_c).
\end{aligned}$$

Then,

a. *the parameter* $\boldsymbol{\xi}_{\delta}$ *can be written as* $\boldsymbol{\xi}_{\delta} = \mathbf{V}_1\boldsymbol{\eta}_{\boldsymbol{\delta}} + \mathbf{V}_2\boldsymbol{\theta}_{\boldsymbol{\delta}}$, *where* $\boldsymbol{\eta}_{\boldsymbol{\delta}} = \mathbf{V}_1'\boldsymbol{\xi}_{\delta}$, $\boldsymbol{\theta}_{\boldsymbol{\delta}} = \mathbf{V}_2'\boldsymbol{\xi}_{\delta}$, $\mathbf{V}_1 = \mathbf{V}_{\boldsymbol{\delta},1}$, *and* $\mathbf{V}_2 = \mathbf{V}_{\boldsymbol{\delta},2}$;

b. *the parameters* $\boldsymbol{\eta}_{\boldsymbol{\delta}}$ *and* $\boldsymbol{\xi}_{\delta}$ *are implicit functions of* $\boldsymbol{\theta}_{\boldsymbol{\delta}}$. *Therefore,* $\partial\boldsymbol{\eta}_{\boldsymbol{\delta}}/\partial\boldsymbol{\theta}_{\boldsymbol{\delta}}'$ *and* $\partial\boldsymbol{\xi}_{\delta}/\partial\boldsymbol{\theta}_{\boldsymbol{\delta}}'$ *exist;*

c. [Original result]. *the first three derivatives of* $\boldsymbol{\delta}$ *with respect to* $\boldsymbol{\theta_\delta}$ *can be written as follows:*

$$\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta} = 2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{D}^{(1)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta} = 2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{V}_2,$$

$$\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} = 2\mathbf{T}_2\left(\mathbf{I}_{q_2} - \mathbf{P}_{\boldsymbol{\xi_\delta}}\right)\left(\mathbf{V}'_2 * \mathbf{V}'_2\right)', \text{ and}$$

$$\mathbf{D}^{(3)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} = 2\mathbf{T}_2\left(\mathbf{I}_{q_2} - \mathbf{P}_{\boldsymbol{\xi_\delta}}\right)\left(\mathbf{D}^{(2)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta}\right)'\mathbf{J}_{\nu_2}, \text{ where}$$

$$\mathbf{D}^{(1)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta} = \mathbf{V}_2, \quad \mathbf{D}^{(2)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} = -\mathbf{D}^{-1}_{\boldsymbol{\xi_\delta}}\mathbf{P}_{\boldsymbol{\xi_\delta}}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta}\right)', \quad \mathbf{P}_{\boldsymbol{\xi_\delta}} = \mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{W}^+_{\boldsymbol{\delta}}\mathbf{C}'_1\mathbf{T}_2,$$

$*$ *is the Khatri-Rao column-wise product,* $\mathbf{J}_{\nu_2}$ *and* $\mathbf{N}_{\nu_2}$ *are defined in Table 56 and* $\mathcal{D}^+(r_c)$ *is defined in Table 57; and*

d. [Original result]. $\mathbf{P}_{\boldsymbol{\xi_\delta}}$ *is a projection operator.*

Proof: [Independent Proofs of (a) (b) and (d)][Proof of (c) With Advisor's Help] *The structure of (26) and (27) are adapted from Theorem 13 in the supplement of Boik [35]. Results of part a and part b were given in* §2.4 *of Boik [34] and the corresponding proofs are given below.*

a. *The parameter* $\boldsymbol{\xi}_\delta$ *can be written in as follows:*

$$\begin{aligned}
\boldsymbol{\xi}_\delta &= \mathbf{V}_{\boldsymbol{\delta}}\mathbf{V}'_{\boldsymbol{\delta}}\boldsymbol{\xi}_\delta \\
&= \begin{pmatrix}\mathbf{V}_{\boldsymbol{\delta},1} & \mathbf{V}_{\boldsymbol{\delta},2}\end{pmatrix}\begin{pmatrix}\mathbf{V}'_{\boldsymbol{\delta},1} \\ \mathbf{V}'_{\boldsymbol{\delta},2}\end{pmatrix}\boldsymbol{\xi}_\delta \\
&= \mathbf{V}_{\boldsymbol{\delta},1}\mathbf{V}'_{\boldsymbol{\delta},1}\boldsymbol{\xi}_\delta + \mathbf{V}_{\boldsymbol{\delta},2}\mathbf{V}'_{\boldsymbol{\delta},2}\boldsymbol{\xi}_\delta \\
&= \mathbf{V}_1\mathbf{V}'_1\boldsymbol{\xi}_\delta + \mathbf{V}_2\mathbf{V}'_2\boldsymbol{\xi}_\delta \\
&= \mathbf{V}_1\boldsymbol{\eta_\delta} + \mathbf{V}_2\boldsymbol{\theta_\delta},
\end{aligned}$$

where $\boldsymbol{\eta_\delta} = \mathbf{V}'_1\boldsymbol{\xi}_\delta$, $\boldsymbol{\theta_\delta} = \mathbf{V}'_2\boldsymbol{\xi}_\delta$, $\mathbf{V}_1 = \mathbf{V}_{\boldsymbol{\delta},1}$, *and* $\mathbf{V}_2 = \mathbf{V}_{\boldsymbol{\delta},2}$

b. *By the definition of* $\odot$, $\boldsymbol{\delta} = \mathbf{T}_2(\boldsymbol{\xi}_\delta \odot \boldsymbol{\xi}_\delta) = \mathbf{T}_2\sum_{i=1}^{q_2}\left(\mathbf{e}^{q_2}_i\boldsymbol{\xi}'_{\boldsymbol{\delta}}\mathbf{e}^{q_2}_i\mathbf{e}^{q_2'}_i\boldsymbol{\xi}_\delta\right)$, *where* $\mathbf{e}^{q_2}_i$ *is defined in Table 56.*

Examine the matrix of derivatives $\frac{\partial}{\partial \boldsymbol{\eta}_{\boldsymbol{\delta}}'}(\mathbf{C}_1'\boldsymbol{\delta} - \mathbf{c}_0)$*:*

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\eta}_{\boldsymbol{\delta}}'}(\mathbf{C}_1'\boldsymbol{\delta} - \mathbf{c}_0) &= \frac{\partial}{\partial \boldsymbol{\eta}_{\boldsymbol{\delta}}'}\mathbf{C}_1'\mathbf{T}_2(\boldsymbol{\xi}_\delta \odot \boldsymbol{\xi}_\delta) \\
&= \frac{\partial}{\partial \boldsymbol{\eta}_{\boldsymbol{\delta}}'}\mathbf{C}_1'\mathbf{T}_2\sum_{i=1}^{q_2}\left(\mathbf{e}_i^{q_2}\boldsymbol{\xi}_{\boldsymbol{\delta}}'\mathbf{e}_i^{q_2}\mathbf{e}_i^{q_2'}\boldsymbol{\xi}_\delta\right) \\
&= \frac{\partial}{\partial \boldsymbol{\eta}_{\boldsymbol{\delta}}'}\mathbf{C}_1'\mathbf{T}_2\sum_{i=1}^{q_2}\left[\mathbf{e}_i^{q_2}(\mathbf{V}_1\boldsymbol{\eta}_{\boldsymbol{\delta}} + \mathbf{V}_2\boldsymbol{\theta}_{\boldsymbol{\delta}})'\mathbf{e}_i^{q_2}\mathbf{e}_i^{q_2'}(\mathbf{V}_1\boldsymbol{\eta}_{\boldsymbol{\delta}} + \mathbf{V}_2\boldsymbol{\theta}_{\boldsymbol{\delta}})\right] \\
&= 2\mathbf{C}_1'\mathbf{T}_2\sum_{i=1}^{q_2}\left[\mathbf{e}_i^{q_2}(\mathbf{V}_1\boldsymbol{\eta}_{\boldsymbol{\delta}} + \mathbf{V}_2\boldsymbol{\theta}_{\boldsymbol{\delta}})'\mathbf{e}_i^{q_2}\mathbf{e}_i^{q_2'}\frac{\partial}{\partial \boldsymbol{\eta}_{\boldsymbol{\delta}}'}(\mathbf{V}_1\boldsymbol{\eta}_{\boldsymbol{\delta}} + \mathbf{V}_2\boldsymbol{\theta}_{\boldsymbol{\delta}})\right] \\
&= 2\mathbf{C}_1'\mathbf{T}_2\sum_{i=1}^{q_2}\left[\mathbf{e}_i^{q_2}(\mathbf{V}_1\boldsymbol{\eta}_{\boldsymbol{\delta}} + \mathbf{V}_2\boldsymbol{\theta}_{\boldsymbol{\delta}})'\mathbf{e}_i^{q_2}\mathbf{e}_i^{q_2'}\mathbf{V}_1\right] \\
&= 2\mathbf{C}_1'\mathbf{T}_2\sum_{i=1}^{q_2}\left(\mathbf{e}_i^{q_2}\boldsymbol{\xi}_{\boldsymbol{\delta}}'\mathbf{e}_i^{q_2}\mathbf{e}_i^{q_2'}\right)\mathbf{V}_1 \\
&= 2\mathbf{C}_1'\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}\mathbf{V}_1 \\
&= 2\mathbf{W}_{\boldsymbol{\delta}}\mathbf{V}_1.
\end{aligned}
$$

Note that if $\mathbf{W}_{\boldsymbol{\delta}}$ *does not have full row-rank, then one or more linear functions of the restrictions* $\mathbf{C}_1'\boldsymbol{\delta} = \mathbf{c}_0$ *are non-informative and can be deleted.*

By the singular value decomposition of $\mathbf{W}_{\boldsymbol{\delta}}$*, one can write* $2\mathbf{W}_{\boldsymbol{\delta}}\mathbf{V}_1$ *as follows:*

$$2\mathbf{W}_{\boldsymbol{\delta}}\mathbf{V}_1 = 2\mathbf{U}_{\boldsymbol{\delta}}\mathbf{D}_{\boldsymbol{\delta}}\mathbf{V}_{\boldsymbol{\delta}}'\mathbf{V}_1 = 2\mathbf{U}_{\boldsymbol{\delta}}\mathbf{D}_{\boldsymbol{\delta},1}\mathbf{V}_{\boldsymbol{\delta},1}'\mathbf{V}_1 = 2\mathbf{U}_{\boldsymbol{\delta}}\mathbf{D}_{\boldsymbol{\delta},1}\mathbf{V}_1'\mathbf{V}_1 = 2\mathbf{U}_{\boldsymbol{\delta}}\mathbf{D}_{\boldsymbol{\delta},1}.$$

It follows that $\mathbf{U}_{\boldsymbol{\delta}}$ *and* $\mathbf{D}_{\boldsymbol{\delta},1}$ *are nonsingular because* $\mathbf{U}_{\boldsymbol{\delta}} \in \mathcal{O}(r_c)$ *and* $\mathbf{D}_{\boldsymbol{\delta},1} \in \mathcal{D}^+(r_c)$*. Therefore, the matrix of derivatives* $\partial\,(\mathbf{C}_1'\boldsymbol{\delta} - \mathbf{c}_0)/\partial\,\boldsymbol{\eta}_{\boldsymbol{\delta}}'$ *is nonsingular. By the implicit function theorem (Callahan [61], 2010, Corollary 6.13, Page 207), the parameter* $\boldsymbol{\eta}_{\boldsymbol{\delta}}$ *is an implicit function of* $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ *and* $\partial\boldsymbol{\eta}_{\boldsymbol{\delta}}/\partial\boldsymbol{\theta}_{\boldsymbol{\delta}}'$ *exists. It also follows that the parameter* $\boldsymbol{\xi}_\delta$ *is an implicit function of* $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ *and* $\partial\boldsymbol{\xi}_\delta/\partial\boldsymbol{\theta}_{\boldsymbol{\delta}}'$ *exists because* $\boldsymbol{\xi}_\delta = \mathbf{V}_1\boldsymbol{\eta}_{\boldsymbol{\delta}} + \mathbf{V}_2\boldsymbol{\theta}_{\boldsymbol{\delta}}$*.*

c. *The first derivative of* $\boldsymbol{\delta}$ *with respect to* $\boldsymbol{\theta_\delta}$ *can be written as*

$$
\begin{aligned}
\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta} &= \frac{\partial}{\partial\,\boldsymbol{\theta}'_{\boldsymbol{\delta}}}\mathbf{T}_2(\boldsymbol{\xi}_\delta \odot \boldsymbol{\xi}_\delta) \\
&= \frac{\partial}{\partial\,\boldsymbol{\theta}'_{\boldsymbol{\delta}}}\mathbf{T}_2\sum_{i=1}^{q_2}\left(\mathbf{e}_i^{q_2}\boldsymbol{\xi}'_{\boldsymbol{\delta}}\mathbf{e}_i^{q_2}\mathbf{e}_i^{q_2'}\boldsymbol{\xi}_\delta\right) \\
&= 2\mathbf{T}_2\sum_{i=1}^{q_2}\left(\mathbf{e}_i^{q_2}\boldsymbol{\xi}'_{\boldsymbol{\delta}}\mathbf{e}_i^{q_2}\mathbf{e}_i^{q_2'}\mathbf{D}^{(1)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta}\right) \\
&= 2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{D}^{(1)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta}.
\end{aligned}
$$

All derivatives of $\mathbf{C}_1'\boldsymbol{\delta}$ *with respect to* $\boldsymbol{\theta_\delta}$ *are* $\mathbf{0}$ *for all* $\boldsymbol{\theta_\delta}$ *in the parameter space because* $\mathbf{C}_1'\boldsymbol{\delta} = \mathbf{c}_0\ \forall\ \boldsymbol{\theta_\delta} \in \Theta_{\boldsymbol{\delta}}$*. It follows that*

$$
\begin{gathered}
\mathbf{C}_1'\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta} = \frac{\partial}{\partial\,\boldsymbol{\theta}'_{\boldsymbol{\delta}}}\mathbf{C}_1'\mathbf{T}_2(\boldsymbol{\xi}_\delta \odot \boldsymbol{\xi}_\delta) = 2\mathbf{C}_1'\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{D}^{(1)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta} = 2\mathbf{W}_{\boldsymbol{\delta}}\mathbf{D}^{(1)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta} = \mathbf{0} \\
\Longrightarrow \mathbf{W}_{\boldsymbol{\delta}}\left(\mathbf{V}_1\mathbf{D}^{(1)}_{\boldsymbol{\eta_\delta};\boldsymbol{\theta}'_\delta} + \mathbf{V}_2\right) = \mathbf{0} \\
\Longrightarrow \mathbf{W}_{\boldsymbol{\delta}}\mathbf{V}_1\mathbf{D}^{(1)}_{\boldsymbol{\eta_\delta};\boldsymbol{\theta}'_\delta} = \mathbf{0} \text{ because } \mathbf{W}_{\boldsymbol{\delta}}\mathbf{V}_2 = \mathbf{U}_{\boldsymbol{\delta}}\mathbf{D}_{\boldsymbol{\delta},1}\mathbf{V}_1'\mathbf{V}_2 = \mathbf{0} \\
\Longrightarrow \mathbf{D}^{(1)}_{\boldsymbol{\eta_\delta};\boldsymbol{\theta}'_\delta} = \mathbf{0} \text{ because } \mathbf{W}_{\boldsymbol{\delta}}\mathbf{V}_1 \text{ is nonsingular} \\
\Longrightarrow \mathbf{D}^{(1)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta} = \mathbf{V}_2 \\
\Longrightarrow \mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta} = 2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{V}_2.
\end{gathered}
$$

The second derivative of $\boldsymbol{\delta}$ *with respect to* $\boldsymbol{\theta_\delta}$ *can be written as*

$$
\begin{aligned}
\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} &= \frac{\partial}{\partial\,\boldsymbol{\theta}'_{\boldsymbol{\delta}}}2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{D}^{(1)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta} \\
&= 2\mathbf{T}_2\sum_{i=1}^{q_2}\left[\mathbf{e}_i^{q_2}\mathbf{e}_i^{q_2'}\mathbf{D}^{(1)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta}\left(\mathbf{I}_{\nu_2}\otimes\mathbf{e}_i^{q_2'}\mathbf{D}^{(1)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta}\right)\right] + 2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{D}^{(2)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} \\
&= 2\mathbf{T}_2\sum_{i=1}^{q_2}\mathbf{e}_i^{q_2}\left[\left(\mathbf{e}_i^{q_2'}\mathbf{D}^{(1)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta}\right)\otimes\left(\mathbf{e}_i^{q_2'}\mathbf{D}^{(1)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta}\right)\right] + 2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{D}^{(2)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} \\
&= 2\mathbf{T}_2\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta}\right)' + 2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{D}^{(2)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} \\
&= 2\mathbf{T}_2\left(\mathbf{V}_2' * \mathbf{V}_2'\right)' + 2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{D}^{(2)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta}.
\end{aligned}
$$

It follows that

$$\mathbf{C}_1'\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} = 2\mathbf{C}_1'\mathbf{T}_2\left(\mathbf{V}_2' * \mathbf{V}_2'\right)' + 2\mathbf{C}_1'\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{D}^{(2)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} = \mathbf{0}$$

because $\mathbf{C}_1'\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} = \mathbf{0}$ *for all* $\boldsymbol{\theta_\delta}$ *in the parameter space.*

$$\Longrightarrow 2\mathbf{C}_1'\mathbf{T}_2\left(\mathbf{V}_2' * \mathbf{V}_2'\right)' + 2\mathbf{W}_{\boldsymbol{\delta}}\mathbf{D}^{(2)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} = \mathbf{0}$$

$$\Longrightarrow 2\mathbf{C}_1'\mathbf{T}_2\left(\mathbf{V}_2' * \mathbf{V}_2'\right)' + 2\mathbf{W}_{\boldsymbol{\delta}}\mathbf{V}_1\mathbf{D}^{(2)}_{\boldsymbol{\eta_\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} = \mathbf{0}$$

$$\Longrightarrow \mathbf{D}^{(2)}_{\boldsymbol{\eta_\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} = -\mathbf{V}_1\mathbf{W}^{+}_{\boldsymbol{\delta}}\mathbf{C}_1'\mathbf{T}_2\left(\mathbf{V}_2' * \mathbf{V}_2'\right)'$$

$$\Longrightarrow \mathbf{D}^{(2)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} = -\mathbf{W}^{+}_{\boldsymbol{\delta}}\mathbf{C}_1'\mathbf{T}_2\left(\mathbf{V}_2' * \mathbf{V}_2'\right)'$$

$$\Longrightarrow \mathbf{D}^{(2)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} = -\mathbf{D}^{-1}_{\boldsymbol{\xi_\delta}}\mathbf{P}_{\boldsymbol{\xi_\delta}}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta'} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta'}\right)'$$

$$\Longrightarrow \mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} = 2\mathbf{T}_2\left(\mathbf{I}_{q_2} - \mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{D}^{-1}_{\boldsymbol{\xi_\delta}}\mathbf{P}_{\boldsymbol{\xi_\delta}}\right)\left(\mathbf{V}_2' * \mathbf{V}_2'\right)'$$

$$\Longrightarrow \mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} = 2\mathbf{T}_2\left(\mathbf{I}_{q_2} - \mathbf{P}_{\boldsymbol{\xi_\delta}}\right)\left(\mathbf{V}_2' * \mathbf{V}_2'\right)'.$$

The third derivative of $\boldsymbol{\delta}$ *with respect to* $\boldsymbol{\theta_\delta}$ *can be written as*

$$\begin{aligned}
\mathbf{D}^{(3)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} &= \frac{\partial}{\partial\,\boldsymbol{\theta}_{\boldsymbol{\delta}}'}\left[2\mathbf{T}_2\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta'} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta'}\right)' + 2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{D}^{(2)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'}\right] \\
&= 2\mathbf{T}_2\sum_{i=1}^{q_2}\mathbf{e}_i^{q_2}\left[\left(\mathbf{e}_i^{q_2'}\mathbf{D}^{(2)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'}\right)\otimes\left(\mathbf{e}_i^{q_2'}\mathbf{D}^{(1)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta'}\right)\right] \\
&+ 2\mathbf{T}_2\sum_{i=1}^{q_2}\mathbf{e}_i^{q_2}\left[\left(\mathbf{e}_i^{q_2'}\mathbf{D}^{(2)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'}\right)\otimes\left(\mathbf{e}_i^{q_2'}\mathbf{D}^{(1)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta'}\right)\right]\left[\mathbf{I}_{\nu_2}\otimes\mathbf{I}_{(\nu_2,\nu_2)}\right] \\
&+ 2\mathbf{T}_2\sum_{i=1}^{q_2}\mathbf{e}_i^{q_2}\left[\left(\mathbf{e}_i^{q_2'}\mathbf{D}^{(1)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta'}\right)\otimes\left(\mathbf{e}_i^{q_2'}\mathbf{D}^{(2)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'}\right)\right] \\
&+ 2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{D}^{(3)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} \\
&= 2\mathbf{T}_2\left(\mathbf{D}^{(2)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta'}\right)'\left(\mathbf{I}_{\nu_2}\otimes 2\mathbf{N}_{\nu_2}\right) \\
&+ 2\mathbf{T}_2\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta'} * \mathbf{D}^{(2)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'}\right)' \\
&+ 2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{D}^{(3)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} \\
&= 2\mathbf{T}_2\left[\left(\mathbf{D}^{(2)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta'}\right)'\left(\mathbf{I}_{\nu_2}\otimes 2\mathbf{N}_{\nu_2}\right) + \left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta'} * \mathbf{D}^{(2)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'}\right)'\right] \\
&+ 2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{D}^{(3)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'},
\end{aligned}$$

where $\mathbf{D}^{(3)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta}$ *can be computed by the following steps. It follows that*

$$\mathbf{C}_1'\mathbf{D}^{(3)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} = 2\mathbf{C}_1'\mathbf{T}_2\left(\mathbf{D}^{(2)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta}\right)'(\mathbf{I}_{\nu_2}\otimes 2\mathbf{N}_{\nu_2})$$
$$+2\mathbf{C}_1'\mathbf{T}_2\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta} * \mathbf{D}^{(2)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta}\right)' + 2\mathbf{C}_1'\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{D}^{(3)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} = \mathbf{0}$$

because $\mathbf{C}_1'\mathbf{D}^{(3)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} = \mathbf{0}$ *for all* $\boldsymbol{\theta_\delta}$ *in the parameter space.*

$$\Longrightarrow \mathbf{C}_1'\mathbf{T}_2\left(\mathbf{D}^{(2)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta}\right)'(\mathbf{I}_{\nu_2}\otimes 2\mathbf{N}_{\nu_2})$$
$$+\mathbf{C}_1'\mathbf{T}_2\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta} * \mathbf{D}^{(2)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta}\right)' + \mathbf{W}_{\boldsymbol{\delta}}\mathbf{D}^{(3)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} = \mathbf{0}$$
$$\Longrightarrow \mathbf{C}_1'\mathbf{T}_2\left(\mathbf{D}^{(2)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta}\right)'(\mathbf{I}_{\nu_2}\otimes 2\mathbf{N}_{\nu_2})$$
$$+\mathbf{C}_1'\mathbf{T}_2\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta} * \mathbf{D}^{(2)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta}\right)' + \mathbf{W}_{\boldsymbol{\delta}}\mathbf{V}_1\mathbf{D}^{(3)}_{\boldsymbol{\eta_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} = \mathbf{0}$$
$$\Longrightarrow$$
$$\mathbf{D}^{(3)}_{\boldsymbol{\eta_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} = -\mathbf{V}_1\mathbf{W}^+_{\boldsymbol{\delta}}\mathbf{C}_1'\mathbf{T}_2\left[\left(\mathbf{D}^{(2)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta}\right)'(\mathbf{I}_{\nu_2}\otimes 2\mathbf{N}_{\nu_2}) + \left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta} * \mathbf{D}^{(2)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta}\right)'\right]$$
$$\Longrightarrow$$
$$\mathbf{D}^{(3)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} = -\mathbf{W}^+_{\boldsymbol{\delta}}\mathbf{C}_1'\mathbf{T}_2\left[\left(\mathbf{D}^{(2)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta}\right)'(\mathbf{I}_{\nu_2}\otimes 2\mathbf{N}_{\nu_2}) + \left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta} * \mathbf{D}^{(2)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta}\right)'\right]$$
$$\Longrightarrow$$
$$\mathbf{D}^{(3)}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} = -\mathbf{D}^{-1}_{\boldsymbol{\xi_\delta}}\mathbf{P}_{\boldsymbol{\xi_\delta}}\left[\left(\mathbf{D}^{(2)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta}\right)'(\mathbf{I}_{\nu_2}\otimes 2\mathbf{N}_{\nu_2}) + \left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta} * \mathbf{D}^{(2)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta}\right)'\right]$$
$$\Longrightarrow \mathbf{D}^{(3)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} = 2\mathbf{T}_2\left[\mathbf{I}_{q_2} - \mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{D}^{-1}_{\boldsymbol{\xi_\delta}}\mathbf{P}_{\boldsymbol{\xi_\delta}}\right]\left(\mathbf{D}^{(2)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta}\right)'\mathbf{J}_{\nu_2}$$
$$\Longrightarrow \mathbf{D}^{(3)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} = 2\mathbf{T}_2\left(\mathbf{I}_{q_2} - \mathbf{P}_{\boldsymbol{\xi_\delta}}\right)\left(\mathbf{D}^{(2)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi_\delta};\boldsymbol{\theta}'_\delta}\right)'\mathbf{J}_{\nu_2}.$$

The first three derivative expressions were checked numerically.

d. It is true that $\mathbf{P}_{\boldsymbol{\xi_\delta}}$ *is a projection operator because*

$$\begin{aligned}\left[\mathbf{P}_{\boldsymbol{\xi_\delta}}\right]^2 &= \left(\mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{W}^+_{\boldsymbol{\delta}}\mathbf{C}_1'\mathbf{T}_2\right)\left(\mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{W}^+_{\boldsymbol{\delta}}\mathbf{C}_1'\mathbf{T}_2\right)\\ &= \mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{W}^+_{\boldsymbol{\delta}}\mathbf{W}_{\boldsymbol{\delta}}\mathbf{W}^+_{\boldsymbol{\delta}}\mathbf{C}_1'\mathbf{T}_2\\ &= \mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{W}^+_{\boldsymbol{\delta}}\mathbf{C}_1'\mathbf{T}_2 \quad \textit{because } \mathbf{W}^+_{\boldsymbol{\delta}}\mathbf{W}_{\boldsymbol{\delta}}\mathbf{W}^+_{\boldsymbol{\delta}} = \mathbf{W}^+_{\boldsymbol{\delta}}\end{aligned}$$

$$= \mathbf{P}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}.$$

$\square$

B.3. Proof of Theorem 3

Theorem 3. [Original result]. *Let $\boldsymbol{\delta}$ be parameterized as model 4 in Table 4. Suppose that the $q \times q_2$ design matrix $\mathbf{T}_2$ has full column-rank. Then $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}$ has full column-rank if and only if $\boldsymbol{\xi}_{\boldsymbol{\delta},i} \neq 0$ for $i = 1, 2, \ldots, q_2$.*

Proof: [Independent Proof]. *First, suppose that $\boldsymbol{\xi}_{\boldsymbol{\delta},i} = 0$ for some $i \in \{1, 2, \ldots, q_2\}$, then $\mathbf{W}_{\boldsymbol{\delta}}\mathbf{e}_i^{q_2} = \mathbf{C}'_1\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}\mathbf{e}_i^{q_2} = \mathbf{0} \Longrightarrow \mathbf{U}_{\boldsymbol{\delta}}\mathbf{D}_{\boldsymbol{\delta},1}\mathbf{V}'_1\mathbf{e}_i^{q_2} = \mathbf{0}$. Therefore, $\mathbf{e}_i^{q_2} \in \mathcal{N}(\mathbf{V}'_1) = \mathcal{R}(\mathbf{V}_2) \Longrightarrow \mathbf{e}_i^{q_2} = \mathbf{V}_2\mathbf{b}$ for some nonzero vector $\mathbf{b}$. By Theorem 2 part c, $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta} = 2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}\mathbf{V}_2$. Accordingly, $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\mathbf{b} = 2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}\mathbf{V}_2\mathbf{b} = 2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}\mathbf{e}_i^{q_2} = \mathbf{0}$, which implies $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}$ does not have full column-rank.*

Second, suppose that $\boldsymbol{\xi}_{\boldsymbol{\delta},i} \neq 0$ for $i = 1, 2, \ldots, q_2$. Let $\mathbf{w}$ be a vector such that $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\mathbf{w} = \mathbf{0} \Longrightarrow 2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}\mathbf{V}_2\mathbf{w} = \mathbf{0} \Longrightarrow \mathbf{D}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}\mathbf{V}_2\mathbf{w} = \mathbf{0}$ because $\mathbf{T}_2$ has full column-rank $\Longrightarrow \mathbf{V}_2\mathbf{w} = \mathbf{0}$ because $\mathbf{D}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}$ has full column-rank $\Longrightarrow \mathbf{w} = \mathbf{0}$ because $\mathbf{V}_2$ has full column-rank. Now it can be concluded that $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}$ has full column-rank. $\square$

B.4. Proof of Theorem 4

Theorem 4. [Original result]. *Let $\boldsymbol{\delta}$ be parameterized as model 4 in Table 4. If $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}$ does not have full column-rank, then $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ is not identified.*

Proof: [Proof With Advisor's Help]. *Suppose that $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}$ does not have full column-rank, then $\boldsymbol{\xi}_{\boldsymbol{\delta},i} = 0$ for some $i \in \{1, 2, \ldots, q_2\}$ by Theorem 3 $\Longrightarrow \mathbf{W}_{\boldsymbol{\delta}}\mathbf{e}_i^{q_2} = \mathbf{C}'_1\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi}_{\boldsymbol{\delta}}}\mathbf{e}_i^{q_2} = \mathbf{0} \Longrightarrow \mathbf{U}_{\boldsymbol{\delta}}\mathbf{D}_{\boldsymbol{\delta},1}\mathbf{V}'_1\mathbf{e}_i^{q_2} = \mathbf{0} \Longrightarrow \mathbf{e}_i^{q_2} \in \mathcal{N}(\mathbf{V}'_1) = \mathcal{R}(\mathbf{V}_2) \Longrightarrow \mathbf{e}_i^{q_2} = \mathbf{V}_2\mathbf{b}$ for some nonzero full column-rank vector $\mathbf{b}$.*

Note that $\mathbf{e}_i^{q_2\prime}\boldsymbol{\xi}_\delta = \mathbf{0} \Longrightarrow \mathbf{e}_i^{q_2\prime}\boldsymbol{\xi}_\delta = \mathbf{b}'\mathbf{V}_2'(\mathbf{V}_1\boldsymbol{\eta_\delta} + \mathbf{V}_2\boldsymbol{\theta_\delta}) = \mathbf{0}$ *by Theorem 2 part a* $\Longrightarrow \mathbf{b}'\boldsymbol{\theta_\delta} = \mathbf{0}$ *because* $\mathbf{V}_2'\mathbf{V}_1 = \mathbf{0}$ *and* $\mathbf{V}_2'\mathbf{V}_2 = \mathbf{I}_{q_2-r_c} \Longrightarrow \boldsymbol{\theta_\delta} = (\mathbf{I}_{q_2-r_c} - \mathbf{H}_\mathbf{b})\,\boldsymbol{\theta_\delta}$ *where* $\mathbf{H}_\mathbf{b} = \text{ppo}(\mathbf{b}) = \mathbf{b}(\mathbf{b}\mathbf{b}')^{-1}\mathbf{b}'$. *Accordingly,* $\boldsymbol{\xi}_\delta$ *can be written as* $\boldsymbol{\xi}_\delta = \mathbf{V}_1\boldsymbol{\eta_\delta} + \mathbf{V}_2\boldsymbol{\theta_\delta} = \mathbf{V}_1\boldsymbol{\eta_\delta} + \mathbf{V}_2(\mathbf{I}_{q_2-r_c} - \mathbf{H}_\mathbf{b})\,\boldsymbol{\theta_\delta} = \mathbf{V}_1\boldsymbol{\eta_\delta} + \mathbf{V}_2^*\boldsymbol{\theta_\delta} = \mathbf{V}_1\boldsymbol{\eta_\delta} + \mathbf{V}_2^*(\boldsymbol{\theta_\delta} + \mathbf{b}\alpha)$, *where* $\mathbf{V}_2^* = \mathbf{V}_2(\mathbf{I}_{q_2-r_c} - \mathbf{H}_\mathbf{b})$ *and* α *is arbitrary because* $\mathbf{V}_2^*\mathbf{b} = \mathbf{0}$. *It follows that* $\boldsymbol{\delta}(\boldsymbol{\theta_\delta}) = \boldsymbol{\delta}(\boldsymbol{\theta}_\delta^*)$, *where* $\boldsymbol{\theta_\delta} \neq \boldsymbol{\theta}_\delta^*$. *By the definition of identification provided in § 1.3.1,* $\boldsymbol{\theta_\delta}$ *is not identified.* □

B.5. Proof of Theorem 5

Theorem 5. [Original result]. *Let* $\boldsymbol{\delta}$ *be parameterized as model 4 in Table 4. If* $\mathbf{T_2}$ *has full column-rank and* $\xi_{\boldsymbol{\delta},i} \neq 0$ *for* $i = 1, 2, \ldots, q_2$, *then*

a. The dimension of $\boldsymbol{\theta_\delta}$ *is* $\nu_2 = q_2 - r_c$

b. A special case of part a: if $\mathbf{T}_2 = \bigoplus_{i=1}^{k}\mathbf{1}_{m_i}$ *and* $\sum_{i=1}^{k} m_i = q$, *then* $\nu_2 = k - r_c$.

Proof: [Independent Proof].

a. By Theorem 3, $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta'}$ *has full column-rank because* $\mathbf{T_2}$ *has full column-rank and* $\xi_{\boldsymbol{\delta},i} \neq 0$ *for* $i = 1, 2, \ldots, q_2$. *The dimension of* $\boldsymbol{\theta_\delta}$ *is equal to the number of the columns in* $\mathbf{V}_2$ *because* $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta'} = 2\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi_\delta}}\mathbf{V}_2$ *and* $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta'}$ *has full column-rank. The number of the columns in* $\mathbf{V}_2$ *is* $q_2 - r_c$ *by Theorem 2. Therefore,* $\nu_2 = q_2 - r_c$.

b. $\mathbf{T}_2 = \bigoplus_{i=1}^{k}\mathbf{1}_{m_i} \Longrightarrow q_2 = k \Longrightarrow \nu_2 = k - r_c$. □

B.5.1. Proof of corollary 5.1

Corollary 5.1. [Original result]. *Let* $\boldsymbol{\delta}$ *be parameterized as model 4 in Table 4. If* $\mathbf{T_2} = \mathbf{I}_q$, *no entry in* $\boldsymbol{\xi}_\delta$ *is* $\mathbf{0}$, $\mathbf{C_1} = \mathbf{1}_q$ *and* $\mathbf{c}_0 = q$, *then* $\mathbf{V}_1 = \boldsymbol{\xi}_\delta/\sqrt{\boldsymbol{\xi}_\delta{}'\boldsymbol{\xi}_\delta}$, $\mathcal{R}(\mathbf{V}_2) =$

$\mathcal{N}(\boldsymbol{\xi}_\delta{}')$, $\boldsymbol{\theta_\delta} = \mathbf{0}$ *and* $\nu_2 = q-1$.

Proof: [Proof With Advisor's Help]. *If* $\mathbf{T_2} = \mathbf{I}_q$, $\mathbf{C}_1 = \mathbf{1}_q$ *and* $\mathbf{c}_0 = q$, *then*

$$\mathbf{W_\delta} = \mathbf{C}_1'\mathbf{T}_2\mathbf{D}_{\boldsymbol{\xi_\delta}} = \mathbf{1}_q'\mathbf{D}_{\boldsymbol{\xi_\delta}} = \boldsymbol{\xi}_\delta{}',$$

where $\mathbf{W_\delta}$ *is a* $1\times q$ *row vector. Accordingly, the singular value decomposition of* $\mathbf{W_\delta}$ *as follows:*

$$\mathbf{W_\delta} = \boldsymbol{\xi}_\delta{}' = \mathbf{U}_1\mathbf{D}_1\mathbf{V}_1',$$

where $\mathbf{U}_1 = 1$, $\mathbf{D}_1 = \sqrt{\boldsymbol{\xi}_\delta{}'\boldsymbol{\xi}_\delta}$, *and* $\mathbf{V}_1 = \boldsymbol{\xi}_\delta/\sqrt{\boldsymbol{\xi}_\delta{}'\boldsymbol{\xi}_\delta}$. *It follows that* $\mathbf{D}_1 = \sqrt{\boldsymbol{\xi}_\delta{}'\boldsymbol{\xi}_\delta} = \sqrt{q}$ *and* $\mathbf{V}_1 = \boldsymbol{\xi}_\delta/\sqrt{\boldsymbol{\xi}_\delta{}'\boldsymbol{\xi}_\delta} = \boldsymbol{\xi}_\delta/\sqrt{q}$ *because* $\mathbf{C}_1'\boldsymbol{\delta} = \mathbf{1}_q'\mathbf{I}_q\boldsymbol{\xi}_\delta{}^{\odot 2} = \boldsymbol{\xi}_\delta{}'\boldsymbol{\xi}_\delta$, $\mathbf{c}_0 = q$, *and* $\mathbf{C}_1'\boldsymbol{\delta} = \mathbf{c}_0$. *It follows that* $\mathbf{V}_2'\boldsymbol{\xi}_\delta = \mathbf{0}$ *because* $\mathbf{V_\delta} = \begin{pmatrix}\mathbf{V}_{\boldsymbol{\delta},1} & \mathbf{V}_{\boldsymbol{\delta},2}\end{pmatrix} \in \mathcal{O}(q)$ *and* $\mathcal{R}(\mathbf{V}_2) = \mathcal{N}(\boldsymbol{\xi}_\delta{}')$. *Accordingly,* $\boldsymbol{\theta_\delta} = \mathbf{0}$ *because Theorem 2 part a that* $\boldsymbol{\theta_\delta} = \mathbf{V}_2'\boldsymbol{\xi}_\delta$. *By Theorem 5 part b,* $\nu_2 = q-1$. □

B.6. Proof of Theorem 6

Theorem 6. [Boik [35], Supplement, 2009, Theorem 1, Page 60].

$$\mathbf{GG}' = \mathbf{I}_q \Longleftrightarrow \mathbf{D}_q'\operatorname{vec}(\mathbf{GG}' - \mathbf{I}_q) = \mathbf{0}.$$

Proof: [Independent Proof].

($\Longrightarrow$) *If* $\mathbf{GG}' = \mathbf{I}_q$, *then* $\operatorname{vec}(\mathbf{GG}' - \mathbf{I}_q) = \mathbf{0}$. *Furthermore,* $\mathbf{D}_q'\operatorname{vec}(\mathbf{GG}' - \mathbf{I}_q) = \mathbf{0}$.

($\Longleftarrow$) *If* $\mathbf{D}_q'\operatorname{vec}(\mathbf{GG}' - \mathbf{I}_q) = \mathbf{0}$, *then* $\mathbf{D}_q(\mathbf{D}_q'\mathbf{D}_q)^{-1}\mathbf{D}_q'\operatorname{vec}(\mathbf{GG}' - \mathbf{I}_q) = \mathbf{0}$. *It follows that* $\mathbf{N}_q\operatorname{vec}(\mathbf{GG}' - \mathbf{I}_q) = \mathbf{0}$. *Accordingly,* $\operatorname{vec}(\mathbf{GG}' - \mathbf{I}_q) = \mathbf{0}$, *because* $\mathbf{GG}' - \mathbf{I}_q$ *is symmetric.*

□

B.7. Proof of Theorem 7

Theorem 7. [Boik [35], Supplement, 2009, Theorem 1, Page 61]. *Define* $\mathbf{W}_\gamma$ *as*

$$\mathbf{W}_\gamma \stackrel{\text{def}}{=} \mathbf{C}_3' \left(\mathbf{\Gamma}\mathbf{\Delta} \otimes \mathbf{\Gamma}\right) 2\mathbf{N}_q^\perp \mathbf{A}_2,$$

where $\mathbf{C}_3'$ *is defined in (39),* $\mathbf{A}_2$ *is defined in (36) and* $\mathbf{N}_p^\perp$ *is defined in Table 56. Assume the* $r_k \times \nu_3^*$ *matrix* $\mathbf{W}_\gamma$ *has full row-rank, where* $r_k = q - 1$ *and* ν_3^* *is defined in (34). Write the singular value decomposition of* $\mathbf{W}_\gamma$ *as follows:*

$$\begin{aligned}
&\mathbf{W}_\gamma = \mathbf{U}_\gamma \mathbf{D}_\gamma \mathbf{V}_\gamma', \text{ where } \mathbf{U}_\gamma \in \mathcal{O}(r_k)\\
&\mathbf{V}_\gamma = \begin{pmatrix} \mathbf{V}_{\gamma,1} & \mathbf{V}_{\gamma,2} \end{pmatrix} \in \mathcal{O}(\nu_3^*), \quad \mathbf{V}_{\gamma,1} \in \mathcal{O}(\nu_3^*, r_k),\\
&\mathbf{D}_\gamma = \begin{pmatrix} \mathbf{D}_{\gamma,1} & \mathbf{0}_{r_k \times (\nu_3^* - r_k)} \end{pmatrix}, \text{ and } \mathbf{D}_{\gamma,1} \in \mathcal{D}^+(r_k).
\end{aligned}$$

Choose $\mathbf{V}_3$ *to be any matrix whose columns form a basis for* $\mathcal{R}(\mathbf{W}_\gamma')$*, then the matrix* $\partial\, \mathbf{h}(\boldsymbol{\eta}_\gamma; \boldsymbol{\theta}_\delta, \boldsymbol{\theta}_\gamma)/\partial\, \boldsymbol{\eta}_\gamma' \Big|_{\mathbf{G}=\mathbf{I}}$ *is nonsingular and* $\boldsymbol{\eta}_\gamma$ *is an implicit function of* $(\boldsymbol{\theta}_\delta, \boldsymbol{\theta}_\gamma)$ *in a neighborhood of* $\mathbf{G} = \mathbf{I}$*.*

Proof: [Proof With Advisor's Help]. *The structure of (44) is adapted from* §18 *in the supplement of Boik [33].*

Write $\mathbf{G}$ *as in (40) and define* $\mathbf{H}$ *as*

$$\begin{aligned}
\mathbf{H} &\stackrel{\text{def}}{=} \frac{\partial\, \mathbf{h}(\boldsymbol{\eta}_\gamma; \boldsymbol{\theta}_\delta, \boldsymbol{\theta}_\gamma)}{\partial\, \boldsymbol{\eta}_\gamma'}\Big|_{\mathbf{G}=\mathbf{I}} = 2\begin{pmatrix} \mathbf{D}_q'\mathbf{A}_1 & \mathbf{D}_q'\mathbf{A}_2\mathbf{V}_3 \\ \mathbf{C}_3'(\mathbf{\Gamma}\mathbf{\Delta} \otimes \mathbf{\Gamma})\mathbf{A}_1 & \mathbf{C}_3'(\mathbf{\Gamma}\mathbf{\Delta} \otimes \mathbf{\Gamma})\mathbf{A}_2\mathbf{V}_3 \end{pmatrix}\\
&= 2\begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix},
\end{aligned}$$

where $\mathbf{H}_{11} = \mathbf{D}_q'\mathbf{A}_1$*,* $\mathbf{H}_{12} = \mathbf{D}_q'\mathbf{A}_2\mathbf{V}_3$*,* $\mathbf{H}_{21} = \mathbf{C}_3'(\mathbf{\Gamma}\mathbf{\Delta} \otimes \mathbf{\Gamma})\mathbf{A}_1$ *and* $\mathbf{H}_{22} = \mathbf{C}_3'(\mathbf{\Gamma}\mathbf{\Delta} \otimes \mathbf{\Gamma})\mathbf{A}_2\mathbf{V}_3$*. Then,*

$$\left|\mathbf{H}\right| = 2^{\frac{q^2+3q-2}{2}}\left|\mathbf{H}_{11}\right|\left|\mathbf{H}_{22} - \mathbf{H}_{21}\mathbf{H}_{11}^{-1}\mathbf{H}_{12}\right|$$

$$
\begin{aligned}
& \quad \textit{because } \dim(\mathbf{H}) \textit{ is } \frac{q^2+3q-2}{2} \times \frac{q^2+3q-2}{2} \\
&= 2^{\frac{q^2+3q-2}{2}} \left| \mathbf{I}_{\frac{q(q+1)}{2}} \right| \left| \mathbf{C}_3'(\mathbf{\Gamma\Delta} \otimes \mathbf{\Gamma})\mathbf{A}_2\mathbf{V}_3 - \mathbf{C}_3'(\mathbf{\Gamma\Delta} \otimes \mathbf{\Gamma})\mathbf{A}_1\mathbf{D}_q'\mathbf{A}_2\mathbf{V}_3 \right| \\
& \quad \textit{because } \mathbf{D}_q'\mathbf{A}_1 = \mathbf{I}_{\frac{q(q+1)}{2}} \textit{ in (42) part (a)} \\
&= 2^{\frac{q^2+3q-2}{2}} \left| \mathbf{C}_3'(\mathbf{\Gamma\Delta} \otimes \mathbf{\Gamma})(\mathbf{A}_2 - \mathbf{A}_1\mathbf{D}_q'\mathbf{A}_2)\mathbf{V}_3 \right| \\
&= 2^{\frac{q^2+3q-2}{2}} \left| \mathbf{C}_3'(\mathbf{\Gamma\Delta} \otimes \mathbf{\Gamma})(\mathbf{I}_{q^2} - \mathbf{K}_{q,q})\mathbf{A}_2\mathbf{V}_3 \right| \\
& \quad \textit{because } \mathbf{A}_1\mathbf{D}_q'\mathbf{A}_2 = \mathbf{K}_{q,q}\mathbf{A}_2 \textit{ in (42) part (b)} \\
&= 2^{\frac{q^2+3q-2}{2}} \left| \mathbf{C}_3'(\mathbf{\Gamma\Delta} \otimes \mathbf{\Gamma})2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{V}_3 \right| \\
&= 2^{\frac{q^2+3q-2}{2}} \left| \mathbf{W}_{\gamma}\mathbf{V}_3 \right|.
\end{aligned}
$$

Note that if $\mathbf{W}_{\gamma}$ *does not have full row-rank, then one or more linear functions of the restrictions* $\mathbf{h}(\boldsymbol{\eta}_{\gamma}; \boldsymbol{\theta}_{\boldsymbol{\delta}}, \boldsymbol{\theta}_{\gamma})$ *in (41) are non-informative and can be deleted.*

It follows that $\mathbf{V}_3 = \mathbf{W}_{\gamma}'\mathbf{B}$*, where* $\mathbf{B}$ *is nonsingular, because columns in* $\mathbf{V}_3$ *form a basis for* $\mathcal{R}(\mathbf{W}_{\gamma}')$ *and* $\mathbf{W}_{\gamma}'$ *has full column-rank. Accordingly,* $\mathbf{W}_{\gamma}\mathbf{V}_3 = \mathbf{W}_{\gamma}\mathbf{W}_{\gamma}'\mathbf{B}$ *is nonsingular because* $\mathbf{W}_{\gamma}\mathbf{W}_{\gamma}'$ *is nonsingular. Therefore,* $\mathbf{H}$ *is nonsingular.*

By the implicit function theorem, if $\mathbf{H}$ *is nonsingular, then* $\boldsymbol{\eta}_{\gamma}$ *is an implicit function of* $(\boldsymbol{\theta}_{\boldsymbol{\delta}}, \boldsymbol{\theta}_{\gamma})$ *in a neighborhood of* $\mathbf{G} = \mathbf{I}$*. Note that* $\mathbf{\Gamma}_0 = \mathbf{\Gamma} \Longleftrightarrow \mathbf{G} = \mathbf{I}_p$ *because* $\mathbf{G} = \mathbf{\Gamma}_0'\mathbf{\Gamma}$. □

B.8. Proof of Theorem 8

Theorem 8. [Boik [35], Supplement, 2009, Theorem 1, Page 61]. $\mathbf{G} = \mathbf{I}_q \Longleftrightarrow \boldsymbol{\theta}_{\gamma} = \mathbf{0}$.

Proof: [Proof With Advisor's Help]. *First, suppose that* $\mathbf{G} = \mathbf{I}_q$*. In (36),* $\mathbf{A}_1 = \sum_{j=1}^{q}\sum_{i=j}^{q} \left(\mathbf{e}_j^q \otimes \mathbf{e}_i^q\right) \mathbf{e}_{f_{ij}}^{h\prime}$*, where* $h = \frac{q(q+1)}{2}$ *and* $f_{ij} = \frac{(2q-j)(j-1)}{2} + i$*. Examine*

$$
\begin{aligned}
\mathbf{A}_1\mathbf{A}_1' \operatorname{vec}\mathbf{G} &= \mathbf{A}_1\mathbf{A}_1' \operatorname{vec}\mathbf{I}_q \\
&= \sum_{j=1}^{q}\sum_{i=j}^{q} \left(\mathbf{e}_j^q \otimes \mathbf{e}_i^q\right) \mathbf{e}_{f_{ij}}^{h\prime} \sum_{m=1}^{q}\sum_{n=m}^{q} \mathbf{e}_{f_{nm}}^{h} \left(\mathbf{e}_m^{q\prime} \otimes \mathbf{e}_n^{q\prime}\right) \operatorname{vec}\mathbf{I}_q
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{j=1}^{q}\sum_{i=j}^{q}\sum_{m=1}^{q}\sum_{n=m}^{q}\left(\mathbf{e}_j^q \otimes \mathbf{e}_i^q\right)\mathbf{e}_{f_{ij}}^{h\prime}\mathbf{e}_{f_{nm}}^{h}\left(\mathbf{e}_m^{q\prime} \otimes \mathbf{e}_n^{q\prime}\right)\operatorname{vec}\mathbf{I}_q \\
&= \sum_{j=1}^{q}\sum_{i=j}^{q}\sum_{m=1}^{q}\sum_{n=m}^{q}\left(\mathbf{e}_j^q \otimes \mathbf{e}_i^q\right)1_{f_{ij}=f_{nm}}\left(\mathbf{e}_m^{q\prime} \otimes \mathbf{e}_n^{q\prime}\right)\operatorname{vec}\mathbf{I}_q \\
&= \sum_{j=1}^{q}\sum_{i=j}^{q}\left(\mathbf{e}_j^q \otimes \mathbf{e}_i^q\right)\left(\mathbf{e}_j^{q\prime} \otimes \mathbf{e}_i^{q\prime}\right)\operatorname{vec}\mathbf{I}_q \\
&\quad \textit{because } f_{ij} = f_{nm} \Longrightarrow m = j \textit{ and } n = i \\
&= \sum_{j=1}^{q}\sum_{i=j}^{q}\left(\mathbf{e}_j^q \otimes \mathbf{e}_i^q\right)\operatorname{vec}\left(\mathbf{e}_i^{q\prime}\mathbf{I}_q\mathbf{e}_j^q\right) = \sum_{j=1}^{q}\sum_{i=j}^{q}\left(\mathbf{e}_j^q \otimes \mathbf{e}_i^q\right)1_{j=i} \\
&= \sum_{j=1}^{q}\left(\mathbf{e}_j^q \otimes \mathbf{e}_j^q\right)\operatorname{vec}1 = \sum_{j=1}^{q}\operatorname{vec}\left(\mathbf{e}_j^q\mathbf{e}_j^{q\prime}\right) = \operatorname{vec}\mathbf{I}_q \\
&= \operatorname{vec}\mathbf{G},
\end{aligned}
$$

where the indicator functions $1_{a=b}$ *are defined as* $1_{a=b} = \begin{cases} 1 & \textit{if} \quad a=b, \\ 0 & \textit{otherwise} \end{cases}.$

It can be concluded that when $\mathbf{G} = \mathbf{I}_q$*,* $\operatorname{vec}\mathbf{G} = \mathbf{A}_1\mathbf{A}_1'\operatorname{vec}\mathbf{G}$*. It follows that* $\boldsymbol{\theta}_\gamma^* = \mathbf{A}_2'(\operatorname{vec}\mathbf{G}) = \mathbf{A}_2'\left(\mathbf{A}_1\mathbf{A}_1'\operatorname{vec}\mathbf{I_q}\right) = \left(\mathbf{A}_2'\mathbf{A}_1\right)\mathbf{A}_1'\operatorname{vec}\mathbf{I_q} = \mathbf{0}$ *because* $\mathbf{A}_2'\mathbf{A}_1 = \mathbf{0}$ *and* $\mathbf{A}_2'(\operatorname{vec}\mathbf{G}) = \boldsymbol{\theta}_\gamma^*$ *given in (36). Accordingly,* $\boldsymbol{\theta}_\gamma = \mathbf{0}$ *because* $\boldsymbol{\theta}_\gamma = \mathbf{V}_4'\boldsymbol{\theta}_\gamma^*$ *defined in (40).*

Second, suppose that $\boldsymbol{\theta}_\gamma = \mathbf{0}$*. Let* $\boldsymbol{\eta}_{\gamma,1} = \mathbf{A}_1'\operatorname{vec}(\mathbf{I}_q)$ *and* $\boldsymbol{\eta}_{\gamma,2} = \mathbf{0}$*, then,* $\mathbf{G} = \mathbf{I}_q$ *because* $\mathbf{A}_1\mathbf{A}_1'\operatorname{vec}(\mathbf{I}_q) = \operatorname{vec}(\mathbf{I}_q)$*. Accordingly,* $\mathbf{G} = \mathbf{I}_q$ *is a solution to* $\mathbf{h}(\boldsymbol{\eta}_\gamma; \boldsymbol{\theta}_\delta, \boldsymbol{\theta}_\gamma) = \mathbf{0}$*. By Theorem 7,* $\boldsymbol{\eta}_\gamma = (\boldsymbol{\eta}_{\gamma,1}' \quad \boldsymbol{\eta}_{\gamma,2}')'$ *is an implicit function of* $(\boldsymbol{\theta}_\delta, \boldsymbol{\theta}_\gamma)$ *because* $\mathbf{W}_\gamma$ *has full row-rank and* $\mathbf{V}_3$ *form a basis for* $\mathcal{R}(\mathbf{W}_\gamma')$*. This implies that* $\mathbf{G} = \mathbf{I}_q$ *is the unique solution* $\mathbf{h}(\boldsymbol{\eta}_\gamma; \boldsymbol{\theta}_\delta, \boldsymbol{\theta}_\gamma) = \mathbf{0}$ *when* $\boldsymbol{\theta}_\gamma = \mathbf{0}$*.* □

B.9. Proof of Theorem 9

Theorem 9. [Boik [33], Supplement, 2010, Theorem 29, Page 86]. *The first three derivatives of* vec $\mathbf{G}$ *in* (40) *with respect to* $\boldsymbol{\theta}_\gamma$*, evaluated at* $\boldsymbol{\theta}_\gamma = \mathbf{0}$ *and* $\boldsymbol{\Gamma}_0 = \boldsymbol{\Gamma}$*, can* *be written as follows:*

$$\begin{aligned}
\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} &= 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{V}_4, \\
\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} &= \left[(\mathbf{I}_{q^2}-\mathbf{P}_{\boldsymbol{\gamma}})\,\mathbf{A}_1\mathbf{D}'_q\mathbf{I}'_{q,3} + 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}^{+}_{\boldsymbol{\gamma}}\mathbf{C}'_3(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\right] \\
&\times \left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\right), \text{ and} \\
\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} &= -\left\{2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}^{+}_{\boldsymbol{\gamma}}\mathbf{C}'_3(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q] + (\mathbf{I}_{q^2}-\mathbf{P}_{\boldsymbol{\gamma}})\,\mathbf{A}_1\mathbf{D}'_q\mathbf{I}'_{q,3}\right\} \\
&\times \left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\right)\mathbf{J}_{21,\nu_3}
\end{aligned}$$

where $\mathbf{P}_{\boldsymbol{\gamma}} = 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}^{+}_{\boldsymbol{\gamma}}\mathbf{C}'_3\,(\boldsymbol{\Gamma\Delta}\otimes\boldsymbol{\Gamma})$, $\mathbf{D}_q$, $\mathbf{I}_{q,3}$, $\mathbf{J}_{21,\nu_3}$, $\mathbf{N}_q^{\perp}$ *are defined in Table 56 and* $\mathbf{W}^{+}_{\boldsymbol{\gamma}}$ *is the Moore-Penrose inverse of* $\mathbf{W}_{\boldsymbol{\gamma}}$*.*

Proof: [Independent Proof]. *Denote the derivatives of* $\boldsymbol{\eta}_{\boldsymbol{\gamma}j}$ *for* $j = 1, 2$ *with respect to* $\boldsymbol{\theta}_{\boldsymbol{\gamma}}$*, evaluated at* $(\boldsymbol{\Gamma}_0 = \boldsymbol{\Gamma},\ \boldsymbol{\theta}_{\boldsymbol{\gamma}} = \mathbf{0})$*, by* $\mathbf{D}^{(1)}_{\boldsymbol{\eta}_{\boldsymbol{\gamma},j};\boldsymbol{\theta}'_{\boldsymbol{\gamma}}}$, $\mathbf{D}^{(2)}_{\boldsymbol{\eta}_{\boldsymbol{\gamma},j};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma}$*, and* $\mathbf{D}^{(3)}_{\boldsymbol{\eta}_{\boldsymbol{\gamma},j};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma}$*. For fixed* $\mathbf{V}_j$*;* $j = 1, 2$ *in* (40)*, the derivatives of* $\mathbf{g}$ *can be written as*

$$\begin{aligned}
&\text{(a) } \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} = \mathbf{A}_1\mathbf{D}^{(1)}_{\boldsymbol{\eta}_{\gamma,1};\boldsymbol{\theta}'_\gamma} + \mathbf{A}_2\mathbf{V}_3\mathbf{D}^{(1)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}'_\gamma} + \mathbf{A}_2\mathbf{V}_4, \\
&\text{(b) } \mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} = \mathbf{A}_1\mathbf{D}^{(2)}_{\boldsymbol{\eta}_{\gamma,1};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} + \mathbf{A}_2\mathbf{V}_3\mathbf{D}^{(2)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma}, \text{ and} \\
&\text{(c) } \mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} = \mathbf{A}_1\mathbf{D}^{(3)}_{\boldsymbol{\eta}_{\gamma,1};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} + \mathbf{A}_2\mathbf{V}_3\mathbf{D}^{(3)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma}.
\end{aligned} \tag{187}$$

Recall $\mathbf{h}(\boldsymbol{\eta}_{\boldsymbol{\gamma}};\boldsymbol{\theta}_{\boldsymbol{\delta}},\boldsymbol{\theta}_{\boldsymbol{\gamma}})$ *in (41), where*

$$\mathbf{h}(\boldsymbol{\eta}_{\boldsymbol{\gamma}};\boldsymbol{\theta}_{\boldsymbol{\delta}},\boldsymbol{\theta}_{\boldsymbol{\gamma}}) = \begin{pmatrix} \mathbf{D}'_q\operatorname{vec}(\mathbf{GG}'-\mathbf{I}_q) \\ \mathbf{C}'_3\operatorname{vec}(\boldsymbol{\Phi}) \end{pmatrix} = \mathbf{0}.$$

It follows from the two constraints in $\mathbf{h}(\boldsymbol{\eta}_{\boldsymbol{\gamma}};\boldsymbol{\theta}_{\boldsymbol{\delta}},\boldsymbol{\theta}_{\boldsymbol{\gamma}})$ *that*

$$\frac{\partial^s\mathbf{D}'_q\operatorname{vec}(\mathbf{GG}'-\mathbf{I}_p)}{\left(\partial\,\boldsymbol{\theta}'_{\boldsymbol{\gamma}}\right)^{\otimes s}} = \mathbf{0} \quad \forall \quad \boldsymbol{\theta}_{\boldsymbol{\gamma}}\in\mathring{N}(\mathbf{0}) \quad \text{and} \quad \frac{\partial^s\mathbf{C}'_3\operatorname{vec}(\boldsymbol{\Phi})}{\left(\partial\,\boldsymbol{\theta}'_{\boldsymbol{\gamma}}\right)^{\otimes s}} = \mathbf{0} \quad \forall \quad \boldsymbol{\theta}_{\boldsymbol{\gamma}}\in\mathring{N}(\mathbf{0}),$$

where $s = 1, 2, 3$.

It is shown that expressions for the derivatives of the constraints, evaluated at $(\boldsymbol{\Gamma}_0 = \boldsymbol{\Gamma},\ \boldsymbol{\theta}_{\boldsymbol{\gamma}} = \mathbf{0})$ *or* $(\mathbf{G} = \mathbf{I}_q)$*, are*

$$\text{(a) } \mathbf{D}'_q\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} = \mathbf{0}, \quad \mathbf{C}'_3\,(\boldsymbol{\Gamma\Delta}\otimes\boldsymbol{\Gamma})\,\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} = \mathbf{0};$$

$$(b)\ \mathbf{D}_q'\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} - \mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right) = \mathbf{0},$$

$$\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\left\{(\boldsymbol{\Delta}\otimes\mathbf{I}_q)\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} - [\mathbf{I}_q\otimes\mathrm{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\right\} = \mathbf{0}; \quad \textit{and}$$

$$(c)\ \mathbf{D}_q'\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} + \mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\mathbf{J}_{21,\nu_3} = \mathbf{0},$$

$$\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma}) \times \tag{188}$$

$$\left\{(\boldsymbol{\Delta}\otimes\mathbf{I}_q)\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} + [\mathbf{I}_q\otimes\mathrm{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\mathbf{J}_{21,\nu_3}\right\} = \mathbf{0},$$

where $\mathbf{J}_{21,\nu_3}$ *is defined in Table 56.*

A short proof for (188) is given before the expressions for $\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}$, $\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}$ *and* $\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}$ *are derived. The following proofs uses* $\mathbf{G}(\boldsymbol{\theta}_{\boldsymbol{\gamma}}=\mathbf{0}) = \mathbf{I}_q$.

188(a):

$$\begin{aligned}
\frac{\partial \mathbf{D}_q'\,\mathrm{vec}\,(\mathbf{G}\mathbf{G}'-\mathbf{I}_q)}{\partial\,\boldsymbol{\theta}_{\boldsymbol{\gamma}}'} &= \mathbf{D}_q'(\mathbf{G}\otimes\mathbf{I}_q)\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'} + \mathbf{D}_q'(\mathbf{I}_p\otimes\mathbf{G})\mathbf{K}_{q,q}\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'} \\
&= \mathbf{D}_q'(\mathbf{G}\otimes\mathbf{I}_q)\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'} + \mathbf{D}_q'\mathbf{K}_{q,q}(\mathbf{G}\otimes\mathbf{I}_q)\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'} \\
&= 2\mathbf{D}_q'(\mathbf{G}\otimes\mathbf{I}_q)\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'} \ \ \textit{because } \mathbf{D}_q'\mathbf{K}_{q,q} = \mathbf{D}_q' \\
\left.\frac{\partial \mathbf{D}_q'\,\mathrm{vec}\,(\mathbf{G}\mathbf{G}'-\mathbf{I}_q)}{\partial\,\boldsymbol{\theta}_{\boldsymbol{\gamma}}'}\right|_{\boldsymbol{\theta}_{\boldsymbol{\gamma}}=\mathbf{0}} &= 2\mathbf{D}_q'\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'} = \mathbf{0} \\
\left.\frac{\partial \mathbf{C}_3'\,\mathrm{vec}(\boldsymbol{\Phi})}{\partial\,\boldsymbol{\theta}_{\boldsymbol{\gamma}}'}\right|_{\boldsymbol{\Gamma}_0=\boldsymbol{\Gamma}} &= \mathbf{C}_3'\left(\boldsymbol{\Gamma}\boldsymbol{\Delta}\otimes\boldsymbol{\Gamma}\right)\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'} + \mathbf{C}_3'\mathbf{K}_{q,q}\left(\boldsymbol{\Gamma}\boldsymbol{\Delta}\otimes\boldsymbol{\Gamma}\right)\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'} \\
&= 2\mathbf{C}_3'\mathbf{N}_q\left(\boldsymbol{\Gamma}\boldsymbol{\Delta}\otimes\boldsymbol{\Gamma}\right)\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'} \\
&= 2\mathbf{C}_3'\left(\boldsymbol{\Gamma}\boldsymbol{\Delta}\otimes\boldsymbol{\Gamma}\right)\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'} \\
&\qquad \textit{because } \mathbf{C}_3' = \mathbf{K}'\mathbf{L}_{21,q}' \textit{ and } \mathbf{L}_{21,q}'\mathbf{N}_q = \mathbf{L}_{21,q}' \\
&= \mathbf{0}
\end{aligned}$$

188(b):

$$\begin{aligned}
&\frac{\partial^2 \mathbf{D}_q'\,\mathrm{vec}\,(\mathbf{G}\mathbf{G}'-\mathbf{I}_q)}{\left(\partial\,\boldsymbol{\theta}_{\boldsymbol{\gamma}}'\right)^{\otimes 2}} \\
&= \frac{\partial\left[2\mathbf{D}_q'(\mathbf{G}\otimes\mathbf{I}_q)\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right]}{\partial\,\boldsymbol{\theta}_{\boldsymbol{\gamma}}'} \quad \textit{based on (188a) result}
\end{aligned}$$

$$
\begin{aligned}
&= 2\mathbf{D}_q'\left[\frac{\partial(\mathbf{G}\otimes\mathbf{I}_q)}{\partial\boldsymbol{\theta}_{\boldsymbol{\gamma}}'}\left(\mathbf{I}_{\nu_3}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)+(\mathbf{G}\otimes\mathbf{I}_q)\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\right]\\
&= 2\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{I}_{q^2}\right)\left(\mathbf{I}_{\nu_3}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\\
&+ 2\mathbf{D}_q'(\mathbf{G}\otimes\mathbf{I}_q)\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\\
&= 2\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)+2\mathbf{D}_q'(\mathbf{G}\otimes\mathbf{I}_q)\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}
\end{aligned}
$$

$$
\begin{aligned}
&\left.\frac{\partial^2\mathbf{D}_q'\operatorname{vec}(\mathbf{G}\mathbf{G}'-\mathbf{I}_q)}{\left(\partial\,\boldsymbol{\theta}_{\boldsymbol{\gamma}}'\right)^{\otimes 2}}\right|_{\boldsymbol{\theta}_{\boldsymbol{\gamma}}=\mathbf{0}}\\
&= 2\mathbf{D}_q'\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}+2\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\\
&= 2\mathbf{D}_q'\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}-2\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\\
&\qquad \textit{because } \mathbf{K}_{q,q}\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}=-\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\\
&= \mathbf{0}
\end{aligned}
$$

$$
\begin{aligned}
&\left.\frac{\partial^2\mathbf{C}_3'\operatorname{vec}(\boldsymbol{\Phi})}{\left(\partial\,\boldsymbol{\theta}_{\boldsymbol{\gamma}}'\right)^{\otimes 2}}\right|_{\boldsymbol{\Gamma}_0=\boldsymbol{\Gamma}}\\
&= 2\mathbf{C}_3'\left(\boldsymbol{\Gamma}\boldsymbol{\Delta}\otimes\boldsymbol{\Gamma}\right)\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\\
&+ 2\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\boldsymbol{\Gamma}]\left(\mathbf{K}_{q,q}\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\\
&= 2\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})(\boldsymbol{\Delta}\otimes\mathbf{I}_q)\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\\
&+ 2\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{K}_{q,q}\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\\
&= 2\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\times\\
&\quad\left\{(\boldsymbol{\Delta}\otimes\mathbf{I}_q)\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}-[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\right\}\\
&= \mathbf{0}
\end{aligned}
$$

188(c):

$$
\begin{aligned}
\frac{\partial^3\mathbf{D}_q'\operatorname{vec}(\mathbf{G}\mathbf{G}'-\mathbf{I}_q)}{\left(\partial\,\boldsymbol{\theta}_{\boldsymbol{\gamma}}'\right)^{\otimes 3}} &= \frac{\partial\left[2\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)+2\mathbf{D}_q'(\mathbf{G}\otimes\mathbf{I}_q)\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\right]}{\partial\,\boldsymbol{\theta}_{\boldsymbol{\gamma}}'}\\
&\qquad \textit{based on (188b) result}\\
&= 2\mathbf{D}_q'\mathbf{I}_{q,3}'\frac{\partial\left(\mathbf{K}_{q,q}\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)}{\partial\,\boldsymbol{\theta}_{\boldsymbol{\gamma}}'}
\end{aligned}
$$

$$
\begin{aligned}
&+ \quad 2\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'} \otimes \mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\right)\\
&+ \quad 2\mathbf{D}_q'(\mathbf{G}\otimes\mathbf{I}_q)\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} \quad \textit{based on (188b) result}\\
&= \quad 2\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right) + 2\mathbf{D}_q'\mathbf{I}_{q,3}'\mathbf{K}_{q^2,q^2}\\
&\times \quad \left(\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} \otimes \mathbf{K}_{q,q}\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)(\mathbf{I}_{\nu_3}\otimes\mathbf{K}_{\nu_3,\nu_3})\\
&+ \quad 2\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'} \otimes \mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\right) + 2\mathbf{D}_q'(\mathbf{G}\otimes\mathbf{I}_q)\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}
\end{aligned}
$$

To simply the expression further, examine $\mathbf{N}_q\mathbf{I}_{q,3}'\mathbf{K}_{q^2,q^2}$*:*

$$
\begin{aligned}
\mathbf{N}_q\mathbf{I}_{q,3}'\mathbf{K}_{q^2,q^2} &= \mathbf{N}_q\left(\mathbf{I}_q\otimes(\operatorname{vec}\mathbf{I}_q)'\otimes\mathbf{I}_q\right)\mathbf{K}_{q^2,q^2}\\
&= \mathbf{N}_q\mathbf{K}_{q,q}\left[\mathbf{I}_q\otimes\sum_{i=1}^{q}\left(\mathbf{e}_i^{q\prime}\otimes\mathbf{e}_i^{q\prime}\right)\otimes\mathbf{I}_q\right]\mathbf{K}_{q^2,q^2}\\
&= \mathbf{N}_q\sum_{i=1}^{q}\left(\mathbf{e}_i^{q\prime}\otimes\mathbf{I}_q\otimes\mathbf{I}_q\otimes\mathbf{e}_i^{q\prime}\right)\\
&= \mathbf{N}_q\sum_{i=1}^{q}\left[(\mathbf{K}_{1,q}\otimes\mathbf{K}_{q,1})\left(\mathbf{I}_q\otimes\mathbf{e}_i^{q\prime}\otimes\mathbf{e}_i^{q\prime}\otimes\mathbf{I}_q\right)(\mathbf{K}_{q,q}\otimes\mathbf{K}_{q,q})\right]\\
&= \mathbf{N}_q\left[\mathbf{I}_q\otimes\sum_{i=1}^{q}\left(\mathbf{e}_i^{q\prime}\otimes\mathbf{e}_i^{q\prime}\right)\otimes\mathbf{I}_q\right](\mathbf{K}_{q,q}\otimes\mathbf{K}_{q,q})\\
&= \mathbf{N}_q\mathbf{I}_{q,3}'(\mathbf{K}_{q,q}\otimes\mathbf{K}_{q,q})\,.
\end{aligned}
$$

Further, $\mathbf{D}_q'\mathbf{I}_{q,3}'\mathbf{K}_{q^2,q^2} = \mathbf{D}_q'\mathbf{I}_{q,3}'(\mathbf{K}_{q,q}\otimes\mathbf{K}_{q,q})$.

Now continue to simply $\partial^3\mathbf{D}_q'\operatorname{vec}(\mathbf{G}\mathbf{G}'-\mathbf{I}_p)/\left(\partial\,\boldsymbol{\theta}_{\boldsymbol{\gamma}}'\right)^{\otimes 3}$*:*

$$
\begin{aligned}
&\frac{\partial^3\mathbf{D}_q'\operatorname{vec}(\mathbf{G}\mathbf{G}'-\mathbf{I}_p)}{\left(\partial\,\boldsymbol{\theta}_{\boldsymbol{\gamma}}'\right)^{\otimes 3}}\\
&= \quad 2\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\\
&+ \quad 2\mathbf{D}_q'\mathbf{I}_{q,3}'\mathbf{K}_{q^2,q^2}\left(\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} \otimes \mathbf{K}_{q,q}\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)(\mathbf{I}_{\nu_3}\otimes\mathbf{K}_{\nu_3,\nu_3})\\
&+ \quad 2\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'} \otimes \mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\right) + 2\mathbf{D}_q'(\mathbf{G}\otimes\mathbf{I}_q)\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\\
&= \quad 2\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right) + 2\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\\
&\times \quad (\mathbf{I}_{\nu_3}\otimes\mathbf{K}_{\nu_3,\nu_3}) + 2\mathbf{D}_q'\mathbf{I}_{q,3}'\mathbf{K}_{q^2,q^2}\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'} \otimes \mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\right)
\end{aligned}
$$

$$
\begin{aligned}
&+\quad 2\mathbf{D}_q'(\mathbf{G}\otimes\mathbf{I}_q)\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\\
&=\quad 2\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\\
&+\quad 2\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)(\mathbf{I}_{\nu_3}\otimes\mathbf{K}_{\nu_3,\nu_3})\\
&+\quad 2\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\mathbf{K}_{\nu_3^2,\nu_3}\\
&+\quad 2\mathbf{D}_q'(\mathbf{G}\otimes\mathbf{I}_q)\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\\
&=\quad 2\mathbf{D}_q'(\mathbf{G}\otimes\mathbf{I}_q)\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}+2\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\mathbf{J}_{21,\nu_3},
\end{aligned}
$$

where $\mathbf{J}_{21,\nu_3}$ *is defined in Table 56. It follows that*

$$
\begin{aligned}
\frac{\partial^3\mathbf{D}_q'\operatorname{vec}(\mathbf{G}\mathbf{G}'-\mathbf{I}_p)}{\left(\partial\,\boldsymbol{\theta}_\gamma'\right)^{\otimes 3}}\Big|_{\boldsymbol{\theta}_\gamma=\mathbf{0}} &= 2\mathbf{D}_q'\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}+2\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\mathbf{J}_{21,\nu_3}\\
&= \mathbf{0}\\
\frac{\partial^3\mathbf{C}_3'\operatorname{vec}(\boldsymbol{\Phi})}{\left(\partial\,\boldsymbol{\theta}_\gamma'\right)^{\otimes 3}}\Big|_{\boldsymbol{\Gamma}_0=\boldsymbol{\Gamma}} &= 2\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})(\boldsymbol{\Delta}\otimes\mathbf{I}_q)\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}+\\
&\quad 2\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\\
&\quad \left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\mathbf{J}_{21,\nu_3}\\
&= \mathbf{0}
\end{aligned}
$$

Now, the proof for (188) is finished. Those results are used for the rest of the proof. It follows from (42), (187a) and (188a) that

$$
\begin{gathered}
2\mathbf{D}_q'\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}=\mathbf{0}\Longrightarrow\mathbf{D}^{(1)}_{\boldsymbol{\eta}_{\gamma,1};\boldsymbol{\theta}_\gamma'}=-\mathbf{D}_q'\mathbf{A}_2\mathbf{V}_3\mathbf{D}^{(1)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma'}-\mathbf{D}_q'\mathbf{A}_2\mathbf{V}_4\Longrightarrow\\
\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}=-\mathbf{A}_1\mathbf{D}_q'\mathbf{A}_2\mathbf{V}_3\mathbf{D}^{(1)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma'}-\mathbf{A}_1\mathbf{D}_q'\mathbf{A}_2\mathbf{V}_4 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{V}_3\mathbf{D}^{(1)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma'}\\
+\mathbf{A}_2\mathbf{V}_3\mathbf{D}^{(1)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma'}+\mathbf{A}_2\mathbf{V}_4\\
\Longrightarrow\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}=2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{V}_3\mathbf{D}^{(1)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma'}+2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{V}_4\\
\mathbf{C}_3'\left(\boldsymbol{\Gamma}\boldsymbol{\Delta}\otimes\boldsymbol{\Gamma}\right)\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}=\mathbf{0}\Longrightarrow\mathbf{C}_3'\left(\boldsymbol{\Gamma}\boldsymbol{\Delta}\otimes\boldsymbol{\Gamma}\right)\left[2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{V}_3\mathbf{D}^{(1)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma'}+2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{V}_4\right]=\mathbf{0}
\end{gathered}
$$

Recall that

$$
\mathbf{W}_{\boldsymbol{\gamma}}=\mathbf{C}_3'\left(\boldsymbol{\Gamma}\boldsymbol{\Delta}\otimes\boldsymbol{\Gamma}\right)2\mathbf{N}_q^{\perp}\mathbf{A}_2\ \textit{and}\ SVD(\mathbf{W}_{\boldsymbol{\gamma}})=\mathbf{U}_{\boldsymbol{\gamma}}\mathbf{D}_{\boldsymbol{\gamma}}\mathbf{V}_{\boldsymbol{\gamma}}'=\mathbf{U}_{\boldsymbol{\gamma}}\mathbf{D}_{\boldsymbol{\gamma},1}\mathbf{V}_{\boldsymbol{\gamma},1}'.
$$

Let $\mathbf{V}_4 = \mathbf{V}_{\boldsymbol{\gamma},2}$ *so that* $\mathbf{W}_{\boldsymbol{\gamma}}\mathbf{V}_4 = \mathbf{0}$*. It follows that*

$$\mathbf{C}_3'\left(\boldsymbol{\Gamma\Delta}\otimes\boldsymbol{\Gamma}\right)\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'} = \mathbf{W}_{\boldsymbol{\gamma}}\mathbf{V}_3\mathbf{D}^{(1)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma'} + \mathbf{W}_{\boldsymbol{\gamma}}\mathbf{V}_4 = \mathbf{0} \Longrightarrow \mathbf{D}^{(1)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma'} = \mathbf{0},$$

because $\mathbf{W}_{\boldsymbol{\gamma}}\mathbf{V}_3$ *has full rank.*

The first derivative of $\operatorname{vec}\mathbf{G} = \mathbf{g}$ *in* (40) *with respect to* $\boldsymbol{\theta}_{\boldsymbol{\gamma}}$*, evaluated at* $\boldsymbol{\theta}_{\boldsymbol{\gamma}} = \mathbf{0}$ *and* $\boldsymbol{\Gamma}_0 = \boldsymbol{\Gamma}$*, can be written as follows:*

$$\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'} = 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{V}_4,$$

because $\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'} = 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{V}_3\mathbf{D}^{(1)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma'} + 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{V}_4$*.*

It follows from (42), (187b) and (188b) that

$$\begin{aligned}
&\mathbf{D}^{(2)}_{\boldsymbol{\eta}_{\gamma,1};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} = \mathbf{D}_q'\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} - \mathbf{D}_q'\mathbf{A}_2\mathbf{V}_3\mathbf{D}^{(2)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\\
&\Longrightarrow \mathbf{D}^{(2)}_{\boldsymbol{\eta}_{\gamma,1};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} = \mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right) - \mathbf{D}_q'\mathbf{A}_2\mathbf{V}_3\mathbf{D}^{(2)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} \quad \textit{using (188b)}\\
&\Longrightarrow \mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} = 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{V}_3\mathbf{D}^{(2)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} + \mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right) \quad \textit{using (187b)}\\
&\Longrightarrow \mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})(\boldsymbol{\Delta}\otimes\mathbf{I}_q)\left[2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{V}_3\mathbf{D}^{(2)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} + \mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\right]\\
&\qquad -\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right) = \mathbf{0}\\
&\Longrightarrow \mathbf{W}_{\boldsymbol{\gamma}}\mathbf{V}_3\mathbf{D}^{(2)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} + \mathbf{C}_3'\left(\boldsymbol{\Gamma\Delta}\otimes\boldsymbol{\Gamma}\right)\mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\\
&\qquad -\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right) = \mathbf{0}\\
&\Longrightarrow \mathbf{D}_{\boldsymbol{\gamma},1}^{-1}\mathbf{U}_{\boldsymbol{\gamma}}'\left[\mathbf{W}_{\boldsymbol{\gamma}}\mathbf{V}_3\mathbf{D}^{(2)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} + \mathbf{C}_3'\left(\boldsymbol{\Gamma\Delta}\otimes\boldsymbol{\Gamma}\right)\mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\right]\\
&\qquad -\mathbf{D}_{\boldsymbol{\gamma},1}^{-1}\mathbf{U}_{\boldsymbol{\gamma}}'\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right) = \mathbf{0}\\
&\Longrightarrow \mathbf{D}^{(2)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} + \mathbf{D}_{\boldsymbol{\gamma},1}^{-1}\mathbf{U}_{\boldsymbol{\gamma}}'\mathbf{C}_3'\left(\boldsymbol{\Gamma\Delta}\otimes\boldsymbol{\Gamma}\right)\mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\\
&\qquad -\mathbf{D}_{\boldsymbol{\gamma},1}^{-1}\mathbf{U}_{\boldsymbol{\gamma}}'\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right) = \mathbf{0}\\
&\qquad \textit{because SVD}(\mathbf{W}_{\boldsymbol{\gamma}}) = \mathbf{U}_{\boldsymbol{\gamma}}\mathbf{D}_{\boldsymbol{\gamma},1}\mathbf{V}_{\boldsymbol{\gamma},1}' \textit{ and } \mathbf{V}_{\boldsymbol{\gamma},1} = \mathbf{V}_3\\
&\Longrightarrow \mathbf{D}^{(2)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} = \mathbf{V}_3'\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)
\end{aligned}$$

$$-\mathbf{V}_3'\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'\left(\boldsymbol{\Gamma\Delta}\otimes\boldsymbol{\Gamma}\right)\mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)$$

because $SVD(\mathbf{W}_{\boldsymbol{\gamma}}^{+})=\mathbf{V}_{\boldsymbol{\gamma},1}\mathbf{D}_{\boldsymbol{\gamma},1}{}^{-1}\mathbf{U}_{\boldsymbol{\gamma}}'$ *and* $\mathbf{V}_3'\mathbf{W}_{\boldsymbol{\gamma}}^{+}=\mathbf{D}_{\boldsymbol{\gamma},1}^{-1}\mathbf{U}_{\boldsymbol{\gamma}}'$.

The second derivative of $\operatorname{vec}\mathbf{G}=\mathbf{g}$ *in* (40) *with respect to* $\boldsymbol{\theta}_{\boldsymbol{\gamma}}$*, evaluated at* $\boldsymbol{\theta}_{\boldsymbol{\gamma}}=\mathbf{0}$ *and* $\boldsymbol{\Gamma}_0=\boldsymbol{\Gamma}$*, can be written as follows:*

$$\begin{aligned}
\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} &= \left(\mathbf{I}_{q^2}-\mathbf{P}_{\boldsymbol{\gamma}}\right)\mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\\
&+ 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\\
&= \left[\left(\mathbf{I}_{q^2}-\mathbf{P}_{\boldsymbol{\gamma}}\right)\mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}'+2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\right]\\
&\times \left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right),
\end{aligned}$$

because $\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}=2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{V}_3\mathbf{D}^{(2)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}+\mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)$, $\mathbf{V}_3\mathbf{V}_3'\mathbf{W}_{\boldsymbol{\gamma}}^{+}=\mathbf{W}_{\boldsymbol{\gamma}}^{+}$, *and* $\mathbf{P}_{\boldsymbol{\gamma}}\stackrel{\text{def}}{=}2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'\left(\boldsymbol{\Gamma\Delta}\otimes\boldsymbol{\Gamma}\right)$ *is a projection operator.*

Also, it follows from (42), (187c) and (188c) that

$$\mathbf{D}^{(3)}_{\boldsymbol{\eta}_{\gamma,1};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}=\mathbf{D}_q'\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}-\mathbf{D}_q'\mathbf{A}_2\mathbf{V}_3\mathbf{D}^{(3)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\qquad\textit{using (187c)}$$

$$\Longrightarrow\mathbf{D}^{(3)}_{\boldsymbol{\eta}_{\gamma,1};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}=-\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\mathbf{J}_{21,\nu_3}-\mathbf{D}_q'\mathbf{A}_2\mathbf{V}_3\mathbf{D}^{(3)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}$$

$$\textit{using (188c)}$$

$$\Longrightarrow\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}=2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{V}_3\mathbf{D}^{(3)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}-\mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\mathbf{J}_{21,\nu_3}$$

$$\Longrightarrow\mathbf{W}_{\boldsymbol{\gamma}}\mathbf{V}_3\mathbf{D}^{(3)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}-\mathbf{C}_3'\left(\boldsymbol{\Gamma\Delta}\otimes\boldsymbol{\Gamma}\right)\mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\mathbf{J}_{21,\nu_3}$$

$$+\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\mathbf{J}_{21,\nu_3}=\mathbf{0}\qquad\textit{using (188c)}$$

$$\Longrightarrow$$

$$\mathbf{D}_{\boldsymbol{\gamma},1}^{-1}\mathbf{U}_{\boldsymbol{\gamma}}'\left[\mathbf{W}_{\boldsymbol{\gamma}}\mathbf{V}_3\mathbf{D}^{(3)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}-\mathbf{C}_3'\left(\boldsymbol{\Gamma\Delta}\otimes\boldsymbol{\Gamma}\right)\mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\mathbf{J}_{21,\nu_3}\right]$$

$$+\mathbf{D}_{\boldsymbol{\gamma},1}^{-1}\mathbf{U}_{\boldsymbol{\gamma}}'\left\{\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\mathbf{J}_{21,\nu_3}\right\}=\mathbf{0}$$

$$\Longrightarrow\mathbf{D}^{(3)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}-\mathbf{D}_{\boldsymbol{\gamma},1}^{-1}\mathbf{U}_{\boldsymbol{\gamma}}'\mathbf{C}_3'\left(\boldsymbol{\Gamma\Delta}\otimes\boldsymbol{\Gamma}\right)\mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\mathbf{J}_{21,\nu_3}$$

$$+\mathbf{D}_{\boldsymbol{\gamma},1}^{-1}\mathbf{U}_{\boldsymbol{\gamma}}'\left\{\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\mathbf{J}_{21,\nu_3}\right\}=\mathbf{0}$$

$$\text{because } SVD(\mathbf{W}_{\boldsymbol{\gamma}}) = \mathbf{U}_{\boldsymbol{\gamma}}\mathbf{D}_{\boldsymbol{\gamma},1}\mathbf{V}'_{\boldsymbol{\gamma},1} \text{ and } \mathbf{V}_{\boldsymbol{\gamma},1} = \mathbf{V}_3$$

$$\Longrightarrow \mathbf{D}^{(3)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} = \mathbf{V}'_3\mathbf{W}^+_{\boldsymbol{\gamma}}\mathbf{C}'_3\left(\boldsymbol{\Gamma}\boldsymbol{\Delta}\otimes\boldsymbol{\Gamma}\right)\mathbf{A}_1\mathbf{D}'_q\mathbf{I}'_{q,3}\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\right)\mathbf{J}_{21,\nu_3}$$
$$-\mathbf{V}'_3\mathbf{W}^+_{\boldsymbol{\gamma}}\left\{\mathbf{C}'_3(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q]\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\right)\mathbf{J}_{21,\nu_3}\right\}$$
$$\text{because } SVD(\mathbf{W}^+_{\boldsymbol{\gamma}}) = \mathbf{V}_{\boldsymbol{\gamma},1}\mathbf{D}_{\boldsymbol{\gamma},1}{}^{-1}\mathbf{U}'_{\boldsymbol{\gamma}} \text{ and } \mathbf{V}'_3\mathbf{W}^+_{\boldsymbol{\gamma}} = \mathbf{D}^{-1}_{\boldsymbol{\gamma},1}\mathbf{U}'_{\boldsymbol{\gamma}},$$

where $\mathbf{J}_{21,\nu_3}$ *is defined in Table 56.*

The third derivative of $\operatorname{vec}\mathbf{G} = \mathbf{g}$ *in* (40) *with respect to* $\boldsymbol{\theta}_{\boldsymbol{\gamma}}$, *evaluated at* $\boldsymbol{\theta}_{\boldsymbol{\gamma}} = \mathbf{0}$ *and* $\boldsymbol{\Gamma}_0 = \boldsymbol{\Gamma}$, *can be written as follows:*

$$\begin{aligned}
\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} &= -\left(\mathbf{I}_{q^2}-\mathbf{P}_{\boldsymbol{\gamma}}\right)\mathbf{A}_1\mathbf{D}'_q\mathbf{I}'_{q,3}\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\right)\mathbf{J}_{21,\nu_3} \\
&- 2\mathbf{N}^\perp_q\mathbf{A}_2\mathbf{W}^+_{\boldsymbol{\gamma}}\mathbf{C}'_3(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q] \\
&\times \left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\right)\mathbf{J}_{21,\nu_3} \\
&= -\left\{2\mathbf{N}^\perp_q\mathbf{A}_2\mathbf{W}^+_{\boldsymbol{\gamma}}\mathbf{C}'_3(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})[\mathbf{I}_q\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\mathbf{I}_q] + \left(\mathbf{I}_{q^2}-\mathbf{P}_{\boldsymbol{\gamma}}\right)\mathbf{A}_1\mathbf{D}'_q\mathbf{I}'_{q,3}\right\} \\
&\times \left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\right)\mathbf{J}_{21,\nu_3},
\end{aligned}$$

because $\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} = 2\mathbf{N}^\perp_q\mathbf{A}_2\mathbf{V}_3\mathbf{D}^{(3)}_{\boldsymbol{\eta}_{\gamma,2};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} - \mathbf{A}_1\mathbf{D}'_q\mathbf{I}'_{q,3}\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\right)\mathbf{J}_{21,\nu_3}$. *The derivative expressions in Theorem 9 were checked numerically.* □

B.10. Proof of Theorem 10

Theorem 10. [Adapted from Boik [33], Supplement, 2010, Theorem 29, Page 84]. *Assume that* $\mathbf{W}_{\boldsymbol{\gamma}}$ *has full row-rank and* $\mathbf{V}_3 = \mathbf{V}_{\boldsymbol{\gamma},1}$. *The derivatives of* $\operatorname{vec}\mathbf{G}$ *in* (40) *with respect to* $\boldsymbol{\theta}_{\boldsymbol{\delta}}$ *and* $\boldsymbol{\theta}_{\boldsymbol{\gamma}}$, *evaluated at* $\boldsymbol{\theta}_{\boldsymbol{\gamma}} = \mathbf{0}$ *and* $\boldsymbol{\Gamma}_0 = \boldsymbol{\Gamma}$, *can be written as follows:*

$$\begin{aligned}
\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta} &= -\mathbf{N}^\perp_q\mathbf{A}_2\mathbf{W}^+_{\boldsymbol{\gamma}}\mathbf{C}'_3(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}, \\
\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} &= \left(\mathbf{I}_{q^2}-\mathbf{P}_{\boldsymbol{\gamma}}\right)\mathbf{A}_1\mathbf{D}'_q\mathbf{I}'_{q,3}\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\right) \\
&+ 2\mathbf{N}^\perp_q\mathbf{A}_2\mathbf{W}^+_{\boldsymbol{\gamma}}\mathbf{C}'_3[\boldsymbol{\Gamma}\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\boldsymbol{\Gamma}]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\right)
\end{aligned}$$

$$
\begin{aligned}
& - \quad \mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\mathbf{L}_{21,q}\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} \\
& - \quad 4\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes\operatorname{vec}'(\mathbf{I}_q)\otimes\boldsymbol{\Gamma}](\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\delta'})\mathbf{N}_{\nu_2}, \\
\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\gamma'} \quad & = \quad (\mathbf{I}_{q^2}-\mathbf{P}_{\boldsymbol{\gamma}})\,\mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}'\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right) \\
& + \quad 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\boldsymbol{\Gamma}]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right) \\
& - \quad 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes\operatorname{vec}'(\mathbf{I}_q)\otimes\boldsymbol{\Gamma}](\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}), \\
\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'} \quad & = \quad -\left\{(\mathbf{I}_{q^2}-\mathbf{P}_{\boldsymbol{\gamma}})\,\mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}'+2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\boldsymbol{\Gamma}]\right\} \\
& \times \quad \left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\delta'}\right)\mathbf{J}_{21,\nu_2} \\
& - \quad 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes\operatorname{vec}'(\mathbf{I}_q)\otimes\boldsymbol{\Gamma}] \\
& \times \quad \left[\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'}\otimes\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta'}\right)+\left(\mathbf{L}_{21,q}\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\delta'}\right)\right]\mathbf{J}_{21,\nu_2} \\
& + \quad 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes\operatorname{vec}'(\mathbf{I}_q)\otimes\operatorname{vec}'(\mathbf{I}_q)\otimes\boldsymbol{\Gamma}] \\
& \times \quad \left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\delta'}\otimes\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\delta'}\right) \\
& \times \quad [(\mathbf{K}_{\nu_2,\nu_2}\otimes\mathbf{I}_{\nu_2})+(\mathbf{I}_{\nu_2}\otimes 2\mathbf{N}_{\nu_2})] \\
& - \quad \mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\mathbf{L}_{21,q}\mathbf{D}^{(3)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'}, \\
\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\gamma'} \quad & = \quad -\left\{(\mathbf{I}_{q^2}-\mathbf{P}_{\boldsymbol{\gamma}})\,\mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}'+2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\boldsymbol{\Gamma}]\right\} \\
& \times \quad \left[\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\delta'}\right)\mathbf{J}_{21,\nu_2\nu_3}+\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\right] \\
& - \quad 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes\operatorname{vec}'(\mathbf{I}_q)\otimes\boldsymbol{\Gamma}] \\
& \times \quad \left[\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta'}\right)\mathbf{J}_{21,\nu_2\nu_3}+\left(\mathbf{L}_{21,q}\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\right] \\
& + \quad 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes\operatorname{vec}'(\mathbf{I}_q)\otimes\operatorname{vec}'(\mathbf{I}_q)\otimes\boldsymbol{\Gamma}] \\
& \times \quad \left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\delta'}\otimes\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\mathbf{J}^{*}_{21,\nu_2\nu_3} \\
& \quad\;\; \mathit{and} \\
\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'} \quad & = \quad -\left\{(\mathbf{I}_{q^2}-\mathbf{P}_{\boldsymbol{\gamma}})\,\mathbf{A}_1\mathbf{D}_q'\mathbf{I}_{q,3}'+2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\boldsymbol{\Gamma}]\right\} \\
& \times \quad \left\{\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\delta',\boldsymbol{\theta}_\gamma'}\otimes\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}\right)\left[\mathbf{I}_{\nu_3^2\nu_2}+\mathbf{I}_{\nu_2}\otimes\mathbf{K}_{\nu_3,\nu_3}\right]-\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'}\right)\right\} \\
& - \quad 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes\operatorname{vec}'(\mathbf{I}_q)\otimes\boldsymbol{\Gamma}](\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta'}\otimes\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}_\gamma',\boldsymbol{\theta}_\gamma'})
\end{aligned}
$$

$$
\begin{aligned}
&+ \quad 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'[\boldsymbol{\Gamma}\otimes \operatorname{vec}'(\mathbf{I}_q)\otimes \operatorname{vec}'(\mathbf{I}_q)\otimes \boldsymbol{\Gamma}] \\
&\times \quad \left(\mathbf{D}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}^{(1)}\otimes \mathbf{L}_{21,q}\mathbf{D}_{\boldsymbol{\delta};\boldsymbol{\theta}_\delta'}^{(1)}\otimes \mathbf{D}_{\mathbf{g};\boldsymbol{\theta}_\gamma'}^{(1)}\right)(\mathbf{K}_{\nu_3,\nu_2}\otimes \mathbf{I}_{\nu_3}),
\end{aligned}
$$

where $\mathbf{P}_{\boldsymbol{\gamma}} = 2\mathbf{N}_q^{\perp}\mathbf{A}_2\mathbf{W}_{\boldsymbol{\gamma}}^{+}\mathbf{C}_3'(\boldsymbol{\Gamma\Delta}\otimes\boldsymbol{\Gamma})$, $\mathbf{I}_{q,3}$, $\mathbf{J}_{21,\nu_3}$, $\mathbf{N}_q^{\perp}$ *are defined in Table 56,* $\mathbf{W}_{\boldsymbol{\gamma}}^{+}$ *is the Moore-Penrose inverse of* $\mathbf{W}_{\boldsymbol{\gamma}}$, $\mathbf{J}_{21,\nu_2\nu_3} = \mathbf{K}_{\nu_2\nu_3,\nu_2} + \mathbf{I}_{\nu_2}\otimes\mathbf{K}_{\nu_3,\nu_2}$ *and* $\mathbf{J}^{*}_{21,\nu_2\nu_3} = \mathbf{I}_{\nu_2^2\nu_3} + \mathbf{K}_{\nu_2,\nu_2}\otimes\mathbf{I}_{\nu_3}$.

Proof: Theorem 10 is based on the results from Theorem 29 in the supplement of Boik [33]. Certain substitutions are made for the use of this thesis. The proof is similar to that of Theorem 9 and is omitted. The derivative expressions in Theorem 10 were checked numerically. □

B.11. Proof of Theorem 11

Theorem 11. [Original result]. *Define* $\mathbf{D}_{\boldsymbol{\xi}_\psi}$ *and* $\mathbf{W}_\psi$ *as*

$$\mathbf{D}_{\boldsymbol{\xi}_\psi} \stackrel{\text{def}}{=} \operatorname{Diag}(\boldsymbol{\xi}_\psi) \text{ and } \mathbf{W}_\psi \stackrel{\text{def}}{=} \mathbf{C}_1'\mathbf{T}_4\mathbf{D}_{\boldsymbol{\xi}_\psi}.$$

Assume that no entry in $\boldsymbol{\xi}_\psi$ *is* $\mathbf{0}$ *and* $\mathbf{C}_1$ *has been chosen such that the* $r_p \times p_2$ *matrix* $\mathbf{W}_\psi$ *has full row-rank. Write the singular value decomposition of* $\mathbf{W}_\psi$ *as follows:*

$$
\begin{aligned}
&\mathbf{W}_\psi = \mathbf{U}_\psi\mathbf{D}_\psi\mathbf{V}_\psi', \text{ where} \\
&\mathbf{U}_\psi \in \mathcal{O}(r_p), \quad \mathbf{V}_\psi = \begin{pmatrix}\mathbf{V}_{\psi,1} & \mathbf{V}_{\psi,2}\end{pmatrix} \in \mathcal{O}(p_2), \\
&\mathbf{V}_{\psi,1} \in \mathcal{O}(p_2, r_p), \quad \mathbf{D}_\psi = \begin{pmatrix}\mathbf{D}_{\psi,1} & \mathbf{0}_{r_p\times(p_2-r_p)}\end{pmatrix}, \text{ and } \mathbf{D}_{\psi,1} \in \mathcal{D}^{+}(r_p).
\end{aligned}
$$

Then,

a. *the parameter* $\boldsymbol{\xi}_\psi$ *can be written as* $\boldsymbol{\xi}_\psi = \mathbf{V}_1\boldsymbol{\eta}_\psi + \mathbf{V}_2\boldsymbol{\theta}_\psi$, *where* $\boldsymbol{\eta}_\psi = \mathbf{V}_1'\boldsymbol{\xi}_\psi$, $\boldsymbol{\theta}_\psi = \mathbf{V}_2'\boldsymbol{\xi}_\psi$, $\mathbf{V}_1 = \mathbf{V}_{\psi,1}$, *and* $\mathbf{V}_2 = \mathbf{V}_{\psi,2}$;

b. *the parameters* $\boldsymbol{\eta}_\psi$ *and* $\boldsymbol{\xi}_\psi$ *are implicit functions of* $\boldsymbol{\theta}_\psi$. *Therefore,* $\partial\boldsymbol{\eta}_\psi/\partial\boldsymbol{\theta}_\psi'$ *and* $\partial\boldsymbol{\xi}_\psi/\partial\boldsymbol{\theta}_\psi'$ *exist;*

c. the first three derivatives of $\boldsymbol{\psi}$ *with respect to* $\boldsymbol{\theta}_\psi$ *can be written as follows:*

$$\mathbf{D}^{(1)}_{\psi;\boldsymbol{\theta}'_\psi} = 2\mathbf{T}_4\mathbf{D}_{\boldsymbol{\xi}_\psi}\mathbf{D}^{(1)}_{\boldsymbol{\xi}_\psi;\boldsymbol{\theta}'_\psi} = 2\mathbf{T}_4\mathbf{D}_{\boldsymbol{\xi}_\psi}\mathbf{V}_2,$$

$$\mathbf{D}^{(2)}_{\psi;\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi} = 2\mathbf{T}_4\left(\mathbf{I}_{p_2} - \mathbf{P}_{\boldsymbol{\xi}_\psi}\right)\left(\mathbf{V}'_2 * \mathbf{V}'_2\right)', \text{ and}$$

$$\mathbf{D}^{(3)}_{\psi;\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi} = 2\mathbf{T}_4\left(\mathbf{I}_{p_2} - \mathbf{P}_{\boldsymbol{\xi}_\psi}\right)\left(\mathbf{D}^{(2)\prime}_{\boldsymbol{\xi}_\psi;\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi}_\psi;\boldsymbol{\theta}'_\psi}\right)'\mathbf{J}_{\nu_4}, \text{ where}$$

$$\mathbf{D}^{(1)}_{\boldsymbol{\xi}_\psi;\boldsymbol{\theta}'_\psi} = \mathbf{V}_2, \quad \mathbf{D}^{(2)}_{\boldsymbol{\xi}_\psi;\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi} = -\mathbf{D}^{-1}_{\boldsymbol{\xi}_\psi}\mathbf{P}_{\boldsymbol{\xi}_\psi}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\xi}_\psi;\boldsymbol{\theta}'_\psi} * \mathbf{D}^{(1)\prime}_{\boldsymbol{\xi}_\psi;\boldsymbol{\theta}'_\psi}\right)', \quad \mathbf{P}_{\boldsymbol{\xi}_\psi} = \mathbf{D}_{\boldsymbol{\xi}_\psi}\mathbf{W}^+_\psi\mathbf{C}'_1\mathbf{T}_4,$$

$*$ *is the Khatri-Rao column-wise product,* $\mathbf{J}_{\nu_4}$ *and* $\mathbf{N}_{\nu_4}$ *are defined in Table 56 and* $\mathcal{D}^+(r_p)$ *is defined in Table 57; and*

d. $\mathbf{P}_{\boldsymbol{\xi}_\psi}$ *is a projection operator.*

Proof: The proof is similar to that of Theorem 2 and is ommited. The derivative expressions in Theorem 11 were checked numerically. □

B.12. Proof of Theorem 12

Theorem 12. [Original result]. *Suppose that the* $p \times p_2$ *design matrix* $\mathbf{T}_4$ *has full column-rank. Then* $\mathbf{D}^{(1)}_{\psi;\boldsymbol{\theta}'_\psi}$ *has full column-rank if and only if* $\boldsymbol{\xi}_{\psi,i} \neq 0$ *for* $i = 1, 2, \ldots, p_2$. *Proof: See proof of Theorem 3.* □

B.13. Proof of Theorem 13

Theorem 13. [Original result]. *If* $\mathbf{D}^{(1)}_{\psi;\boldsymbol{\theta}'_\psi}$ *does not have full column-rank, then* $\boldsymbol{\theta}_\psi$ *is not identified.*

Proof: See proof of Theorem 4. □

B.14. Proof of Theorem 14

Theorem 14. [Original result]. *If* $\mathbf{T_4}$ *has full column-rank and* $\boldsymbol{\xi}_{\psi,i} \neq 0$ *for* $i = 1, 2, \ldots, p_2$*, then*

a. The dimension of $\boldsymbol{\theta_\psi}$ *is* $\nu_4 = p_2 - r_p$

b. A special case of part a: if $\mathbf{T}_4 = \bigoplus_{i=1}^{k} \mathbf{1}_{m_i}$ *and* $\sum_{i=1}^{k} m_i = p$*, then* $\nu_4 = k - r_p$*.*

Proof: See proof of Theorem 5. □

B.15. Proof of Theorem 15

Theorem 15. [Original result]. *First, second and third derivatives of* $\boldsymbol{\sigma}$ *with respect to* $\boldsymbol{\theta_\lambda}$*,* $\boldsymbol{\theta_\delta}$*,* $\boldsymbol{\theta_\gamma}$ *and* $\boldsymbol{\theta_\psi}$*, evaluated at* $\mathbf{G} = \mathbf{I}_q$*, are listed as follows:*

$$
\begin{aligned}
\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\lambda} &= 2\mathbf{N}_p(\boldsymbol{\Lambda\Phi} \otimes \mathbf{I}_p)\mathbf{W}_2, \\
\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\delta} &= 2\mathbf{N}_p(\boldsymbol{\Lambda\Gamma\Delta} \otimes \boldsymbol{\Lambda\Gamma})\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta} + (\boldsymbol{\Lambda\Gamma} \otimes \boldsymbol{\Lambda\Gamma})\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}, \\
\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\gamma} &= 2\mathbf{N}_p(\boldsymbol{\Lambda\Gamma\Delta} \otimes \boldsymbol{\Lambda\Gamma})\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}, \\
\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi} &= \mathbf{L}_{21,p}\mathbf{D}^{(1)}_{\boldsymbol{\psi};\boldsymbol{\theta}'_\psi}, \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\lambda,\boldsymbol{\theta}'_\lambda} &= 2\mathbf{N}_p[\mathbf{I}_p \otimes \operatorname{vec}'(\boldsymbol{\Phi}) \otimes \mathbf{I}_p]\left[\mathbf{K}_{p,q}\mathbf{W}_2 \otimes \mathbf{W}_2\right], \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\lambda} &= 4\mathbf{N}_p[\boldsymbol{\Lambda} \otimes \operatorname{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_p]\mathbf{N}_q(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}) \\
&\times \left\{\left[(\boldsymbol{\Delta} \otimes \mathbf{I}_q)\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta} + 1/2\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\right] \otimes \mathbf{W}_2\right\}, \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\lambda} &= 4\mathbf{N}_p[\boldsymbol{\Lambda} \otimes \operatorname{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_p]\left[\mathbf{N}_q(\boldsymbol{\Gamma\Delta} \otimes \boldsymbol{\Gamma})\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} \otimes \mathbf{W}_2\right], \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\lambda} &= \mathbf{0}_{p^2 \times \nu_1\nu_4}, \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} &= 2\mathbf{N}_p[\boldsymbol{\Lambda\Gamma} \otimes \operatorname{vec}'(\mathbf{I}_q) \otimes \boldsymbol{\Lambda\Gamma}](\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta})2\mathbf{N}_{\nu_2} \\
&+ 2\mathbf{N}_p(\boldsymbol{\Lambda\Gamma\Delta} \otimes \boldsymbol{\Lambda\Gamma})\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} \\
&- 2\mathbf{N}_p[\boldsymbol{\Lambda\Gamma} \otimes \operatorname{vec}'(\boldsymbol{\Delta}) \otimes \boldsymbol{\Lambda\Gamma}]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\right)
\end{aligned}
$$

$$
\begin{aligned}
& + \quad (\boldsymbol{\Lambda\Gamma} \otimes \boldsymbol{\Lambda\Gamma})\mathbf{L}_{21,q}\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta}, \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\delta} &= 2\mathbf{N}_p(\boldsymbol{\Lambda\Gamma\Delta} \otimes \boldsymbol{\Lambda\Gamma})\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\delta} \\
&- 2\mathbf{N}_p[\boldsymbol{\Lambda\Gamma} \otimes \mathrm{vec}'(\boldsymbol{\Delta}) \otimes \boldsymbol{\Lambda\Gamma}]\Big(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\Big) \\
&- 2\mathbf{N}_p[\boldsymbol{\Lambda\Gamma} \otimes \mathrm{vec}'(\mathbf{I}_q) \otimes \boldsymbol{\Lambda\Gamma}]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} \otimes \mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\right), \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\delta} &= \mathbf{0}_{p^2\times\nu_2\nu_4}, \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} &= 2\mathbf{N}_p(\boldsymbol{\Lambda\Gamma\Delta} \otimes \boldsymbol{\Lambda\Gamma})\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} \\
&- 2\mathbf{N}_p[\boldsymbol{\Lambda\Gamma} \otimes \mathrm{vec}'(\boldsymbol{\Delta}) \otimes \boldsymbol{\Lambda\Gamma}]\Big(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\Big), \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\gamma} &= \mathbf{0}_{p^2\times\nu_3\nu_4}, \\
\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi} &= \mathbf{L}_{21,p}\mathbf{D}^{(2)}_{\boldsymbol{\psi};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi}, \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\lambda,\boldsymbol{\theta}'_\lambda,\boldsymbol{\theta}'_\lambda} &= \mathbf{0}_{p^2\times\nu_1^3}, \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\lambda,\boldsymbol{\theta}'_\lambda} &= 2\mathbf{N}_p[\mathbf{I}_p \otimes \mathrm{vec}'(\mathbf{I}_q) \otimes \mathrm{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_p](\mathbf{K}_{pq,q^2} \otimes \mathbf{I}_{pq}) \\
&\times \left\{\Big[2\mathbf{N}_q(\boldsymbol{\Gamma\Delta} \otimes \boldsymbol{\Gamma})\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta} + (\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\Big] \otimes \mathbf{I}_{p^2q^2}\right\} \\
&\times (\mathbf{I}_{\nu_2} \otimes \mathbf{K}_{p,q}\mathbf{W}_2 \otimes \mathbf{W}_2), \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\lambda,\boldsymbol{\theta}'_\lambda} &= 4\mathbf{N}_p[\mathbf{I}_p \otimes \mathrm{vec}'(\mathbf{I}_q) \otimes \mathrm{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_p](\mathbf{K}_{pq,q^2} \otimes \mathbf{I}_{pq}) \\
&\times \Big[\mathbf{N}_q(\boldsymbol{\Gamma\Delta} \otimes \boldsymbol{\Gamma})\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} \otimes \mathbf{I}_{p^2q^2}\Big](\mathbf{I}_{\nu_3} \otimes \mathbf{K}_{p,q}\mathbf{W}_2 \otimes \mathbf{W}_2), \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\lambda,\boldsymbol{\theta}'_\lambda} &= \mathbf{0}_{p^2\times\nu_1^2\nu_4}, \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\lambda} &= 4\mathbf{N}_p[\boldsymbol{\Lambda} \otimes \mathrm{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_p]\,[\mathbf{N}_q(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})\mathbf{A} \otimes \mathbf{W}_2], \ \textit{where} \\
\mathbf{A} &= -[\mathbf{I}_q \otimes \mathrm{vec}'(\boldsymbol{\Delta}) \otimes \mathbf{I}_q]\Big(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\Big) \\
&+ [\mathbf{I}_q \otimes \mathrm{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_q]\left(\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta}\right) \\
&+ (\boldsymbol{\Delta} \otimes \mathbf{I}_q)\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} \\
&- \mathbf{N}_q[\mathbf{I}_q \otimes \mathrm{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_q]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta} \otimes \mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\right) + 1/2\mathbf{L}_{21,q}\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta}, \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\lambda} &= 4\,(\mathbf{K}_{\nu_3,\nu_2} \otimes \mathbf{I}_{\nu_1p^2})\,\mathbf{N}_p[\boldsymbol{\Lambda} \otimes \mathrm{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_p]\,[\mathbf{N}_q(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})\mathbf{B} \otimes \mathbf{W}_2], \ \textit{where} \\
\mathbf{B} &= (\boldsymbol{\Delta} \otimes \mathbf{I}_q)\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\gamma} - [\mathbf{I}_q \otimes \mathrm{vec}'(\boldsymbol{\Delta}) \otimes \mathbf{I}_q]\Big(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\Big)
\end{aligned}
$$

$$
\begin{aligned}
&+ && [\mathbf{I}_q \otimes \mathrm{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_q] \left(\mathbf{L}_{21,q} \mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} \right), \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\lambda} &= && \mathbf{0}_{\nu_1\nu_2\nu_4}, \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\lambda} &= && 4\mathbf{N}_p[\boldsymbol{\Lambda} \otimes \mathrm{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_p] \\
&\times && \mathbf{N}_q \left[(\boldsymbol{\Gamma}\boldsymbol{\Delta} \otimes \boldsymbol{\Gamma}) \mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} - [\boldsymbol{\Gamma} \otimes \mathrm{vec}'(\boldsymbol{\Delta}) \otimes \boldsymbol{\Gamma}] \left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} \right) \right] \otimes \mathbf{W}_2, \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\lambda} &= && \mathbf{0}_{\nu_1\nu_3\nu_4}, \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\lambda} &= && \mathbf{0}_{\nu_1\nu_4^2}, \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} &= && 2\mathbf{N}_p(\boldsymbol{\Lambda}\boldsymbol{\Gamma} \otimes \boldsymbol{\Lambda}\boldsymbol{\Gamma})\mathbf{C}, \ \textit{where} \\
\mathbf{C} &= && -[\mathbf{I}_q \otimes \mathrm{vec}'(\mathbf{I}_q) \otimes \mathrm{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_q] \left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta} \otimes \mathbf{L}_{21,q} \mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta} \right) \\
&\times && [(\mathbf{K}_{\nu_2,\nu_2} \otimes \mathbf{I}_{\nu_2}) + (\mathbf{I}_{\nu_2} \otimes 2\mathbf{N}_{\nu_2})] \\
&+ && [\mathbf{I}_q \otimes \mathrm{vec}'(\boldsymbol{\Delta}) \otimes \mathbf{I}_q] \left(\mathbf{K}_{q,q} \mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta} \right) \mathbf{J}_{21,\nu_2} \\
&+ && [\mathbf{I}_q \otimes \mathrm{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_q] \\
&\times && \left[\left(\mathbf{K}_{q,q} \mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} \otimes \mathbf{L}_{21,q} \mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta} \right) + \left(\mathbf{L}_{21,q} \mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta} \right) \right] \mathbf{J}_{21,\nu_2} \\
&+ && (\boldsymbol{\Delta} \otimes \mathbf{I}_q) \mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} + 1/2 \mathbf{L}_{21,q} \mathbf{D}^{(3)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta}, \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\gamma} &= && 2\mathbf{N}_p(\boldsymbol{\Lambda}\boldsymbol{\Gamma} \otimes \boldsymbol{\Lambda}\boldsymbol{\Gamma})\mathbf{E}, \ \textit{where} \\
\mathbf{E} &= && [\mathbf{I}_q \otimes \mathrm{vec}'(\boldsymbol{\Delta}) \otimes \mathbf{I}_q] \left(\mathbf{K}_{q,q} \mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\gamma} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta} \right) (\mathbf{K}_{\nu_2\nu_3,\nu_2} + \mathbf{I}_{\nu_2} \otimes \mathbf{K}_{\nu_3,\nu_2}) \\
&+ && [\mathbf{I}_q \otimes \mathrm{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_q] \left(\mathbf{K}_{q,q} \mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\gamma} \otimes \mathbf{L}_{21,q} \mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta} \right) \\
&\times && (\mathbf{K}_{\nu_2\nu_3,\nu_2} + \mathbf{I}_{\nu_2} \otimes \mathbf{K}_{\nu_3,\nu_2}) \\
&+ && (\boldsymbol{\Delta} \otimes \mathbf{I}_q) \mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\gamma} - [\mathbf{I}_q \otimes \mathrm{vec}'(\mathbf{I}_q) \otimes \mathrm{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_q] \\
&\times && \left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta} \otimes \mathbf{L}_{21,q} \mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} \right) \left[\mathbf{I}_{\nu_2^2\nu_3} + (\mathbf{K}_{\nu_2,\nu_2} \otimes \mathbf{I}_{\nu_3}) \right] \\
&+ && [\mathbf{I}_q \otimes \mathrm{vec}'(\boldsymbol{\Delta}) \otimes \mathbf{I}_q] \left(\mathbf{K}_{q,q} \mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} \right) \\
&+ && [\mathbf{I}_q \otimes \mathrm{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_q](\mathbf{L}_{21,q} \mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}), \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\delta} &= && \mathbf{0}_{\nu_2^2\nu_4}, \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} &= && 2\mathbf{N}_p(\boldsymbol{\Lambda}\boldsymbol{\Gamma} \otimes \boldsymbol{\Lambda}\boldsymbol{\Gamma})\mathbf{F}, \ \textit{where}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{F} &= [\mathbf{I}_q \otimes \text{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_q](\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta} \otimes \mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma}) \\
&- [\mathbf{I}_q \otimes \text{vec}'(\boldsymbol{\Delta}) \otimes \mathbf{I}_q](\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\delta} \otimes \mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma}) + (\boldsymbol{\Delta} \otimes \mathbf{I}_q)\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} \\
&- [\mathbf{I}_q \otimes \text{vec}'(\mathbf{I}_q) \otimes \text{vec}'(\mathbf{I}_q) \otimes \mathbf{I}_q]\left(\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} \otimes \mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\right) \\
&\times (\mathbf{K}_{\nu_3,\nu_2} \otimes \mathbf{I}_{\nu_3}) \\
&+ [\mathbf{I}_q \otimes \text{vec}'(\boldsymbol{\Delta}) \otimes \mathbf{I}_q](\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\delta,\boldsymbol{\theta}'_\gamma} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}) \\
&\times \left[\mathbf{I}_{\nu_2\nu_3^2} + (\mathbf{I}_{\nu_2} \otimes \mathbf{K}_{\nu_3,\nu_3})\right], \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\delta} &= \mathbf{0}_{\nu_2\nu_3\nu_4}, \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\delta} &= \mathbf{0}_{\nu_2\nu_4^2}, \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} &= 2\mathbf{N}_p(\boldsymbol{\Lambda}\boldsymbol{\Gamma} \otimes \boldsymbol{\Lambda}\boldsymbol{\Gamma})\mathbf{D}, \ \textit{where} \\
\mathbf{D} &= [\mathbf{I}_q \otimes \text{vec}'(\boldsymbol{\Delta}) \otimes \mathbf{I}_q]\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} \otimes \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\right)\mathbf{J}_{21,\nu_3} + (\boldsymbol{\Delta} \otimes \mathbf{I}_q)\mathbf{D}^{(3)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma}, \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\gamma,\boldsymbol{\theta}'_\gamma} &= \mathbf{0}_{\nu_3^2\nu_4}, \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\gamma} &= \mathbf{0}_{\nu_3\nu_4^2}, \\
\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi} &= \mathbf{L}_{21,p}\mathbf{D}^{(3)}_{\boldsymbol{\psi};\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi,\boldsymbol{\theta}'_\psi},
\end{aligned}
$$

where the derivatives of $\boldsymbol{\delta}$*,* $\mathbf{g}$ *and* $\boldsymbol{\psi}$ *with respective to* $\boldsymbol{\theta}_{\boldsymbol{\delta}}$*,* $\boldsymbol{\theta}_{\boldsymbol{\gamma}}$*,* $\boldsymbol{\theta}_{\boldsymbol{\psi}}$ *respectively were given in Theorem 2, Theorem 10, Theorem 9 and Theorem 11 and* $\mathbf{0}_{a\times b}$ *represents a matrix of zeros with dimension* $a \times b$*.*

Proof: The derivative expressions in Theorem 15 were checked numerically. □

B.16. Proof of Theorem 16

Theorem 16. [Boik [35], 2010, Theorem 4, Page 13]. *The first three derivatives of* $\mathbf{g}_*$ *in* (55) *with respect to* $\boldsymbol{\theta}_{\boldsymbol{\gamma}*}$*, evaluated at* $\boldsymbol{\theta}_{\boldsymbol{\gamma}*} = \mathbf{0}$*, can be written as follows:*

$$
\begin{aligned}
\mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}} &= 2\mathbf{N}_q^\perp\mathbf{A}_2, \\
\mathbf{D}^{(2)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\gamma*}} &= \mathbf{A}_1\mathbf{D}'_q\mathbf{I}'_{q,3}\left(\mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}} \otimes \mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}}\right), \ \textit{and}
\end{aligned}
$$

$$\mathbf{D}^{(3)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\gamma*}} = -\mathbf{A}_1\mathbf{D}'_q\mathbf{I}'_{q,3}\left(\mathbf{K}_{q,q}\mathbf{D}^{(2)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\gamma*}}\otimes\mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}}\right)\mathbf{J}_{21,\nu_3^*}$$

where $\mathbf{I}_{q,3}$, $\mathbf{D}_q$ *and* $\mathbf{J}_{21,\nu_3}$ *are defined in Table 56.*

Proof: The proof is omitted because it is similar to the proof of Theorem 9. □

B.17. Proof of Theorem 17

Theorem 17. [Adapted from Boik [34], 2003, Page 689, & Supplement, Page 26].

(a.) *The first derivative of* $F(\boldsymbol{\Sigma},\mathbf{S})$ *in (66) with respect to* $\boldsymbol{\theta}$*, evaluated at* $\mathbf{G}=\mathbf{I}_q$*, is as follows:*

$$\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}} \stackrel{\text{def}}{=} \left.\frac{\partial F(\boldsymbol{\Sigma},\mathbf{S})}{\partial\,\boldsymbol{\theta}}\right|_{\mathbf{G}=\mathbf{I}_q} = -\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma}),$$

where $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}$ *is given in (48).*

(b.) *The second derivative of* $F(\boldsymbol{\Sigma},\mathbf{S})$ *in (66) with respect to* $\boldsymbol{\theta}$ *and* $\boldsymbol{\theta}'$*, evaluated at* $\mathbf{G}=\mathbf{I}_q$*, is as follows:*

$$\begin{aligned}\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'} &\stackrel{\text{def}}{=} \left.\frac{\partial^2 F(\boldsymbol{\Sigma},\mathbf{S})}{\partial\,\boldsymbol{\theta}\otimes\partial\,\boldsymbol{\theta}'}\right|_{\mathbf{G}=\mathbf{I}_q}\\ &= \mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} + 2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right]\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\\ &- \mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}\left[\mathbf{I}_\nu\otimes\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right],\end{aligned}$$

where $\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'} = \operatorname{dvec}\left(\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'}, p^2\nu, \nu\right)$ *and* $\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'}$ *is given in (49).*

(c.) *The third derivative of* $F(\boldsymbol{\Sigma},\mathbf{S})$ *in (66) with respect to* $\boldsymbol{\theta}'$, $\boldsymbol{\theta}'$ *and* $\boldsymbol{\theta}$*, evaluated at* $\mathbf{G}=\mathbf{I}_q$*, is as follows:*

$$\begin{aligned}\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}} &\stackrel{\text{def}}{=} \left.\frac{\partial^3 F(\boldsymbol{\Sigma},\mathbf{S})}{\partial\,\boldsymbol{\theta}'\otimes\partial\,\boldsymbol{\theta}'\otimes\partial\,\boldsymbol{\theta}}\right|_{\mathbf{G}=\mathbf{I}_q}\\ &= -4\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]'\left(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right)\mathbf{N}_\nu\\ &+ 2\mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}\left\{\mathbf{I}_\nu\otimes\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]\right\}\mathbf{N}_\nu\end{aligned}$$

$$
\begin{aligned}
&+ \quad \left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]' \mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'} \\
&- \quad 2\left\{\operatorname{vec}\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right] \otimes \left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]\right\}' \\
&\times \quad \left(\mathbf{I}_{p\nu} \otimes \mathbf{K}_{p,p} \otimes \mathbf{I}_p\right)\left[\mathbf{I}_\nu \otimes \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
&- \quad 4\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]' \\
&\times \quad \left\{\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathbf{N}_p\left[\mathbf{I}_p \otimes (\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right] \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right\} \\
&- \quad \mathbf{D}^{(3)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta},\boldsymbol{\theta}'}\left[\mathbf{I}_{\nu^2} \otimes \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
&+ \quad 4\mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}\left(\mathbf{I}_\nu \otimes \boldsymbol{\Sigma}^{-1} \otimes \operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)' \\
&\times \quad \left[\left(\mathbf{I}_\nu \otimes \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right)\mathbf{N}_\nu \otimes \operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
&+ \quad 2\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]'\left[\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'} \otimes \operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right],
\end{aligned}
$$

where $\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'} = \operatorname{dvec}\left(\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'}, p^2\nu, \nu\right)$, $\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta},\boldsymbol{\theta}'} = \operatorname{dvec}\left(\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}'}, p^2\nu^2, \nu\right)$, $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}$, $\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'}$, *and* $\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}'}$ *are given in (50).*

Also, define $\mathbf{D}^{(3)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}',\boldsymbol{\theta}'}$ *as*

$$
\mathbf{D}^{(3)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}',\boldsymbol{\theta}'} \stackrel{\text{def}}{=} \left.\frac{\partial^3 F(\boldsymbol{\Sigma},\mathbf{S})}{\partial\,\boldsymbol{\theta} \otimes \partial\,\boldsymbol{\theta}' \otimes \partial\,\boldsymbol{\theta}'}\right|_{\mathbf{G}=\mathbf{I}_q},
$$

and $\mathbf{D}^{(3)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}',\boldsymbol{\theta}'} = \mathbf{D}^{(3)}_{F;\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}}$.

Proof: [Independent Proof]. *The expressions of* $\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}$, $\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}$ *and* $\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}}$ *in Theorem 17 can be easily derived from the results provided by Boik [34]. For completeness, the relevant proofs for those results are given below.*

(a.) *By Lemma 1 part (a), it can be concluded that* $\operatorname{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) = \operatorname{vec}'\mathbf{S}\operatorname{vec}\boldsymbol{\Sigma}^{-1}$. *By Lemma 1 part (a) and (b), it can be concluded that*

$$
\frac{\partial \ln|\boldsymbol{\Sigma}|}{\partial\,\theta_i} = \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\,\theta_i}\right) = \operatorname{vec}'\boldsymbol{\Sigma}^{-1}\operatorname{vec}\left(\frac{\partial\boldsymbol{\Sigma}}{\partial\,\theta_i}\right).
$$

Accordingly,

$$
\frac{\partial \ln|\boldsymbol{\Sigma}|}{\partial\,\boldsymbol{\theta}'} = \operatorname{vec}'\boldsymbol{\Sigma}^{-1}\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial\,\boldsymbol{\theta}'}.
$$

Examine

$$
\begin{aligned}
\frac{\partial F(\boldsymbol{\Sigma},\mathbf{S})}{\partial\,\boldsymbol{\theta}'} &= \frac{\partial}{\partial\,\boldsymbol{\theta}'}\left[\mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) + \ln|\boldsymbol{\Sigma}|\right] \\
&= \frac{\partial}{\partial\,\boldsymbol{\theta}'}\,\mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) + \frac{\partial}{\partial\,\boldsymbol{\theta}'}\ln|\boldsymbol{\Sigma}| \\
&= \frac{\partial}{\partial\,\boldsymbol{\theta}'}\left(\mathrm{vec}'\mathbf{S}\,\mathrm{vec}\,\boldsymbol{\Sigma}^{-1}\right) + \frac{\partial}{\partial\,\boldsymbol{\theta}'}\ln|\boldsymbol{\Sigma}| \\
&= \frac{\partial}{\partial\,\boldsymbol{\theta}'}\left(\mathrm{vec}'\mathbf{S}\,\mathrm{vec}\,\boldsymbol{\Sigma}^{-1}\right) + \frac{\partial}{\partial\,\boldsymbol{\theta}'}\ln|\boldsymbol{\Sigma}| \\
&= \mathrm{vec}'\mathbf{S}\frac{\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}^{-1}}{\partial\,\boldsymbol{\theta}'} + \frac{\partial}{\partial\,\boldsymbol{\theta}'}\ln|\boldsymbol{\Sigma}| \\
&= -\mathrm{vec}'\mathbf{S}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\frac{\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}}{\partial\,\boldsymbol{\theta}'} + \mathrm{vec}'\boldsymbol{\Sigma}^{-1}\frac{\mathrm{vec}\boldsymbol{\Sigma}}{\partial\,\boldsymbol{\theta}'} \\
&\quad\ \textit{using Lemma 1 part (c).}
\end{aligned}
$$

The first derivative of $F(\boldsymbol{\Sigma},\mathbf{S})$ in (66) with respect to $\boldsymbol{\theta}$ is as follows:

$$
\begin{aligned}
\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}} &= \frac{\partial F(\boldsymbol{\Sigma},\mathbf{S})}{\partial\,\boldsymbol{\theta}} = \left(\frac{\partial F(\boldsymbol{\Sigma},\mathbf{S})}{\partial\,\boldsymbol{\theta}'}\right)' \\
&= -\frac{\partial \mathrm{vec}'\boldsymbol{\Sigma}}{\partial\,\boldsymbol{\theta}}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}\,\mathbf{S} + \frac{\partial \mathrm{vec}'\boldsymbol{\Sigma}}{\partial\,\boldsymbol{\theta}}\,\mathrm{vec}\,\boldsymbol{\Sigma}^{-1} \\
&= -\frac{\partial \mathrm{vec}'\boldsymbol{\Sigma}}{\partial\,\boldsymbol{\theta}}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}\,\mathbf{S} + \frac{\partial \mathrm{vec}'\boldsymbol{\Sigma}}{\partial\,\boldsymbol{\theta}}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}\,\boldsymbol{\Sigma} \\
&= -\frac{\partial \mathrm{vec}'\boldsymbol{\Sigma}}{\partial\,\boldsymbol{\theta}}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\left(\mathrm{vec}\,\mathbf{S} - \mathrm{vec}\,\boldsymbol{\Sigma}\right) \\
&= -\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma}),
\end{aligned}
$$

where $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}$ is given in (48).

(b.) Examine

$$
\begin{aligned}
\mathbf{D}^{(2)}_{F;\boldsymbol{\theta}',\boldsymbol{\theta}} &= \frac{\partial^2 F(\boldsymbol{\Sigma},\mathbf{S})}{\partial\,\boldsymbol{\theta}'\otimes\partial\,\boldsymbol{\theta}} = \frac{\partial}{\partial\,\boldsymbol{\theta}'}\otimes\left(\frac{\partial F(\boldsymbol{\Sigma},\mathbf{S})}{\partial\,\boldsymbol{\theta}}\right) \\
&= \frac{\partial}{\partial\,\boldsymbol{\theta}'}\otimes\left(-\left[\frac{\partial \mathrm{vec}'\boldsymbol{\Sigma}}{\partial\,\boldsymbol{\theta}}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right]\right) \\
&= -\frac{\partial^2 \mathrm{vec}'\boldsymbol{\Sigma}}{\partial\,\boldsymbol{\theta}'\otimes\partial\,\boldsymbol{\theta}}\left[\mathbf{I}_\nu\otimes\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
&\quad - \frac{\partial \mathrm{vec}'\boldsymbol{\Sigma}}{\partial\,\boldsymbol{\theta}}\left[\frac{\partial\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)}{\partial\,\boldsymbol{\theta}'}\right]\left[\mathbf{I}_\nu\otimes\mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right]
\end{aligned}
$$

$$
\begin{aligned}
&+ \quad \frac{\partial \mathrm{vec}'\boldsymbol{\Sigma}}{\partial\,\boldsymbol{\theta}} \left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right) \left(\frac{\partial\,\mathrm{vec}\,\boldsymbol{\Sigma}}{\partial \boldsymbol{\theta}'}\right) \\
&= \quad -\mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'} \left[\mathbf{I}_\nu \otimes \left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
&- \quad \mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \left[\frac{\partial\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)}{\partial\,\boldsymbol{\theta}'}\right] \left[\mathbf{I}_\nu \otimes \mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
&+ \quad \mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \\
&= \quad -\mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'} \left[\mathbf{I}_\nu \otimes \left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
&+ \quad \mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \left[\boldsymbol{\Sigma}^{-1}\otimes \mathrm{vec}'(\boldsymbol{\Sigma}^{-1})\otimes\boldsymbol{\Sigma}^{-1}\right] \left[\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
&+ \quad \mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\mathbf{K}_{p,p} \left[\boldsymbol{\Sigma}^{-1}\otimes \mathrm{vec}'(\boldsymbol{\Sigma}^{-1})\otimes\boldsymbol{\Sigma}^{-1}\right] \left[\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
&+ \quad \mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \\
&\quad\quad \textit{using Lemma 1 part (d)} \\
&= \quad -\mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'} \left[\mathbf{I}_\nu \otimes \left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
&+ \quad 2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \left[\boldsymbol{\Sigma}^{-1}\otimes \mathrm{vec}'(\boldsymbol{\Sigma}^{-1})\otimes\boldsymbol{\Sigma}^{-1}\right] \left[\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
&+ \quad \mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \\
&\quad\quad \textit{because } \mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\mathbf{K}_{p,p} = \mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \\
&= \quad \mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \\
&+ \quad 2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \left[\boldsymbol{\Sigma}^{-1}\otimes \boldsymbol{\Sigma}^{-1}(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right]\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \\
&- \quad \mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'} \left[\mathbf{I}_\nu \otimes \left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right].
\end{aligned}
$$

The second derivative of $F(\boldsymbol{\Sigma},\mathbf{S})$ *with respect to* $\boldsymbol{\theta}$ *and* $\boldsymbol{\theta}'$ *is as follows:*

$$
\begin{aligned}
& \quad \mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'} = \mathbf{K}_{\nu,1}\mathbf{D}^{(2)}_{F;\boldsymbol{\theta}',\boldsymbol{\theta}}\mathbf{K}_{\nu,1} = \mathbf{D}^{(2)}_{F;\boldsymbol{\theta}',\boldsymbol{\theta}} \\
&= \quad \mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} + 2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \left[\boldsymbol{\Sigma}^{-1}\otimes \boldsymbol{\Sigma}^{-1}(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right]\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \\
&- \quad \mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'} \left[\mathbf{I}_\nu \otimes \left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right]
\end{aligned}
$$

where $\nu = \nu_1+\nu_2+\nu_3+\nu_4$, $\mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'} = \mathrm{dvec}\left(\mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'}, p^2\nu, \nu\right)$ *and* $\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'}$ *is given in (49).*

(c.) Based on the result from part (b), it is necessary to go through the following steps before deriving the expression for $\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}}$.

Part I

$$\begin{aligned}
&\frac{\partial}{\partial\,\boldsymbol{\theta}'}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\\
=\;& \mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}\left\{\mathbf{I}_{\nu}\otimes\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]\right\}\\
-\;& 2\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]'\left(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right) \qquad (189)\\
+\;& \left[\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]'\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'}
\end{aligned}$$

using Lemma 1 part (d), Lemma 2 (2) and $\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\mathbf{K}_{p,p}=\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}$.

Part II

$$\begin{aligned}
&\frac{\partial}{\partial\,\boldsymbol{\theta}'}\left\{-\mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}\left[\mathbf{I}_{\nu}\otimes\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right]\right\}\\
=\;& -\mathbf{D}^{(3)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta},\boldsymbol{\theta}'}\left[\mathbf{I}_{\nu^2}\otimes\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right]\\
+\;& \mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}\mathbf{K}_{\nu,p^2}\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\mathbf{I}_{\nu}\right]\left(\mathbf{I}_{\nu}\otimes\mathbf{K}_{1,\nu}\right)\\
+\;& 2\mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}\left(\mathbf{I}_{\nu}\otimes\boldsymbol{\Sigma}^{-1}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)'\\
& \left[\left(\mathbf{I}_{\nu}\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right)\mathbf{K}_{\nu,\nu}\otimes\operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right]
\end{aligned}$$

using Lemma 2 (2) and (7)

$$\begin{aligned}
=\;& -\mathbf{D}^{(3)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta},\boldsymbol{\theta}'}\left[\mathbf{I}_{\nu^2}\otimes\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right]\\
+\;& \mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}\left\{\mathbf{I}_{\nu}\otimes\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]\right\}\mathbf{K}_{\nu,\nu} \qquad (190)\\
+\;& 2\mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}\left(\mathbf{I}_{\nu}\otimes\boldsymbol{\Sigma}^{-1}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)'\\
& \left[\left(\mathbf{I}_{\nu}\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right)\mathbf{K}_{\nu,\nu}\otimes\operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right].
\end{aligned}$$

Part III

$$2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\frac{\partial}{\partial\,\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\right]\left[\mathbf{I}_{\nu}\otimes\mathbf{I}_{p}\otimes(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right]\left(\mathbf{I}_{\nu}\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right)$$

using Lemma 2 (2)

$$
\begin{aligned}
= \quad & -2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\boldsymbol{\Sigma}^{-1}\otimes\mathrm{vec}'(\boldsymbol{\Sigma}^{-1})\otimes\boldsymbol{\Sigma}^{-1}\right]\left(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\mathbf{I}_{p^2}\right)\\
\times \quad & \left[\mathbf{I}_\nu\otimes\mathbf{I}_p\otimes(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right]\left(\mathbf{I}_\nu\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right)\\
- \quad & 2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\boldsymbol{\Sigma}^{-1}\otimes\mathrm{vec}'(\boldsymbol{\Sigma}^{-1})\otimes\boldsymbol{\Sigma}^{-1}\right]\left(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\mathbf{K}_{p,p}\right)\\
\times \quad & \left[\mathbf{I}_\nu\otimes\mathbf{I}_p\otimes(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right]\left(\mathbf{I}_\nu\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right)
\end{aligned}
$$

using Lemma 1 part (d) and $\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\mathbf{K}_{p,p}=\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}$

$$
\begin{aligned}
= \quad & -2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\boldsymbol{\Sigma}^{-1}\otimes\mathrm{vec}'(\boldsymbol{\Sigma}^{-1})\otimes\boldsymbol{\Sigma}^{-1}\right]\\
\times \quad & \left[\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\mathbf{I}_p\otimes(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right]\left(\mathbf{I}_\nu\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right)\\
- \quad & 2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\boldsymbol{\Sigma}^{-1}\otimes\mathrm{vec}'(\boldsymbol{\Sigma}^{-1})\otimes\boldsymbol{\Sigma}^{-1}\right]\\
\times \quad & \left\{\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\mathbf{K}_{p,p}\left[\mathbf{I}_p\otimes(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right]\right\}\left(\mathbf{I}_\nu\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right)\\
= \quad & -2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\boldsymbol{\Sigma}^{-1}\otimes\mathrm{vec}'(\boldsymbol{\Sigma}^{-1})\otimes\boldsymbol{\Sigma}^{-1}\right]\\
\times \quad & \left\{\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\left[\mathbf{I}_p\otimes(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right]\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right\}\\
- \quad & 2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\boldsymbol{\Sigma}^{-1}\otimes\mathrm{vec}'(\boldsymbol{\Sigma}^{-1})\otimes\boldsymbol{\Sigma}^{-1}\right]\\
\times \quad & \left\{\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\mathbf{K}_{p,p}\left[\mathbf{I}_p\otimes(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right]\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right\}\\
= \quad & -4\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\mathrm{vec}\,\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]'\\
\times \quad & \left\{\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\mathbf{N}_p\left[\mathbf{I}_p\otimes(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right]\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right\}.
\end{aligned}
\tag{191}
$$

$$
\begin{aligned}
& 2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\left[\frac{\partial}{\partial\,\boldsymbol{\theta}'}\left(\mathbf{I}_p\otimes\mathrm{vec}'\boldsymbol{\Sigma}^{-1}\otimes\mathbf{I}_p\right)\right]\\
\times \quad & \left[\mathbf{I}_\nu\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right]\\
= \quad & 2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\left(\mathbf{I}_1\otimes\mathbf{K}_{p,p}\right)\left[\frac{\partial\,\mathrm{vec}'\boldsymbol{\Sigma}^{-1}}{\partial\,\boldsymbol{\theta}'}\otimes\mathbf{I}_{p^2}\right]\\
\times \quad & \left(\mathbf{I}_\nu\otimes\mathbf{K}_{p^3,p}\right)\left[\mathbf{I}_\nu\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right]\\
= \quad & -2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\left\{\left[\left(\mathbf{I}_\nu\otimes\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}}\right]'\otimes\mathbf{I}_{p^2}\right\}\\
\times \quad & \left(\mathbf{I}_\nu\otimes\mathbf{K}_{p^3,p}\right)\left[\mathbf{I}_\nu\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right]
\end{aligned}
$$

$$
\begin{aligned}
& \quad \textit{using Lemma 1 part (e)} \\
= & \; -2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\left\{\operatorname{vec}'\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]\otimes\mathbf{I}_{p^2}\right\} \\
\times & \; \left(\mathbf{I}_\nu\otimes\mathbf{K}_{p^3,p}\right)\left[\mathbf{I}_\nu\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
& \quad \textit{using Lemma 2 (3)} \\
= & \; -2\left\{\left[\operatorname{vec}\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]\otimes\mathbf{I}_{p^2}\right]\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]\right\}' \\
\times & \; \left(\mathbf{I}_\nu\otimes\mathbf{K}_{p^3,p}\right)\left[\mathbf{I}_\nu\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
= & \; -2\left\{\operatorname{vec}\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]\otimes\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]\right\}' \\
\times & \; \left(\mathbf{I}_{p\nu}\otimes\mathbf{K}_{p,p}\otimes\mathbf{I}_p\right)\left[\mathbf{I}_\nu\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right]. \qquad (192)
\end{aligned}
$$

The above results in (191) and (192) are useful for the following derivation result.

$$
\begin{aligned}
& \frac{\partial}{\partial\,\boldsymbol{\theta}'}\left\{2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right]\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right\} \\
= & \; \frac{\partial}{\partial\,\boldsymbol{\theta}'}\left\{2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right]\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right\} \\
= & \; \frac{\partial}{\partial\,\boldsymbol{\theta}'}\left\{2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\left[(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right]\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right\} \\
= & \; 2\mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}\left(\mathbf{I}_\nu\otimes\boldsymbol{\Sigma}^{-1}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)'\left[\mathbf{I}_\nu\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
- & \; 4\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]' \\
\times & \; \left\{\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\mathbf{N}_p\left[\mathbf{I}_p\otimes(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right]\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right\} \\
- & \; 2\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]'\mathbf{K}_{p^2,p^2}\left(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right)\mathbf{K}_{\nu,\nu}\mathbf{K}_{\nu,\nu} \\
- & \; 2\left\{\operatorname{vec}\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]\otimes\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]\right\}' \\
\times & \; \left(\mathbf{I}_{p\nu}\otimes\mathbf{K}_{p,p}\otimes\mathbf{I}_p\right)\left[\mathbf{I}_\nu\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
+ & \; 2\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]'\left[\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'}\otimes\operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
& \quad \textit{using equation 191 and Lemma 2 (1), (2), (5), (8)} \\
= & \; 2\mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}\left(\mathbf{I}_\nu\otimes\boldsymbol{\Sigma}^{-1}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)'\left[\mathbf{I}_\nu\otimes\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\otimes\operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
- & \; 4\left[\left(\boldsymbol{\Sigma}^{-1}\otimes\operatorname{vec}\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]'
\end{aligned}
$$

$$
\begin{aligned}
&\times \quad \left\{\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathbf{N}_p \left[\mathbf{I}_p \otimes (\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right] \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right\} \\
&- \quad 2\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]' \left(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right) \mathbf{K}_{\nu,\nu} \\
&- \quad 2\left\{\operatorname{vec}\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right] \otimes \left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]\right\}' \\
&\times \quad \left(\mathbf{I}_{p\nu} \otimes \mathbf{K}_{p,p} \otimes \mathbf{I}_p\right) \left[\mathbf{I}_\nu \otimes \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
&+ \quad 2\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]' \left[\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'} \otimes \operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \qquad (193)
\end{aligned}
$$

using Lemma 2 (7).

Part IV The derivation results in (189), (190) and (193) given in the previous parts composes the expression of $\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}}$.

$$
\begin{aligned}
\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}} &= \left.\frac{\partial^3 F(\boldsymbol{\Sigma},\mathbf{S})}{\partial\boldsymbol{\theta}' \otimes \partial\boldsymbol{\theta}' \otimes \partial\boldsymbol{\theta}}\right|_{\mathbf{G}=\mathbf{I}_q} \\
&= \frac{\partial}{\partial\boldsymbol{\theta}'}\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right] \\
&+ \frac{\partial}{\partial\boldsymbol{\theta}'}\left\{-\mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'} \left[\mathbf{I}_\nu \otimes \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right]\right\} \\
&+ \frac{\partial}{\partial\boldsymbol{\theta}'}\left\{2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \left[\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right] \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right\} \\
&= -4\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]' \left(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right) \mathbf{N}_\nu \\
&+ 2\mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'} \left\{\mathbf{I}_\nu \otimes \left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]\right\} \mathbf{N}_\nu \\
&+ \left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]' \mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'} \\
&- 2\left\{\operatorname{vec}\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right] \otimes \left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]\right\}' \\
&\times \left(\mathbf{I}_{p\nu} \otimes \mathbf{K}_{p,p} \otimes \mathbf{I}_p\right) \left[\mathbf{I}_\nu \otimes \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
&- 4\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]' \\
&\times \left\{\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathbf{N}_p \left[\mathbf{I}_p \otimes (\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right] \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right\} \\
&- \mathbf{D}^{(3)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta},\boldsymbol{\theta}'} \left[\mathbf{I}_{\nu^2} \otimes \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right] \\
&+ 4\mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'} \left(\mathbf{I}_\nu \otimes \boldsymbol{\Sigma}^{-1} \otimes \operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)' \\
&\times \left[\left(\mathbf{I}_\nu \otimes \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right) \mathbf{N}_\nu \otimes \operatorname{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right]
\end{aligned}
$$

$$+ \quad 2\left[\left(\mathbf{\Sigma}^{-1}\otimes \operatorname{vec}\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]'\left[\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'}\otimes \operatorname{vec}(\mathbf{S}-\mathbf{\Sigma})\right]$$

using the expressions given in (189), (190) and (193),

where $\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'} = \operatorname{dvec}\left(\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'}, p^2\nu, \nu\right)$,
$\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta},\boldsymbol{\theta}'} = \operatorname{dvec}\left(\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}'}, p^2\nu^2, \nu\right)$, $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}$, $\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}'}$, *and* $\mathbf{D}^{(3)}_{\boldsymbol{\sigma};\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}'}$ *are given in (50).*

Lastly, notice that $\mathbf{D}^{(3)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}',\boldsymbol{\theta}'} = \mathbf{K}_{\nu,1}\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}}\mathbf{K}_{\nu^2,1} = \mathbf{D}^{(3)}_{F;\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}}$. □

B.18. Proof of Theorem 18

Theorem 18. [A Well-Known Result]. *Rewrite* $\mathbf{\Lambda}_*$ *as* $\mathbf{\Lambda}_* = (\mathbf{\Lambda}_1^{*\prime} \quad \mathbf{\Lambda}^{*\prime})'$, *where* $\mathbf{\Lambda}_1^*$ *is a* $q\times q$ *nonsingular matrix. Without loss of generality,* $\mathbf{\Lambda}_1^*$ *in* $\mathbf{\Lambda}_* = (\mathbf{\Lambda}_1^{*\prime} \quad \mathbf{\Lambda}^{*\prime})'$ *can be replaced by a* $q\times q$ *identity matrix* $\mathbf{I}_q$ *so that* $\mathbf{\Lambda}_* = (\mathbf{I}_q \quad \mathbf{\Lambda}_2^{*\prime})'$, *where* $\mathbf{\Lambda}_2^*$ *is* $p_1\times q$ *with* $p_1 = p-q$.

Proof: [Independent Proof]. *Notice that* $\mathbf{\Lambda}_1^{*-1}$ *exists because* $\mathbf{\Lambda}_1^*$ *is a nonsingular matrix. Examine*

$$\begin{aligned}
\mathbf{\Lambda}_*\mathbf{f} &= \begin{pmatrix}\mathbf{\Lambda}_1^*\\ \mathbf{\Lambda}^*\end{pmatrix}\mathbf{f} = \begin{pmatrix}\mathbf{\Lambda}_1^*\\ \mathbf{\Lambda}^*\end{pmatrix}\left(\mathbf{\Lambda}_1^{*-1}\mathbf{\Lambda}_1^*\right)\mathbf{f}\\
&= \begin{pmatrix}\mathbf{\Lambda}_1^*\mathbf{\Lambda}_1^{*-1}\\ \mathbf{\Lambda}^*\mathbf{\Lambda}_1^{*-1}\end{pmatrix}\mathbf{\Lambda}_1^*\mathbf{f} = \begin{pmatrix}\mathbf{\Lambda}_1^*\mathbf{\Lambda}_1^{*-1}\\ \mathbf{\Lambda}^*\mathbf{\Lambda}_1^{*-1}\end{pmatrix}\mathbf{f}^*\\
&= \begin{pmatrix}\mathbf{I}_q\\ \mathbf{\Lambda}_2^*\end{pmatrix}\mathbf{f}^*,
\end{aligned}$$

where $\mathbf{\Lambda}_2^* = \mathbf{\Lambda}^*\mathbf{\Lambda}_1^{*-1}$ *and* $\mathbf{f}^* = \mathbf{\Lambda}_1^*\mathbf{f}$.

It follows that $\mathbf{\Lambda}_* = (\mathbf{\Lambda}_1^{*\prime} \quad \mathbf{\Lambda}^{*\prime})'$ *can be written as* $\mathbf{\Lambda}_* = (\mathbf{I}_q \quad \mathbf{\Lambda}_2^{*\prime})'$ *provided that* $\mathbf{f}^* = \mathbf{\Lambda}_1^*\mathbf{f}$. □

B.19. Proof of Theorem 19

Theorem 19. [An Application of a Well-Known Result]. *Define* $\boldsymbol{\theta}_{\boldsymbol{\phi}*}$ *as* $\boldsymbol{\theta}_{\boldsymbol{\phi}*} = \operatorname{vech}({\boldsymbol{\Phi}_*}^{-1})$ *and* $\boldsymbol{\theta}_{\boldsymbol{\psi}_2}$ *as* $\boldsymbol{\theta}_{\boldsymbol{\psi}_2} \stackrel{\text{def}}{=} \operatorname{diag}(\boldsymbol{\Psi}_2)$*, where* $\operatorname{vech}({\boldsymbol{\Phi}_*}^{-1})$ *and* $\operatorname{diag}(\boldsymbol{\Psi}_2)$ *are as in Table 56. Assume that* $\mathbf{S}_{\boldsymbol{zx}}$ *and* $\mathbf{S}_{\boldsymbol{zz}}$ *are matrices of constants.*

(a.) The approximate generalized least squares (GLS) estimator of $\boldsymbol{\theta}_{\boldsymbol{\phi}*}$ *is*

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{\phi}*} = \left(\mathbf{X}_{1.2}'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{X}_{1.2}\right)^{-1}\mathbf{X}_{1.2}'\widehat{\boldsymbol{\Omega}}^{-1}\operatorname{vec}\mathbf{S}_{\boldsymbol{zz}},$$

where $\mathbf{X}_{1.2} = \left[\mathbf{I}_{p_1^2} - \mathbf{L}_{21,p_1}\left(\mathbf{L}_{21,p_1}'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{L}_{21,p_1}\right)^{-1}\mathbf{L}_{21,p_1}'\widehat{\boldsymbol{\Omega}}^{-1}\right](\mathbf{S}_{\boldsymbol{zx}}\otimes\mathbf{S}_{\boldsymbol{zx}})\,\mathbf{D}_q$*,* $\widehat{\boldsymbol{\Omega}} = \mathbf{S}_{\boldsymbol{zz}}\otimes\mathbf{S}_{\boldsymbol{zz}}$*,* $\mathbf{D}_q$ *and* $\mathbf{L}_{21,p_1}$ *are defined in Table 56.*

(b.) The approximate ordinary least squares (OLS) estimator of $\boldsymbol{\theta}_{\boldsymbol{\phi}*}$ *is*

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{\phi}*} = \left(\mathbf{X}_{1.2}'\mathbf{X}_{1.2}\right)^{-1}\mathbf{X}_{1.2}'\operatorname{vec}\mathbf{S}_{\boldsymbol{zz}},$$

where $\mathbf{X}_{1.2} = \left(\mathbf{I}_{p_1^2} - \mathbf{L}_{22,p_1}\right)(\mathbf{S}_{\boldsymbol{zx}}\otimes\mathbf{S}_{\boldsymbol{zx}})\,\mathbf{D}_q.$

Proof: [Proof With Advisor's Help].

(a.) It can be concluded that $\operatorname{vec}\boldsymbol{\Phi}_*^{-1} = \mathbf{D}_q\boldsymbol{\theta}_{\boldsymbol{\phi}*}$ *and* $\operatorname{vec}\boldsymbol{\Psi}_2 = \mathbf{L}_{21,p_1}\boldsymbol{\theta}_{\boldsymbol{\psi}_2}$ *because* $\boldsymbol{\theta}_{\boldsymbol{\phi}*} = \operatorname{vech}(\boldsymbol{\Phi}_*^{-1})$ *and* $\boldsymbol{\theta}_{\boldsymbol{\psi}_2} = \operatorname{diag}(\boldsymbol{\Psi}_2)$*, where* $\mathbf{D}_q$ *and* $\mathbf{L}_{21,p_1}$ *are defined in Table 56.*

It follows from (74) that

$$\operatorname{vec}\mathbf{S}_{\boldsymbol{zz}} \approx (\mathbf{S}_{\boldsymbol{zx}}\otimes\mathbf{S}_{\boldsymbol{zx}})\operatorname{vec}\boldsymbol{\Phi}_*^{-1} + \operatorname{vec}\boldsymbol{\Psi}_2. \tag{194}$$

Accordingly, equation 194 can be written as

$$\operatorname{vec}\mathbf{S}_{\boldsymbol{zz}} \approx (\mathbf{S}_{\boldsymbol{zx}}\otimes\mathbf{S}_{\boldsymbol{zx}})\,\mathbf{D}_q\boldsymbol{\theta}_{\boldsymbol{\phi}*} + \mathbf{L}_{21,p_1}\boldsymbol{\theta}_{\boldsymbol{\psi}_2}. \tag{195}$$

Equation 195 is a linear regression model because $\mathbf{S}_{\boldsymbol{zx}}$ *is treated as a matrix of constants. Specifically, in (195),* $\operatorname{vec}\mathbf{S}_{\boldsymbol{zz}}$ *is a vector of response variables,*

$(\mathbf{S}_{\boldsymbol{zx}} \otimes \mathbf{S}_{\boldsymbol{zx}})\,\mathbf{D}_q$ *and* $\mathbf{L}_{21,p_1}$ *are treated as model matrices, and* $\boldsymbol{\theta}_{\boldsymbol{\phi}*}$ *and* $\boldsymbol{\theta}_{\boldsymbol{\psi}_2}$ *are parameter vectors.*

The following approximate generalized least squares estimator of $\boldsymbol{\theta}_{\boldsymbol{\phi}*}$, $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\phi}*}$, *is based on definition 6.2 on page 187 of Hocking [62].*

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{\phi}*} = \left(\mathbf{X}_{1.2}'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{X}_{1.2}\right)^{-1}\mathbf{X}_{1.2}'\widehat{\boldsymbol{\Omega}}^{-1}\operatorname{vec}\mathbf{S}_{\boldsymbol{zz}}, \tag{196}$$

where $\mathbf{X}_{1.2} = \left[\mathbf{I}_{p_1^2} - \mathbf{L}_{21,p_1}\left(\mathbf{L}_{21,p_1}'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{L}_{21,p_1}\right)^{-1}\mathbf{L}_{21,p_1}'\widehat{\boldsymbol{\Omega}}^{-1}\right](\mathbf{S}_{\boldsymbol{zx}} \otimes \mathbf{S}_{\boldsymbol{zx}})\,\mathbf{D}_q$, $\widehat{\boldsymbol{\Omega}} = \mathbf{S}_{\boldsymbol{zz}} \otimes \mathbf{S}_{\boldsymbol{zz}}$, $\mathbf{D}_q$ *and* $\mathbf{L}_{21,p_1}$ *are defined in Table 56.*

Note that $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\phi}*}$ *is not a conventional GLS estimator because the model matrix* $\mathbf{X}_{1.2}$ *and the weight matrix* $\widehat{\boldsymbol{\Omega}}$ *in (196) are treated as matrices of constants.*

(b.) Based on part (a), let $\widehat{\boldsymbol{\Omega}} = \mathbf{I}_{p_1^2}$. *The approximate ordinary least squares estimator of* $\boldsymbol{\theta}_{\boldsymbol{\phi}*}$ *is*

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{\phi}*} = \left(\mathbf{X}_{1.2}'\mathbf{X}_{1.2}\right)^{-1}\mathbf{X}_{1.2}'\operatorname{vec}\mathbf{S}_{\boldsymbol{zz}}, \tag{197}$$

where $\mathbf{X}_{1.2} = \left[\mathbf{I}_{p_1^2} - \mathbf{L}_{21,p_1}\left(\mathbf{L}_{21,p_1}'\mathbf{L}_{21,p_1}\right)^{-1}\mathbf{L}_{21,p_1}'\right](\mathbf{S}_{\boldsymbol{zx}} \otimes \mathbf{S}_{\boldsymbol{zx}})\,\mathbf{D}_q$. *From Lemma 1 part (f),* $\mathbf{X}_{1.2}$ *can be simplified as* $\mathbf{X}_{1.2} = \left(\mathbf{I}_{p_1^2} - \mathbf{L}_{22,p_1}\right)(\mathbf{S}_{\boldsymbol{zx}} \otimes \mathbf{S}_{\boldsymbol{zx}})\,\mathbf{D}_q$. □

B.20. Proof of Theorem 20

Theorem 20. [Original result].

$$\begin{aligned}
\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_{\lambda*}'} &= 2\mathbf{N}_p(\boldsymbol{\Lambda}_*\boldsymbol{\Phi}_* \otimes \mathbf{I}_p)\mathbf{W}_2,\\
\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_{\phi*}'} &= (\boldsymbol{\Lambda}_* \otimes \boldsymbol{\Lambda}_*)\mathbf{D}_q,\\
\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_{\psi*}'} &= \mathbf{L}_{21,p}\mathbf{D}^{(1)}_{\boldsymbol{\psi};\boldsymbol{\theta}_{\psi*}'}.
\end{aligned}$$

Proof: The derivative expressions in Theorem 20 were checked numerically. □

B.21. Proof of Theorem 21

Theorem 21. [Original result]. *First and second derivatives of the $(q-1)$-dimensional constraints function $\mathbf{z}(\boldsymbol{\theta}_*)$ in (125) with respect to $\boldsymbol{\theta}_{\boldsymbol{\lambda}*}$, $\boldsymbol{\theta}_{\boldsymbol{\delta}*}$, $\boldsymbol{\theta}_{\boldsymbol{\gamma}*}$ and $\boldsymbol{\theta}_{\boldsymbol{\psi}*}$, evaluated at $\mathbf{G}_* = \mathbf{I}_q$, are listed as follows:*

$$
\begin{aligned}
\mathbf{D}^{(1)}_{\mathbf{z};\boldsymbol{\theta}'_{\lambda*}} &= \mathbf{0}_{q-1\times\nu_{*1}}, \\
\mathbf{D}^{(1)}_{\mathbf{z};\boldsymbol{\theta}'_{\delta*}} &= \mathbf{C}'\mathbf{L}'_{21,q}(\boldsymbol{\Gamma}\mathbf{G}\otimes\boldsymbol{\Gamma}\mathbf{G})\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_{\delta}} = \mathbf{C}'\mathbf{L}'_{21,q}(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_{\delta}}, \\
\mathbf{D}^{(1)}_{\mathbf{z};\boldsymbol{\theta}'_{\gamma*}} &= \mathbf{C}'\mathbf{L}'_{21,q}2\mathbf{N}_q\left(\boldsymbol{\Gamma}\mathbf{G}\boldsymbol{\Delta}\otimes\boldsymbol{\Gamma}\right)\mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}} = 2\mathbf{C}'\mathbf{L}'_{21,q}(\boldsymbol{\Gamma}\boldsymbol{\Delta}\otimes\boldsymbol{\Gamma})\mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}}, \\
\mathbf{D}^{(1)}_{\mathbf{z};\boldsymbol{\theta}'_{\psi*}} &= \mathbf{0}_{q-1\times\nu_{*4}}, \\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\lambda*},\boldsymbol{\theta}'_{\lambda*}} &= \mathbf{0}_{q-1\times\nu_{*1}\nu_{*1}}, \\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\delta*},\boldsymbol{\theta}'_{\lambda*}} &= \mathbf{0}_{q-1\times\nu_{*1}\nu_{*2}}, \\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\lambda*}} &= \mathbf{0}_{q-1\times\nu_{*1}\nu_{*3}}, \\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\psi*},\boldsymbol{\theta}'_{\lambda*}} &= \mathbf{0}_{q-1\times\nu_{*1}\nu_{*4}}, \\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\delta*},\boldsymbol{\theta}'_{\delta*}} &= \mathbf{C}'\mathbf{L}'_{21,q}(\boldsymbol{\Gamma}\mathbf{G}\otimes\boldsymbol{\Gamma}\mathbf{G})\mathbf{L}_{21,q}\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_{\delta},\boldsymbol{\theta}'_{\delta}} = \mathbf{C}'\mathbf{L}'_{21,q}(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\mathbf{L}_{21,q}\mathbf{D}^{(2)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_{\delta},\boldsymbol{\theta}'_{\delta}}, \\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\delta*}} &= -2\mathbf{C}'\mathbf{L}'_{21,q}[\boldsymbol{\Gamma}\otimes\operatorname{vec}'(\mathbf{I}_q)\otimes\boldsymbol{\Gamma}]\left(\mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}}\otimes\mathbf{L}_{21,q}\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_{\delta}}\right), \\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\psi*},\boldsymbol{\theta}'_{\delta*}} &= \mathbf{0}_{q-1\times\nu_{*2}\nu_{*4}}, \\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\gamma*}} &= 2\mathbf{C}'\mathbf{L}'_{21,q}(\boldsymbol{\Gamma}\boldsymbol{\Delta}\otimes\boldsymbol{\Gamma})\mathbf{D}^{(2)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\gamma*}} \\
&- 2\mathbf{C}'\mathbf{L}'_{21,q}[\boldsymbol{\Gamma}\otimes\operatorname{vec}'(\boldsymbol{\Delta})\otimes\boldsymbol{\Gamma}]\left(\mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}}\otimes\mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}}\right), \\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\psi*},\boldsymbol{\theta}'_{\gamma*}} &= \mathbf{0}_{q-1\times\nu_{*3}\nu_{*4}}, \ \textit{and} \\
\mathbf{D}^{(2)}_{\mathbf{z};\boldsymbol{\theta}'_{\psi*},\boldsymbol{\theta}'_{\psi*}} &= \mathbf{0}_{q-1\times\nu_{*4}\nu_{*4}},
\end{aligned}
$$

where $\mathbf{D}^{(1)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*}}$ and $\mathbf{D}^{(2)}_{\mathbf{g}_*;\boldsymbol{\theta}'_{\gamma*},\boldsymbol{\theta}'_{\gamma*}}$ were given in Theorem 16.

Proof: *The proof is omitted. The derivative expressions in Theorem 21 were checked numerically.* □

B.22. Theorem 22

Theorem 22. [Boik [33], 2011, Details on Lagrange Algorithm, Theorem 1, Page 3]. *Assuming that a solution to the Lagrange equations in (127) exists in which the Lagrange multipliers are finite and the required inverses exist, one solution (the Moore-Penrose solution) to (134) yields the following modified Newton update,*

$$\widehat{\boldsymbol{\omega}}_{i+1} = \begin{pmatrix} \hat{\boldsymbol{\theta}}_{*(i+1)} \\ \widehat{\boldsymbol{\zeta}}_{(i+1)} \end{pmatrix} = \begin{pmatrix} \widehat{\boldsymbol{\theta}}_{*i} - \alpha \mathbf{D}^{(1)+}_{\mathbf{z};\hat{\boldsymbol{\theta}}'_{*i}} \mathbf{z}(\widehat{\boldsymbol{\theta}}_{*i}) \\ \widehat{\boldsymbol{\zeta}}_i \, (1-\alpha) \end{pmatrix}$$

$$+\alpha \begin{pmatrix} -\widehat{\mathbf{F}}_i^{\perp} \left(\widehat{\mathbf{F}}_i^{\perp\prime} \widehat{\mathbf{A}}_{i,11} \widehat{\mathbf{F}}_i^{\perp} \right)^{-1} \widehat{\mathbf{F}}_i^{\perp\prime} \\ \mathbf{D}^{(1)+\prime}_{\mathbf{z};\hat{\boldsymbol{\theta}}'_{*i}} \left[\mathbf{I}_{\nu*} - \widehat{\mathbf{A}}_{i,11} \widehat{\mathbf{F}}_i^{\perp} \left(\widehat{\mathbf{F}}_i^{\perp\prime} \widehat{\mathbf{A}}_{i,11} \widehat{\mathbf{F}}_i^{\perp} \right)^{-1} \widehat{\mathbf{F}}_i^{\perp\prime} \right] \end{pmatrix} \left(\mathbf{D}^{(1)}_{F;\hat{\boldsymbol{\theta}}_{*i}} - \widehat{\mathbf{A}}_{i,11} \mathbf{D}^{(1)+}_{\mathbf{z};\hat{\boldsymbol{\theta}}'_{*i}} \mathbf{z}(\widehat{\boldsymbol{\theta}}_{*i}) \right),$$

where $\widehat{\mathbf{U}}_i \widehat{\mathbf{D}}_i \widehat{\mathbf{F}}'_i$ *is the full-rank SVD* $\left(\mathbf{D}^{(1)}_{\mathbf{z};\hat{\boldsymbol{\theta}}'_{*i}}\right)$, $\widehat{\mathbf{U}}_i \in \mathcal{O}(q-1)$, $\widehat{\mathbf{D}}_i \in \mathcal{D}^+(q-1)$, $\widehat{\mathbf{F}}_i \in \mathcal{O}(\nu^*, (q-1))$, $q-1 = rank \left(\mathbf{D}^{(1)}_{\mathbf{z};\hat{\boldsymbol{\theta}}'_{*i}}\right)$, ν^* *is defined in (57),* $\mathbf{D}^{(1)+}_{\mathbf{z};\hat{\boldsymbol{\theta}}'_{*i}}$ *and* $\widehat{\mathbf{F}}_i^{\perp}$ *are defined in Table 56, and* $\mathcal{O}(\nu^*, (q-1))$ *is defined in Table 57.*

B.23. Theorem 23

Theorem 23. [Slutsky's Theorem]. *Let* $\mathbf{t}_n$ *be a random* $p \times 1$ *vector that satisfies* $\mathbf{t}_n \xrightarrow{\text{dist}} \mathbf{t}$ *as* $n \to \infty$, *where* $\mathbf{t}$ *is a random* $p \times 1$ *vector. Suppose that* $\mathbf{a}_n$ *is a random* $k \times 1$ *vector and that* $\mathbf{B}_n$ *is a random* $k \times p$ *matrix that satisfy* $\mathbf{a}_n \xrightarrow{\text{prob}} \mathbf{a}$ *and* $\mathbf{B}_n \xrightarrow{\text{prob}} \mathbf{B}$, *where* $\mathbf{a}$ *is a* $k \times 1$ *vector of constants and* $\mathbf{B}$ *is a* $k \times p$ *matrix of constants. Then*

1. $\mathbf{a}_n + \mathbf{B}_n \mathbf{t}_n \xrightarrow{\text{dist}} \mathbf{a} + \mathbf{B}\mathbf{t}$

2. $\mathbf{a}_n + \mathbf{B}_n^{-1} \mathbf{t}_n \xrightarrow{\text{dist}} \mathbf{a} + \mathbf{B}^{-1}\mathbf{t}$, *if* $k = p$ *and* $\mathbf{B}^{-1}$ *exists.*

B.24. Theorem 24

Theorem 24. [A Well-Known Result].

(1.) If g is a continuous function and $\mathbf{t}_n \xrightarrow{\text{dist}} \mathbf{t}$, *then* $g(\mathbf{t}_n) \xrightarrow{\text{dist}} g(\mathbf{t})$.

(2.) If g is a continuous function and $\mathbf{t}_n \xrightarrow{\text{prob}} \mathbf{t}$, *then* $g(\mathbf{t}_n) \xrightarrow{\text{prob}} g(\mathbf{t})$.

B.25. Theorem 25

Theorem 25. [Delta Method]. *Let $\mathbf{t}_n$ be a random $p \times 1$ vector with asymptotic distribution* $\sqrt{n}(\mathbf{t}_n - \boldsymbol{\theta}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \boldsymbol{\Omega}_{\mathbf{t}})$. *Suppose that $\mathbf{f}(\mathbf{t}_n)$ is a vector-valued differentiable function. Then*

$$\sqrt{n}\left[\mathbf{f}(\mathbf{t}_n) - \mathbf{f}(\boldsymbol{\theta})\right] \xrightarrow{\text{dist}} \mathbf{N}\left[\mathbf{0}, \mathbf{D}(\boldsymbol{\theta})\boldsymbol{\Omega}_{\mathbf{t}}\mathbf{D}(\boldsymbol{\theta})'\right],$$

where $\mathbf{D}(\boldsymbol{\theta}) = \left.\frac{\partial\, \mathbf{f}(\mathbf{t}_n)}{\partial\, \mathbf{t}_n'}\right|_{\mathbf{t}_n = \boldsymbol{\theta}}$.

B.26. Theorem 26

Theorem 26. [A Well-Known Result]. *Let $\mathbf{H}_{\mathbf{x}}$ be the projection operator that projects onto $\mathcal{R}(\mathbf{X})$ along $\mathcal{N}(\mathbf{X}')$, that is,* $\mathbf{H}_{\mathbf{x}} = \text{ppo}(\mathbf{X}) = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-}\mathbf{X}'$, *where* $\text{ppo}(\mathbf{X})$ *is defined in Table 56. Then,* $\mathbf{H}_{\mathbf{x}}\mathbf{X} = \mathbf{X}$, $\mathbf{H}_{\mathbf{x}} = \mathbf{H}_x'$ *and* $\text{rank}(\mathbf{H}_{\mathbf{x}}) = \text{tr}(\mathbf{H}_{\mathbf{x}})$.

B.27. Proof of Theorem 27

Theorem 27. [Asymptotic Distribution of $\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})$][Muirhead [45], 1982, Theorem 1.2.17, Page 19]. *The model* $\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{F}\boldsymbol{\Lambda}' + \mathbf{E}$ *can be rewritten as*

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{Z},$$

where $\mathbf{Z} = \mathbf{F}\boldsymbol{\Lambda}' + \mathbf{E}$ *is an $N \times p$ matrix of random errors with* $E[\mathbf{Z}] = \mathbf{0}_{N\times p}$ *and* $Var(\text{vec}\,\mathbf{Z}) = \boldsymbol{\Sigma} \otimes \mathbf{I}_N$. *Denote the i^{th} row of $\mathbf{Z}$ by $\mathbf{z}_i'$ and assume that* $\mathbf{z}_i \overset{\text{iid}}{\sim} (\mathbf{0}, \boldsymbol{\Sigma})$ *with finite fourth moments, where* $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$.

The asymptotic distribution of $\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})$ *is*

$$\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \boldsymbol{\Omega}),$$

where $\mathbf{s} = \operatorname{vec}\mathbf{S}$, $\boldsymbol{\sigma} = \operatorname{vec}\boldsymbol{\Sigma}$, $\boldsymbol{\Omega} = \mathrm{E}\left[\mathbf{z}_i\mathbf{z}_i' \otimes \mathbf{z}_i\mathbf{z}_i'\right] - \boldsymbol{\sigma}\boldsymbol{\sigma}'$, *and* $\mathbf{S}$ *is the* $p \times p$ *sample covariance matrix.*

Proof: [Proof With Advisor's Help]. *The proof of Theorem 27 is composed of two parts, Part I and Part II. Part I is to show that* $\sqrt{n}\operatorname{vec}\left(n^{-1}\mathbf{Z}'\mathbf{Z}-\boldsymbol{\Sigma}\right) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \boldsymbol{\Omega})$ *and Part II is to show* $\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \boldsymbol{\Omega})$ *based on part I, where* $n = N - r_x$ *and* $\operatorname{rank}(\mathbf{X}) = r_x$.

Part I: Consider the $p^2 \times 1$ *vector* $(\mathbf{z}_i \otimes \mathbf{z}_i) \overset{\text{iid}}{\sim} (\operatorname{vec}\boldsymbol{\Sigma}, \boldsymbol{\Omega})$ *for* $i = 1, \ldots, N$, *where* $\boldsymbol{\Omega} = \operatorname{Var}(\mathbf{z}_i \otimes \mathbf{z}_i)$ *and examine*

$$\begin{aligned}
\boldsymbol{\Omega} = \operatorname{Var}(\mathbf{z}_i \otimes \mathbf{z}_i) &= \mathrm{E}\left[(\mathbf{z}_i \otimes \mathbf{z}_i)(\mathbf{z}_i \otimes \mathbf{z}_i)'\right] - \mathrm{E}\left[\mathbf{z}_i \otimes \mathbf{z}_i\right]\mathrm{E}\left[\mathbf{z}_i' \otimes \mathbf{z}_i'\right] \\
&= \mathrm{E}\left[\mathbf{z}_i\mathbf{z}_i' \otimes \mathbf{z}_i\mathbf{z}_i'\right] - (\operatorname{vec}\boldsymbol{\Sigma})(\operatorname{vec}'\boldsymbol{\Sigma}) \\
&= \mathrm{E}\left[\mathbf{z}_i\mathbf{z}_i' \otimes \mathbf{z}_i\mathbf{z}_i'\right] - \boldsymbol{\sigma}\boldsymbol{\sigma}' < \infty, \ \textit{because} \qquad (198) \\
\operatorname{vec}\left[\mathrm{E}\left(\mathbf{z}_i\mathbf{z}_i' \otimes \mathbf{z}_i\mathbf{z}_i'\right)\right] &= \mathrm{E}\left\{\operatorname{vec}\left[(\mathbf{z}_i \otimes \mathbf{z}_i)(\mathbf{z}_i \otimes \mathbf{z}_i)'\right]\right\} \\
&= \mathrm{E}\left(\mathbf{z}_i \otimes \mathbf{z}_i \otimes \mathbf{z}_i \otimes \mathbf{z}_i\right) < \infty.
\end{aligned}$$

By the multivariate central limit theorem on page 48 of Shao [63], if the $p^2 \times 1$ *vector* $(\mathbf{z}_i \otimes \mathbf{z}_i) \overset{\text{iid}}{\sim} (\operatorname{vec}\boldsymbol{\Sigma}, \boldsymbol{\Omega})$ *for* $i = 1, \ldots, N$ *with* $\boldsymbol{\Omega} < \infty$, *then*

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i \otimes \mathbf{z}_i) - \operatorname{vec}\boldsymbol{\Sigma}\right) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \boldsymbol{\Omega}), \qquad (199)$$

where $\operatorname{vec}\boldsymbol{\Sigma} = \mathrm{E}\left[1/N\sum_{i=1}^{N}(\mathbf{z}_i \otimes \mathbf{z}_i)\right]$ *and* $\boldsymbol{\Omega} = \mathrm{E}\left[\mathbf{z}_i\mathbf{z}_i' \otimes \mathbf{z}_i\mathbf{z}_i'\right] - \boldsymbol{\sigma}\boldsymbol{\sigma}'$.

Part I is completed by examining $\sqrt{n}\,\text{vec}\,(n^{-1}\mathbf{Z}'\mathbf{Z}-\mathbf{\Sigma})$ *as follows:*

$$\begin{aligned}\sqrt{n}\,\text{vec}\left(n^{-1}\mathbf{Z}'\mathbf{Z}-\mathbf{\Sigma}\right) &= \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{N}(\mathbf{z}_i\otimes\mathbf{z}_i)-\text{vec}\,\mathbf{\Sigma}\right]\\ &\qquad \textit{because}\;\; \text{vec}\,(\mathbf{Z}'\mathbf{Z})=\sum_{i=1}^{N}(\mathbf{z}_i\otimes\mathbf{z}_i)\\ &= \frac{\sqrt{n}}{\sqrt{N}}\sqrt{N}\left[\frac{N}{n}\frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i\otimes\mathbf{z}_i)-\text{vec}\,\mathbf{\Sigma}\right]\end{aligned} \tag{200}$$

Note that $\lim_{n\to\infty}\frac{\sqrt{n}}{\sqrt{N}}=\lim_{n\to\infty}\sqrt{\frac{n}{n+r_x}}=\lim_{n\to\infty}\sqrt{\frac{1}{1+r_x/n}}=1$ *because* $r_x=O(1)$. *Similarly, it can be shown that* $\lim_{n\to\infty}\frac{N}{n}=\lim_{n\to\infty}\frac{n+r_x}{n}=\lim_{n\to\infty}\left(1+\frac{r_x}{n}\right)=1$.

Based on (200), it can be concluded that

$$\sqrt{n}\,\text{vec}\left(n^{-1}\mathbf{Z}'\mathbf{Z}-\mathbf{\Sigma}\right)\xrightarrow{\text{dist}}\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i\otimes\mathbf{z}_i)-\text{vec}\,\mathbf{\Sigma}\right)\xrightarrow{\text{dist}}\mathbf{N}(\mathbf{0},\mathbf{\Omega}), \tag{201}$$

by Slutsky's Theorem 23.

Part II: The $p\times p$ *sample covariance matrix* $\mathbf{S}$ *can be computed as follows:*

$$\begin{aligned}\mathbf{S} &= n^{-1}\mathbf{Y}'(\mathbf{I}-\mathbf{H_x})\mathbf{Y}=n^{-1}(\mathbf{XB}+\mathbf{Z})'(\mathbf{I}-\mathbf{H_x})(\mathbf{XB}+\mathbf{Z})\\ &\qquad \textit{where}\quad \mathbf{H_x}=\text{ppo}(\mathbf{X})\\ &= n^{-1}\left[\mathbf{B}'\mathbf{X}'(\mathbf{I}-\mathbf{H_x})\mathbf{XB}+\mathbf{B}'\mathbf{X}'(\mathbf{I}-\mathbf{H_x})\mathbf{Z}\right]\\ &+ n^{-1}\left[\mathbf{Z}'(\mathbf{I}-\mathbf{H_x})\mathbf{XB}+\mathbf{Z}'(\mathbf{I}-\mathbf{H_x})\mathbf{Z}\right]\\ &= n^{-1}\mathbf{Z}'(\mathbf{I}-\mathbf{H_x})\mathbf{Z}\\ &\qquad \textit{by Theorem 26}\\ &= n^{-1}\mathbf{Z}'\mathbf{Z}-n^{-1}\mathbf{Z}'\mathbf{H_x}\mathbf{Z}.\end{aligned} \tag{202}$$

Factor $\mathbf{H_x}$ *as* $\mathbf{H_x} = \mathbf{U}\mathbf{U}'$, *where* $\mathbf{U} \in \mathcal{O}(N, r_x)$ *and* $\mathcal{O}(N, r_x)$ *is defined in Table 57. Then* $\mathbf{Z}'\mathbf{H_x}\mathbf{Z} = \mathbf{T}'\mathbf{T}$, *where* $\mathbf{T} = \mathbf{U}'\mathbf{Z}$. *It can be concluded that*

$$\begin{aligned} &\mathrm{E}\,(\mathrm{vec}\,\mathbf{T}) = \mathrm{E}\,[\mathrm{vec}\,(\mathbf{U}'\mathbf{Z})] = \mathrm{vec}\,\mathrm{E}\,(\mathbf{U}'\mathbf{Z}) = \mathbf{0} \quad \textit{because} \quad \mathrm{E}\,(\mathbf{Z}) = \mathbf{0}, \\ &\textit{and} \quad \mathrm{Var}\,(\mathrm{vec}\,\mathbf{T}) = \mathrm{Var}\,[(\mathbf{I}_p \otimes \mathbf{U}')\,\mathrm{vec}\,\mathbf{Z}] = \mathbf{\Sigma} \otimes \mathbf{I}_{r_x}, \end{aligned} \tag{203}$$

because $\mathbf{U}'\mathbf{U} = \mathbf{I}_{r_x}$ *and* $Var(\mathrm{vec}\,\mathbf{Z}) = \mathbf{\Sigma} \otimes \mathbf{I}_N$. *Furthermore,*

$$\begin{aligned} \mathrm{E}\left(\|\mathbf{T} - \mathrm{E}(\mathbf{T})\|^2\right) &= \mathrm{E}\left(\|\mathbf{T}\|^2\right) \quad \textit{because} \quad \mathrm{E}\,(\mathbf{T}) = \mathbf{U}'\,\mathrm{E}\,(\mathbf{Z}) = \mathbf{0} \\ &= \mathrm{E}\,(\mathrm{vec}'\mathbf{T}\,\mathrm{vec}\,\mathbf{T}) \quad \textit{because} \quad \|\mathbf{T}\|^2 = \mathrm{vec}'\mathbf{T}\,\mathrm{vec}\,\mathbf{T} \\ &= \mathrm{E}\,(\mathrm{vec}'\mathbf{T}\mathbf{I}_{pr_x}\,\mathrm{vec}\,\mathbf{T}) \\ &= \mathrm{tr}\,[\mathbf{I}_{pr_x}\,\mathrm{Var}\,(\mathrm{vec}\,\mathbf{T})] + \mathrm{E}'\,(\mathrm{vec}\,\mathbf{T})\,\mathbf{I}_{pr_x}\,\mathrm{E}\,(\mathrm{vec}\,\mathbf{T}) \\ &= \mathrm{tr}\,[\mathbf{I}_{pr_x}\,(\mathbf{\Sigma} \otimes \mathbf{I}_{r_x})] + \mathbf{0}' \times \mathbf{I}_{pr_x} \times \mathbf{0} \\ &\qquad \textit{because of}\ (203) \\ &= \mathrm{tr}(\mathbf{\Sigma})\,\mathrm{tr}(\mathbf{I}_{r_x}) = r_x\,\mathrm{tr}\,(\mathbf{\Sigma}), \quad \textit{and} \\ \sqrt{\mathrm{E}\,(\|\mathbf{T} - \mathrm{E}(\mathbf{T})\|^2)} &= \sqrt{r_x\,\mathrm{tr}\,(\mathbf{\Sigma})} = O(1), \end{aligned}$$

because $r_x = O(1)$ *and* $\mathrm{tr}\,(\mathbf{\Sigma})$ *is fixed. Accordingly,* $\mathbf{T} = O_p(1)$ *because* $\mathbf{T} - \mathrm{E}(\mathbf{T}) = O_p(1)$ *and* $\mathrm{E}\,(\mathbf{T}) = \mathbf{0}$.

Based on $\mathbf{S}$ *in (202), write* $\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})$ *as*

$$\begin{aligned} \sqrt{n}(\mathbf{s} - \boldsymbol{\sigma}) &= \sqrt{n}\,\mathrm{vec}\left(n^{-1}\mathbf{Z}'\mathbf{Z} - n^{-1}\mathbf{Z}'\mathbf{H_x}\mathbf{Z} - \mathbf{\Sigma}\right) \\ &= \sqrt{n}\,\mathrm{vec}\left(n^{-1}\mathbf{Z}'\mathbf{Z} - \mathbf{\Sigma}\right) - \sqrt{n}\,\mathrm{vec}\left(n^{-1}\mathbf{Z}'\mathbf{H_x}\mathbf{Z}\right) \\ &= \sqrt{n}\,\mathrm{vec}\left(n^{-1}\mathbf{Z}'\mathbf{Z} - \mathbf{\Sigma}\right) - n^{-1/2}\,\mathrm{vec}\,(\mathbf{T}'\mathbf{T}) \\ &\quad \textit{because} \quad \mathbf{Z}'\mathbf{H_x}\mathbf{Z} = \mathbf{T}'\mathbf{T} \\ &= \sqrt{n}\,\mathrm{vec}\left(n^{-1}\mathbf{Z}'\mathbf{Z} - \mathbf{\Sigma}\right) - n^{-1/2}O_p(1) \\ &\quad \textit{because} \quad \mathbf{T} = O_p(1) \\ &= \sqrt{n}\left(n^{-1}\mathbf{Z}'\mathbf{Z} - \mathbf{\Sigma}\right) + O_p(n^{-1/2}), \end{aligned} \tag{204}$$

because $n^{-1/2}O_p(1) = O_p(n^{-1/2})$.

Therefore, $\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})$ *and* $\sqrt{n}\,\mathrm{vec}\,(n^{-1}\mathbf{Z}'\mathbf{Z}-\boldsymbol{\Sigma})$ *are asymptotically equal in distribution by Slutsky's Theorem 23. From the result in Part I that* $\sqrt{n}\,(n^{-1}\mathbf{Z}'\mathbf{Z}-\boldsymbol{\Sigma}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0},\boldsymbol{\Omega})$, *it can be concluded that* $\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0},\boldsymbol{\Omega})$. □

B.27.1. Proof of corollary 27.1

Corollary 27.1. [Asymptotic Distribution of $\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})$ Under Normality][Magnus and Neudecker [47], 1979, Corollary 4.2, Page 394 and Muirhead [45], 1982, Corollary 1.2.18, Page 19]. *Assume that* $\mathbf{z}_i \stackrel{\text{iid}}{\sim} \mathbf{N}(\mathbf{0},\boldsymbol{\Sigma})$, *where* $\mathbf{z}_i'$ *is the* i^{th} *row of* $\mathbf{Z}$ *in (135). The asymptotic distribution of* $\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})$ *is*

$$\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0},\boldsymbol{\Omega}),$$

where $\mathbf{s} = \mathrm{vec}\,\mathbf{S}$, $\boldsymbol{\sigma} = \mathrm{vec}\,\boldsymbol{\Sigma}$ *and* $\boldsymbol{\Omega} = 2\mathrm{N}_p\,(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})$.

Proof: [Independent Proof]. *Based on the proof provided in Theorem 27, the proof of Corollary 27.1 is completed by providing the expression of* $\boldsymbol{\Omega}$ *under* $\mathbf{z}_i \stackrel{\text{iid}}{\sim} \mathbf{N}(\mathbf{0},\boldsymbol{\Sigma})$. *Denote* $\mathbf{z}_i$ *by* $\mathbf{z}$ *in this proof for notation convenience. Accordingly, the moment generating function of* $\mathbf{z}$ *is* $M_{\mathbf{z}}(\mathbf{t}) = e^{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}$, *where* $\mathbf{t}\colon p\times 1$.

Examine

$$\begin{aligned}
\frac{\partial M_{\mathbf{z}}(\mathbf{t})}{\partial \mathbf{t}} &= \frac{\partial\left(\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right)}{\partial \mathbf{t}} e^{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}} = \boldsymbol{\Sigma}\mathbf{t}e^{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}} \\
\frac{\partial^2 M_{\mathbf{z}}(\mathbf{t})}{\partial\,\mathbf{t}\otimes\partial\,\mathbf{t}} &= (\mathrm{vec}\,\boldsymbol{\Sigma})e^{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}} + (\mathbf{I}_p\otimes\boldsymbol{\Sigma}\mathbf{t})\,\frac{\partial\left(\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right)}{\partial \mathbf{t}} e^{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}} \\
&= (\mathrm{vec}\,\boldsymbol{\Sigma})e^{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}} + (\boldsymbol{\Sigma}\mathbf{t}\otimes\boldsymbol{\Sigma}\mathbf{t})\,e^{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}} \\
\frac{\partial^3 M_{\mathbf{z}}(\mathbf{t})}{\partial\,\mathbf{t}\otimes\partial\,\mathbf{t}\otimes\partial\,\mathbf{t}'} &= [\mathbf{I}_1\otimes(\mathrm{vec}\,\boldsymbol{\Sigma})]\,\mathbf{t}'\boldsymbol{\Sigma}\left(\mathbf{I}_p\otimes e^{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}\right) \\
&\quad + \frac{\partial\,(\boldsymbol{\Sigma}\mathbf{t}\otimes\boldsymbol{\Sigma}\mathbf{t})}{\partial \mathbf{t}'}\left(\mathbf{I}_p\otimes e^{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}\right) \\
&\quad + [\mathbf{I}_1\otimes(\boldsymbol{\Sigma}\mathbf{t}\otimes\boldsymbol{\Sigma}\mathbf{t})]\,\mathbf{t}'\boldsymbol{\Sigma}\left(\mathbf{I}_p\otimes e^{\frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}\right)
\end{aligned}$$

$$
\begin{aligned}
&= (\operatorname{vec}\mathbf{\Sigma})\mathbf{t}'\mathbf{\Sigma}\left(\mathbf{I}_p \otimes e^{\frac{1}{2}\mathbf{t}'\mathbf{\Sigma}\mathbf{t}}\right) \\
&\quad + [(\mathbf{\Sigma}\otimes\mathbf{\Sigma}\mathbf{t}) + \mathbf{K}_{p,p}(\mathbf{\Sigma}\otimes\mathbf{\Sigma}\mathbf{t})]\left(\mathbf{I}_p \otimes e^{\frac{1}{2}\mathbf{t}'\mathbf{\Sigma}\mathbf{t}}\right) \\
&\quad + (\mathbf{\Sigma}\mathbf{t}\otimes\mathbf{\Sigma}\mathbf{t})\,\mathbf{t}'\mathbf{\Sigma}\left(\mathbf{I}_p \otimes e^{\frac{1}{2}\mathbf{t}'\mathbf{\Sigma}\mathbf{t}}\right) \\
&= [(\operatorname{vec}\mathbf{\Sigma})\mathbf{t}'\mathbf{\Sigma} + 2\mathbf{N}_p(\mathbf{\Sigma}\otimes\mathbf{\Sigma}\mathbf{t}) + (\mathbf{\Sigma}\mathbf{t}\otimes\mathbf{\Sigma}\mathbf{t})\,\mathbf{t}'\mathbf{\Sigma}] \\
&\quad \times \left(\mathbf{I}_p \otimes e^{\frac{1}{2}\mathbf{t}'\mathbf{\Sigma}\mathbf{t}}\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^4 M_{\mathbf{z}}(\mathbf{t})}{\partial\,\mathbf{t}\otimes\partial\,\mathbf{t}\otimes\partial\,\mathbf{t}'\otimes\partial\,\mathbf{t}'} &= [(\operatorname{vec}\mathbf{\Sigma})(\operatorname{vec}'\mathbf{\Sigma}) + 2\mathbf{N}_p\mathbf{K}_{p,p}(\mathbf{\Sigma}\otimes\mathbf{\Sigma})]\left(\mathbf{I}_{p^2} \otimes e^{\frac{1}{2}\mathbf{t}'\mathbf{\Sigma}\mathbf{t}}\right) \\
&\quad + [(\mathbf{\Sigma}\otimes\mathbf{\Sigma}\mathbf{t}) + \mathbf{K}_{p,p}(\mathbf{\Sigma}\otimes\mathbf{\Sigma}\mathbf{t})]\left(\mathbf{I}_p \otimes e^{\frac{1}{2}\mathbf{t}'\mathbf{\Sigma}\mathbf{t}}\mathbf{t}'\mathbf{\Sigma}\right) \\
&\quad + (\mathbf{\Sigma}\mathbf{t}\otimes\mathbf{\Sigma}\mathbf{t})\,(\operatorname{vec}'\mathbf{\Sigma})\left(\mathbf{I}_{p^2} \otimes e^{\frac{1}{2}\mathbf{t}'\mathbf{\Sigma}\mathbf{t}}\right) \\
&\quad + [(\operatorname{vec}\mathbf{\Sigma})\mathbf{t}'\mathbf{\Sigma} + 2\mathbf{N}_p(\mathbf{\Sigma}\otimes\mathbf{\Sigma}\mathbf{t}) + (\mathbf{\Sigma}\mathbf{t}\otimes\mathbf{\Sigma}\mathbf{t})\,\mathbf{t}'\mathbf{\Sigma}] \\
&\quad \times \left(\left[\mathbf{t}'\mathbf{\Sigma}\left(\mathbf{I}_p \otimes e^{\frac{1}{2}\mathbf{t}'\mathbf{\Sigma}\mathbf{t}}\right)\right]\otimes\mathbf{I}_p\right),
\end{aligned}
$$

where $\mathbf{N}_p = \frac{1}{2}[\mathbf{I}_{p^2} + \mathbf{K}_{p,p}]$.

Evaluate $\partial^4 M_{\mathbf{z}}(\mathbf{t})/(\partial\,\mathbf{t}\otimes\partial\,\mathbf{t}\otimes\partial\,\mathbf{t}'\otimes\partial\,\mathbf{t}')$ *at* $t = 0$ *yields*

$$
\operatorname{E}[\mathbf{z}_i\mathbf{z}_i'\otimes\mathbf{z}_i\mathbf{z}_i'] = \left.\frac{\partial^4 M_{\mathbf{z}}(\mathbf{t})}{\partial\,\mathbf{t}\otimes\partial\,\mathbf{t}\otimes\partial\,\mathbf{t}'\otimes\partial\,\mathbf{t}'}\right|_{\mathbf{t}=0} = (\operatorname{vec}\mathbf{\Sigma})(\operatorname{vec}'\mathbf{\Sigma}) + 2\mathbf{N}_p(\mathbf{\Sigma}\otimes\mathbf{\Sigma}).
$$

Accordingly, $\mathbf{\Omega} = \operatorname{E}[\mathbf{z}_i\mathbf{z}_i'\otimes\mathbf{z}_i\mathbf{z}_i'] - (\operatorname{vec}\mathbf{\Sigma})(\operatorname{vec}'\mathbf{\Sigma}) = 2\mathbf{N}_p(\mathbf{\Sigma}\otimes\mathbf{\Sigma})$. □

B.28. Proof of Theorem 28

Theorem 28. [Asymptotic Distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$][Browne [48], 1974, Proposition 6, Page 13 and Browne [49], 1984, Proposition 2, Page 67]. *Let* $\hat{\boldsymbol{\theta}}$ *be the minimum discrepancy estimator of* $\boldsymbol{\theta}$ *in* $F(\mathbf{\Sigma}(\boldsymbol{\theta}),\mathbf{S})$*, where* $F(\mathbf{\Sigma}(\boldsymbol{\theta}),\mathbf{S})$ *is given in (66). Assume that* $\hat{\boldsymbol{\theta}}$ *is a consistent root of* $F(\mathbf{\Sigma}(\boldsymbol{\theta}),\mathbf{S})$ *and* $\operatorname{rank}(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}) = \nu$*, where* $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}$ *is given in (48). The asymptotic distribution of* $\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ *is*

$$
\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0},\mathbf{\Omega}_{\boldsymbol{\theta}}),
$$

where $\boldsymbol{\Omega}_{\boldsymbol{\theta}} = \mathbf{A}'_{\boldsymbol{\theta}}\boldsymbol{\Omega}\mathbf{A}_{\boldsymbol{\theta}}$, $\mathbf{A}_{\boldsymbol{\theta}} = \left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]^{-1}$, *and* $\boldsymbol{\Omega} = \mathrm{E}\left[\mathbf{z}_i\mathbf{z}_i'\otimes\mathbf{z}_i\mathbf{z}_i'\right] - \boldsymbol{\sigma}\boldsymbol{\sigma}'$ *is given in Theorem 27.*

Proof: [Proof With Advisor's Help]. *Define* $\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}_1,\boldsymbol{\theta}_1',\boldsymbol{\theta}_1'}$ *as* $\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}_1,\boldsymbol{\theta}_1',\boldsymbol{\theta}_1'} = \mathbf{D}^{(3)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}',\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_1}$, *where* $\mathbf{D}^{(3)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}',\boldsymbol{\theta}'}$ *is defined in Theorem 17 and* $\boldsymbol{\theta}_1 = \alpha\hat{\boldsymbol{\theta}} + (1-\alpha)\boldsymbol{\theta}$ *for some* $\alpha\in(0,1)$.

Expand $\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ *in a Taylor series about* $\hat{\boldsymbol{\theta}}=\boldsymbol{\theta}$, *where* $\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}$ *is defined in Theorem 17. It follows from* $\hat{\boldsymbol{\theta}}$ *is a solution to* $\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}} = \mathbf{0}$ *that*

$$\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \mathbf{0} = \mathbf{D}^{(1)}_{F;\boldsymbol{\theta}} + \mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) + \frac{1}{2}\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}_1,\boldsymbol{\theta}_1',\boldsymbol{\theta}_1'}\left[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\otimes(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right], \tag{205}$$

where $\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}$ *is defined in Theorem 17.*

Accordingly,

$$\begin{aligned}
\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}} &= -\left\{\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) + \frac{1}{2}\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}_1,\boldsymbol{\theta}_1',\boldsymbol{\theta}_1'}\left[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\otimes(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right]\right\};\\
\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}} &= -\left\{\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'} + \frac{1}{2}\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}_1,\boldsymbol{\theta}_1',\boldsymbol{\theta}_1'}\left[\mathbf{I}_\nu\otimes(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right]\right\}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right);\ \textit{and}\\
(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) &= -\left\{\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'} + \frac{1}{2}\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}_1,\boldsymbol{\theta}_1',\boldsymbol{\theta}_1'}\left[\mathbf{I}_\nu\otimes(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right]\right\}^{-1}\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}.
\end{aligned}\tag{206}$$

That is,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) = -\left\{\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'} + \frac{1}{2}\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}_1,\boldsymbol{\theta}_1',\boldsymbol{\theta}_1'}[\mathbf{I}_\nu\otimes(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})]\right\}^{-1}\sqrt{n}\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}, \tag{207}$$

where $\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}} = -\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma})$.

Note that

$$\begin{aligned}
\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'} &= \mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\\
&\quad + 2\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}(\mathbf{S}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\right]\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\\
&\quad - \mathbf{D}^{(2)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}\left[\mathbf{I}_\nu\otimes\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathrm{vec}(\mathbf{S}-\boldsymbol{\Sigma})\right]\qquad\textit{and}\\
E\left(\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}\right) &= \mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}.
\end{aligned}\tag{208}$$

Based on (208), it is readily shown that

$$\mathrm{vec}(\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}) - \mathrm{E}\left[\mathrm{vec}\left(\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}\right)\right] = \textit{A Matrix of Constants}\times(\mathbf{s}-\boldsymbol{\sigma}). \tag{209}$$

It can be concluded from (209) that

$$\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'} - \mathrm{E}\left[\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}\right] = O_p(n^{-1/2}), \tag{210}$$

because $\mathbf{s} - \boldsymbol{\sigma} = O_p(n^{-1/2})$ *by Theorem 27.*

Similarly, based on the expressions of $\mathbf{D}^{(3)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}',\boldsymbol{\theta}'}$ *and* $\mathrm{E}(\mathbf{D}^{(3)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}',\boldsymbol{\theta}'})$ *in Theorem 17, it can be concluded that*

$$\mathbf{D}^{(3)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}',\boldsymbol{\theta}'} - \mathrm{E}(\mathbf{D}^{(3)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}',\boldsymbol{\theta}'}) = A\ Matrix\ of\ Constants \times \left[\mathbf{I}_{\nu^2} \otimes (\mathbf{s} - \boldsymbol{\sigma})\right], \tag{211}$$

where $[\mathbf{I}_{\nu^2} \otimes (\mathbf{s} - \boldsymbol{\sigma})] = O_p(n^{-1/2})$ *because* $\mathbf{s} - \boldsymbol{\sigma} = O_p(n^{-1/2})$ *and* $\mathbf{I}_{\nu^2}$ *is fixed. Further,* $\mathbf{D}^{(3)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}',\boldsymbol{\theta}'} = \mathrm{E}(\mathbf{D}^{(3)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}',\boldsymbol{\theta}'}) + O_p(n^{-1/2}) = O(1) + O_p(n^{-1/2}) = O_p(1)$ *based on (211). Note that* $\boldsymbol{\theta}_1 = \alpha\hat{\boldsymbol{\theta}} + (1-\alpha)\boldsymbol{\theta} = \alpha(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + \boldsymbol{\theta} \xrightarrow{\text{dist}} \boldsymbol{\theta}$ *by Slutsky's Theorem 23 because* $\alpha \in (0,1)$ *and* $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = o_p(1)$*. Further, by Theorem 24, it can be concluded that* $\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}_1,\boldsymbol{\theta}_1',\boldsymbol{\theta}_1'} \xrightarrow{\text{dist}} \mathbf{D}^{(3)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}',\boldsymbol{\theta}'}$ *because* $\boldsymbol{\theta}_1 \xrightarrow{\text{dist}} \boldsymbol{\theta}$*. That is,* $\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}_1,\boldsymbol{\theta}_1',\boldsymbol{\theta}_1'} = O_p(1)$.

Accordingly, (207) can be written as follows:

$$\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) &= -\left\{\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'} + \frac{1}{2}\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}_1,\boldsymbol{\theta}_1',\boldsymbol{\theta}_1'}\left[\mathbf{I}_\nu \otimes (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right]\right\}^{-1} \sqrt{n}\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}} \\
&= -\left[E\left(\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}\right) + Op(n^{-\frac{1}{2}}) + Op(1)o_p(1)\right]^{-1} \sqrt{n}\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}} \\
&\qquad \textit{because (210)}\ , \mathbf{D}^{(3)}_{F;\boldsymbol{\theta}_1,\boldsymbol{\theta}_1',\boldsymbol{\theta}_1'} = O_p(1),\ \textit{and}\ \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = o_p(1) \\
&= -\left[E\left(\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}\right) + Op(n^{-\frac{1}{2}}) + o_p(1)\right]^{-1} \sqrt{n}\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}, \\
&= -\left[E\left(\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}\right) + o_p(1)\right]^{-1} \sqrt{n}\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}, \\
&\xrightarrow{\text{dist}} -\left[E\left(\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}\right)\right]^{-1} \sqrt{n}\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}},
\end{aligned}$$

by Slutsky's Theorem 23.

Examine $\lim_{n\to\infty} \mathrm{Var}\left\{\left[-E\left(\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}\right)\right]^{-1} \sqrt{n}\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}\right\}$ *as follows:*

$$\begin{aligned}
&\lim_{n\to\infty} \mathrm{Var}\left\{\left[-E\left(\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}\right)\right]^{-1} \sqrt{n}\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}\right\} \\
= &\lim_{n\to\infty} \mathrm{Var}\left\{\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]^{-1} \mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\sqrt{n}\,\mathrm{vec}(\mathbf{S} - \boldsymbol{\Sigma})\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \left\{\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]^{-1}\right\}'\boldsymbol{\Omega}\times \\
&\quad \left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]^{-1} \\
&= \mathbf{A}'_{\boldsymbol{\theta}}\boldsymbol{\Omega}\mathbf{A}_{\boldsymbol{\theta}},
\end{aligned}
$$

where $\mathbf{A}_{\boldsymbol{\theta}} = \left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]^{-1}$*, and* $\boldsymbol{\Omega}$ *is given in Theorem 27.*

By Slutsky's Theorem 23 and Theorem 27, it can be concluded that

$$
\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0},\boldsymbol{\Omega}_{\boldsymbol{\theta}}), \tag{212}
$$

where $\boldsymbol{\Omega}_{\boldsymbol{\theta}} = \lim_{n\to\infty}\operatorname{Var}\left\{\left[-E\left(\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}\right)\right]^{-1}\sqrt{n}\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}\right\} = \mathbf{A}'_{\boldsymbol{\theta}}\boldsymbol{\Omega}\mathbf{A}_{\boldsymbol{\theta}}.$ □

B.28.1. Proof of corollary 28.1

Corollary 28.1. [Asymptotic Distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ Under Normality][Browne [49], 1984, Proposition 5, Page 76]. *Assume that* $\mathbf{z}_i \overset{\text{iid}}{\sim} \mathbf{N}(\mathbf{0},\boldsymbol{\Sigma})$*, where* $\mathbf{z}'_i$ *is the* i^{th} *row of* $\mathbf{Z}$ *in (135). Then the asymptotic distribution of* $\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ *is*

$$
\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0},\boldsymbol{\Omega}_{\boldsymbol{\theta}}),
$$

where $\boldsymbol{\Omega}_{\boldsymbol{\theta}} = 2\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]^{-1}$.

Proof: [Independent Proof]. *From Theorem 27.1, under* $\mathbf{z}_i \overset{\text{iid}}{\sim} \mathbf{N}(\mathbf{0},\boldsymbol{\Sigma})$, $\boldsymbol{\Omega} = 2\mathbf{N}_p\left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}\right)$.

The expression of $\boldsymbol{\Omega}_{\boldsymbol{\theta}}$ *in Theorem 28 can be simplified as follows:*

$$
\begin{aligned}
\boldsymbol{\Omega}_{\boldsymbol{\theta}} &= \mathbf{A}'_{\boldsymbol{\theta}}\boldsymbol{\Omega}\mathbf{A}_{\boldsymbol{\theta}} \\
&= \mathbf{A}'_{\boldsymbol{\theta}}2\mathbf{N}_p\left(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}\right)\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]^{-1} \\
&= \mathbf{A}'_{\boldsymbol{\theta}}2\mathbf{N}_p\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]^{-1} \\
&= 2\mathbf{A}'_{\boldsymbol{\theta}}\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]^{-1}
\end{aligned}
$$

$$because\ \mathbf{N}_p \operatorname{vec}\boldsymbol{\Sigma} = \operatorname{vec}\boldsymbol{\Sigma}\ and\ \mathbf{N}_p\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} = \partial\mathbf{N}_p\boldsymbol{\sigma}/\partial\boldsymbol{\theta}' = \partial\boldsymbol{\sigma}/\partial\boldsymbol{\theta}' = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}$$

$$\begin{aligned} &= 2\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]^{-1}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \\ &\times \left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]^{-1} \\ &= 2\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\right]^{-1}. \end{aligned}$$

□

B.29. Proof of Theorem 29

Theorem 29. [Asymptotic Distribution of $\sqrt{n}(\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda})$][Original result]. *The asymptotic distribution of* $\sqrt{n}(\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda})$ *is*

$$\sqrt{n}(\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \mathbf{D}^{(1)}_{\boldsymbol{\lambda};\boldsymbol{\theta}'_\lambda}\mathbf{E}'_{1,\nu}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\mathbf{E}_{1,\nu}\mathbf{D}^{(1)\prime}_{\boldsymbol{\lambda};\boldsymbol{\theta}'_\lambda}),$$

where $\mathbf{D}^{(1)}_{\boldsymbol{\lambda};\boldsymbol{\theta}'_\lambda}$ *is given in (24),* $\mathbf{E}_{1,\nu_1}$ *is the* 1st *submatrix of*
$\mathbf{I}_\nu = (\mathbf{E}_{1,\nu_1}\ \ \mathbf{E}_{2,\nu_2}\ \ \mathbf{E}_{3,\nu_3}\ \ \mathbf{E}_{4,\nu_4})$, $\dim(\mathbf{E}_{j,\nu_j}) = \nu\times\nu_j$ *and* $\boldsymbol{\Omega}_{\boldsymbol{\theta}}$ *is given in Theorem 28.*
Proof: [Independent Proof]. *By Theorem 28 and Delta Method Theorem 25, it is readily shown that*

$$\sqrt{n}(\widehat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \mathbf{D}^{(1)}_{\boldsymbol{\lambda};\boldsymbol{\theta}'_\lambda}\mathbf{E}'_{1,\nu}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\mathbf{E}_{1,\nu}\mathbf{D}^{(1)\prime}_{\boldsymbol{\lambda};\boldsymbol{\theta}'_\lambda}),$$

because $\boldsymbol{\theta}_{\boldsymbol{\lambda}} = \mathbf{E}'_{1,\nu}\boldsymbol{\theta}$. □

B.30. Proof of Theorem 30

Theorem 30. [Asymptotic Distribution of $\sqrt{n}(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta})$][Original result]. *The asymptotic distribution of* $\sqrt{n}(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta})$ *is*

$$\sqrt{n}(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\mathbf{E}'_{2,\nu}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\mathbf{E}_{2,\nu}\mathbf{D}^{(1)\prime}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}),$$

where $\mathbf{E}_{2,\nu_2}$ *is the* 2nd *submatrix of* $\mathbf{I}_\nu = (\mathbf{E}_{1,\nu_1}\ \ \mathbf{E}_{2,\nu_2}\ \ \mathbf{E}_{3,\nu_3}\ \ \mathbf{E}_{4,\nu_4})$, $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}$ *is given in Theorem 2,* $\dim(\mathbf{E}_{j,\nu_j}) = \nu\times\nu_j$ *and* $\boldsymbol{\Omega}_{\boldsymbol{\theta}}$ *is given in Theorem 28.*

Proof: [Independent Proof]. *By Theorem 28 and Delta Method Theorem 25, it is readily shown that*

$$\sqrt{n}(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\mathbf{E}'_{2,\nu}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\mathbf{E}_{2,\nu}\mathbf{D}^{(1)'}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}),$$

because $\boldsymbol{\theta}_{\boldsymbol{\delta}} = \mathbf{E}'_{2,\nu}\boldsymbol{\theta}$. □

B.31. Proof of Theorem 31

Theorem 31. [Asymptotic Distribution of $\sqrt{n}\,\mathrm{vec}(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})$][Original result]. *The asymptotic distribution of* $\sqrt{n}\,\mathrm{vec}(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})$ *is*

$$\sqrt{n}\,\mathrm{vec}(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \mathbf{A}_{\boldsymbol{\theta}_{\delta\gamma}}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\mathbf{A}'_{\boldsymbol{\theta}_{\delta\gamma}}),$$

where $\mathbf{A}_{\boldsymbol{\theta}_{\delta\gamma}} = (\mathbf{I}_q \otimes \boldsymbol{\Gamma})\left(\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}\mathbf{E}'_{2,\nu_2} + \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}\mathbf{E}'_{3,\nu_3}\right)$, $\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta}$ *is given in Theorem 2,* $\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma}$ *is given in Theorem 9,* $\mathbf{E}_{3,\nu_3}$ *is the* 3^{rd} *submatrix of* $\mathbf{I}_\nu = (\mathbf{E}_{1,\nu_1}\ \ \mathbf{E}_{2,\nu_2}\ \ \mathbf{E}_{3,\nu_3}\ \ \mathbf{E}_{4,\nu_4})$, $\dim(\mathbf{E}_{j,\nu_j}) = \nu \times \nu_j$, *and* $\boldsymbol{\Omega}_{\boldsymbol{\theta}}$ *is given in Theorem 28.*

Proof: [Independent Proof]. *Recall that in Chapter 2,* $\mathbf{G}(\boldsymbol{\theta}_\delta, \boldsymbol{\theta}_\gamma) = \boldsymbol{\Gamma}'_0\boldsymbol{\Gamma}$, *where* $\boldsymbol{\Gamma}_0$ *is in a small neighborhood of* $\boldsymbol{\Gamma}$. *Let* $\widehat{\mathbf{G}}$ *be the estimator of* $\mathbf{G}$ *such that* $\widehat{\boldsymbol{\Gamma}} = \boldsymbol{\Gamma}\widehat{\mathbf{G}}$, *where* $\widehat{\boldsymbol{\Gamma}}$ *minimizes* $F(\boldsymbol{\Sigma}(\boldsymbol{\theta}), \mathbf{S})$. *For notational convenience, denote* $(\boldsymbol{\theta}'_{\boldsymbol{\delta}}\ \ \mathbf{0}')'$ *by* $\boldsymbol{\theta}_{\delta\gamma}$ *and* $\left(\widehat{\boldsymbol{\theta}}'_{\boldsymbol{\delta}}\ \ \widehat{\boldsymbol{\theta}}'_\gamma\right)'$ *by* $\widehat{\boldsymbol{\theta}}_{\delta\gamma}$, *where* $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta}} = \mathbf{E}'_{2,\nu_2}\hat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{\theta}}_\gamma = \mathbf{E}'_{3,\nu_3}\hat{\boldsymbol{\theta}}$, *and* $\hat{\boldsymbol{\theta}}$ *is the minimizer of* $F(\boldsymbol{\Sigma}(\boldsymbol{\theta}), \mathbf{S})$. *Note that* $\mathrm{vec}(\widehat{\boldsymbol{\Gamma}}) = \mathrm{vec}(\boldsymbol{\Gamma}\widehat{\mathbf{G}}) = (\mathbf{I}_q \otimes \boldsymbol{\Gamma})\,\mathrm{vec}(\widehat{\mathbf{G}}) = (\mathbf{I}_q \otimes \boldsymbol{\Gamma})\mathbf{g}(\widehat{\boldsymbol{\theta}}_{\delta\gamma})$, *where* $\mathbf{g}(\widehat{\boldsymbol{\theta}}_{\delta\gamma}) = \mathrm{vec}(\widehat{\mathbf{G}})$.

The Taylor series expansion of $\mathbf{g}(\boldsymbol{\theta}_{\delta\gamma})$ *about* $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta}} = \boldsymbol{\theta}_{\boldsymbol{\delta}}$ *and* $\widehat{\boldsymbol{\theta}}_\gamma = \mathbf{0}$ *is*

$$\begin{aligned}\mathbf{g}(\widehat{\boldsymbol{\theta}}_{\delta\gamma}) &= \mathbf{g}(\boldsymbol{\theta}_{\delta\gamma}) + \frac{\partial \mathbf{g}}{\partial\,\boldsymbol{\theta}'_{\delta\gamma}}\left(\widehat{\boldsymbol{\theta}}_{\delta\gamma} - \boldsymbol{\theta}_{\delta\gamma}\right) \\ &\quad + \frac{1}{2}\frac{\partial^2 \mathbf{g}}{\partial\,\boldsymbol{\theta}'_{\delta\gamma_1} \otimes \partial\,\boldsymbol{\theta}'_{\delta\gamma_1}}\left[\left(\widehat{\boldsymbol{\theta}}_{\delta\gamma} - \boldsymbol{\theta}_{\delta\gamma}\right)\right] \otimes \left[\left(\widehat{\boldsymbol{\theta}}_{\delta\gamma} - \boldsymbol{\theta}_{\delta\gamma}\right)\right],\end{aligned} \tag{213}$$

where $\partial^2\mathbf{g}/\left(\partial\,\boldsymbol{\theta}'_{\delta\gamma_1} \otimes \partial\,\boldsymbol{\theta}'_{\delta\gamma_1}\right)$ *is* $\partial^2\mathbf{g}/\left(\partial\,\boldsymbol{\theta}'_{\delta\gamma} \otimes \partial\,\boldsymbol{\theta}'_{\delta\gamma}\right)$ *evaluated at* $\boldsymbol{\theta}_{\delta\gamma_1} = \alpha\widehat{\boldsymbol{\theta}}_{\delta\gamma} + (1 - \alpha)\boldsymbol{\theta}_{\delta\gamma}$ *for some* $\alpha \in (0, 1)$.

Examine $\partial^2\mathbf{g}/\left(\partial\,\boldsymbol{\theta}'_{\boldsymbol{\delta\gamma}}\otimes\partial\,\boldsymbol{\theta}'_{\boldsymbol{\delta\gamma}}\right)$ *as follows:*

$$\begin{aligned}\frac{\partial^2\mathbf{g}}{\partial\,\boldsymbol{\theta}'_{\boldsymbol{\delta\gamma}}\otimes\partial\,\boldsymbol{\theta}'_{\boldsymbol{\delta\gamma}}} &= \frac{\partial}{\partial\,\boldsymbol{\theta}'_{\boldsymbol{\delta\gamma}}}\left[\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta}}\quad\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\gamma}}\right]\\ &= \left[\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta},\boldsymbol{\theta}'_{\delta}}\quad\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta},\boldsymbol{\theta}'_{\gamma}}\quad\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_{\gamma},\boldsymbol{\theta}'_{\delta}}\quad\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_{\gamma},\boldsymbol{\theta}'_{\gamma}}\right] = A\ Matrix\ of\ Constants\end{aligned}\tag{214}$$

where the expressions for $\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta},\boldsymbol{\theta}'_{\delta}}$, $\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta},\boldsymbol{\theta}'_{\gamma}}$ *and* $\mathbf{D}^{(2)}_{\mathbf{g};\boldsymbol{\theta}'_{\gamma},\boldsymbol{\theta}'_{\gamma}}$ *are given in Theorem 9 and Theorem 10.*

Note that $\boldsymbol{\theta}_{\boldsymbol{\delta\gamma}_1} = \alpha\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta\gamma}} + (1-\alpha)\boldsymbol{\theta}_{\boldsymbol{\delta\gamma}} = \alpha(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta\gamma}} - \boldsymbol{\theta}_{\boldsymbol{\delta\gamma}}) + \boldsymbol{\theta}_{\boldsymbol{\delta\gamma}} \xrightarrow{\text{dist}} \boldsymbol{\theta}_{\boldsymbol{\delta\gamma}}$ *by Slutsky's Theorem 23 because* $\alpha\in(0,1)$ *and* $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta\gamma}} - \boldsymbol{\theta}_{\boldsymbol{\delta\gamma}} = o_p(1)$. *Further, by Theorem 24,* $\partial^2\mathbf{g}/\left(\partial\,\boldsymbol{\theta}'_{\boldsymbol{\delta\gamma}_1}\otimes\partial\,\boldsymbol{\theta}'_{\boldsymbol{\delta\gamma}_1}\right) \xrightarrow{\text{dist}} \partial^2\mathbf{g}/\left(\partial\,\boldsymbol{\theta}'_{\boldsymbol{\delta\gamma}}\otimes\partial\,\boldsymbol{\theta}'_{\boldsymbol{\delta\gamma}}\right)$ *because* $\boldsymbol{\theta}'_{\boldsymbol{\delta\gamma}_1} \xrightarrow{\text{dist}} \boldsymbol{\theta}'_{\boldsymbol{\delta\gamma}}$. *Accordingly, it can be concluded that* $\partial^2\mathbf{g}/\left(\partial\,\boldsymbol{\theta}'_{\boldsymbol{\delta\gamma}_1}\otimes\partial\,\boldsymbol{\theta}'_{\boldsymbol{\delta\gamma}_1}\right) = O_p(1)$.

Further,

$$\begin{aligned}&\sqrt{n}\,\mathrm{vec}(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\\
=\ &(\mathbf{I}_q\otimes\boldsymbol{\Gamma})\sqrt{n}\left[\mathbf{g}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta\gamma}})-\mathbf{g}(\boldsymbol{\theta}_{\boldsymbol{\delta\gamma}})\right]\\
=\ &(\mathbf{I}_q\otimes\boldsymbol{\Gamma})\left[\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta}}\quad\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\gamma}}\right]\sqrt{n}\begin{pmatrix}\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}-\boldsymbol{\theta}_{\boldsymbol{\delta}}\\ \widehat{\boldsymbol{\theta}}_{\boldsymbol{\gamma}}-\mathbf{0}\end{pmatrix}\\
+\ &\frac{(\mathbf{I}_q\otimes\boldsymbol{\Gamma})}{2\sqrt{n}}\frac{\partial^2\mathbf{g}}{\partial\,\boldsymbol{\theta}'_{\boldsymbol{\delta\gamma}_1}\otimes\partial\,\boldsymbol{\theta}'_{\boldsymbol{\delta\gamma}_1}}\left[\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta\gamma}}-\boldsymbol{\theta}_{\boldsymbol{\delta\gamma}}\right)\right]\otimes\left[\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta\gamma}}-\boldsymbol{\theta}_{\boldsymbol{\delta\gamma}}\right)\right]\\
=\ &(\mathbf{I}_q\otimes\boldsymbol{\Gamma})\left[\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta}}\quad\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\gamma}}\right]\sqrt{n}\begin{pmatrix}\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}-\boldsymbol{\theta}_{\boldsymbol{\delta}}\\ \widehat{\boldsymbol{\theta}}_{\boldsymbol{\gamma}}-\mathbf{0}\end{pmatrix}+O_p(n^{-\frac{1}{2}})\\
&\textit{because}\quad\partial^2\mathbf{g}/\left(\partial\,\boldsymbol{\theta}'_{\boldsymbol{\delta\gamma}_1}\otimes\partial\,\boldsymbol{\theta}'_{\boldsymbol{\delta\gamma}_1}\right)=O_p(1)\quad\textit{and}\quad\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta\gamma}}-\boldsymbol{\theta}_{\boldsymbol{\delta\gamma}}\right)=O_p(1)\\
=\ &(\mathbf{I}_q\otimes\boldsymbol{\Gamma})\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta}}\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}-\boldsymbol{\theta}_{\boldsymbol{\delta}})+(\mathbf{I}_q\otimes\boldsymbol{\Gamma})\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\gamma}}\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\gamma}}-\mathbf{0})+O_p(n^{-\frac{1}{2}})\\
=\ &(\mathbf{I}_q\otimes\boldsymbol{\Gamma})\left[\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta}}\mathbf{E}'_{2,\nu_2}+\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\gamma}}\mathbf{E}'_{3,\nu_3}\right]\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})+O_p(n^{-\frac{1}{2}}),\end{aligned}$$

where $\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\gamma}}$ *is given in Theorem 9 and* $\mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_{\delta}}$ *is given in Theorem 10.*

By Slutsky's Theorem 23, Delta Method Theorem 25, and Theorem 28, it is readily shown that

$$\sqrt{n}\,\mathrm{vec}(\widehat{\mathbf{\Gamma}} - \mathbf{\Gamma}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \mathbf{A}_{\boldsymbol{\theta}_{\boldsymbol{\delta\gamma}}} \mathbf{\Omega}_{\boldsymbol{\theta}} \mathbf{A}'_{\boldsymbol{\theta}_{\boldsymbol{\delta\gamma}}}),$$

where $\mathbf{A}_{\boldsymbol{\theta}_{\boldsymbol{\delta\gamma}}} = (\mathbf{I}_q \otimes \mathbf{\Gamma}) \left(\mathbf{D}^{(1)}_{\boldsymbol{\delta};\boldsymbol{\theta}'_\delta} \mathbf{E}'_{2,\nu_2} + \mathbf{D}^{(1)}_{\mathbf{g};\boldsymbol{\theta}'_\gamma} \mathbf{E}'_{3,\nu_3} \right)$ *and* $\mathbf{\Omega}_{\boldsymbol{\theta}}$ *is given in Theorem 28.* $\square$

B.32. Proof of Theorem 32

Theorem 32. [Asymptotic Distribution of $\sqrt{n}(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi})$][Original result]. *The asymptotic distribution of* $\sqrt{n}(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi})$ *is*

$$\sqrt{n}(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \mathbf{D}^{(1)}_{\boldsymbol{\psi};\boldsymbol{\theta}'_\psi} \mathbf{E}'_{4,\nu} \mathbf{\Omega}_{\boldsymbol{\theta}} \mathbf{E}_{4,\nu} \mathbf{D}^{(1)\prime}_{\boldsymbol{\psi};\boldsymbol{\theta}'_\psi}),$$

where $\mathbf{E}_{4,\nu_1}$ *is the* 4^{th} *submatrix of* $\mathbf{I}_\nu = (\mathbf{E}_{1,\nu_1} \;\; \mathbf{E}_{2,\nu_2} \;\; \mathbf{E}_{3,\nu_3} \;\; \mathbf{E}_{4,\nu_4})$, $\mathbf{D}^{(1)}_{\boldsymbol{\psi};\boldsymbol{\theta}'_\psi}$ *is given in Theorem 11,* $\dim(\mathbf{E}_{j,\nu_j}) = \nu \times \nu_j$, *and* $\mathbf{\Omega}_{\boldsymbol{\theta}}$ *is given in Theorem 28.*

Proof: [Independent Proof]. *By Theorem 28 and Delta Method Theorem 25, it is readily shown that*

$$\sqrt{n}(\widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0}, \mathbf{D}^{(1)}_{\boldsymbol{\psi};\boldsymbol{\theta}'_\psi} \mathbf{E}'_{4,\nu} \mathbf{\Omega}_{\boldsymbol{\theta}} \mathbf{E}_{4,\nu} \mathbf{D}^{(1)\prime}_{\boldsymbol{\psi};\boldsymbol{\theta}'_\psi}),$$

because $\boldsymbol{\theta}_{\boldsymbol{\psi}} = \mathbf{E}'_{4,\nu}\boldsymbol{\theta}$. $\square$

B.33. Proof of Theorem 33

Theorem 33. [Adapted from Boik, Panishkan, and Hyde [35], 2010, Page 131]. *Consider a test of* $\mathbf{H}_0 : \boldsymbol{\theta} \in \mathbf{\Theta}_0$ *against* $\mathbf{H}_a : \boldsymbol{\theta} \in \mathbf{\Theta}_a$, *where* $\mathbf{\Theta}_0 \cap \mathbf{\Theta}_a = \emptyset$. *Suppose that under* $\mathbf{H}_0$, *constraints are placed on* $\boldsymbol{\theta} \in \mathbf{\Theta}$, *where* $\mathbf{\Theta} = \mathbf{\Theta}_0 \cup \mathbf{\Theta}_a$. *If* $\mathbf{H}_0$ *is true, then, by construction, there exist a value* $\boldsymbol{\theta}_0 \in \mathbf{\Theta}_0$ *and a value* $\boldsymbol{\theta}_a \in \mathbf{\Theta}$ *so that* $\mathbf{\Sigma}(\boldsymbol{\theta}_0) = \mathbf{\Sigma}(\boldsymbol{\theta}_a) = \mathbf{\Sigma}$ *and* $\boldsymbol{\theta}_a = g(\boldsymbol{\theta}_0)$. *Assume that*

(a.) $\hat{\boldsymbol{\theta}}_0$ *and* $\hat{\boldsymbol{\theta}}_a$ *are consistent roots of* $F(\mathbf{\Sigma}(\boldsymbol{\theta}), \mathbf{S})$, *where* $\hat{\boldsymbol{\theta}}_0$ *and* $\hat{\boldsymbol{\theta}}_a$ *are defined in (153);*

(b.) $\operatorname{rank}(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_0}) = \nu_0$ *and* $\operatorname{rank}(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}) = \nu_a$, *where* $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_0} = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$, $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a} = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_a}$, *and* $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}$ *is given in (48); and*

(c.) The p^* *distinct elements in* $\mathbf{z}_i\mathbf{z}_i'$ *have a positive definite covariance matrix. That is,* $\operatorname{Var}[\operatorname{vech}(\mathbf{z}_i\mathbf{z}_i')]$ *is positive definite, where* $\operatorname{vech}(\mathbf{z}_i\mathbf{z}_i')$ *is defined in Table 56.*

Define p^*, $\mathbf{H}_p$, $\mathbf{A}_p$, $\mathbf{P}_0$, *and* $\mathbf{P}_a$ *as follows:*

$$
\begin{aligned}
&p^* = p(p+1)/2, \quad \mathbf{H}_p = (\mathbf{D}_p'\mathbf{D}_p)^{-1}\mathbf{D}_p', \quad \mathbf{A}_p = \mathbf{D}_p'\boldsymbol{\Omega}\mathbf{D}_p,\\
&\mathbf{P}_0 = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_0}\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_0}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_0}\right]^{-1}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_0}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right), \text{ and}\\
&\mathbf{P}_a = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}\right]^{-1}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right),
\end{aligned}
$$

where $\mathbf{D}_p$ *is a* $p^2\times p^*$ *duplication matrix defined in Table 56, and* $\boldsymbol{\Omega} = \operatorname{E}\left[\mathbf{z}_i\mathbf{z}_i'\otimes\mathbf{z}_i\mathbf{z}_i'\right] - \boldsymbol{\sigma}\boldsymbol{\sigma}'$ *is given in Theorem 27.*

Then,

(1.) Define ν_d *as* $\nu_d = \nu_a - \nu_0$, *where* $\nu_0 = \dim(\boldsymbol{\Theta}_0)$ *and* $\nu_a = \dim(\boldsymbol{\Theta})$. $\mathbf{P}_a - \mathbf{P}_0$ *is a projection operator and* $\operatorname{rank}(\mathbf{P}_a - \mathbf{P}_0) = \nu_d$.

(2.) $\mathbf{A}_p$ *is positive definite and can be diagonalized as* $\mathbf{A}_p = \mathbf{U}_p\mathbf{V}_p\mathbf{U}_p'$, *where* $\mathbf{U}_p \in \mathcal{O}(p^*)$, $\mathbf{V}_p = \operatorname{Diag}(v_1, v_2, \ldots, v_{p^*})$, *and* $v_i > 0$ *for* $i = 1, 2, \ldots, p^*$. *Define* $\mathbf{V}_p^{1/2}$ *as* $\mathbf{V}_p^{1/2} = \operatorname{Diag}(\sqrt{v_1}, \sqrt{v_2}, \ldots, \sqrt{v_{p^*}})$ *and* $\mathbf{A}_p^{1/2}$ *as* $\mathbf{A}_p^{1/2} = \mathbf{U}_p\mathbf{V}_p^{1/2}\mathbf{U}_p'$. *Then,* $\mathbf{A}_p^{1/2}$ *is positive definite.*

(3.) Define $\mathbf{A}$ *as* $\mathbf{A} = \mathbf{A}_p^{1/2}\mathbf{H}_p\left[\frac{1}{2}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)(\mathbf{P}_a - \mathbf{P}_0)\right]\mathbf{H}_p'\mathbf{A}_p^{1/2}$. *The nonzero eigenvalues of* $\mathbf{A}$ *is the same as the nonzero eigenvalues of* $\boldsymbol{\Omega}\left[\frac{1}{2}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)(\mathbf{P}_a - \mathbf{P}_0)\right]$ *and* $\mathbf{A}$ *can be diagonalized as* $\mathbf{A} = \mathbf{U}\mathbf{V}\mathbf{U}'$, *where* $\mathbf{U}$ *is a* $p^*\times\nu_d$ *semi-orthogonal matrix,* $\{v_i\}_{i=1}^{\nu_d}$ *are the nonzero eigenvalues of* $\mathbf{A}$, *and* $\mathbf{V} = \operatorname{Diag}(v_1, \ldots, v_{\nu_d})$.

(4.) Given $\mathbf{H}_0$ *is true,* $\mathbf{X}^2 = n\left[F(\hat{\boldsymbol{\Sigma}}_0, \mathbf{S}) - F(\hat{\boldsymbol{\Sigma}}_a, \mathbf{S})\right]$ *can be written as*

$$\mathbf{X}^2 = \frac{1}{2}\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)(\mathbf{P}_a - \mathbf{P}_0)\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) + O_p(n^{-\frac{1}{2}}).$$

(5.) Given $\mathbf{H}_0$ *is true,* $\mathbf{X}^2 \xrightarrow{\text{dist}} \sum_{i=1}^{\nu_d} v_i w_i^2$*, where* $w_i^2 \overset{\text{iid}}{\sim} \chi^2_1$ *for* $i = 1, \ldots, \nu_d$*.*

Proof: [Proof With Advisor's Help].

(1.) Note that

$$\mathbf{P}_a = \mathbf{P}_a^2, \quad \mathbf{P}_0 = \mathbf{P}_0^2, \quad \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'} = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\left(\frac{\partial\,\boldsymbol{\theta}_a}{\partial\,\boldsymbol{\theta}_0'}\right)$$

because $\boldsymbol{\theta}_a = g(\boldsymbol{\theta}_0)$*, and*

$$\begin{aligned}
&\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\right]^{-1}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\\
=\;&\left(\frac{\partial\,\boldsymbol{\theta}_a'}{\partial\,\boldsymbol{\theta}_0}\right)\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\right]^{-1}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\\
=\;&\left(\frac{\partial\,\boldsymbol{\theta}_a'}{\partial\,\boldsymbol{\theta}_0}\right)\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\\
=\;&\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'}.
\end{aligned} \tag{215}$$

Based on (215), it can be concluded that $\mathbf{P}_0\mathbf{P}_a = \mathbf{P}_0$*. Further,* $\mathbf{P}_a\mathbf{P}_0 = \mathbf{P}_0$ *because* $\mathbf{P}_0 = \mathbf{P}_0^2$ *and* $\mathbf{P}_0(\mathbf{P}_a\mathbf{P}_0 - \mathbf{P}_0) = \mathbf{0}$*.*

Accordingly,

$$(\mathbf{P}_a - \mathbf{P}_0)^2 = \mathbf{P}_a + \mathbf{P}_0 - \mathbf{P}_0\mathbf{P}_a - \mathbf{P}_a\mathbf{P}_0 = \mathbf{P}_a - \mathbf{P}_0. \tag{216}$$

That is, $\mathbf{P}_a - \mathbf{P}_0$ *is a projection operator.*

By 4.5.12 on page 220 in Meyer [43], $\operatorname{rank}(\mathbf{P}_0) = \operatorname{rank}(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'}) = \nu_0$ *and* $\operatorname{rank}(\mathbf{P}_a) = \operatorname{rank}(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}) = \nu_a$*. Note that* $\mathbf{P}_a$ *and* $\mathbf{P}_0$ *are projection operators because* $\mathbf{P}_a = \mathbf{P}_a^2$ *and* $\mathbf{P}_0 = \mathbf{P}_0^2$*. By Theorem 26,*

$$\operatorname{tr}(\mathbf{P}_0) = \operatorname{rank}(\mathbf{P}_0) = \nu_0 \quad \textit{and} \quad \operatorname{tr}(\mathbf{P}_a) = \operatorname{rank}(\mathbf{P}_a) = \nu_a. \tag{217}$$

Further, $\operatorname{rank}(\mathbf{P}_a - \mathbf{P}_0) = \operatorname{tr}(\mathbf{P}_a - \mathbf{P}_0) = \operatorname{tr}(\mathbf{P}_a) - \operatorname{tr}(\mathbf{P}_0) = \nu_a - \nu_0 = \nu_d$ *because* $\mathbf{P}_a - \mathbf{P}_0$ *is a projection operator.*

(2.) It is readily shown that $\operatorname{Var}[\operatorname{vech}(\mathbf{z}_i\mathbf{z}_i')] = \mathbf{H}_p \operatorname{Var}[\operatorname{vec}(\mathbf{z}_i\mathbf{z}_i')]\,\mathbf{H}_p' = \mathbf{H}_p\boldsymbol{\Omega}\mathbf{H}_p'$ *because* $\boldsymbol{\Omega} = \operatorname{Var}[\operatorname{vec}(\mathbf{z}_i\mathbf{z}_i')]$ *as shown in (198) of Theorem 27. By the assumption that* $\operatorname{Var}[\operatorname{vech}(\mathbf{z}_i\mathbf{z}_i')]$ *is positive definite,* $\mathbf{H}_p\boldsymbol{\Omega}\mathbf{H}_p'$ *is positive definite.*

Let $\mathbf{t}$ *be any* p^* *vector, note that* $\mathbf{A}_p = \mathbf{A}_p'$*, where* $\mathbf{A}_p = \mathbf{D}_p'\boldsymbol{\Omega}\mathbf{D}_p$*, and*

$$\begin{aligned} \mathbf{t}'\mathbf{A}_p\mathbf{t} &= \mathbf{t}'(\mathbf{D}_p'\mathbf{D}_p)(\mathbf{D}_p'\mathbf{D}_p)^{-1}\mathbf{A}_p(\mathbf{D}_p'\mathbf{D}_p)^{-1}(\mathbf{D}_p'\mathbf{D}_p)\mathbf{t} \\ &= \mathbf{t}'(\mathbf{D}_p'\mathbf{D}_p)\left[\mathbf{H}_p\boldsymbol{\Omega}\mathbf{H}_p'\right](\mathbf{D}_p'\mathbf{D}_p)\mathbf{t} > 0 \end{aligned} \tag{218}$$

for all $\mathbf{t} \neq \mathbf{0}$*, because* $\mathbf{H}_p\boldsymbol{\Omega}\mathbf{H}_p'$ *is positive definite. Thus,* $\mathbf{A}_p$ *is positive definite.*

By Diagonability Theorem on page 284, Lemma 3 on page 307, and § 11.6 property (c) on page 290 in Searle [36], $\mathbf{A}_p$ *can be diagonalized as* $\mathbf{A}_p = \mathbf{U}_p\mathbf{V}_p\mathbf{U}_p'$*, where* $\mathbf{U}_p \in \mathcal{O}(p^*)$*,* $\mathbf{V}_p = \operatorname{Diag}(v_1, v_2, \ldots, v_{p^*})$*, and* $v_i > 0$ *for* $i = 1, 2, \ldots, p^*$*.*

By the definition of $\mathbf{A}_p^{1/2}$*, it can be concluded that* $\mathbf{A}_p^{1/2}$ *is positive definite because* $\mathbf{A}_p^{1/2} = \mathbf{A}_p^{1/2'}$ *and* $\sqrt{v_i} > 0$ *for* $i = 1, 2, \ldots, p^*$*.*

(3.) Denote $\frac{1}{2}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)(\mathbf{P}_a - \mathbf{P}_0)$ *by* $\mathbf{B}$ *and* $\mathbf{H}_p'\mathbf{A}_p^{1/2}$ *by* $\mathbf{C}$*. Accordingly,* $\mathbf{A}$ *can be written as* $\mathbf{A} = \mathbf{C}'\mathbf{B}\mathbf{C}$*. Note that* $\mathbf{N}_p\boldsymbol{\Omega} = \mathbf{N}_p \operatorname{E}[\mathbf{z}_i\mathbf{z}_i' \otimes \mathbf{z}_i\mathbf{z}_i'] - \mathbf{N}_p\boldsymbol{\sigma}\boldsymbol{\sigma}' = \boldsymbol{\Omega}$ *because* $\mathbf{N}_p(\mathbf{z}_i \otimes \mathbf{z}_i) = \mathbf{z}_i \otimes \mathbf{z}_i$ *and* $\mathbf{K}_{p,p}\boldsymbol{\sigma} = \boldsymbol{\sigma}$*. Further,*

$$\mathbf{C}'\mathbf{C}\mathbf{B} = \mathbf{H}_p'\mathbf{D}_p'\boldsymbol{\Omega}\mathbf{D}_p\mathbf{H}_p\mathbf{B} = \mathbf{N}_p\boldsymbol{\Omega}\mathbf{N}_p\mathbf{B} = \boldsymbol{\Omega}\mathbf{N}_p\mathbf{B} = (\mathbf{N}_p\boldsymbol{\Omega})'\,\mathbf{B} = \boldsymbol{\Omega}\mathbf{B}.$$

Accordingly, the nonzero eigenvalues of $\mathbf{A}$ *are the same as the nonzero eigenvalues of* $\boldsymbol{\Omega}\mathbf{B}$ *because* $\mathbf{A} = \mathbf{C}'\mathbf{B}\mathbf{C}$ *and* $\boldsymbol{\Omega}\mathbf{B} = \mathbf{C}'\mathbf{C}\mathbf{B}$*.*

Examine

$$\operatorname{rank}(\mathbf{A}) \;=\; \operatorname{rank}\left[\mathbf{H}_p\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)(\mathbf{P}_a - \mathbf{P}_0)\mathbf{H}_p\right]$$

$$
\begin{aligned}
&= \operatorname{rank}[\mathbf{D}_p' \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) (\mathbf{P}_a - \mathbf{P}_0)\mathbf{D}_p] \\
&\geq \operatorname{rank}[\mathbf{N}_p \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) (\mathbf{P}_a - \mathbf{P}_0)\mathbf{N}_p] \\
&\quad \textit{because } \mathbf{N}_p = \mathbf{D}_p(\mathbf{D}_p'\mathbf{D}_p)^{-1}\mathbf{D}_p' \\
&\quad \textit{and Theorem 2.10. (a) in Schott [44] says} \operatorname{rank}(\mathbf{A}) \geq \operatorname{rank}(\mathbf{AB}) \\
&= \operatorname{rank}[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{N}_p(\mathbf{P}_a - \mathbf{P}_0)\mathbf{N}_p] \\
&\quad \textit{because } \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{N}_p = \mathbf{N}_p \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \\
&= \operatorname{rank}[\mathbf{N}_p(\mathbf{P}_a - \mathbf{P}_0)\mathbf{N}_p] \quad \textit{because } \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \textit{ has full rank} \\
&= \operatorname{rank}(\mathbf{P}_a - \mathbf{P}_0),
\end{aligned}
$$

because $\mathbf{N}_p\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}$ *and* $(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\,\mathbf{N}_p = \mathbf{N}_p\,(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})$.

Further, it can be concluded that $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{P}_a - \mathbf{P}_0) = \nu_d$ *because* $\operatorname{rank}(\mathbf{A}) \leq \operatorname{rank}(\mathbf{P}_a - \mathbf{P}_0)$ *and* $\operatorname{rank}(\mathbf{A}) \geq \operatorname{rank}(\mathbf{P}_a - \mathbf{P}_0)$.

By Theorem 4.9 in Schott [44], the number of nonzero eigenvalues of $\mathbf{A}$ *is* $\operatorname{rank}(\mathbf{A}) = \nu_d$, *because* $\mathbf{A} = \mathbf{A}'$ *is diagonalizable. Further,* $\mathbf{A}$ *can be written as*

$$
\mathbf{A} = \mathbf{U}\mathbf{V}\mathbf{U}', \tag{219}
$$

where $\mathbf{U}$ *is a* $p^* \times \nu_d$ *semi-orthogonal matrix,* $\{v_i\}_{i=1}^{\nu_d}$ *are the nonzero eigenvalues of* $\mathbf{A}$, *and* $\mathbf{V} = \operatorname{Diag}(v_1, \ldots, v_{\nu_d})$.

(4.) For notational convenience, rewrite $F(\hat{\boldsymbol{\Sigma}}_0, \mathbf{S})$ *and* $F(\hat{\boldsymbol{\Sigma}}_a, \mathbf{S})$ *in (155) as* $\mathbf{F}(\hat{\boldsymbol{\theta}}_0)$ *and* $\mathbf{F}(\hat{\boldsymbol{\theta}}_a)$, *respectively. Define* $\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}_0}$ *and* $\mathbf{D}^{(2)}_{F;\boldsymbol{\theta}_0,\boldsymbol{\theta}_0'}$ *as follows:*

$$
\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}_0} = \mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \quad \textit{and} \quad \mathbf{D}^{(2)}_{F;\boldsymbol{\theta}_0,\boldsymbol{\theta}_0'} = \mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},
$$

where $\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}$ *and* $\mathbf{D}^{(2)}_{F;\boldsymbol{\theta},\boldsymbol{\theta}'}$ *are defined in Theorem 17. Given that* $\mathbf{H}_0$ *is true, then, by (205), (208) and (210) in Theorem 28, it can be concluded that*

$$\begin{aligned}
&\sqrt{n}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0) \\
&= -[\mathrm{E}(\mathbf{D}^{(2)}_{F;\boldsymbol{\theta_0},\boldsymbol{\theta_0'}})]^{-1}\left[\sqrt{n}\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}_0}\right] + O_p(n^{-\frac{1}{2}}) \\
&= \left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta_0'}}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta_0'}}\right]^{-1}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta_0'}}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) \\
&+ O_p(n^{-\frac{1}{2}}).
\end{aligned} \tag{220}$$

Given that $\mathbf{H}_0$ *is true, expand* $\mathbf{F}(\hat{\boldsymbol{\theta}}_0)$ *in a Taylor series around* $\hat{\boldsymbol{\theta}}_0 = \boldsymbol{\theta}_0$ *as follows:*

$$\begin{aligned}
\mathbf{F}(\hat{\boldsymbol{\theta}}_0) &= \mathbf{F}(\boldsymbol{\theta}_0) + (\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0)'\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}_0} + \frac{1}{2}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0)'\mathbf{D}^{(2)}_{F;\boldsymbol{\theta_0},\boldsymbol{\theta_0'}}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0) \\
&+ \frac{1}{3!}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0)'\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_1}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0)^{\otimes 2} \\
&= \mathbf{F}(\boldsymbol{\theta}_0) - \frac{1}{n}[\sqrt{n}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0)']\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta_0'}}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})\right] \\
&+ \frac{1}{2}[\sqrt{n}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0)']\mathbf{D}^{(2)}_{F;\boldsymbol{\theta_0},\boldsymbol{\theta_0'}}[\sqrt{n}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0)] + O_p(n^{-\frac{3}{2}}) \\
&= \mathbf{F}(\boldsymbol{\theta}_0) - \frac{1}{n}\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{P}_0\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) \\
&+ \frac{1}{2n}[\sqrt{n}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0)']\mathbf{D}^{(2)}_{F;\boldsymbol{\theta_0},\boldsymbol{\theta_0'}}[\sqrt{n}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0)] + O_p(n^{-\frac{3}{2}}) \\
&= \mathbf{F}(\boldsymbol{\theta}_0) - \frac{1}{n}\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{P}_0\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) \\
&+ \frac{1}{2n}[\sqrt{n}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0)']\,\mathrm{E}(\mathbf{D}^{(2)}_{F;\boldsymbol{\theta_0},\boldsymbol{\theta_0'}})[\sqrt{n}(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0)] + O_p(n^{-\frac{3}{2}}) \quad \textit{by (210)} \\
&= \mathbf{F}(\boldsymbol{\theta}_0) - \frac{1}{n}\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{P}_0\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) \\
&+ \frac{1}{2n}\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{P}_0\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) + O_p(n^{-\frac{3}{2}}) \quad \textit{by (208)} \\
&= \mathbf{F}(\boldsymbol{\theta}_0) \\
&- \frac{1}{2n}\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{P}_0\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) \\
&+ O_p(n^{-\frac{3}{2}}),
\end{aligned} \tag{221}$$

where $\mathbf{D}^{(3)}_{F;\boldsymbol{\theta}',\boldsymbol{\theta}',\boldsymbol{\theta}}$ *is given in Theorem 17,* $\boldsymbol{\theta}_1 = \alpha\hat{\boldsymbol{\theta}}_0 + (1-\alpha)\boldsymbol{\theta}_0$*, and* $(\hat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0)^{\otimes 2}$ *is defined in Table 56.*

Given that $\mathbf{H}_0$ *is true,* $\mathbf{F}(\boldsymbol{\theta}_0) = \mathbf{F}(\boldsymbol{\theta}_a)$ *because* $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_a) = \boldsymbol{\Sigma}$*. The expansion of* $\mathbf{F}(\hat{\boldsymbol{\theta}}_a)$ *can be conducted in a similar manner as shown above and the expression for* $\mathbf{F}(\hat{\boldsymbol{\theta}}_a)$ *have the same form as those under* $\mathbf{H}_0$*, except that* $\mathbf{P}_a$ *is substituted for* $\mathbf{P}_0$*. That is,*

$$\mathbf{F}(\hat{\boldsymbol{\theta}}_a) = \mathbf{F}(\boldsymbol{\theta}_0) - \frac{1}{2n}\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{P}_a\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) + O_p(n^{-\frac{3}{2}}). \quad (222)$$

Further, based on (221) and (222), given that $\mathbf{H}_0$ *is true,*
$\mathbf{X}^2 = n\left[\mathbf{F}(\hat{\boldsymbol{\theta}}_0) - \mathbf{F}(\hat{\boldsymbol{\theta}}_a)\right]$ *can be written as*

$$\mathbf{X}^2 = \frac{1}{2}\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})'\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)(\mathbf{P}_a - \mathbf{P}_0)\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) + O_p(n^{-\frac{1}{2}}). \quad (223)$$

(5.) The proof for part (5.) is adapted from the proof of Theorem 5 on page 9 in the supplement of Boik [34].
Define $\mathbf{z}$ *as* $\mathbf{z} \stackrel{\text{def}}{=} \mathbf{A}_p^{-1/2}\mathbf{D}_p'\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})$*, where* $\mathbf{z}$ *is a* $p^*\times 1$ *random vector.It can be concluded that* $\mathbf{D}_p\mathbf{H}_p\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) = \sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})$*. Further,*

$$\begin{aligned}\mathbf{H}_p'\mathbf{A}_p^{1/2}\mathbf{z} &= \mathbf{H}_p'\mathbf{D}_p'\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) = \mathbf{D}_p\mathbf{H}_p\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) \\ &= \mathbf{N}_p\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) = \sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}),\end{aligned} \quad (224)$$

because $\mathbf{D}_p\mathbf{H}_p = \mathbf{N}_p$ *and* $\mathbf{K}_{p,p}(\mathbf{s}-\boldsymbol{\sigma}) = \mathbf{s}-\boldsymbol{\sigma}$*.*

Examine $\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})'\left[1/2\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)(\mathbf{P}_a - \mathbf{P}_0)\right]\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})$ *in* $\mathbf{X}^2$ *of (223) as follows:*

$$\begin{aligned}&\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})'\left[1/2\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)(\mathbf{P}_a - \mathbf{P}_0)\right]\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) \\ &= \mathbf{z}'\mathbf{A}_p^{1/2}\mathbf{H}_p\left[1/2\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)(\mathbf{P}_a - \mathbf{P}_0)\right]\mathbf{H}_p'\mathbf{A}_p^{1/2}\mathbf{z} \quad \textit{by (224)} \\ &= \mathbf{z}'\mathbf{A}\mathbf{z} \quad \textit{by the definition of } \mathbf{A} \textit{ in part (3.).}\end{aligned} \quad (225)$$

By Theorem 27, $\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) \xrightarrow{\text{dist}} \mathbf{N}(\mathbf{0},\boldsymbol{\Omega})$, *where* $\boldsymbol{\Omega} = \mathrm{E}\left[\mathbf{z}_i\mathbf{z}_i' \otimes \mathbf{z}_i\mathbf{z}_i'\right] - \boldsymbol{\sigma}\boldsymbol{\sigma}'$. *Denote* $\mathbf{U}'\mathbf{z}$ *by* $\mathbf{w}$, *where* $\mathbf{U}$ *is defined in part (3.). By Slutsky's Theorem 23, it can be concluded that* $\mathbf{w} \xrightarrow{\text{dist}} \mathbf{N}\left[\mathbf{0},\mathbf{I}_{\nu_d}\right]$ *because* $\mathbf{z} \xrightarrow{\text{dist}} \mathbf{N}\left(\mathbf{0},\mathbf{I}_{p^*}\right)$. *Accordingly, it is readily shown that*

$$f(\mathbf{w}) \xrightarrow{\text{dist}} \prod_{i=1}^{\nu_d} f(w_i), \tag{226}$$

where $f(\mathbf{w})$ *is the probability density function of* $\mathbf{w}$, *and the probability density function of* w_i *is standard normal distribution,* $N[0,1]$, *for* $i=1,\ldots,\nu_d$.

Further, by definition 4.6.2. in Casella and Berger [3], $w_i \overset{\text{iid}}{\sim} N(0,1)$, *for* $i = 1,\ldots,\nu_d$. *By definition 3.1.3. in Muirhead [45] and Theorem 4.6.5. in Casella and Berger [3],* $w_i^2 \overset{\text{iid}}{\sim} \chi_1^2$ *for* $i=1,\ldots,\nu_d$.

By (219) and (225), given that $\mathbf{H}_0$ *is true,* $\mathbf{X}^2$ *in (223) can be simplified as*

$$\mathbf{X}^2 = \mathbf{z}'\mathbf{A}\mathbf{z} + O_p(n^{-1/2}) = \mathbf{z}'\mathbf{U}\mathbf{V}\mathbf{U}'\mathbf{z} + O_p(n^{-1/2}) = \mathbf{w}'\mathbf{V}\mathbf{w} + O_p(n^{-1/2}). \tag{227}$$

Further, by Slutsky's Theorem 23, (226) and (227), the null distribution of $\mathbf{X}^2$ *is*

$$\mathbf{X}^2 \xrightarrow{\text{dist}} \sum_{i=1}^{\nu_d} v_i w_i^2, \tag{228}$$

where $w_i^2 \overset{\text{iid}}{\sim} \chi_1^2$ *for* $i=1,\ldots,\nu_d$. □

B.33.1. Proof of corollary 33.1

Corollary 33.1. [Adapted from Boik, Panishkan, and Hyde [35], 2010, Page 131].*Consider a goodness-of-fit test of* $\mathbf{H}_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$. *Assume that the* $p \times p$ *sample covariance matrix* $\mathbf{S}$ *is invertible. The test statistic for the goodness-of-fit test is*

$$X^2 = nF(\hat{\boldsymbol{\Sigma}}_0,\mathbf{S}),$$

where $F(\boldsymbol{\Sigma}, \mathbf{S})$ *is defined in (66), and* $\hat{\boldsymbol{\Sigma}}_0$ *is defined in (153). Reject* $\mathbf{H}_0$*, for large values of* $\mathbf{X}^2$*. Furthermore, the asymptotic null distribution of* X^2 *is* $\mathbf{X}^2 \xrightarrow{\text{dist}} \sum_{i=1}^{\nu_d} v_i w_i^2$*, where* $\mathbf{X}^2 = \frac{1}{2}\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma})' \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) (\mathbf{I}_{p^2} - \mathbf{P}_0)\sqrt{n}(\mathbf{s} - \boldsymbol{\sigma}) + O_p(n^{-\frac{1}{2}})$*,* $\nu_d = p(p+1)/2 - \nu_0$*, and* $w_i^2 \stackrel{\text{iid}}{\sim} \chi_1^2$ *for* $i = 1, \ldots, \nu_d$*.*

Proof: [Proof With Advisor's Help]. *In the goodness-of-fit test, a saturated model for* $\boldsymbol{\Sigma}$ *is unconstrained. Accordingly, the number of parameters under a saturated model,* ν_a*, is* $p(p+1)/2$*, because* $\operatorname{vec}(\boldsymbol{\Sigma}) = \mathbf{D}_p \operatorname{vech}(\boldsymbol{\Sigma})$ *and* $\dim(\operatorname{vech} \boldsymbol{\Sigma}) = [p(p+1)/2] \times 1$*. Under a saturated model,*

$$\mathbf{P}_a = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a} \left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}\right]^{-1} \mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right),$$

can be simplified as $\mathbf{N}_p$*, which can be shown by the following four steps.*

step 1. *It can be concluded that* $\mathcal{R}(\mathbf{N}_p) = \mathcal{R}(\mathbf{D}_p)$ *and* $\mathcal{N}(\mathbf{N}_p) = \mathcal{N}(\mathbf{D}'_p)$ *because* $\mathbf{N}_p = \operatorname{ppo}(\mathbf{D}_p)$ *that projects onto* $\mathcal{R}(\mathbf{D}_p)$ *along* $\mathcal{N}(\mathbf{D}'_p)$*, where* $\mathcal{R}(\cdot)$ *and* $\mathcal{N}(\cdot)$ *are defined in Table 57.*

step 2. *Show that* $\mathcal{R}(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}) = \mathcal{R}(\mathbf{D}_p)$*.*

It can be concluded that $\mathcal{R}(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}) \subseteq \mathcal{R}(\mathbf{N}_p)$ *because* $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a} = \mathbf{N}_p \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}$*. Furthermore,* $\mathcal{R}(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}) = \mathcal{R}(\mathbf{N}_p)$ *because* $\operatorname{rank}(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}) = \nu_a = p(p+1)/2 = \operatorname{rank}(\mathbf{N}_p)$*. Thus,* $\mathcal{R}(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}) = \mathcal{R}(\mathbf{D}_p)$ *by* $\mathcal{R}(\mathbf{N}_p) = \mathcal{R}(\mathbf{D}_p)$*.*

step 3. *Show that* $\mathcal{N}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\right) = \mathcal{N}(\mathbf{D}'_p)$*.*

For any vector $\mathbf{u} \in \mathcal{N}(\mathbf{N}_p)$*,* $\mathbf{N}_p\mathbf{u} = \mathbf{0}$*. Thus,* $\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{N}_p \mathbf{u} = \mathbf{0}$*. Note that* $\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{N}_p \mathbf{u} = \mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{u}$*, because* $\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a} = \mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a} \mathbf{N}_p$ *and* $\mathbf{N}_p \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) = \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{N}_p$*. It can be concluded that* $\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \mathbf{u} = \mathbf{0}$*. Thus,* $\mathcal{N}(\mathbf{N}_p) \subseteq \mathcal{N}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\right)$*. Furthermore,* $\mathcal{N}(\mathbf{N}_p) = \mathcal{N}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\right)$ *because rank*$(\mathbf{N}_p) = p(p+$

$1)/2 = \operatorname{rank}(\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}) = \operatorname{rank}(\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}))$. *Therefore,* $\mathcal{N}(\mathbf{D}'_p) = \mathcal{N}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\right)$ *because* $\mathcal{N}(\mathbf{N}_p) = \mathcal{N}(\mathbf{D}'_p)$.

step 4. Show that $\mathbf{P}_a = \mathbf{N}_p$.

Note that $\mathbf{P}_a$ *projects onto* $\mathcal{R}(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a})$ *along* $\mathcal{N}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\right)$. *It can be concluded that* $\mathbf{P}_a$ *projects onto* $\mathcal{R}(\mathbf{D}_p)$ *along* $\mathcal{N}(\mathbf{D}'_p)$, *because* $\mathcal{R}(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}) = \mathcal{R}(\mathbf{D}_p)$ *and* $\mathcal{N}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\right) = \mathcal{N}(\mathbf{D}'_p)$. *Furthermore,* $\mathbf{P}_a = \mathbf{N}_p$ *because* $\mathbf{N}_p = \operatorname{ppo}(\mathbf{D}_p)$ *is unique.*

The following proof is divided into two parts. The first part is to derive the expression of the test statistic X^2*; and the second part is to derive the asymptotic null distribution of* $\mathbf{X}^2$.

Part I Provided that the model for $\boldsymbol{\Sigma}$ *is saturated, set* $\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}} = \mathbf{0}$ *to obtain the minimum discrepancy estimator of* $\boldsymbol{\Sigma}$ *in* $F(\boldsymbol{\Sigma}, \mathbf{S})$, $\widehat{\boldsymbol{\Sigma}}$, *where* $\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}}$ *is defined in Theorem 17. That is,*

$$\mathbf{D}^{(1)}_{F;\boldsymbol{\theta}_a} = -\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}(\mathbf{S} - \boldsymbol{\Sigma}) = \mathbf{0}.$$

This means that $\operatorname{vec}(\mathbf{S} - \boldsymbol{\Sigma}) \in \mathcal{N}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\right)$. *By the result shown in step 3. that* $\mathcal{N}(\mathbf{N}_p) = \mathcal{N}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})\right)$, *it can be concluded that* $\operatorname{vec}(\mathbf{S} - \boldsymbol{\Sigma}) \in \mathcal{N}(\mathbf{N}_p)$. *Therefore,* $\mathbf{N}_p \operatorname{vec}(\mathbf{S} - \boldsymbol{\Sigma}) = \mathbf{0}$. *Furthermore,* $\operatorname{vec}(\mathbf{S} - \boldsymbol{\Sigma}) = \mathbf{0}$ *because* $\mathbf{N}_p \operatorname{vec}(\mathbf{S} - \boldsymbol{\Sigma}) = \operatorname{vec}(\mathbf{S} - \boldsymbol{\Sigma})$. *Thus, if the model for* $\boldsymbol{\Sigma}$ *is saturated, then* $\widehat{\boldsymbol{\Sigma}} = \mathbf{S}$.

Accordingly, the discrepancy test statistic X^2 *defined in (153) is*

$$X^2 = n\left[F(\hat{\boldsymbol{\Sigma}}_0, \mathbf{S}) - F(\mathbf{S}, \mathbf{S})\right] = nF(\hat{\boldsymbol{\Sigma}}_0, \mathbf{S}),$$

because $F(\mathbf{S}, \mathbf{S}) = 0$. *Based on the definition of* X^2, $\mathbf{H}_0$ *is rejected for large values of* X^2.

Part II Based on part (4.) in Theorem 33, given $\mathbf{H}_0$ *is true,*

$\mathbf{X}^2 = n\left[F(\hat{\mathbf{\Sigma}}_0, \mathbf{S}) - F(\hat{\mathbf{\Sigma}}_a, \mathbf{S})\right]$ *can be written as*

$$\mathbf{X}^2 = \frac{1}{2}\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma})'\left(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1}\right)(\mathbf{I}_{p^2}-\mathbf{P}_0)\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) + O_p(n^{-\frac{1}{2}}),$$

because $\mathbf{P}_a = \mathbf{N}_p$, $(\mathbf{s}-\boldsymbol{\sigma})'\mathbf{N}_p = (\mathbf{s}-\boldsymbol{\sigma})'$,

and $(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\,\mathbf{N}_p = \mathbf{N}_p\,(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})$.

Accordingly, the asymptotic null distribution of $\mathbf{X}^2$ *is* $\mathbf{X}^2 \xrightarrow{\text{dist}} \sum_{i=1}^{\nu_d} v_i w_i^2$, *where* $\nu_d = p(p+1)/2 - \nu_0$, *and* $w_i^2 \overset{\text{iid}}{\sim} \chi_1^2$ *for* $i = 1, \ldots, \nu_d$. □

B.33.2. Proof of corollary 33.2

Corollary 33.2. [Adapted from Boik, Panishkan, and Hyde [35], 2010, Page 131]. *Provided that* $v_i \approx v$ *for all* i*, the Satterthwaite approximation [51] to the null distribution of* X^2 *is*

$$\lim_{n\to\infty} cX^2 \dot\sim \chi_f^2,$$

where $c = \operatorname{tr}(\mathbf{J})/\operatorname{tr}(\mathbf{J}^2)$, $\mathbf{X}^2 = n[F(\hat{\mathbf{\Sigma}}_0, \mathbf{S}) - F(\hat{\mathbf{\Sigma}}_a, \mathbf{S})]$, $f = [\operatorname{tr}(\mathbf{J})]^2/\operatorname{tr}(\mathbf{J}^2)$, *and* $\mathbf{J} = \frac{1}{2}\mathbf{\Omega}\,(\mathbf{\Sigma}^{-1}\otimes\mathbf{\Sigma}^{-1})\,(\mathbf{P}_a - \mathbf{P}_0)$. *Reject* $\mathbf{H}_0$ *for large values of* cX^2.

Proof: [Proof With Advisor's Help]. *By part (5) in Theorem 33, given* $\mathbf{H}_0$ *is true,* $\mathbf{X}^2 \xrightarrow{\text{dist}} \sum_{i=1}^{\nu_d} v_i w_i^2$, *where* $w_i^2 \overset{\text{iid}}{\sim} \chi_1^2$ *for* $i = 1, \ldots, \nu_d$.

Given that $v_i \approx v$ *for all* i*, suppose that there exist* c *and* f *such that* $\lim_{n\to\infty} cX^2 \dot\sim \chi_f^2$*, then it can be concluded that*

$$\lim_{n\to\infty} \operatorname{E}(cX^2) = \operatorname{E}\left(c\sum_{i=1}^{\nu_d} v_i w_i^2\right) = f, \quad \text{and}$$

$$\lim_{n\to\infty} \operatorname{Var}(cX^2) = \operatorname{Var}\left(c\sum_{i=1}^{\nu_d} v_i w_i^2\right) = 2f. \tag{229}$$

To solve c *and* f*, (229) can be expanded as follows:*

$$f = \operatorname{E}\left(c\sum_{i=1}^{\nu_d} v_i w_i^2\right) = c\sum_{i=1}^{\nu_d} v_i \operatorname{E}\left(w_i^2\right) = c\sum_{i=1}^{\nu_d} v_i * 1 = c\operatorname{tr}(\mathbf{J}),$$

$$2f \;=\; \operatorname{Var}\left(c\sum_{i=1}^{\nu_d} v_i w_i^2\right) = c^2\left(\sum_{i=1}^{\nu_d} v_i^2 * 2\right) = 2c^2\operatorname{tr}(\mathbf{J}^2), \tag{230}$$

because $w_i^2 \overset{\text{iid}}{\sim} \chi_1^2$, $\sum_{i=1}^{\nu_d} v_i = \operatorname{tr}(J)$, *and* $\sum_{i=1}^{\nu_d} v_i^2 = \operatorname{tr}(J^2)$. *From (230), it can be concluded that* $2c\operatorname{tr}(\mathbf{J}) = 2c^2\operatorname{tr}(\mathbf{J}^2)$, *furthermore,* $c\,[c\operatorname{tr}(\mathbf{J}^2) - \operatorname{tr}(\mathbf{J})] = 0$. *Thus,* $c = \operatorname{tr}(\mathbf{J})/\operatorname{tr}(\mathbf{J}^2)$ *and* $f = [\operatorname{tr}(\mathbf{J})]^2/\operatorname{tr}(\mathbf{J}^2)$, *because* $f = c\operatorname{tr}(\mathbf{J})$ *and* $f > 0$. $\square$

B.34. Proof of Theorem 34

Theorem 34. [Adapted from Boik, Panishkan, and Hyde [35], 2010, Page 131]. *Consider a test of* $\mathbf{H}_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ *against* $\mathbf{H}_a : \boldsymbol{\theta} \in \boldsymbol{\Theta}_a$, *where* $\boldsymbol{\Theta}_0 \cap \boldsymbol{\Theta}_a = \emptyset$. *Suppose that under* $\mathbf{H}_0$, *constraints are placed on* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, *where* $\boldsymbol{\Theta} = \boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_a$. *If* $\mathbf{H}_0$ *is true, then, by construction, there exist a value* $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}_0$ *and a value* $\boldsymbol{\theta}_a \in \boldsymbol{\Theta}$ *so that* $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_a) = \boldsymbol{\Sigma}$ *and* $\boldsymbol{\theta}_a = g(\boldsymbol{\theta}_0)$.

Define $\mathbf{P}_{B,0}$, $\mathbf{P}_{B,a}$, $\widehat{\mathbf{P}}_{B,0}$, $\widehat{\mathbf{P}}_{B,a}$, *and* $SSE(\hat{\boldsymbol{\theta}}_0)$ *as follows:*

$$\begin{aligned}
&\mathbf{P}_{B,0} = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'}\boldsymbol{\Omega}_n^+\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'}\right)^{-1}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'}\boldsymbol{\Omega}_n^+,\\
&\mathbf{P}_{B,a} = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\boldsymbol{\Omega}_n^+\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\right)^{-1}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\boldsymbol{\Omega}_n^+,\\
&\widehat{\mathbf{P}}_{B,0} = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}_0'}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}_0'}\widehat{\boldsymbol{\Omega}}_n^+\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}_0'}\right)^{-1}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}_0'}\widehat{\boldsymbol{\Omega}}_n^+,\\
&\widehat{\mathbf{P}}_{B,a} = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}_a'}\left(\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}_a'}\widehat{\boldsymbol{\Omega}}_n^+\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}_a'}\right)^{-1}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}_a'}\widehat{\boldsymbol{\Omega}}_n^+,\\
&SSE(\hat{\boldsymbol{\theta}}_0) = n(\mathbf{s} - \widehat{\boldsymbol{\sigma}}_0)'\widehat{\boldsymbol{\Omega}}_n^+(\mathbf{I}_{p^2} - \widehat{\mathbf{P}}_{B,0})(\mathbf{s} - \widehat{\boldsymbol{\sigma}}_0), \quad \textit{and}\\
&SSE(\hat{\boldsymbol{\theta}}_a) = n(\mathbf{s} - \widehat{\boldsymbol{\sigma}}_a)'\widehat{\boldsymbol{\Omega}}_n^+(\mathbf{I}_{p^2} - \widehat{\mathbf{P}}_{B,a})(\mathbf{s} - \widehat{\boldsymbol{\sigma}}_a),
\end{aligned}$$

where $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'} = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$, $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'} = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_a}$, $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}_0'} = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0}$, $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\hat{\boldsymbol{\theta}}_a'} = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_a}$, $\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'}$ *is given in (48),* $\boldsymbol{\Omega}_n^+$ *and* $\widehat{\boldsymbol{\Omega}}_n^+$ *are the Moore-Penrose generalized inverses of* $\boldsymbol{\Omega}_n$ *and* $\widehat{\boldsymbol{\Omega}}_n$, *respectively,* $\boldsymbol{\Omega}_n$ *is defined in (141),* $\widehat{\boldsymbol{\Omega}}_n$ *is defined in (142),* $\widehat{\boldsymbol{\sigma}}_0 = \operatorname{vec}\hat{\boldsymbol{\Sigma}}_0$, $\widehat{\boldsymbol{\sigma}}_a = \operatorname{vec}\hat{\boldsymbol{\Sigma}}_a$, *and* $\hat{\boldsymbol{\Sigma}}_0$ *and* $\hat{\boldsymbol{\Sigma}}_a$ *are defined in (153). The Moore-*

Penrose generalized inverses of $\mathbf{\Omega}_n$ *and* $\widehat{\mathbf{\Omega}}_n$ *can be written as* $\mathbf{D}_p(\mathbf{D}_p'\mathbf{\Omega}_n\mathbf{D}_p)^{-1}\mathbf{D}_p'$ *and* $\mathbf{D}_p(\mathbf{D}_p'\widehat{\mathbf{\Omega}}_n\mathbf{D}_p)^{-1}\mathbf{D}_p'$*, respectively.*

Browne's residual-based statistic is

$$SSE_B = SSE(\hat{\boldsymbol{\theta}}_0) - SSE(\hat{\boldsymbol{\theta}}_a),$$

where $SSE(\hat{\boldsymbol{\theta}}_0)$ *and* $SSE(\hat{\boldsymbol{\theta}}_a)$ *are defined in (158). Reject* $\mathbf{H}_0$*, for large values of* SSE_B*. Given* $\mathbf{H}_0$ *is true,* $SSE_B \xrightarrow{\text{dist}} \chi^2_{\nu_d}$*, where* $\nu_d = \nu_a - \nu_0$*,* ν_0 *and* ν_a *are defined in (153).*

Proof: [Proof With Advisor's Help]. *Let* $\mathbf{M} = \mathbf{D}_p(\mathbf{D}_p'\widehat{\mathbf{\Omega}}_n\mathbf{D}_p)^{-1}\mathbf{D}_p'$*. According to the definition of* $\widehat{\mathbf{\Omega}}_n$ *in (142), it is true that*

$$\begin{aligned}
\widehat{\mathbf{\Omega}}_n &= \widehat{\mathbf{\Omega}}_n\mathbf{N}_p = \mathbf{N}_p\widehat{\mathbf{\Omega}}_n \\
\mathbf{M}\widehat{\mathbf{\Omega}}_n &= \mathbf{D}_p(\mathbf{D}_p'\widehat{\mathbf{\Omega}}_n\mathbf{D}_p)^{-1}\mathbf{D}_p'\widehat{\mathbf{\Omega}}_n\mathbf{N}_p = \mathbf{D}_p(\mathbf{D}_p'\widehat{\mathbf{\Omega}}_n\mathbf{D}_p)^{-1}\mathbf{D}_p'\widehat{\mathbf{\Omega}}_n\mathbf{D}_p\mathbf{H}_p = \mathbf{N}_p, \text{ and} \\
\widehat{\mathbf{\Omega}}_n\mathbf{M} &= \mathbf{N}_p\widehat{\mathbf{\Omega}}_n\mathbf{D}_p(\mathbf{D}_p'\widehat{\mathbf{\Omega}}_n\mathbf{D}_p)^{-1}\mathbf{D}_p' = \mathbf{H}_p{}'\mathbf{D}_p'\widehat{\mathbf{\Omega}}_n\mathbf{D}_p(\mathbf{D}_p'\widehat{\mathbf{\Omega}}_n\mathbf{D}_p)^{-1}\mathbf{D}_p' = \mathbf{N}_p,
\end{aligned}$$

where $\mathbf{H}_p$ *is defined in (157) and* $\mathbf{N}_p = \mathbf{D}_p\mathbf{H}_p = \mathbf{H}_p{}'\mathbf{D}_p'$*. Furthermore, it can shown that* $\mathbf{M}$ *is the Moore-Penrose generalized inverses of* $\widehat{\mathbf{\Omega}}_n$ *as follows:*

(1.) $\mathbf{M}\widehat{\mathbf{\Omega}}_n\mathbf{M} = \mathbf{N}_p\mathbf{M} = \mathbf{N}_p\mathbf{D}_p(\mathbf{D}_p'\widehat{\mathbf{\Omega}}_n\mathbf{D}_p)^{-1}\mathbf{D}_p' = \mathbf{M}$,

(2.) $\widehat{\mathbf{\Omega}}_n\mathbf{M}\widehat{\mathbf{\Omega}}_n = \mathbf{N}_p\widehat{\mathbf{\Omega}}_n = \widehat{\mathbf{\Omega}}_n$,

(3.) $\left(\mathbf{M}\widehat{\mathbf{\Omega}}_n\right)' = \mathbf{N}_p = \mathbf{M}\widehat{\mathbf{\Omega}}_n$, *and*

(4.) $\left(\widehat{\mathbf{\Omega}}_n\mathbf{M}\right)' = \mathbf{N}_p = \widehat{\mathbf{\Omega}}_n\mathbf{M}$.

The proof for $\mathbf{D}_p(\mathbf{D}_p'\mathbf{\Omega}_n\mathbf{D}_p)^{-1}\mathbf{D}_p'$ *is the Moore-Penrose generalized inverses of* $\mathbf{\Omega}_n$ *is similar to the above, and therefore omitted.*

Given that $\mathbf{H}_0$ *is true, expand* $\sqrt{n}(\mathbf{s} - \widehat{\boldsymbol{\sigma}}_0)$ *in a Taylor series around* $\hat{\boldsymbol{\theta}}_0 = \boldsymbol{\theta}_0$ *as follows:*

$$\sqrt{n}(\mathbf{s} - \widehat{\boldsymbol{\sigma}}_0)$$

$$
\begin{aligned}
&= \sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) - \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'}\sqrt{n}(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta}_0) - \frac{\sqrt{n}}{2}(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta}_0)'\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_1}(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta}_0) \\
&\quad \textit{because } \boldsymbol{\sigma}(\boldsymbol{\theta}_0)=\boldsymbol{\sigma}(\boldsymbol{\theta}_a)=\boldsymbol{\sigma} \textit{ given that } \mathbf{H}_0 \textit{ is true} \\
&= \sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) \\
&- \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'}\left[\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'}\right]^{-1}\mathbf{D}^{(1)\prime}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'}\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) + O_p(n^{-\frac{1}{2}}) \\
&\quad \textit{because of (220)} \\
&- \frac{\sqrt{n}}{2}(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta}_0)'\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_1}(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta}_0) \\
&= \sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) - \mathbf{P}_0\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) - \frac{\sqrt{n}}{2}(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta}_0)'\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_1}(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta}_0) + O_p(n^{-\frac{1}{2}}) \\
&= \left(\mathbf{I}_{p^2}-\mathbf{P}_0\right)\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) - \frac{1}{2\sqrt{n}}\sqrt{n}(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta}_0)'\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_1}\sqrt{n}(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta}_0) + O_p(n^{-\frac{1}{2}}) \\
&= \left(\mathbf{I}_{p^2}-\mathbf{P}_0\right)\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) + O_p(n^{-\frac{1}{2}}) \quad \textit{because } \sqrt{n}(\hat{\boldsymbol{\theta}}_0-\boldsymbol{\theta}_0)=O_p(1),
\end{aligned}
$$

where $\mathbf{D}^{(2)}_{\boldsymbol{\sigma};\boldsymbol{\theta},\boldsymbol{\theta}'}$ *is defined in Theorem 17, and* $\boldsymbol{\theta}_1=\alpha\hat{\boldsymbol{\theta}}_0+(1-\alpha)\boldsymbol{\theta}_0$.

Provided that $\mathbf{H}_0$ *is true, the expansion of* $\sqrt{n}(\mathbf{s}-\widehat{\boldsymbol{\sigma}}_a)$ *can be conducted in a similar manner as shown above and the expression for* $\sqrt{n}(\mathbf{s}-\widehat{\boldsymbol{\sigma}}_a)$ *have the same form as* $\sqrt{n}(\mathbf{s}-\widehat{\boldsymbol{\sigma}}_0)$*, except that* $\mathbf{P}_a$ *is substituted for* $\mathbf{P}_0$*. That is,*

$$
\sqrt{n}(\mathbf{s}-\widehat{\boldsymbol{\sigma}}_a) = \left(\mathbf{I}_{p^2}-\mathbf{P}_a\right)\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) + O_p(n^{-\frac{1}{2}}). \tag{231}
$$

Define $\mathbf{B}_0$ *and* $\mathbf{B}_a$ *as*

$$
\begin{aligned}
\mathbf{B}_0 &= \left(\mathbf{I}_{p^2}-\mathbf{P}_0\right)'\boldsymbol{\Omega}_n^+\left(\mathbf{I}_{p^2}-\mathbf{P}_{B,0}\right)\left(\mathbf{I}_{p^2}-\mathbf{P}_0\right), \quad \textit{and} \\
\mathbf{B}_a &= \left(\mathbf{I}_{p^2}-\mathbf{P}_a\right)'\boldsymbol{\Omega}_n^+\left(\mathbf{I}_{p^2}-\mathbf{P}_{B,a}\right)\left(\mathbf{I}_{p^2}-\mathbf{P}_a\right), \quad \textit{respectively.}
\end{aligned}
$$

Provided that $\mathbf{H}_0$ *is true, by Slutsky's Theorem 23,*

$$
\begin{aligned}
SSE(\hat{\boldsymbol{\theta}}_0) &\xrightarrow{\text{dist}} n(\mathbf{s}-\boldsymbol{\sigma})'\left(\mathbf{I}_{p^2}-\mathbf{P}_0\right)'\boldsymbol{\Omega}_n^+\left(\mathbf{I}_{p^2}-\mathbf{P}_{B,0}\right)\left(\mathbf{I}_{p^2}-\mathbf{P}_0\right)(\mathbf{s}-\boldsymbol{\sigma}) \\
&= n(\mathbf{s}-\boldsymbol{\sigma})'\mathbf{B}_0(\mathbf{s}-\boldsymbol{\sigma}), \\
SSE(\hat{\boldsymbol{\theta}}_a) &\xrightarrow{\text{dist}} n(\mathbf{s}-\boldsymbol{\sigma})'\left(\mathbf{I}_{p^2}-\mathbf{P}_a\right)'\boldsymbol{\Omega}_n^+\left(\mathbf{I}_{p^2}-\mathbf{P}_{B,a}\right)\left(\mathbf{I}_{p^2}-\mathbf{P}_0\right)(\mathbf{s}-\boldsymbol{\sigma}) \qquad (232) \\
&= n(\mathbf{s}-\boldsymbol{\sigma})'\mathbf{B}_a(\mathbf{s}-\boldsymbol{\sigma}), \quad \textit{and} \\
SSE_B = SSE(\hat{\boldsymbol{\theta}}_0) - SSE(\hat{\boldsymbol{\theta}}_a) &\xrightarrow{\text{dist}} n(\mathbf{s}-\boldsymbol{\sigma})'\left(\mathbf{B}_0-\mathbf{B}_a\right)(\mathbf{s}-\boldsymbol{\sigma}),
\end{aligned}
$$

where $SSE(\hat{\boldsymbol{\theta}}_0)$ *and* $SSE(\hat{\boldsymbol{\theta}}_a)$ *are defined in (158),* $\widehat{\mathbf{P}}_{B,0} \xrightarrow{\text{dist}} \mathbf{P}_{B,0}$ *and* $\widehat{\mathbf{P}}_{B,a} \xrightarrow{\text{dist}} \mathbf{P}_{B,a}$ *because* $\widehat{\boldsymbol{\Omega}}_n$ *is a consistent estimator of* $\boldsymbol{\Omega}_n$*.*

To derive the asymptotic null distribution of SSE_B*, simplifications of* $\mathbf{B}_0$ *and* $\mathbf{B}_a$ *is needed. To simplify* $\mathbf{B}_0$ *and* $\mathbf{B}_a$*, it helps to realize that*

$$
\begin{aligned}
&\mathbf{P}_{B,0}^2 = \mathbf{P}_{B,0}, \quad \mathbf{P}_{B,a}^2 = \mathbf{P}_{B,a}, \quad \mathbf{P}_{B,0}\mathbf{P}_0 = \mathbf{P}_0, \quad \mathbf{P}_{B,a}\mathbf{P}_a = \mathbf{P}_a, \\
&\boldsymbol{\Omega}_n^+\mathbf{P}_{B,0} = \mathbf{P}_{B,0}'\boldsymbol{\Omega}_n^+, \quad \boldsymbol{\Omega}_n^+\mathbf{P}_{B,a} = \mathbf{P}_{B,a}'\boldsymbol{\Omega}_n^+, \quad \mathbf{P}_{B,a}\mathbf{P}_{B,0} = \mathbf{P}_{B,0}, \qquad (233) \\
&\textit{and } \mathbf{P}_{B,0}\mathbf{P}_{B,a} = \mathbf{P}_{B,0}, \ \textit{because } \boldsymbol{\theta}_a = g(\boldsymbol{\theta}_0) \ \textit{and } \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_0'} = \mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}_a'}\left(\frac{\partial\,\boldsymbol{\theta}_a}{\partial\,\boldsymbol{\theta}_0'}\right),
\end{aligned}
$$

where $\mathbf{P}_0$ *and* $\mathbf{P}_a$ *are defined in (157). Furthermore, it can be shown that*

$$
\begin{aligned}
\left(\mathbf{P}_{B,a}-\mathbf{P}_{B,0}\right)^2 &= \mathbf{P}_{B,a}-\mathbf{P}_{B,0}, \\
\left[\boldsymbol{\Omega}_n^+\left(\mathbf{P}_{B,a}-\mathbf{P}_{B,0}\right)\boldsymbol{\Omega}\right]^2 &= \boldsymbol{\Omega}_n^+\left(\mathbf{P}_{B,a}-\mathbf{P}_{B,0}\right)\boldsymbol{\Omega} \qquad (234) \\
\textit{and } \operatorname{tr}\left[\boldsymbol{\Omega}_n^+\left(\mathbf{P}_{B,a}-\mathbf{P}_{B,0}\right)\boldsymbol{\Omega}\right] &= \nu_a - \nu_0 = \nu_d,
\end{aligned}
$$

as follows:

$$
\begin{aligned}
\left(\mathbf{P}_{B,a}-\mathbf{P}_{B,0}\right)^2 &= \mathbf{P}_{B,a}^2 - \mathbf{P}_{B,a}\mathbf{P}_{B,0} - \mathbf{P}_{B,0}\mathbf{P}_{B,a} + \mathbf{P}_{B,0}^2 \\
&= \mathbf{P}_{B,a} - \mathbf{P}_{B,0} - \mathbf{P}_{B,0} + \mathbf{P}_{B,0} \\
&= \mathbf{P}_{B,a} - \mathbf{P}_{B,0} \\
\left[\boldsymbol{\Omega}_n^+\left(\mathbf{P}_{B,a}-\mathbf{P}_{B,0}\right)\boldsymbol{\Omega}\right]^2 &= \boldsymbol{\Omega}_n^+\left(\mathbf{P}_{B,a}-\mathbf{P}_{B,0}\right)\boldsymbol{\Omega}\boldsymbol{\Omega}_n^+\left(\mathbf{P}_{B,a}-\mathbf{P}_{B,0}\right)\boldsymbol{\Omega}
\end{aligned}
$$

$$
\begin{aligned}
&= (\mathbf{P}_{B,a} - \mathbf{P}_{B,0})' \, \boldsymbol{\Omega}_n^+ \boldsymbol{\Omega} \boldsymbol{\Omega}_n^+ (\mathbf{P}_{B,a} - \mathbf{P}_{B,0}) \, \boldsymbol{\Omega} \\
&\quad \textit{because } \boldsymbol{\Omega}_n^+ \mathbf{P}_{B,a} = \mathbf{P}'_{B,a} \boldsymbol{\Omega}_n^+ \textit{ and } \boldsymbol{\Omega}_n^+ \mathbf{P}_{B,0} = \mathbf{P}'_{B,0} \boldsymbol{\Omega}_n^+ \\
&= (\mathbf{P}_{B,a} - \mathbf{P}_{B,0})' \, \boldsymbol{\Omega}_n^+ (\mathbf{P}_{B,a} - \mathbf{P}_{B,0}) \, \boldsymbol{\Omega} \\
&= \boldsymbol{\Omega}_n^+ (\mathbf{P}_{B,a} - \mathbf{P}_{B,0}) (\mathbf{P}_{B,a} - \mathbf{P}_{B,0}) \, \boldsymbol{\Omega} \\
&= \boldsymbol{\Omega}_n^+ (\mathbf{P}_{B,a} - \mathbf{P}_{B,0}) \, \boldsymbol{\Omega} \\
&\quad \textit{because } (\mathbf{P}_{B,a} - \mathbf{P}_{B,0})^2 = \mathbf{P}_{B,a} - \mathbf{P}_{B,0}
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{tr}\left[\boldsymbol{\Omega}_n^+ (\mathbf{P}_{B,a} - \mathbf{P}_{B,0}) \, \boldsymbol{\Omega}\right] &= \operatorname{tr}\left[(\mathbf{P}_{B,a} - \mathbf{P}_{B,0}) \, \boldsymbol{\Omega} \boldsymbol{\Omega}_n^+\right] \\
&= \operatorname{tr}(\mathbf{P}_{B,a} - \mathbf{P}_{B,0}) \\
&\quad \textit{by the definitions of } \mathbf{P}_{B,a} \textit{ and } \mathbf{P}_{B,0} \\
&= \operatorname{tr}(\mathbf{P}_{B,a}) - \operatorname{tr}(\mathbf{P}_{B,0}) \\
&= \operatorname{rank}(\mathbf{P}_{B,a}) - \operatorname{rank}(\mathbf{P}_{B,0}) \\
&\quad \textit{because } \mathbf{P}_{B,a}^2 = \mathbf{P}_{B,a} \textit{ and } \mathbf{P}_{B,0}^2 = \mathbf{P}_{B,0} \\
&= \nu_a - \nu_0 \\
&= \nu_d,
\end{aligned}
$$

because $\nu_a = \operatorname{rank}(\mathbf{P}_a) = \operatorname{rank}(\mathbf{P}_{B,a}\mathbf{P}_a) \le \operatorname{rank}(\mathbf{P}_{B,a}) \le \operatorname{rank}(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_a}) = \nu_a$ *and* $\nu_0 = \operatorname{rank}(\mathbf{P}_0) = \operatorname{rank}(\mathbf{P}_{B,0}\mathbf{P}_0) \le \operatorname{rank}(\mathbf{P}_{B,0}) \le \operatorname{rank}(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'_0}) = \nu_0$.

Simplifications of $\mathbf{B}_0$ *can be obtained by using the above results through examining*

$$
\begin{aligned}
\mathbf{B}_0 &= (\mathbf{I}_{p^2} - \mathbf{P}_0)' \, \boldsymbol{\Omega}_n^+ (\mathbf{I}_{p^2} - \mathbf{P}_{B,0}) (\mathbf{I}_{p^2} - \mathbf{P}_0) \\
&= \boldsymbol{\Omega}_n^+ (\mathbf{I}_{p^2} - \mathbf{P}_{B,0}) + \left[\boldsymbol{\Omega}_n^+ \mathbf{P}_{B,0} \mathbf{P}_0 - \boldsymbol{\Omega}_n^+ \mathbf{P}_0\right] \\
&\quad + \left[\mathbf{P}_0' \boldsymbol{\Omega}_n^+ \mathbf{P}_{B,0} - \mathbf{P}_0' \boldsymbol{\Omega}_n^+\right] + \left[\mathbf{P}_0' \boldsymbol{\Omega}_n^+ \mathbf{P}_0 - \mathbf{P}_0' \boldsymbol{\Omega}_n^+ \mathbf{P}_{B,0} \mathbf{P}_0\right] \\
&= \boldsymbol{\Omega}_n^+ (\mathbf{I}_{p^2} - \mathbf{P}_{B,0}) + \mathbf{0} + \mathbf{0} + \mathbf{0} \\
&= \boldsymbol{\Omega}_n^+ (\mathbf{I}_{p^2} - \mathbf{P}_{B,0}), \quad \textit{similarly,} \\
\mathbf{B}_a &= \boldsymbol{\Omega}_n^+ (\mathbf{I}_{p^2} - \mathbf{P}_{B,a}).
\end{aligned}
\tag{235}
$$

By (232) and (235), given $\mathbf{H}_0$ is true, it can be concluded that

$$SSE_B = SSE(\hat{\boldsymbol{\theta}}_0) - SSE(\hat{\boldsymbol{\theta}}_a) \xrightarrow{\text{dist}} n(\mathbf{s}-\boldsymbol{\sigma})'\boldsymbol{\Omega}_n^+ \left(\mathbf{P}_{B,a} - \mathbf{P}_{B,0}\right)(\mathbf{s}-\boldsymbol{\sigma}).$$

Furthermore, by Theorem 1.3.6 part (1) on page 10 of Christensen [52], Theorem 27, and (234), it is readily shown that the asymptotic null distribution of SSE_B is

$$SSE_B \xrightarrow{\text{dist}} \chi^2_{\nu_d}.$$

$\square$

B.35. Proof of Theorem 35

Theorem 35. [Original result]. *Let $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\boldsymbol{\theta}_{\boldsymbol{\lambda}})$, $\boldsymbol{\Phi} = \boldsymbol{\Phi}(\boldsymbol{\theta}_{\boldsymbol{\delta}}, \boldsymbol{\theta}_{\boldsymbol{\gamma}})$ and $\boldsymbol{\Psi} = \boldsymbol{\Psi}(\boldsymbol{\theta}_{\psi})$, where $\dim(\boldsymbol{\Lambda}) = p \times q$, $\dim(\boldsymbol{\Phi}) = q \times q$ and $\dim(\boldsymbol{\Psi}) = p \times p$. Let $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Psi}$ represent a $p \times p$ covariance matrix, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_\lambda' \quad \boldsymbol{\theta}_\delta' \quad \boldsymbol{\theta}_\gamma' \quad \boldsymbol{\theta}_\psi')'$ and $\dim(\boldsymbol{\theta}) = \nu$. Write $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ as $\boldsymbol{\Sigma}(\boldsymbol{\theta}) : \mathbb{R}^\nu \to \mathbb{R}^{p\times p}$. Define $\boldsymbol{\Theta}_1$ as $\boldsymbol{\Theta}_1 = \left\{\boldsymbol{\theta}\middle|\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Psi} \quad \text{and} \quad \operatorname{diag}(\boldsymbol{\Phi}(\boldsymbol{\theta}_{\boldsymbol{\delta}}, \boldsymbol{\theta}_{\boldsymbol{\gamma}})) = \mathbf{1}_q\right\}$ and $\boldsymbol{\Theta}_2$ as $\boldsymbol{\Theta}_2 = \left\{\boldsymbol{\theta}\middle|\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Psi} \quad \text{and} \quad \boldsymbol{\Phi}(\boldsymbol{\theta}_{\boldsymbol{\delta}}, \boldsymbol{\theta}_{\boldsymbol{\gamma}}) \text{ is a correlation matrix}\right\}$. Define $\boldsymbol{\Sigma}_1$ as $\boldsymbol{\Sigma}_1 : \boldsymbol{\Theta}_1 \to \mathbb{R}^{p\times p}$ and $\boldsymbol{\Sigma}_2$ as $\boldsymbol{\Sigma}_2 : \boldsymbol{\Theta}_2 \to \mathbb{R}^{p\times p}$. Let $\hat{\boldsymbol{\theta}}_1 = (\hat{\boldsymbol{\theta}}_{\lambda 1}' \quad \hat{\boldsymbol{\theta}}_{\delta 1}' \quad \hat{\boldsymbol{\theta}}_{\gamma 1}' \quad \hat{\boldsymbol{\theta}}_{\psi 1}')'$ be the minimizer of $F(\boldsymbol{\Sigma_1}(\boldsymbol{\theta}), \mathbf{S})$ and $\hat{\boldsymbol{\theta}}_2 = (\hat{\boldsymbol{\theta}}_{\lambda 2}' \quad \hat{\boldsymbol{\theta}}_{\delta 2}' \quad \hat{\boldsymbol{\theta}}_{\gamma 2}' \quad \hat{\boldsymbol{\theta}}_{\psi 2}')'$ be the minimizer of $F(\boldsymbol{\Sigma_2}(\boldsymbol{\theta}), \mathbf{S})$. If $\boldsymbol{\Phi}(\hat{\boldsymbol{\theta}}_{\delta 1}, \hat{\boldsymbol{\theta}}_{\gamma 1})$ is a correlation matrix, then $F(\boldsymbol{\Sigma_1}(\hat{\boldsymbol{\theta}}_{\mathbf{1}}), \mathbf{S}) = F(\boldsymbol{\Sigma_2}(\hat{\boldsymbol{\theta}}_{\mathbf{2}}), \mathbf{S})$.*

Proof: [Independent Proof]. *If $\boldsymbol{\Phi}(\hat{\boldsymbol{\theta}}_{\delta 1}, \hat{\boldsymbol{\theta}}_{\gamma 1})$ is a correlation matrix, then there exists $\hat{\boldsymbol{\theta}}_2^* \in \boldsymbol{\Theta}_2$ such that $\boldsymbol{\Sigma}_1(\hat{\boldsymbol{\theta}}_1) = \boldsymbol{\Sigma}_2(\hat{\boldsymbol{\theta}}_2^*)$. Accordingly, $F(\boldsymbol{\Sigma_1}(\hat{\boldsymbol{\theta}}_{\mathbf{1}}), \mathbf{S}) = F(\boldsymbol{\Sigma_2}(\hat{\boldsymbol{\theta}}_{\mathbf{2}}^*), \mathbf{S})$. By the definition of $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$, $\boldsymbol{\Sigma}_2(\hat{\boldsymbol{\theta}}_2)$ is obtained under stricter constraints to minimize $F(\boldsymbol{\Sigma}, \mathbf{S})$ than $\boldsymbol{\Sigma}_1(\hat{\boldsymbol{\theta}}_1)$. Therefore, $F(\boldsymbol{\Sigma_1}(\hat{\boldsymbol{\theta}}_{\mathbf{1}}), \mathbf{S}) \leq F(\boldsymbol{\Sigma_2}(\hat{\boldsymbol{\theta}}_{\mathbf{2}}), \mathbf{S})$. It can be concluded that $F(\boldsymbol{\Sigma_1}(\hat{\boldsymbol{\theta}}_{\mathbf{1}}), \mathbf{S}) = F(\boldsymbol{\Sigma_2}(\hat{\boldsymbol{\theta}}_{\mathbf{2}}), \mathbf{S})$ because $\hat{\boldsymbol{\theta}}_2$ is the minimizer of the discrepancy function $F(\boldsymbol{\Sigma_2}(\boldsymbol{\theta}), \mathbf{S})$ and $F(\boldsymbol{\Sigma_2}(\hat{\boldsymbol{\theta}}_{\mathbf{2}}^*), \mathbf{S}) \leq F(\boldsymbol{\Sigma_2}(\hat{\boldsymbol{\theta}}_{\mathbf{2}}), \mathbf{S})$.* $\square$

Proof of Lemma 1

Lemma 1. *Let* $\mathbf{A}$ *be a* $r \times s$ *matrix and* $\mathbf{B}$ *be a* $s \times r$ *matrix. Some useful results are the following.*

(a.) [Searle [36], 1982, Theorem 2, Page 333] $\operatorname{tr}(\mathbf{AB}) = \operatorname{vec}'(\mathbf{A}') \operatorname{vec}(\mathbf{B})$.

(b.) [Searle [36], 1982, equation 39, Page 337] $\partial \ln(|\mathbf{A}|)/\partial y = \operatorname{tr}\left[\mathbf{A}^{-1}\left(\partial \mathbf{A}/\partial y\right)\right]$ *for symmetric and non-symmetric* $\mathbf{A}$*, where* $\mathbf{A}$ *is a* $r \times r$ *matrix.*

(c.) [A Useful Result] $\partial \operatorname{vec}\boldsymbol{\Sigma}^{-1}/\partial \boldsymbol{\theta}' = -\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \partial \operatorname{vec}\boldsymbol{\Sigma}/\partial \boldsymbol{\theta}'$.

(d.) [A Useful Result] $\partial\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)/\partial \boldsymbol{\theta}' =$
$-\left[\boldsymbol{\Sigma}^{-1} \otimes \operatorname{vec}'(\boldsymbol{\Sigma}^{-1}) \otimes \boldsymbol{\Sigma}^{-1}\right]\left(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathbf{I}_{p^2}\right)$
$-\mathbf{K}_{p,p}\left[\boldsymbol{\Sigma}^{-1} \otimes \operatorname{vec}'(\boldsymbol{\Sigma}^{-1}) \otimes \boldsymbol{\Sigma}^{-1}\right]\left(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathbf{K}_{p,p}\right)$.

(e.) [A Useful Result] $\partial \operatorname{vec}\boldsymbol{\Sigma}^{-1}/\partial \boldsymbol{\theta} = -\left(\mathbf{I}_\nu \otimes \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \partial \operatorname{vec}\boldsymbol{\Sigma}/\partial \boldsymbol{\theta}$.

(f.) [A Useful Result] $\mathbf{L}_{21,p}\mathbf{L}'_{21,p} = \mathbf{L}_{22,p}$ *and* $\mathbf{L}'_{21,p}\mathbf{L}_{21,p} = \mathbf{I}_p$*, where* $\mathbf{L}_{21,p}$ *is as Table 56.*

Proof: [Independent Proof].

(a.) *Partition* $\mathbf{A}$ *in terms of its rows and partition* $\mathbf{B}$ *in terms of its columns. That is,*

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \\ \vdots \\ \mathbf{a}'_r \end{pmatrix} \text{ and } \mathbf{B} = \begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_r \end{pmatrix},$$

where $\mathbf{a}_i$ *and* $\mathbf{b}_i$ *are* $s \times 1$ *vectors for* $i = 1, \ldots, r$*. Accordingly,* $\operatorname{tr}(\mathbf{AB}) = \sum_{i=1}^{r} \mathbf{a}'_i \mathbf{b}_i$.

It follows that $\operatorname{vec}'(\mathbf{A}')\operatorname{vec}(\mathbf{B}) = \sum_{i=1}^{r} \mathbf{a}_i'\mathbf{b}_i$ *because*

$$\operatorname{vec}(\mathbf{A}') = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_r \end{pmatrix}, \quad \operatorname{vec}'(\mathbf{A}') = \begin{pmatrix} \mathbf{a}_1' & \mathbf{a}_2' & \cdots & \mathbf{a}_r' \end{pmatrix} \text{ and } \operatorname{vec}(\mathbf{B}) = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_r \end{pmatrix}.$$

Therefore, $\operatorname{tr}(\mathbf{AB}) = \operatorname{vec}'(\mathbf{A}')\operatorname{vec}(\mathbf{B})$.

(b.) Searle [36] showed that the derivative of the log determinant of a matrix with respect to a scalar can be written as $\partial \ln(|\mathbf{A}|)/\partial y = \operatorname{tr}\left[\mathbf{A}^{-1}\left(\partial \mathbf{A}/\partial y\right)\right]$.

(c.) It is follows that

$$\frac{\partial \operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\right)}{\partial \boldsymbol{\theta}'} = \mathbf{0} = \left(\boldsymbol{\Sigma} \otimes \mathbf{I}_p\right)\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}^{-1}}{\partial \boldsymbol{\theta}'} + \left(\mathbf{I}_p \otimes \boldsymbol{\Sigma}^{-1}\right)\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\theta}'}$$

because $\operatorname{vec}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\right) = \operatorname{vec}\mathbf{I}_p = \left(\boldsymbol{\Sigma} \otimes \mathbf{I}_p\right)\operatorname{vec}\boldsymbol{\Sigma}^{-1} = \left(\mathbf{I}_p \otimes \boldsymbol{\Sigma}^{-1}\right)\operatorname{vec}\boldsymbol{\Sigma}$.

$$\Longrightarrow \frac{\partial \operatorname{vec}\boldsymbol{\Sigma}^{-1}}{\partial \boldsymbol{\theta}'} = -\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}}{\partial \boldsymbol{\theta}'}.$$

(d.) Examine

$$\begin{aligned} \left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) &= \left(\mathbf{I}_p \otimes \operatorname{vec}'\mathbf{I}_p \otimes \boldsymbol{\Sigma}^{-1}\right)\left(\operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_{p^2}\right) \\ &= \mathbf{K}_{p,p}\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathbf{K}_{p,p} \\ &= \mathbf{K}_{p,p}\left(\mathbf{I}_p \otimes \operatorname{vec}'\mathbf{I}_p \otimes \boldsymbol{\Sigma}^{-1}\right)\left(\operatorname{vec}\boldsymbol{\Sigma}^{-1} \otimes \mathbf{K}_{p,p}\right). \end{aligned}$$

Accordingly,

$$\begin{aligned} &\frac{\partial\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)}{\partial \boldsymbol{\theta}'} \\ &= \left(\mathbf{I}_p \otimes \operatorname{vec}'\mathbf{I}_p \otimes \boldsymbol{\Sigma}^{-1}\right)\left(\frac{\partial \operatorname{vec}\boldsymbol{\Sigma}^{-1}}{\partial \boldsymbol{\theta}'} \otimes \mathbf{I}_{p^2}\right) \end{aligned}$$

$$
\begin{aligned}
&+ \quad \mathbf{K}_{p,p}\left(\mathbf{I}_p \otimes \text{vec}'\mathbf{I}_p \otimes \boldsymbol{\Sigma}^{-1}\right)\left(\frac{\partial\,\text{vec}\,\boldsymbol{\Sigma}^{-1}}{\partial\,\boldsymbol{\theta}'} \otimes \mathbf{K}_{p,p}\right) \\
&= \quad -\left(\mathbf{I}_p \otimes \text{vec}'\mathbf{I}_p \otimes \boldsymbol{\Sigma}^{-1}\right)\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathbf{I}_{p^2}\right] \\
&- \quad \mathbf{K}_{p,p}\left(\mathbf{I}_p \otimes \text{vec}'\mathbf{I}_p \otimes \boldsymbol{\Sigma}^{-1}\right)\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathbf{K}_{p,p}\right] \\
&= \quad -\left(\mathbf{I}_p \otimes \text{vec}'\mathbf{I}_p \otimes \boldsymbol{\Sigma}^{-1}\right)\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \otimes \mathbf{I}_{p^2}\right]\left(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathbf{I}_{p^2}\right) \\
&- \quad \mathbf{K}_{p,p}\left(\mathbf{I}_p \otimes \text{vec}'\mathbf{I}_p \otimes \boldsymbol{\Sigma}^{-1}\right)\left[\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right) \otimes \mathbf{I}_{p^2}\right]\left(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathbf{K}_{p,p}\right) \\
&= \quad -\left(\mathbf{I}_p \otimes \text{vec}'\mathbf{I}_p \otimes \boldsymbol{\Sigma}^{-1}\right)\left[\boldsymbol{\Sigma}^{-1} \otimes \left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_p\right) \otimes \mathbf{I}_p\right]\left(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathbf{I}_{p^2}\right) \\
&- \quad \mathbf{K}_{p,p}\left(\mathbf{I}_p \otimes \text{vec}'\mathbf{I}_p \otimes \boldsymbol{\Sigma}^{-1}\right)\left[\boldsymbol{\Sigma}^{-1} \otimes \left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_p\right) \otimes \mathbf{I}_p\right]\left(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathbf{K}_{p,p}\right) \\
&= \quad -\left[\boldsymbol{\Sigma}^{-1} \otimes \text{vec}'(\boldsymbol{\Sigma}^{-1}) \otimes \boldsymbol{\Sigma}^{-1}\right]\left(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathbf{I}_{p^2}\right) \\
&- \quad \mathbf{K}_{p,p}\left[\boldsymbol{\Sigma}^{-1} \otimes \text{vec}'(\boldsymbol{\Sigma}^{-1}) \otimes \boldsymbol{\Sigma}^{-1}\right]\left(\mathbf{D}^{(1)}_{\boldsymbol{\sigma};\boldsymbol{\theta}'} \otimes \mathbf{K}_{p,p}\right).
\end{aligned}
$$

(e.) It is follows that

$$\frac{\partial\,\text{vec}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma})}{\partial\,\boldsymbol{\theta}} = \mathbf{0} = (\mathbf{I}_\nu \otimes \boldsymbol{\Sigma} \otimes \mathbf{I}_p)\,\frac{\partial \text{vec}\boldsymbol{\Sigma}^{-1}}{\partial\,\boldsymbol{\theta}} + \left(\mathbf{I}_\nu \otimes \mathbf{I}_p \otimes \boldsymbol{\Sigma}^{-1}\right)\frac{\partial \text{vec}\boldsymbol{\Sigma}}{\partial\,\boldsymbol{\theta}}$$

because $\text{vec}\,(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}) = \text{vec}\,\mathbf{I}_p = (\boldsymbol{\Sigma} \otimes \mathbf{I}_p)\,\text{vec}\,\boldsymbol{\Sigma}^{-1} = (\mathbf{I}_p \otimes \boldsymbol{\Sigma}^{-1})\,\text{vec}\,\boldsymbol{\Sigma}$.

$$\Longrightarrow \frac{\partial \text{vec}\boldsymbol{\Sigma}^{-1}}{\partial\,\boldsymbol{\theta}} = -\left(\mathbf{I}_\nu \otimes \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\frac{\partial \text{vec}\boldsymbol{\Sigma}}{\partial\,\boldsymbol{\theta}}.$$

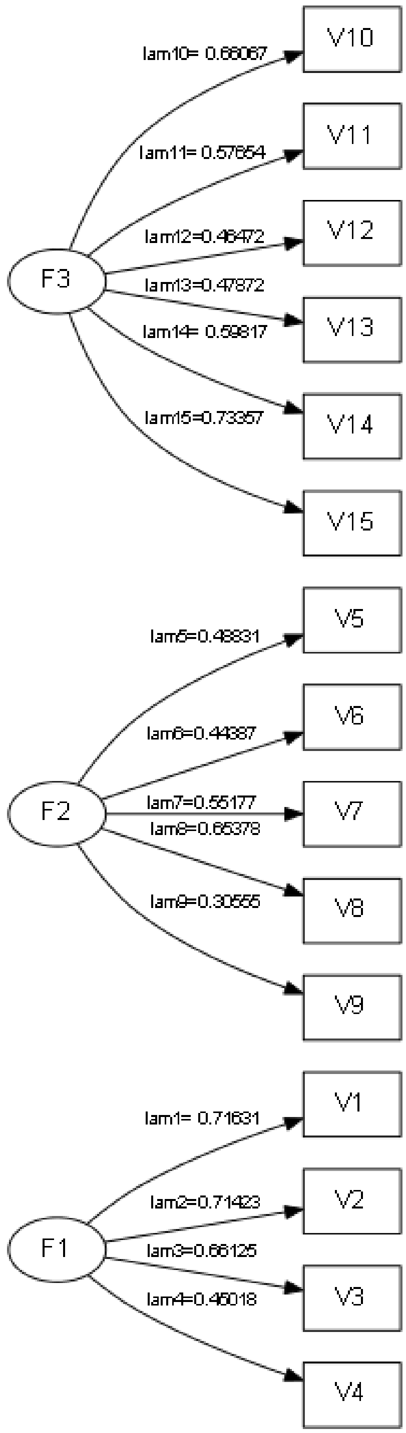
(f.) Finally,

$$
\begin{aligned}
\mathbf{L}_{21,p}\mathbf{L}'_{21,p} &= \sum_{j=1}^{p}\left(\mathbf{e}^p_j \otimes \mathbf{e}^p_j\right)\mathbf{e}^{p\prime}_j \sum_{i=1}^{p}\mathbf{e}^p_i\left(\mathbf{e}^{p\prime}_i \otimes \mathbf{e}^{p\prime}_i\right) \\
&= \sum_{j=1}^{p}\left(\mathbf{e}^p_j \otimes \mathbf{e}^p_j\right)\left(\mathbf{e}^{p\prime}_j \otimes \mathbf{e}^{p\prime}_j\right) \\
&= \mathbf{L}_{22,p}. \\
\mathbf{L}'_{21,p}\mathbf{L}_{21,p} &= \sum_{i=1}^{p}\mathbf{e}^p_i\left(\mathbf{e}^{p\prime}_i \otimes \mathbf{e}^{p\prime}_i\right)\sum_{j=1}^{p}\left(\mathbf{e}^p_j \otimes \mathbf{e}^p_j\right)\mathbf{e}^{p\prime}_j \\
&= \sum_{j=1}^{p}\mathbf{e}^p_j\mathbf{e}^{p\prime}_j = \mathbf{I}_p.
\end{aligned}
$$

$\square$

APPENDIX C

DIAGRAM FOR LEVY'S DATA


lam10= 0.66067
V10
lam11= 0.57654
V11
lam12=0.46472
V12
F3
lam13=0.47872
V13
lam14= 0.59817
V14
lam15=0.73357
V15
lam5=0.48831
V5
lam6=0.44387
V6
F2
lam7=0.55177
V7
lam8=0.65378
V8
lam9=0.30555
V9
lam1= 0.71631
V1
lam2=0.71423
V2
F1
lam3=0.66125
V3
lam4=0.45018
V4